Instructors' Guide for

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. *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.*

Contents

PREFACE

As instructors should appreciate, this module covers the three scientific areas that provide evidence bearing on causation in the "toxic tort" or environmental disease context. Those three areas are epidemiology, toxicology, and the still rapidly developing field of genetics. ***We envision that the module would be distributed to students.***

To cover all of the materials in the module comprehensively would take roughly 24 hours. The Introduction to Causation, Epidemiology and the basics of Specific Causation (omitting the sections on genetics) would be approximately 10–12 hours. The materials on Toxicology could consume 6–8 hours. The materials on genetics (Sections V. F. 2 and 3 and V. G.) would require in the range of 6–9 hours.

We realize that instructors who are not teaching these materials as a stand-alone course will not have that amount of time in which to introduce students to these sciences. For instructors who want to distill the materials in the module down to approximately eight hours, we would suggest covering Sections I and II (1 hour), and the following subsections in Sections III (3 hours), IV (2 to 3 hours), and V (2 hours).

Section III

A
B. 1. 2. a. and b.
D.
F. (introductory materials) and 2 and 3; and
G.

Section IV

A.

Section V

F.2 and F.3 *or* F.2 and G *or* F.3 and G.

Some instructors seeking to cut down coverage may want to teach only one of the three substantive scientific areas in this module. In that case, we would recommend beginning with the two introductory sections and then, depending on how much time the instructor wants to devote, either to the full coverage of the selected scientific areas addressed above or the condensed coverage of each scientific area provided immediately above. In short, instructors should be able to develop a coherent syllabus for teaching subsets of these materials in as few as 6 to 8 hours or as many as 24 hours.

For instructors seeking to further condense coverage of epidemiology, we recommend starting with Subsection II.A. of the module, which presents the core concept of the "but-for" test for causation and the concomitant point that there are many causes of any outcome as reflected in Slides 1 and 2 below. Beyond that, students should understand the concepts of general causation and specific causation covered in Subsection II.F.1.-2. because each of the three of the sciences discussed later are keyed to them. *Stubbs* at the end of Section II is a nice case to present the problem of competing causes but we would recommend that instructors with limited time skip over the critique of the evidence in *Stubbs*, which foreshadows the differences between good and bad statistical evidence addressed in Section III. Within Section III, the cohort and case-control study designs addressed in Subsections III.B. 2a. and b. are the primary ones of interest in toxic tort cases. Case-control studies are not intuitive and require longer for students to understand. The outputs of epidemiologic studies, relative risks, odds ratios, attributable proportion of risk and standardized rate ratios, covered in Subsection III.D. are central to evaluating the meaning of a study with regard to causation, although instructors cramped for time could skip much of the math in these subsections. The subsection on adjustment could be eliminated as could the details of significance testing in Subsection F.1., although students should understand that random error can be responsible for a study outcome in part or in whole. The concepts of bias and confounding should not take a great deal of time and one might wrap up this section by explaining the existence of the Hill criteria, in Subsection III.G. and the need for them, *Brock*—an embarrassment in which the court errs in explaining what significance testing does—is a lovely case to impress students with the pitfalls of not understanding science but may be a luxury for those with limited time. Instructors who do not wish to cover toxicology or genetics could close off the course with as little as a few minutes to explain Subsections V.A.–B., or an hour by proceeding to Subsections V.D.–E. but omitting the Racette abstract and affiliated matter, or several hours by including Subsections V.C., F.1., and H.

Condensing the toxicology materials can be accomplished by teaching less than all of Section IV.A. and by teaching none or only some of Section IV.B. Sections IV.A.1. through IV.A.5. provide the essential building blocks for an understanding of toxicology. An instructor who wishes to spend a minimal amount of time on this section should focus attention on these materials. Additional time may be saved by excluding a discussion of the Tyl study in Section IV.A.1. One should be able to cover these sections in a day or two at most. Sections IV.A.6. and IV.A.7. address two key issues (types of evidence and extrapolation) that are at the heart of many toxic tort cases. Covering them as well will add another day or two. One could skip over the discussion of clinical trials in Section IV.A.6., as some of this is covered in the epidemiology materials. An instructor could stop here and teach none of the cases in Section IV.B.

Adding all the materials in Section IV.B. will add considerable time. A complete discussion of each case with references back to the materials in Section IV.A., will take at least another day or two per case. If one wishes to shorten this section, the best first step is to skip the second case: *U.S. v. Tsosie*. It does not address the fundamental causation questions that are at the heart of this entire module. One could choose to teach just one of the other two

cases: *Chlorine Chemical* or *Johnson*. If one's overall focus is more on public law, then one should choose the *Chlorine Chemical* case, while if one is focused on private law the *Johnson* case is the better choice. Unfortunately, however, in order for the students to fully appreciate either of these cases they will have had to read almost of the materials in Section IV.A. Specifically, Section IV.A.7. is crucial to understanding either case. In addition, for students to appreciate the *Johnson* case they will need to have read the materials on the *Daubert* decision in Section II.E.

To condense the material on specific causation, genetics, toxicogenomics, and molecular epidemiology (Sections V.F. and V.G.), instructors have a number of options. First, instructors might wish to skip Section V.F.1 or to assign only the three introductory paragraphs of that section. The excerpt from the *Estate of George* court opinion reinforces material covered earlier (particularly Sections II.D. through II.F., III.D. through III.F., III.I., and V.B. through V.E.) and introduces the concept of supplementing group-based data with information about a particular plaintiff to attempt to estimate the probability of causation in an individual case. The excerpt is pregnant with, but does not directly confront, the genetic variability issues addressed in the subsections that follow it. Thus instructors wishing to move quickly to the science of genetics could omit *Estate of George* or could limit class discussion of the case to the question: "To what extent should characteristics of individual plaintiffs be considered an 'adjustment' to population-based relative risks derived from epidemiologic study?"

Subsections V.F.2. and V.F.3. discuss how genetic and genomic data might be used to assess whether genetic variations make an individual plaintiff more or less susceptible to the toxic effect of an exposure (Section V.F.2.) or to developing a disease or condition irrespective of any exposure (Section V.F.3.). These subsections are arranged in this sequence because of the module's focus on alleged toxic effects and because they flow well in this sequence after *Estate of George*. For some students, inherited susceptibility to disease may be an easier concept than inherited susceptibility to toxic effects. An instructor who chooses not to teach *Estate of George* might begin with the first 10 paragraphs of Section V.F.2., which provide an overview of genetics beginning with the structure of DNA, and then teach Section V.F.3. before returning to complete Section V.F.2. Instructors could rely on the reading to provide the basic genetics material, using class time instead to focus on application of genetic data to toxic tort causation. Another time-saving technique would be to teach V.F.3. without discussing the excerpt from the *Bowen* court opinion. Bowen highlights two important concepts: (1) the possibility that inherited genetic susceptibility and toxic exposure may be competing causes of a disease, and (2) the importance of distinguishing between genetic variations that cause disease and genetic variations that cause disease in concurrence with a toxic exposure. Working through the case's procedural complexities, however, may require a bit of effort in class, which instructors who wish to save time might choose to forgo.

Section V.G. addresses the frontiers of research that seeks to identify biomarkers that can distinguish cases of disease caused by an alleged toxic exposure from cases of disease not caused by that exposure (even if the plaintiff received the exposure). The text of this section is relatively short, but it is relatively dense as well, and the concepts involved are qualitatively

different from what precedes it. Therefore, an instructor who wishes to reduce the amount of material covered and prefers a tighter focus on the role of genetics in disease causation could omit Section V.G. and teach only Sections V.F.2. and V.F.3. On the other hand, an instructor who wishes to include the material on biomarkers could use either genetic susceptibility to disease or genetic susceptibility to toxic effect as a sufficient illustration of the role of genetic variability and the principles of genomic research. In that case the instructor would teach either Sections V.F.2. and V.G., or the first 10 paragraphs of Section V.F.2. followed by Sections V.F.3. and V.G.

At a number of points throughout the module, we have built inquiries about fundamental matters into the text. Instructors could use classroom time to go over those with students, assuring their understanding of the concepts explained in the module. At the same time, we have included numerous "study questions" in these teaching notes, likely more than any individual instructor will have time to cover. Thus, we leave it to each instructor to decide which study questions to employ in teaching this module.

We have prepared a number of slides that might be useful in illustrating and discussing the module. We have identified where we believe these slides would be most usefully employed and displayed each of them in these Teaching Materials when discussing them for ease of reference. ***The content of the slides is our own except where indicated otherwise.*** The slides are also contained in a separate PowerPoint file available with this module. Instructors should feel free to modify these slides to suit their individual purposes. Some may decide to develop additional slides, and we would be grateful if you would share those with us so that we might employ them in future iterations of this module.

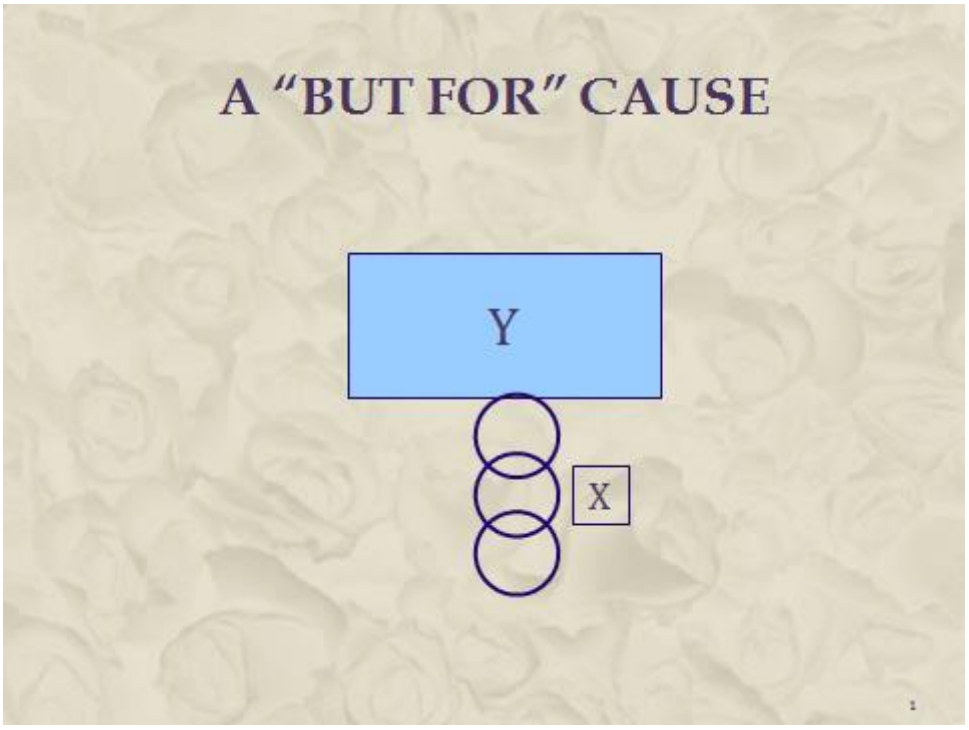
We have constructed these materials envisioning that students will read a section of the materials in advance of the class in which they are covered along with selected study questions the instructor intends to address in class. Class would involve the instructor lecturing on the important points in the assigned materials—learning and fully absorbing the materials will take considerably more than merely reading them in our estimation. We also recommend that the classroom environment be one conducive to student questions during the lecture as it is difficult to know what aspects with which they will struggle. Study questions would involve class discussion led by the instructor, which have the additional advantage of providing the instructor as well as the students with feedback about students' understanding. Some questions could, if the instructor wishes, be used for student assessment. We have also prepared a comprehensive assessment tool, the "Gold River Hypothetical," that instructors can use at the end of the module or the course in which the module is taught.

Section II. An Introduction to Causation

Section II.A. Sine Qua Non and Substantial Factor

Slide 1 sets forth a causal chain model that we will continue to employ throughout this module.

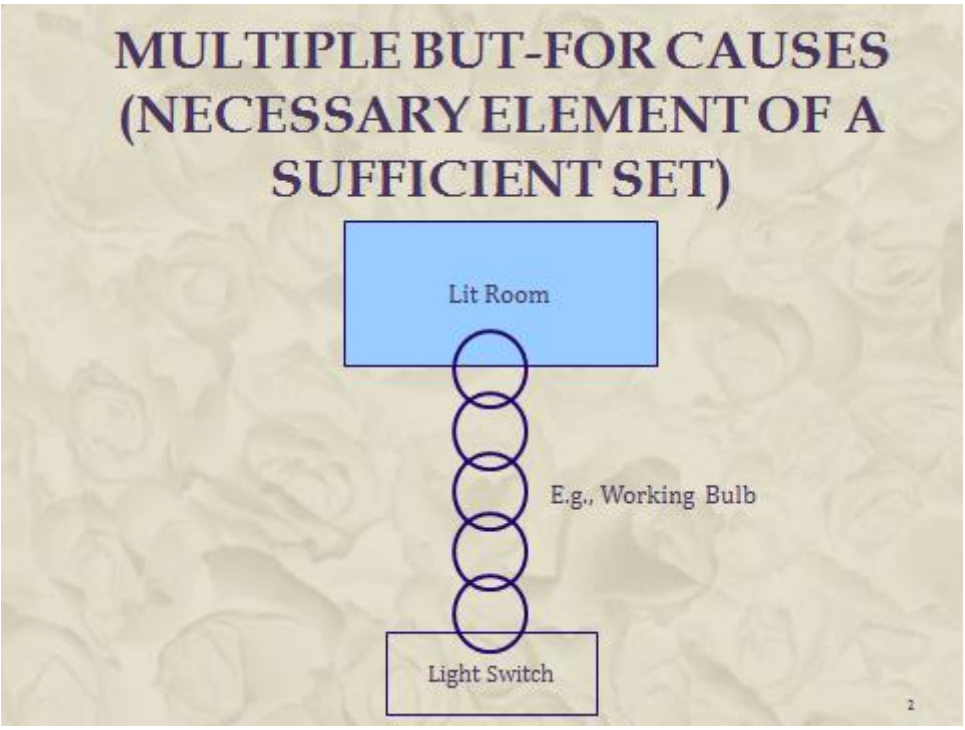
Slide 1 A "But For" Cause



The idea behind this slide is to show the framing of the causal question—the relevant act "X" and the relevant outcome "Y." Then, the circle represents a necessary link in the causal chain. With all of the links in place, Y will occur due to X (and the other necessary causes signified by the other links (circles) in the factual cause chain. Without those other links, Y will not occur, despite the presence of X.

We make the point of multiple necessary causes just discussed more explicit in Slide 2.

Slide 2 Multiple But-for Causes

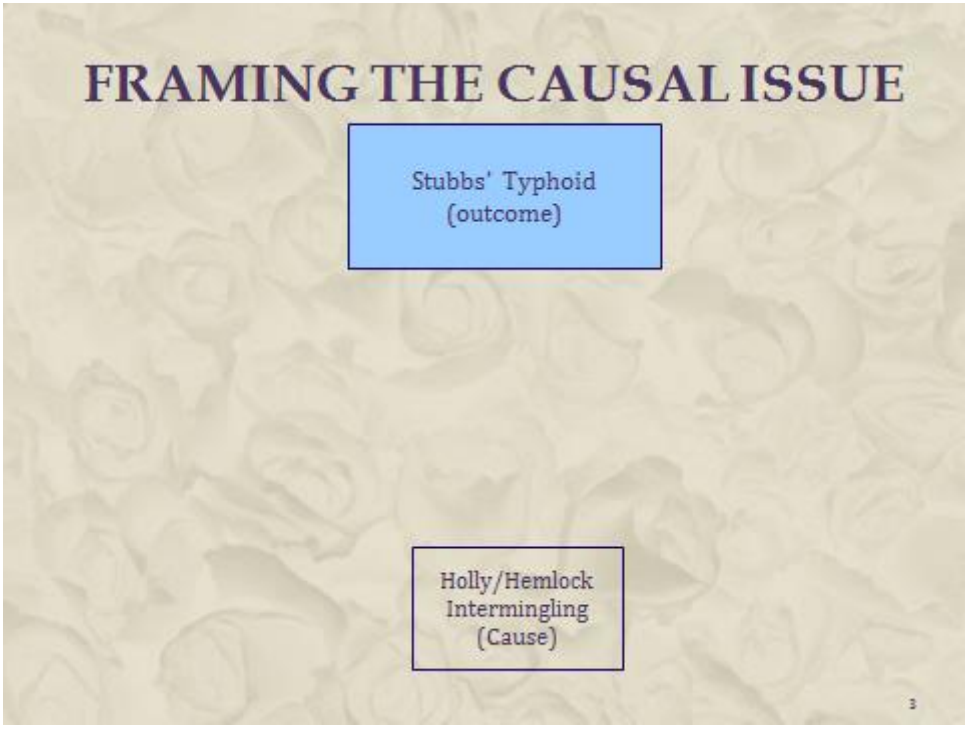


In this slide, the relevant act we are interested in is turning on the light switch and the effect of interest is a lit room. But there are other necessary conditions for the outcome, including a working light bulb, wiring between the switch and the bulb, a supply of electricity, etc. These other necessary conditions, in addition to the act of turning the switch, are represented by the circles in the slide.

Section II.B. Framing the Causal Inquiry

Slide 3, keyed to the *Stubbs* case, which comes later in the module, could be employed here to illustrate the simple point of the need for outcome and cause identification.

Slide 3 Framing the Causal Issue



An example of a cognizable harm that is not a discrete injury is a lost opportunity (or lost chance) for a better outcome that is recognized by many courts in the medical malpractice context when the probability that the malpractice caused the adverse outcome is 50% or less. Courts have re-conceptualized the harm as the lost opportunity. See, e.g., *Matsuyama v. Birnbaum*, 890 N.E.2d 819 (Mass. 2008). Death is always an accelerated harm because we all suffer death eventually.

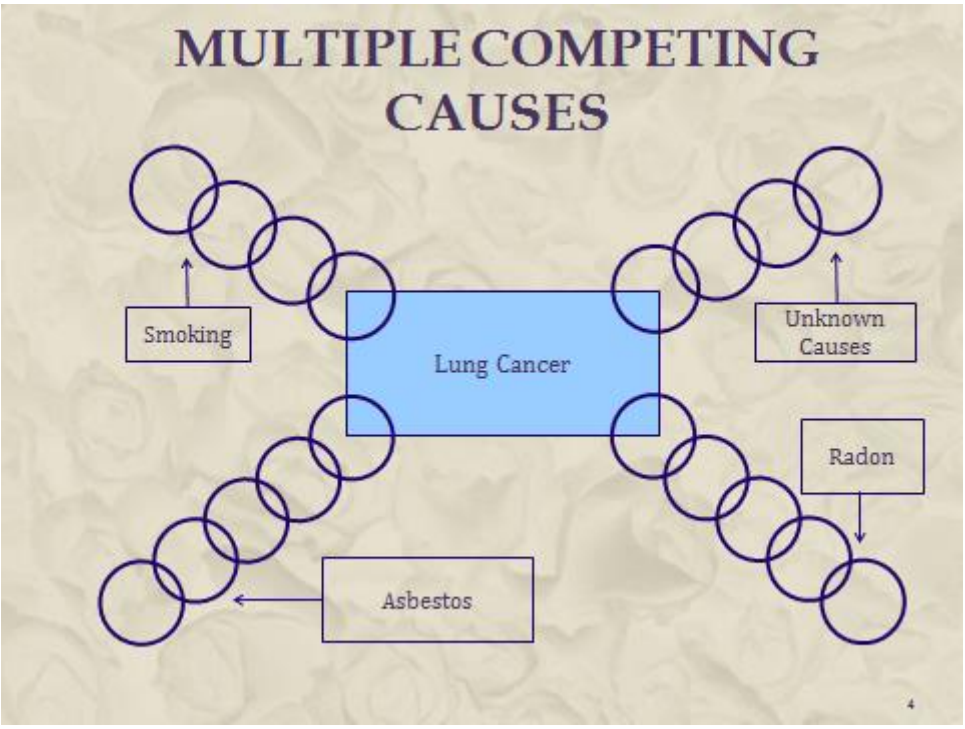
Section II.C. The Necessity of Interference in Making Causal Assessments

This material is relatively straightforward and does not require much in the way of discussion. The instructor might challenge students by dropping a pencil and stating, "You can all see that my hand opening was the cause of that pencil hitting the ground. What inference is required?" We hope that students will respond to the challenge by stating that while they could see all of the physical events and, indeed, understand the mechanism of gravity in the causal chain, nobody actually "saw" causation from the hand opening to the pencil lying on the ground.

Section II.D. The Difficulty of Causal Interference for Toxic Torts

Slide 4 illustrates multiple competing causes:

Slide 4 Multiple Competing Causes



For the sake of simplicity, we have not addressed the synergy between asbestos and smoking. Because the risk of lung cancer when one is exposed to both is greater than the additive risk of the two, there are some lung cancer cases caused by *both* smoking and asbestos. For the sake of clarity, we omit that matter at this point. Some instructors may want to address this here. The relative risk of lung cancer for those exposed to asbestos is roughly 5.0 and the relative risk for smokers is approximately 12. The relative risk for those who have both risk factors should be 17 if there is no interaction. But the actual relative risk is around 60—implying that while among lung cancer victims, some are due only to smoking or asbestos, most (43/60) are due to the combined exposure to both agents. These matters are developed in Section V.E. infra.

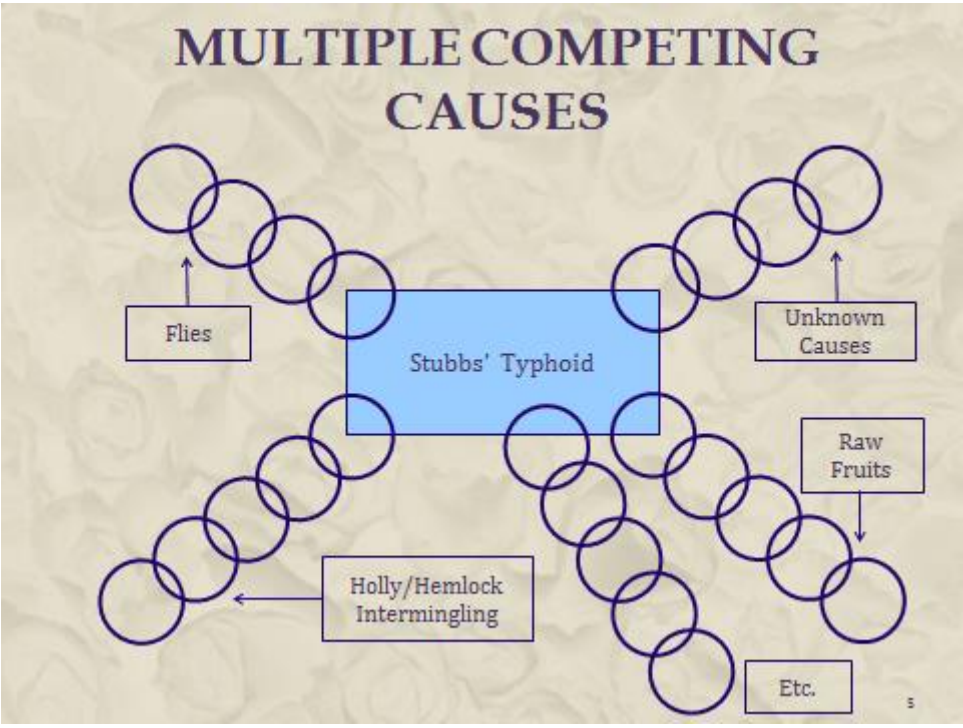
Stubbs v. City of Rochester, 124 N.E. 137 (N.Y. 1919):

Stubbs presents a multiple-competing-causes situation, although each of the "causes" identified by the court are actually exposure vectors by which a victim may be infected with the

typhoid bacillus. While exposure is a necessary link and therefore a "cause" of the plaintiff's disease, we would tend to think of the typhoid bacillus as the relevant cause of typhoid. For the sake of simplicity and clarity, instructors may decide to simply treat the competing exposure vectors as competing causes, as the *Stubbs* court did and as the next slide does.

Slide 5 illustrates multiple competing causes, specific to the *Stubbs* case:

Slide 5 Multiple Competing Causes



The primitive nature of the statistical evidence employed in this case provides an opportunity to think critically about their use and to appreciate some basic principles of epidemiology (study question 1). It also provides a platform to reinforce the materials that the students have already read (study questions 1–4 and 6), an opportunity for a preview of coming attractions (study question 5), and finally, a review of the civil standard of proof of a preponderance of the evidence (study question 9), which plays an important role for the threshold required in a study finding an increased risk of disease.

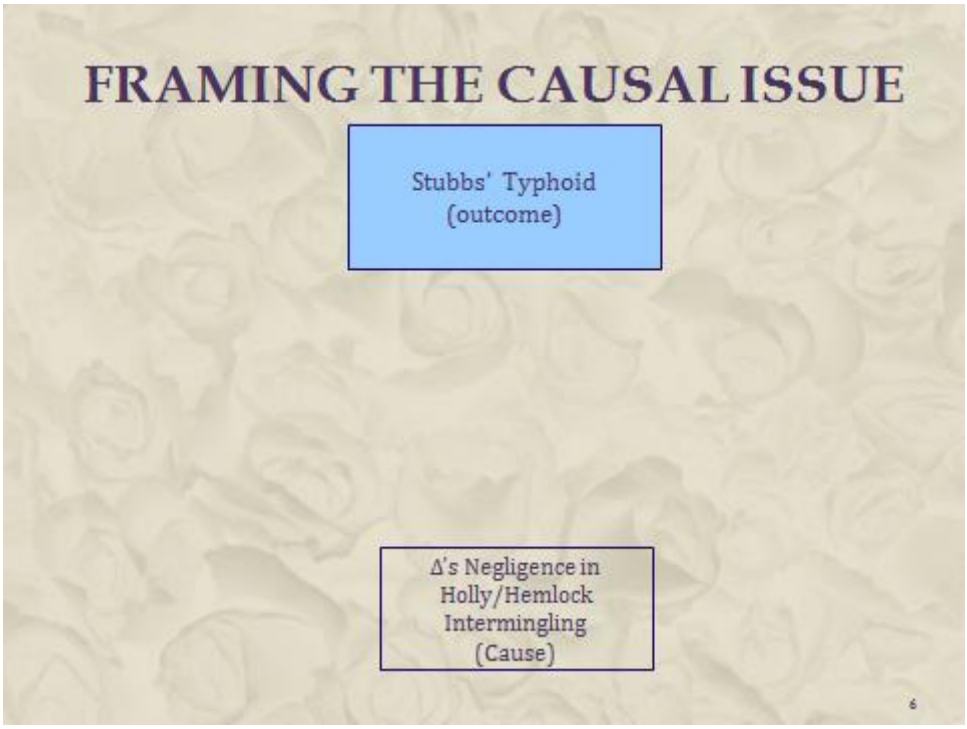
Study Questions:

1. Frame the causal question that must be resolved in this negligence claim.

   Whether the negligence of the City of Rochester in permitting the Holly and

Hemlock water systems to become intermingled caused plaintiff's typhoid fever and the consequences of his contracting that disease. Slide 6 illustrates this framing.

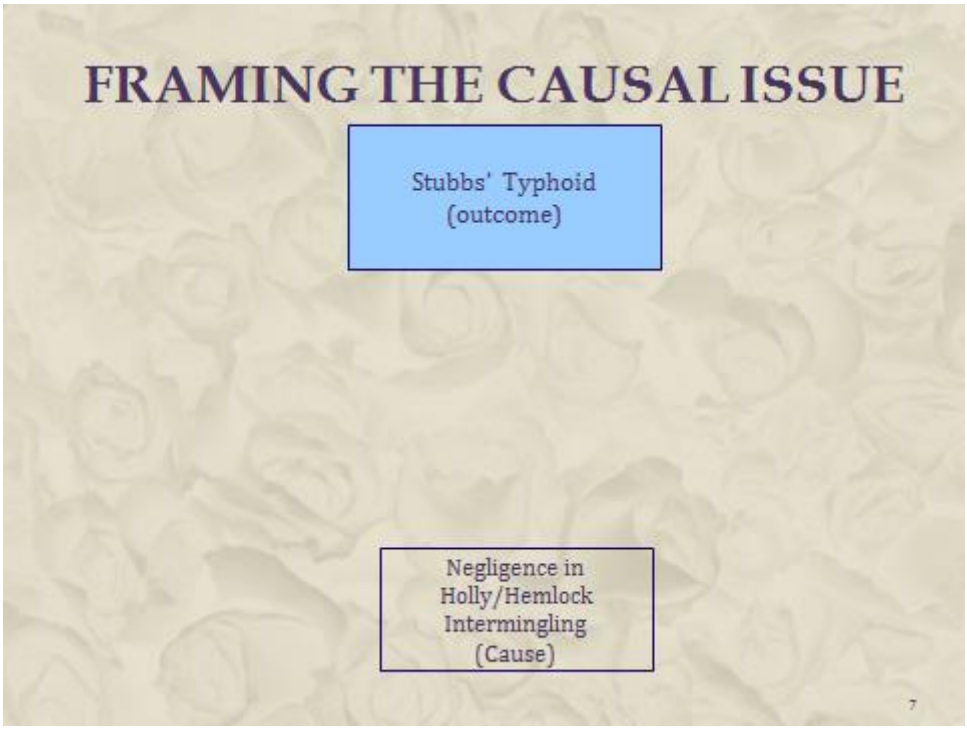
Slide 6 Framing the Causal Issue



2. Frame the agent-disease causal question that exists in this case. In what way are these two causal inquiries different?

Whether contaminated Hemlock water drunk by plaintiff was the source of his being infected with the typhoid bacteria that caused his typhoid (Slide 7).

Slide 7 Framing the Causal Issue



As explained above, *Stubbs* is not about competing causes, although the court presents the issue in that fashion. In fact, typhoid fever is a signature for exposure to, as the court puts it, the typhoid bacillus (the proper name today is Salmonella typhi). The real issue is Stubbs' source of exposure to the bacteria that caused his disease. This is of no matter, though, for purposes of trying to determine that source through statistical methodology and the differential etiology methodology (see infra SectionV.H.4. on differential etiology).

3. The court states that the minimum latency period is 2 to 3 weeks. What evidence about the maximum latency period for typhoid fever would have been helpful for resolving the causal issue in *Stubbs*?

Contamination began in May, although it does not appear that it was discovered until July, when the unusual condition of the water was noticed. Stubbs contracted the disease on September 6th, 4 months later. If the maximum latency period was, say, 3 months, arguably Stubbs would have developed the disease too late for Hemlock water to be the cause. This is only arguable because he was continually exposed through the time when he contracted typhoid fever, and he might not have been infected by the bacteria until some later point after his initial exposures. Fairly good information about

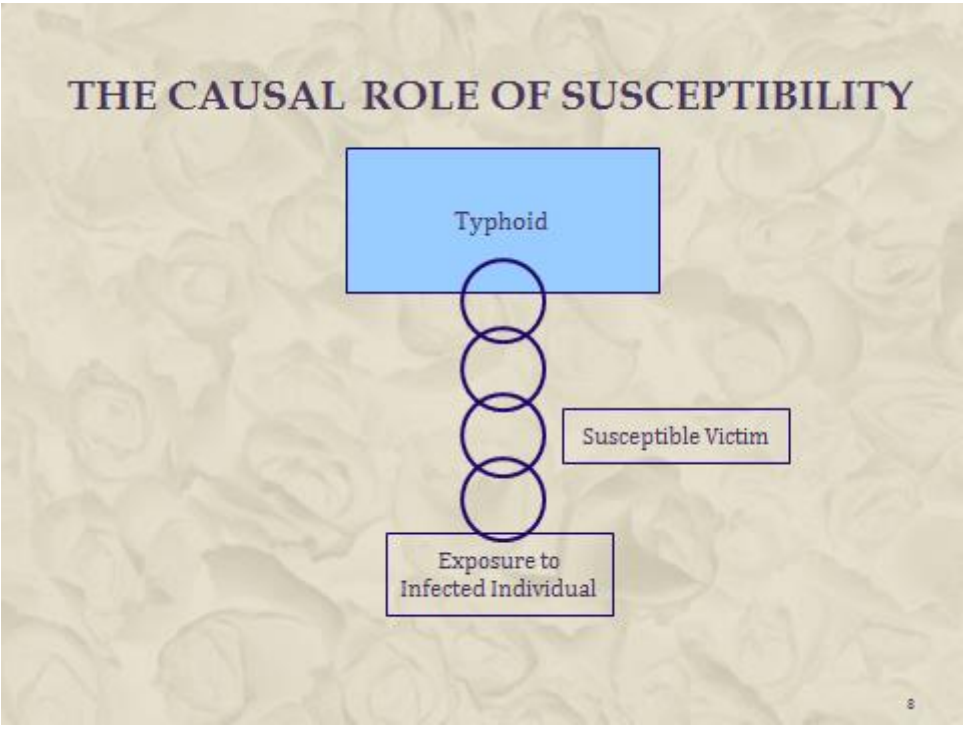
the latency window for childhood-neurological problems has informed causal determinations in the vaccine court set up by the Childhood Vaccine Act, for those suffering adverse events after receiving a vaccine, as revealed in *Burchett v. Secretary of Health & Human Services*, No. 12-119V, 2014 WL 2465194 (Fed. Cl. May 13, 2014):

> (3) Althen Prong Three: Proximate Temporal Relationship
>
> Both parties' experts agree that the timing for a viral source for petitioner's recurrent GBS is medically appropriate. Dr. Rubenstein testified that, in cases of recurrent GBS, onset can take place within two to three days to six weeks. Tr. 68, 71. Dr. Kohrman opined that onset of recurrent GBS happens either from four to five days to six weeks. Resp't's Ex. D at 6-7.
>
> The medical literature the parties submitted addresses the issue of timing. The authors of the Slade article stated that "8 of the confirmed cases [of Guillain-Barré syndrome] were within the 4- to 42-day window of biological plausibility." Resp't's Ex. K at 5; Resp't's Ex. D at 6-7. Likewise, the "Institute of Medicine determined that the plausible range of post-exposure latency for GBS to be 5 days to 6 weeks." Resp't's Ex. D at 6. In the Mossberg article, the authors found that the time from the triggering infection to the onset of RGBS [recurrent GBS] showed a tendency to shorten in successive episodes…" Resp't's Ex. H at 4.
>
> As noted above, the onset of petitioner's GBS was March 31, 2010. The parties stipulate that petitioner was diagnosed with and treated for viral gastroenteritis on March 19, 2010. Jt. Sub. at 1. Thus, onset took place between nine and fourteen days after petitioner was diagnosed with viral gastroenteritis.
>
> Regarding the upper respiratory infection, the undersigned has found that petitioner experienced the URI approximately three weeks prior to the onset of her GBS on March 31, 2014. Thus, the timeframe as supplied by both petitioner's and respondent's expert is also medically appropriate for a URI.
>
> In light of the above, the undersigned finds that respondent has provided preponderant proof that the onset of petitioner's recurrent GBS occurred within a medically timeframe in relation to one of her antecedent infections.

4. The court states that typhoid fever is contagious and that contact with someone already suffering from typhoid fever can cause it in someone "who has a predilection for" infection. Explain how susceptibility might play a causal role in typhoid fever. Where it does, is the cause of typhoid the victim's susceptibility or the exposure to someone with the disease?

Slide 8 reveals that both contagious exposure and individual susceptibility are but-for causes of typhoid fever. From that perspective, we cannot pick out "the cause," because there are multiple but-for causes required for the outcome.

Slide 8 The Causal Role of Susceptibility



Students might be inclined to tell a story about why an individual is susceptible, for example that he or she has a compromised immune system that is unable to ward off the bacteria once there has been exposure. (Perhaps, but there must be more to the story because otherwise, susceptibility would not be limited to exposure to those with the disease.) But this attempt to explain reveals the role of mechanism evidence in attempting to determine causation. Here, the mechanism story is nothing more than largely uninformed guesswork.

5. In listing the different causes of typhoid fever, the court states that there are unknown causes. What are the implications of the existence of unknown causes for plaintiff's efforts to prove that Hemlock water contamination caused his typhoid fever? Is the existence of such unknown causes an insurmountable obstacle to plaintiff proving causation? Why or why not?

These questions anticipate discussion of differential etiology in Section V.H.4., infra, on specific causation. Differential etiology, while used in some medical contexts, is not so much a scientific matter as it is one of logic. Consider Slide 5. Eliminating raw fruits as a possible cause of plaintiffs' lung cancer makes each of the

other remaining potential causes more likely. If we could eliminate all competing causes save one, we would, by process of elimination, have identified the cause of the victim's lung cancer. This "elimination of competing causes" methodology is widely used in many disciplines when seeking a cause—auto mechanics use it determine the cause of a car not starting and detectives use it in investigating a crime.

6. The court details evidence of the unwholesomeness of the Hemlock drinking water during the months of July through September. Of what significance is this evidence to the causal issue in *Stubbs*?

   This question is designed to reinforce the short explanation about exposure as a critical element of agent-disease causation. Stubbs must show that he was exposed to contaminated water for which the City of Rochester was responsible, and this evidence is responsive to that burden. Note that there is no discussion of how much contaminated water Stubbs drank, which would address the matter of dose. Here, infection with the typhoid bacteria does not appear to be dose dependent (or at least does not require anything more than some low threshold of exposure). Finally, note in connection with this question that the evidence about Stubbs obtaining all of his drinking water from contaminated sources is significant not only because of establishing exposure, but as an aspect of the differential etiology effort to eliminate other competing causes.

7. The plaintiff provided a variety of statistical evidence to support his causation burden of proof. Consider each of these different efforts. In what way is this evidence supportive of the plaintiff's claim? How might the defendant have pointed out weaknesses in each of these statistical datums?

   We intend this question to be an introduction to epidemiologic principles, but some instructors may find this too advanced at this stage in the module. Discussion could be deferred until the presentation of epidemiology in these materials. Alternatively, the instructor could decide to use lecture to explain a critique of the statistics.

   a. The number of typhoid cases in Rochester was 223, an excess of 50 over any of the prior 9 years.

      Several interesting critiques:

      (1) *Incidence rates*. What we do not know is what the population was in Rochester in those earlier years. If the population was growing rapidly,

the increase of 50 may not reflect an increased incidence, that is, the proportion of the population contracting the disease within a specified time period, in 1910, the year of intermingling.

(2) *The difficulties with secular trend data*. This data compares disease across different periods, one without exposure (1900–1909) and with exposure (1910) to attempt to discern whether the exposure is associated with an increase in disease. The difficulty is that there may be other changes that occurred that may be responsible for the increase. If risk factors are known, it may be possible to account for them, but recall here that some causes are unknown. And it is not just changes in risk factors. Perhaps the change responsible for any increase in typhoid fever was improved diagnostic techniques, such that earlier cases were not diagnosed.

(3) *Random phenomena*. This is a classic cluster—an excess of disease in a localized area. Clusters can be evidence that something is going on but they can also occur randomly. Beginning in 1980, there was a well-publicized cluster of cancers among players for the New York Giants whose games were played in the stadium in the meadowlands of New Jersey where a variety of hazardous waste sites existed. Four players contracted cancer; two of them died. Ultimately, a study concluded that environmental factors played no role in the cancers. (Joseph F. Sullivan, "Athletes' Cancers a Coincidence, Study of Meadowlands Site Finds," *New York Times* [July 15, 1989]). A nice way to explain clusters is to ask students to imagine a large board in the shape of the United States onto which 1,000 pennies (representing disease) are dropped. Those pennies would not end up spread evenly across the board. Instead, because of randomness, some clusters would exist although nothing on the board would have had a role in that skewed distribution. When students confront random error in the epidemiology section, they will better appreciate the tools—statistical significance testing—that scientists of all stripes employ to assess whether their data reveals something real, or merely reflects randomness. Statistical analysis of this data would provide some sense of the plausibility that this increase in disease was due to randomness.

b. Of the 223 cases of typhoid, 180 occurred in 4 months from August through November and only 43 during the remaining 8 months.

Aside from the possibility of random error (unlikely, given the magnitude of the increase), it seems something is going on during these four months. But what? A number of the other causes listed by the court seem likely to be

more prevalent during the summer months. More data on what the distribution of typhoid over the course of the year in years other than 1910 would be helpful.

c. Fifty-eight others who drank Hemlock water during the period of contamination also contracted typhoid. This datum is interesting because it begins examining groups that are different in a single way relevant to the study question at hand, a central tenet of epidemiology. The 58 others were exposed to contaminated Hemlock water, while others in Rochester presumably were not. The problem, however, is that all we have is an absolute number without having a denominator (the total number of those exposed to contaminated Hemlock water) that would permit calculation of an incident rate for the typhoid disease among those exposed. Even then, we would need to compare that incidence to the incidence among a group (controls) who were similar to the exposed group, save that they were not exposed to contaminated Hemlock water.

An instructor might want to vary the data to set up a case-control study design: Thus, 100 people treated for typhoid during this period were matched with 100 people who were not, and the rate of exposure to contaminated Hemlock water for each group was determined and compared. On balance, our assessment is that it is too early to start with the subject, which we will come back to when discussing different epidemiologic types of studies.

8. What other data would have been useful in assessing the causal issue in *Stubbs*?

Ideally, we would have a control group identical (or at least similar) to those who were consuming contaminated Hemlock water but were not, themselves, exposed to that water. Then we would compare the incidence of typhoid over the same period of August through November in the control group with the incidence in the group that was exposed to contaminated Hemlock water. The effects of contamination would have revealed themselves in a higher incidence of typhoid among those exposed to the contaminated water if a causal relationship existed. This would dovetail with what is known as a cohort-study design for an epidemiologic study. The basic insight is to compare two groups in which the only difference is exposure to the suspected agent and then to compare the rate of disease to see if it is different in the two groups. The difficulties with these "observational" studies are laid out in the module in Section III.B.2., infra.

In addition, for item 7a, the results of statistical testing for random error would be useful to have.

As stated in 7b above, information about the distribution of typhoid cases over the course of other years would provide important perspective in assessing whether this increase could be attributed to contamination of Hemlock water.

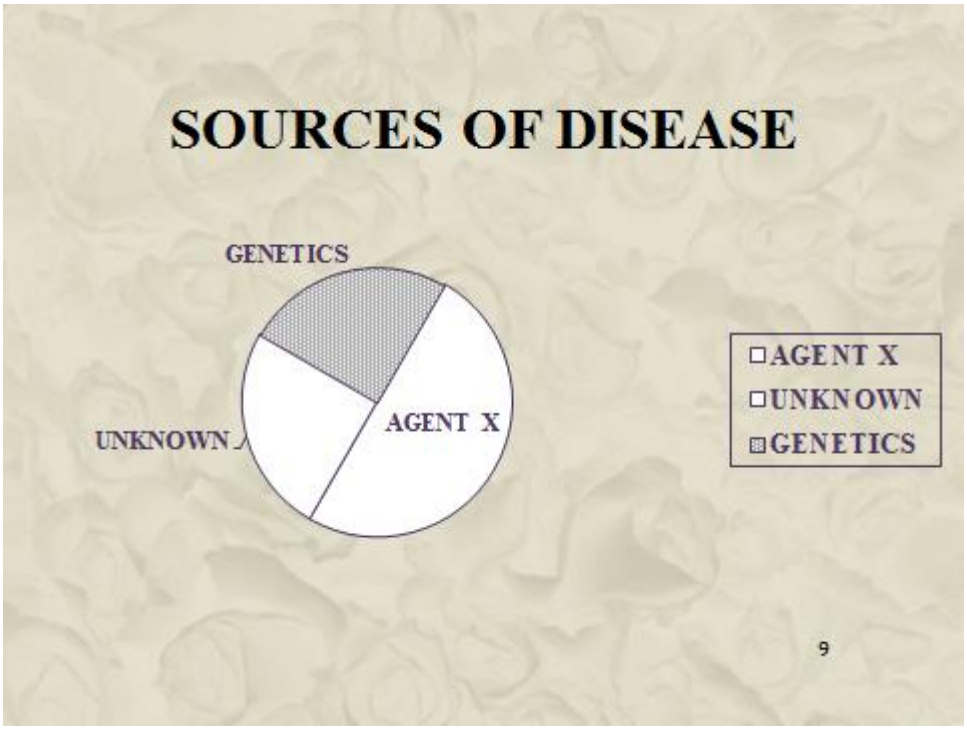
Again, as stated above in 7c, we would like to know not *some number* of those who drank contaminated Hemlock water and suffered from typhoid fever, but rather, the incidence of typhoid among those who drank Hemlock water and a comparison incidence to those who did not drink the contaminated water. An instructor might vary the data to set up talking about a case-control design, that is: 100 people treated for typhoid during this period were matched with 100 people who were not, and the rate of exposure to contaminated Hemlock water for each was determined and compared (but it seems too early to get this sophisticated).

9. How many of the other exposure routes must the plaintiff eliminate in order to satisfy his burden of proof?

The court rejects defendant's claims that plaintiff must eliminate all competing potential causes of his disease in order to meet his burden of proof on causation, instead stating that he merely had to demonstrate "reasonable certainty" that the Hemlock water caused the disease. This is a good time to talk about the civil standard of proof, which requires proof by a preponderance of the evidence. How is the "reasonable certainty" standard adopted by the court different from preponderance of the evidence (or, "more likely than not")? Reasonable medical (or scientific) certainty was adopted by many courts as the requisite threshold for the admission of expert testimony, but many of those courts equate that standard with the preponderance standard. See Restatement (Third) of Torts: Liability for Physical and Emotional Harm Section 28, cmt. e & cmt. c, rptrs. note. The preponderance standard plays an important role in determining whether epidemiology studies are adequate to prove specific causation. See Section V.B., infra.

Slide 9 illustrates in simplified form the role of unknown causes, which cannot be eliminated. However, other known causes sometimes can—in this slide, genetics might be eliminated. Note that if genetics is eliminated, Agent X is responsible for more than half of the remaining disease, indeed for two-thirds of it, which would satisfy the preponderance of the evidence standard. We should also recognize that the risk profile for plaintiff may be different because, for example, he was exposed to a significantly higher dose than those in the study that produced the risk proportions shown in the slide. More on this in Section V on specific causation, infra.

Slide 9 Sources of Disease



10. What is the significance of the evidence that P had his own drinking cup at work?

This addresses the question of whether other exposures might be responsible for Stubbs' typhoid. If a co-worker was infected from some source other than the contaminated Hemlock water and Stubbs drank from the co-worker's cup, then that would constitute a competing source of disease that Stubbs would not be able to rule out. Of course, it could be that the co-worker's cup was contaminated because of Hemlock water, in which case it would not matter that the co-worker's cup rather than Stubbs' cup was the exposure vector for the typhoid bacteria from the contaminated Hemlock water.

Section II.E. The Role of *Daubert* and its Progeny

This section is self-explanatory.

Section II.F. General Causation, Specific Causation, and Signature Diseases

One might attempt to reinforce the concept of general causation by asking students about why this concept is almost never used for classical traumatic-injury cases, such as the

broken arm suffered in an automobile accident. For most traumatic-injury cases, we have a good enough sense of the biological mechanism that we simply take general causation for granted. As Hart and Honoré put it (without consideration of toxic tort cases), "the lawyer . . . is concerned to identify particular causes with the aid of established causal laws or accepted generalizations." But, even in this context, we need to be satisfied that there is either a theory by which general causation exists or statistical evidence establishing its existence.

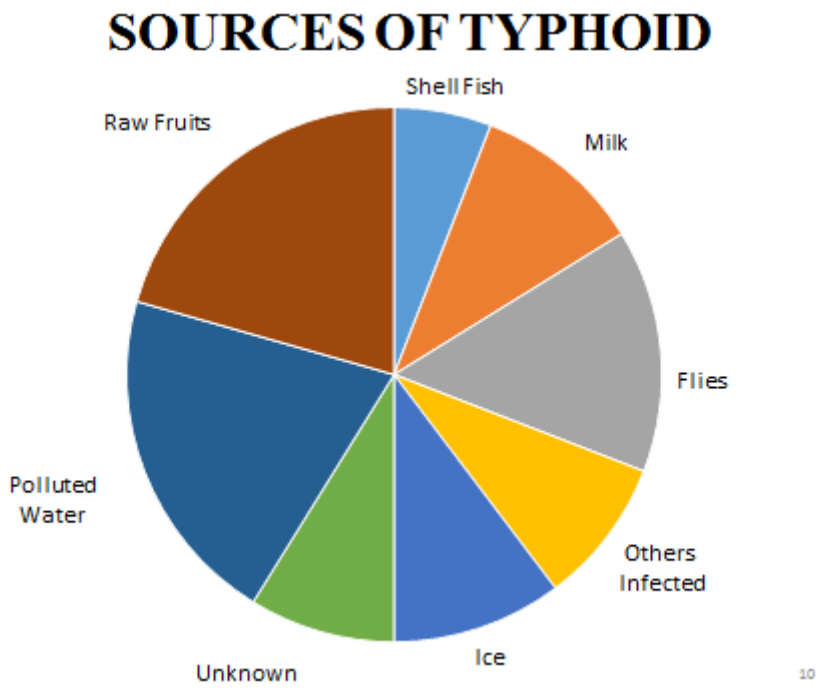
Study Question:

1. Was the causal issue in *Stubbs* one of general causation or specific causation?

   The court relates a litany of potential known causes of typhoid. Each of them addresses general causation, including exposure to water contaminated with the typhoid bacteria. Stubbs' causal burden was to show that it was contaminated Hemlock water, rather than other known or unknown causes, that were responsible for his typhoid.

   Slides 10 and 11 are designed to illustrate, by reference to *Stubbs*, the idea of a signature disease.

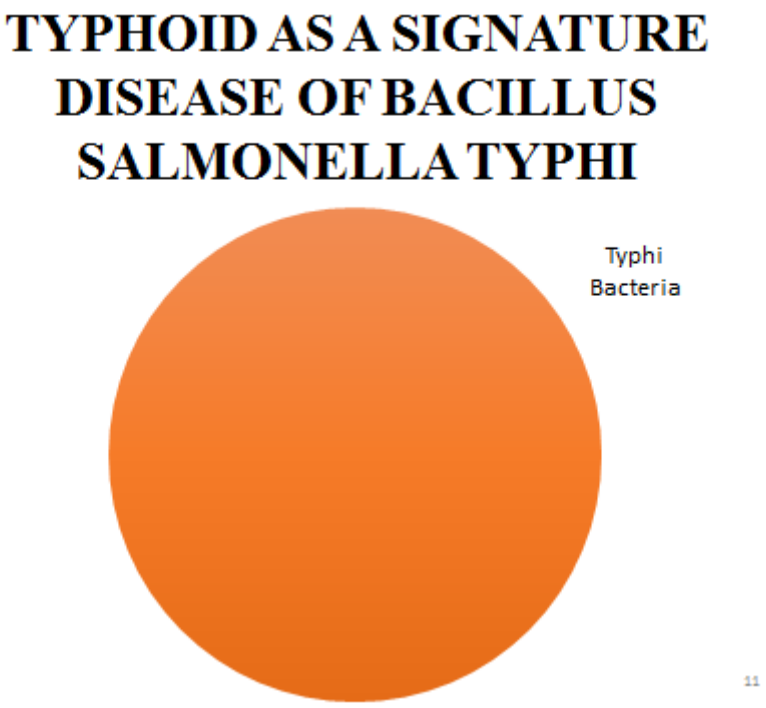
Slide 10 Sources of Typhoid



Slide 10 shows the risk factors for typhoid as related by the court. Each one is a competing cause of Mr. Stubbs' typhoid, at least until that one can be ruled out as a cause. Specific causation—connecting Stubbs' exposure to contaminated Hemlock water—was the issue in *Stubbs*.

Slide 11 goes on to show a single cause of typhoid, thereby collapsing the general-causation and specific-causation questions.

Slide 11 Typhoid as a Signature Disease of Bacillus Salmonella Typhi



The instructor should advert to the explanation in the module that the issue in *Stubbs* was not agent-disease causation by connecting the defendant to the agent that caused Stubbs' typhoid. So, the "causes" shown in Slide 10 are vectors for exposure by which the agent could have gotten to Stubbs and the issue was whether one of those vectors, contaminated water from the Hemlock system, was the one responsible for Stubbs' disease.

Study Question:

1. Is there any causal issue that remains when plaintiff has contracted a signature disease?

As revealed in Slide 11 and in the discussion above, a signature disease obviates any question about general causation and specific causation, but causal issues relating to the connection between defendant's negligence (or other tortious conduct) remain, as *Stubbs* reveals. Other sources of infection with the Typhi Bacteria exist that would not be the product of Holly-Hemlock intermingling and therefore would not be the legal responsibility of the defendant in *Stubbs*. Market share liability is another example where, despite a signature disease, as occurred with the drug DES, which caused vaginal adenocarcinoma in young women, there remained the question of connecting the DES that caused a given plaintiff's disease with its manufacturer.

Section II.G. Causation Standards in Tort and in the Regulatory Arena

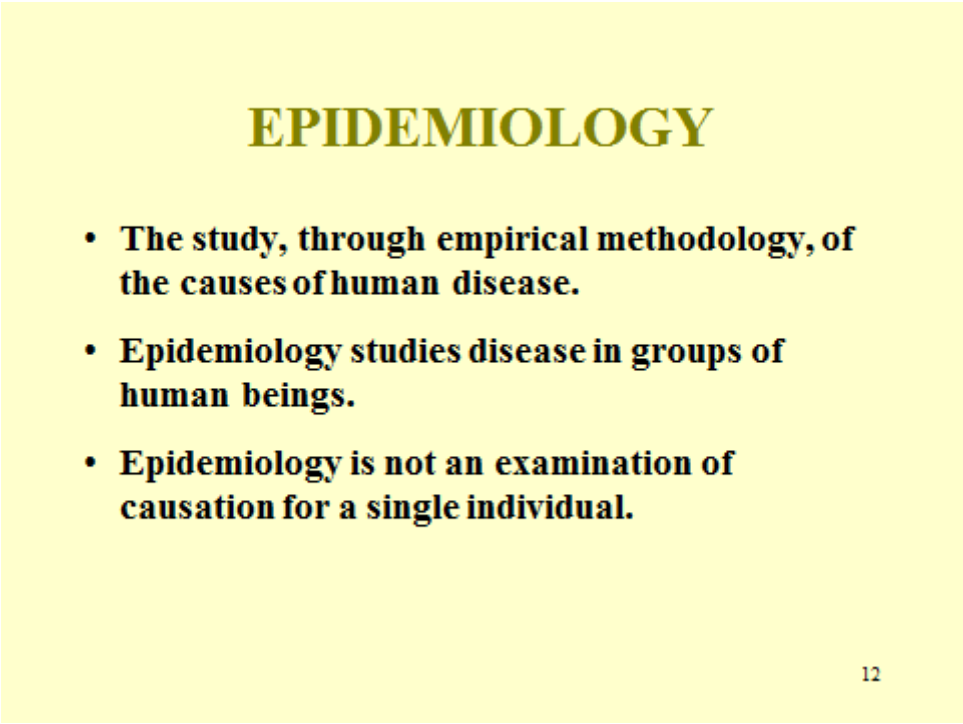
This section is self-explanatory.

SECTION III. EPIDEMIOLOGY

Significant portions of this material are from the chapter on Epidemiology continued in the Federal Judicial Center and National Research Council's *Reference Manual on Scientific Evidence* (3d ed. 2011). That document is widely respected as authoritative by both state and federal courts and frequently cited and relied upon in judicial opinions.

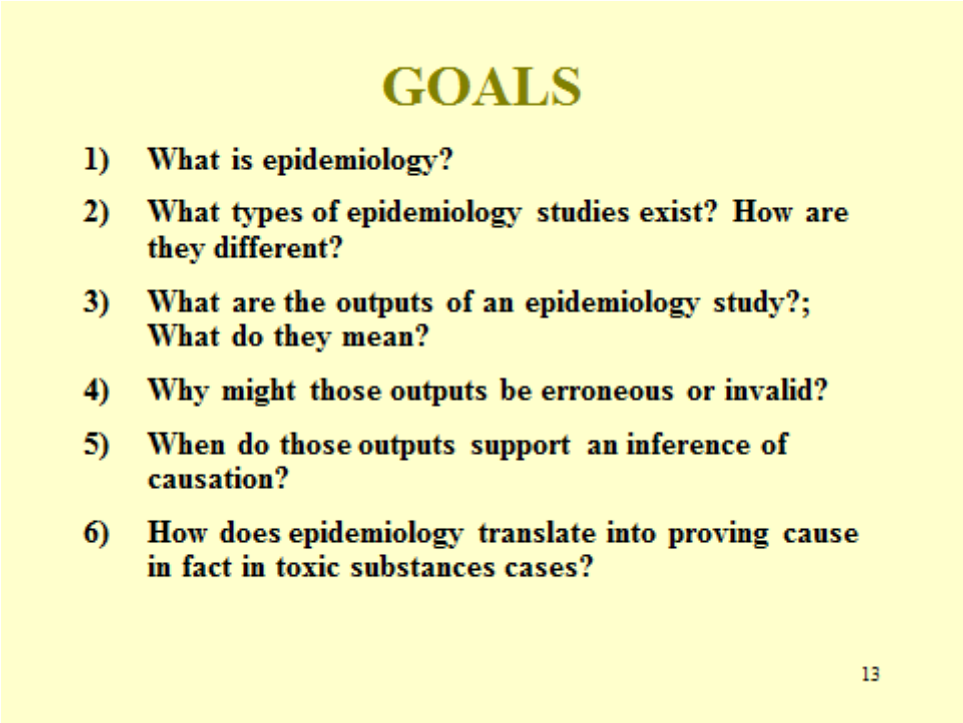
Slide 12 contains a stylized definition of epidemiology along with reinforcement that it is about general causation and not specific causation.

Slide 12 Epidemiology



Slide 13 displays the goals for this segment of the module on Epidemiology.

Slide 13 Goals



Section III.A. Introduction

A basic principle that all students should appreciate is that epidemiology addresses the general-causation question. Epidemiology studies disease in groups of humans and does not attempt to address the question of whether a toxic agent caused a specific individual's disease (specific causation). Indeed, the concept of specific causation was developed by courts, rather than by scientists. Nevertheless, a substantial body of legal precedent has developed that addresses the use of epidemiologic evidence to prove causation for an individual litigant through probabilistic means, and the law that developed in these cases is discussed later in these module materials.

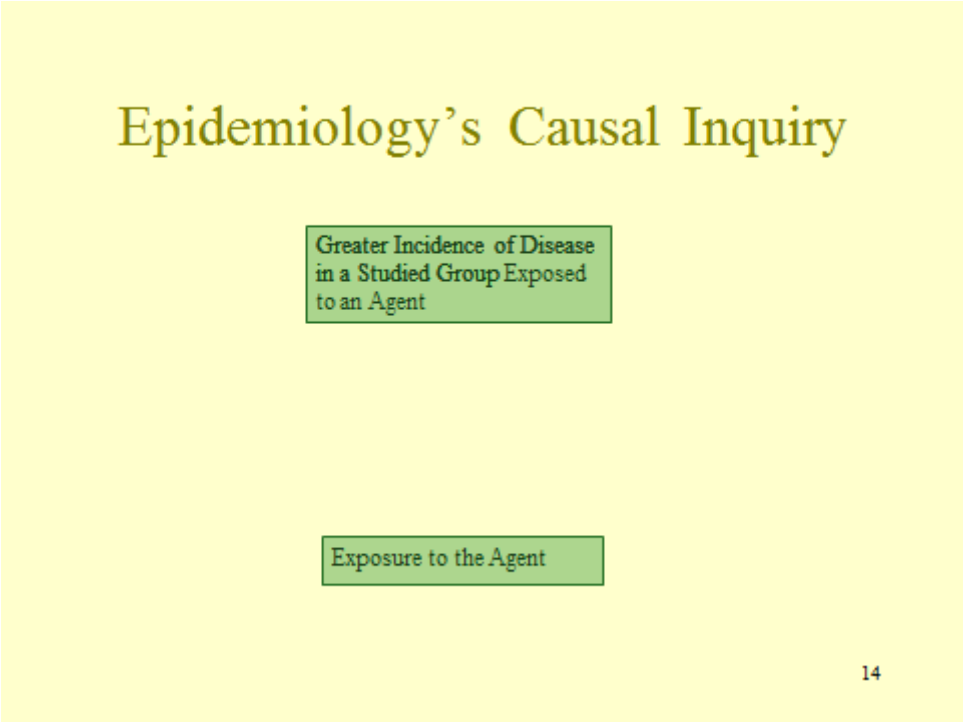
Study Question:

1. Frame the causal question that epidemiologists study.

   Epidemiology speaks of cause in the sense of inquiring whether among a group of individuals, exposure to the agent of interest—which is the difference between

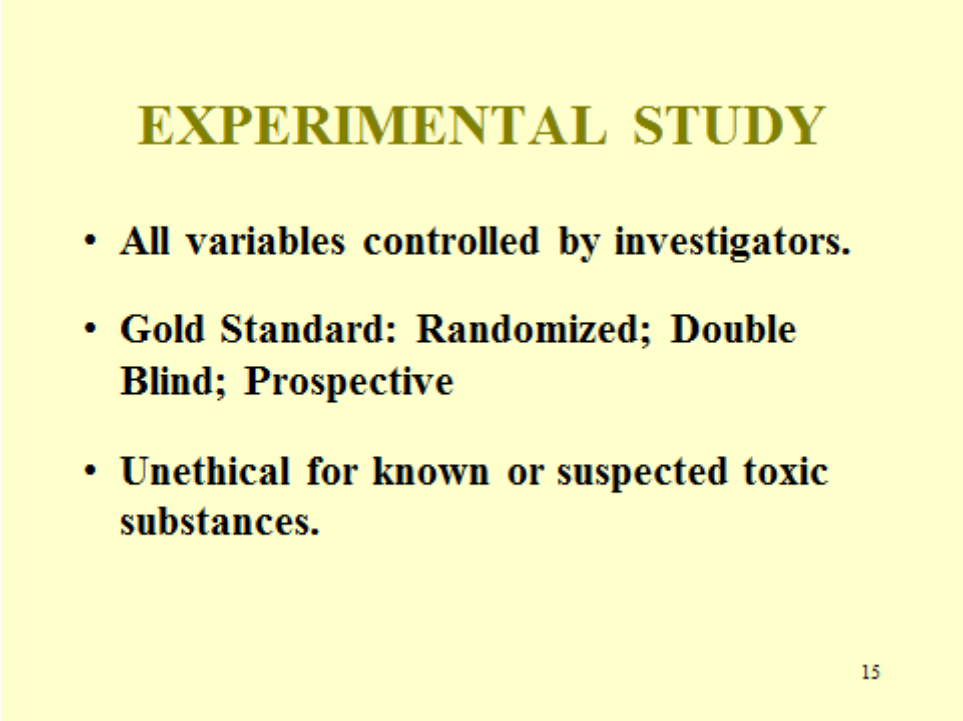
the two groups—is a necessary link in the causal chain that leads to more disease in the group. This is general causation. Epidemiologists do not address specific causation in their scientific work, although they may testify about it. This is displayed in Slide 14.

Slide 14 Epidemiology's Causal Inquiry



Section III.B. Experimental and Observational Studies

Slide 15 Experimental Study



This slide is designed to highlight the characteristics of an experiment and provide a comparison to observational studies, in which none of these characteristics exist.

Study Question:

1. The first line of Slide 15 is not quite accurate. What is wrong with it?

   "Some variables," rather than "all variables" are controlled by investigators. Instructors may want to point this out, ask students what is inaccurate, or revise the slide. Even for randomized clinical trials, investigators do not control the food each subject eats, the use of cellphones, and myriad other aspects of subjects' lives that might affect the risks of disease to which they are subject. Random distribution of these factors might minimize their impact, particularly if the study group is large. If the study is prospective and there are other suspected risk factors for the disease being studied, data can be obtained to enable statistical adjustment for that factor. In some cases, similar adjustments may be possible even for retrospective studies.

But, in the end, human-experimental studies are not as free of biases as are animal studies in which investigators can control all environmental risk to which the animals are exposed.

Indeed, the truly ideal study would not only involve the characteristics mentioned above, but would also put all of the subjects in a big box with a controlled environment that was the same for each group, providing exactly the same food, water, exercise, and other characteristics.

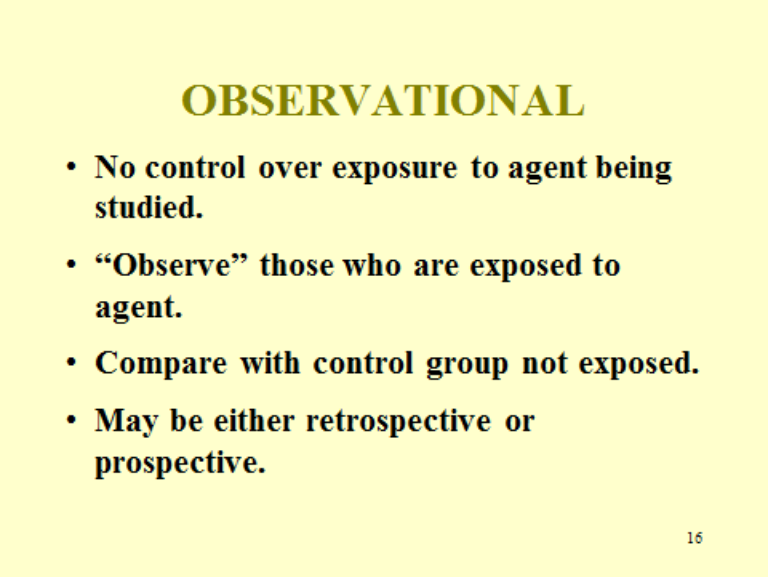
We obviously cannot do that with human beings, although we can with animal studies and that is one of their advantages over human studies.

Indeed, ethical limitations prevent experimental studies of suspected toxic agents. Consider the ethical difficulty of conducting a prospective, experimental study of exposure to arsenic or asbestos. Plainly, it would be improper for an epidemiologist to deliberately expose study subjects to a known or even suspected carcinogen or other toxic substance.

In this situation, scientists must accept and study those who have been or are being exposed to the agent and compare them to those who have not been exposed. These are not experiments; rather, they are what is known as observational studies. Cigarette smokers, asbestos workers, or those who work in industrial settings where chemicals are used as solvent or for other purposes are all people who have been exposed to suspected toxic agents and can therefore be "observed" for exposure and disease status.

Thus, virtually all epidemiologic work with toxic substances is of the *observational*-study type (Slide 16) rather than of the experimental type. Students may be confused because even in experimental studies, researchers conduct "observations," so "observational" here is a term of art.

Slide 16 Observational



The significant difference between experimental and observational studies is that in observational studies, the investigators do not control exposure to the agent, rather, they rely on what evidence is available—asbestos fiber counts, a smoker's report on how much he smokes, a mother's recollection that she drank an average of four alcoholic drinks per week during her pregnancy, or similar evidence.[1]

This section reveals one of the weaknesses of observational studies compared to experiments. The way in which exposure and dose is determined may have some degree of error in it, and our study outcome will necessarily be affected by that error in measuring the dosage of exposure.

Nor, in observational studies, do epidemiologists control for other factors that may also have a causal role: diet, stress, genes, and exposure to other toxicants. If those factors are known or suspected to have a causal role, they can be measured. But unlike a rat study, epidemiologists do not control many variables, such as diet, in their study populations. However, they may be able to control for some risk factors by using certain selection criteria of those included in the study—for example, by excluding smokers from a study or, even if they are included, gathering data about whether a participant is a smoker and thereby enable researchers to assess its impact on the outcome.

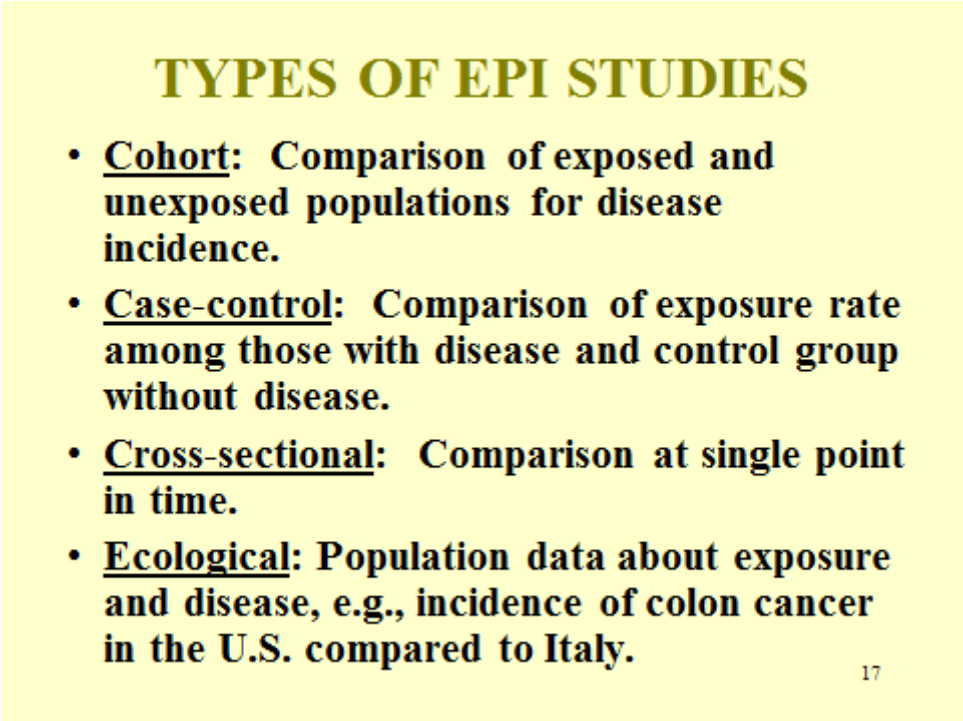
Observational studies may be either retrospective or prospective. Dr. Irving Selikoff's studies of asbestos insulation workers present an example of retrospective studies. Students

[1] Exposure may be acute, as with ingestion of a drug or Bhopal, or it may be chronic, as is the case with many of the toxics mentioned above. Measuring exposure in the chronic case also involves estimating the point at which a trigger dose has been reached so as to know when to start looking for an effect.

might be asked why he chose a retrospective cohort study design; the answer is the lengthy latency period for cancer that has to be accounted for in a prospective study. Alternatively, an example of a prospective study is provided by the famous heart study of men in Framingham Massachusetts that began in 1948 and continues today with an offspring and third-generation cohort.

There are four different types of epidemiological studies identified in Slide 17, but two of them are most important for purposes of toxic substances cases: cohort[2] and case-control.

Slide 17 Types of EPI Studies



Before talking about those study designs in detail, we think that mention of cross-sectional studies and ecological studies is worthwhile so students understand why those studies are not best for examining the causal question asked by tort law, but also appreciate that they can play a role. An expert witness in the Bendectin litigation did an ecological study that compared the incidence of birth defects prior to 1983 while Bendectin was on the market with the period after it was removed from the market, and found no difference in the incidence between those two periods. Thus, an ecological study does not measure exposure or disease in

[2] The retrospective cohort study relies on historical exposure and a follow-up period. In that kind of study, the investigator is relying on others' measures of exposure and diagnosis of disease for ascertainment. The reason for doing this is that it is much cheaper than studies where the investigator is contemporaneously following the cohort and controls.

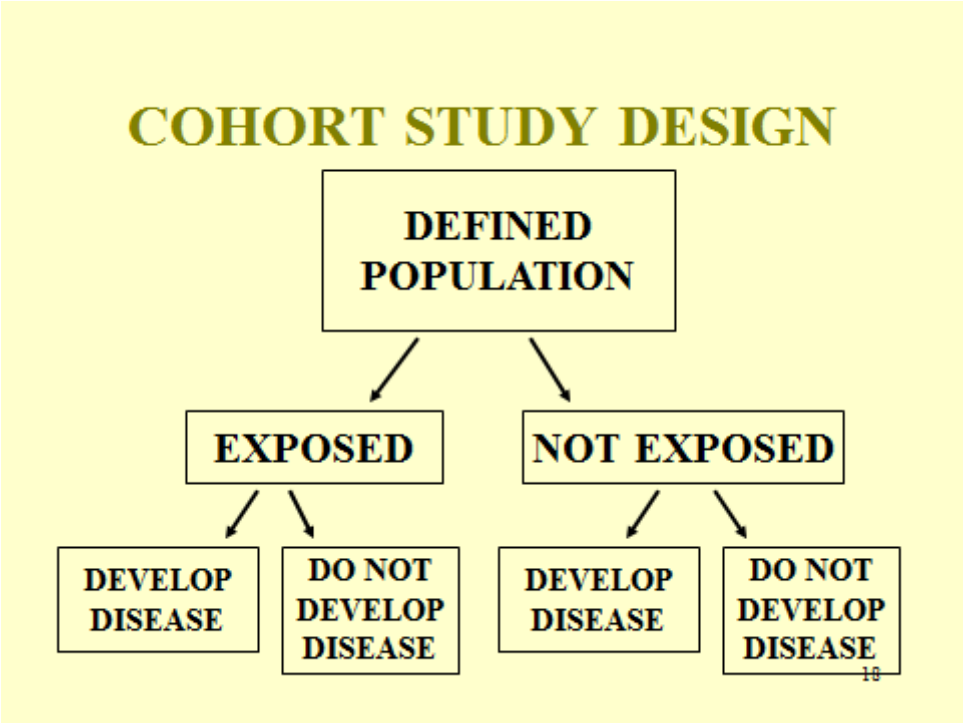
individuals, but instead employs group or national statistics to determine exposure rates and diseases. The difficulty with secular-trend ecological studies of this sort is that something else may happen during the relevant time period that affects the incidence of the studied disease.

Another difficulty with ecological studies discussed in the module is illustrated with the following hypothetical study. Suppose we examine the proportion of lawyers in both the United States and Japan, and then the incidence of mental disease in both. We find that Japan, with only one-fourth the rate of lawyers, has only 10% of the rate of mental disease as that in the United States. Would it be a fair inference from that study that lawyering causes mental disease? The difficulty is that there may be other differences between the United States and Japan that fully explain the difference in mental disease and, indeed, without individual data, it may be that lawyers in both countries suffer mental disease at the same rate. Ecological studies can be the basis for interesting hypotheses that then should be explored in cohort and case-control studies.

In cross-sectional studies, the agent-exposure and disease status are determined in study participants at a single point in time. The problem with these studies is when exposure is not constant. Asking students about the following hypothetical study illustrates this point. Consider a survey of the class right now to determine which of you ate cold cereal for breakfast and which did not, followed by a quiz on algebra to see which kind of breakfast is most conducive to understanding algebra. What is wrong with this study design?

Slide 18 replicates Figure 1 in the module, but may be useful to instructors in pointing out the salient aspects of a cohort study design.

Slide 18 Cohort Study Design



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.

Slide 19 contains Table III-1 from the module and should be useful in connection with the study questions below.

Slide 19 Cross-Tabulation of Exposure by Disease Status

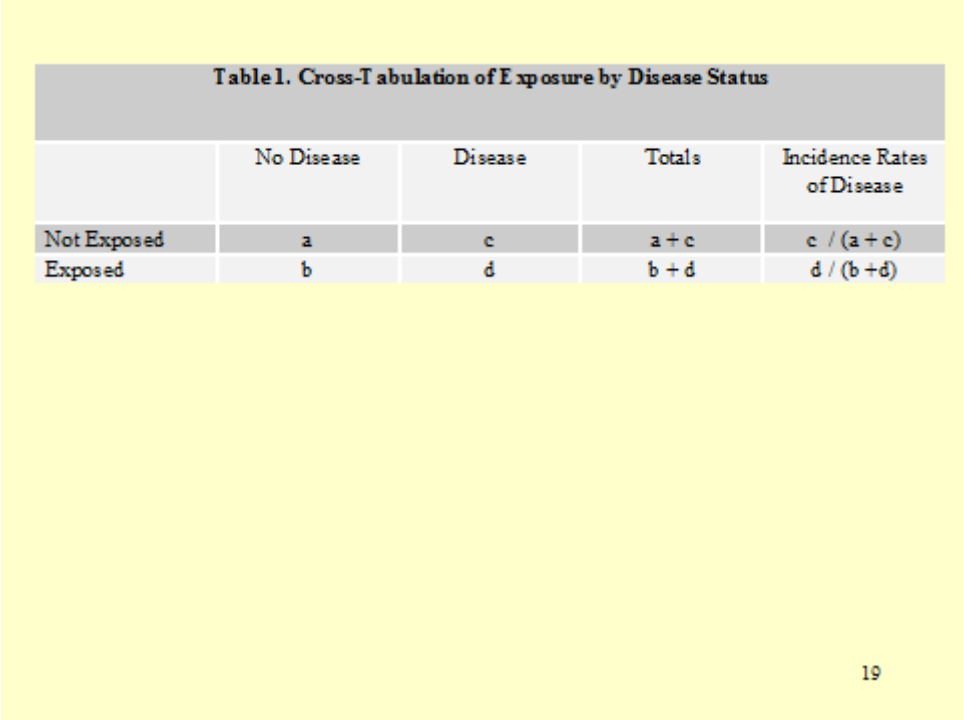


Table 1. Cross-Tabulation of Exposure by Disease Status

	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)

19

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.

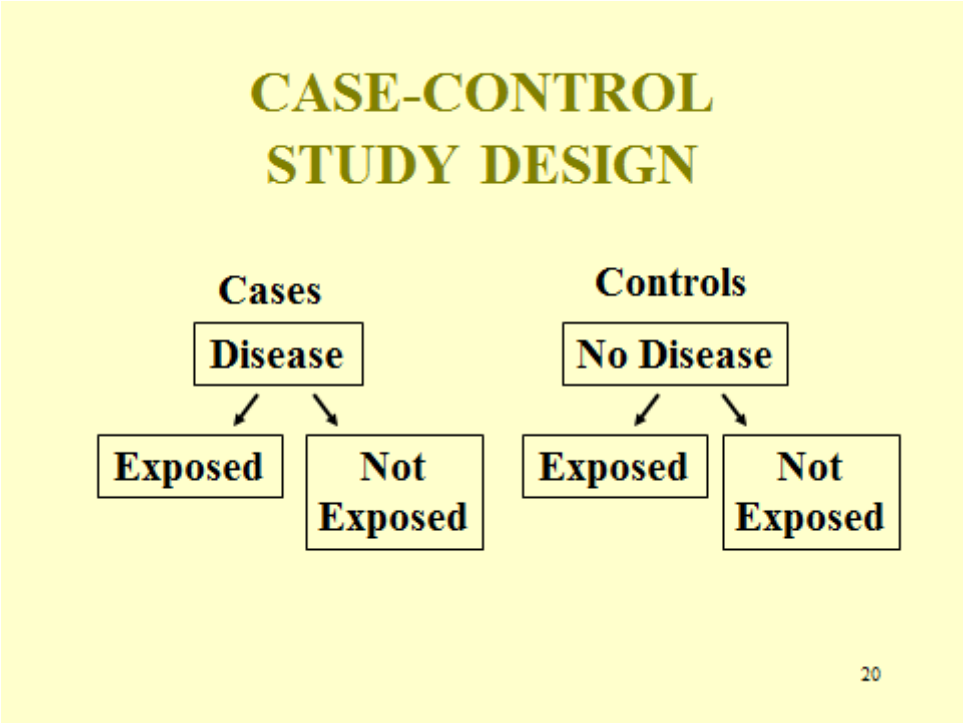
Study Questions:

1. How will the data from Table 1 be used to determine if there is an "association," that is, a difference in disease rates in those exposed and those not exposed?

   The last column contains the incidence rate for the control and exposed cohorts and we would compare them to decide whether an association exists.

   Slide 20 replicates Figure 2 in the module and, similarly, may be of use in class during discussion of a case-control study design.

Slide 20 Case-Control Study Design



SOURCE: Adapted from 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.

Instructors should spend some time in class explaining the case-control study design. It is not intuitive, and students may take some time to wrap their heads around it. It may assist students to emphasize that a case-control study starts with a cohort of "cases"—those with the disease of interest—and then follow with a question about who these cases are being compared with. Finally, what is the independent variable that researchers measure when conducting such a study? Students should appreciate that this is exposure to the studied agent. Table 2 in Slide 21 shows this format in similar form to that in Table 1 for cohort studies.

Slide 21 Cross-Tabulation of Disease by Exposure Status

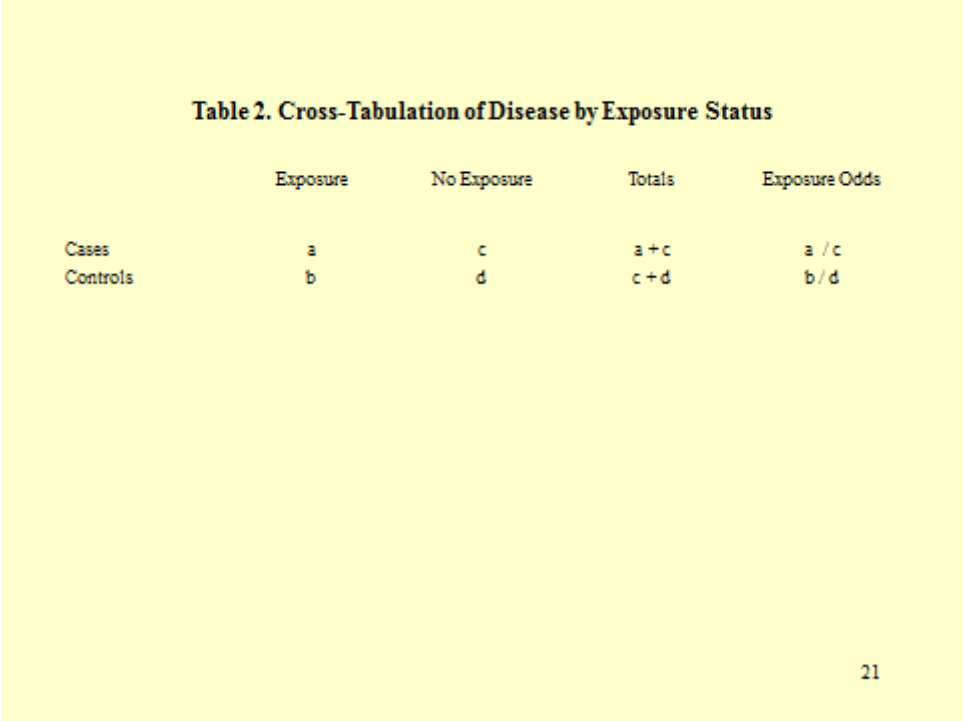


Table 2. Cross-Tabulation of Disease by Exposure Status

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

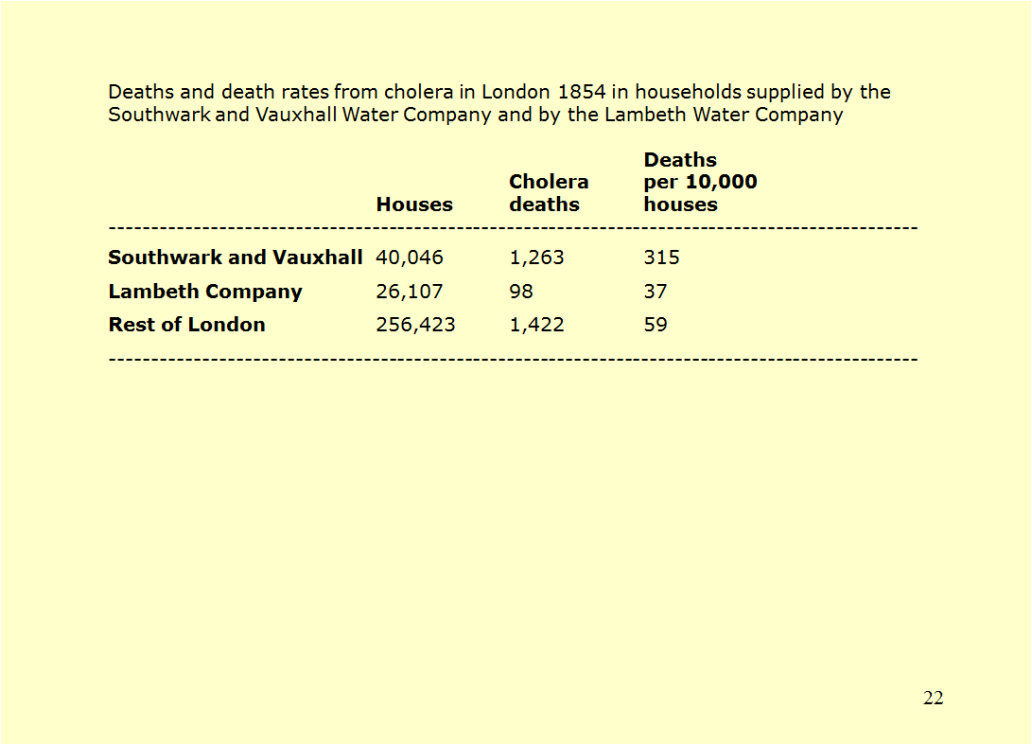
21

SOURCE: Courtesy of the authors.

Instructors may want to demonstrate how one compares the odds of exposure of the two cohorts and the way in which this becomes the cross product: a x d / b x c.

2. Consider the following data (Slide 22), which was gathered by John Snow during a cholera outbreak in London in the mid-19th century.

Slide 22 Deaths and Death Rates from Cholera in London 1854



Deaths and death rates from cholera in London 1854 in households supplied by the Southwark and Vauxhall Water Company and by the Lambeth Water Company

	Houses	Cholera deaths	Deaths per 10,000 houses
Southwark and Vauxhall	40,046	1,263	315
Lambeth Company	26,107	98	37
Rest of London	256,423	1,422	59

22

SOURCE: Republished with permission of Oxford University Press – Journals, from W. Winkelstein, Jr., "A New Perspective on John Snow's Communicable Disease Theory," *Am J Epidemiol*. 142 (9 Suppl): S8 (1995); permission conveyed through Copyright Clearance Center, Inc.

What type of epidemiologic study design would this data support? Is there a control group? Which one? What does this data suggest about the risk of contracting cholera based on the system supplying one's water?

Snow's hypothesis was that water from Southwark and Vauxhall was responsible for the cholera epidemic. So, he determined the incidence of cholera morbidity for 1,854 in that water district, Lambeth, and in the rest of London. If his hypothesis was correct, he should have found an increased incidence of morbidity in Southwark and Vauxhall compared to the other two, which served as control groups to Southwark and Vauxwall, where something indeed was in the water. The incidence of death in Southwark and Vauxhall was about nine times that of one control group, and about 5 times higher than in the second control. This data would support a cohort study.

3. Population data have been gathered that show that the frequency of mortality due

to bladder cancer for men is higher in New York City, New Jersey, New England, and industrialized areas bordering the Great Lakes. Incidence rates of bladder cancer in men are about 2 per 100,000. What kind of study is this comparison? Why? What hypothesis about the causes of bladder cancer seems appropriate to investigate? How would you carry out a study to assess the validity of your hypothesis? What is missing from the above information that you would like to know in answering the questions above?

First, instructors may want to query students to establish that the predicate for this inquiry is an ecological study, because population rather than individual data is its basis. Fashioning a hypothesis requires imagination: what is different about those geographic areas from others in the United States? One might think pollution or, alternatively, industrialization that involves the use of chemicals in production. Note one other aspect: the incidence of the effect—bladder cancer—is quite low. We are not going to find a lot of instances of bladder cancer. We address the problem of "power" below, and the difficulties of conducting a useful study when incidence is rare. Students should, nevertheless, appreciate that case-control studies are the preferable design in these circumstances. Thus, one might design a study by obtaining information about some number of bladder cancer victims in an industrialized geographic area. Once those "cases" are identified, controls who are similar to the cases but who do not have bladder cancer would be identified. Both the cases and the controls would be assessed for exposure to industrial chemicals (the investigator would have to set the measure and conditions for "exposure") perhaps initially through interviews. Recall bias in this situation would not seem to be a major problem and, thus, confirmatory evidence of exposure may not be required.

We think it too soon to give students a complete epidemiologic study to read (that will come later). But there is one abstract that instructors might want to distribute at this point for discussion in class:

Abstract from Susan S. Jick et al., *The Risk of Cataract among Users of Inhaled Steroids*, 12 Epidemiology 229 (2001).

This abstract is available at:

http://journals.lww.com/epidem/pages/articleviewer.aspx?year=2001&issue=03000&article=00016&type=abstract.

Study Questions:

1. Did the authors of this study conduct a cohort, case-control, or some other type of epidemiologic study?

   The authors state that they conducted a retrospective observational-cohort study. By gathering data about the evidence of disease in those exposed and not exposed to inhaled corticosteroidal use, they were conducting a cohort study. But they also speak of a nested case-control analysis, which may be confusing to students. But reading that part of the abstract, we hope some students will appreciate that, after gathering data on those with cataracts, the author examined those over 40 years old to determine their exposure rate to corticosteroids and compared that rate to a matched control group who did not have cataracts. Hence, the reference to a rate ratio.

2. The authors state that they "controlled" for diabetes, hypertension, and smoking. What does that mean and why do you think they did that?

   Diabetes, hypertension, and smoking are all associated with an increased risk for cataracts. Risk factors can be controlled for in two ways. First, all people with the identified risk factor may be excluded from the study entirely. Alternatively, subgroups can be created to control for each risk factor so that people with hypertension, for example, are only compared to other people with hypertension. Failing to control for risk factors can result in an observed association that is not representative of any real association.

Section III.C. Determining Exposure and Measuring Dose

Study Question:

1. Several studies show that physical exercise has two different effects on the risk of heart attacks. For those engaged in vigorous exercise, there is an increased risk of heart attack immediately following the exercise. However, regular exercise over a lengthy period protects against heart attacks. Why might this be?

   The point of this exercise (no pun intended) is to have students start thinking about biological mechanism. As the module explains, biological mechanism is important for determining the appropriate measure of exposure. Thus, for increased risk, the exposure a researcher would want to study is activity immediately preceding a heart attack, while a study of its protective effect would look at regular exercise over an extended period in proximity to a heart attack but not necessarily at that immediate moment. How would we know this before the studies described above? Through a hypothesis about the mechanisms by which heart attacks occur.

Sometimes those hypotheses are supported by good or decent evidence; sometimes they are no more than informed speculation.[3]

The current best theory about why those who exercise are at increased risk of heart attacks begins with understanding that plaque buildup in arteries can restrict blood flow through those arteries. When there is not enough blood flow to meet that demanded by tissues or organs, tissue death results (that is what a heart attack is). When exercising, the heart requires more blood flow, so someone who was not having clinical symptoms while at rest but who has sufficient plaque buildup may have a mismatch between blood supply and demand, which results in heart attack.

Exercise's cardio-protective effects are thought, but without solid evidentiary support, to result from exercise stimulating white blood cells, which make up part of plaque buildup (along with cholesterol, and platelets) to begin a process that results in those white blood cells chewing up and destroying the cholesterol inside plaque thereby reducing the size of the buildup.

Study questions in conjunction with the Rothman quotation on exposure, Section III.C. of the module:

1. Why would one be unable to measure benzo[*a*]pyrene inhaled in a case-control study?

   The primary difficulty is that a case-control study looks retrospectively at the exposure after cases have developed disease. Even if one could, in planning a follow-up study, employ a device to measure the amount of inhaled chemical (the excerpt either does not consider that possibility or discounts it as technologically or practically infeasible), this could not be done after the fact of exposure.

2. What impediments to such a measurement would exist in a case-control study?

   It seems unlikely that there would be any objective evidence of extent of exposure by a subject in the study, thus leaving the researcher with only the recollection of the subject. This raises concerns both about recall bias, which acts differentially, and measurement bias—errors by subjects in recalling the matters that Rothman identifies in the excerpt, which would be nondifferential.

3. In what kind of study would one best be able to measure the benzo[*a*]pyrene inhaled by a study subject?

[3] From 1966 to 2003, over 30 hypotheses for the mechanism of thalidomide teratogenicity were floated. See J.M. Hansen & C. Harris, "A novel hypothesis for thalidomide-induced limb teratogenesis: redox misregulation of the NF-kappaB pathway," *6 Antiox. Redox. Sig. 1* (2004).

A follow-up cohort study would minimize measurement error as subjects could be tasked with keeping track of the number and type of cigarettes smoked. While measurement error would not be eliminated, it would be reduced over what would occur in a retrospective study.

4. Why would one want exposure information for "a period covering many decades"?

   As the excerpt states, we do not know what the relevant dose of cigarette smoke is for developing disease. Is it: (1) peak dose; (2) total dose, that is, dose/time period times total time of exposure; (3) total dose only during the early years of exposure; (4) for those who stop smoking, how does cessation affect the risk due to prior exposure; or (5) some other measure of dose?

5. Assume that a follow-up study examined the relationship between higher education and brain cancer. Participants consisted of those who had just graduated from high school and "followed" for 10 years to determine the incidence of brain cancer in those with 1, 2, 3, and 4 years of higher education as well as those who did not attend college. The researchers reported that they found no association, that is, no difference in disease among all 5 groups. What concern(s) would you have about the validity of the study?

   The main point of this question is that, given the long latency period before almost all cancers develop, the study did not follow participants long enough to discover whether cancer develops from higher education. The Rothman quotation in the module makes this point. Study design requires taking into account all that is known biologically about the disease, and we know that the latency period for cancer is typically in the decades.

   Students may also raise difficulties with measuring how much education study participants actually received. Some measurement error may occur, but there is no reason to believe that it is systematically skewed in one direction or another.

The module refers to occupational studies in which investigators employ a specific job as the independent variable rather than specific suspected toxins. This, of course, creates difficulties in toxic-tort cases because there may be exposure to multiple potential toxins at a job site and each is a competing cause of disease, unless there is an evidentiary basis for eliminating one or more as a general cause, which is rare. These difficulties are revealed in *Knight v. Kirby Inland Marine Inc.*, 482 F.3d 347, 353 (5th Cir. 2007), in which the court affirmed the exclusion of plaintiff's causation expert in part because studies on which he relied were occupational studies that did not isolate and examine each of the potential toxic chemicals to which employees were exposed:

The other case-control study, L. Hardell et al., [plaintiffs' expert] relied on for his testimony on Hodgkin's lymphoma suffers from similar flaws. That study focused on organic solvents as a class, including a wide-range of chemicals to which appellants were never exposed. Of all the organic solvents the study controlled for, it could not determine which led to an increased risk of cancer. Thus, the study concluded that "[i]t is possible . . . that exposure to organic solvents may be relevant for the development of Hodgkin's disease" and that "it is possible that exposure to organic solvents may promote the development of Hodgkin's disease irrespective of other, still unknown, etiologic factors." See L. Hardell & N.O. Bengtsson, "Epidemiological Study of Socioeconomic Factors and Clinical Findings in Hodgkin's Disease, and Reanalysis of Previous Data Regarding Chemical Exposure," *Br. J. Cancer* 48, 218 (1983). We cannot say that the district court's assessment of this study was clearly erroneous or that its exclusion of it as unreliable for general causation was an abuse of discretion. The study does not provide a reliable basis for the opinion that the types of chemicals appellants were exposed to could cause their particular injuries in the general population.

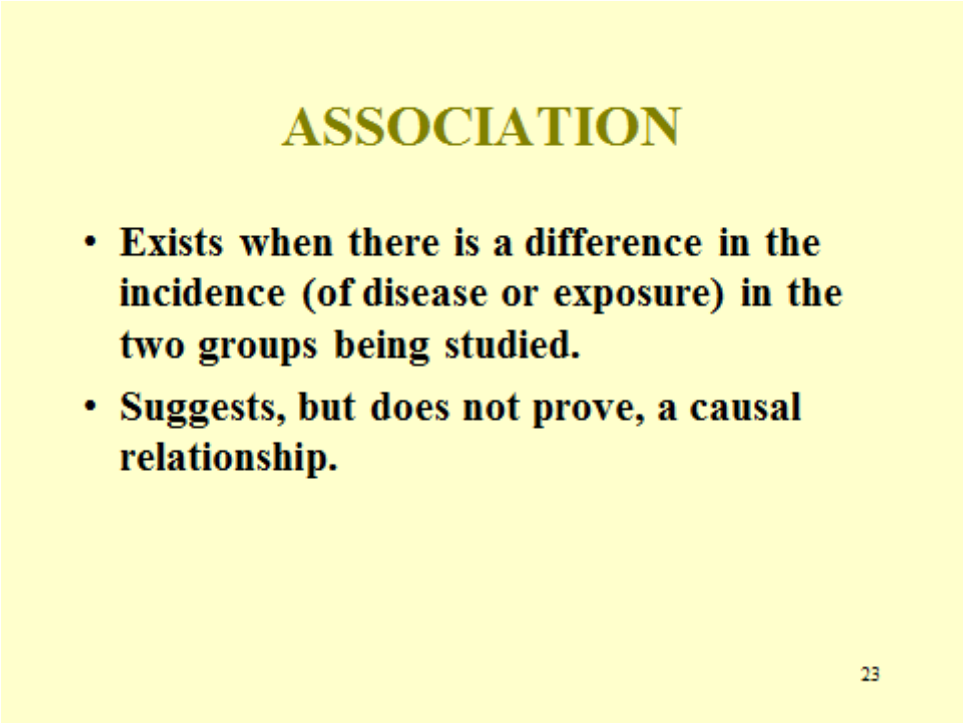
Section III.D. The Outputs of Studies: Relative Risks, Odds Ratios, Attributable Proportion of Risk, and Standardized Rate Ratios

Study Question:

1. What does the existence of an association mean with regard to the existence of a causal relationship?

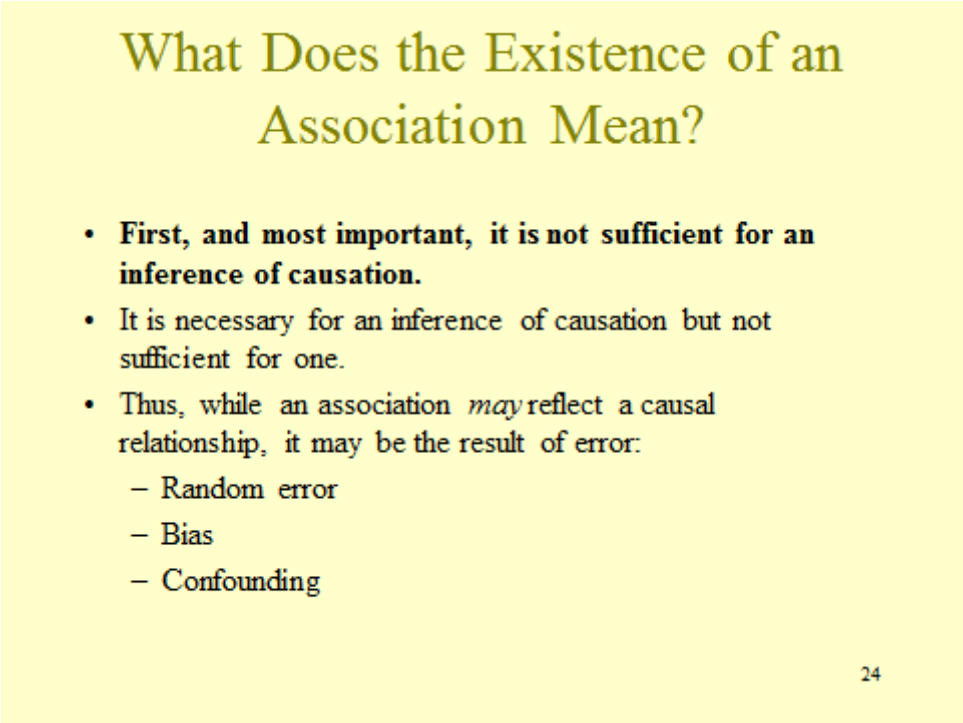
   An explanation of what an association is—simply a difference in disease between exposed and unexposed cohorts in a cohort study or a difference in exposure rates between cases and controls in a case-control study—is contained in Slide 23.

Slide 23 Association



In the introductory paragraph to the Outputs of Studies section, the connection between association and causation is discussed—this deserves some emphasis as nonscientists often tend to reason directly from association to causation. This entails the logical fallacy known as *post hoc ergo propter hoc*. Slide 24 is designed to assist in making this point in class.

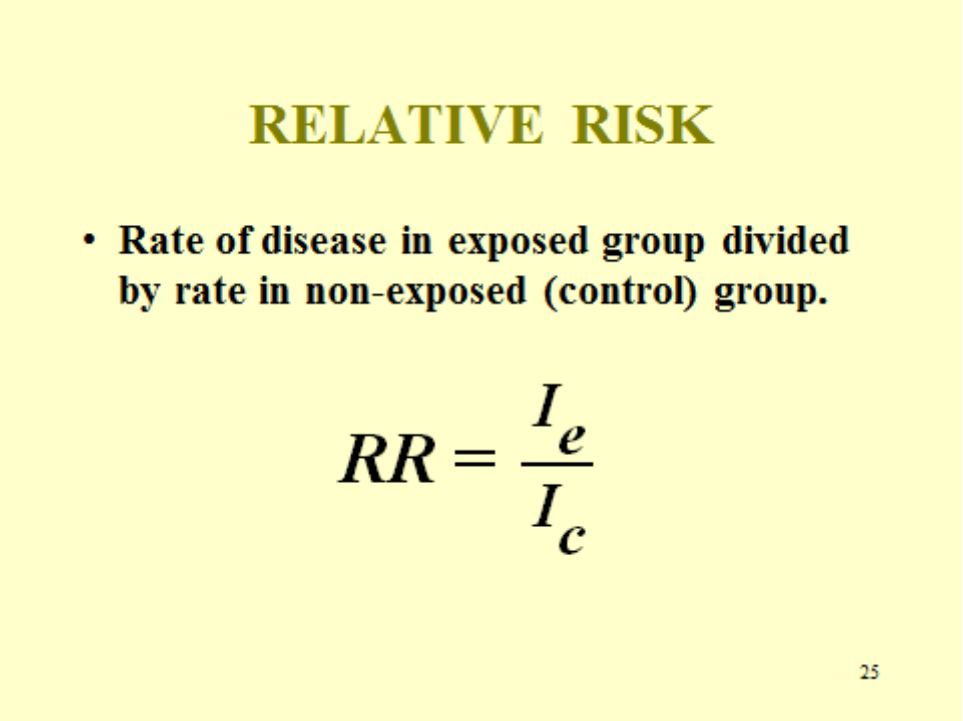
Slide 24 What Does the Existence of an Association Mean?



Numerous cases make this point, e.g, *Henricksen v. ConocoPhillips Co.*, 605 F. Supp. 2d 1142, 1176 (E.D. Wash. 2009); *Sharkey v. Sterling Drug, Inc.*, 600 So. 2d 701, 712 (La. Ct. App. 1992) ("As you are well aware, an epidemiologic association does not prove causation. However, causation may be strongly indicated by association when no other explanation for the relationship is found . . . ."); *Cornell v. 360 West 51st Street Realty*, LLC, 9 N.E.3d 884, 898 (N.Y. 2014) ("[The expert witness] repeatedly equated association with causation. In so doing, he departed from the generally accepted methodology for evaluating epidemiologic evidence when determining whether exposure to an agent causes a harmful effect or disease."). If you think legal precedent will be helpful in emphasizing this point (it is remarkable how our sense of legal precedent results in courts citing other cases for epidemiologic principles, even when scientific sources are readily available), feel free to construct a slide using these or similar cases.

We begin the discussion of study outputs with relative risk, a straightforward concept employed for cohort studies. Students should not have great difficulty understanding the concept or even, given incident rates, calculating one. The formula for calculating a relative risk is contained in Slide 25. Instructors should ensure that students understand what these variables mean and what they are doing when they calculate a relative risk.

Slide 25 Relative Risk



Slide 26 demonstrates from raw data drawn from Table 1 how the variables in Slide 25 are determined.

Slide 26 Cross-Tabulation of Exposure by Disease Status

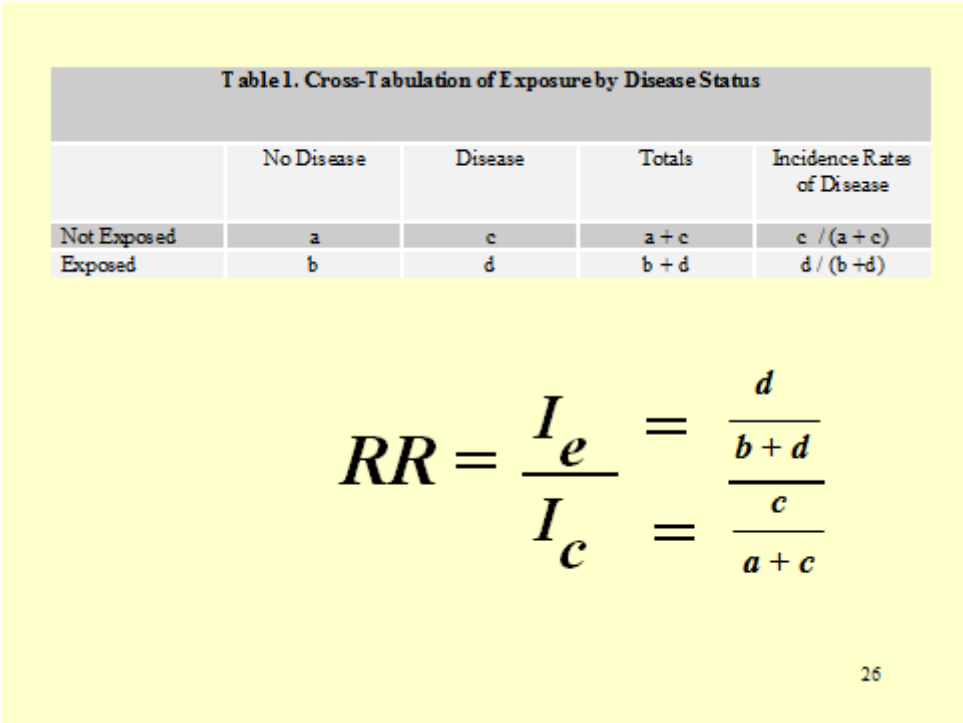


Table 1. Cross-Tabulation of Exposure by Disease Status

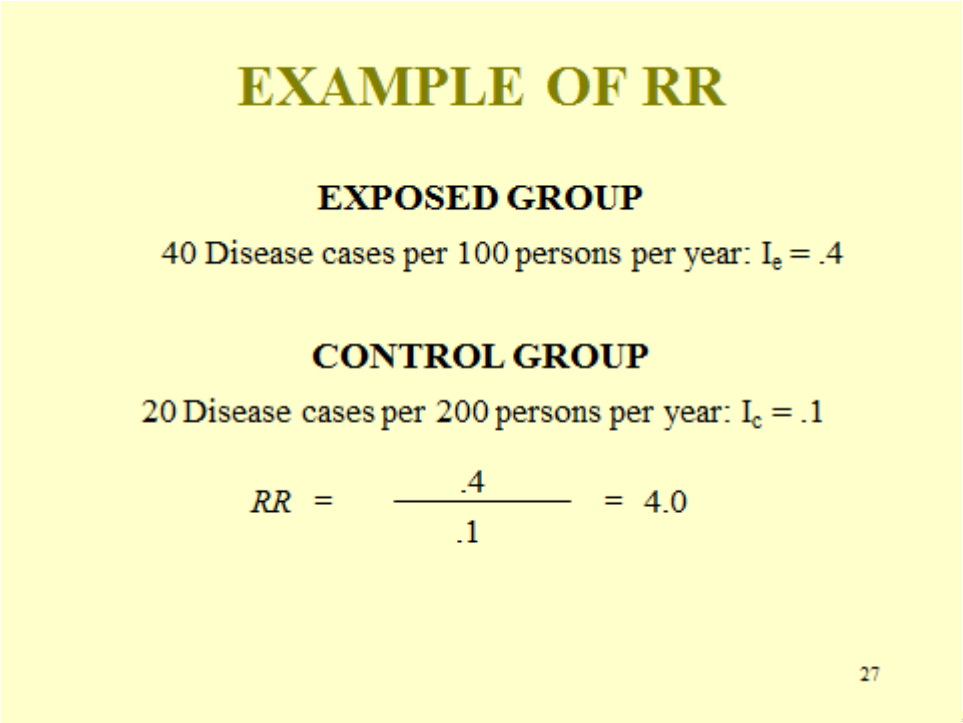
	No Disease	Disease	Totals	Incidence Rates of Disease
Not Exposed	a	c	a + c	c / (a + c)
Exposed	b	d	b + d	d / (b +d)

$$RR = \frac{I_e}{I_c} = \frac{\frac{d}{b+d}}{\frac{c}{a+c}}$$

26

We run through the example provided in the module in class using Slide 27. Instructors may want to assign an additional problem for students to work through on their own.

Slide 27 Example of RR



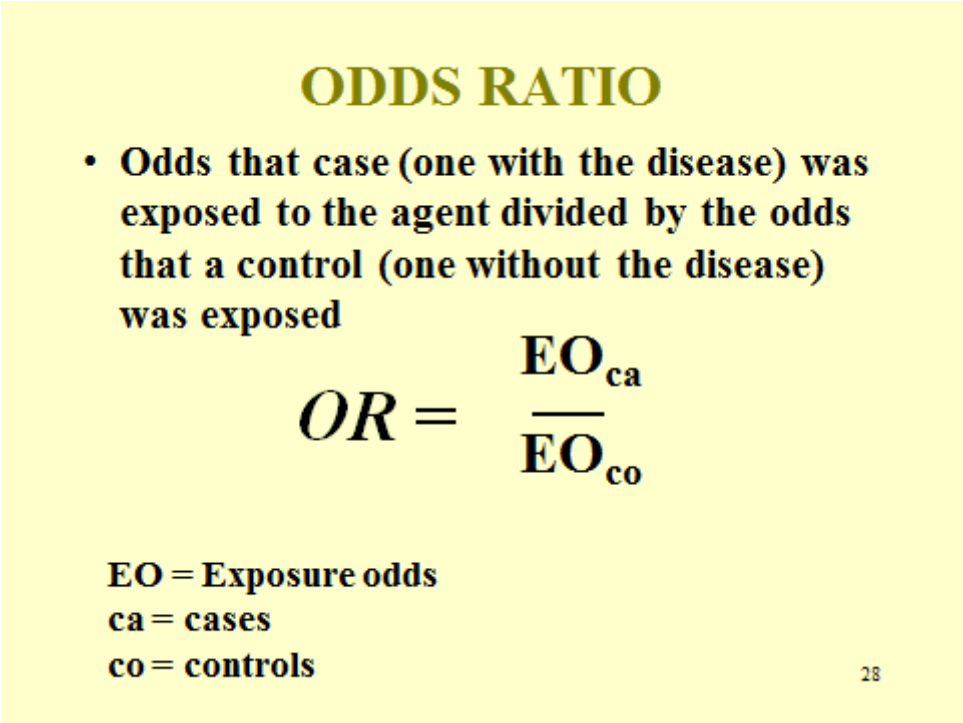
Study Question:

1. Recall the concept of a "signature disease," discussed supra Section III.D.3. What is the relative risk for a signature disease?

   The relative risk for a signature disease would be infinite because the incidence of disease in the unexposed group would be zero—recall that a signature disease, by definition, requires exposure to the signature agent before the disease occurs. Those in unexposed cohort, therefore, would not contract any of the disease in question. The important point to appreciate is that the relative risk is so high that, once the agent is identified—think Thalidomide or DES—it explains all (or virtually all) of the disease that is found.

   The odds ratio, the predominant measure of association used for case-control studies, is explained in the module and in Slide 28.

Slide 28 Odds Ratio



Slide 29 shows the determination of the two exposure odds variables (for the cases and controls) based on the raw data in Table 2.

Slide 29 Cross-Tabulation of Disease by Exposure Status

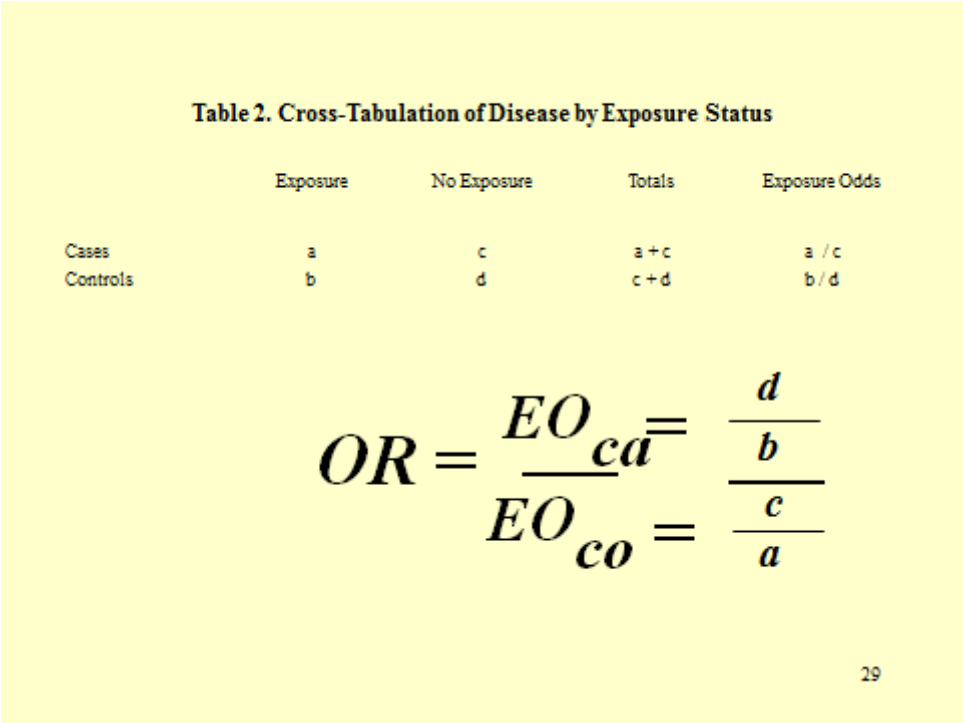


Table 2. Cross-Tabulation of Disease by Exposure Status

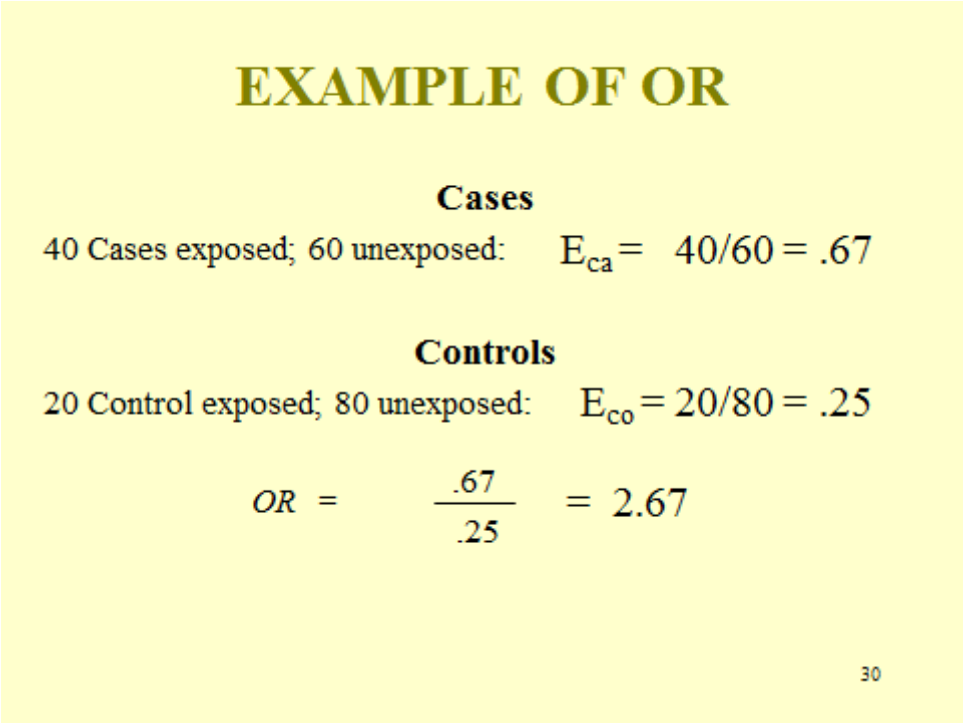
	Exposure	No Exposure	Totals	Exposure Odds
Cases	a	c	a + c	a / c
Controls	b	d	c + d	b / d

$$OR = \frac{EO_{ca}}{EO_{co}} = \frac{\frac{d}{b}}{\frac{c}{a}}$$

29

We conduct the arithmetic required to calculate an odds ratio from the data in Table 6 in Slide 30. Again, assigning another problem should assist students to better absorb this idea.

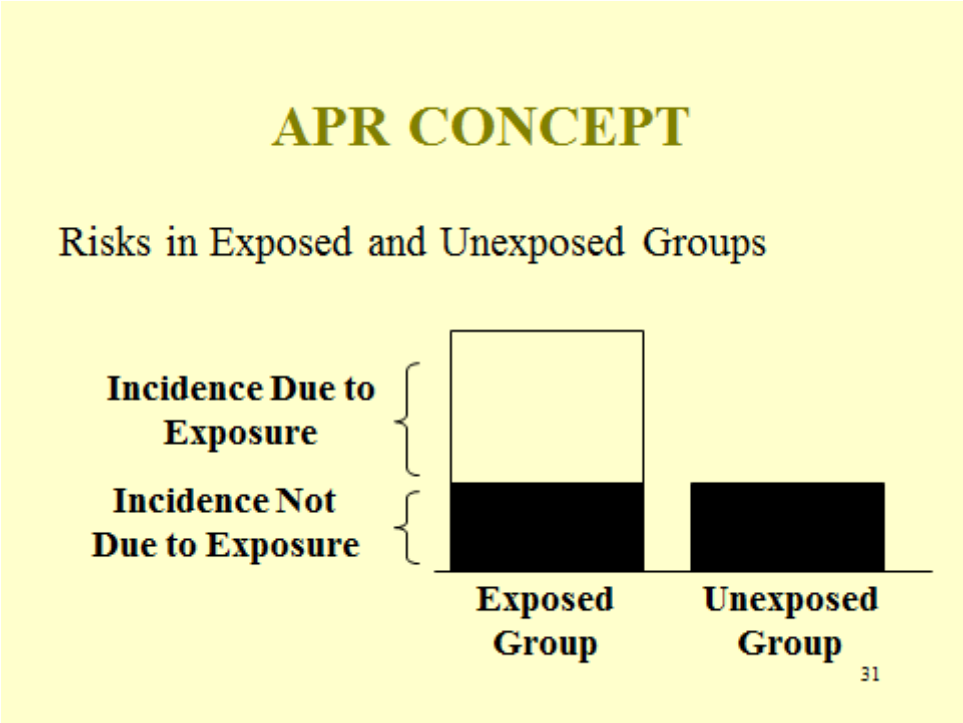
Slide 30 Example of OR



The attributable-risk measure has important public health uses. We explain it here both because epidemiology studies frequently report the AR and because it is helpful when we get to specific causation and the role that epidemiology can play in proving specific causation.

Slide 31 is designed to present the AR concept visually so as to enable students to grasp what AR is about.

Slide 31 APR Concept

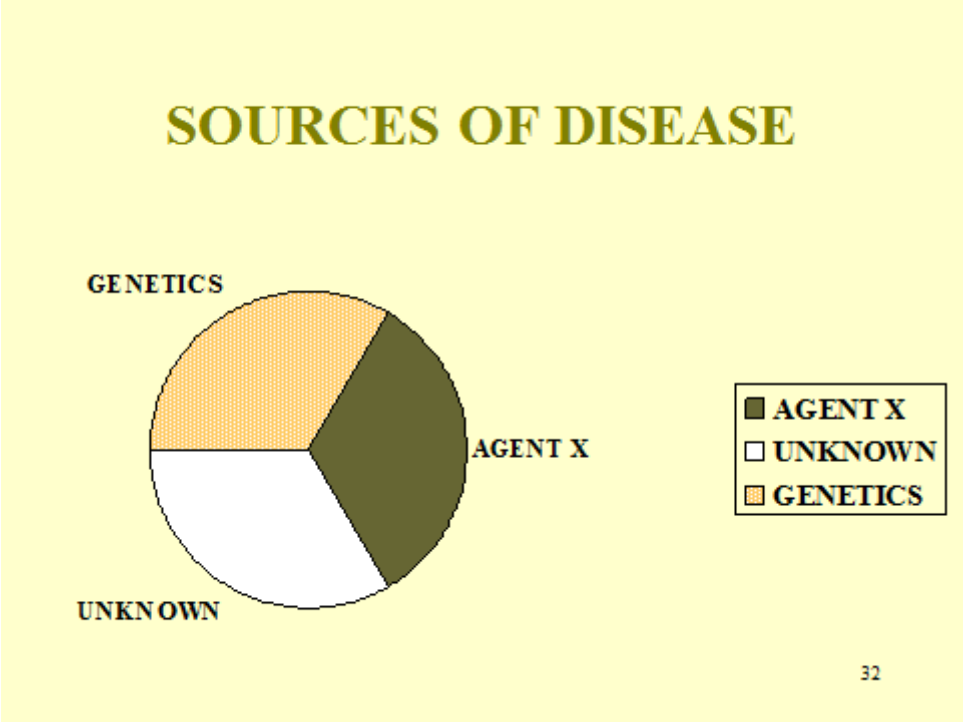


SOURCE: National Research Council. *Reference Manual on Scientific Evidence: Third Edition*. Washington, DC: The National Academies Press, 2011, p. 570. Copyright © 2011 National Academy of Sciences.

The column on the left reflects the incidence of disease in the exposed cohort of a cohort study. The incidence of disease in the control group is the column on the left, which is also transported over to the exposed column so that students can see that the darkened portion of the left column reflects the incidence of disease that would have occurred in the exposed cohort due to causes other than the agent. The white remainder of the left column reflects the proportion of disease attributable to agent exposure and is the attributable risk.

Attributable risk is also shown in a pie chart in Slide 32. We posit that genetics, Agent X and an unknown cause or causes are responsible for all of the disease. The segment of the pie labeled “Agent X” is the attributable risk due to Agent X.

Slide 32 Sources of Disease



In connection with Slide 32, students could be asked what the AR is if all three segments are the same size. Since all are equal, they each reflect a 33% association with the disease and, thus, the AR would be 33%. The attributable proportion of risk expresses the percentage of disease in the exposed population that is due to the agent of interest. Here that is .33, or to put it another way, 33% of this disease in society is caused by Agent X.

Study Question:

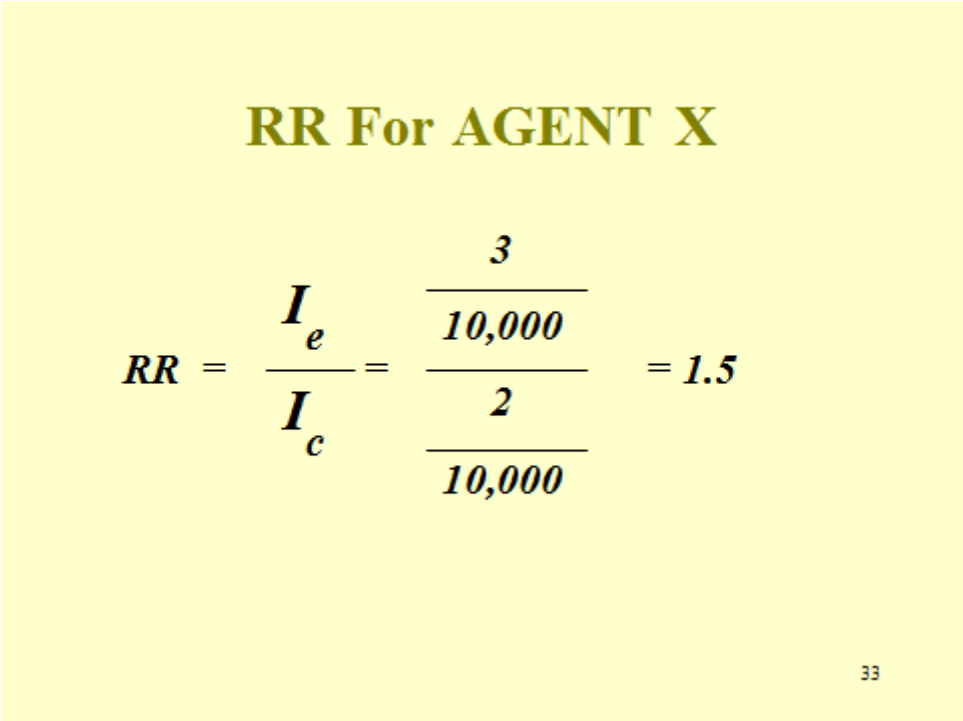
1. What is the relative risk for Agent X in Slide 32?

   One-third of the disease in the pie chart is due to exposure, which means that among those unexposed, the incidence of disease would be two-thirds that of the exposed group with genetics causing half and unknown causes responsible for the other half.

   In the exposed group, the incidence of excess disease would be one-half of what it was in the unexposed group. This is shown in Slide 32, where the incidence of disease in those exposed to agent X is 3/10,000, while the incidence in those unexposed is 2/10,000, which results in a relative risk of 1.5. This then reveals the formula for calculating an AR from the relative risk in Slide 33, which is keyed to the pie chart in Slide 31. The relative risk for Agent X is 1.5—that is to say, the incidence

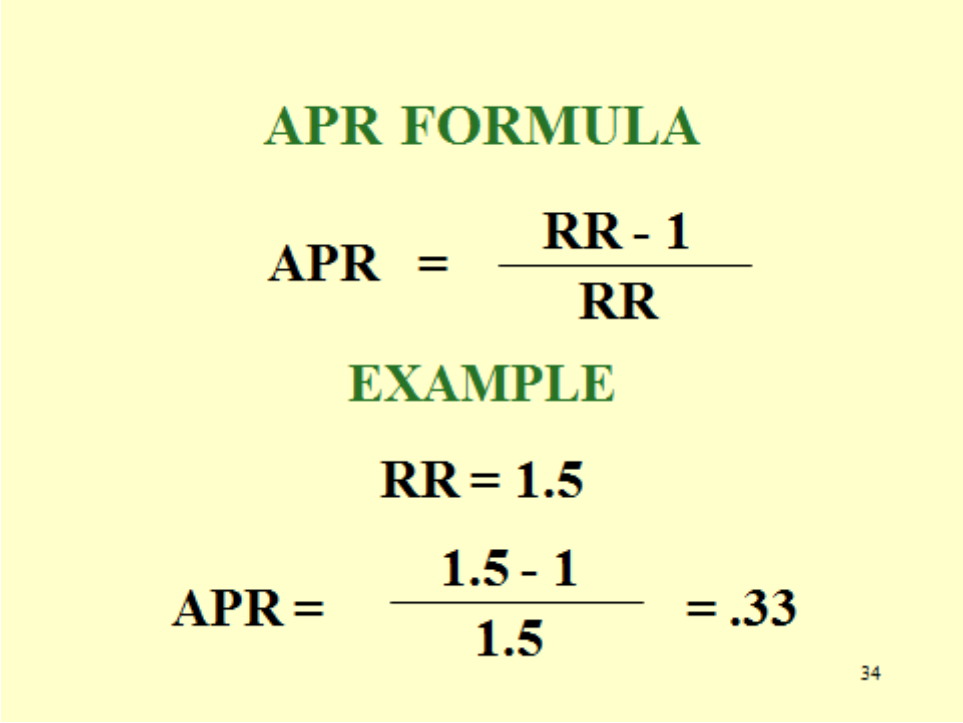
of disease of a group exposed to Agent X will be 50% higher than the incidence in a control group, which does not have exposure to Agent X. We work through this exercise again, graphically, in Slides 144 through 147, below.

Slide 33 RR for Agent X



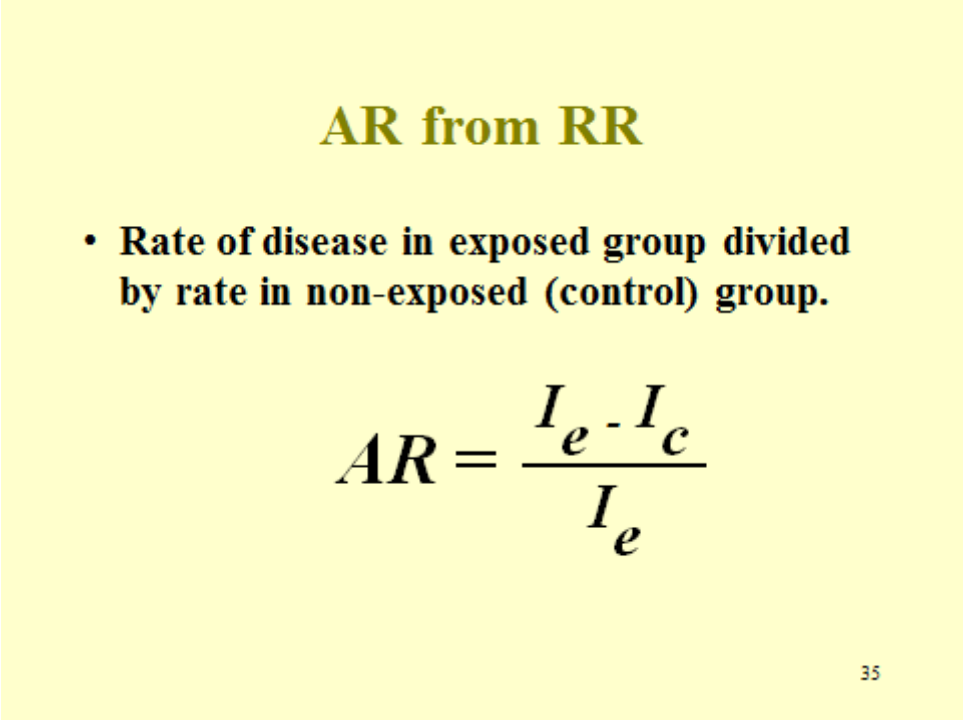
Slide 34 shows the formula for determining the attributable risk from the relative risk. Instructors would want to explain, in presenting Slide 33, why this formula is what it is: Recall that AR reflects the proportion of risk of a disease in a group or society. RR – 1 reflects the excess of risk of disease due to the agent in question. That excess is then divided by the total risk for disease in the population; in other words, the attributable risk provides the proportion of risk due to the agent.

Slide 34 APR Formula



Another way to illustrate this would be to work from the incidence of disease in the exposed and unexposed cohorts as contained in Slide 35. One could show how the formula in Slide 34 is identical to the formula in Slide 35 by substituting for the RR in Slide 34 the equivalent I_e / I_c and employing a bit of arithmetical manipulation—but, unless students are curious, we are not inclined to set off on this collateral matter.

Slide 35 AR from RR



The following abstract should be distributed:

Abstract from Ana M. Garcia et al., "Parental Agricultural Work and Selected Congenital Malformations," 149 *Am. J. Epidemiology* 64 (1999), available at http://www.ncbi.nlm.nih.gov/pubmed/9883795.

Study Questions:

1. What is the difference between the odds ratio the authors report in the abstract above and a relative risk? Why did the authors use an odds ratio to report their results?

   As explained in the module, because a case-control study does not measure the total population of those exposed to the agent from which the cases emerge, nor the total number of unexposed, an incidence rate cannot be calculated. Instead, case-control studies employ an odds ratio that also addresses whether an association exists and, when the disease in question is rare, numerically approximates the relative risk.

2. What weaknesses exist in this study (as reflected in the abstract) that might be exploited by a defendant manufacturer of a pesticide if sued by a child born with a cardiovascular birth defect who seeks to prove causation based on the study?

   There are two problems here. First, this study examines occupational exposure to any (and all) agricultural pesticides. Different pesticides may have different teratogenic profiles, including not being one. By lumping all exposures together, the study does not permit determination of whether a specific pesticide is a teratogen and, if so, how much it increases the risk. Caveat: If all pesticides use the same active ingredient in comparable concentrations, this concern may well be negated. The second problem is that the abstract does not report separately on the magnitude of the odds ratio for cardiovascular defects, only stating that other types of defects were primarily responsible for the increased risk. However, reference to the full study (in Appendix A) reveals that exposure among case-mothers was 4, while among control-mothers was 2, resulting in an odds ratio of 2. Although the authors did not report a p-value, it was almost certainly considerably above .05.

3. Design your own case-control study (by "design," we mean only pick a disease and an agent or agents that are suspected to cause the disease) of a disease or outcome in which you are interested or make up yourself. Provide made-up data that you might have found while conducting the study, then calculate the odds ratio and the attributable proportion of risk. What is the relationship between the odds ratio you found and the relative risk for the same association? Be prepared to present your work in class.

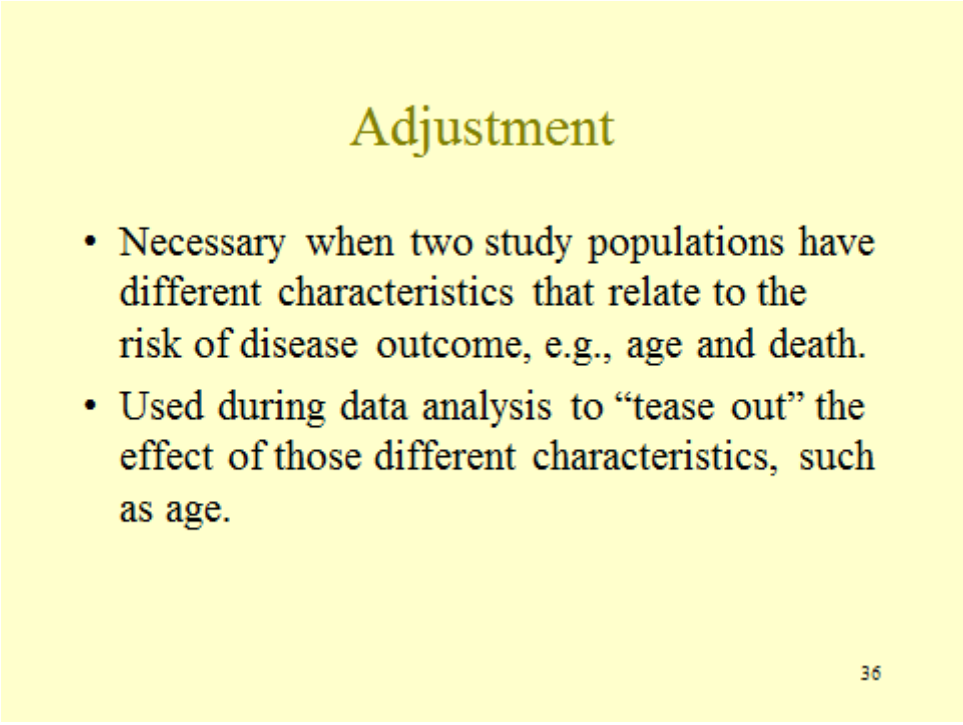
   We assigned this project on a group basis in advance of class with the instruction that students should come to class prepared to present their project. At the beginning of class, we paired off two groups and each of the two groups presented their study to the other group for discussion and critique before selecting a few groups to present their study to the whole class during which questions and critiques were encouraged. We found this exercise very helpful. For some students, unclear about the difference between a cohort and case-control study, the difference was brought home. For others, the study design foreshadowed issues of selection bias and exposure measure that will heighten appreciation for these matters when we get to them. In a couple of cases, students explored a study in which they found a protective effect, and this afforded an opportunity to discuss the meaning of an odds ratio (or relative risk) of less than 1.0, that is, a negative association that implies the agent has a protective effect against disease—precisely what researchers hope to find in clinical trials of new drugs or medical treatments. Finally, a number of students foundered in calculating an APR from their odds ratio because the formula provided in Slide 34 requires incidence measures, which are

not available in case studies. Thus, to calculate an AR for a case study and an odds ratio, students must recall that the odds ratio approximates the relative risk when the incidence of disease is low. Thus, substituting the odds ratio for the relative risk in the formula in Slide 33 would enable an estimation for the AR. If the incidence of disease in a study is substantial, the odds ratio and relative risk will no longer be a close approximation, but rather the odds ratio generally becomes progressively larger than the relative risk as the incidence of disease becomes larger.

Section III.E. Adjustment for Study Groups that Are Not Comparable: Standardized Mortality and Morbidity Ratios

Our goals are relatively modest for this subject. We want students to understand the need for adjustment and why certain studies require adjustment. We also want them to appreciate that standardized mortality and morbidity ratios are equivalent to relative risks. Slide 36 serves as a brief explanation of adjustment and the opportunity to discuss what types of studies will require adjustment. Because we have not yet covered confounding, explaining the relationship of these risk factors to confounding seems premature, but the instructor may want to alert students that we will return to this problem of different risk factors existing in comparison-study groups, an avoidable risk when random selection is unavailable.

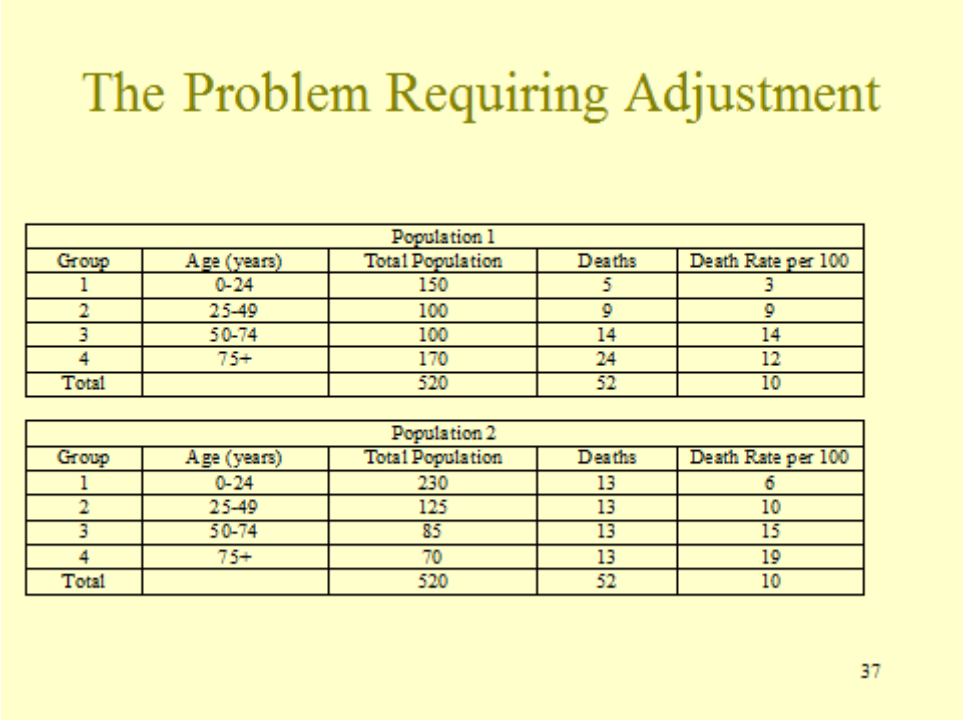
Slide 36 Adjustment



Before proceeding, students should appreciate that epidemiologic studies that involve matched cases and controls or exposed and unexposed cohorts will not require adjustment (and therefore not report SMRs) because the process of matching removes differences between the two study populations that affect risk. When, however, disease incidence is determined through a reference population (e.g., based on the state as suggested in the module), there can be a difference in risk characteristics between the two study groups. Ecological studies that examine disease in populations and do not gather individual data are the most susceptible to requiring adjustment. Indeed, the example used in the module about death rates in Alaska and Florida is an ecological study.

Slide 37, which consists of two tables from the module, brings home why adjustment is required. It reveals that the crude incidence rate of death in Florida and Alaska is the same.

Slide 37 The Problem Requiring Adjustment



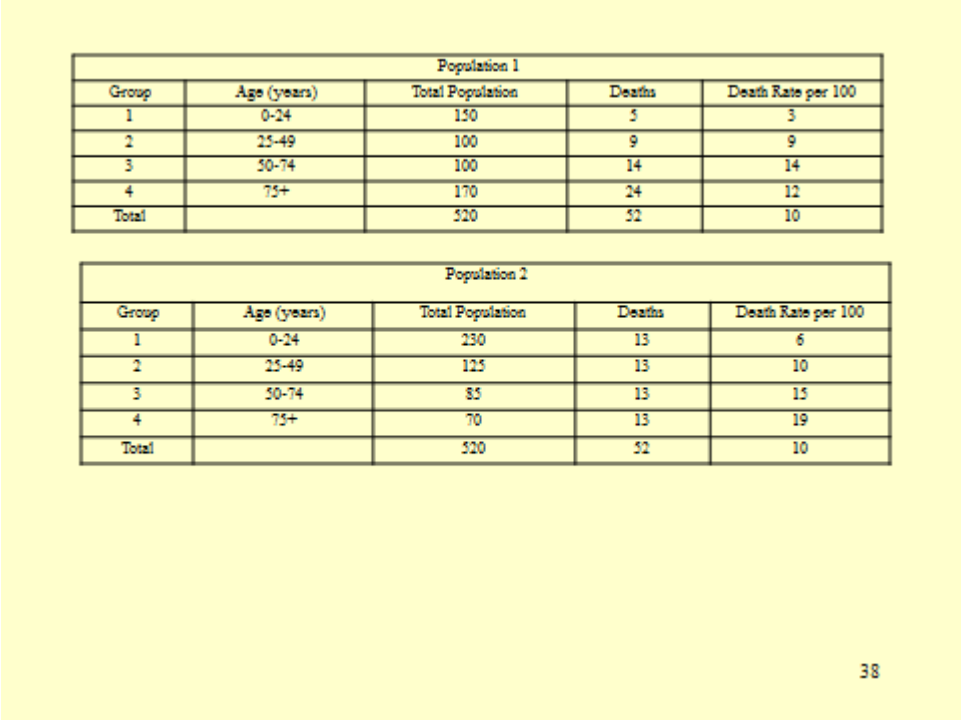
Population 1				
Group	Age (years)	Total Population	Deaths	Death Rate per 100
1	0-24	150	5	3
2	25-49	100	9	9
3	50-74	100	14	14
4	75+	170	24	12
Total		520	52	10

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

We might observe in class with this slide displayed: “The rate of death in each population is the same, so if these were study populations where we were examining the effect of contaminated water—to which Population 2 was our exposed group and Population 1 was our unexposed group—we would say that we do not have an association. With no association, we should move on to our next study and hope that we find something in that one because this one is a dead end. Right?”

Wrong, as Slide 38 (drawn from the next two tables in the module) reveals.

Slide 38 Age Distribution Across 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

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

38

As the module observes, Population 2 has a higher age-specific death rate in each identified age subgroup. How can this be? Population 2 should have a higher overall crude death rate than Population 1 because Population 2 has a higher age-specific death rate in every single age group. We hope that students will see that the problem is that there are very different age distributions in the two populations and the crude rate in Population 2 is driven down because of the small population of older people. Thus, because of this skewed population distribution, the crude rates of mortality are not comparable. When we stratify mortality by age, as in Slide 39, we see that death rates in each age group in Population 2 is higher, revealing that there is an association between exposure to contaminated water and mortality.

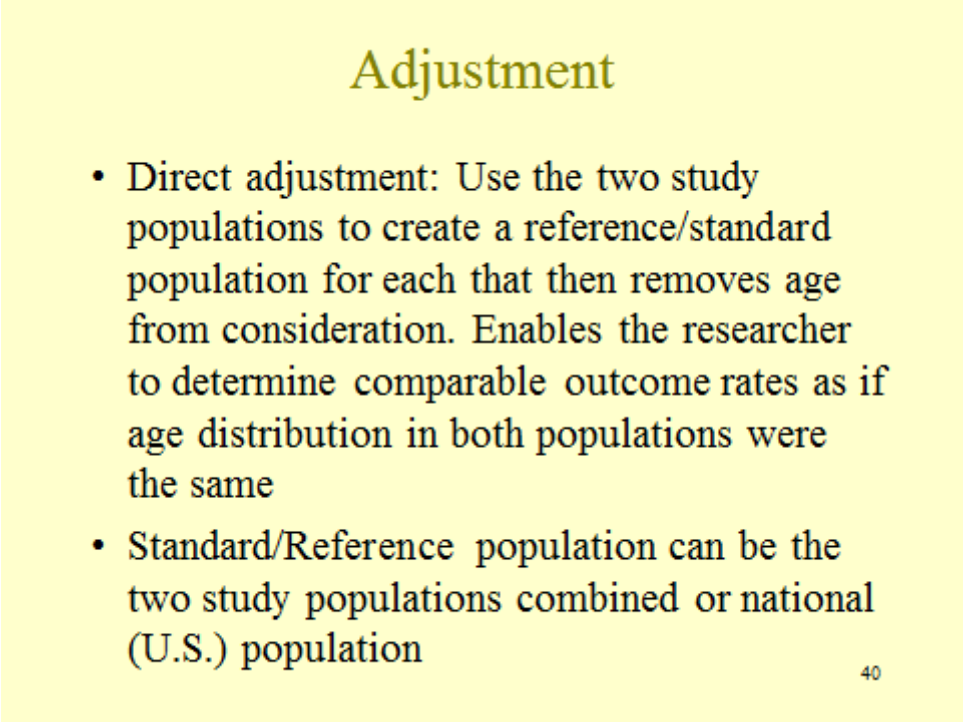
Slide 39 Death Rates in Population 1 and Population 2

Group	Age	Population age-specific death rate	Population 2 age-specific death rate
1	0-24	3	6
2	25-49	9	10
3	50-74	14	15
4	75+	12	19

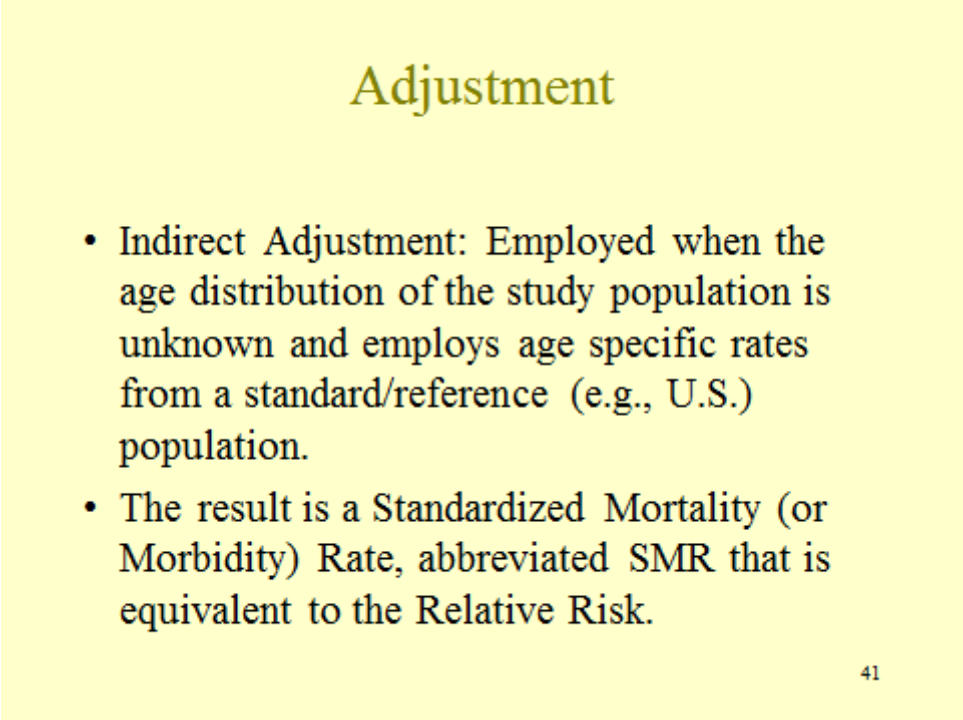
39

Direct adjustment is explained in Slide 40 and indirect adjustment in Slide 41.

Slide 40 Direct Adjustment



Slide 41 Indirect Adjustment



Study Questions:

1. What difficulty exists with comparing crude death rates? What can be done to minimize this problem?

   Crude death rates do not take into account any differences in distributions of risk factors in a population. For example, if two populations have very different age distributions, it may appear that one population has a drastically higher death rate at first glance. When one of the populations has a higher percentage of elderly people, that population is likely to have a higher death rate than one with fewer elderly people. The populations can be adjusted according to age so that the comparable death rate can be determined for each population as though each had the same age distribution.

2. What are the goals of adjustment?

   Adjustment allows researchers to appropriately combine data, make appropriate comparisons among groups, and make accurate inferences so that appropriate

conclusions about whether an association exists may be drawn.

3. Remember that specific rates can be calculated for characteristics other than age. What characteristics do you think may be helpful in an analysis of mortality?

   It can be helpful to adjust for any factor that affects the risk of death or disease. These factors include age, gender, race, occupation, and income. Populations often differ in the distribution of these factors. Adjustment allows us to make valid comparisons of two populations that have different compositions.

4. For instructors who want to go into more detail or assign an adjustment problem, we lay out another adjustment hypothetical below.

   Consider the age-adjusted mortality rate for partners in law firms based on gender.

		Male Partners			**Female Partners**		
		1	2	3	4	5	6
Group	Age	Population	# of Deaths	Death Rate per 100	Population	# of Deaths	Death Rate
1	0-24	5,000	115	2.3	3,000	60	2
2	25-49	20,000	800	4	7,000	170	2.4
3	50-74	4,000	2,200	55	6,000	2,000	33.3
4	74+	200	180	90	900	470	52.2
Total		29,200	3,295	11	16,900	2,690	16

SOURCE: Courtesy of the authors.

Crude death rates per 100 people are:

Male: 11
Female: 16

Comparison of the crude rates suggests that female partners are more likely than male partners to die. However, the age-specific death rates per 100 are lower for female partners in each age group. How is this possible? Look at the age distributions in the two populations. Over 85% of male partners are under the age of 50 while less than 60% of female partners are under the age of 50.

Group	Age	Male death rate per 100	Female death rate per 100	Reference population	Expected deaths- male	Expected deaths- female
1	0-24	2.3	2	8,000	184	160
2	25-49	4	2.4	27,000	1,080	648
3	50-74	55	33.3	10,000	5,500	3,330
4	74+	90	52.2	1,100	990	574
Total				46,100	7,754	4,712

	Male Partners	Female Partners
Age-Adjusted Rate	$\frac{7,754}{46,100}$	$\frac{4,712}{46,100}$
Deaths/ 100 population	16.8	10.2

SOURCE: Courtesy of the authors.

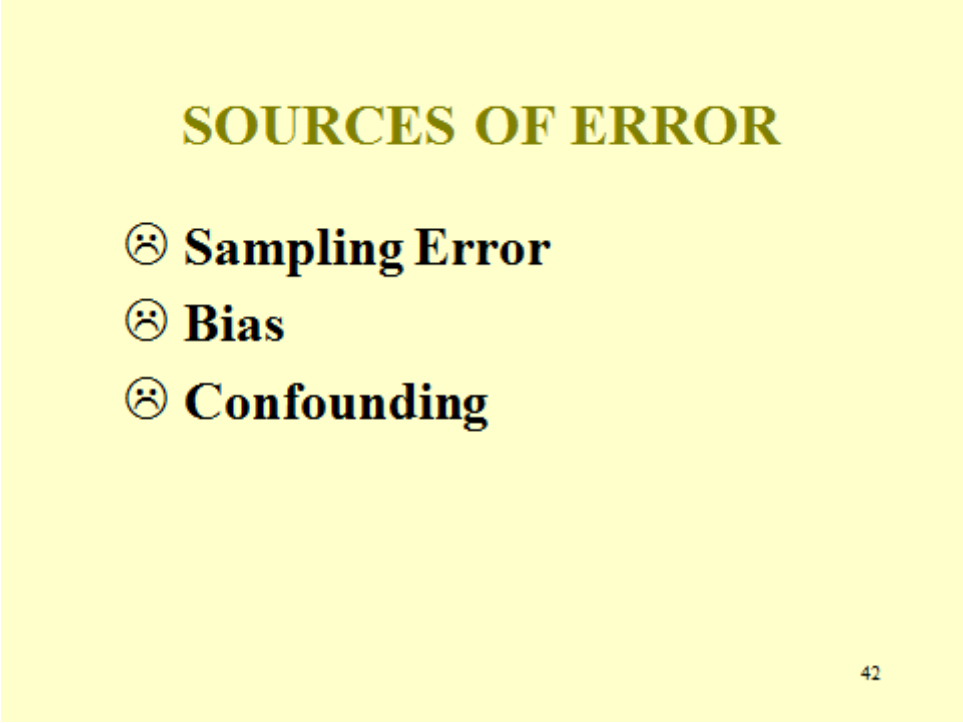
The age-adjusted rate is 16.8 for male partners and 10.2 for female partners. Unlike the crude death rates, males now have a higher mortality rate than females.

(Ie/ Ic – Ic/Ic)

SECTION III.F. SOURCES OF ERROR

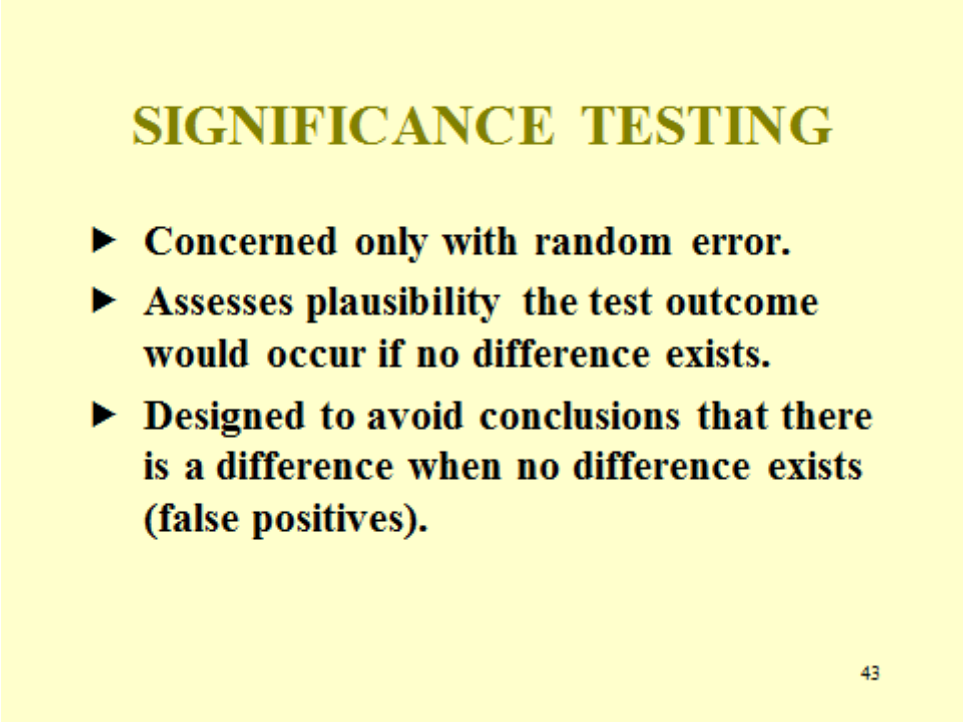
We begin with a short overview of the sources of error in an epidemiology study. Slide 42 merely lists these if the instructor wants to have a visual while covering the preliminaries on these concepts.

Slide 42 Sources of Error



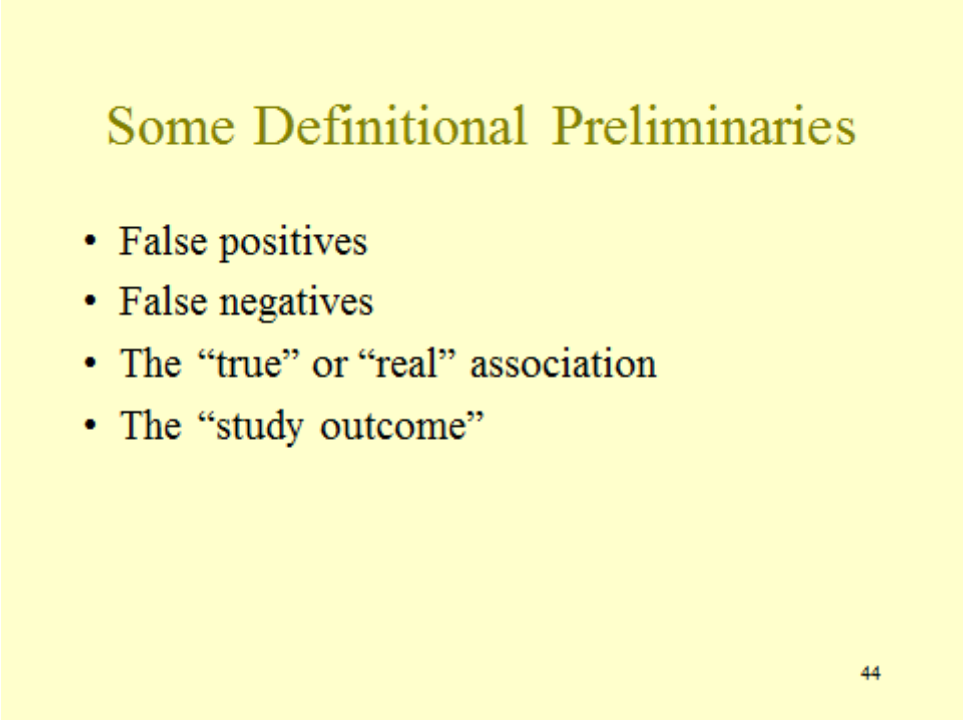
The slide on significance testing covers the basics of what significance testing is about. Many students will have had some exposure to this concept from a social sciences statistics course. Instructors may want to revise or supplement Slide 43 to suit their individual purposes.

Slide 43 Significance Testing



Slide 44 introduces four concepts that students should understand before proceeding with a discussion of what a p-value means and does not mean.

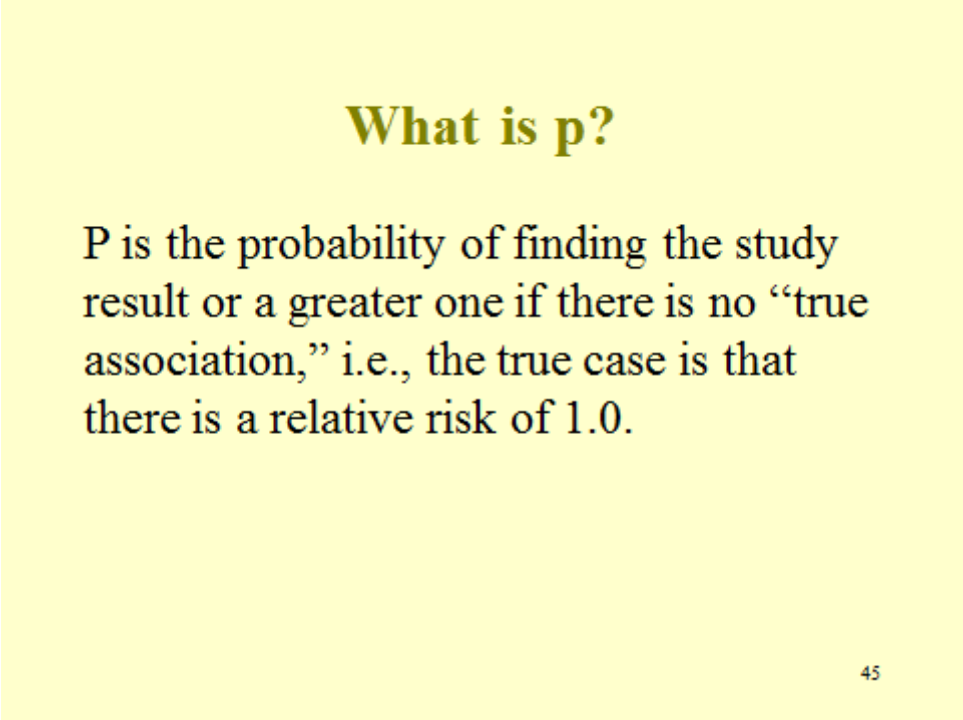
Slide 44 Some Definitional Preliminaries



It may be helpful when presenting these concepts to have students focus on Table 5 in the module, which employs both the false positive and false negative concepts.

So what is p? Once students have the concepts in Slide 44 down, Slide 45 provides an explanation.

Slide 45 What Is p?

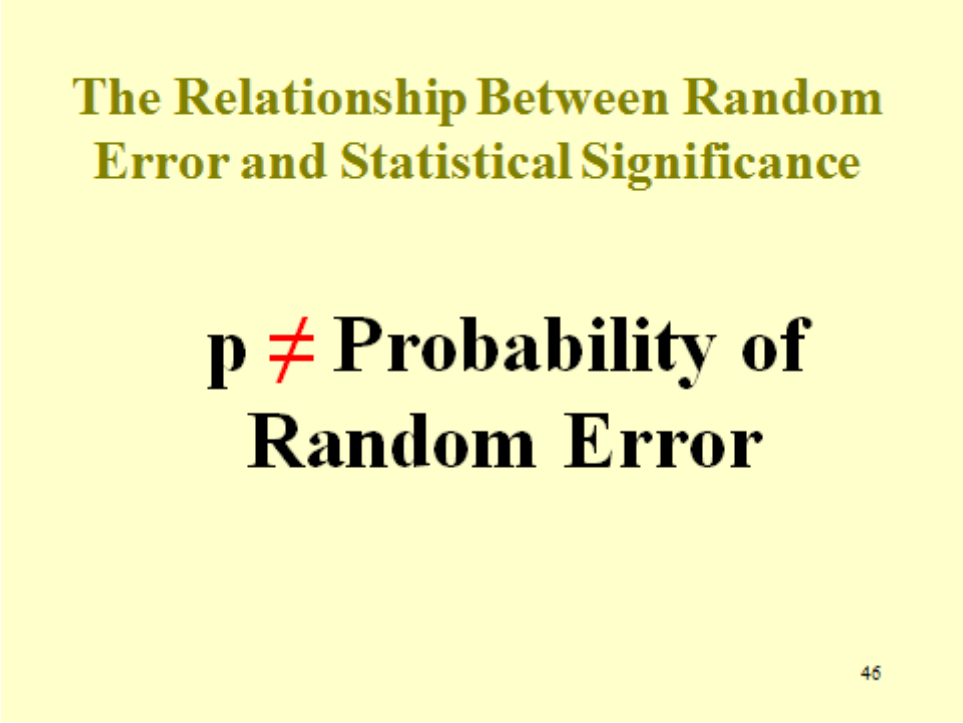


One way to begin to have students appreciate that a p-value is not about the probability that random error produced an erroneous association is to ask them in class what the probability is that some study's outcome with a specified p-value was due to random error. The answer is that it cannot be calculated because conventional statistical methods are unable to calculate that value (Bayesian methods might be used but are employed infrequently by epidemiologists).[4] Thus, the point of this question (and answer) is to emphasize that p-values are not about the probability of random error, but about the probability that would result if, in true (but unknown) fact, there was no association between the agent and disease. Slide 46 makes this point in emphatic terms and may be useful in driving this point home.[5]

[4] Although the frequency of such applications seems to be increasing. For a brief explanation of the use of Bayesian methods for determining the probability of random errors, see Steven N. Goodman & Jesse A. Berlin, "The Use of Predicted Confidence Intervals When Planning Experiments and the Misuse of Power When Interpreting Results," 121 *Annals Internal Med.* 200, 202 (1994).

[5] For an explanation of common errors made by scientists with regard to statistical significance testing, including the one in Slide 45, see Sander Greenland et al., "Statistical Test, *P*-Values, Confidence Intervals, and Power: A Guide to Misinterpretations," *The American Statistician, Online Supplement 1* (2016),

Slide 46 The Relationship between Random Error and Statistical Significance



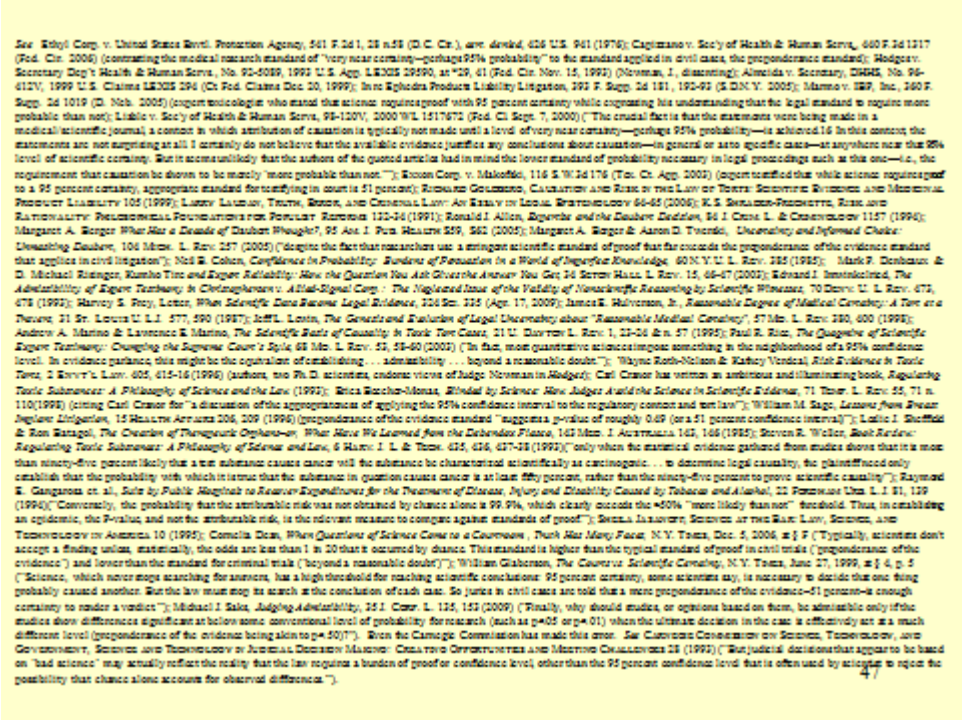
The module explains a difficult matter that is contrary to intuition. The level of significance adopted for significance testing cannot be analogized to the legal standard of proof. It may be expecting too much for students to understand why this is the case, but the module attempts that explanation. Critical points to appreciate if the instructor seeks to have students understand include:

- P-values are used to determine whether the null hypothesis should be accepted or rejected based on whether random error could have produced the result. To the strictest significant testers, a study's results are thrown out as not disproving the null hypothesis when the study results are not statistically significant. By contrast, statistically significant results are at least provisionally accepted as disproving the null.
- The p-value does not tell us the probability that the study's results were produced by chance. Consequently, the complement of the p-value does not tell us the probability that the study's results were free of chance error. Indeed, β, the probability of a false positive, is not the complement of the p-value.

Instructors may decide not to delve into the explanation and simply be declarative that the two cannot be compared. The Kaye characterization in the module is wonderfully evocative

on this point and Slide 47, cataloguing others who have made the error, may provide an incentive for students to absorb this matter even if they cannot fully understand it.

Slide 47 Mistaking p Values as Analogous to the Standard of Proof


See Ethyl Corp. v. United States Envtl. Protection Agency, 541 F.2d 1, 28 n.58 (D.C. Cir.), *cert. denied*, 426 U.S. 941 (1976); Cagliano v. Sec'y of Health & Human Servs., 460 F.3d 1317 (Fed. Cir. 2006) (contrasting the medical research standard of "very near certainty—perhaps 95% probability" to the standard applied in civil cases, the preponderance standard); Hodges v. Secretary Dep't Health & Human Servs., No. 92-5089, 1992 U.S. App. LEXIS 29590, at *29, 41 (Fed. Cir. Nov. 15, 1992) (Newman, J., dissenting); Almeida v. Secretary, DHHS, No. 96-612V, 1999 U.S. Claims LEXIS 296 (Ct Fed. Claims Dec. 20, 1999); In re Ephedra Products Liability Litigation, 393 F. Supp. 2d 181, 192-93 (S.D.N.Y. 2005); Manno v. IMP, Inc., 360 F. Supp. 2d 1019 (D. Neb. 2005) (expert toxicologist who stated that science requires proof with 95 percent certainty while expressing his understanding that the legal standard to require more probable than not); Libby v. Sec'y of Health & Human Servs., 98-120V, 2000 WL 1517672 (Fed. Cl. Sept. 7, 2000) ("The crucial fact is that the statements were being made in a medical/scientific journal, a context in which attribution of causation is typically not made until a level of very near certainty—perhaps 95% probability—is achieved.16 In this context, the statements are not surprising at all. I certainly do not believe that the available evidence justifies any conclusions about causation—in general or as to specific cases—at anywhere near that 95% level of scientific certainty. But it seems unlikely that the authors of the quoted article had in mind the lower standard of probability necessary in legal proceedings such as this one—i.e., the requirement that causation be shown to be merely 'more probable than not.'"); Exxon Corp. v. Makofski, 116 S.W.3d 176 (Tex. Ct. App. 2003) (expert testified that while science requires proof to a 95 percent certainty, appropriate standard for testifying in court is 51 percent); Richard Goldberg, Causation and Risk in the Law of Torts: Scientific Evidence and Medicinal Product Liability 105 (1999); Lucky Luciano, Truth, Error, and Criminal Law: An Essay in Legal Epistemology 64-65 (2006); K.S. Shrader-Frechette, Risk and Rationality: Philosophical Foundations for Populist Reforms 132-36 (1991); Ronald J. Allen, *Expertise and the Daubert Decision*, 84 J. Crim. L. & Criminology 1157 (1994); Margaret A. Berger, *What Has a Decade of Daubert Wrought?*, 95 Am. J. Pub. Health S59, S61 (2005); Margaret A. Berger & Aaron D. Twerski, *Uncertainty and Informed Choice: Unsticking Daubert*, 104 Mich. L. Rev. 257 (2005) ("despite the fact that researchers use a stringent scientific standard of proof that far exceeds the preponderance of the evidence standard that applies in civil litigation"); Neil B. Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. Rev. 385 (1985); Mark F. Grady & D. Michael Risinger, Rumbo Tire *and Expert Reliability: How the Question You Ask Gives the Answer You Get*, 36 Seton Hall L. Rev. 15, 66-67 (2005); Edward J. Imwinkelried, *The Admissibility of Expert Testimony in Christophersen v. Allied-Signal Corp.: The Neglected Issue of the Validity of Nonscientific Reasoning by Scientific Witnesses*, 70 Denv. U. L. Rev. 673, 678 (1993); Harvey S. Perry, Letter, *When Scientific Data Become Legal Evidence*, 216 Sci. 335 (Apr. 17, 2009); James E. Halverson, Jr., *Reasonable Degree of Medical Certainty: A Tort on a Tortoise*, 21 St. Louis U. L.J. 577, 590 (1987); Jeff L. Lewin, *The Genesis and Evolution of Legal Uncertainty about "Reasonable Medical Certainty"*, 57 Mo. L. Rev. 380, 400 (1993); Andrew A. Marino & Lawrence E. Marino, *The Scientific Basis of Causality in Toxic Tort Cases*, 21 U. Dayton L. Rev. 1, 23-26 & n. 57 (1995); Paul R. Rice, *The Quagmire of Scientific Expert Testimony: Changing the Supreme Court's Style*, 68 Mo. L. Rev. 53, 59-60 (2003) ("In fact, most quantitative sciences impose something in the neighborhood of a 95% confidence level. In evidence parlance, this might be the equivalent of establishing . . . admissibility . . . beyond a reasonable doubt."); Wayne Roth-Nelson & Kelsey Vendzel, *Risk Evidence in Toxic Torts*, 2 Envtl. Law. 605, 615-16 (1996) (authors, two Ph.D. scientists, endorse views of Judge Newman in *Hodges*); Carl Cranor has written an ambitious and illuminating book, *Regulating Toxic Substances: A Philosophy of Science and the Law*. (1993); Erica Beecher-Monas, *Blinded by Science: How Judges Avoid the Science in Scientific Evidence*, 71 Temp. L. Rev. 55, 71 n. 110(1998) (citing Carl Cranor for "a discussion of the appropriateness of applying the 95% confidence interval to the regulatory context and tort law"); William M. Sage, *Lessons from Breast Implant Litigation*, 15 Health Affairs 206, 209 (1996) (preponderance of the evidence standard "suggests a p-value of roughly 0.69 (or a 51 percent confidence interval)"); Leslie J. Sheffield & Ron Satzgali, *The Creation of Therapeutic Orphans—or: What Have We Learned from the Bendectin Fiasco*, 143 Mem. J. Australia 163, 166 (1985); Steven R. Weller, *Book Review: Regulating Toxic Substances: A Philosophy of Science and Law*, 6 Harv. J. L. & Tech. 425, 426, 427-28 (1993)("only when the statistical evidence gathered from studies shows that it is more than ninety-five percent likely that a test substance causes cancer will the substance be characterized scientifically as carcinogenic . . . to determine legal causality, the plaintiff need only establish that the probability with which it is true that the substance in question causes cancer is at least fifty percent, rather than the ninety-five percent to prove scientific causality"); Raymond S. Gangarosa et. al., *Safe by Public Mandate vs Reserve Expenditures for the Treatment of Disease, Injury and Disability Caused by Tobacco and Alcohol*, 22 Forensic Utra. L. J. 81, 139 (1994)("Conversely, the probability that the attributable risk was not obtained by chance alone is 99.9%, which clearly exceeds the >50% "more likely than not" threshold. Thus, in establishing an epidemic, the P-value, and not the attributable risk, is the relevant measure to compare against standards of proof"); Sheila Jasanoff, Science at the Bar: Law, Science, and Technology in America 10 (1995); Cornelia Dean, *When Questions of Science Come to a Courtroom*, *Truth Has Many Faces*, N.Y. Times, Dec. 5, 2006, at § F ("Typically, scientists don't accept a finding unless, statistically, the odds are less than 1 in 20 that it occurred by chance. This standard is higher than the typical standard of proof in civil trials ("preponderance of the evidence") and lower than the standard for criminal trials ("beyond a reasonable doubt")."); William Glaberson, *The Courts vs. Scientific Certainty*, N.Y. Times, June 27, 1999, at § 4, p. 5 ("Science, which never stops searching for answers, has a high threshold for reaching scientific conclusions: 95 percent certainty, some scientists say, is necessary to decide that one thing probably caused another. But the law must stop its search at the conclusion of each case. So juries in civil cases are told that a mere preponderance of the evidence—51 percent—is enough certainty to render a verdict."); Michael J. Saks, *Judging Admissibility*, 25 J. Corp. L. 135, 153 (2009) ("Finally, why should studies, or opinions based on them, be admissible only if the studies show differences significant at below some conventional level of probability for research (such as p=.05 or p=.01) when the ultimate decision in the case is effectively set at a much different level (preponderance of the evidence being akin to p=.50)?"). When the Carnegie Commission has made this error. *See* Carnegie Commission on Science, Technology, and Government, Science and Technology in Judicial Decision Making: Creating Opportunities and Meeting Challenges 28 (1992) ("But judicial decisions that appear to be based on "bad science" may actually reflect the reality that the law requires a burden of proof or confidence level, other than the 95 percent confidence level that is often used by scientists to reject the possibility that chance alone accounts for observed differences.").

47

In re *Ephedra Products Liability Litigation*, 393 F. Supp. 2d 181 (S.D.N.Y. 2005)

With regard to the question in the module, the court is quite simply wrong in: (1) stating that the p-value "means that there is one chance in twenty that a result showing increased risk was caused by sampling error . . . ." and that a $p \leq .05$ means that "no more than one chance in twenty of a finding [of] a false association due to sampling error"; and (2) comparing a p-value to the preponderance of the evidence standard. On the latter error, the court is one of the many listed in Slide 47 that has made this error. On the former error, the court fails to appreciate that significance testing is not about the probability of an erroneous result, but is instead about whether the study result is erroneous if there is indeed no true association. The court's statement about confidence intervals (which are covered immediately following *In re Ephedra*) is incorrect in referring to its measure as the "same risk of sampling error," but correct about the implications of confidence intervals including the value of 1.0 with regard to a lack of statistical significance. Unlike much of legal reasoning, there simply is no wiggle room about the correctness of analysis with statistical methods.

Study Questions:

1. Why, after discussing the preponderance of the evidence legal standard, does Judge Rakoff countenance use of "reasonable degree of scientific certainty"?

   This case provides an opportunity to explain to students that science has no such standard of proof. This concept was made up by courts as the standard for the admissibility of expert witness testimony to emphasize to those experts the importance of their testimony in court. However, as the Third Restatement of Torts explains in comment e to Section 28:

   > In an effort to screen expert opinions that are speculative, some courts have employed a requirement that an expert testify that an opinion is held to a "reasonable degree of medical [or scientific] certainty" for it to be admissible. This phrase implies a standard different from the preponderance requirement, suggests reliance on medical or scientific standards for proof, and seems to impose a high threshold for the opinion to be admissible. Some courts avoid the last-mentioned implication by employing instead a requirement that the expert testify that an opinion is held to a "reasonable medical [or scientific] probability."
   >
   > Requiring an expert to state that an opinion is held to a medical or scientific certainty is problematic because the medical and scientific communities have no such "reasonable certainty" standard. Thus for an expert to understand this standard, meaning must be provided by the attorney who hired the expert, by the expert's imagination, or by some other source outside the legal system. The implication that the reasonable certainty standard requires something more than the preponderance standard is belied by courts that have provided a definition of or explained the phrase—the vast majority of those courts state that the standard is equivalent to the usual preponderance requirement.[6]

2. The court refers to a study that had a "fivefold increased risk of hemorrhagic stroke" but that is not "statistically significant." How can you have a 5-fold increase in the risk that is not significant?

   Well, you can. As explained in the module, statistical significance is not about the magnitude or the association found in a study. Rather, this concept only addresses the matter of random error in the study result.

3. If, as previously discussed, case-control studies are particularly useful because they

[6] Restatement of the Law Third, Torts: Liability for Physical and Emotional Harm copyright © 2010 by The American Law Institute. Reproduced with permission for educational use only. All rights reserved.

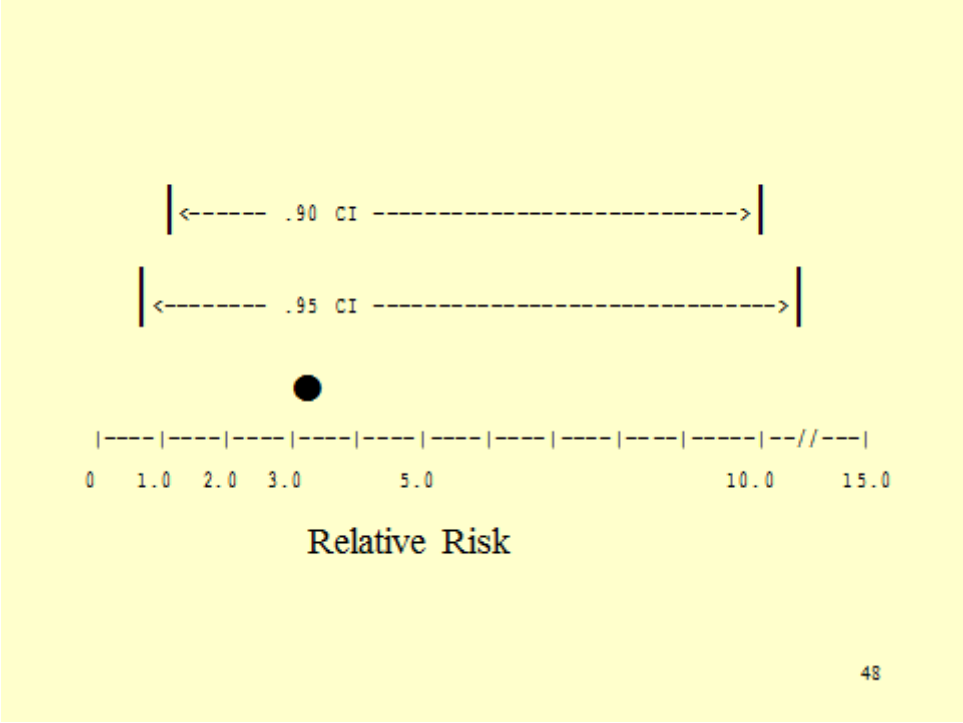
find associations when the incidence of disease is small, why were the researchers discussed in the case unable to find statistically significant results? Would a cohort study have been better to employ in these circumstances?

The point is that if exposure is rare, regardless of the incidence of stroke, researchers will be unable to determine if a difference in exposure exists between cases and controls. A cohort study would be a better study design, assuming you could create an exposed cohort of ephedra uses that would be large enough to pursue differences in the incidence of stroke between that cohort and an unexposed cohort.

Confidence Intervals

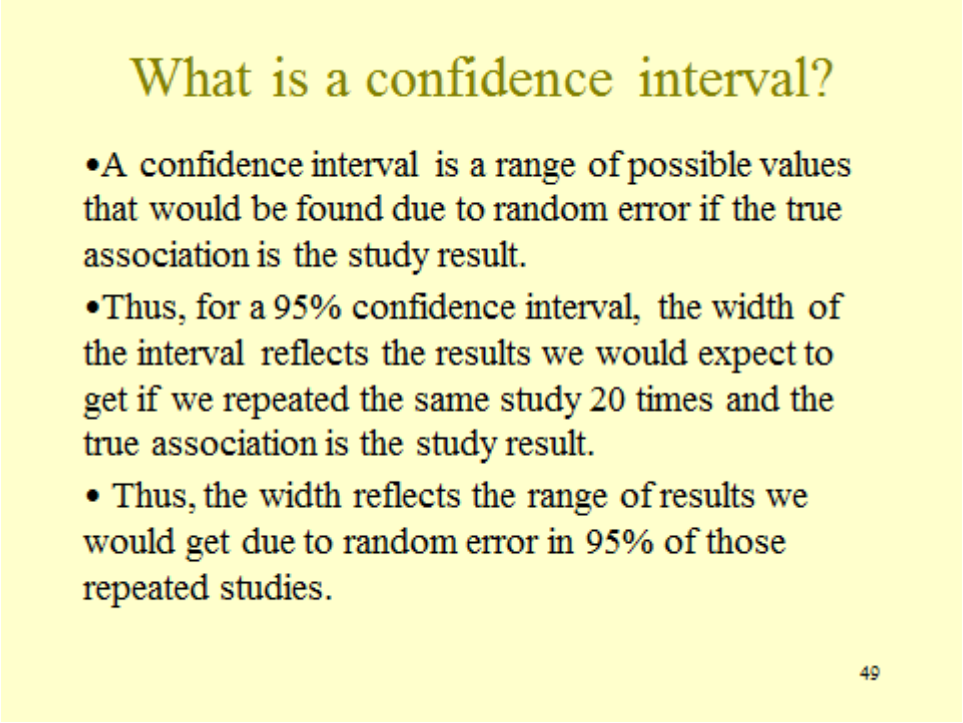
Slide 48 contains a hypothetical and primitive confidence interval that an instructor might use to test whether students understand the textual material on this subject.

Slide 48 Relative Risk



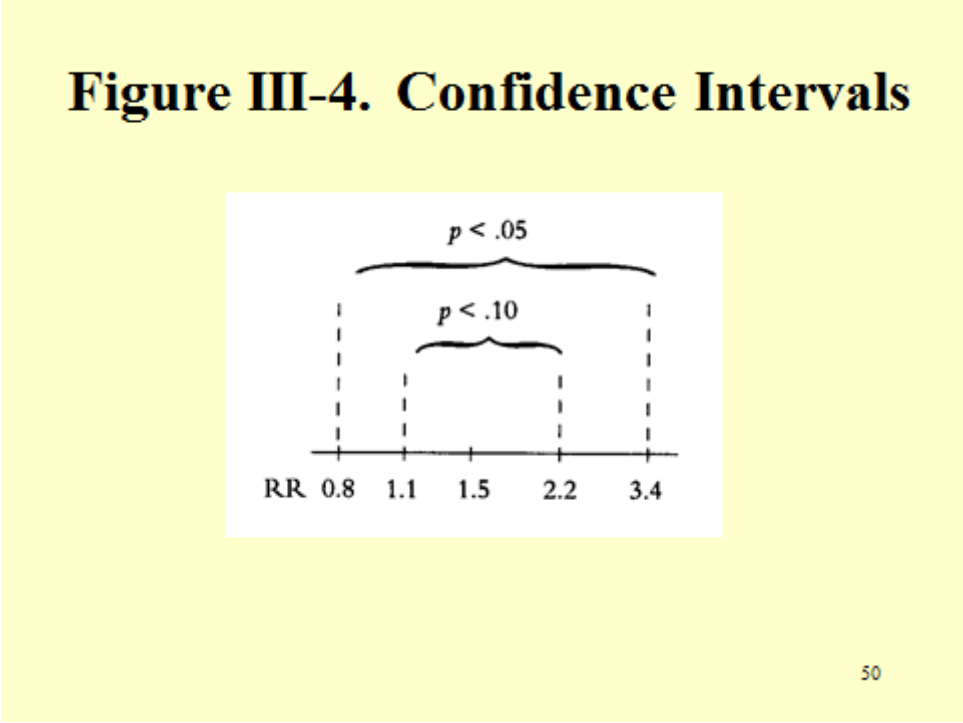
After displaying Slide 48 or perhaps even before, Slide 49 contains an outline of what a confidence interval is.

Slide 49 What Is a Confidence Interval?



There is a great deal packed into Slide 49, and instructors should tarry if necessary to be sure students have absorbed these concepts. Proceeding to Figure III-4 in the module, contained in Slide 50, affords an opportunity to test their understanding and the relationship between a confidence interval and statistically significant study results.

Slide 50 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.

Study Questions (about slide 50):

1. What are the bounds of the 95% CI? The 90% CI?

   The boundaries for the 95% confidence interval are .8 and 3.4. The boundaries for the 90% confidence interval are 1.1 and 2.2.

2. Why is the 90% CI narrower than the 95% CI?

   Because the 90% CI reflects the range of results we would expect if we repeated the study 10 times. By contrast, the 95% CI reflects the results we would expect if we repeated the study 20 times. With more repetitions, there is more opportunity for extreme results.

3. How can one tell whether the study results (what is the study result, anyway?) are statistically significant?

We should not let the CI limits obscure that the study found a relative risk of 1.5 (denoted by the hash mark on the horizontal access). Study results are statistically significant at the specified p-level when the lower bound of the CI is greater than 1.0, which is the same as saying that if the null hypothesis is true, we would expect to find the study result less than one time in 20.

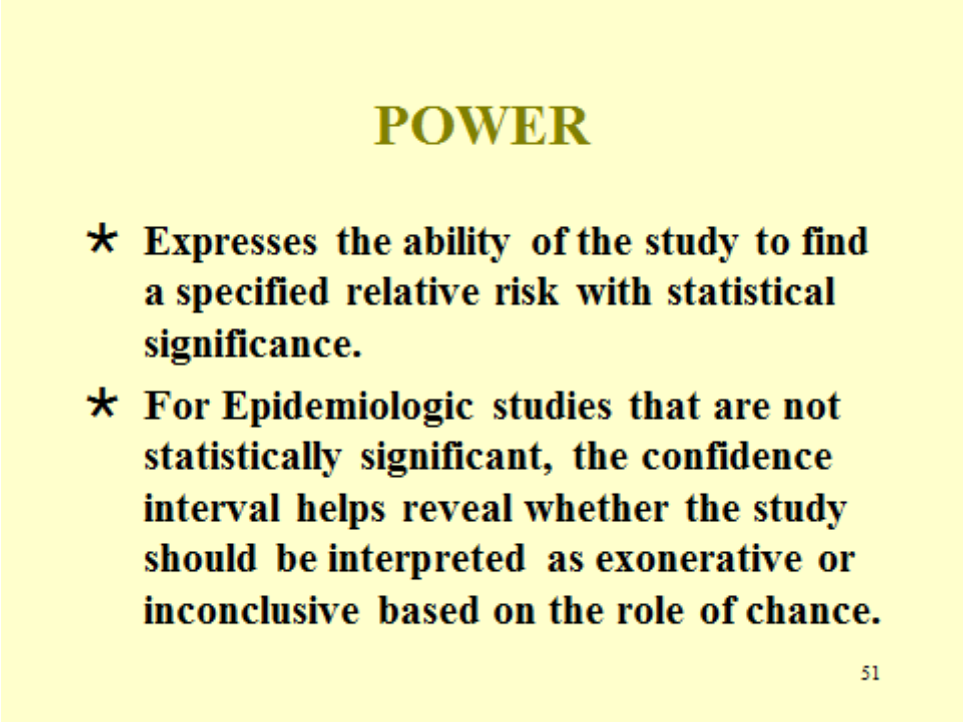
False Negatives

The material on false negatives requires little further elucidation. The main points are (1) statistical significance testing is about avoiding false-positive results; (2) more stringent avoidance of false-positive results increases the likelihood of false-negative results; and (3) beta is not simply the mathematical complement of alpha.

Power

Slide 51 describes power and how it is used in study design. After a study is completed, a confidence interval can be employed to determine which results are statistically compatible with the study results and which are not. (For technical reasons, calculating power after a study is completed is problematic.)

Slide 51 Power



We use a study concerning which of two baseball players is the better hitter to explain the concept of power and the difficulty of finding a statistically significant result when the disease being studied is rare or when the increased risk by the studied agent is small.

Imagine a study designed to determine which of two baseball players is the better hitter. If the study found that Player A hit .200 and Player B hit .300, we might be tempted to conclude that Player B is the better hitter. After all, that is quite a large difference in batting averages.

But suppose that our study consists of a sample of 10 at-bats by each—meaning that Player A had 2 hits and Player B had 3 hits. If in their very next at-bats, Player A gets a hit and Player B strikes out, they have exactly the same average.

The possibility that there is no difference in their batting abilities and the results are just an artifact of 10 at-bats looms large—the difference in their batting successes found in the study is not statistically significant. Indeed, the difference we found here is far from being statistically significant.

Does that mean that we should conclude that the batters are of equal ability? Most emphatically not! That the study failed to find a statistically significant difference does not mean that the two hitters are equal, that is, that the null hypothesis is true.

Here, because of the small number of events surveyed, this study was

exceedingly unlikely to find any real difference that might exist with statistical significance—or to put it another way, its power was quite low. Consequently, its evidentiary value, should the relative merits of the two batters be relevant, is quite limited. This is an inconclusive study, rather than one that supports the proposition that the two batters are of equal ability.

On the other hand, if we looked at 1,000 at-bats per hitter and discovered that one hit .300 and another hit .200, we would be much more confident that we would like our favorite team to trade the latter hitter for the former one—chance is an unlikely explanation for the disparity in batting averages if they are really equally good hitters.

Because of the larger number of events in the sample, the study had greater power to find a statistically significant difference in their batting averages.

Now, consider poor batters, indeed very, very poor ones—that is, Player A hits .020 and Player B hits .030. The power of our study is considerably reduced even though the sample size remains the same—this is equivalent to disease that occurs infrequently, such as a specific birth defect.

To put the point another way, it is far harder to detect causes of very rare diseases than it is to detect the causes of more commonly occurring ones. The number of episodes—disease—as well as the sample size are important determinants of the power of a study.

And not only is it hard to detect causes of rare events, it is virtually impossible to detect, with any degree of accuracy, those agents that only increase the incidence of disease by a small amount.

That is, if the background rate of disease is very high compared to the amount that is caused by the agent being studied, we are most unlikely to be able to identify the agent as having a causal relationship.

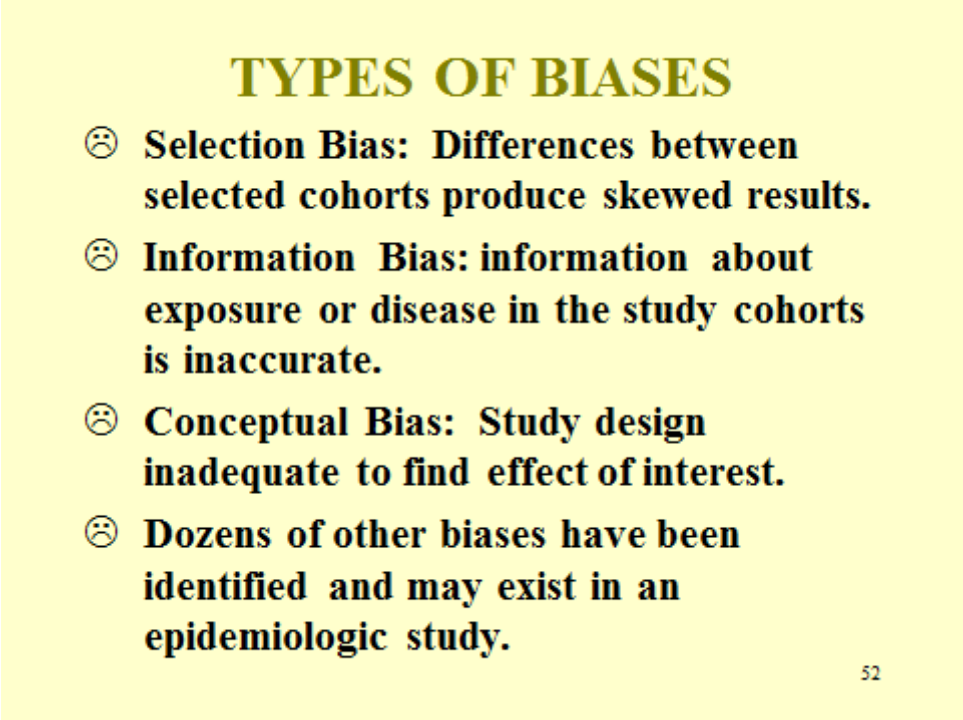
To go back to our baseball analogy, if Player A hits .025 and Player B hits .0255, it is most unlikely that we will obtain a statistically significant outcome in our study, or that other variables—like the different parks that they play in, or the different abilities of the pitcher they face—are not responsible for any difference we find, rather than a real difference in their hitting abilities.

Bias

We start off the discussion of bias by reminding students that usage of “bias” in science is different from its legal usage. Scientists use bias to mean any methodological error that may result in an inaccurate outcome. Here, bias does not mean that someone has an interest in the outcome, or an axe to grind, or some invidious purpose.

Slide 52 contains the three major biases discussed in the Student Materials, along with an explanation that there are many more that have been identified. Indeed, one article purported to identify over 100 different biases! Students might be asked if they can come up with examples of each of these three different biases, or an example of another kind of bias.

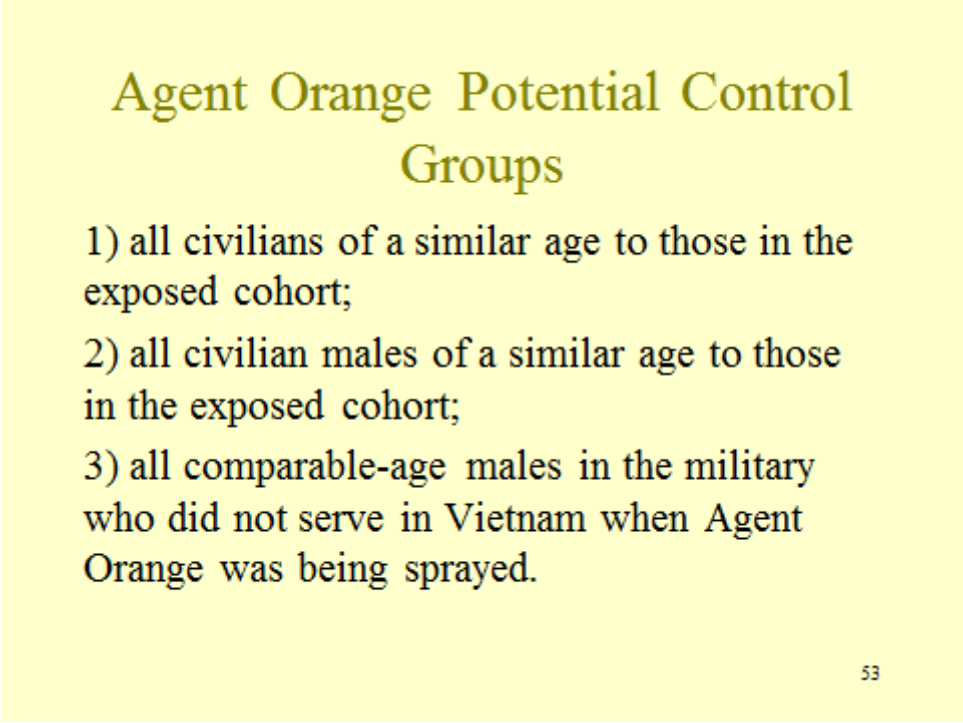
Slide 52 Types of Biases



Study Questions:

1. In a study of the health effects of exposure to Agent Orange, a herbicide used during the Vietnam War, the exposed cohort consisted of young men who were enlisted or drafted into military service. Consider (Slide 53) potential control groups for that cohort study: (1) all civilians of a similar age to those in the exposed cohort; (2) all civilian males of a similar age to those in the exposed cohort; or (3) all comparable-age males in the military who did not serve in Vietnam when Agent Orange was being sprayed.

Slide 53 Agent Orange Potential Control Groups



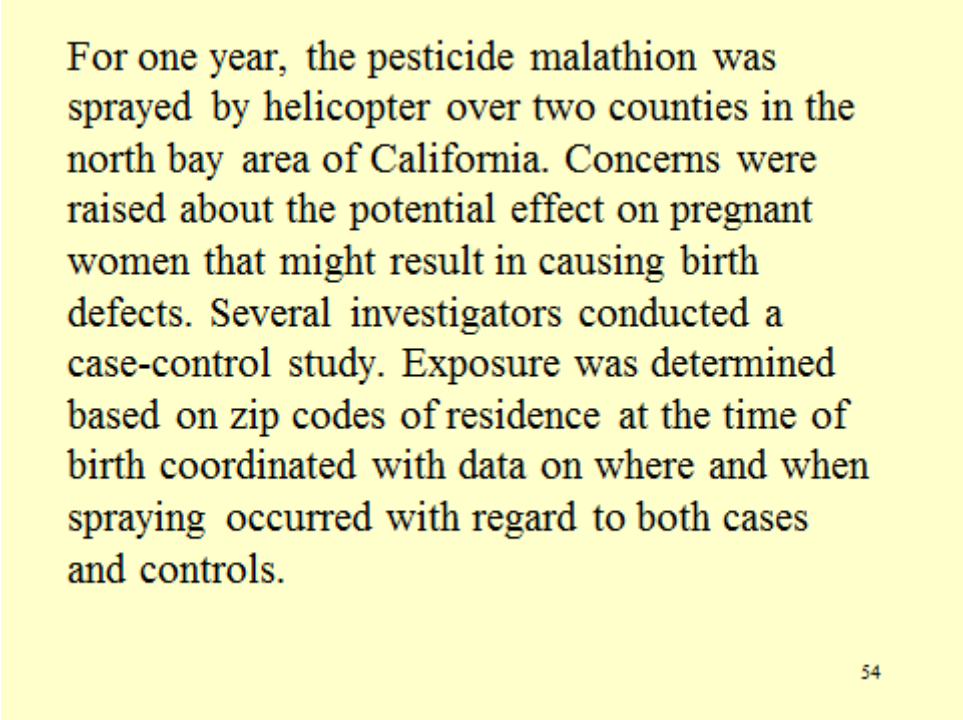
Which of these three choices do you think is preferable? Why? Can you think of a control group that you think would be preferable to the one you selected from the three above?

a. Comparing the mortality rate of the exposed cohort and that of a control group made up of civilians might have resulted in error due to conceptual problems that lead to selection bias. (Recall and remind students that very few females served in Vietnam and none in combat, where most were exposed to Agent Orange.) If the outcome being studied were testicular cancer, including women in the control group would differentially result in a lower and biased disease rate in the controls, thereby producing a spurious association or exaggerating one that might exist.
b. Limiting the control group to males avoids the selection bias in the first proposed control group. However, because many who did not serve in Vietnam were excluded because of health reasons, this control group raises concerns about selection bias because of differences in the health status of exposed and control cohorts. Failing to account for health status as an independent variable tends to mask or understate any association between exposure and disease in studies in which the exposed cohort is healthier.
c. This control group avoids the conceptual and selection biases of the first two

proposed control groups. There may still be some residual concern that those who did not serve in Vietnam but elsewhere may have been exposed to other toxins or risk factors where they served that were different from those that existed in Vietnam, apart from Agent Orange.

2. For 1 year, the pesticide malathion was sprayed by helicopter over two counties in the north bay area of California. Concerns were raised about malathion's potential effect on pregnant women that might result in causing birth defects. Several investigators conducted a case-control study.[7] Exposure was determined based on zip codes at the time of birth, coordinated with data on where and when spraying occurred with regard to both cases and controls. Refer to Slide 54.

Slide 54 Malathion Exposure



3. What concerns would you express to the investigators about their measure of exposure?

[7] Duncan C. Thomas et al., "Reproductive Outcomes in Relation to Malathion Spraying in the San Francisco Bay Area, 1981–1982," 3 *Epidemiology* 32 (1992).

This question raises the issue of information bias, which addresses inaccuracies in measuring either disease or exposure status: here, the issue of error in exposure classification. The criterion for classifying the exposed cohort raises a number of concerns. Residence at the time of birth may not be the same as residence during pregnancy (or, more importantly, at the time during pregnancy when birth defects can be caused by environmental agents). In addition, pregnant women who lived in one place may have worked in another where pesticide spraying was different from the area of their residence. However, all of these potential classification errors are nondifferential, and so we would expect, random error aside, to tend to bias the results toward no association.

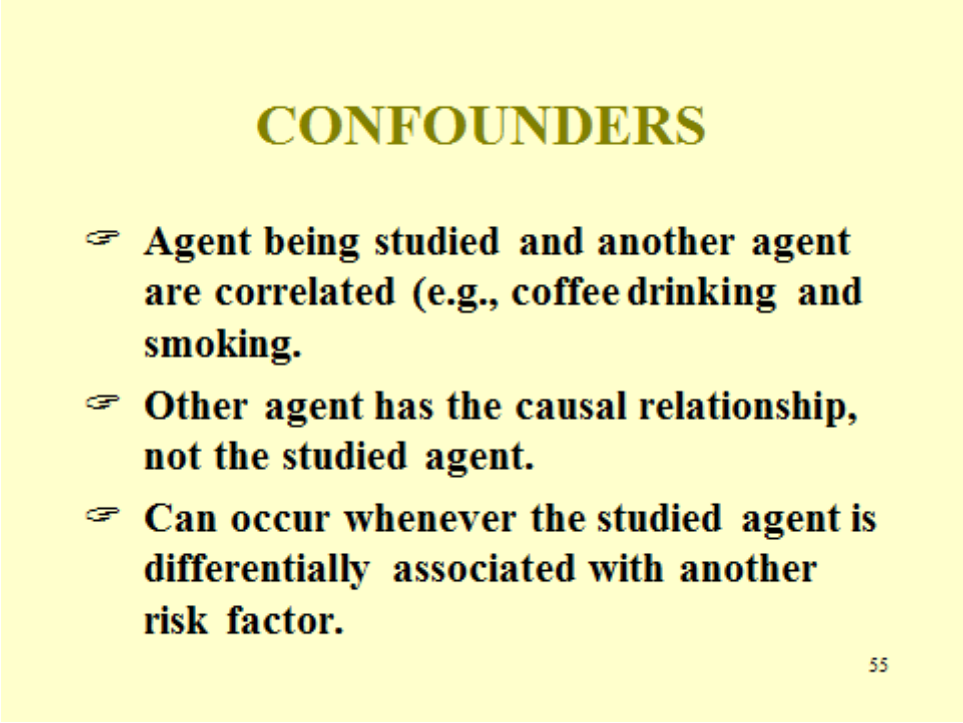
4. The text on information biases mentions the use of monitoring devices to measure exposure and dose. In what kind of studies might measuring-device data be practical to use?

Prospective cohort studies are the ones in which monitoring is feasible. In addition, the exposure must be one that can actually be monitored, for example, radiation exposure for those who whose work environment involves exposure to radiation. Or radon exposure for those whose homes have excess radon.

Confounding

For presentation of the concept of confounding, we use Slide 55.

Slide 55 Confounders



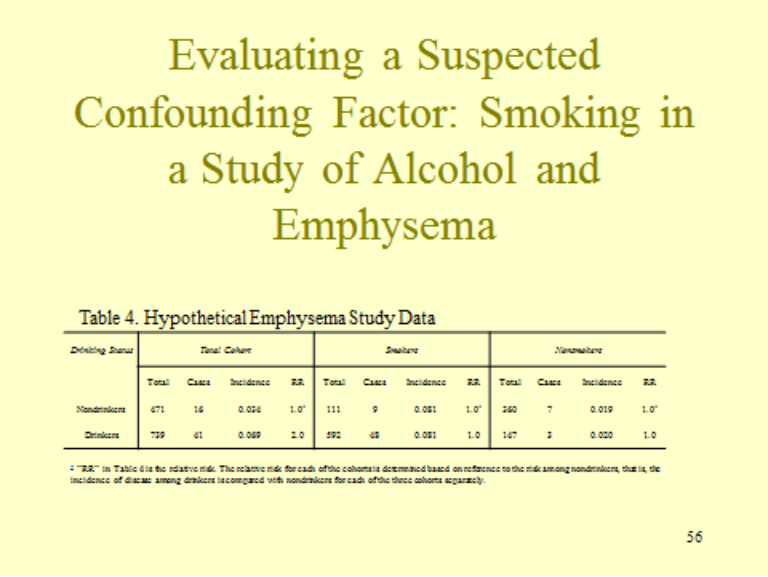
This is the point at which illustrations of confounding are helpful in students' absorption of the concept. The text employs hair color and its relationship to death. Instructors may want to make up their own additional illustrations for use in class. Here is another one that we use:

> One study of Bendectin found a statistically significant association between Bendectin and a birth defect known as pyloric stenosis.
>
> Pyloric stenosis is a condition in a newborn that involves a defect in the valve to his or her stomach and reveals itself with projectile vomiting. The child has trouble getting nutrition because of this condition, but once diagnosed, a relatively straightforward operation can fix it.
>
> One might have concluded that the association between Bendectin and pyloric stenosis is causal. But some researchers cautioned that the real relationship may be between morning sickness and pyloric stenosis.
>
> If so, then we would find an association between Bendectin and pyloric stenosis, even though Bendectin is not at all causal. Why?
>
> Because women with morning sickness are much more likely to use Bendectin than women who do not have morning sickness. Thus, even if Bendectin has no causal role, it shows up more frequently because morning sickness, a confounder, is differentially associated with use of Bendectin.

Instructors might ask students to make up their own examples of confounding or assign that task to groups of students during class with each group presenting their example and time for the remainder of the class (and the instructor) to comment, critique, or laud.

Confounding factors can be controlled for if identified by researchers during the study design. Controlling requires not only identification, but also gathering data about those confounding factors during that phase of the study. Table 6 in the module (reproduced as Slide 56) reveals how, through stratification based on the suspected confounding factor of smoking, researchers can control for confounding and assess whether there is a real association between the studied exposure and outcome.

Slide 56 Evaluation a Suspected Confounding Factor: Smoking in a Study of Alcohol and Emphysema



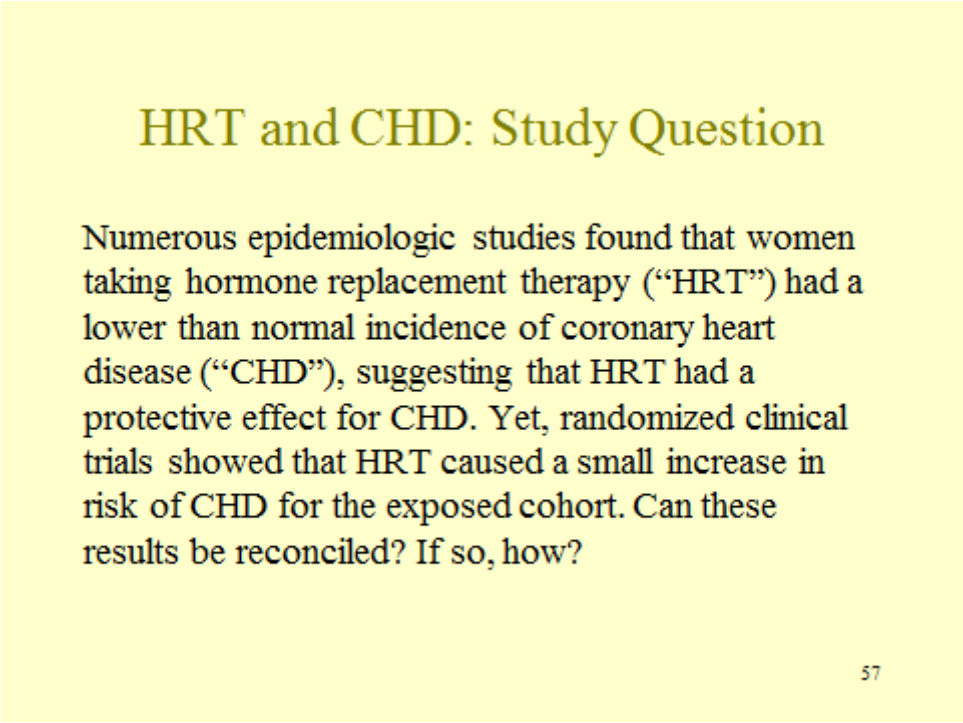
The key matter revealed by Table 6 is that when study participants are stratified by the suspected confounder, as in the second and third columns, the association between alcohol washes out. Only because of the higher incidence of emphysema among smokers and the greater proportion of smokers in the drinking (exposed) cohort does an association between drinking and emphysema appear.

Study Question:

1. Numerous epidemiologic studies found that women taking hormone replacement therapy (HRT) had a lower than normal incidence of coronary heart disease (CHD), suggesting that HRT had a protective effect for CHD. Yet, randomized clinical trials showed that HRT caused a small increase in risk of CHD for the exposed cohort. Can

these results be reconciled? If so, how? Refer to Slide 57.

Slide 57 HRT and CHD: Study Question

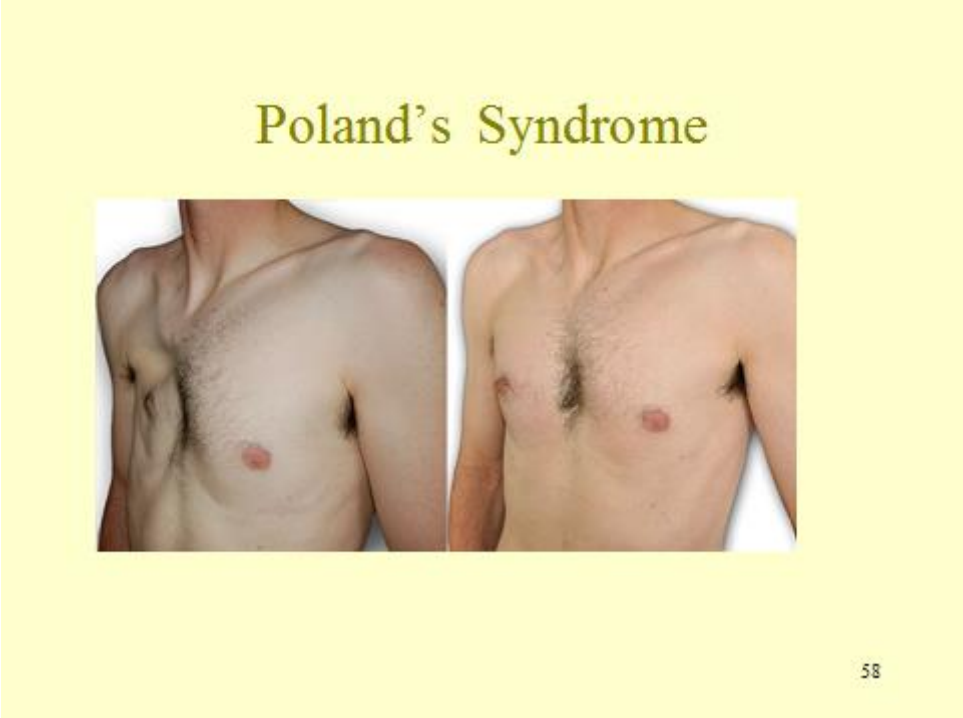


When the results of the randomized trials were reanalyzed, researchers realized that the effects found were confounded by socioeconomic status and the better exercise and healthy life style of those with greater financial assets and education. When the results were stratified based on socioeconomic status, the protective effects of CHD disappeared.

Brock v. Merrell Dow Pharmaceuticals, Inc., 874 F.2d 307 (5th Cir. 1989):

The *Brock* case is included because it is the best example of a court's failing to understand epidemiology. We employ it here because students have the tools to appreciate the court's erroneous understanding of significance testing. Might we also hope that reading a reported opinion by federal court of appeals with these problems will serve as motivation to better understand the course materials? Slide 58 illustrates Poland's syndrome, the birth defect at issue in *Brock*.

Slide 58 Poland's Syndrome



SOURCE:
https://commons.wikimedia.org/wiki/File:Photo_AvantApres_Poland_Homme_02_700x400.jpg
Images are in the public domain and licensed under the Creative Commons Attribution-Share Alike 4.0 International license.

Study Questions:

1. The court states that error could occur if there were a difference in the existence of other risk factors for birth defects between the exposed cohort and the control cohort. If that were the case, would that comprise sampling error, bias, or confounding? If bias, what type of bias? Can you think of a strategy to employ in the study design to minimize the possibility of this error?

   A difference in risk factors could occur due to sampling error, several biases, and confounding. Pure bad luck (sampling error) could result in more genetic risk for heart disease existing in the exposed cohort than in the control group. The smaller the cohorts, the greater the risk of this phenomenon. Selection bias could produce a difference in risk factors: earlier we refer to a study in which women with hysterectomies are over-represented in the exposed cohort of a study examining cervical cancer. This means that the control group would be at a greater risk of

cancer than the exposed cohort. The healthy-worker bias is another example of differential risk factors, once again resulting in the occupationally exposed cohort being at a lower risk of disease than the control group. Finally, confounding is about a difference in risk factors between control and cohort due to the confounding factor being associated with the exposure.

The strategy in all cases would be to identify as many risk factors for disease and attempt in the study design to avoid those that can be avoided and adjust for those that cannot during data analysis.

2. The court also states that error would occur if there were a dependence between variables, such as women taking Bendectin being more likely to smoke than women who did not. Would such a dependence cause error? If so, what kind of error?

   If smoking were a risk factor for birth defects, then this dependence would produce error, as one would find a higher incidence of birth defects in the exposed cohort due solely to smoking. This error is classified as confounding, but can be controlled for by gathering data on smoking among study participants and assessing if there is a difference between the exposed and control cohorts and, if so, adjusting for it as explained in the section of module on confounding.

3. The court identifies selective recall as a potential source of error. In what kind of study would selective recall be of concern? In what direction would such a source of error affect the results?

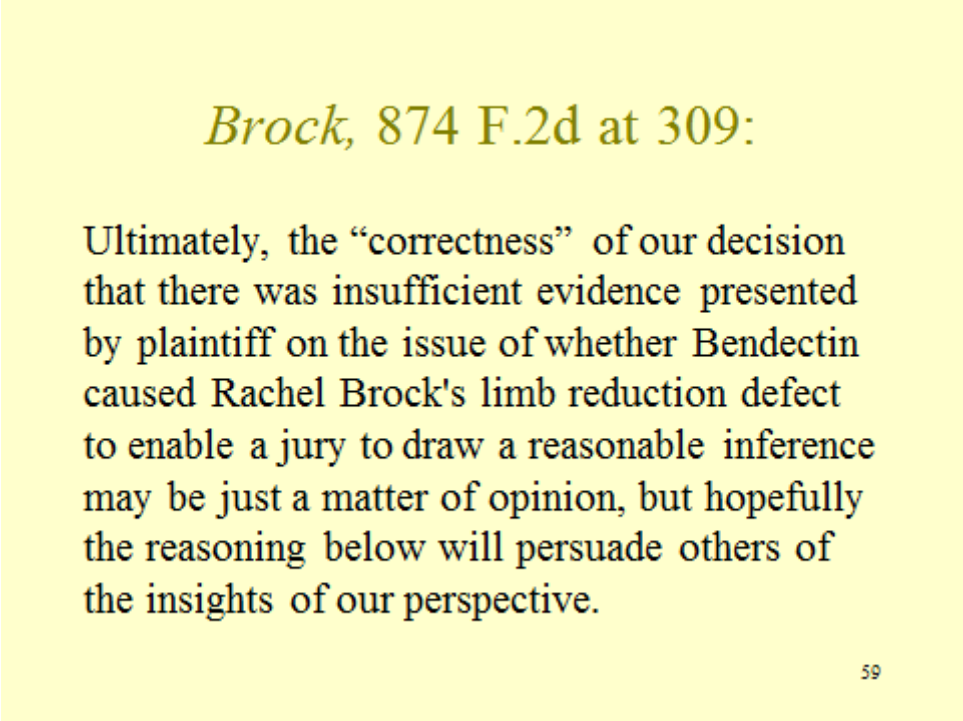
   Selective recall is of concern in retrospective studies in which there is no documented source of information about exposure. Prospective studies have the advantage of gathering exposure during the follow-up phase of the study, thereby avoiding reliance on the memories of study participants who have the disease of interest. Case-control studies, which are always retrospective because they begin by identifying those with the disease of interest, are particularly susceptible to recall bias.

4. The court writes that significance testing can address all of the potential sources of error in an observational study: sampling error, bias, and confounding, and hence adopts a requirement that plaintiff proffer statistically significant studies to satisfy his or her burden. Is the court's premise for requiring statistically significant studies correct?

   It is not. As students should appreciate at this point, significance testing addresses only random error and has nothing to say about bias or confounding. But beyond this erroneous understanding of significance testing, *Brock* reveals a broader

over-emphasis on error due to sampling error (which can be conveyed numerically) and error due to bias and confounding (which often must be assessed qualitatively).

Slide 59 *Brock* Excerpt

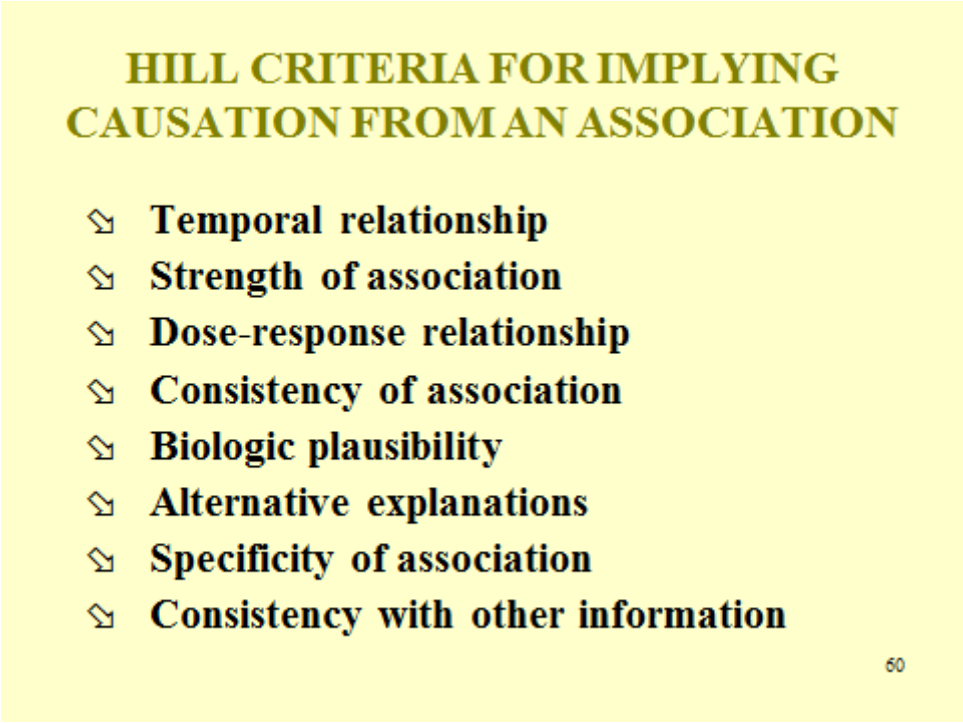


5. At the outset of its opinion, the court expresses its wish (Slide 59) that its opinion will persuade the reader of the conclusion that plaintiffs had not introduced sufficient evidence of causation. Are you persuaded by the court's explanation?

   The irony is deep. The court seeks to be persuasive but embarrasses itself with its misunderstanding of epidemiology.

Section III.G. Assessing Whether an Association is Causal or Spurious

Slide 60 Hill Criteria for Implying Causation from an Association



The Hill criteria (Slide 60) are merely a template to assist in making a judgment about whether a true causal relationship should be inferred from the association found in the study. Alternatively, the result may be spurious in the sense of being an artifact of bias, confounding, or sampling error, or some combination of them.

Whether to draw a causal inference from an association based on these criteria is a matter of informed judgment, not scientific method. There is no algorithm that exists for using these criteria to make that judgment. As a result, this is a matter about which scientists can sometimes reasonably disagree.

The Hill factors are not a basis for determining, in the absence of an association, that causation exists. Most courts seem to have come to appreciate this in the face of experts who attempt to use Hill factors to bridge a gap in existing epidemiologic evidence.

Section III.G.1. Temporal Relationship

The only one of these factors that is a sine qua non for an inference of causation is time. Effects cannot occur before their causes, by definition. Thus, a Martian who suddenly dropped into a cancer ward might remark on the incidence of smoking among cancer victims and conclude: "Cancer must cause people to smoke." But if we look at the timing, all of those cancer victims were smoking before they suffered from cancer.

The other role of time is when we know what the latency period is for developing the disease. If the exposure occurred before the latency period window began, that is, this class develops lung cancer right after finals at a higher rate than control class B and our inquiry is whether the association between taking toxic torts and lung cancer is causal. No, because we know lung cancer's latency period is decades, not months. Similarly, on the other end of the time spectrum are vaccine cases where the infection period is known and the outcome occurs only after a longer period from vaccine administration. Another case, similar to the *Swine Flu Immunization* case discussed in the module, is *Shirley v. Secretary of Department of Health & Human Services*, No. 01-537V, 2006 WL 5606263, at *4 (Fed. Cl. Aug. 7, 2006). Vaccinations seem to be the predominant, if not exclusive, province of sufficiently precise latency-period information of use in toxic tort litigation.

Study Question:

1. In *Alder v. Bayer Corp.*, AGFA Division, 61 P.3d 1068, 1090 (Utah 2002), the court, in giving credence to temporality for purposes of inferring causation, wrote: "If a bicyclist falls and breaks his arm, causation is assumed without argument because of the temporal relationship between the accident and the injury." By contrast, in *Glastetter v. Novartis Pharmaceuticals Corp.*, 252 F.3d 986, 990 (8th Cir. 2001), the court took a very different approach to the role of temporality. The case involved a mother who suffered a stroke after taking a drug to suppress lactation. With regard to case reports, the court wrote: "Though case reports demonstrate a temporal association between Parlodel and stroke, or stroke-precursors, that association is not scientifically valid proof of causation." Can these two approaches to temporality be reconciled and, if so, how?

   An important distinction is the role of competing causes. In the case of the bicycle accident, no competing cause is evident. By contrast, there are numerous other potential causes that might explain the plaintiff's stroke, some of which may have occurred well prior to the stroke. In addition, we understand the biological mechanism involved in broken limbs better than we understand the mechanism in strokes and the circumstances in the bicycle accident are consistent with that biological mechanism knowledge.

Section III.G.2. The Strength of the Association

This is the magnitude of the relative risk. This factor is not based on the notion that weak associations occur less frequently than stronger ones. Rather, it is based on the idea that sources of error (sampling, biases, and confounding) are less likely to produce a substantial error in the association of a study and that, therefore, when a strong association is found, it is more likely real rather than spurious. This factor thus overlaps with the "alternative explanations" one.

Indeed, many epidemiologists resist making a causal attribution for relative risks under 3.0.[8] The risk of error or confounding is just too great to trust such small-study results unless there is significant confirmatory evidence, such as studies with higher exposure finding a stronger association, toxicology evidence, or a well-supported biologic mechanism explanation.

Study Question:

1. For some causal relationships—such as the one between thalidomide and limb-reduction birth defects and between in utero exposure to DES and vaginal adenocarcinoma in young women—researchers were prepared to declare that a causal relationship existed once the common exposure (thalidomide and DES exposure) were discovered among the victims. Why was it so easy to find these causal relationships while others require decades and multiple studies?

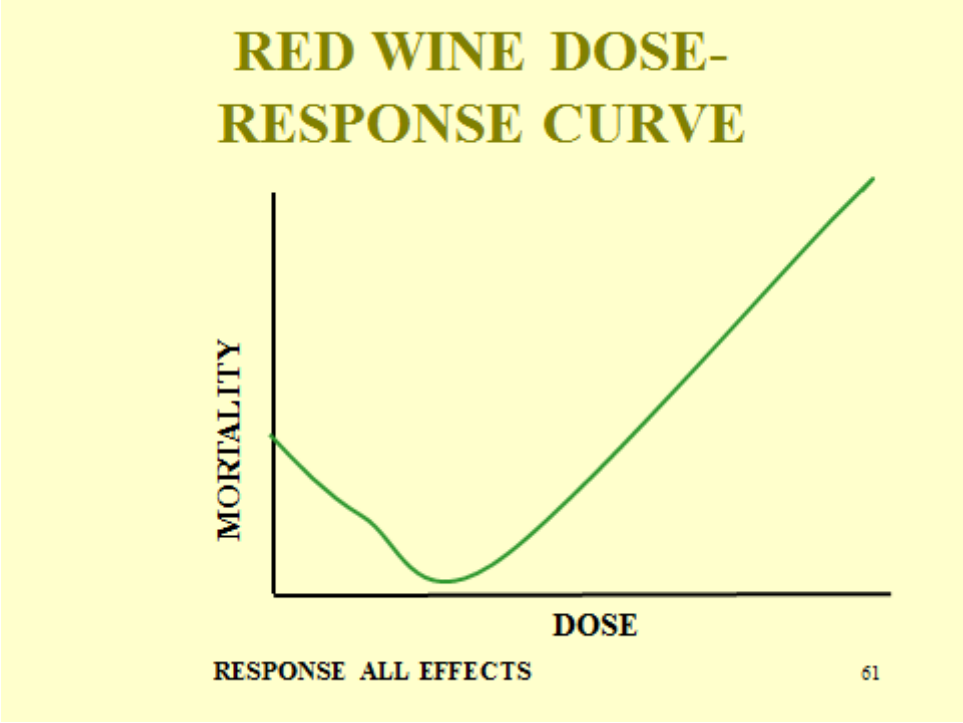
   In both cases, the outcomes are signature diseases (or nearly so, in the case of limb-reduction birth defects). Thus, relative risks are very large, indeed approaching infinity, which reflects that there are no competing causes of the disease that might be responsible instead of the identified agent.

Section III.G.3. Dose-Response Relationship

Instructors may want to defer addressing dose-response relationships until the toxicology materials, which address this matter in more detail. We just note there that there are two types of dose-response relationships. One is the severity of the disease in relation to the dose. Progressive diseases reflect this relationship, while others, such as cancer, do not. The other dose-response relationship is between dose and the incidence of disease. In general, the greater the dose, the higher the incidence of disease. Intuitively, we assume that the relationship will be linear, but there many relationships that are supra- or sub-linear, others with threshold effect, and even some that are U shaped, as in Slide 61.

[8] Gary Taubes, "Epidemiology Faces its Limits," 269 *Sci*. 164 (1995) (explaining views of several epidemiologists to seriously consider a threshold relative risk of 3.0 as establishing a causal relationship).

Slide 61 Red Wine Dose-Response Curve



Study Questions:

1. Consider the testimony of experts in a toxic-tort case in which there were epidemiology studies demonstrating an association assessed to be causal for those taking 400 mg/day of the drug:

   Dr. Doherty, and to some extent Dr. Rymer, also rely on studies of Kebraska 400 mg/d to support their opinion of causation at 200 mg/d. Dr. Doherty acknowledges that dose matters with Celebrex, and he takes the relative risk point estimate for 400 mg/d and halves it to support his opinion that Kebraska at 200 mg/d can cause a heart attack.[9]

2. Assess the reliability of these experts' extrapolation of the risk that exists when a dose is halved.

   The methodology here ignores both that a threshold-effect dose may exist and that the dose-response relationship may not be linear. Many supra-linear and sub-

[9] This is a modestly revised version of two experts' testimony in *In re Bextra & Celebrex Marketing Sales Practices & Product Liability Litigation*, 524 F. Supp. 2d 1166, 1180 (N.D. Cal. 2007).

linear dose-response relationships exist in addition to thresholds, as detailed in the materials on toxicology.

The court intuitively appreciated that something was amiss:

> When the Court asked Dr. Doherty if there is anything in the scientific literature to support such primitive extrapolation, he failed to identify any scientific support for his method other than his own judgment. He also admitted that there is no way of knowing what the confidence interval is for 200 mg/d under his unique methodology. Such an unscientific, untested methodology cannot support the proffered opinion of causation at 200 mg/d, especially where, as here, Dr. Doherty agrees with all the other experts that there is a dose effect with Celebrex.

Section III.G.4. Consistency of Association

Are there other studies of this relationship? Are their findings similar or inconsistent? Where a substantial number of studies agree in their findings (and this does not mean they find the exact same magnitude of effect), an outlier may be just that and should not be given undue weight.

Study Question:

1. How should we evaluate the consistency factor if some studies find a statistically significant effect and others do not?

   We would not have thought that anyone with any knowledge of this area would have difficulty with this question—the answer is that these studies are not inconsistent simply because some are statistically significant and others are not, but simply reflect that some were better powered than others. However, the third edition of the Rothman text observes that this mistake "is so common that it deserves special mention."[10]

Section III.G.5. Biological Plausibility

This factor examines consistency with knowledge about the mechanism of the disease being studied. High cholesterol as a factor in heart disease is biologically plausible because it is found in the plaques that clog arteries. Mechanism evidence is highly specific to the disease in question and usually quite inaccessible to non-specialists.

[10] Kenneth J. Rothman et al., *Modern Epidemiology* 27 (3d ed. 2008).

Section III.G.6. Alternative Explanations

Are there biases, study design problems, or potential confounders that may provide a better explanation for the association than causation? Of course, we should recognize that those alternative explanations may not be responsible for the entirety of any effect found, but only influenced it in one direction or another.

Section III.G.7. Specificity

Most agents cause a single disease or series of biologically related diseases. Many nonscientists do not appreciate this. Some toxic-tort cases attempt to demonstrate cancer causation by using a study that finds that the agent causes a different cancer or causes an overall increase in the cancer rate, but that does not address the specific cancer from which the plaintiff suffers. Without evidence that the two cancers operate with the same mechanism (often such evidence is unavailable), these efforts are likely to be unsuccessful.

Study Question:

1. Researchers sometimes are successful in subcategorizing a type of cancer into several different forms of that cancer. Those different subtypes can then be diagnosed as such, which facilitates creating or refining treatment regimens for each of the subcategories. Why is such research welcomed by lawyers who represent defendants in toxic-tort cases?

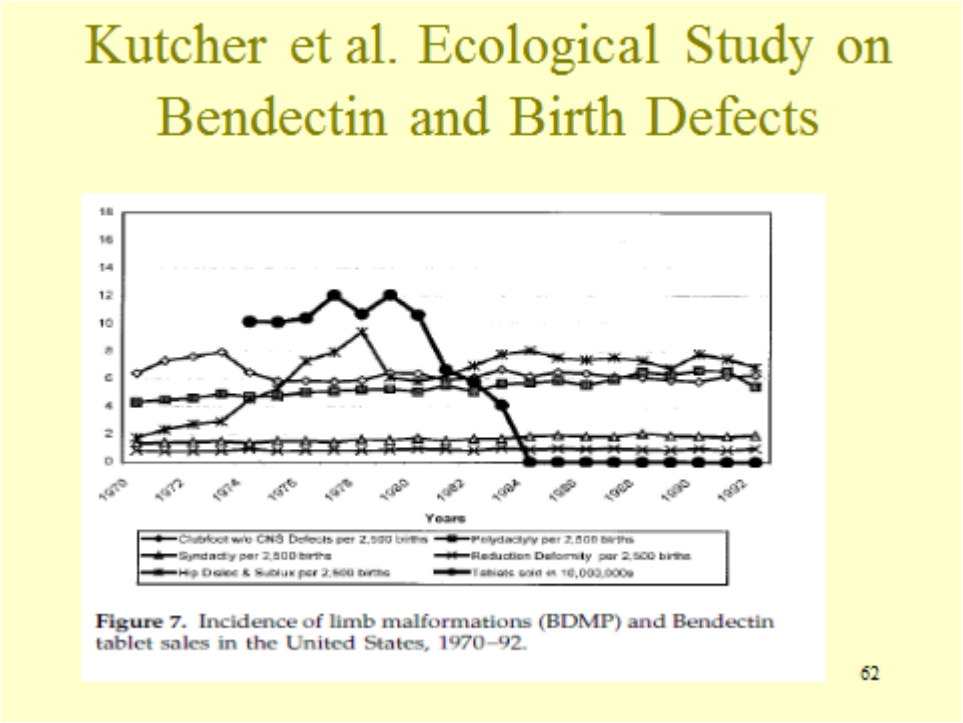
   As diseases are refined into subtypes, defendants can argue that studies of the root disease are not relevant to a particular subtype. With subtypes that are rare, it is unlikely that plaintiffs will have epidemiology, which is difficult to conduct successfully when the disease is rare. Just this phenomenon occurred in an important First Circuit case involving benzene and leukemia. See *Milward v. Acuity Specialty Products Group*, Inc., 639 F.3d 11 (1st Cir. 2011).

Section III.G.8. Consistency with Other Information/Ceasing Exposure

How might a study of tobacco use and cancer mortality have been of use to a researcher in the middle of the 20th century who was studying the association between cigarette smoking and lung cancer? The point is that ecological studies may exist and should be taken into account in deciding whether to draw a causal inference. The Lamm exhibit explained in the module was quite dramatic because of the drastic decrease in Bendectin exposure and the virtually unchanged rate of limb-reduction birth defects. However, as explained in the discussion of ecological studies Section III.B.2.d., supra, caution is required because of the possibility that

some other factor—rather than the one of interest—explains any change in the incidence of disease in the population. We include Slide 62, containing Figure 7 of the Kutcher study, in case instructors want to display it in class.

Slide 62 Kutcher et al. Ecological Study on Bendectin and Birth Defects



SOURCE: Jeffrey S. Kutcher, "Bendectin and Birth Defects II: Ecological Analyses," 67 *Clin. & Molecular Teratology* 88 (2003). © 2003 Wiley-Liss, Inc.

Study Questions:

1. How can cessation of exposure to an agent support a causal relationship?

   Following a significant decrease in the incidence of smoking in California, there were significantly fewer cases of lung cancer. The association between cessation of smoking and the decrease in the incidence of lung cancer either suggests there may be a causal relationship, or provides evidence in support of an epidemiology study based on individual data that finds such an effect. Of course, before drawing any such conclusion, the researcher should inquire into whether some other change in the population during this time led to the decrease in lung cancer.

2. What role might a latency period play in evaluating a causal relationship following

cessation of exposure?

It is important to know the latency period for the disease you are evaluating so you will know when to assess the effects of cessation. For example, if the latency period for developing lung cancer is 25 years, you cannot assess the effects of smoking cessation after only 10 years.

Section III.H. Multiple Studies and Meta-analysis

Our inclination would be to focus the meta-analysis discussion on the Li abstract,[11] which is in the module and reproduced below in Slide 63.

Slide 63 Li Abstract

The abstract discusses the process of conducting a meta-analysis. Gathering and screening studies is well described. A case-crossover study design is used for short exposures that produce acute results. Each subject is both a case and control, so that they are both exposed and disease determined. Then after the acute disease is resolved and exposure ends, they are observed to determine the incidence of disease without exposure. Obviously, this

[11] The abstract is Copyright © 2014, Bioscientifica, Ltd.

study design is inappropriate for chronic diseases. “Case-crossover,” which sounds like a variation on a case-control study, is misleading in that respect as exposure is the independent variable and incidence rates of disease are the dependent variable, as with cohort studies.

Note the stunningly high p-values for very small effects—not a single relative risk exceeds 1.2, yet each one is statistically significant with very small ps. The confidence intervals for these results are exceedingly narrow. Of course, other sources of error could have produced these results, but when multiple, independent studies are involved, one is less concerned about bias and confounding.

Study Questions:

1. In what way can a meta-analysis provide information that other studies cannot?

   Often, published studies will have inconsistent results. This can be due to a number of reasons, including random sampling error, bias, and confounding. These studies are often not only inconsistent, but also may fail to find any statistical significant result because of insufficient numbers of subjects. Meta-analysis allows us to look at the results from multiple studies, assess the study design of each study, and pool the results to arrive at a single figure that can be used to assess a potential association.

2. What are the potential pitfalls of using a meta-analysis for observational studies?

   Meta-analysis is ideally used for randomized-controlled trials. As previously discussed, these studies are not always feasible and we must resort to observational studies.

   However, observational studies have inherent biases and often varying study designs.

3. Li et al. originally identified 925 articles, but only 36 were reviewed in-depth. Why would the authors choose to only examine 36 out of 925 studies?

   Articles are often excluded due to overlapping records, for being obvious or irrelevant studies, for being review articles rather than reports of new research, or for poor study design.

4. How can a study that finds as small of relative risks as those in this study produce a statement that the outcome is “significantly associated with an increased risk.”

   To repeat what we have already said on this subject in Section III.E.1., supra, “significance” in this context is not about the magnitude of effect—these effects are

quite small, involving an increase in mortality ranging from under 1% to just over 12%. Rather, significance addresses the statistical stability of these results and, as the lower bound of each of the confidence intervals being above 1.0 reveals, all of these results are statistically significant.

5. Would any of those who were exposed to air pollution and died from diabetes-related diseases be able successfully to sue the entity responsible for the air pollution (assume only one such entity and that its conduct in producing the pollution was tortious)?

   We use this as a prelude to the section on specific causation. It affords an opportunity to introduce students to the relationship between the preponderance of the evidence standard of proof and relative risks.

Section III.I. Reading an Epidemiologic Study

We chose the Demers study because it is a fairly standard occupational-epidemiologic study that affords an opportunity to review a number of concepts covered in these materials. It also provides an opportunity to consider how a court, confronted with one of the causal issues investigated in the study, read the study—the *Lindquist* case follows. The study also plays a role, less significantly, in the *Estate of George* case in Section V.F.1., infra.

The study questions below are basic ones for any consumer of an epidemiologic study and are adapted from Alexander M. Walker, "Reporting the Results of Epidemiologic Studies," 76 *Am. J. Pub. Health* 556 (1986).

Study Questions:

1. What causal relationship(s) were the authors investigating, that is, the dependent variable(s) and the independent variable(s)? How many relationships did the study investigate?

   To start with the last inquiry above, one might respond with one or quite a few. The study examined the incidence of mortality (deaths per the 45-year follow-up period) for firefighters. But the authors examined the cause of death (and thereby the incidence of disease) for participants. As Table 2 reveals, there were approximately 19 different cancers and roughly 14 other diseases examined in this study. The independent variable, that is, the cause being examined, was exposure to fire smoke.

2. What is the composition of the exposed, case, and control groups? In a case-control study, how are the controls selected and matched to the cases?

This is a cohort study. The study explains that this is a follow up of an earlier retrospective cohort study, but students should be able to discern this from the abstract, which explains that the study examined the effect of fire smoke on firefighters (the exposed cohort) by comparing them with national mortality data (one control group) and police officers (the other control group).

3. How was relevant data obtained?

A variety of sources provided the data for this study. To determine who was in the exposed cohort of firefighters, records of the fire department and pension boards of the three cities involved were used. For identification of the police control group, similar records were employed. State and federal records were employed to determine deaths and cause of death for those in each of these groups. White males in the United States constituted the other control group, and this data was obtained from NIOSH. Given the sources of data and the very high completion rate reported, this aspect of the study does not raise significant concerns about bias.

4. What results did the study find?

Demers et al. examined over 30 different diseases responsible for death in the study group. The independent variable was exposure to fire smoke. It is worth observing that with such a large number of investigations, the probability is that the study will find at least one statistically significant relationship that is due purely to random error. Examining large numbers of relationships without some specific hypothesis about why an association might exist is sometimes known colloquially (and pejoratively) as data dredging. Most notably, from Table 2, the authors found an SMR of .81 (a protective effect) for mortality due to all causes. Although this result is statistically significant, one suspects that a healthy-worker bias could have been responsible for it. Yet, when the control group is police officers, the protective effect remains. The penultimate substantive paragraph of the study explains reasons why the results in comparison to police officers may not be entirely reliable. There are any number of other results in this study that students may identify and instructors may wish to discuss.

5. What confounders were considered and what are the results of the study adjusting for those confounders?

As discussed in the study, good health may be differentially associated with firefighters compared to the general population. This is a confounding effect that would distort the true association to a lower level, which may explain some of the

protective effects found in Table 2. Rather than gathering data on the state of health of each worker (students should see why that would have been difficult, if not impossible), the authors employed an alternative control group—police officers—in which any health effects should wash out.

The study also explains that the police-officer control group may have been confounded by its exposure to unique occupational risks faced by police officers (certainly fatal attacks by criminals is a confounding risk, although one that is probably quite minimal—fatal shootings of police officers average about 50 per year).

6. What is the p-value for any association found in the study? What boundaries exist on the confidence intervals for the study outcome? What level of association appears unlikely due to random error in light of the study results, that is, associations beyond the confidence intervals)?

   Numerous associations exist for discussion purposes. Instructors might attempt to get students to express and explain their reasons for accepting or discounting a specific association found in this study.

7. What potential biases do the investigators identify and explain? Are there other potential biases the study does not discuss?

   Information bias is explained by the authors: cause of death in death certificates may not have been accurate, in which case the results would be biased, although it may be non-differential if the reference population was subject to the same inaccuracies.

Lindquist v. City of Jersey City Fire Department, 814 A.2d 1069 (N.J. 2003):

The *Lindquist* case is paired with the Demers study because the court did some extracurricular research, locating and reading the Demers study and relying on it in the course of reversing the lower court and reinstating plaintiff's workers' compensation award. The court's reliance on Demers provided an opportunity to assess how well nonscientists consume science—a subject that students in this course should embrace. It also presented an occasion to consider the various causal roles of exposure to another environmental risk factor—plaintiff smoked three-fourths of a pack of cigarettes per day for 22 years, approximately 17-pack years of cigarette exposure—but one for which the defendant is not responsible. Finally, it provided a rich opportunity to clarify vague and potentially confusing judicial language about scientific research. We begin with the last of those three items.

As a preliminary matter, it may be wise to clarify that chronic obstructive pulmonary disease (COPD), with which plaintiff was diagnosed, is an umbrella term for a variety of specific

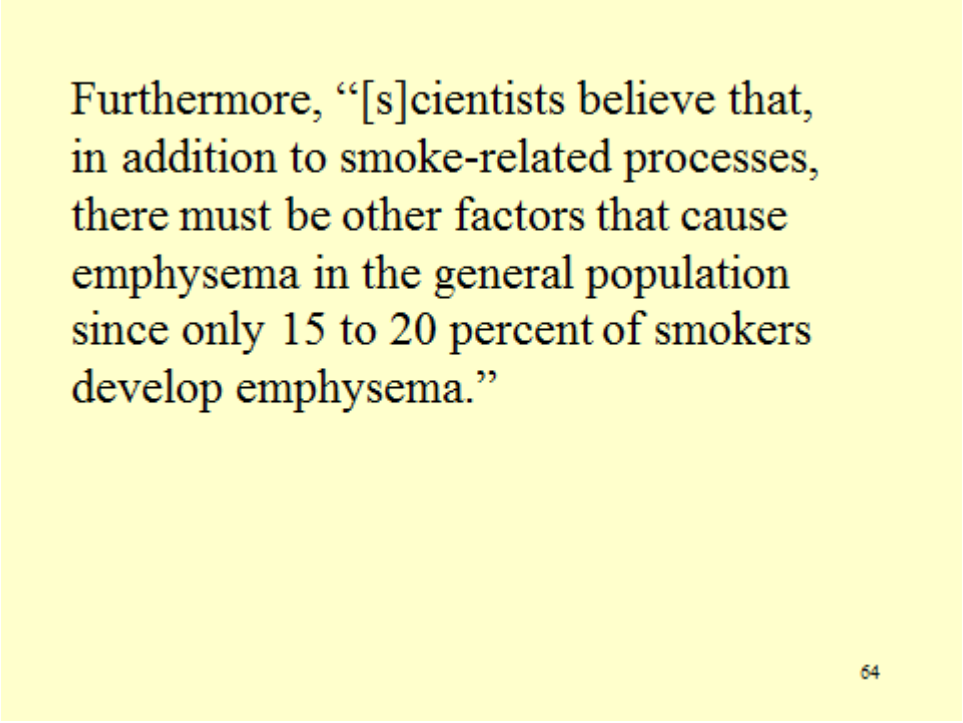
diseases, including emphysema. The Demers study examined the relationship between firefighting and both emphysema and COPD. As the ICD-9 codes (International Statistical Classification of Diseases and Related Health Problems) in those tables reveal, emphysema was removed from the umbrella COPD classification for separate assessment.

Consider that plaintiff was exposed to fire smoke for 23 years and was diagnosed with emphysema in 1995 at the age of 47. We should also appreciate that the Demers study is of mortality, not disease. Life expectancy of those diagnosed with emphysema is not well studied, but one study suggests an overall median survival for those with severe emphysema to be in the range of 4 years, with variation depending on a number of factors, especially age.[12]

Study Questions:

1. The *Lindquist* court states (Slide 64):

Slide 64 *Lindquist* Excerpt

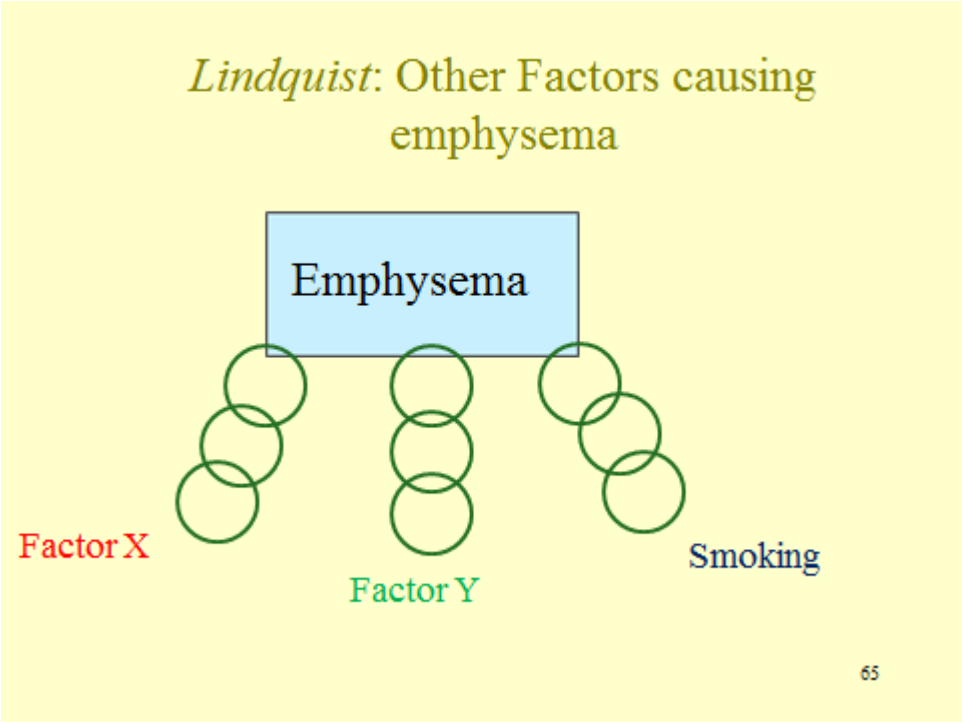


What, precisely, does the court mean?

[12] Fernando J. Martinez et al., "Predictors of Mortality in Patients with Emphysema and Severe Airflow Obstruction," 173 *Am. J. Respiratory & Critical Care Med.* 1326 (2006).

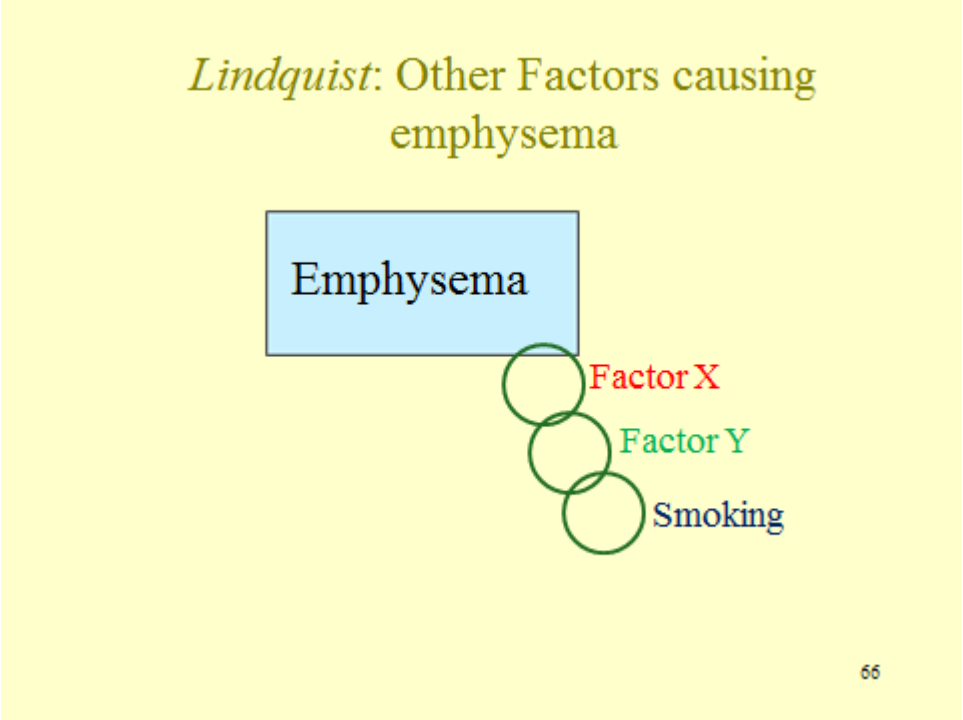
One might reasonably understand this statement (and the court seems to think it means this, as shown by the last paragraph of the opinion) to mean that which is depicted in Slide 65, which shows other competing causal chains that each contain a factor.

Slide 65 *Lindquist*: Other Factors Causing Emphysema



But there is a logical fallacy here. Saying that only 20% of smokers contract emphysema only means that smoking is not a sufficient cause and that other factors are required before a smoker contracts emphysema, as depicted in Slide 66. Thus, the reason only a fraction of smokers contract emphysema is because one or both of factors X and Y are missing from the majority of those who smoke.

Slide 66 *Lindquist*: Other Factors Causing Emphysema



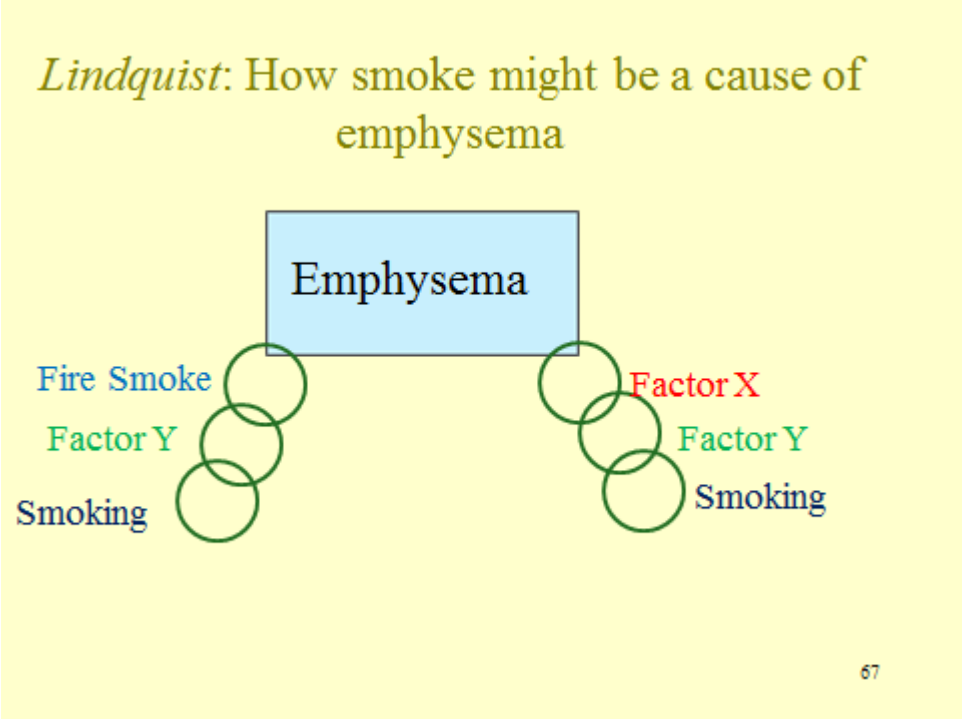
The court appears to endorse the causal relationship reflected in Slide 65 because that would leave room for fire smoke to be a cause of emphysema. However, that explanation, aside from it not following from the statement about what scientists believe, runs into the cross-examination concession by Dr. Eisenstein that he could not cite any studies in which nonsmoking firefighters developed emphysema. One explanation, of course, is that no such studies exist because no researchers attempted to investigate this relationship. Another is that there some low-powered studies, but they are inconclusive rather than exonerative. But even if it is true that no nonsmoking firefighters contract emphysema, might fire smoke still be a cause of emphysema? This leads to the next study question.

2. Assume, as suggested by Dr. Eisenstein's testimony, that no nonsmoking firefighters develop emphysema. Does that mean that fire smoke is not a cause of emphysema?

Defense counsel obviously argued that it did, and the court seems to have some sympathy for that notion—although it is inconsistent with the causal scenario shown in Slide 65 and with which the court dallies. Instead, however, consider the circumstance shown in Slide 67, in which all causal chains leading to emphysema contain smoking, but different ones contain other factors, one of which could be fire

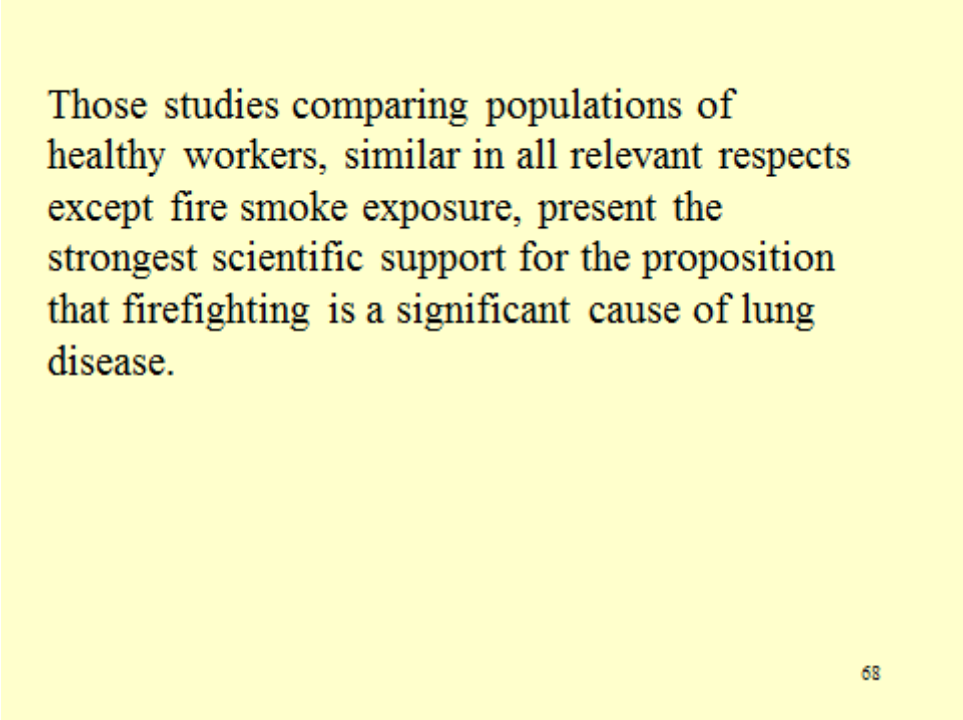
smoke. Thus, while emphysema is a signature disease for smoking that does not logically rule out other causes being necessary for some incidence of emphysema.

Slide 67 *Lindquist*: How Smoke Might Be a Cause of Emphysema



3. The court interprets the Demers study and its predecessor to "present the strongest scientific support for the proposition that firefighting is a significant cause of lung cancer" (Slide 68) and relies to a considerable extent on it to conclude that plaintiff should prevail. Assess the court's use of the Demers study.

Slide 68 *Lindquist* Excerpt re Demers Study



Slide 69 contains Table 2 from the Demers study with the line for emphysema highlighted. It reveals a relative risk of 1.19 with a confidence interval of .72–1.83, fairly narrow but not statistically significant at a .05 level. (The actual p-value is not disclosed.) One might think that the magnitude of any association was skewed lower by the healthy-worker bias, but this is thin gruel for a general-causation conclusion. As stated in the section on inferring causation supra, most epidemiologists want multiple observational studies that find a minimum relative risk of 3 before being willing to draw a causal inference. The result improves a bit in Table 3 (in Slide 70), which employs police officers as the control group, but not by much. The relative risk increases to 1.45, but with a wider confidence interval. Given the concerns about information bias discussed by the authors, as well as the inverted U shape of the dose-response curve for emphysema in Table 6 (in Slide 71) (but note the authors' speculation that older, susceptible firefighters may have left the workplace due to disability), we are skeptical of the court's reliance on the Demers study for its conclusion. Indeed, we suspect that 9 out of 10 courts confronted with the Demers study would have decided the *Lindquist* case differently.

Slide 69 Seattle, Portland, and Tacoma Firefighter Mortality: 1945-89

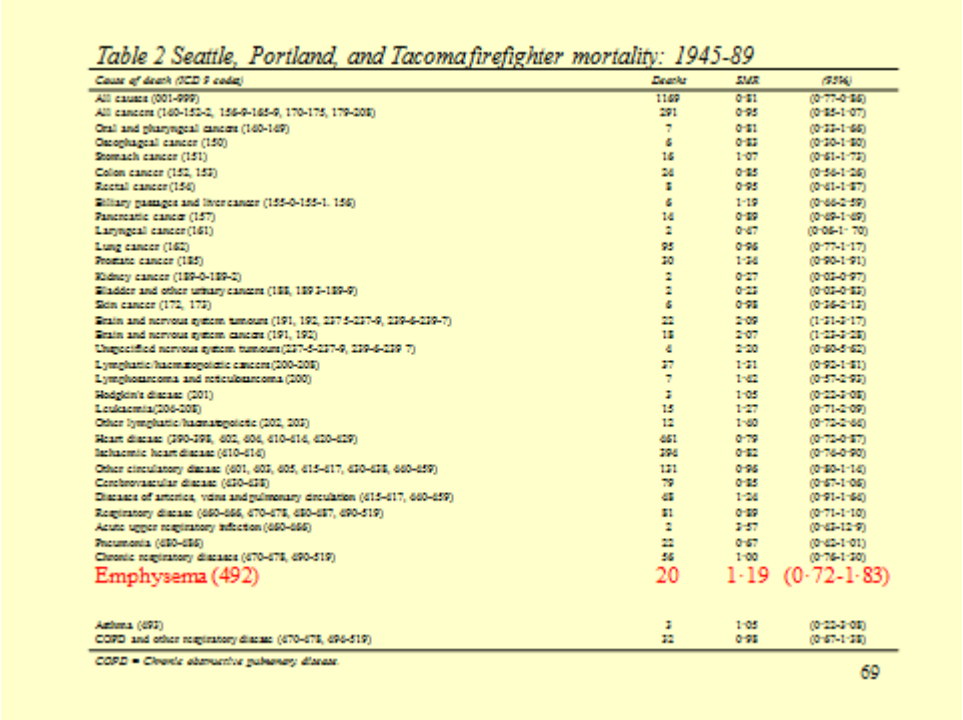


Table 2 Seattle, Portland, and Tacoma firefighter mortality: 1945-89

Cause of death (ICD 9 code)	Deaths	SMR	(95%)
All causes (001-999)	1169	0·81	(0·77-0·86)
All cancers (140-152-2, 154-9-165-9, 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)	26	0·85	(0·56-1·26)
Rectal cancer (154)	9	0·95	(0·61-1·87)
Biliary passages and liver cancer (155·0-155·1. 156)	6	1·19	(0·66-2·59)
Pancreatic cancer (157)	14	0·89	(0·49-1·49)
Laryngeal cancer (161)	2	0·67	(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 and 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)	27	1·31	(0·92-1·81)
Lymphosarcoma and reticulosarcoma (200)	7	1·62	(0·67-3·33)
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	1·60	(0·72-2·66)
Heart disease (390-398, 402, 404, 410-414, 420-429)	661	0·79	(0·72-0·87)
Ischaemic heart disease (410-414)	296	0·82	(0·76-0·90)
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	(0·91-1·66)
Respiratory disease (460-466, 470-478, 480-487, 490-519)	81	0·89	(0·71-1·10)
Acute upper respiratory infection (460-466)	2	2·57	(0·42-12·9)
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, 496-519)	32	0·98	(0·67-1·38)

COPD = Chronic obstructive pulmonary disease.

69

SOURCE: 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.

Slide 70 Seattle, Portland, and Tacoma Firefighter Mortality Compared with Police and Police Mortality Compared with United States White Male Rates: 1945-89

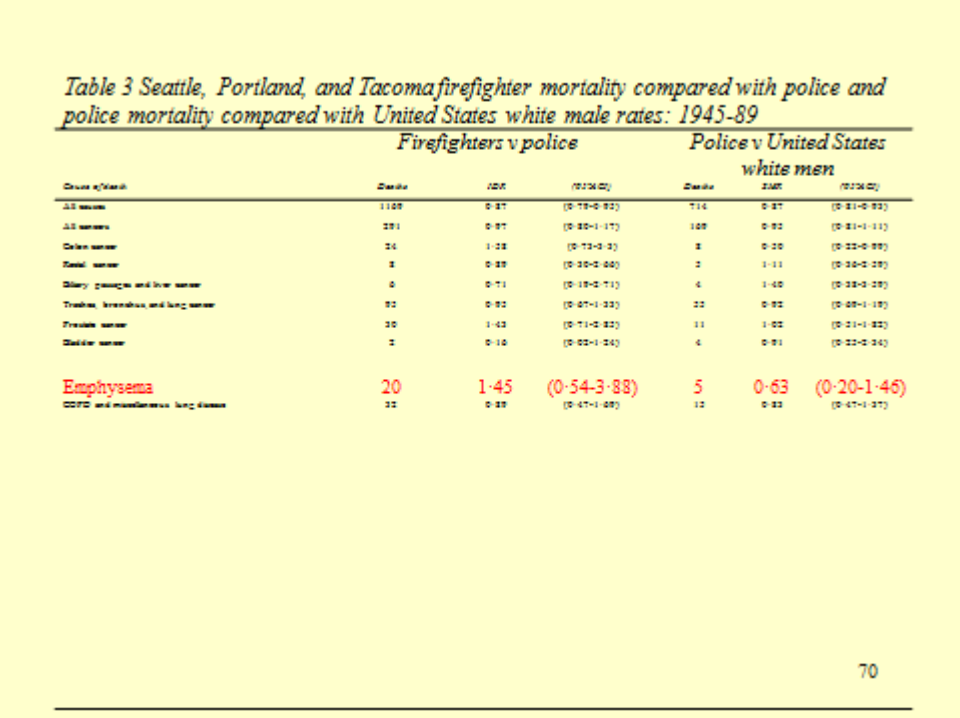


Table 3 Seattle, Portland, and Tacoma firefighter mortality compared with police and police mortality compared with United States white male rates: 1945-89

	Firefighters v police			*Police v United States white men*		
Cause of death	Deaths	SMR	(95% CI)	Deaths	SMR	(95% CI)
All causes	1169	0·87	(0·79-0·93)	716	0·87	(0·81-0·93)
All cancers	291	0·97	(0·86-1·17)	169	0·92	(0·81-1·11)
Colon cancer	24	1·28	(0·73-3·3)	8	0·50	(0·22-0·98)
Rectal cancer	8	0·59	(0·30-0·99)	3	1·11	(0·36-2·58)
Biliary passages and liver cancer	6	0·71	(0·19-2·71)	4	1·45	(0·39-3·69)
Trachea, bronchus, and lung cancer	93	0·93	(0·67-1·33)	33	0·82	(0·56-1·15)
Prostate cancer	30	1·43	(0·71-2·82)	11	1·02	(0·51-1·82)
Bladder cancer	2	0·16	(0·02-1·24)	4	0·91	(0·22-2·34)
Emphysema	20	1·45	(0·54-3·88)	5	0·63	(0·20-1·46)
COPD and miscellaneous lung disease	33	0·89	(0·47-1·66)	13	0·83	(0·47-1·37)

70

SOURCE: 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.

Slide 71 Seattle, Portland, and Tacoma Firefighter Mortality by Duration of Exposed Employment: 1945–89

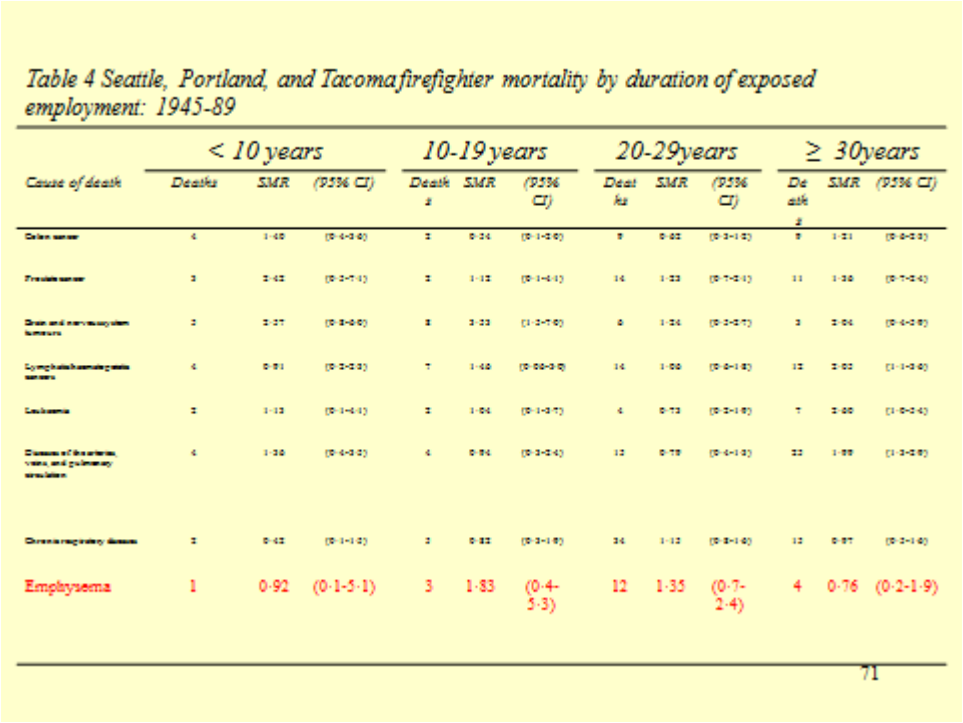


Table 4 Seattle, Portland, and Tacoma firefighter mortality by duration of exposed employment: 1945-89

	< 10 years			10-19 years			20-29 years			≥ 30 years		
Cause of death	*Deaths*	*SMR*	*(95% CI)*	*Deaths*	*SMR*	*(95% CI)*	*Deaths*	*SMR*	*(95% CI)*	*Deaths*	*SMR*	*(95% CI)*
[illegible]	4	1·40	(0·4–3·6)	2	0·54	(0·1–2·0)	9	0·65	(0·3–1·2)	9	1·21	(0·6–2·3)
[illegible]	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)
[illegible]	2	2·27	(0·3–8·2)	3	3·23	(1·2–7·0)	6	1·24	(0·5–2·7)	3	2·04	(0·4–5·9)
[illegible]	4	0·81	(0·2–2·1)	7	1·43	(0·58–2·9)	14	1·00	(0·6–1·6)	12	2·02	(1·1–3·6)
[illegible]	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)
[illegible]	4	1·36	(0·4–3·5)	4	0·84	(0·2–2·1)	12	0·78	(0·4–1·3)	22	1·88	(1·2–2·8)
[illegible]	2	0·42	(0·1–1·5)	2	0·82	(0·2–1·9)	24	1·12	(0·8–1·6)	12	0·87	(0·5–1·5)
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)

71

SOURCE: 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.

4. The court states, in its concluding paragraph: "We emphasize 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." What does the court mean with regard to an agent "contribut[ing] to a material degree"?

   Got us. Doses of a toxin that are super-sufficient for the outcome—multiple sufficient causes—aside, we do not know how one can talk relatively about causes when using a model that requires a cause be necessary (but for) for the outcome. Such necessary causes either are or are not necessary for the outcome, and one cannot speak of them in degrees; like pregnancy, you either are or you are not. And, if the court is not employing necessity as its model for causation, we do not know what its alternative is. That reflects the mushy and standardless usage of "substantial factor" for causation that we put aside at the outset of these materials for the reasons stated there.

SECTION IV. TOXICOLOGY

Section IV.A. The Science of Toxicology

Section IV.A.1. Introduction

Three themes underlie much of the discussion in this section. First, as the opening paragraph indicates, a key point to be made in the early part of the toxicology discussion is that this discipline, unlike epidemiology, focuses on how something causes harm (or benefit) to an organism. A second theme is the emphasis on the external validity problems that exist when we attempt to apply the results of many toxicology results to humans. Finally, at several points, the chapter visits the thorny question of what to do with the no-threshold assumption usually made with respect to substances that are carcinogens.

As in the epidemiology section, we have attempted to keep the module relatively note-free, but we have included notes to many points raised there in these materials. This section draws heavily on Karen Stine &Thomas Brown, *Principles of Toxicology* (3d ed. 2015); Bernard D. Goldstein & Mary Sue Henifin, "Reference Guide on Toxicology," in Federal Judicial Center (ed.), *Reference Manual on Scientific Evidence* (3d ed. 2011); and Bernard Goldstein, "Toxicology," in David Faigman et al. (eds.), *Modern Scientific Evidence: The Law and Science of Expert Testimony*, at ch. 22 (2014–2015). The instructor may wish to have students research the toxic effects of some substances using the U.S. National Library of Medicine's TOXNET web pages, http://toxnet.nlm.nih.gov/ (last visited Apr. 20, 2016), and the TOXLINE embedded database, http://toxnet.nlm.nih.gov/newtoxnet/toxline.htm (last visited Apr. 20, 2016).

This section is divided into two parts. Part I provides an overview of toxicology. We present these materials without cases interspersed because we believe the toxicology issues presented in appellate opinions are best appreciated after the student has gained an overview of the field. Part II then presents three cases designed to provide an example of the use of toxicological evidence in the regulatory arena, in a criminal case, and in a civil case. Some instructors may wish to skip the second case, a criminal case discussing what the courts call retrograde extrapolation. While it does involve the use of knowledge of toxicokinetics, it does not address a question of causation, that is, whether some substance caused an injury.

At the end of each section of Part I, we present a set of questions concerning the materials in this part. The instructor may wish to raise these questions after completing each section or defer and ask the questions at the end of Part I as a general review of the materials. There are, of course, many other questions one might ask, and the instructor should consider expanding on this list. We hope the answer to nearly all of these questions is clear in the materials, but nevertheless, we have provided answers to each.

At the end of the introduction, we present an excerpt of the Tyl study to give students a flavor of the nature of toxicological research. We include the excerpt here to give students a flavor of an actual toxicological study. Some instructors may wish to discuss the Tyl et al.

excerpt here. On the other hand, some may wish to defer a discussion until students get to the in vivo discussion later in the materials.

The two tables presented at the outset contain only part of the information found in the article. They focus entirely on skeletal defects, although the article also reports on external and visceral defects. In addition, the authors discuss and report on what they call "variations," which are differences among individuals that do not rise to the level of malformations. The classification of differences as either malformations or variations was already in place prior to the current study. The authors found no differences among exposure groups with respect to variations.

The tables also exclude skeletal examinations that did not produce significant effects. For example, there were no group differences with respect to posterior limbs and, therefore, we did not include these results in the first table. The authors looked for differences between male and female fetuses, but found none. Therefore, we also excluded this data. All of this was done to focus student attention on the main issues discussed below.

If the instructor does wish to spend some time with the two tables, it certainly would be worth explaining a bit about what is reported, especially the standard error values provided in the first table. The standard error of the mean quantifies how precisely you know the true mean of the population. It takes into account both the value of the standard error and the sample size. Here is one explanation of the SEM taken from the web:

> What is the standard error of the mean?
>
> The standard error of the mean (SEM) estimates the variability between sample means that you would obtain if you took multiple samples from the same population. The standard error of the mean estimates the variability between samples whereas the standard deviation measures the variability within a single sample.
>
> For example, you have a mean delivery time of 3.80 days with a standard deviation of 1.43 days based on a random sample of 312 delivery times. These numbers yield a standard error of the mean of 0.08 days (1.43 divided by the square root of 312). Had you taken multiple random samples of the same size and from the same population the standard deviation of those different sample means would be around 0.08 days.
>
> Use the standard error of the mean to determine how precisely the mean of the sample estimates the population mean. Lower values of the standard error of the mean indicate more precise estimates of the population mean. Usually, a larger standard deviation will result in a larger standard error of the mean and a less precise estimate. A larger sample size will result in a smaller standard error of the mean and a more precise estimate.

A more thorough discussion with useful examples may be found in John H. McDonald, *Handbook of Biological Statistics, Standard Error of the Mean* (3d ed. 2014), http://www.biostathandbook.com/standarderror.html.

The SEM serves the same purpose as a confidence interval discussed in the epidemiology materials. One should also point out that in the first table, the unit of analysis is the litter, not the fetus.

Substantively, one may wish to point out how much larger the doses given to the rats were than the dose taken by pregnant women. Later in these materials, we discuss the reasons for this decision and the problems it presents when attempting to use in vivo research to assess human risk.

Conceptually, the central issue confronting the authors of this study is how to interpret the greater incidence of defects among the animals receiving the highest doses. One may wish to ask the students the following two questions.

Study Questions:

1. What are the competing hypotheses for the malformations exhibited by fetuses whose mothers were exposed to 800 mg/kg/day?

   The answer, of course, is whether the defects are caused by specific teratogenic properties of Bendectin or whether they are caused by maternal and fetal toxicity caused by the very high dose of Bendectin received by some dams. That is, they are a result of the fact that such a large dose is caused by, among other things, weight loss and lack of proper nutrition for both the mother and fetus, and this in turn caused relatively greater malformations in the fetuses.

2. How did the researchers attempt to resolve which hypothesis is correct? Here, one should point the students to the particulars of the authors' discussion.

   First, the authors discussed the various signs of maternal and fetal toxicity. Most obvious, of course, is the number of dams that died when receiving the highest dose and other dams that simply reabsorbed their fetuses. Remaining dams had a lower body weight and lower water intake during treatment period but greater water intake after the treatment ceased.

   Fetuses at higher dose rates were more likely to have a short thirteenth rib, but this defect has been observed in other studies where maternal toxicity was present.

   Most statistically significant effects occurred at higher doses. The one effect observed at the 200 mg/kg/day dose is questionable because the effect was very small and the results may be due to intercoder variability because of differences in visual acuity. (Ideally, the same individual would have coded all fetuses or at least each coder should have coded an equal number of fetuses from each exposure

group. Thus, the problem of intercoder variations introduces another confounder into the study.) More importantly, perhaps, the authors question the biological relevance of this small effect, especially because while there is evidence of incomplete ossification, there is no indication of abnormal ossification (the type of thing one would observe with limb-reduction defects) "in any district of the fetal skeleton."

The last prong of their justification is the authors' ANCOVA (analysis of covariance) analysis. The ANCOVA statistic discussed in the paper can be explained relatively easily. ANCOVA evaluates whether population means of a dependent variable (e.g., number of skeletal malformations) are equal across levels of a categorical independent variable (e.g., the dose of Bendectin given to the rats), while statistically controlling for the effects of other continuous variables, called covariates, (e.g., fetal body weight), that may also be explaining differences in the dependent variable. An ANCOVA analysis decomposes the variance in the dependent variable into variance explained by the covariate(s), variance explained by the categorical independent variable, and residual variance. ANCOVA can be thought of as "adjusting" the dependent variable by the group means of the continuous variable(s). Although the authors perhaps should have said more about the results of this analysis, it appears that they found a significant reduction in mean differences among dose levels when controlling for body weight. (As an aside, one could discuss whether the authors adequately report their results in this paper.). As part of this discussion, they note that Bendectin seems to have had a greater effect on fetal weight than ossification.

At the end of the day, one might ask the students if the authors' interpretation is correct and Bendectin's effects are all due to toxicity. Even if one were to conclude that there might be some residual doubt, what is clear is that this study suggests that at 200 mg/kg/day, that is, 100 to 200 times the dose taken by pregnant women, no malformations were reliably observed and in none of the fetuses did the researcher observe malformations of a type similar to the limb-reduction defects Bendectin allegedly caused in the children of mothers who took the drug.

Section IV.A.2. Toxic Effects

A more complete description of the mutational theory of carcinogenesis and the several steps thought to be necessary before neoplastic cells develop into a malignancy can be found in Stine & Brown, supra, which discusses these stages and competing theories of cancer causation.

It is important to understand that while some substances cause a single type of injury, some cause multiple injuries. As the module notes, tobacco is a prime example. For a general review of the myriad of illnesses associated with tobacco use, see Joaquin Barnoya et al., "Tobacco," in Faigman et al. (eds.), supra, at ch. 25.

The DES litigation has turned primarily on a purely legal issue: whether one can sue a group of manufacturers even when, because of the passage of time, one cannot discover which defendant actually supplied the product to the plaintiff's mother.

As the module notes, cigarette smoke is a particularly toxic soup. Some sources list even more than 60 or 70 potential carcinogens. A list of these pollutants in a particularly useful discussion of exposure to second-hand smoke can be found in James Repace, "Exposure to Secondhand Smoke," in Wayne Ott et al. (eds.), *Exposure Analysis* (2007).

Thalidomide poses an interesting case that may be worth mentioning. It is one of the most devastating teratogens yet discovered, but it also has proven to be a useful drug in the treatment of multiple myeloma and some of the symptoms of leprosy. The lesson to be learned from this example is that organisms at different stages react differently to substances and, perhaps more importantly, most drugs are designed to affect organisms and organ systems. If they did not do so, they would be no more than a placebo. But unfortunately, it is quite rare that every effect is beneficial in every circumstance.

The cases presented in the law section below involve substances that affect humans in at least one of the following ways: *Chlorine Chemical*—cancer; *Tsosie*—central nervous system; *Johnson*—fibrosis. The instructor may wish to return to this list when discussing each case.

Study Question:

1. What are the distinct ways xenobiotics cause harm to organisms?

   The module lists the following effects. Xenobiotics may be carcinogens and teratogens, causing birth defects. They may also cause fibrosis. They may cause damage to the cardiovascular system. Finally, they may affect the central nervous system. This is not an exhaustive list, but it provides a flavor of the varieties of injuries caused by toxic substances.

Section IV.A.3. The "Laws" of Toxicology

From the point of view of the law-science interface, the most important tenet is the first: the dose makes the poison. At this point, it may be worth expanding on this tenet in several ways. First, there is the distinction between acute and chronic exposures. With respect to both, there is a dose-response relationship, but the two kinds of exposure may produce very different results. Some adverse effects, such as fibrosis, are unlikely from acute exposures. (This point is explored in depth in the *Johnson* case presented below.)

Second, drugs present an interesting variation on this tenet. Beneficial drugs generally have little or no beneficial effect at very low doses. There is some dose, which may vary among users, at which drugs reach their maximum beneficial effect. Second- and third-phase clinical trials—discussed later in the module—are designed in part to zero in on this dose. But as dose increases beyond this level, the drug may show increased adverse effects. For a case on point,

where the individual was given an overdose of Danocrine and developed primary pulmonary hypertension, see *Zuchowicz v. United States*, 140 F.3d 381 (2d Cir. 1998).

It may also be worth mentioning that a number of toxicologists believe that endocrine-disrupting chemicals, one of which is DES, do not obey this "law." See Laura N. Vandenberg et al., "Hormones and Endocrine-Disrupting Chemicals: Low-Dose Effects and Nonmonotonic Dose Responses," 33 *Endocrine Revs.* 378 (2012). We return to this issue later in the module.

The benzene standard can be found at Occupational Exposure to Benzene, 52 Fed. Reg. 34,460 (Sept. 11, 1987) (to be codified at 29 C.F.R. pt. 1910). Later in the Teaching Materials (but not the module), we discuss a Supreme Court case dealing with this standard. At this point or later, the instructor may wish to point students to the OSHA web page that discusses benzene standards. Occupational Safety & Health Admin., U.SDep't of Labor, Safety and Health Topics: Benzene, https://www.osha.gov/SLTC/benzene/standards.html (last visited Apr. 20, 2016).

If the first tenet is particularly important in legal contexts, the second is at the heart of toxicology itself. An overarching toxicology goal is to understand the specific ways each substance interacts with organisms to produce its effect. This involves an understanding of the chemical and physical structure of the substance and the biology of the organism, which increasingly involves an understanding of genetics.

The last tenet justifies research on animals as a window into the effect of substances on humans, but of course the devil is in the details and as discussed below, there are many reasons why different species react differently to toxins.

Study Question:

1. What are the three basic tenets of toxicology?

   (1) The dose makes the poison; (2) Xenobiotics produce specific patterns of biological effects due to the unique chemical structure of the agent and the laws of biology that govern the organism's response; and (3) humans are members of the animal kingdom and, therefore, by studying Xenobiotic effects on animals we may better understand their effects on humans. Note, however, that at many places in the following discussion, we point out ways in which animal models fail to replicate effects on humans.

Section IV.A.4. Areas of Toxicological Research

The "forensic sciences" in the criminal law area have come under sustained attack for a failure to employ basic scientific methods to test the validity of their practices. Much of forensic toxicology stands outside this critique. Nevertheless, many questions of interest to law remain unanswered. Some of them are outlined in the discussion surrounding the *Tsosie* case presented below.

A useful overview of the state of environmental toxicology may be found in Comm. on Toxicity Testing & Assessment of Envtl. Agents et al., *Toxicity Testing in the 21st Century: A Vision and a Strategy* (2007), available online and as a free PDF file at http://www.nap.edu/catalog/11970/toxicity-testing-in-the-21st-century-a-vision-and-a. The book addresses many issues raised in these materials, including problems of extrapolation and the fact that we do not presently have the resources to test the safety of most chemicals used in society today. Some instructors may wish to look over this book and assign some readings, especially the materials on the committee's vision for the future of toxicological testing.

In some ways, pharmacological toxicology is the most well-developed area of toxicology. This is so because human testing is possible. We lay out the process of new drug testing in the module. However, these materials do not present a drug case, in part because these cases more often than not involve epidemiological data as well as toxicological evidence.

Study Question:

1. Name and describe the basic subdisciplines of toxicology.

   This section sets out three broad areas of toxicology research: forensic toxicology, environmental toxicology, and pharmaceutical toxicology. Of course there are many other ways to parse the discipline, and we discuss other categorizations below.

Section IV.A.5. Toxicokinetics & Toxicodynamics

a. Pharmacokinetics and Toxicokinetics

These two areas of investigation are closely tied to similar enterprises undertaken with respect to drugs: pharmacokinetics and pharmacodynamics. A useful, recent book discussing toxicokinetics and pharmacokinetics is Mehdi Boroujerdi, *Pharmacokinetics and Toxicokinetics* (2015). This discussion provides a general overview and makes no attempt to draw distinctions between these related areas.

The ADME cycle is often the first topic nontoxicologists are exposed to, and for good reason. An understanding of absorption, distribution, metabolism, and excretion is a fundamental building block to understand, among other things: the primary reason in vitro research results may differ from in vivo results, why chemicals with similar structures may not have the same effects on humans, and why findings in non-human animals are not always translatable to humans.

i. Absorption

The brief discussion of dermal absorption comes in part from Alesia Ferguson et al., "Dermal Exposure, Uptake, and Dose," in Wayne Ott et al. (eds.), *Exposure Analysis* 255 (2007). The entire book is a useful resource on exposure. The reference to the bioavailability of dioxins comes from Thomas Umbreit et al., "Bioavailability of Dioxin in Soil from A 2, 5, 5-T Manufacturing Site," 232 *Sco.* 497 (1986). For an overview of the issue of the bioavailability of soil and sediment contaminants, see Nat'l Research Council, *Bioavailability of Contaminants in Soils and Sediments: Processes, Tools, and Applications* (2003), http://www.nap.edu/read/10523/.

ii. Distribution

This subsection basically explains itself. In later discussions of cases presented below, one may wish to return to this subsection and ask what the substance's effects say about its ability to cross various barriers. This point is obvious in the drunk-driving case presented below.

iii. Metabolism

For an overview of metabolism in toxicology, see R.A. Kemper et al., "Metabolism: A Determinant of Toxicity," in A. Wallace Hayes (ed.), *Principles and Methods of Toxicology* 103–78 (5th ed. 2008).

The risk posed by metabolites is particularly important with respect to therapeutic drugs. For two papers that discuss this issue and suggest ways in which we may gain a better understanding of these processes and hopefully a better ability to predict and guard against the adverse effects of metabolites during the drug development process, see Thomas Bailee, "Metabolism and Toxicity of Drugs: Two Decades of Progress in Industrial Drug Metabolism," 21 *Chemical Res. Toxicology* 129 (2008) and B. Kevin Park et al., "Drug Metabolism and Drug Toxicity," 9 *InflammoPharmacology* 183 (2001).

A further discussion of metabolism and how genes may affect metabolic processes appears in the Teaching Materials on genes as modifiers of toxic susceptibility. See Slides 134 and 135.

The discussion of benzene metabolism comes from Robert Snyder, "Xenobiotic Metabolism and the Mechanism(s) of Benzene Toxicity," 36 *Drug Metabolism Revs.* 531, 547 (2004). Benzene metabolism is explored in substantial depth in the "expert report," found in Section IV.F.2. of the module, and the discussion of this report in the accompanying Teaching Materials below. One may wish to use some of that discussion here to emphasize that metabolism

may vary substantially not only across species, but also within species depending on genetic makeup.

One study examining the metabolism of thalidomide in animal species is Changa Tantibanchachai & Joanna Yang, "Studies of Thalidomide's Effects on Rodent Embryos from 1962–2008*," in Embryo Project Encyclopedia* (2014).

For a discussion of these effects in the elderly, see S. Shi & U. Klotz, "Age-Related Changes in Pharmacokinetics," 12 *Current Drug Metabolism* 601 (2011). They note, for example that because aging is associated with some reduction in first-pass metabolism, bioavailability of some drugs may be increased. Moreover, aging is associated with body fat increases and decreases in total body water. Therefore, water-soluble (hydrophilic) drugs have a smaller apparent volume of distribution, while lipid-soluble (lipophilic) drugs have an increased volume, with lengthened half-life. Because of age-related illnesses, there is larger inter-individual variability in the pharmacokinetics of drugs and other xenobiotics.

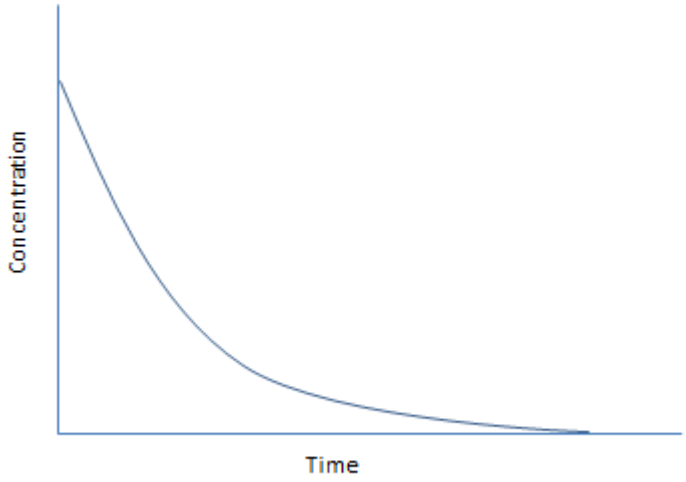
iii. Elimination

Again, this subsection is basically self-explanatory. Later in the section, when we present a case on how one may calculate blood alcohol concentrations (BAC), we do return to the issue of why alcohol is eliminated at a steady state from the body. As this discussion indicates, elimination is particularly important when our focus is on how a substance impairs function.

The first two figures (Slides 72 and 73) illustrate what are called "zero-order" and "first-order" elimination rates. Figure 3 (Slide 74) indicates the relationship between absorption and elimination. This figure is taken from Stine & Brown, supra, at 25.

Slide 72 Elimination Rate Affected by Concentration

Figure IV-1.

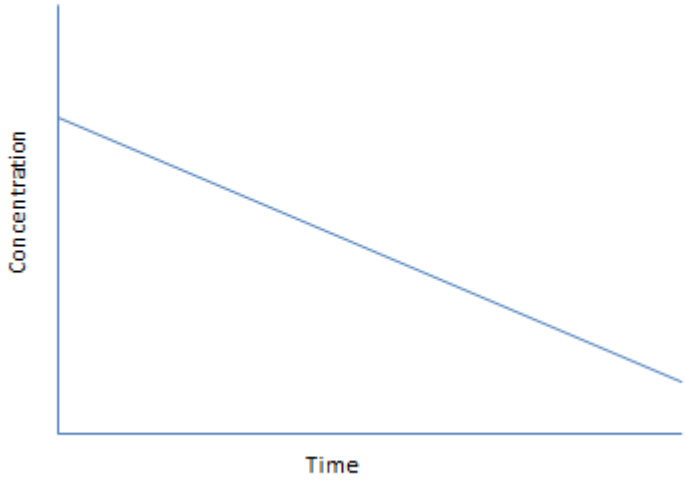


Elimination Rate affected by Concentration

72

Slide 73 Elimination Is a Constant Over Time

Figure IV-2.

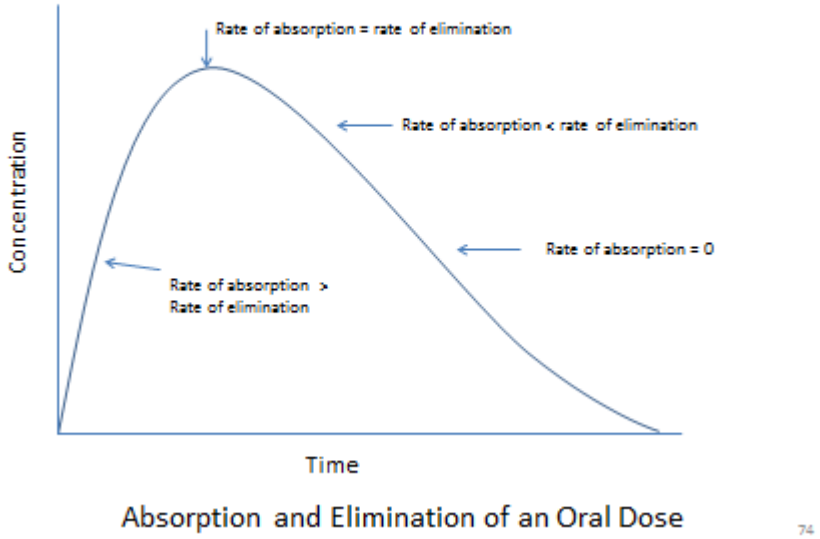


Elimination is a Constant Over Time

73

Slide 74 Absorption and Elimination of an Oral Dose

Figure IV-3.



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.

Study Questions:

1. What are the four basic processes involved in toxicokinetics?

   Absorption, Distribution, Metabolism, Excretion.

With respect to each of these processes, here are some additional questions.

2. What is the difference between exposure and dose?

   The answer addresses the general question of bioavailability. The dose is the actual amount of a substance entering the organism. With the same level of exposure, the dose received may vary due to many factors. One might wish to refer to examples such as HazMat suits designed to dramatically reduce dosage when exposed to a high concentration of a toxic substance.

3. How does this relate to the concept of bioavailability?

Bioavailability measures the effective dose after the substance has been ingested into the body.

4. What factors affect the distribution of xenobiotics within the body?

The answer, of course, is the four things listed in this subsection under (i), (ii), (iii), and (iv). Note this paragraph addresses distribution through the bloodstream. As the second paragraph notes, this is not the only method of distribution.

5. What is the importance of the blood-brain barrier?

It keeps some Xenobiotics from reaching most of the central nervous system.

6. What are two major first stage metabolic processes?

They are Hydrolysis and Oxidation.

7. What is one reason Xenobiotics produce harmful metabolites in one species and not another?

One reason is due to different specific proteins in each species.

8. What is the difference between first-order processes and zero-order processes of elimination?

In first-order processes, the rate of elimination is related to the concentration of the substance: higher concentrations lead to more rapid elimination and the rate of elimination slows as concentration lessens. In zero-order processes, toxicants are eliminated at a steady rate, usually due to some limiting factor on the rate of elimination. See the relevant figures.

9. Name a factor that affects the speed of elimination of a substance from the body.

Lipophilicity. Lipophilic substances are less quickly eliminated.

b. Toxicodynamics

For an overview and general introduction to mechanistic toxicology, see Urs A. Boelsterli, *Mechanistic Toxicology: The Molecular Basis of How Chemicals Disrupt Biological Targets* (2003). This section draws heavily from his book. Our discussion of this topic is quite basic, in part because a more complete discussion

of this topic requires a fair degree of understanding of biological mechanisms that may be beyond many who are using the module. However, the above book and many others provide the interested instructor with additional useful examples.

The hemoglobin example comes from Stine & Brown, supra, at 185.

The gasoline example is taken from Boelsterli, supra at 4.

The sources for the thalidomide discussion include: Boelsterli, supra at 8–12; Stine & Brown, supra, at 143; James Kim & Anthony Scialli, "Thalidomide: The Tragedy of Birth Defects and the Effective Treatment of Disease," 122 *Toxicological Sci.* 1 (2011); Herbert Schumacher et al., "A Comparison of the Teratogenic Activity of Thalidomide in Rabbits and Rats," 160 *J. Pharmacology & Experimental Therapeutics* 189 (1967); and Josef Warkany, "Why I Doubted that Thalidomide Was the Cause of the Epidemic of Limb Defects of 1959 to 1961," 38 *Teratology* 217 (1988).

Study Question:

1. How does gasoline produce kidney tumors in rats, but not in mice or humans?

The answer is that gasoline alters a plasma protein unique to rats. The protein slowly degrades and accumulated in kidney tubular cells, causing them lethal damage. The repeated cycle of tubular cell damage and regeneration leads to greater cell proliferation which is a mechanism of carcinogenesis. Note how putting this long answer in one place demonstrates the complex set of processes that lead to adverse effects.

Section IV.A.6. Types of Research

a. Predictive Toxicology

This section is designed primarily as an overview of the types of toxicological investigation. Our focus is on in vitro and in vivo testing. One thing we have *not* done is engage in the ongoing debate concerning the relative virtues of in vitro and in vivo approaches and the desire of many to reduce the use of animals in toxicological research.

Some instructors may wish to devote more time to predictive research. One resource for those so inclined is the U.S. Environmental Protection Agency's National Center for Computational Toxicology (http://www.epa.gov/comptox/index.html (last updated Apr. 1, 2016)), which is developing new software and methods for predictive toxicology. For a useful review of computational methods in drug development, see

Gregory Sliwoski et al., "Computational Methods in Drug Discovery," 66 *Pharmacological Revs.* 334 (2014).

The data on the cost of new drug development come from Christopher Paul Adams & Van Vu Brantner, "Spending on New Drug Development," 19 *Health Economics* 130 (2010); Joseph A. DiMasi et al., "Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs," 47 *J. Health Econ.* 20 (2016).

The benzene example comes from Goldstein & Henifin, supra, at 663 fn.82.

Study Question:

1. Why do chemicals of similar structure sometimes fail to produce similar effects on organisms?

   One answer is different metabolic pathways. But, of course, this is not the only reason.

b. In Vitro Research

One general resource is Anna Bal-Price & Paul Jennings et al., *In Vitro Toxicology Systems* (2014). See also *Richard Coico, Immunology: A Short Course* (7th ed. 2015). The dermatology example comes from D. Jírová et al., "Comparison of Human Skin Irritation Patch Test Data with In Vitro Skin Irritation Assays and Animal Data," 62 *Contact Dermatitis* 109 (2010).

Study Question:

1. What are the advantages and disadvantages of in vitro testing?

   Advantages: Cost, simplicity, and use of human cells when human research is possible. Disadvantages: External validity.

c. In Vivo Research

The toxicology materials began, of course, with an example of in vivo research and the instructor may wish to revisit the materials when covering this section. As we noted there, some instructors may wish to discuss the excerpt at the end of this section. Doing so permits the instructor to provide a concrete example to several of the issues discussed in the section.

Faoro v. Secretary of Health & Human Services, No. 10-704V, 2016 WL 675491 (Fed. Cl. Jan. 29, 2016), a case arising under the Childhood Vaccine Act, provides a useful example of the use of knockout animals. It was not disputed in *Faoro* that the

petitioner's child was born with a mutation of her SCN1A gene and that she had a seizure disorder known as Dravat syndrome. The full facts of the case are better discussed in the genomics material, but the gist of the debate is the claimant's position that the genetic mutation created susceptibility in an individual and could result in a seizure disorder depending on the occurrence of a "second hit" in the gene, resulting from environment interaction. The DTaP vaccine, was the alleged second hit because it causes fever or inflammation and there is some evidence that fevers may trigger seizures in children with Dravat syndrome. However, the government argued that while, at most, the vaccine might have triggered the early onset of Dravat syndrome in children with the SCN1A mutation, there was still no evidence that a vaccination before or after the onset of the syndrome affected the clinical course of the disease.

In support of this position, the government expert cited a study—John C. Oakley et al., "Temperature- and Age-Dependent Seizures in a Mouse Model of Severe Myoclonic Epilepsy in Infancy," 106 *Proc. Nat'l Acad. Sci.* U.S. Am. 3994 (2009)—involving genetically altered mice in which one copy of the SCN1A gene was deleted. At birth, the mice did not have seizures when exposed to heat, but as they aged, heat could produce seizures in one group of mice. Another group was not exposed to heat, but subsequently began to seize on their own, and both groups exhibited features typically seen in humans with Dravat syndrome. The disease developed and progressed with or without any trigger. A second study, not reported in the opinion, reproduced this effect. Christine S. Cheah et al., "Specific Deletion of NaV1.1 Sodium Channels in Inhibitory Interneurons Causes Seizures and Premature Death in a Mouse Model of Dravet Syndrome," 109 *Proc. Nat'l Acad. Sci. U.S. Am.* 14,646 (2012).

One may wish to ask students how persuasive this study is in supporting the government's position that the vaccines did not cause the claimant's injury. One should emphasize that the study's persuasiveness turns on the degree to which the mice and animals share a gene and the degree to which the genetic mutation produces similar injury. In fact, the relevance of the two studies cited here is supported by the additional fact that in both species, the mutation in question interferes with sodium channels in the same way.

For an overview of the use of genetically altered mice, see the web page for the International Mouse Phenotype Consortium, http://www.mousephenotype.org/ (last visited Apr. 20, 2016), whose stated goal is, "to discover functional insight for every gene by generating and systematically phenotyping 20,000 knockout mouse strains."

Most of this section focuses on dose issues. Before turning to this topic, however, the instructor may wish to review the choice of animal models and review why randomized experiments are superior to other methods with respect to internal validity. It has been our experience that students are well served by revisiting this notion at several points in a course such as this.

The section emphasizes the distinction between acute and chronic exposures. This point is important in the *Johnson* case presented below. Importantly, evidence on

the effects of acute exposure may or may not be particularly relevant when the issue at stake is one of chronic exposure.

The discussion of maximum tolerated dose (MTD) is designed to stress upon the students that there is no easy solution to the high doses given to animals in most in vivo research and thus the extrapolation issues discussed below are an inevitable result of the practicalities of animal testing. A useful resource on dose selection is Food & Drug Admin., *Guidance for Industry: S1C(R2) Dose Selection for Carcinogenicity Studies* (2008), http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm074919.pdf. The example on sample size, incidence frequency, and statistical significance is taken, albeit slightly changed, from Goldstein & Henifin, supra. To aid in the discussion of dose, we reproduce Table IV-1 from the module as Slide 75.

Slide 75 Acute Oral LD_{50} for Various Chemicals

Table IV-1

Acute Oral LD_{50} 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	

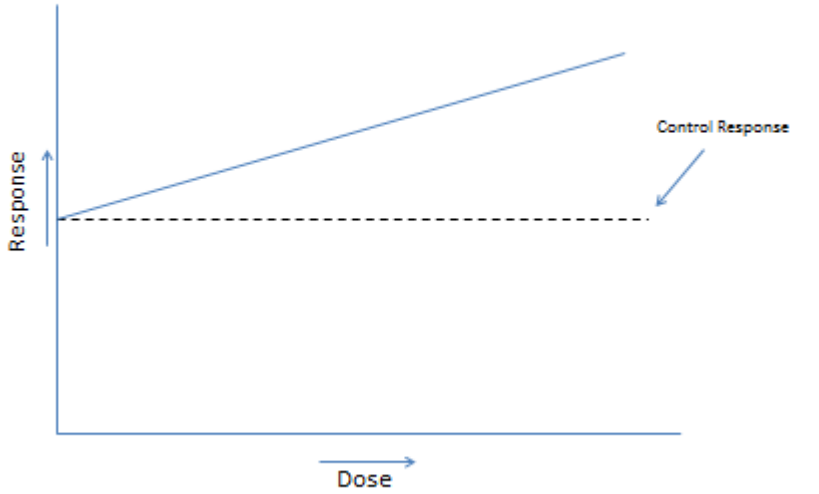
SOURCE: Republished with permission of National Science Teachers Association, from Nancy M. Trautmann, *Assessing Toxic Risk*, Student's Edition, 2001; permission conveyed through Copyright Clearance Center, Inc.

The no-threshold assumption (compare Figure IV-4 and IV-5, Slides 76 and 77) with respect to carcinogens is discussed in Martin W. Gehlhaus III et al., "Approaches to Cancer Assessment in EPA's Integrated Risk Information System," 254 *Toxicology & Applied Pharmacology* 170 (2011). The correctness of the LNT model is most vigorously

debated with respect to the effects of ionizing radiation. One paper opposing this view is Maurice Tubiana et al., "The Linear No-Threshold Relationship Is Inconsistent with Radiation Biologic and Experimental Data," 251 *Radiology* 13 (2009). But see Richard William Harbron, "Cancer Risks from Low Dose Exposure to Ionising Radiation—Is the Linear No-threshold Model Still Relevant," 18 *Radiography* 28 (2012). The *Chlorine Chemical* case presented below deals with this issue in the context of the particular ways in which substances may cause cancer.

Slide 76 Linear No-threshold (LNT) Model

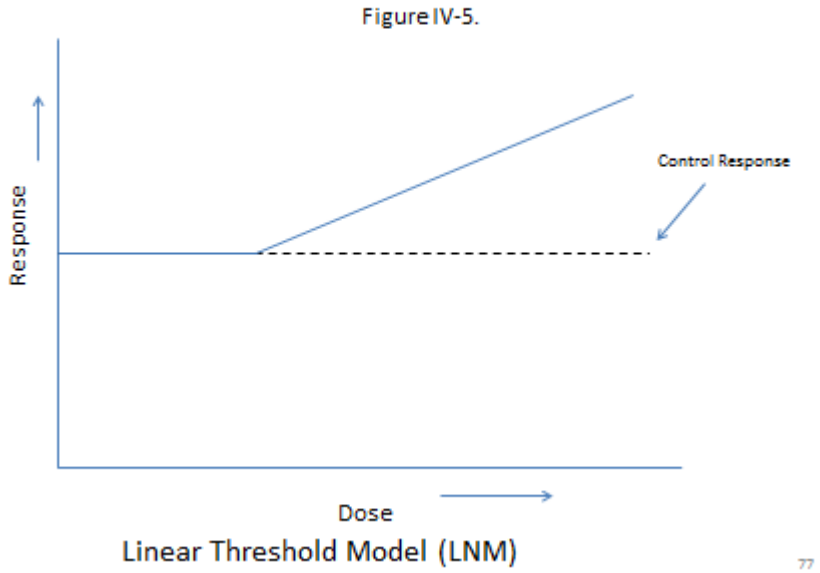
Figure IV-4.



Linear No-threshold (LNT) Model

76

Slide 77 Linear Threshold Model (LNM)



The no-threshold assumption has caused a considerable amount of controversy in lawsuits where a defendant has exposed an individual to minimal amounts of some carcinogen. We discuss some of these cases below in these materials (but not in the module) as part of the general discussion on the admissibility of expert evidence in toxic-tort cases.

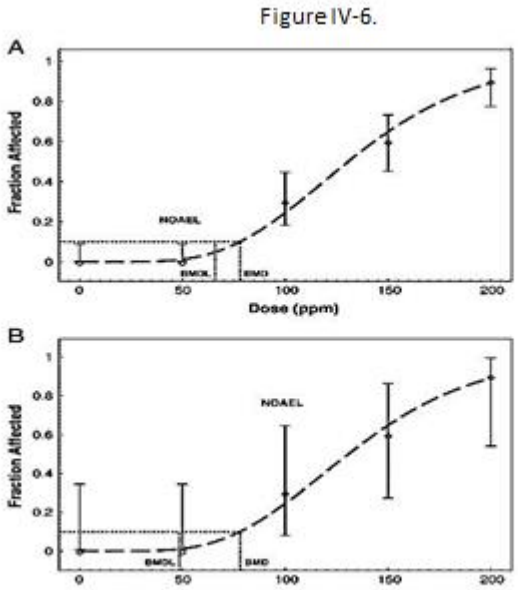
Toxic substances that are not carcinogenic are assumed to have a threshold, and thus effort must be made calculate what that threshold might be. This basically means finding a dose at which there is no statistically significant difference between exposed and control groups. But the instructor should emphasize that the absence of a statistically significant difference does not mean that there is no effect. Depending on the sample size of a study, there may be an effect that is simply undetected—a Type II error—a topic reviewed in the epidemiology materials.

The discussion of BMD models is taken from J. Allen Davis et al., "Introduction to Benchmark Dose Methods and U.S. EPA's Benchmark Dose Software (BMDS) Version 2.1.1," 245 *Toxicology & Applied Pharmacology* 181 (2011). Figure 6 and Table IV-2 (Slides 78 and 79), taken from Davis et al., demonstrates these advantages. Following is their explanation of the Figure and the Table:

[Figure 4] "illustrates the effect sample size has on NOAEL and BMD and BMDL estimation. The incidence of effect at each dose group is represented by diamonds. The error bars represent the 95% confidence limits for the incidence. The fitted model is represented by the dashed line and the BMD and BMDL estimates are represented by the dotted line. Graph A represents a dataset with 50 animals per dose group; the NOAEL has been determined to be 50 ppm as it is the highest dose to fail to reach statistical significance (see [Table IV-2]); the BMD (corresponding to a BMR of 10% extra risk) is determined to be 78 ppm and the BMDL is 66 ppm (see [Table IV-2]). Graph B represents a dataset with only 10 animals per dose group, but with the same percentage of animals responding at each dose as Graph A. In this case, due to the decreased statistical ability to detect a treatment effect, the NOAEL has been determined to be 100 ppm. The BMDL in this situation has decreased to 48 ppm because the 95% lower-bound confidence limit is larger, reflecting the increased variability and uncertainty in the data due to the smaller sample size. Therefore, as sample size decreases, resulting in decreased power to detect treatment effects, the NOAEL method returns higher POD estimates whereas the BMD approach returns lower, more health protective, PODs. Note the increase in the size of the confidence limits around the observed incidence between Graphs A and B."[13]

[13] Reprinted from *Toxicology and Applied Pharmacology* 254(2) , J. Allen Davis, Jeffrey S. Gifta, 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.

Slide 78 BMD versus NOAEL Estimations



78

SOURCE: Reprinted from *Toxicology and Applied Pharmacology* 254(2), J. Allen Davis, Jeffrey S. Gifta, 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.

Slide 79 BMD versus NOAEL Calculations

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
	0	0	1.00			
	50	0	1.00			
50	100	15	< 0.001	50	78.05	65.85
	150	30	< 0.001			
	200	45	< 0.001			
	0	0	1.00			
	50	0	1.00			
10	100	3	0.105	100	78.05	48.40
	150	6	0.005			
	200	9	< 0.001			

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

79

SOURCE: Reprinted from *Toxicology and Applied Pharmacology* 254(2), J. Allen Davis, Jeffrey S. Gifta, 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.

Study Questions:

1. What are the advantages and disadvantages of in vivo testing?

   Advantages: true experiments and being able to observe toxicokinetics. Disadvantages: high dose, and therefore external validity. Although this module generally sidesteps ethical questions with respect to animal testing, the instructor may wish to discuss this issue in the context of in vitro versus in vivo research.

2. What are the advantages of the BMD approach to establishing a threshold dose?

   A better dose-response curve and a more conservative estimate of threshold dose when data are of lower quality.

3. When is a no-threshold model used?

Generally, a no-threshold model is used when the outcome under investigation is cancer. But, as the *Chlorine Chemical* case presented in the module indicates, this is not always appropriate.

d. Clinical Trials

A discussion of clinical trials could be placed solely in the epidemiology section as they do involve human subjects. However, it seems worthwhile to reprise the discussion here as well to highlight the fact that when it comes to drugs, research on animals is often a poor predictor of a drug's effects on humans. The key observation is that nearly two-thirds of the drugs that enter Phase II trials do not move on to Phase III. The statement that 12% of drugs entering clinical trials are approved comes from DiMasi et al., supra.

The rehash of why researchers employ randomization and double blinding may not be necessary for most students, but we have found that the concepts are not always easily grasped. By the way, there is considerable controversy concerning the size of the placebo effect in clinical trials. See J. Hunsley et al., "Interpreting the Magnitude of the Placebo Effect: Mountain or Molehill?" 63 *J. Clinical Psychol.* 391 (2007); Bruce E. Wampold et al., "Placebo Is Powerful: Estimating Placebo Effects in Medicine and Psychotherapy from Randomized Clinical Trials," 61 *J. Clinical Psychol.* 835 (2005).

The formal role of Phase IV trials has been greatly expanded with the passage of the Food and Drug Administration Amendments Act (FDAAA) of 2007, Pub. L. No. 110-85, 121 Stat. 823 (2007). The statute may be found at http://www.fda.gov/RegulatoryInformation/Legislation/SignificantAmendmentstotheFDCAct/FoodandDrugAdministrationAmendmentsActof2007/FullTextofFDAAALaw/default.htm. For a useful discussion of the statute, see Barbara J. Evans, "Seven Pillars of a New Evidentiary Paradigm: The Food, Drug, and Cosmetic Act Enters the Genomic Era," 85 *Notre Dame L. Rev.* 419, 498 (2010).

Study Questions:

1. What are the objectives of each clinical trial phase?

Phase I: to observe how a drug interacts with the human body and establish initial dosing schemes. Phase II: to assess whether a drug is beneficial and provide additional safety data. Phase III: to further assess efficacy, establish dosage, and detect less common side effects that went undetected in earlier phases. Phase IV: to compare a drug with other drugs already in the market, monitor a drug's long-term effectiveness and impact on a patient's quality of life, and determine the cost-effectiveness of a drug therapy relative to other traditional and new therapies.

2. Why should we use randomized, blinded, clinical trials with placebos?

Note that a full answer addresses each of the three aspects: randomization, blinding, and the use of placebos.

Randomization attempts to control for unknown confounders by distributing them equally across the values of the independent variables. For example, if we did not know that being a smoker affected the efficacy of a drug, randomization would ideally place an equal percentage of smokers and nonsmokers in the treatment and control groups, thus neutralizing the independent effect of smoking. The value of randomization is that there may be numerous unknown confounders and ideally, randomization can simultaneously neutralize their effect on the observed correlation between the drug and the outcome under investigation. The ability of randomization to achieve this result is contingent, in part, on the study sample size.

The use of a placebo is designed to guard against the "placebo effect," that is, the effect that may occur when people believe that they have been given a beneficial treatment.

Double blinding guards against what is sometimes referred to as "construct validity." While one may think that one is observing an effect between A and B, one may be observing an effect between X and B. The placebo effect is an example of the problem of construct validity and placebo controls help us distinguish between the effect of a drug (A) and a placebo effect (X) on a clinical outcome. The other blinding—blinding the treating health-care provider as to whether the patient has received the drug or a placebo—guards against the possibility that we are observing an effect based on the provider's expectations, rather than the effect of a drug.

Section IV.A.7. Extrapolation

This section focuses on a topic that has been referred to frequently: how to extrapolate data from one situation in order to draw a conclusion about a different situation. A valuable overview of the topic may be found in J.V. Rodricks et al., "Quantitative Extrapolation in Toxicology," in A. Wallace Hayes (ed.), *Principles and Methods of Toxicology* 365 (5th ed. 2008).

The central idea holding this section together is *external validity*. This concept remains important throughout the remainder of the section. Many legal opinions, including some discussed below, rely heavily on an external validity analysis, whether or not they expressly use the term.

a. Dose Extrapolations

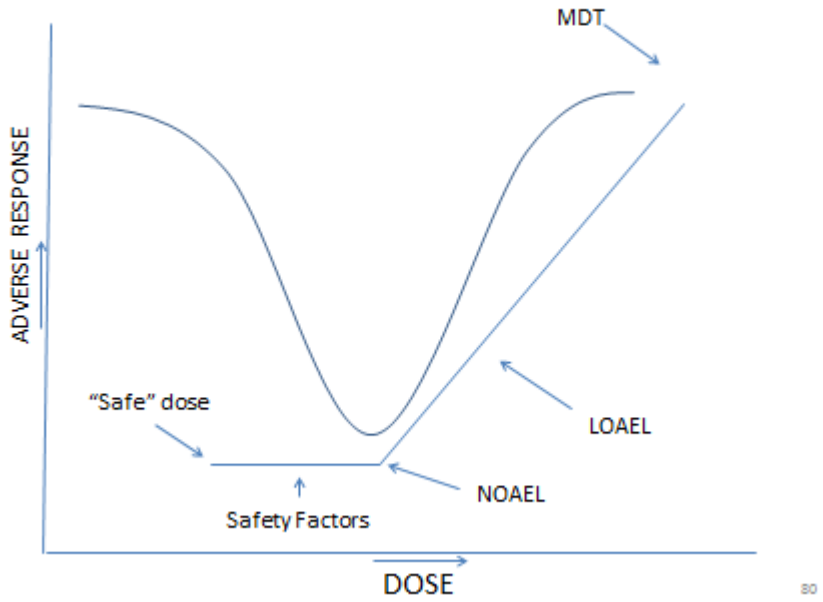
The calculation of RfDs and RfCs is often a matter of controversy in litigation concerning regulatory standards. The *Chlorine Chemical* case presented below involves the EPAs RfC for chloroform. For the EPA's discussion of how reference doses and reference concentrations are established, see U.S. Envt'l Protection Agency, *Review of the Reference Dose and Reference Concentration Processes Document* (2002),

http://www2.epa.gov/risk/review-reference-dose-and-reference-concentration-processes-document.

The nonmonotonic models illustrated in Figures IV-7 and IV-8 (Slides 80 and 81) should be kept distinct. Those who argue for a nonmonotonic curve for endocrine-disrupting chemicals see this as an exception to the general rule. Proponents of hormesis, however, claim to see this in many situations. The discussion of endocrine-disrupting chemicals and Figure IV-7 are taken from Laura N. Vandenberg et al., "Hormones and Endocrine-Disrupting Chemicals: Low-Dose Effects and Nonmonotonic Dose Responses," 33 *Endocrine Revs.* 378 (2012). The primary proponent of this position is Edward Calabrese. See Mark P. Mattson & Edward J. Calabrese, *Hormesis: A Revolution in Biology, Toxicology and Medicine* (2010). One should be careful not to confuse the hormesis hypothesis with the fact that substances may be beneficial with respect to some organs but harmful to others.

Slide 80 Non-monotonic Dose-Response Curve With Greater Low-Dose Risk

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.

Slide 81 Hormesis Effect

Figure IV-8

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

The summary paragraph in this subsection relies on the discussion in Davis et al., supra.

b. Extrapolation Across Species

Whether the history of saccharin counts as regulatory successes or failures depends on one's point of view. The cyclamate experience, however, once again underlines the downside of assuming that any smoke with respect to animal data suggests a fire with respect to human adverse effects. The saccharin discussion comes from Joanne Zurlo and Robert A. Squire, "Is Saccharin Safe? Animal Testing Revisited," 90(1) *J. Nat'l Cancer Inst.* 2 (1998). The National Toxicology Program Report on Carcinogens' delisting of saccharin can be found at Sylvia M. Burwell, U.S. Dep't of Health & Human Servs. Sec'y, *13th Report on Carcinogens* (2014), http://ntp.niehs.nih.gov/pubhealth/roc/roc13/index.html.

Study Questions:

1. What extrapolations are required when using animal data to produce human effects?

   One must extrapolate both species size differences and dosage.

2. What factors are used to establish RfDs and RfCs?

   In addition to dosage and the establishment of a NOAEL or BMR are factors based on the uncertainty involved in the extrapolation process, such as dose and species differences.

c. Susceptibility

One may introduce this discussion by asking students when and where they have heard people say, "Why me?" The examples in this section provide a partial answer to the question, but it is worth noting that for many situations, there is still no toxicological or specific genetic answer. The cigarette example comes from Barnoya et al., supra, at ch. 25:29. The ways in which groups metabolize alcohol is discussed in Nat'l Inst. on Alcohol Abuse & Alcoholism, *The Genetics of Alcohol Metabolism: Role of Alcohol Dehydrogenase and Aldehyde Dehydrogenase Variants*, http://pubs.niaaa.nih.gov/publications/arh301/3-4.htm (last visited Apr. 20, 2016).

A useful reference to the problem of cumulative exposures is Comm. on Improving Risk Analysis Approaches Used by the U.S. EPA et al., *Science and Decisions: Advancing Risk Assessment*, at ch. 7 (2009).

Study Question:

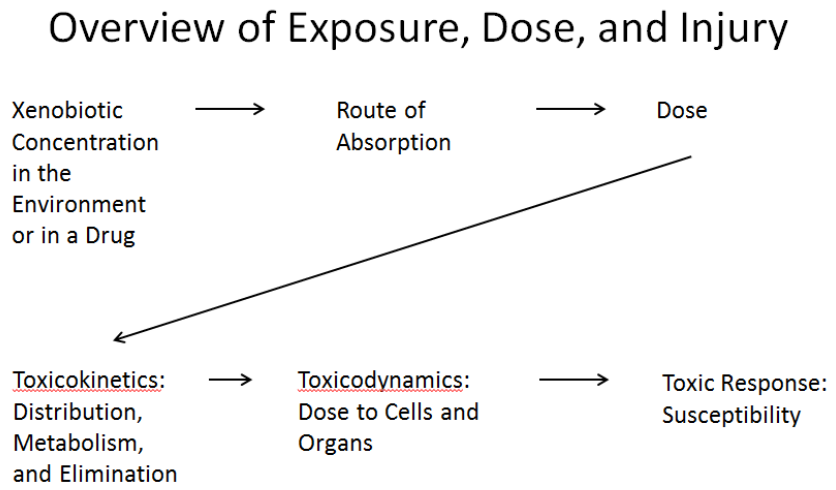
1. What are some of the causes of differences in susceptibility?

   To name some: genetic, environmental, and toxicokinetic effects, such as the rate of absorption and elimination.

The section concludes with a figure (Figure IV-9, Slide 82) depicting an overview of exposure, dose, and injury.

Slide 82 Overview of Exposure, Dose, and Injury

Figure IV-9



d. The Quality of In Vivo Studies

The discussion of external validity and internal validity relies heavily on Thomas D. Cook & Donald T. Campbell, *Quasi-Experimentation: Design and Analysis Issues for Field Testing* 37 (1979).

Figure IV-9 offers a graphical summary of much of the preceding discussion.

The failure of animal studies to execute a blinded-randomized design is an important cautionary tale. The main thing students should take from this discussion is that one must always look behind any study's findings and examine its methods.

We took the data on the frequency of studies using inadequate designs from V. Bebarta et al., "Emergency Medicine Animal Research: Does Use of Randomization and Blinding Affect the Results?" 10 *Acad. Emergency Med.* 684 (2003). See also Nicolas A. Crossley et al., "Empirical Evidence of Bias in the Design of Experimental Stroke Studies: A Metaepidemiologic Approach," 39 *Stroke* 929 (2008); John P. A. Ioannidis, "Extrapolating from Animals to Humans," 4 *Sci. & Translational Med.* 151 (2012).

For another useful article, this time focusing on publication bias and the resultant excess of significant result findings in the published literature, see Konstantinos K. Tsilidis et al., "Evaluation of Excess Significance Bias in Animal Studies of

Neurological Diseases," 11 *PloS Biology* e1001609 (2013). Following is the abstract of this study:

> Animal studies generate valuable hypotheses that lead to the conduct of preventive or therapeutic clinical trials. We assessed whether there is evidence for excess statistical significance in results of animal studies on neurological disorders, suggesting biases. We used data from meta-analyses of interventions deposited in Collaborative Approach to Meta-Analysis and Review of Animal Data in Experimental Studies (CAMARADES). The number of observed studies with statistically significant results (O) was compared with the expected number (E), based on the statistical power of each study under different assumptions for the plausible effect size. We assessed 4,445 datasets synthesized in 160 meta-analyses on Alzheimer disease (n = 2), experimental autoimmune encephalomyelitis (n = 34), focal ischemia (n = 16), intracerebral hemorrhage (n = 61), Parkinson disease (n = 45), and spinal cord injury (n = 2). 112 meta-analyses (70%) found nominally ($p \leq 0.05$) statistically significant summary fixed effects. Assuming the effect size in the most precise study to be a plausible effect, 919 out of 4,445 nominally significant results were expected versus 1,719 observed ($p < 10$). Excess significance was present across all neurological disorders, in all subgroups defined by methodological characteristics, and also according to alternative plausible effects. Asymmetry tests also showed evidence of small-study effects in 74 (46%) meta-analyses. Significantly effective interventions with more than 500 animals, and no hints of bias were seen in eight (5%) meta-analyses. Overall, there are too many animal studies with statistically significant results in the literature of neurological disorders. This observation suggests strong biases, with selective analysis and outcome reporting biases being plausible explanations, and provides novel evidence on how these biases might influence the whole research domain of neurological animal literature.[14]

Study Question:

1. According to empirical work, what is the consequence of relying on in vivo drug studies that do not use appropriate designs?

   They are more likely to find positive effects of potential new drugs on animals. The instructor may wish to relate this again to the concept of construct validity.

Section IV.B. Toxicology and the Law

This section focuses on the role of toxicology in cases addressing how substances affect humans. Toxicology information does play a role in other areas as well. See *Eli Lilly & Co. v.*

[14] Text Copyright © 2013 Tsilidis et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License.

Zenith Goldline Pharmaceuticals, Inc., 364 F. Supp. 2d 820 (S.D. Ind. 2005) (patent infringement litigation).

Section IV.B.1. The Use of Toxicology in Regulations

We provided the citation to each of the regulatory statutes mentioned in the module because some instructors may wish to delve deeper into a particular statute other than the Clean Drinking Water statute, which is the subject of the case below.

For a discussion of the European REACH program, see John S. Applegate, "Synthesizing TSCA and REACH: Practical Principles for Chemical Regulation Reform," 35 *Ecology* L.Q. 721 (2008) for a comparison of the two approaches. The jury still seems to be out on how effective REACH will be in removing dangerous chemicals from use. Much turns, of course, on the degree to which regulators favor avoiding Type II errors over Type I errors.

The House of Representatives proposed legislation updating TSCA, which may be found at The TSCA Modernization Act, H.R. 2576, 114th Cong. (2015), https://www.gpo.gov/fdsys/pkg/BILLS-114hr2576eas/pdf/BILLS-114hr2576eas.pdf. The bill has passed both the House and the Senate, which the Senate has sent back to the House with changes. H.R.2576—TSCA Modernization Act of 2015, Congress.gov, https://www.congress.gov/bill/114th-congress/house-bill/2576 (last visited Apr. 22, 2016).

The data on untested chemicals is from Goldstein & Henifin, supra, at 648.

The CERCLA RfC standard may be found at Office of Emergency &Remedial Response, U.S. Envt'l Protection Agency, *Risk Assessment Guidance for Superfund: Volume I—Human Health Evaluation Manual (Part B, Development of Risk-Based Preliminary Remediation Goals)* ch. 2.6. (1991), http://www2.epa.gov/risk/risk-assessment-guidance-superfund-rags-part-b.

Chlorine Chemistry Council case:

Perhaps one should begin the discussion of this case by making it clear that in the vast majority of situations, the courts affirm agency decisions. Although it is beyond the scope of this module, a brief discussion of "Chevron deference" might be worthwhile. For nonlawyer instructors, the term Chevron deference comes from a United States Supreme Court case, *Chevron, U.S.A., Inc. v. Natural Resources Defense Council, Inc.*, 467 U.S. 837 (1984), providing for judicial deference to agency decision making. The degree of deference actually given to agency rulings is another matter altogether, as the *Chemistry Council* case indicates. For one article attempting to sort out the cases, see William N. Eskridge, Jr. & Lauren E. Baer, "The Continuum of Deference: Supreme Court Treatment of Agency Statutory Interpretations from *Chevron* to *Hamdan*," 96 *Geo. L.J.* 1083 (2008).

As the *Chlorine Chemistry Council* case suggests, various regulatory statutes may each have unique provisions that influence judicial review. *Industrial Union Department, AFL-CIO v. American Petroleum Institute*, 448 U.S. 607 (1980) is another case in which the Court concluded that the agency, in this case, the Occupational Safety and Health Administration (OSHA),

attempted to impose a benzene exposure standard of 1 ppm. Many of the arguments advanced in the Chlorine Chemistry case were presented there as well.

The two key provisions with which the court considered were 29 U.S.C. Section 652(8) and 29 U.S.C. Section 655(b)(5). The former defines an occupational safety and health standard as follows:

> The term "occupational safety and health standard" means a standard which requires conditions, or the adoption or use of one or more practices, means, methods, operations, or processes, reasonably necessary or appropriate to provide safe or healthful employment and places of employment.

The latter states the following with respect to toxic materials:

> The Secretary, in promulgating standards dealing with toxic materials or harmful physical agents under this subsection, shall set the standard which most adequately assures, to the extent feasible, on the basis of the best available evidence, that no employee will suffer material impairment of health or functional capacity even if such employee has regular exposure to the hazard dealt with by such standard for the period of his working life. Development of standards under this subsection shall be based upon research, demonstrations, experiments, and such other information as may be appropriate. In addition to the attainment of the highest degree of health and safety protection for the employee, other considerations shall be the latest available scientific data in the field, the feasibility of the standards, and experience gained under this and other health and safety laws.

At the heart of the case was the government's position that even though there was no data concerning the effects of exposure at or below 10 ppm, "whenever a carcinogen is involved, OSHA will presume that no safe level of exposure exists in the absence of clear proof establishing such a level and will accordingly set the exposure limit at the lowest level feasible." 448 U.S. at 624. As the court noted, given this policy, it was irrelevant to OSHA whether there was any evidence of a leukemia risk at 10 ppm because there was no evidence that there was *not* some risk.

In a particularly instructive part of the opinion, Justice Stevens noted that the issue ultimately is one of burdens of proof:

> Given the conclusion that the Act empowers the Secretary to promulgate health and safety standards only where a significant risk of harm exists, the critical issue becomes how to define and allocate the burden of proving the significance of the risk in a case such as this, where scientific knowledge is imperfect and the precise quantification of risks is therefore impossible. The Agency's position is that there is substantial evidence in the record to support its conclusion that there is no absolutely safe level for a carcinogen and that, therefore, the burden is properly on industry to prove, apparently beyond a shadow of a doubt, that there is a safe level for benzene

> exposure. The Agency argues that, because of the uncertainties in this area, any other approach would render it helpless, forcing it to wait for the leukemia deaths that it believes are likely to occur before taking any regulatory action.
>
> We disagree. As we read the statute, the burden was on the Agency to show, on the basis of substantial evidence, that it is at least more likely than not that long-term exposure to 10 ppm of benzene presents a significant risk of material health impairment. Ordinarily, it is the proponent of a rule or order who has the burden of proof in administrative proceedings. See 5 U.S.C. § 556(d). In some cases involving toxic substances, Congress has shifted the burden of proving that a particular substance is safe onto the party opposing the proposed rule. The fact that Congress did not follow this course in enacting the Occupational Safety and Health Act indicates that it intended the Agency to bear the normal burden of establishing the need for a proposed standard.
>
> Id. at 652-53.

Ultimately, with new and more complete data, OSHA did successfully lower the standard to 1 ppm. It justified this rule in a Federal Register notice that reviewed some of the epidemiology and toxicology data. Occupational Exposure to Benzene, 52 Fed. Reg. 34,460, 34,662–63 (Sept. 11, 1987) (to be codified at 29 C.F.R. pt. 1910). That standard has remained in place. Note that, as in the *Chlorine Chemistry* case, the standard is not zero. Given the widespread presence of benzene in many industrial settings, such a standard would not be feasible.

Setting aside feasibility, is the reduction to 1 ppm justifiable on the basis of a cost-benefit analysis? Most statutes do not require the government to make such calculations; an exception being the Safe Drinking Water Act, but it is hard to imagine a complete risk assessment without explicitly assessing costs and benefits.

OSHA presented estimates of the cost to move from existing baselines to the new standard in seven key industries. In 1983 dollars, it estimated total annualized costs to be $24 million. 52 Fed. Reg. at 34,516. As to benefits,

> [b]ased on OSHA's preferred estimates of risk, the final standard would result in 85 fewer leukemia deaths per 1000 workers exposed at the current 10 ppm level for a working lifetime. Based on OSHA's current estimate of exposures, the final standard would save a minimum of 230 leukemia deaths over a working lifetime (45-year period). Using the API preferred estimate, the lowest estimate of risk reduction, the standard would result in 7.4 fewer leukemia deaths per 1000.
>
> 52 Fed. Reg. at 34,460–01.

Of course, these are not the only benefits. There would also be a reduction in deaths due to aplastic anemia—the EPA estimated the reduction would result in 92 fewer deaths from this cause over the same period—and other diseases associated with benzene. One may wish to

ask students if this makes the reduction a good deal. The answer to that question requires us to set a price on the value of a human life.

Focusing on inflation alone, $24 million in 1983 dollars would be something in excess of $57 million today. This value is based solely on changes in the Consumer Price Index. This is not the only way to calculate the current value of money. For a valuable discussion of various ways to calculate the present value of an amount from the past, see Samuel H. Williamson, *Seven Ways to Compute the Relative Value of a U.S. Dollar Amount—1774 to Present*, MeasuringWorth.com (2016), http://www.measuringworth.com/uscompare/. What value should we place on the lives saved, that is, the value of a statistical life? The EPA established a central estimate of $7.4 million in 2006 and recommended that this figure be updated to the year of the analysis. This figure would be close to $9 million in 2015 dollars or about $3.7 million in 1983 dollars, again using the Consumer Price Index—but one should keep in mind that in 1983, the EPA had not yet adopted its current method of calculating the value of a statistical life. See Nat'l Ctr. for Envtl. Econ., U.S. Envtl. Protection Agency, *Frequently Asked Questions on Mortality Risk Valuation*, http://yosemite.epa.gov/EE%5Cepa%5Ceed.nsf/webpages/MortalityRiskValuation.html (last updated May 11, 2016).

If we add the 230 leukemia lives and the 92 aplastic-anemia lives over 45 years, the total is 322 lives saved. The total 45-year cost (in 1983 dollars) would be a bit more than $1 billion. And if we were to divide this by the 332 lives, the cost of each life would be approximately $3.5 million each, less than the $3.7-million value of each statistical life. Of course, this whole exercise is misleading because the value of life and the cost of implementing the new standard will change over time. Nevertheless, one might note that this particular answer to the question "is it worth it" is "yes, by a little bit." An instructor who is interested may wish to spin out some other hypotheticals. For example, if one used the *American Petroleum Institute*'s calculation of lives saved, the cost of each life saved would be much higher, clearly higher than the EPA's value of a statistical life. One may wish to ask students whether it is even worthwhile talking about the issue in this way, and if not, whether there is some alternative that is superior. The Viscusi article cited in the notes following *Chlorine Chemistry* offers citations to the value of statistical-life literature. Professor Viscusi is perhaps the leading scholar in this area. Related articles include: W. Kip Viscusi & Joseph E. Aldy, "The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World," 27 *J. Risk and Uncertainty* 5 (2003); and W. Kip Viscusi, "Pricing Lives for Corporate Risk Decisions," 68 *Vand. L. Rev.* 1117 (2015).

Notes following *Chlorine Chemical*:

With respect to the *Chlorine Chemical* case itself, the instructor may wish to ask students why the court and the EPA were uncertain as to whether a LND model was correct. The answer to this question turns on how chloroform causes cancer. The best evidence, according to the court, is that the cancer is not caused directly through gene mutation, but rather through a process that begins because chloroform is toxic to cells—cytotoxicity. Cell

death leads to cell proliferation, which ultimately produces cancer through a process similar to that described above when we discussed gasoline-induced cancer in rats. The line of reasoning is that if the dose of chloroform is sufficiently low, then there will not be substantial cell proliferation and thus no tumor. Some of the evidence supporting this view can be found in Michael V. Templin et al., "Chloroform-Induced Cytotoxicity and Regenerative Cell Proliferation in the Kidneys and Liver of BDF_1 Mice," 108 *Cancer Letters* 225 (1996), available at http://www.sciencedirect.com/science/article/pii/0304383596042346.

See also Byron E. Butterworth et al., "The Role of Regenerative Cell Proliferation in Chloroform-Induced Cancer," 82–83 *Toxicology Letters* 23 (1995). At the end of the abstract to this article, the authors note:

> These observations are consistent with a substantial database that indicates that tumor induction by chloroform occurs via a non-genotoxic-cytotoxic mode of action and is secondary to organ-specific toxicity. These data further support the premise that doses that do not induce regenerative cell proliferation do not present an increased risk of cancer.

For a general review of the evidence on the toxicity of chloroform available at the turn of the century, see Toxicological Profile for Chloroform, published in 1997 by the Department of Health and Human Services, which can be found at http://www.atsdr.cdc.gov/toxprofiles/tp6.pdf. The British Health Protection Agency released a statement in 2007 concerning chloroform, "Chloroform: Toxicological Overview," that may be found at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/338535/Chloroform_Toxicological_Overview.pdf. There, it makes the following statement:

> Chlorinated drinking water typically contains chloroform, other trihalomethanes and a wide variety of other disinfection by-products. There has been considerable epidemiological research into the question of associations between chlorinated drinking-water and various diseases, and this research continues. The results have raised concern that chlorination by-products in drinking-water may increase the risk of certain cancers. However, no conclusions can be drawn specifically about chloroform from these studies. The expert advisory Committee on Carcinogenicity of Chemicals in Food, Consumer Products and the Environment (COC) considered that the evidence was inconclusive, but advised that efforts to minimize exposure to chlorination by-products remain appropriate, providing that they do not compromise the efficiency of disinfection of drinking-water (see the statement at http://www.advisorybodies.doh.gov.uk/coc/drink.htm). The committee will be reviewing these issues again shortly. The International Agency for Research on Cancer (IARC) has concluded that there is inadequate evidence in humans for the carcinogenicity of chloroform but sufficient evidence in experimental animals for the carcinogenicity of chloroform. It is classified as possibly carcinogenic to humans (Group 2B).

Perhaps the most noteworthy comment in this statement is the reality that whatever carcinogenic potential the ingestion of chloroform may have, it must be balanced against the efficiency of disinfection of drinking water. One may wish to ask students about this issue and how, exactly, the EPA attempted to address it.

If time permits, one could assign the *City of Waukesha* case and compare how the court handled these two claims. One can also pose the questions following the case to the students.

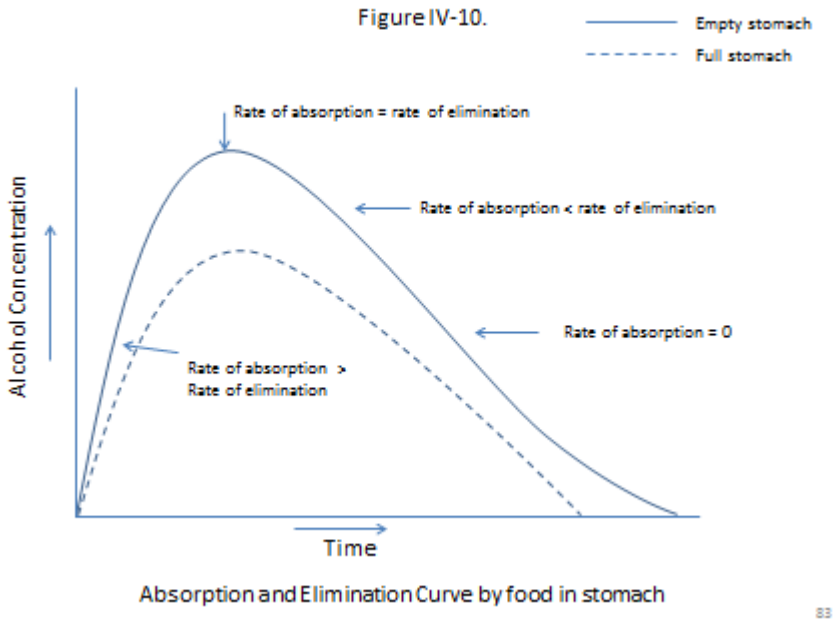
Section IV.B.2. Toxicology in Litigation

a. Toxicology in Criminal Cases

As we noted above, some instructors may wish to skip this case. While it offers an instructive example of how the law uses knowledge of toxicokinetics, it does not deal with the question of whether a particular xenobiotic caused the individual's condition. This issue, however, has become increasingly important in cases that involve some drug other than ethanol. Thus far, toxicology's ability to estimate with any precision the central-nervous-system effects of other recreational drugs, is quite limited.

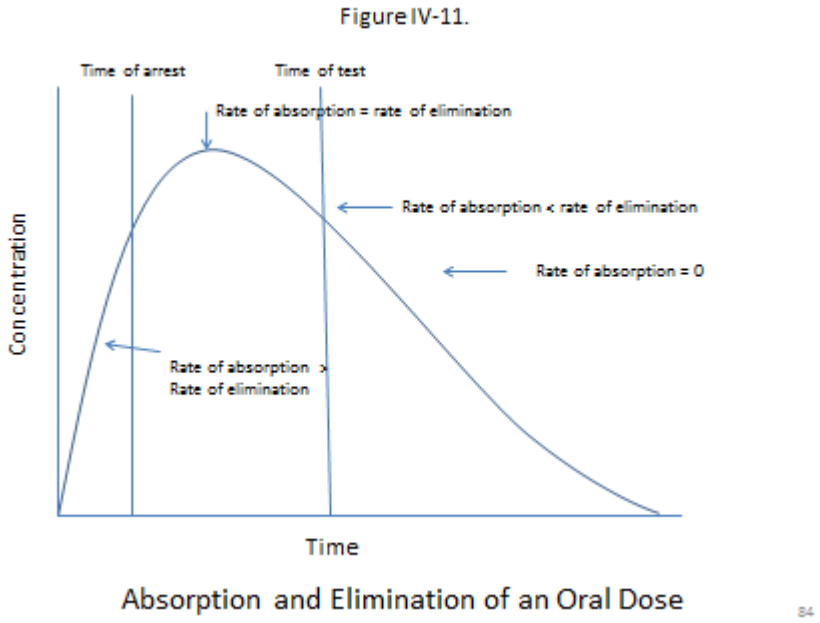
The discussion of alcohol testing is taken in large part from Valentine & Valentine, "Alcohol Testing," in Faigman et al. (eds.), *5 Modern Scientific Evidence: The Law and Science of Expert Testimony* ch. 39 (2015–2016). The materials here and the *Tsosie* case provide the instructor with an opportunity to review pharmacokinetics in general. At the beginning of this discussion, the instructor may wish to refer students back to Figure IV-3. The module includes a Figure IV-10, which is a variation on Figure IV-3, indicating how a full stomach may affect both the rate of absorption and the maximum BAC. Following the case, the instructor may use Figures 10 and 11 (Slides 83 and 84) to explore the reasonableness of a retrograde extrapolation under various scenarios. For a more advanced discussion of alcohol pharmacokinetics, see Ake Norberg et al., "Role of Variability in Explaining Ethanol Pharmacokinetics: Research and Forensic Applications," 42 *Clinical Pharmacokinetics* 1 (2003).

Slide 83 Absorption and Elimination Curve by Foot in Stomach



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.

Slide 84 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.

The *Tsosie* Case:

The retrograde extrapolation issue allows the instructor to discuss a recurring issue in many areas of scientific evidence. Should courts ever establish a causation requirement that the parties very often will be unable to meet? The Montana Supreme Court addressed this issue in State v. McGowan, 139 P.3d 841 (Mont. 2006). The court held that the state need not present retrograde-extrapolation evidence to establish a person's alcohol concentration at the time he or she was driving under the state's DUI per se statute. It recognized,

> the impossible burden that requiring retrograde extrapolation evidence would place on the state. Retrograde extrapolation would require evidence that the state would rarely be able to acquire because of a defendant's constitutional right to remain silent. Specifically, the state would need to ascertain information wholly within the defendant's knowledge, such as when, and in what amounts, the defendant consumed alcohol before driving. Further, the rate of absorption of alcohol varies greatly among individuals, with studies indicating that a

> person reaches their peak blood alcohol level anywhere from 14 to 138 minutes after drinking. Additionally, the amount of food consumed by the defendant affects the rate of absorption. We therefore agree with other jurisdictions that have concluded that the legislature "could not have intended to place such impossible roadblocks in the way of the State in prosecuting [DUI per se] cases."

Asking the impossible of litigants is a recurring concern in toxic-tort litigation. For example, in many situations, there can be no epidemiological evidence and, therefore, creating a requirement that the plaintiff provide it is tantamount to a rule that the plaintiff loses.

Setting aside specific statutory requirements, the answer to the issue in these cases turns on two questions. First, did the defendant have any opportunity to consume alcohol between the time of driving and the time of the test? Obviously, if the answer is yes, then a presumption that the defendant's BAC while driving was as high as or higher than that indicated by the test is unwarranted. Thus, it is important for the state to show that the defendant had no opportunity to consume alcohol after arrest.

Second, was the BAC on the test greater or less than legally permissible? If it was greater, then the defendant's BAC at the time of driving could have been lower only if the defendant was still in the absorption phase at the time the test was taken.

If, on the other hand, the defendant's BAC on the test was less than legally permissible, proof that the BAC was above permissible levels while the defendant was operating the vehicle would require an extrapolation back to the time he was driving. And in this case, the extrapolation is appropriate only if there is evidence that the individual was already in the elimination phase when he was driving. (Again, one may use Figures IV-11 and IV-12 (Slides 83 and 84) to make this point.)

The New Mexico Supreme Court was sensitive to this issue in *State v. Downey*, 195 P.3d 1244 (N.M. 2008). It excluded the state's expert's retrograde extrapolation because it was predicated on factual assumptions unsupported by the record, and thus was unreliable and inadmissible. The court articulated the following:

> Given that Smock did not have the facts necessary to plot Defendant's placement on the BAC curve, he could not express a reasonably accurate conclusion regarding the fact in issue: whether Defendant was under the influence of intoxicating liquor at the time of the collision. Smock's testimony did not "fit" the facts of the present case because he simply assumed for the purpose of his relation-back calculations that Defendant had ceased.

Downey, 195 P.3d at 1252.

The court distinguished the case from an earlier case, *State v. Hughey*, 163 P.3d 470, 475 (N.M. 2007), noting that in the latter case, the expert witness "possessed sufficient facts to support her conclusion that the defendant was at her peak alcohol level at the time the accident occurred."

In *Tsosie* and many other retrograde-extrapolation cases, the expert offered a range of possible BACs. Few of the cases are explicit about the assumptions the expert made about rates of absorption and elimination. However, as should be clear from the *Tsosie* case, estimating absorption rates and peak BAC is more difficult than estimating a BAC at an earlier point in time, given that it is known that throughout the period in question the subject was in the elimination phase. The instructor may wish to alter the facts in the main case to explore with the students these various scenarios.

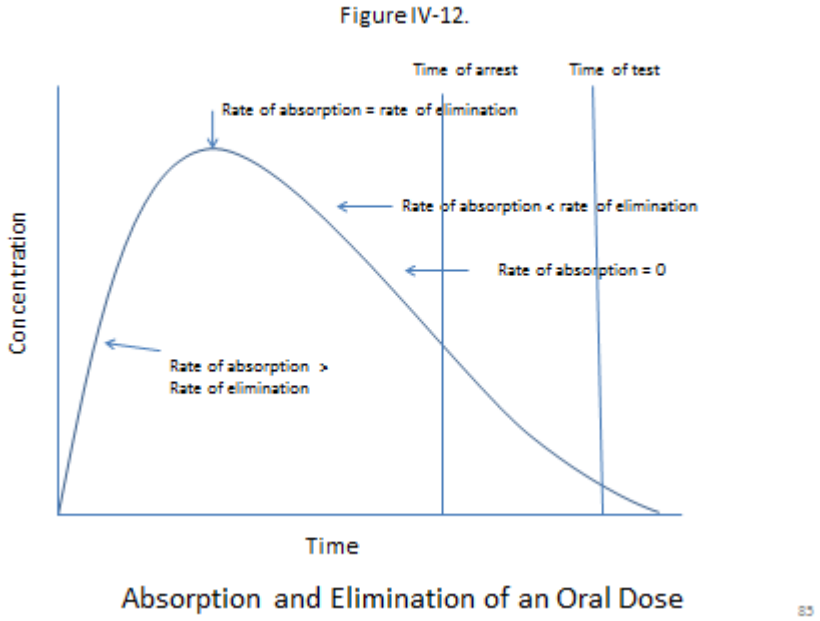
The problem of timing is even more difficult when we are confronted with impairment due to some other substance, be it marijuana, cocaine, amphetamines, or some new, synthetic drug. Problems exist at two levels. First, there is no simple test to quantify the amount of a substance in the individual. Urine tests, for example, indicate whether some drug metabolites are present, but quantification often requires more sophisticated and expensive tests. Second, even when there is evidence concerning the amount of a substance in the body, precise measurement of the amount in the body at the time of an accident is at best difficult. For example, with respect to marijuana use, there is some ability to estimate the time of last use, but this is not the same as the precision possible with respect to alcohol. See Marylyn A. Huestis et al., "Estimating the Time of Last Cannabis Use from Plasma Delta9-Tetrahydrocannabinol and 11-Nor-9-Carboxy-Delta9-Tetrahydrocannabinol Concentrations," 51 *Clinical Chemistry* 2289 (2005). Not even this level of precision is possible with respect to many other drugs. The general point that these examples raise is that the pharmacokinetics of other drugs is less well understood.

Nevertheless, some courts have permitted expert testimony akin to a retrograde extrapolation in drug cases. For example, in *Bekendam v. State*, 441 S.W.3d 295 (Tex. Crim. App. 2014), the court permitted a retrograde extrapolation in a cocaine case. The expert used the half-life of the drug to estimate the time the drug was ingested. And he testified as to how the drug was metabolized and the effect the drug would have on the central nervous system. All of this feels a bit suspicious, but it only highlights the special difficulty that courts face when a defendant is charged with driving under the influence of something other than alcohol.

Notes following *Tsosie*:

Note 2 asks what assumptions the expert made. The primary assumption, of course, is that the defendant was in the elimination phase when last driving. But the expert also made assumptions about the rate of elimination and created bounds around her estimate of the defendant's BAC. One may wish to make it clear to the students that the reasonableness of the expert's opinion turns on the upper and lower bounds concerning elimination rates that the expert used to create a confidence band. See Figure IV-11 (Slide 84, above) and Figure IV-12 (Slide 85).

Slide 85 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.

b. Toxicology in Civil Cases

As we noted in the module, the *Johnson* case raises three important 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 and proof of causation in a private cause of action. Following are additional materials with respect to each of these topics.

Johnson Case:

i. The Admissibility and/or Sufficiency of Various Types of Toxicological Evidence to Prove Causation

External validity challenges to toxicological evidence

It is very unlikely that a court will permit an expert to base an opinion solely on structure-activity or in vitro research, and several cases have expressed such views.

As to structure-activity, one court offered this comment:

> Most of the evidence on which Dr. Jafek relies in expressing the opinion that the zinc in Cold-Eeze causes anosmia is derived from studies of zinc sulfate. The active ingredient in Cold-Eeze, however, is zinc gluconate. This difference is potentially important, because "even minor deviations in molecular structure can radically change a particular substance's properties and propensities."
>
> Polski v. Quigley Corp., No. 04-CV-4199 (PJS/JJG), 2007 WL 2580550, *6 (D. Minn. Sept. 5, 2007), *aff'd*, 538 F.3d 836 (8th Cir. 2008).

Note 1 following the *Johnson* opinion asks students to classify and assess the "class of chemicals" analysis in the opinion. They should be lead to understand that this is a structure-activity issue with respect to MBTC. As the materials presented below indicate, there is a fair amount of research on hydrochloric acid, but very little on MBTC. Most of the research on this latter chemical is on other, closely related tin-based chemicals, and it appears that some of them with similar structure may be more toxic than MBTC.

In vitro research fares little better in court. In *Wade-Greaux v. Whitehall Laboratories*, Inc., the plaintiff's expert based her conclusion, in part, on chick embryo research. The court focused on pharmacokinetic issues in rejecting the testimony. It noted:

> Such chick embryo studies are an in vitro, not an in vivo, animal model, and do not replicate a mammalian, let alone a human, exposure . . . . At least four things happen in the exposure of a human fetus that do not occur in the chick embryo model: (I) absorption of the agent by the

> mother, (ii) distribution of the agent throughout the maternal and fetal systems, (iii) metabolism of the agent by the maternal system and (iv) elimination of the agent by the mother and fetus, thereby limiting the duration of the exposure. In chick embryo studies, the agent is dropped directly onto the embryo and is not subjected to maternal metabolism, is not distributed throughout the maternal and embryonic systems, and is not eliminated from the embryonic system, resulting in constant exposure until birth.
>
> *Wade-Greaux v. Whitehall Laboratories, Inc.*, 874 F. Supp. 1441, 1456–57 (V.I. 1994), *aff'd*, 46 F.3d 1120 (3d Cir. 1994).

In a well-reasoned and detailed discussion of in vitro research, the court in *In re Rezulin Products Liability Litigation*, noted that the reasonableness of extrapolating from in vitro studies to human injuries turns on the type of cell on which the in vitro experiment was performed and the dose to which the cells were exposed. The plaintiffs claimed that Rezulin caused their liver injury, but the in vitro studies upon which they relied either did not involve healthy human liver cells or exposed the cells to higher concentrations than would occur in a living human. *In re Rezulin Products Liability Litigati*on, 369 F. Supp. 2d 398, 429 (S.D.N.Y. 2005).

These cases provide interesting examples for discussion. However, it is unlikely that a toxicologist would conclude that a substance causes harm to an individual based solely on these types of information.

Research based on animal studies present a much more complex picture.

Toxicology in the form of animal studies came to the fore in the litigation surrounding the herbicide Agent Orange, which was widely used in the Vietnam war to destroy jungle canopies. In the years following the war, military personnel who were exposed to the herbicide sued manufacturers of Agent Orange, alleging that their exposure caused them to suffer from various health problems. The district court certified a class action, but the defendants in the following case had opted out of the class, seeking a suit on the merits for their particular injuries. Judge Jack Weinstein, one of the foremost evidence scholars of his time, excluded the plaintiff's expert testimony, and in doing so, made the following statement about the use of animal studies in litigation.

> The many studies on animal exposure to Agent Orange, even plaintiffs' expert concedes, are not persuasive in this lawsuit. In a jointly-authored article, Dr. Ellen K. Silbergeld writes that "laboratory animal studies . . . are generally viewed with more suspicion than epidemiological studies, because they require making the assumption that chemicals behave similarly in different species." Hall & Silbergeld, "Reappraising

> Epidemiology: A Response to Mr. Dore," 7 Harv.Envtl.L.Rev. 441, 442–43 (1983). Dr. Silbergeld further notes that "[a]nimal studies are aimed at discovering a dose-response relationship, while epidemiological studies show an association between exposure and disease." *Id.* at 443 n. 18.
>
> There is no evidence that plaintiffs were exposed to the far higher concentrations involved in both the animal and industrial exposure studies. *Cf. In re "Agent Orange" Product Liability Litigation,* 597 F.Supp. 740, 782 (E.D.N.Y.1984). The animal studies are not helpful in the instant case because they involve different biological species. They are of so little probative force and are so potentially misleading as to be inadmissible. See Fed. R. Ev. 401–03. They cannot be an acceptable predicate for an opinion under Rule 703.
>
> *In re "Agent Orange" Product Liability Litigation*, 611 F. Supp. 1223, 1241 (E.D.N.Y. 1985):

Although this extreme view has been parroted in other cases, most courts adopt a more balanced view. For example, in *Metabolife International, Inc. v. Wornick*, the court of appeals reversed a trial court ruling that as a matter of law animal studies are inadmissible "due to the uncertainties in extrapolating from effects on mice and rats to humans," and concluded that the "district court erred in rejecting the animal studies proffered by Metabolife merely because of the species gap." *Metabolife International, Inc. v. Wornick*, 264 F.3d 832, 842 (9th Cir. 2001).

When testimony based on animal studies is declared inadmissible, there is usually some problem with the research in addition to the basic external validity problem created by extrapolation across species. Most frequently, the problem is one of three kinds: (a) a lack of similarity between the substance in the case and the substance in the studies; (b) a lack of similarity between the injury in the case and the injury in the studies; or (c) a lack of similarity in the dose rate found in the studies and the does rate in the case.

For example, with respect to a lack of similarity between the substance in the case and the substance in the studies, the court in *In re Accutane Products Liability*, 511 F. Supp. 2d 1288, 1294 (M.D. Fla. 2007), *aff'd*, Rand v. Hoffman-LaRoche Inc., 291 Fed. App'x 249 (11th Cir. 2008), noted:

> Dr. Fogel also uses two rat studies to support his conclusion. The rat studies, one done in Poland and another in Russia, were actually studies of high doses of vitamin A, not Accutane. . . . He makes no effort to analogize the effect of high doses of vitamin A to the effect of Accutane on rats, much less on the human body. That is obviously important because, while Accutane is one of many "retinoids," i.e., a group of

> compounds derivative of vitamin A, the human body reacts to Accutane and vitamin A in different ways.

Ruff v. Ensign-Bickford Industries, Inc., 168 F. Supp. 2d 1271 (D. Utah 2001) provides an illustrative example of a lack of similarity between the injury in the case and the injury in the studies. The court granted the defendant's motion to exclude part of the testimony that linked RDX to the plaintiff's injury. The expert came to this conclusion based on animal research:

> The [animal] study shows that female mice contracted liver cancer when injected with RDX. Plaintiffs have not shown how this correlates with their lymphomas or why it is scientifically valid to extrapolate from the study that liver cancer in female mice are predictive of human lymphomas. While the Court agrees with plaintiffs that the Lish study is entitled to some weight, the Court cannot conclude that Dr. Weisenburger's opinion regarding RDX, which is primarily founded on the Lish study, is based on sufficient facts and data. [The revised Fed. R. Evid. 702 now requires that an expert's testimony be "based upon sufficient facts or data."] Accordingly, defendants' motion in limine is granted as it relates to Dr. Weisenburger's conclusion that RDX causes NHL in humans.
>
> Id. at 1280.

On the other hand, the court admitted the same expert's testimony concerning Hydrazine. This testimony was also based on animal studies which indicated that Hydrazines cause NHL in mice, the same type of cancer that afflicted the plaintiffs in this case.

Dose differences are central to several rulings in litigation surrounding Parodel, a drug given to postpartum women to suppress lactation. In *Soldo v. Sandoz Pharmaceuticals Corp.*, 244 F. Supp. 2d 434, 530 (W.D. Pa. 2003), the district court judge expressed the following reservations concerning the results of animal studies presented in the case:

> For example, the testimony of Drs. Petro and Kulig was based in part on their review of various animal studies in which bromocriptine was tested and in which certain allegedly vasoconstrictive phenomena (e.g., necrosis of the tips of dog ears, necrosis of the tips of rat tails, arterial constriction in the hind limb of a dog) were observed. NPC demonstrated—and Dr. Petro admitted—that the doses administered to these animals were hundreds and thousands of times higher than would obtain in a woman using Parlodel (r) for the prevention of lactation.

Dunn v. Sandoz Pharmaceuticals Corp., 275 F. Supp. 2d 672 (M.D.N.C. 2003), another Parlodel case, expressed similar reservations:

> An additional difficulty in using animal studies in an attempt to establish causation in human beings is "that the high doses customarily used in animal studies" make extrapolating the effect on much lower doses in humans very difficult to determine. Reference Manual on Scientific Evidence 346 (2d ed.2000).

Of course, because most animal studies employ doses greater than human dosage, the difficult question confronting courts is how high of a dosage is too high to permit an expert to base an admissible conclusion on the animal research? To this question, the cases offer no clear answer.

Some animal studies exhibit more than one of these problems. A well-known example is a study discussed in *General Electric Co. v. Joiner*, 522 U.S. 136, 144–45 (1997). There, Chief Judge Rehnquist made the following observation concerning an animal study proffered by the plaintiff, who alleged that he suffered from lung cancer due to exposure to polychlorinated biphenyls (PCBs).

> The District Court agreed with petitioners that the animal studies on which respondent's experts relied did not support his contention that exposure to PCB's had contributed to [the plaintiff's] cancer. The studies involved infant mice that had developed cancer after being exposed to PCB's. The infant mice in the studies had had massive doses of PCB's injected directly into their peritoneums or stomachs. Joiner was an adult human being whose alleged exposure to PCB's was far less than the exposure in the animal studies. The PCB's were injected into the mice in a highly concentrated form. The fluid with which Joiner had come into contact generally had a much smaller PCB concentration of between 0-to-500 parts per million. The cancer that these mice developed was alveologenic adenomas; Joiner had developed small-cell carcinomas. No study demonstrated that adult mice developed cancer after being exposed to PCB's. One of the experts admitted that no study had demonstrated that PCB's lead to cancer in any other species.
>
> Respondent failed to reply to this criticism. Rather than explaining how and why the experts could have extrapolated their opinions from these seemingly far-removed animal studies, respondent chose "to proceed as if the only issue [was] whether animal studies can ever be a proper foundation for an expert's opinion." 864 F.Supp., at 1324. Of course, whether animal studies can ever be a proper foundation for an expert's

> opinion was not the issue. The issue was whether these experts' opinions were sufficiently supported by the animal studies on which they purported to rely. The studies were so dissimilar to the facts presented in this litigation that it was not an abuse of discretion for the District Court to have rejected the experts' reliance on them.

According to Justice Rehnquist, this study failed on both the injury involved and the dose administered to the animals' grounds. One may wish to give this excerpt to the class and ask the students to ascertain the study's shortcomings.

Dealing with the "no-threshold" assumption

The injury in the *Johnson* case was fibrosis, the type of injury for which most toxicologists would conclude that there is likely to be some threshold below which exposure does not cause harm (Figure IV-5). Agencies, following in the footsteps of most toxicologists, routinely assume there is no safe threshold with respect to substances that are known carcinogens. As we noted when discussing the *Chlorine Chemistry* and *Industrial Union* cases, this poses dilemmas for courts wrestling with the reasonableness of agency standards. The assumption has also created difficulties for courts wrestling with causation in toxic tort cases.

Plaintiffs have advanced a no-threshold theory in several cases. However, the theory has not been well received. Typical of judicial responses is the comment from *Parker v. Mobil Oil Corp.*, 16 A.D.3d 648, (N.Y. App. Div. 2005), *aff'd on other grounds*, 857 N.E.2d 1114 (N.Y. 2006) ("the scientific reliability of this methodology has flatly been rejected as merely a hypothesis."). *Sutera v. Perrier Group of America Inc.*, 986 F. Supp. 655 (D. Mass. 1997) also involved benzene exposure. See also *National Bank of Commerce v. Associated Milk Producers*, Inc., 22 F. Supp. 2d 942 (E.D. Ark. 1998), *aff'd*, 191 F.3d 858 (8th Cir. 1999) (aflatoxin M–1 (AFM)); *Burleson v. Texas Department of Criminal Justic*e, 393 F.3d 577 (5th Cir. 2004) (radiation).

This issue has been the source of a substantial number of opinions addressing the "any exposure" or "every exposure" arguments presented by plaintiffs in asbestos cases where the quantity of the exposure attributable to some defendants is unknown, but relatively small. See Joseph Sanders, The "Every Exposure" Cases and the Beginning of the Asbestos Endgame, 88 Tul. L. Rev. 1153 (2014). The most direct form of this argument is that any exposure constitutes a substantial factor with respect to the plaintiff's asbestosis or mesothelioma. See, example.g., *Comardelle v. Pennsylvania General Insurance Co*., 76 F. Supp. 3d 628, 632, 634 (E.D. La. 2015):

> Accordingly, Dr. Hammar opines based on this "every exposure" theory that "if Comardelle was exposed to asbestos released from Amchem or Benjamin Foster adhesives, . . . those exposures [would] have been a substantial contributing cause of his disease. . . ."
>
> Although there may be no known safe level of asbestos exposure, this does not support Dr. Hammar's leap to the conclusion that therefore every exposure Comardelle had to asbestos must have been a substantial contributing cause of his mesothelioma.

The problem in these cases is not that the no-threshold assumption is wrong; it is that plaintiffs have encouraged their experts to invoke the idea of a "substantial factor." The phrase "substantial factor" is a legal term of art, usually meaning that if something is found to be a substantial factor, the plaintiff has met her burden of proof on causation. By equating the no-threshold assumption with the legal idea of a substantial factor, plaintiffs hoped to both sidestep the necessity of quantifying the dose to which plaintiff was exposed by defendant and to neutralize epidemiologic evidence that fails to demonstrate a connection between work in certain occupations and asbestos-related diseases. For example, with respect to auto mechanics who claim that exposure to asbestos while changing brake pads caused their asbestos-related disease, see Michael Goodman et al., "Mesothelioma and Lung Cancer Among Motor Vehicle Mechanics: A Meta-analysis," 48 *Annals Occupational Hygiene* 309 (2004):

> The meta-analysis for Tier I (higher quality) and Tier II (lower quality) studies of mesothelioma yielded RR estimates of 0.92 (95% CI 0.55–1.56) and 0.81 (95% CI 0.52–1.28), respectively. Further stratification according to exposure characterization did not affect the results. The meta-analysis for lung cancer produced RR estimates of 1.07 (95% CI 0.88–1.31) for Tier I and 1.17 (95% CI 1.01–1.36) for Tier II. When the lung cancer analysis was limited to studies that used adequate control for smoking, the resulting RR estimate was 1.09 (95% CI 0.92–1.28). Based on these findings, we conclude that employment as a motor vehicle mechanic does not increase the risk of developing mesothelioma. Although some studies showed a small increase in risk of lung cancer among motor vehicle mechanics, the data on balance do not support a conclusion that lung cancer risk in this occupational group is related to asbestos exposure.

ii. The Relationship between Toxicological and Epidemiological Evidence

As discussed in the epidemiology section and suggested by the meta-analysis on auto mechanics, where there is a body of epidemiological evidence, courts are much less receptive to expert opinions based on other types of toxicological evidence. Courts rarely say that epidemiological evidence is required before the plaintiff can prevail, but they sometimes come close to saying that in the face of a large body of contrary epidemiology, appeals to other toxicology evidence are unavailing. Some of the Bendectin cases came to this conclusion. In *Raynor v. Merrell Pharmaceuticals Inc*., 104 F.3d 1371 (D.C. Cir. 1997), the court used a Rule 702 analysis to conclude that the plaintiff's expert testimony was inadmissible and therefore defendant was entitled to summary judgment. The court concluded that when epidemiological evidence is to the contrary, it is not methodologically sound to draw an inference from chemical structure, in vivo animal studies, and in vitro studies that Bendectin causes human birth defects.

A slightly more modest statement can be found in *Norris v. Baxter Healthcare Corp.*, a silicone breast implant case.

> This is not a case where there is no epidemiology. It is a case where the body of epidemiology largely finds no association between silicone breast implants and immune system diseases. We are not holding that epidemiological studies are always necessary in a toxic tort case. We are simply holding that where there is a large body of contrary epidemiological evidence, it is necessary to at least address it with evidence that is based on medically reliable and scientifically valid methodology.

Norris v. Baxter Healthcare Corp., 397 F.3d 878, 882 (10th Cir. 2005).

Both the Bendectin and the silicone-breast-implant cases are, however, at the very tail end of the spectrum when it comes to the number of quality epidemiological studies pointing toward the conclusion that there is not a causal relationship. (The Bendectin research is summarized in Joseph Sanders, *Bendectin On Trial: A Study of Mass Tort Litigation* (1998). The silicone-breast-implant research is thoroughly discussed in Institute of Medicine et al., *Safety of Silicone Breast Implants* (2000)).

When there is very little epidemiological evidence or when there is no epidemiological evidence at all, courts are more amenable to other evidence. In *Ruff v. Ensign-Bickford Industries, Inc.*, 168 F. Supp. 2d 1271, 1281 (D. Utah 2001), even though there was one negative epidemiology study, the court admitted the plaintiffs' expert's causation testimony based on multiple animal studies showing that the same substance to which the plaintiffs were exposed caused the same disease from which the plaintiffs suffered. In *In re Heparin*

Products Liability Litigation, 803 F. Supp. 2d 712 (N.D. Ohio 2011), the court emphasized this distinction:

> Courts have rejected non-epidemiological evidence as unreliable where there is an overwhelming body of epidemiological evidence to the contrary. . . . Here, however, there is no such overwhelming body of contrary epidemiological evidence. Defendants point to two epidemiological studies, neither of which were designed to determine whether there was an association between contaminated heparin and any of the conditions identified in defendants' motion for summary judgment. Absence of proof is not proof of absence, and while these studies do not provide support for plaintiffs' theories, neither do they contradict them. I will not, therefore, exclude plaintiffs' evidence on these grounds.

In this regard, it is worth focusing on the fact that in the *Johnson* case, the court noted that there was no epidemiological evidence linking exposure to MBTC or HCl and restrictive lung disease and pulmonary fibrosis. What the court did not say is that there were no epidemiologic studies at all. Nor was it clear how such studies could be done because, as the opinion itself noted, the only way one would suffer from substantial exposure would be because of an industrial accident. As the WHO review cited below noted:

> Very few data are available on the effects of organotins in humans. Of the reported unintentional occupational exposures, none has an estimate of exposure concentration. Exposure was largely via the inhalation route, with some possibility of dermal exposure. Neurological effects were the most commonly reported, and these can persist for long periods.

iii. The Relationship between Regulatory Standards and Proof of Causation in a Private Cause of Action

The *Johnson* opinion reflects the predominant point of view in the case law. A useful statement of this position may be found in *McClain v. Metabolife International, Inc.*, 401 F.3d 1233 (11th Cir. 2005). The plaintiffs argued that the use of an herbal weight-loss supplement containing ephedrine and caffeine caused ischemic strokes in three of the users and a heart attack in the other. The court noted that one expert put great weight on an FDA proposal to restrict the sale and distribution of herbal supplements containing ephedrine. It commented:

> [] O'Donnell's use of FDA data and recommendations raises a more subtle methodological issue in a toxic tort case. The issue involves identifying and contrasting the type of risk assessment that a government agency follows for establishing public health guidelines versus an expert analysis of toxicity and causation in a toxic tort case.
>
> The Reference Manual on Scientific Evidence explains that:
>
> proof of risk and proof of causation entail somewhat different questions because risk assessment frequently calls for a cost-benefit analysis. The agency assessing risk may decide to bar a substance or product if the potential benefits are outweighed by the possibility of risks that are largely unquantifiable because of presently unknown contingencies. Consequently, risk assessors may pay heed to any evidence that points to a need for caution, rather than assess the likelihood that a causal relationship in a specific case is more likely than not.
>
> Id. at 1249. See Margaret A. Berger, The Supreme Court's "Trilogy on the Admissibility of Expert Testimony," in Federal Judicial Center (ed.), *Reference Manual on Scientific Evidence* 33 (2d ed. 2000). Obviously, in a toxic tort case, the court must focus on assessing causation, not on a cost-benefit analysis for restricting the sale and use of a drug.

Similar comments can be found in *Mitchell v. Gencorp Inc.*, 165 F.3d 778 (10th Cir. 1999) and other cases. For example, in *Parker v. Mobil Oil Corp.,* 857 N.E.2d 1114 (N.Y. 2006), the court stated that "standards promulgated by regulatory agencies as protective measures are inadequate to demonstrate legal causation."

iv. Additional Information on the Toxic Nature of the Chemicals in the *Johnson* Case

Some instructors may find the facts of the *Johnson* case useful for going into greater detail about the available research on the toxic effects of MBTC and HCL. Following are materials gathered from the cited web pages.

For one review of the health effects of MBTC and related chemicals, see Stuart Dobson et al., World Health Org., *Mono- and Disubstituted Methyltin, Butyltin, and Octyltin Compounds* (2006), http://www.inchem.org/documents/cicads/cicads/cicad73.pdf. The page explains the purpose of using this chemical in greater detail than is reported in the *Johnson* opinion. According to this report, the use of monobutyltin

trichloride in glass coating is widespread and of long duration. The report states that:

> Approximately 700 tonnes per year of monobutyltin trichloride are used in hot-end coating of glass bottles, and a further 60–100 tonnes per year are used in the coating of flat glass [in Great Britain]. This technique was developed as an alternative to coating with tin tetrachloride. Hot glass products are exposed to hot air containing monobutyltin trichloride liquid and vapour. On the glass surface, the atomized liquid and vapour react to form tin oxide, which strengthens the glass, filling any "micro-cracks" in the glass.
>
> The above process is well established, having been introduced around 35 years ago, and is reported to be universally applied in the glass industry. It produces a surface that is more resistant to scratching and splintering. It should be noted that this process does not leave any residue of organic tin on the glass surface, since it is all converted to tin oxide through heating to over 400 °C.
>
> Industry has indicated that the number of production lines undertaking coating of glass bottles with monobutyltin trichloride is around 500, with an estimated 2000 production lines worldwide.

The report reviews the in vitro and in vivo research on these substances. However, almost all the animal research reported there involved ingestion of the substance in question and, therefore, does not address the plaintiff's complaint.

A CDC review of the toxicity of butyltins may be found at Centers for Disease Control & Prevention, *Biologic Effects of Exposure*, http://www.cdc.gov/niosh/pdfs/77-115b.pdf (last visited Apr. 22, 2016) and Centers for Disease Control & Prevention, Health Effects, http://www.atsdr.cdc.gov/toxprofiles/tp55-c3.pdf (last visited Apr. 22, 2016). These reviews do report inhalation studies, however, once again, most of the research involves oral exposure and none of the reported research focuses specifically on MBTC—which is generally thought to be less toxic than other similar chemicals, e.g. dibutyltin trichloride, tributyltin trichloride, and many other similar chemicals.

OSHA establishes a "target concentration" of all butyltin trichlorides at 0.1 mg/m^3, but this is not based on a starting LOAEL for inhalation exposure, apparently because there is insufficient data to calculate a LOAEL. Occupational Safety & Health Admin., U.S. Dep't of Labor, *Butyltin Trichloride*, https://www.osha.gov/dts/sltc/methods/partial/id217sg/id217sg.html (last visited Apr. 22, 2016).

Not surprisingly, there is better data with respect to hydrochloric acid (hydrogen chloride). A useful starting place is U.S. Envt'l Protection Agency, *Hydrochloric Acid (Hydrogen Chloride): Hazard Summary* (2000), http://www3.epa.gov/airtoxics/hlthef/hydrochl.html.

This report refers to the Integrated Risk Information System review of risks posed by the inhalation of the hydrogen chloride. See Integrated Risk Information System (IRIS) et al., *Chemical Assessment Summary: Hydrogen Chloride*; CASRN 7647-01-0 (1991), http://cfpub.epa.gov/ncea/iris/iris_documents/documents/subst/0396_summary.pdf. The IRIS discusses the process by which the EPA established its LOAEL and RfC for *chronic* exposures. The LOAEL(HEC), that is, a LOAEL adjusted for dosimetric differences across species to a human-equivalent concentration, established from animal studies (including the baboon study discussed in the case) is 6.1 mg/m^3 and the RfC is 0.2 mg/m^3. In arriving at the RfC, the EPA used an uncertainty factor of 300, comprised of a factor of 3 for interspecies differences, 10 for interspecies extrapolations, and 10 to extrapolate from a LOAEL to a NOAEL. Of course, these values are not exactly on point for the *Johnson* case, which involved an acute, or more specifically, two acute exposures. With respect to acute exposures, the "OSHA permissible exposure limit" (PEL) is 7 mg/m^3, which is equivalent to 5 ppm. See Occupational Safety & Health Admin., U.S. Dep't of Labor, *Hydrogen Chloride*, https://www.osha.gov/dts/chemicalsampling/data/CH_246300.html (last revised Nov. 16, 2012).

v. Notes Following the *Johnson* Opinion

Note 1 is designed to encourage students to address the value of structure-activity evidence. As the above material indicates, MBTC appears to be less toxic than some quite similar chemicals, including dibutyltin trichloride and tributyltin trichloride. As an aside relating to the value of *in vitro* testing, see E. Dinant Kroese et al., "Evaluation of an Alternative In Vitro Test Battery for Detecting Reproductive Toxicants in a Grouping Context," 55 *Reproductive Toxicology* 11 (2015). The paper analyzes a battery of tests that were able to correctly distinguish non- or weak-developmental toxicants, including monobutyltin trichloride, from structurally related developmental toxicants, including tributyltin chloride.

Notes 2 and 3 address the baboon and rat studies referred to in the opinion. There are several issues raised by these studies. Most fundamentally, they pose complex extrapolation problems. In addition, both studies—but especially the rat study—involved research on chronic exposure, not acute exposure. One may wish to discuss the difference with the students and ask

whether fibrosis injuries are more likely to occur from acute or chronic exposures. Presumably, this is the primary reason the trial court disregarded the rat study. Finally, there is a question of thresholds. Unlike the *Chlorine Chemistry* case, this case involves an injury for which there is commonly believed to be a threshold (expressed as ppm or mg/m^3) below which injury does not occur. As the above materials discuss, the EPA does in fact establish a LOAEL for HCL inhalation. One may wish to engage students in a discussion of what effect this should have on the case. At the very least, students should recognize that here, the plaintiff must say something about dose.

As Note 4 and the above materials indicate, there are no epidemiological studies showing a relationship between these chemicals and toxic injury because there are no epidemiological studies, period. An instructor may wish to ask students to reflect on the disingenuous nature of the court's epidemiology comment, which could easily be read to mean that there are epidemiology studies that fail to report a toxic effect.

Regarding Note 5, the OSHA issue allows one to open up a discussion on the difference between precautionary standards used in the regulatory arena and the civil law requirement that the plaintiff prove causation by a preponderance of the evidence. Students will see this more clearly if one walks them through the process of first establishing a LOAEL, and then setting a RfC through the use of rather arbitrary multipliers.

Note 7 allows the instructor to point out that when a causal chain from "cause" to "effect" is of very short duration, lay opinion based on temporal order often suffices. However, this is hardly ever the case in toxic torts and is thus precisely why experts play such a central role.

In a concurring opinion, Judge Reavley argued that the case should go to the jury on the chronic-injury issue as well as the acute-injury issue. Here is his opinion:

> REAVLEY, J., concurring:
>
> I agree that the summary judgment should be reversed, but I disagree with the ruling to deny the trier of fact the testimony of these highly qualified expert witnesses. There are fact issues, primarily the extent of exposure of the plaintiff to the chemical vapors and the diagnosis of his ailment. There may be a disagreement between the experts about what, if anything, it would take for inhalation of these vapors to damage the lungs to the extent of progressive disease. If that bears on the decision of the diagnosis and is in question, the trier of fact needs the assistance of these experts.

It is simply incredible to me to decide that Dr. Schlesinger is an unreliable source for any scientific question in this case. And the same is true of the other expert witnesses appearing in this record. But certainly Dr. Schlesinger is an eminent authority on respiratory toxicology and the study of the adverse effects of exposure to inhaled chemicals. Relying on the diagnosis made by Dr. Grodzin that the patient suffers from restrictive lung disease, Dr. Schlesinger explains the path of physiological response of the lungs inhaling these damaging vapors. And no one will deny that MBTC and hydrochloric acid, if inhaled to some extent, will damage the lungs. The majority in this opinion accept that testimony for "acute injuries," but decide not to allow proof of any relation to progressive disease.

The district court excluded the testimony of Dr. Schlesinger because he could not cite fully tested and peer reviewed studies proving that hydrochloric acid can cause restrictive lung disease. That significant damage may be done is merely factual information. The possible extent of damage from breathing these chemicals may be at issue and will require the testimony of fully qualified and experienced experts, of which Dr. Schlesinger is surely one. There are no studies to meet the requirement of the district court, and that is not surprising. How would that study be designed and conducted, by obtaining a large population of people to breathe this chemical vapor or that vapor in this volume or that volume, then to have their lung function tested and maybe biopsied? Where would so many persons be found to be subjected to this?

Studies that the district court required may prove helpful in determining the reliability of a particular scientific "theory or technique," *Daubert v. Merrell Dow Pharmaceuticals, Inc.,* 509 U.S. 579, 113 S.Ct. 2786, 125 L.Ed.2d 469 (1993), but that is no checklist and reliability may depend on "the nature of the issue, the expert's particular expertise, and the subject of his testimony." *Kumho Tire Company v. Carmichael,* 526 U.S. 137, 119 S.Ct. 1167, 1175, 143 L.Ed.2d 238 (1999). The trial court erred in excluding testimony that is possibly relevant and clearly reliable.

The gist of Judge Reavley's opinion is not entirely clear. Part of it is simply his assertion that the experts were well qualified, but surely the majority did not disagree with this point. The more difficult issue in the concurring opinion is Judge Reavely's apparent position that because the plaintiff was hurt and because the experts did about as well as anyone could, given the available data, that should have been good enough to get to a jury. The instructor may wish to summarize Judge Reavley's opinion to point out that ultimately, judges must exercise their judgment

as to how much warrant an expert must have for an opinion before the expert is permitted to testify. On the general issue of "how good is good enough?" see Joseph Sanders, "Expert Witness Ethics," 76 *Fordham L. Rev.* 1539 (2007).

Section IV.C. Differential Diagnosis and Specific Causation

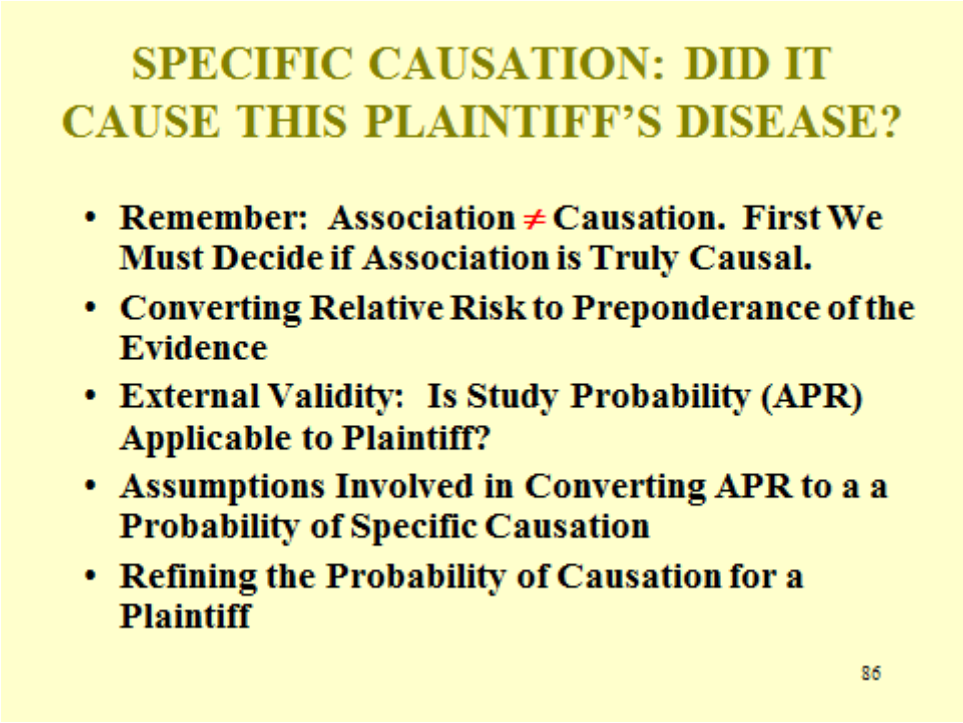
The following materials in Section IV.C. do not directly relate to the module. They are designed to provide additional resources that the instructor may or may not wish to use. Their purpose is to act as a bridge from this material to the material in Section V. on specific causation.

SECTION V. SPECIFIC CAUSATION

Section V.A. Introduction

Slide 86 explains in its title what specific causation is and provides an outline of the issues addressed in this section of the module.

Slide 86 Specific Causation: Did It Cause This Plaintiff's Disease?



A point made in this section (but perhaps without emphasis) is that before addressing specific causation, general causation must be established. Thus, in the absence of evidence on general causation, experts are generally not permitted to testify as to specific causation. An omitted portion of the opinion in *Johnson*, contained at the end of the module in the section on toxicology, addresses this matter, and instructors may want to use what follows to emphasize the importance of general causation.

Johnson v. Arkema, Incorporated

United States Court of Appeals, Fifth Circuit, 2012.
685 F.3d 452.

* * *

B.

Johnson next contests the district court's limitation of Dr. Grodzin's opinion, which prevented Dr. Grodzin from expressing his conclusion that MBTC and HCl caused Johnson's lung disease.[18] In reaching his causation conclusion, Dr. Grodzin's research and analysis essentially mirrored Dr. Schlesinger's save for one key distinction: Dr. Grodzin performed a "differential diagnosis" of Johnson. Accordingly, we need only consider whether, given the existence of Dr. Grodzin's differential diagnosis, the district court's exclusion of Dr. Grodzin's opinion constitutes the abuse of discretion.

As the Fourth Circuit has observed:

> 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 symptoms 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.

Westberry v. Gislaved Gummi AB, 178 F.3d 257, 262 (4th Cir.1999) (quoting *Kannankeril v. Terminix Int'l, Inc.,* 128 F.3d 802, 807 (3d Cir.1997)). Many courts have found that a properly performed differential diagnosis can yield a reliable expert opinion.

However, the results of a differential diagnosis are far from reliable *per se.* In *Moore,* for example, after conducting a differential diagnosis, the expert diagnosed the plaintiff with RADS. *Moore,* 151 F.3d at 273. The expert also concluded that the plaintiff's RADS was caused by certain chemicals to which the plaintiff was exposed based on his analysis of MSDS warnings, his examination and testing of the plaintiff, and the close temporal proximity between the plaintiff's exposure and subsequent injury. *Id.*

[18] Dr. Grodzin currently serves as the Medical Director for the Denton Medical Services Pulmonary Rehabilitation Center. Arkema does not challenge Dr. Grodzin's credentials.

Despite the expert's differential diagnosis, we held that the district judge did not abuse its discretion in excluding the expert's causation testimony because he failed to present reliable scientific support showing that the chemicals at issue could actually cause RADS.

Furthermore, *Moore* illustrates that an expert may not rely on a differential diagnosis to circumvent the requirement of general causation. * * * As we explained in *Knight*:

> General causation is whether a substance is capable of causing a particular injury or condition in the general population, while specific causation is whether a substance caused a particular individual's injury. Evidence concerning specific causation in toxic tort cases is admissible only as a follow-up to admissible general-causation evidence. Thus, there is a two-step process in examining the admissibility of causation evidence in toxic tort cases. First, the district court must determine whether there is general causation. Second, if it concludes that there is admissible general-causation evidence, the district court must determine whether there is admissible specific-causation evidence.

Knight v. Kirby Inland Marine Inc., 482 F.3d 347, 351 (5th Cir. 2007). Thus, before courts can admit an expert's differential diagnosis, which, by its nature, only addresses the issue of specific causation, the expert must first demonstrate that the chemical at issue is actually capable of harming individuals in the general population, thereby satisfying the general causation standard.

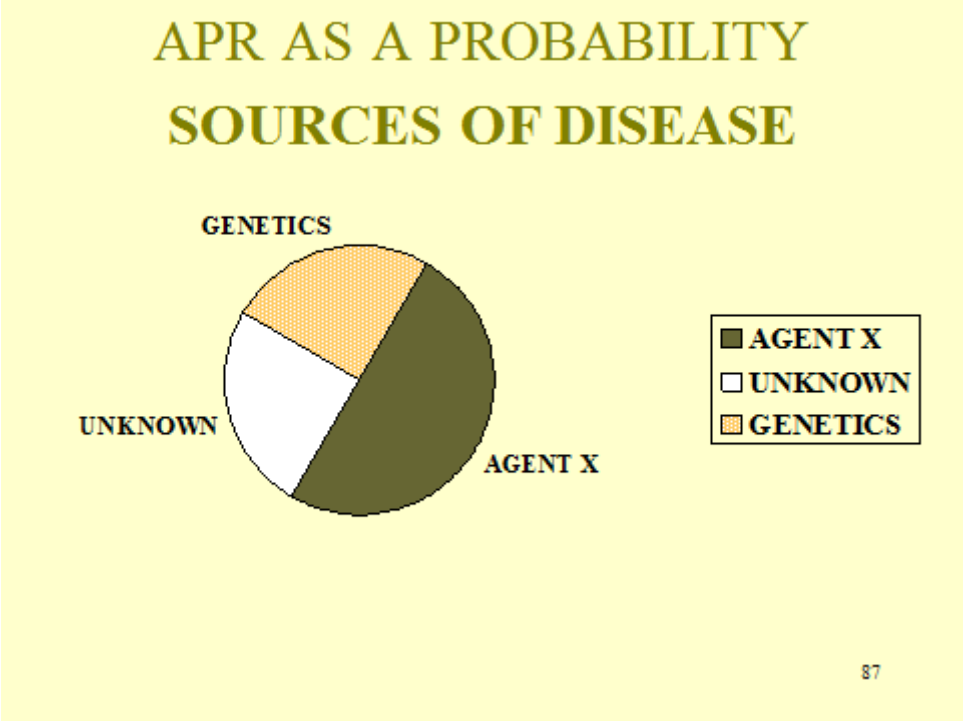
Here, like in *Moore*, Dr. Grodzin's differential diagnosis is based on the presumption that MBTC and HCl are actually capable of causing restrictive lung disease and pulmonary fibrosis in the general population. Dr. Grodzin has not presented any reliable or relevant scientific evidence to bolster this presumption. Instead, Dr. Grodzin essentially relied on the same scientific evidence and reached the same conclusions as Dr. Schlesinger. As we have explained, the district court did not abuse its discretion in excluding Dr. Schlesinger's opinion, thus negating Dr. Schlesinger's ability to satisfy the general causation requirement. Consequently, the fact that Dr. Grodzin conducted a differential diagnosis does not save his opinion from the same fate as Dr. Schlesinger's opinion. * * * The district court did not abuse its discretion in excluding Dr. Grodzin's causation opinion because, irrespective of the differential diagnosis, Dr. Grodzin is unable to satisfy the general causation requirement.

Section V.B. The Logic of Relative Risks Greater than 2.0

Students should appreciate that the probability that specific causation exists—all other things being equal, which they are not (more on this later)—corresponds to the attributable proportion of risk, a frequently used epidemiologic concept. Slide 87 depicts three sources of disease in a Venn diagram. The diagram reveals that half of the disease in the exposed group is associated with Agent X. The first step in determining the probability of specific causation would be to apply that datum to an individual in the study, which for Agent X is 50%. This probability stands right at the cusp of the preponderance of the evidence standard, which

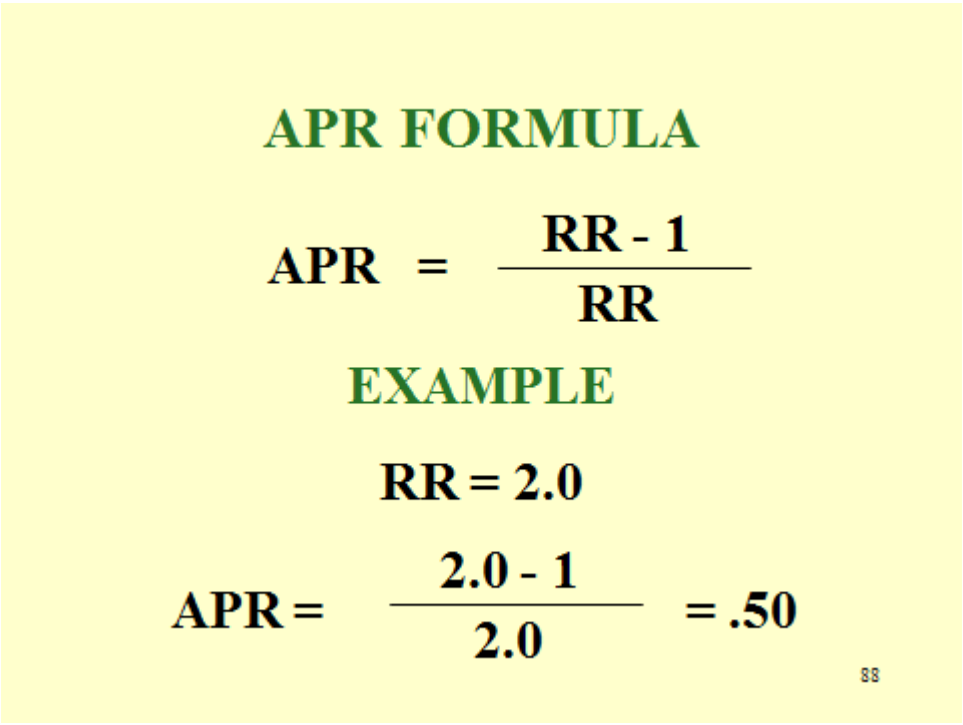
requires a bit more than 50%. Slide 88, a revised version of a slide employed earlier, displays the APR formula and illustrates that: the excess amount of disease attributable to the exposure (RR - 1) divided by the total incidence of the disease (RR)[15] is the average probability (more on this later as well) that any person in the exposed cohort contracted her disease because of the agent and, if the findings are externally valid (discussed below), reflects the average probability for other individuals who were exposed as well.

Slide 87 APR as a Probability: Sources of Disease



[15] The RR is the relative risk found in a study that was assessed to have identified a true causal relationship of the magnitude of the study's RR.

Slide 88 APR Formula



In this section, we have leapt into the difficult question of what role various relative risks play in answering the specific-causation question. However, the specific-causation question often arises in situations where there is no useful epidemiologic evidence. If a court concludes that there is sufficient evidence on general causation, the parties must proceed to the specific-causation issue. In these situations, of which there are many, the dispute is almost always about exposure to other possible causes and whether the plaintiff has successfully eliminated them as a possible cause of his particular ailment.

Following the *Johnson* excerpt, we include a specific-causation problem, based on *Best v. Lowe's Home Centers, Inc.*, 563 F.3d 171 (6th Cir. 2009). The instructor may wish to use this hypothetical here, even before addressing the matter of the role of relative risk in proving specific causation. One might also use it as part of Section V.F.1. on the Role of Information About the Individual Plaintiff.

The *Best* opinion represents a fairly typical differential-diagnosis discussion in situations where there is no epidemiology and where there is no genetic information available to permit placing the plaintiff in a subgroup of individuals based on anything except exposure to various toxic substances. All that is available is some evidence of other exposures that constitute competing causes.

The *Best* court states the following standard for a proper differential diagnosis:

> We hereby adopt the following differential-diagnosis test. . . . A medical-causation opinion in the form of a doctor's differential diagnosis is reliable and admissible where the doctor (1) objectively ascertains, to the extent possible, the nature of the patient's injury, ("A physician who evaluates a patient in preparation for litigation should seek more than a patient's self-report of symptoms or illness and . . . should . . . determine that a patient is ill and what illness the patient has contracted."), (2) "rules in" one or more causes of the injury using a valid methodology, and (3) engages in "standard diagnostic techniques by which doctors normally rule out alternative causes" to reach a conclusion as to which cause is most likely.

One might wish to point out that this test itself misstates the problem in an important way. In step (3) it suggests that there are "standard diagnostic techniques by which doctors normally rule out alternative causes." Perhaps the court is thinking about differential diagnosis as it is routinely used by physicians, that is, choosing among disease when the patient presents with a set of symptoms, not what is better termed as "differential etiology," that is, choosing among causes of a known disease. For this task, most physicians (including the expert in the problem) do not have standard techniques.

The basic problem confronting all specific-causation evidence is our frequent inability to extrapolate from a general finding to a particular individual. Recent work discussing this problem includes A. Philip Dawid et al., "Fitting Science into Legal Contexts: Assessing Effects of Causes or Causes of Effects?" 43 *Soc. Methods & Res.* 359 (2014); David L. Faigman et al., "Group to Individual (G2i) Inference in Scientific Expert Testimony," 81 *U. Chi. L. Rev.* 417 (2014); and Joseph Sanders, "Proof of Individual Causation in Toxic Tort and Forensic Cases," 75 *Brook. L. Rev.* 1367 (2010).

Specific-Causation Hypothetical

Donald Summer visited his local pool supply store in June of 2003. He located the pool product he was looking for, Aqua EZ Super Clear Clarifier (Aqua EZ), but when he lifted it from the shelf, an unknown quantity of the contents splashed onto his face and clothing. Allegedly, the container had been punctured accidently by a pool supply store employee who had opened the shipping box containing several bottles of Aqua EZ earlier that day. Summer went to a hospital emergency room on the day of the spill. Four months later, he sought treatment from Dr. Moore, who had practiced medicine as an otolaryngologist for more than two decades and also held a master's degree in chemical engineering.

Summer described the incident at the pool supply store to Dr. Moore. He said that the spilled product had a strong odor, and that immediately thereafter, he had suffered from irritation and burning of his skin, irritation to his nasal passages and mouth, dizziness, and shortness of breath. Summer also reported that he experienced clear drainage from his nose following the spill and that he eventually lost his sense of smell completely. Dr. Moore was

unable to inspect Summer's mucous membranes for physical damage because they are located too far inside the nasal passages to permit visual examination.

Summer did not see Dr. Moore again until January 2007. At that time, Summer was experiencing rhinitis—otherwise known as a runny or stuffy nose—with swelling and decreased airflow. He reported that, during the 3½-year period since the spill incident, he had struggled with rhinitis, anosmia, and dizzy spells. In April 2008, Dr. Moore administered to Summer the University of Pennsylvania Smell Identification Test (UPSIT), a standardized test of olfactory function. The test involves various sample chemicals, requiring the test subject to choose one of four descriptions of each sample's scent. Summer scored a 6 on the test, a score consistent with complete anosmia.

Dr. Moore testified in his deposition that "[l]oss of smell is caused by either a virus, an accident, tumors to the brain, surgery into the brain, or exposure to chemicals." He also conceded that sometimes anosmia is idiopathic, meaning that it occurs for unknown reasons, and that some medications can cause a loss of the sense of smell. Dr. Moore proceeded to list the following medications that Summer reported taking at the time of his chemical exposure: aspirin, Atenolol, Effexor, hydrochlorothiazide, Lescol, Lotensin, moxamorphin, OxyContin, Protonix, and Remeron. Dr. Moore stated that Atenolol and Lotensin are for blood pressure; aspirin, moxamorphin, and OxyContin are for pain; Effexor is for depression; hydrochlorothiazide is a fluid pill; and Protonex is for the stomach. He was unfamiliar with the drug Lescol. Referring to all of the medications, he stated that "[i]n my practice, with the patients that I have seen . . . over the years . . ., I have never seen an anosmia caused from the use of these medications." He also said that he had looked up all of the medications except Lescol in the course of his practice. Dr. Moore was unable to list the general types of medications that can cause a loss of the sense of smell.

The pool supply store provided Summer's attorney with a one-page document identifying the pool chemical as Aqua EZ. After receiving this document, Summer's attorney obtained a Material Safety Data Sheet (MSDS) prepared by Ciba Specialty Chemicals Corporation, the supplier of the active ingredient in Aqua EZ. Dr. Moore reviewed the MSDS, which described the characteristics of the active ingredient.

The relevant ingredient is an organic cationic polyelectrolyte. Specifically, the compound is a homopolymer with the name 2-Propen-1-aminium, N, N-dimethyl-N-2-propenyl-chloride. The MSDS identifies the chemical as "hazardous" and states that "[p]rolonged or repeated contact may cause eye and skin irritation." Primary routes of entry for the compound are listed as "Ingestion, Skin, Inhalation, Eyes." According to the MSDS, if the chemical is inhaled, the person should be "[r]emove[d] to fresh air, if not breathing give artificial respiration. If breathing is difficult, give oxygen and get immediate medical attention." The Handling Instructions state: "Do not inhale. . . . Use only with adequate ventilation." Under the heading "Engineering Controls," the MSDS instructs: "Work in well ventilated areas. Do not breathe vapors or mist." The MSDS also notes that "Acute Inhalation Toxicity" for the compound has not been determined. Dr. Moreno later reviewed a second MSDS, published by Sigma-Aldrich, another supplier of the relevant compound. That MSDS confirmed that the compound is

"irritating to the mucous membrane and upper respiratory tract" and that it "[m]ay be harmful if inhaled."

Dr. Moore concluded, based on the MSDS information, that the inhalation of Aqua EZ has the potential to cause damage to the nasal and sinus mucosa and the nerve endings of the olfactory bulb. According to Dr. Moore, the culprit components of the polymer in question include a chlorine derivative and an ammonium derivative. His opinion was that "a chemical burn can cause a loss of smell on a time basis" due to "scarring of the tissue," and reported that he has treated other chemical exposures with anosmic side effects following exposure to chlorine derivatives. Dr. Moore did not know the precise amount of the offending chemical that Summer had been exposed to, nor was he able to determine the threshold level of exposure that could cause harm. Dr. Moore summarized his diagnosis regarding causation this way:

> The patient had an accident, chemical was spilled, the patient cannot smell. If we have any trust in the patient at all, all I can say is he cannot smell. I did test him; his test was positive in the fact that he was anosmic. All I can tell you is that exposure to the-the only exposure that he had at the time that I talked to him was exposure to this chemical. There was nothing else in his history that dictated the fact that he was anosmic otherwise.

In short, because of the temporal relationship between Summer's exposure to the chemical and the onset of his symptoms, in conjunction with a principled effort to eliminate other possible causes of anosmia, Dr. Moore formed the opinion that the inhalation of Aqua EZ caused Summer to lose his sense of smell.

Study Question:

1. Assess this "differential diagnosis"

    a. Has the expert sufficiently "ruled in" the spill as a possible cause of Mr. Summer's anosmia? What evidence did he present on this question? What other evidence would you wish to have?

    b. What are the other possible causes of Mr. Summer's anosmia? Did the expert effectively rule them out, that is, exactly what did he say with respect to these causes?

    c. Is a temporal relationship sufficient to prove causation in this case? Why or why not?

Section V.B.1. Instructor Notes for Specific-Causation Hypothetical

In the hypothetical, the expert must rule in Aqua EZ and rule out other causes. Consider each step in turn.

What evidence ties Aqua EZ to Summer's injury? Summer's expert surely would like to have epidemiological studies or, failing this, even good animal studies. However, given the accidental nature of this type of injury, there is no way to develop a body of epidemiology. Moreover, even animal studies are difficult to conduct when the injury's end point—loss of smell—is at least somewhat subjective. Thus, the expert must try to make due with other evidence. What evidence would that be?

The answer is the Material Safety Data Sheet (MSDS) on the active ingredient in the Aqua EZ pool clarifier. This, of course, is the very type of data that was rejected in the *Johnson* case. It may be worth reviewing the court's discussion of MSDSs in that opinion. (Note, by the way, that the *Best* court seems to have accepted this evidence without question.).

The MSDS may tell us the concentration of the ingredient in Aqua EZ, but it cannot tell us very much about the amount to which the plaintiff was exposed. Thus, the problem allows the instructor to reprise the dose discussion from earlier in the toxicology materials.

The temporal order question is somewhat complicated on these facts because it is not quite clear when and to what degree Mr. Summer lost his sense of smell. One should point out to students that a firm diagnosis of anosmia did not occur until April, 2008, nearly 5 years after the chemical spilled on Mr. Summer.

The expert ruled out all other causes of Mr. Summer's injury. One is struck, however, by the very large number of drugs Mr. Summer was taking. Here is that list:

1. Atenolol and Lotensin for high blood pressure.
2. Aspirin, moxamorphin, and OxyContin for pain (moxamorphin is not listed in the PDF but it appears to be a morphine-based pain drug as is OxyContin).
3. Effexor for depression.
4. Hydrochlorothiazide, described as a "fluid pill" in the hypothetical, which is apparently used as a diuretic.
5. Protonix for the stomach (the drug inhibits gastric acid secretion).
6. Lescol—The expert was not familiar with this drug. It is a drug taken by individuals with high cholesterol levels.

A number of articles list anti-depressants, antihypertensive medication, and lipid-lowering drugs as implicated in the loss of smell. See, example.g., Norman M. Mann, "Management of Smell and Taste Problems," 69 *Cleveland Clinic J. Med.* 329 (2002).

Note that the expert did not purport to do an investigation of drug-related anosmia. He only said that "[i]n my practice, with the patients that I have seen . . . over the years . . ., I have never seen an anosmia caused from the use of these medications."

One may wish to ask students to do their own search for the relationship between prescription drugs and anosmia.

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

Judge Rothstein confronted the matter of external validity in a different context from that which scientists would, but her opinion reveals that this is a matter of reasonable inference. Adversarial experts, of course, are going to draw the inference in favor of the side that presents them to testify. So, Judge Rothstein was faced with deciding for herself whether this was a reasonable inference, as she did for other matters presented (with regard to other issues of external validity—application to plaintiffs with heart disease—she barred plaintiffs' experts from testifying). Would it not have been better if a (or several) neutral experts could have provided their expertise and assisted the judge?

In this excerpt, Judge Rothstein appeared most concerned with whether doctors and scientists make the same extrapolations that plaintiffs' experts proposed to make. Rather than focusing on the mechanisms of this disease and differences that might have existed between the studied population and the plaintiffs that would affect the disease process, the court cited to sources that assumed external validity, although primarily because there was no better evidence available.

This may reveal a difference between medical and legal standards of proof. In medicine, doctors frequently balance the costs and benefits of acting with those of not acting. If not acting can result in serious adverse consequences, doctors may decide to act even though there is great uncertainty about the benefits of doing so. A judge might label this decision as "impermissible speculation" and conclude that the party with the burden of proof has failed to meet her burden of production.

Study Questions:

1. In deciding whether the results of the HSP could be applied to those who were older than the study participants, Judge Rothstein stated that she "sees no reason why the increasing risk of stroke would render the HSP . . . unreliable as opposed to this age group." In what way might the increased risk of stroke in the older population affect the external validity of HSP to this population? Would any such effect render Judge Rothstein's conclusion that HSP should apply to this population erroneous?

   A greater risk of stroke in the older population means that it is subject to additional risk factors or competing causes. This would reduce the attributable proportion of risk for PPA and, thus, the probability that PPA caused an older plaintiff's hemorrhagic stroke. Given the magnitude of the effect (OR = 16.58 or APR = 93.9%), it is unlikely that this effect would make a difference, but this is a good

exercise for students to work through.

2. What external-validity issue would exist for the U.S. population based on a study of the effects of Agent Orange on the military during the Vietnam War?

   There are two issues here. One is that the population studied would be male, raising a similar question to the one in the PPA case. The second is that military personnel are healthier than the general population, such that any risk found in the healthier population might be greater in the less healthy one.

Section V.D. Heterogeneity: The Effect on Employing the Average Outcome

The three figures in the module illustrate the point explained at the conclusion of this section about heterogeneity in the study population. Figure V-1 illustrates perfect homogeneity—which is often assumed—while figure V-3 reflects considerable heterogeneity.

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

The Racette excerpt is included because the study outcome appears to support an inference of an acceleration effect due to exposure to the smoke from burning welding rods to which welders are exposed. But this abstract also provides fodder for students to exercise critical reading and thinking skills that they have developed in studying this module.

First, the abstract states that the "authors performed a case-control study." Recall that a case-control study employs a case group of those with the disease under investigation and a control group without the disease. In this study, the three cohorts all had parkinsonism, so a case-control study this was not. Nevertheless, the study design may have been appropriate to compare symptomology between welding-associated parkinsonism and idiopathic parkinsonism.

Second, note that this study tells us nothing about whether welding is a risk factor for, or even associated with, parkinsonism. (To be fair to the authors, their statement of what they were investigating did not identify such a purpose.) Because all three cohorts have the disease, we can't look at the comparative incidence of disease in those exposed and those not exposed (a cohort study), which would be the best way to investigate such a relationship. And, as explained above, we can't look at the comparative incidence of exposure among a disease and non-disease group because all study subjects have the disease.

Third, the welders' age at onset is considerably younger than others who developed the disease due to other causes and thus one might be inclined to conclude that welding fumes accelerate the onset of parkinsonism. Indeed, the authors wrote "we speculate that the younger age at onset may be attributable to the effects of an accelerating agent from welding in a potentially "at-risk" patient who might otherwise develop PD with a later age at onset." But

the difficulty with that assessment is that we don't know about differences that exist between the welder cohort and the other cohorts (one non-welding cohort was matched for age and gender) that may affect when parkinsonism develops. In a critique of the conclusions of studies that compare age-at-onset, Ken Rothman, a well-regarded epidemiologist, wrote:

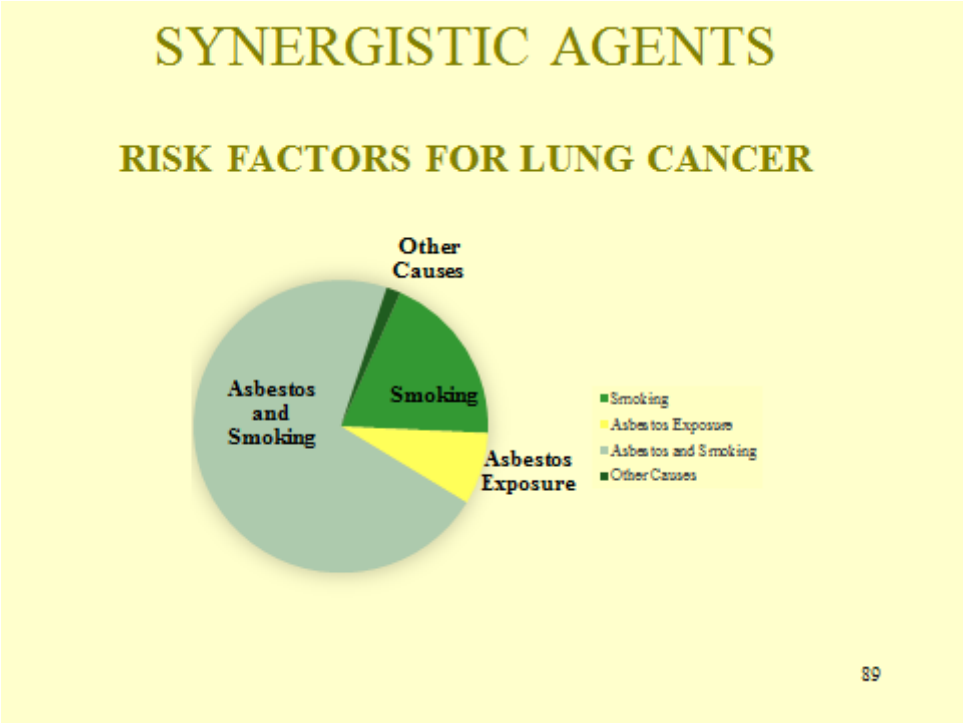
> Suppose we study two groups of people and look at the average age at death among those who die. In group A, the average age of death is 4 years; in group B, it is 28 years. Can we say that being a member of group A is riskier than being a member of group B? We cannot. * * *. Suppose that group A comprises nursery school students and group B comprises military commandos. It would be no surprise that the average age at death of people who are currently military commandos is 28 years or that the average age of people who are currently nursery students is 4 years. * * *[16]

Thus, if one non-welding cohort included those past retirement age and if development of parkinsonism is age-related, then we would expect to find a younger age at onset among the younger welding group than the older non-exposed group.

Slide 89 illustrates interaction with a pie chart that depicts the attributable proportion of risk for lung cancer. As chart reveals, the overwhelming proportion of lung cancer in a population exposed to both asbestos and smoking is caused by the combination of the two. Thus, if mechanism evidence does not enable distinguishing lung cancers caused by one or the other, the overwhelming probability is that the cancer was cause by both together.

[16] Kenneth J. Rothman, *Introduction to Epidemiologic Thinking*, in EPIDEMIOLOGY: AN INTRODUCTION 5 (Oxford K. Rothman ed., 2d ed. 2012).

Slide 89 Synergistic Agents: Risk Factors for Lung Cancer



Study Questions:

1. If acceleration does appear to occur, how should the framing of the causal inquiry be changed?

   Students will likely struggle with this question, but it provides a good opportunity to return to the basic matter of causal framing, discussed in Section II.B., supra. Rather than an effect of causing the disease, the effect would be subjecting the victim to the disease for some period of time, reflecting the amount by which the disease was accelerated.[17] Death, for example, is always an outcome that is accelerated as we know that all eventually die. Thus, the causal question with all mortality is how many years was the victim's life shortened.

[17] A defendant who only accelerates the occurrence of harm, say, chronic back pain, that would have occurred independently to the plaintiff at a later time is not liable for the same amount of damages as a defendant who causes a lifetime of chronic back pain. *See* David A. Fischer, "Successive Causes and the Enigma of Duplicated Harm," 66 *Tenn. L. Rev.* 1127, 1127 (1999); Michael D. Green, "The Intersection of Factual Causation and Damages," 55 *DePaul L. Rev.* 671 (2006).

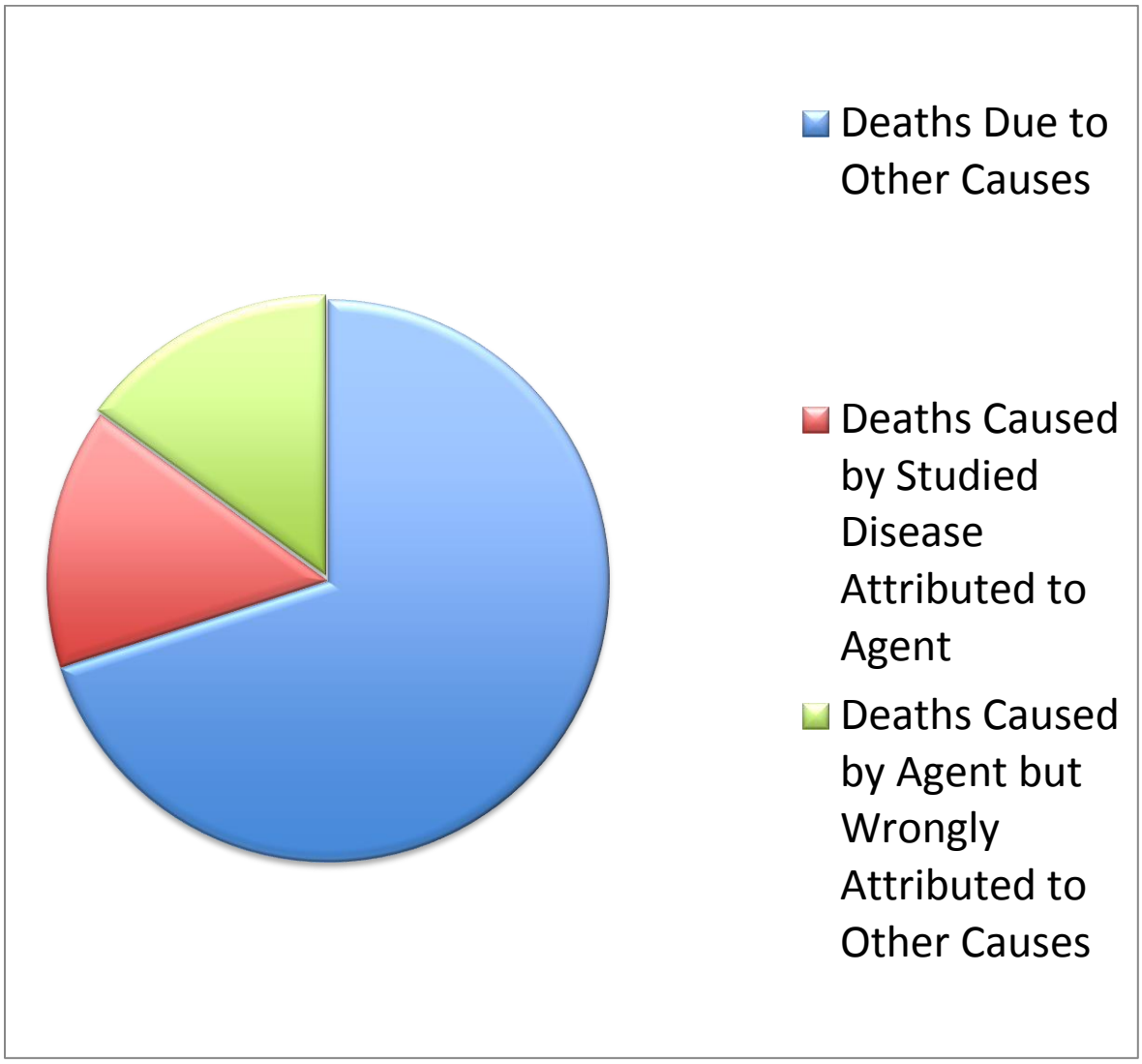
2. Explain the reasoning behind the explanation of the effect of the last two assumptions identified in this section: (a) no other fatal diseases; and (b) no protective effect for a subpopulation.

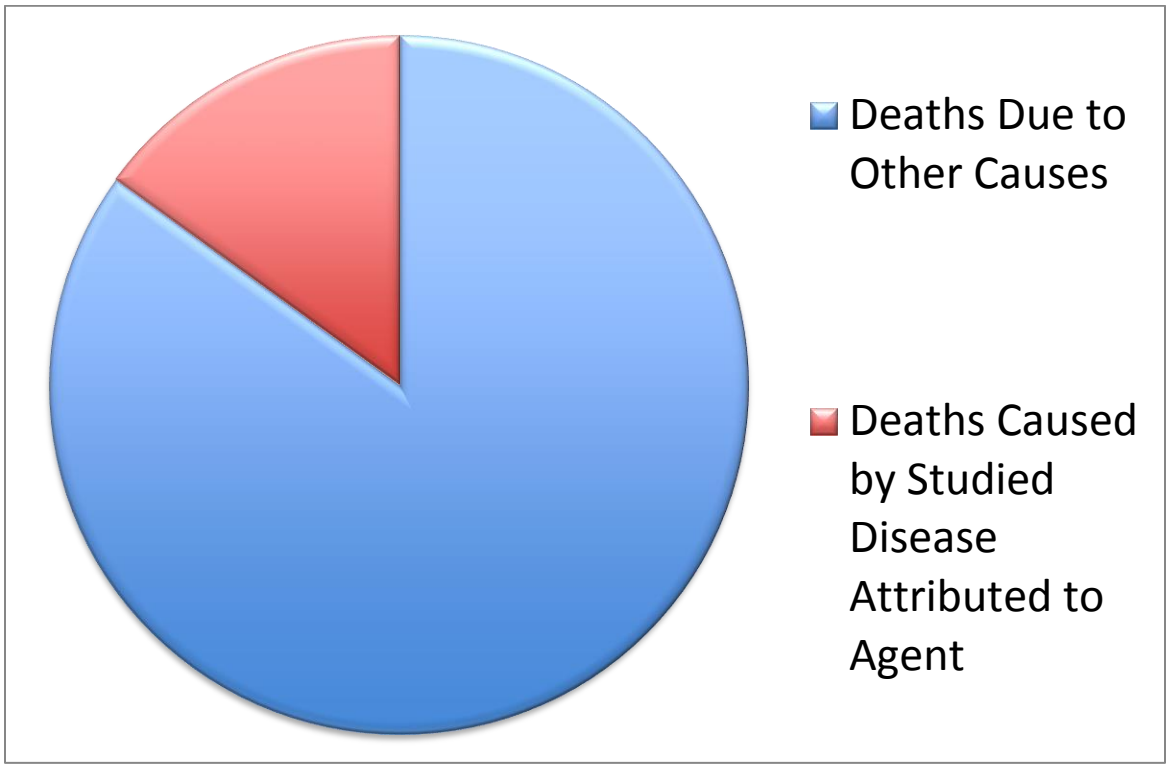
    a. If the agent causes other fatal diseases, then the sum total of its increased risk of death are the deaths due to the disease under study and the deaths due to the other disease. But because only deaths due to the studied disease are attributed to the agent, those other diseases are, in essence, attributed to some other cause. The three pie charts below illustrate:

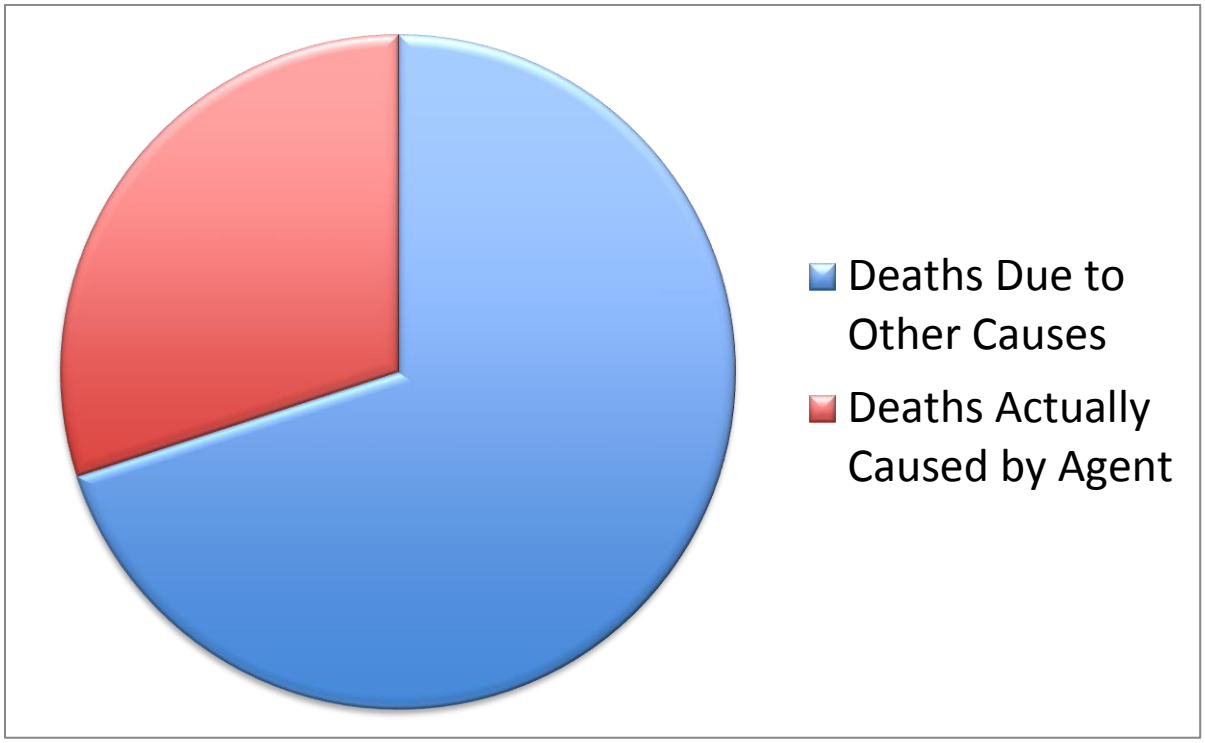
    The first chart shows 15% of deaths attributed to the studied agent (in red) but another 15% of deaths not attributed to the studied agent (in green) because those deaths occurred through a different and non-studied disease.

    The second chart reveals the effect of misclassifying deaths due to the non-studied disease as due to other causes. All of the 15% of deaths due to the non-studied disease are classified as due to other causes.

    The final chart reveals the proper proportion of mortality due to the agent being studied.


Deaths Due to Other Causes
Deaths Caused by Studied Disease Attributed to Agent
Deaths Caused by Agent but Wrongly Attributed to Other Causes


Deaths Due to Other Causes
Deaths Caused by Studied Disease Attributed to Agent


Deaths Due to Other Causes
Deaths Actually Caused by Agent

b. If the agent has a protective effect for a sub-population, the outcome of the study will be the sum of the increased risk in the susceptible population less the decreased risk in the protected subgroup. This will understate the increased risk for the susceptible population because of the subtraction of disease in the protected population.

Section V.F. Refining the Probability of Causation through Subgrouping, Including Genetic Subgrouping and Molecular Epidemiology

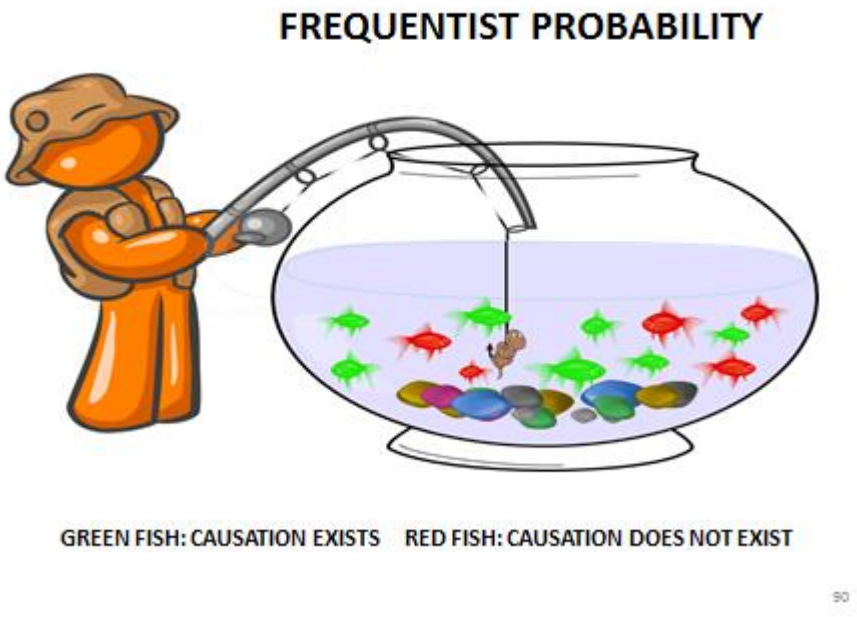
By this point, students should understand the vexatious disconnect between tort law's insistence on causal inference about individual cases of disease and the inherently group-based data produced by epidemiologic study. Many courts bridge this gap by using doubling of relative risk (a group-based property) as a proxy for more-likely-than-not specific causation (a property of an individual case). This section explores whether information about an individual plaintiff can affect the inference from group-based risk to specific causation.

The introductory paragraph drives home the difference between probability as a measure of the frequency of an event in a population and probability as a measure of the likelihood that a factual proposition is true. If we accept that any particular case of disease either was or was not caused by a particular toxic exposure, then the "true"—but unknowable—probability of causation in each case is either 0.00 or 1.00. The law tells jurors that they must decide which it is, but that they need not be certain—they must only decide which way the evidence preponderates, that is, which conclusion is more likely than not. This requires the jury to apply a subjectivist concept of probability. Relative risk, by contrast, provides a frequentist probability—if the relative risk is greater than two, then it is more likely than not that a randomly selected person who has been exposed and is sick is a case of "true" causation. This distinction is at the heart of courts' early, and to some extent continuing, mistrust of group-based evidence as support for causal inference. It has also been the subject of much scholarly commentary. For two examples, see Glen O. Robinson, "Multiple Causation in Tort Law: Reflections on the DES Cases," 68 *Va. L. Rev.* 713 (1982); and Vern R. Walker, "Preponderance, Probability and Warranted Factfinding," 62 *Brook. L. Rev.* 1075 (1996).

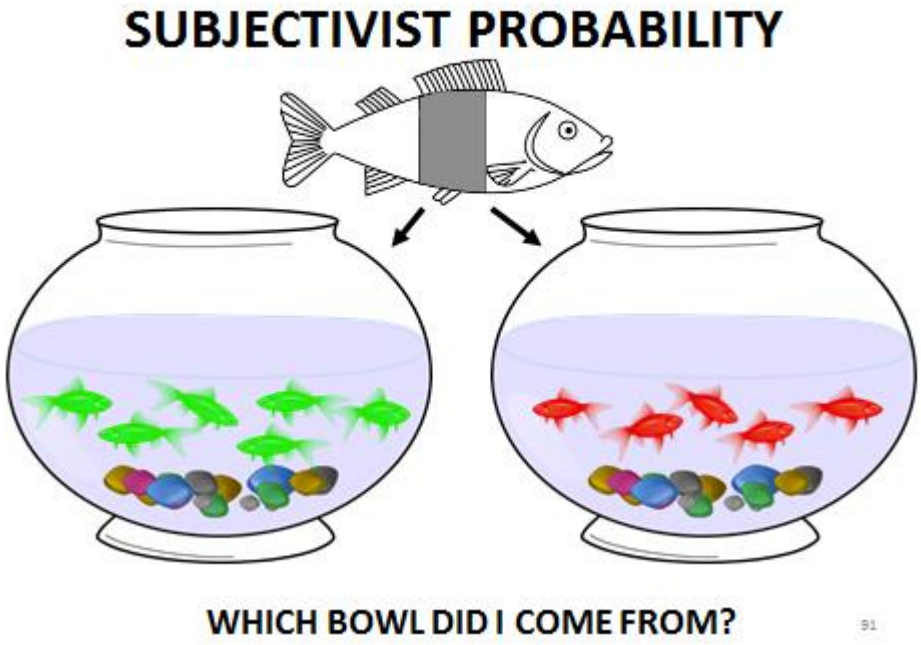
Slides 90 and 91 illustrate the difference between frequentist and subjectivist views of probability, respectively. Slide 90 shows a person fishing out of a bowl with red and green fish. When a fish takes the bait, what is the probability that a green fish will go in the creel? Unless there is reason to believe that fish of one color or the other are more likely to get hooked, the answer can be readily computed from the proportion of fish of each color. In our particular example, with seven green fish representing "true" causation cases and five red fish representing "false" causation cases, any randomly selected case is more likely than not a green fish or "true" case. In Slide 91, by contrast, we have a colorless fish that came out of either a

bowl full of green fish or a bowl full of red fish, which again represent “true” and “false” causation cases, respectively. To determine which bowl the fish came from would require a technique that revealed its color. Failing that, a factfinder might use various clues to infer that the fish more likely than not came from the bowl full of green “true” cases.

Slide 90 Frequentist Probability



Slide 91 Subjectivist Probability



The introductory paragraph concludes with two examples of the effect of all-or-nothing recovery using doubling of relative risk as the dividing line between proof and non-proof of causation. That rule would lead to these results: if all exposed sick people sued when the relative risk was 3.0, all would recover even though only two-thirds were cases of "true" causation; conversely, if all exposed sick people sued when the relative risk was 1.8, none would recover even though 44% were cases of "true" causation.

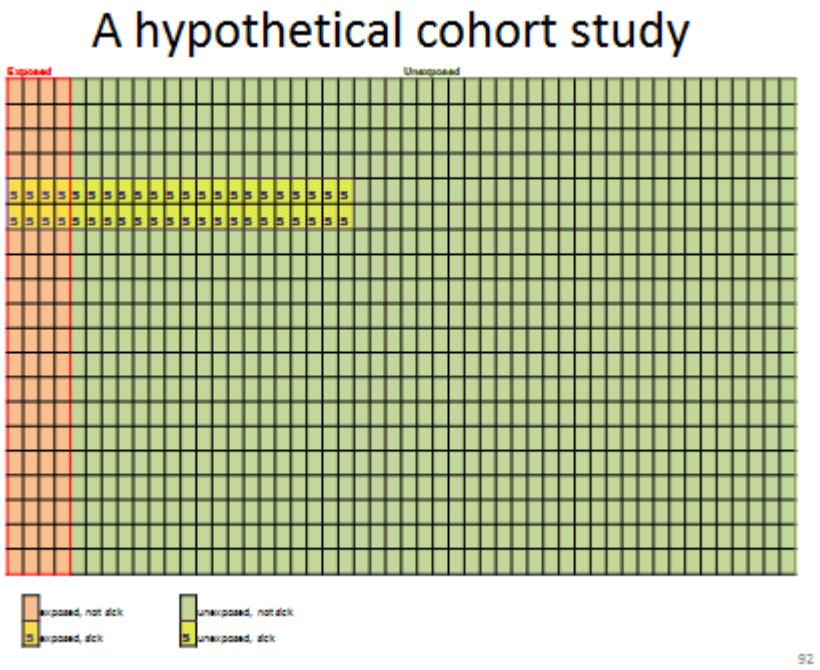
Section V.F.1. The Role of Information about the Individual Plaintiff

We begin with the intuitive concept that from a subjectivist viewpoint, obtaining particularistic information about an individual plaintiff can make it more or less likely that toxic exposure caused the plaintiff's disease—in other words, that information about a particular plaintiff can be legally relevant. The *Stubbs* illustration makes this clear: one would be more confident in our inference that Mr. Stubbs contracted typhoid by drinking contaminated water if one knew that he had not encountered many of the other possible sources of typhoid bacteria. The *Best* hypothetical in the Teaching Materials at the end (Section III.A.) of the toxicology section presents this situation. The instructor could just as well use that hypothetical at the beginning of this subsection.

We then apply the same idea to a frequentist probability derived from group-based studies. The key concept is that relative risk measures the *average* difference in risk between the exposed group and the unexposed group in a cohort study design, but does not mean that exposure increases every person's risk by that average amount. Slides 92 through 95 demonstrate this in hypothetical-epidemiologic cohort studies. (Note that these slides, which are for illustrative purposes only, do not consider whether any increased risk reported in the hypothetical results would be statistically significant).

In Slide 92, each cell in the grid represents a study participant; the grid is 20 x 50, for a total of 1,000 cells. Of these 1,000 cells, 80 red cells represent 80 individuals exposed to the toxic substance at issue; 920 green cells represent 920 unexposed individuals. Yellow cells with a bold purple "S" represent 44 individuals sick with the disease in question: 8 in the exposed group and 36 in the unexposed group. Students should be able to calculate that the relative risk is (8/80) / (36/920) = 0.1 / 0.039 = 2.56, above the 2.0 threshold.

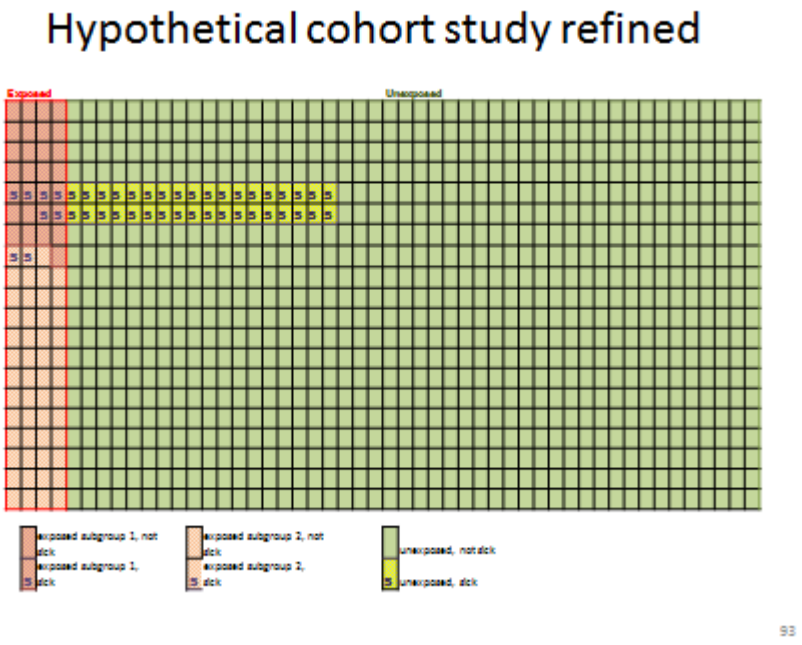
Slide 92 A Hypothetical Cohort Study



Slide 93 disaggregates the exposed group into two subgroups based on some unknown characteristic that alters the amount of risk conferred by exposure. (The remainder of this section will consider what some of those unknown characteristics might be). The overall relative risk for exposure is the same as in Slide 80. But exposed subgroup 1 (darker shading) includes 29 exposed individuals, of whom 6 are sick. Exposed subgroup 2 (lighter shading)

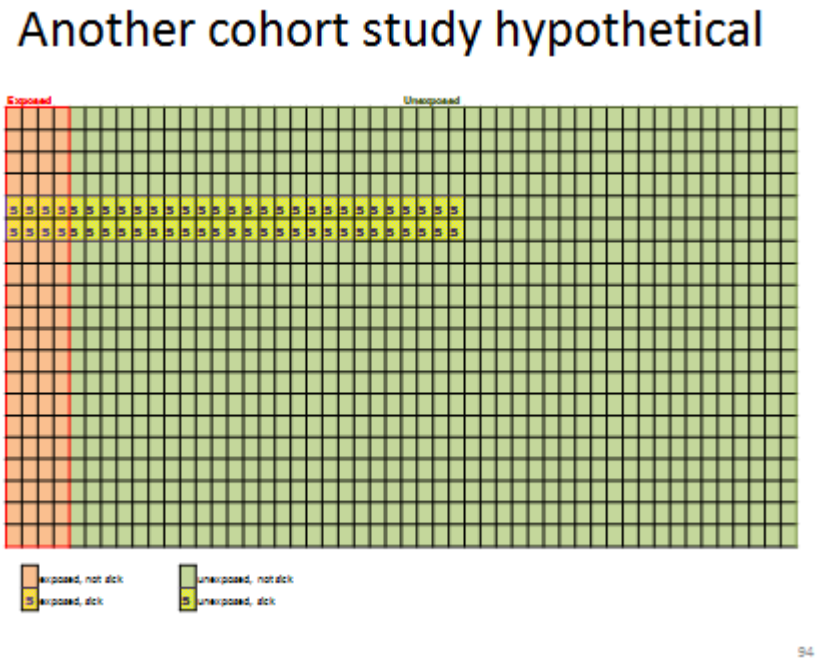
includes 51 exposed individuals, of whom 2 are sick. For subgroup 1, the relative risk is (6/29) / (36/920) = 5.29, well above the 2.0 threshold. For subgroup 2, the relative risk is (2/51) / (36/920) = 1.00, indicating no increased risk at all. Convincing evidence that a plaintiff was a member of subgroup 1 would presumably lead to a finding of no causation even though the study found that exposure, overall, more than doubled the risk.

Slide 93 Hypothetical Cohort Study Refined

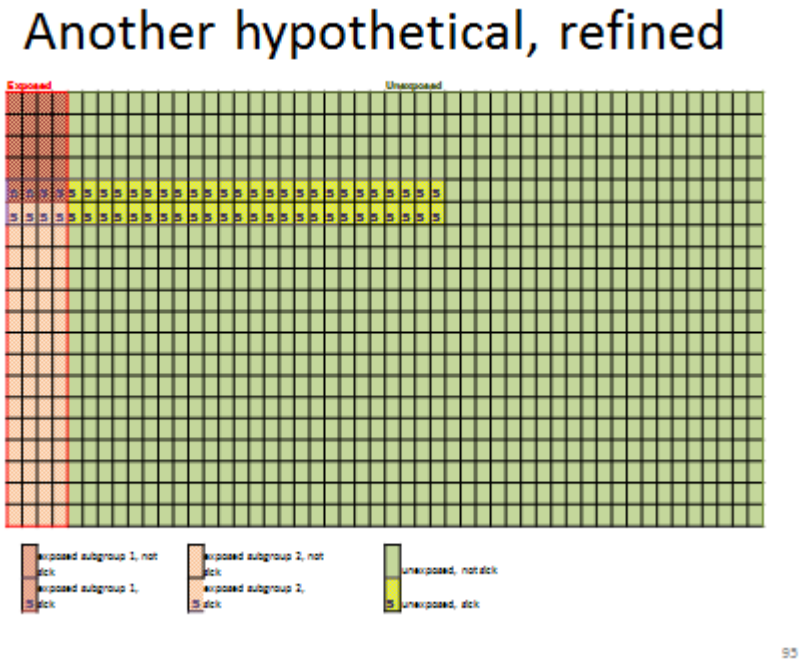


Slides 94 and 95 present a converse situation. In Slide 94, the same 8 out of 80 exposed individuals are sick, but 50 of the 920 unexposed individuals are sick. The relative risk is (8/80) / (50/920) = 1.84, below the 2.0 threshold. Slide 95 divides this hypothetical exposed group into two subgroups. Subgroup 1 (darker shading) includes 20 exposed individuals, 4 of whom are sick; subgroup 2 (lighter shading) includes 60 exposed individuals, 4 of whom are sick. For subgroup 1, the relative risk is (4/20) / (60/920) = 3.68; for subgroup 2, the relative risk is (4/60) / (60/920) = 1.23. Convincing evidence that a plaintiff was a member of subgroup 1 would presumably lead to a finding of causation, even though the study found that exposure, overall, resulted in less than a doubling of risk.

Slide 94 Another Cohort Study Hypothetical



Slide 95 Another Hypothetical, Refined



Students may notice—and if they do not, the instructor should prompt them to consider—that in Slide 93 and Slide 95, only the exposed group, and not the group of unexposed subjects, has been subdivided. A well-designed study would measure the risk-modifying characteristic in both groups and compute relative risks that took into account both exposure and the other characteristic that modified the risk of exposure. But of course, if the epidemiologic data reported separate, stratified relative risks for different subgroups of exposed and unexposed people and the plaintiff could confidently be placed in a particular subgroup, then the parties, factfinder, and court would focus on the relative risk for that subgroup. Far more typically, however, such subgrouping is unavailable or incomplete. Epidemiologic studies more or less inevitably group together people who may vary with respect to some factors that may affect the risk that exposure creates for those individuals. Thus, as a practical matter, the focus with respect to such factors is on the individual plaintiff. In the absence of highly stratified epidemiologic data, to what extent may a plaintiff (or defendant) argue that the plaintiff's unique circumstances made exposure more (or less) harmful to that plaintiff than to the average subject in the relevant epidemiologic studies? The Vermont Supreme Court's opinion in *Estate of George v. Vermont League of Cities and Towns* explores this question.

Estate of George v. Vermont League of Cities and Towns, 993 A.2d 367 (2010):

Estate of George involved a claim for workers' compensation by the estate of a firefighter who died of non-Hodgkin's lymphoma (NHL). Thus, the precise legal question was whether George's NHL was "a personal injury arising out of and in the course of employment" under Vt. Stat. Ann. tit. 21, § 618(a)(1) (2015). Students will readily perceive, however, that the case fits the classic toxic-tort mold: the claimant does not argue that being a firefighter causes NHL, but that exposure to a potpourri of toxic gases while fighting fires does.

Estate of George provides a good opportunity for students to consolidate their understanding of the "more than doubling of risk" threshold and to subject that threshold to critical analysis. The Vermont Supreme Court's approval of the trial court's use of doubling of risk as a "benchmark" is critical to its approval of the trial court's decision to exclude the claimant's causation experts and thus the affirmance of the trial court's grant of summary judgment against the claimant.

One might start by asking what evidence the claimant attempted to introduce to prove the allegation that the claimant's employment as a firefighter caused his NHL. The answer, of course, is expert-opinion testimony based entirely on epidemiologic studies. The majority mentions eight studies on which the claimant's experts relied. (As both the majority and dissenting opinions noted, these studies were cited in a wide-ranging meta-analysis of the epidemiology of cancer in firefighters published by the claimant's expert witness Dr. LeMasters and coauthors. See *Estate of George*, 993 A.2d at 373, 384 (Reiber, C.J., dissenting). The meta-analysis, Grace K. LeMasters et al., "Cancer Risk Among Firefighters: A Review and Meta-analysis of 32 Studies," 48 J. *Occupational & Envt'l Med.* 1189 (2006), is available at

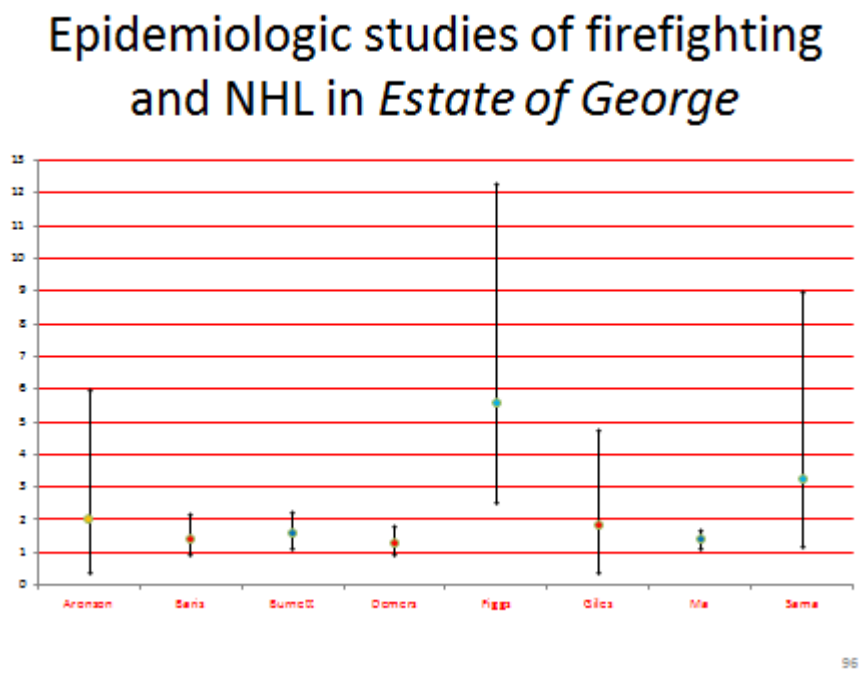
https://www.iaff.org/hs/PDF/Cancer%20Risk%20Among%20Firefighters%20-%20UC%20Study.pdf. The studies used different endpoints for their risk comparisons and a variety of terminology for the disease being studied, but these differences are not consequential to the court's analysis.

The results of these eight studies (only some of which can be found in the court opinion) are summarized in Slide 96.[18] The slide graphically shows the problem the claimant faced. The eight studies are listed in alphabetical order along the *x* axis, with relative risk ratios on the *y* axis. The colored point for each study shows the study's reported estimate of the risk ratio of firefighting for NHL. The vertical bar extending from each point is the 95% confidence interval of the risk estimate. If the lower end of the bar is less than 1, then the increased risk the study found was not statistically significant at the 95% level.

[18] The authors drew this chart based on reviewing each of the named studies. They did not rely on results reported in a review paper. Sources of the point estimates and 95% CI's are as follows:

Aronson, K.J. et al, "Mortality Among Fire Fighters in Metropolitan Toronto," 26 *Am. J. Industrial Med*. 89-101 (1994), p. 93 Table IV; Baris, D et al, "Cohort Mortality Study of Philadelphia Firefighters," 39 A*m. J. Industrial Med.* 463-476 (2001), p. 467 Table II; Burnett, CA et al, "Mortality Among Fire Fighters: A 27 State Survey," 26 *Am. J. Industrial Med.* 831-833 (1994), p. 832 Table I; Demers, PA et al, "Mortality among firefighters from three northwestern United States cities," 49 *Brit. J. Industrial Med.* 664-670 (1992), p. 666 Table 2; Figgs, LW et al, United States Non-Hodgkin's Lymphoma Surveillance by Occupation 1984-1989: A Twenty-Four State "Death Certificate Study," 27 *Am. J. Industrial Med*. 817-835 (1995), p. 822 Table II; Giles, GG & Staples, M, "Cancer Incidence in Melbourne Metropolitan Fire Brigade Members," 1980-1989, 5 *Health Reports* 1993 33-38 (1993), p. 35 Table 1 (available from Statistics Canada, https://www.researchgate.net/publication/14865536_Cancer_incidence_in_Melbourne_Metropolitan_Fire_Brigade_members_1980-1989); Ma,F et al, "Race-specific Cancer Mortality in US Firefighters: 1984 to 1993," 40 *J. Occupational & Envtl. Med.* 1134-1138 (2005), Abstract; and Sama, SR et al, "Cancer Incidence Among Massachusetts Firefighters," 18 *Am. J. Industrial Med.* 47-54 (1990), p. 47 Abstract.

Slide 96 Epidemiologic Studies of Firefighting and NHL in Estate of George



All eight studies found some increased risk of NHL for firefighters (indicated by point estimates greater than 1). But four of those estimates were not statistically significant: Baris, Demers, and Giles, which reported less-than-doubled risk that was not statistically significant (red points); and Aronson, which reported a (barely) more-than-doubled risk that was not statistically significant (orange point). Of the four studies that found statistically significant risk increases, two (Burnett and Ma, dark blue points) reported less-than-doubled risk. Only two studies (Figgs and Sama, bright blue points) reported increased risks that were both more-than-doubled and statistically significant.

The instructor might ask whether the trial court judgment is justified simply by the numerical predominance of studies that failed to find a statistically significant more-than-doubled risk. Students should recall the technique of meta-analysis, Section III.G., supra, for pooling results from multiple studies—and also the limitations of that technique. By increasing the number of data points, meta-analysis can increase the statistical power of the analysis and reduce the likelihood of a Type II (false negative) error.

Certainly, the count of studies seemed to impress the majority of the court. The majority opinion treated dismissively Dr. Guidotti's attempt to assess qualitatively what the studies meant when applied to someone with as long as a firefighting career as the claimant. But it is not easy to discern the precise basis on which the court upheld the exclusion of Dr. Guidotti's opinion. Did the court find Dr. Guidotti's qualitative reasoning unreliable, or only his attempt to link to the quantitative doubling of risk threshold? The court presented its reasoning

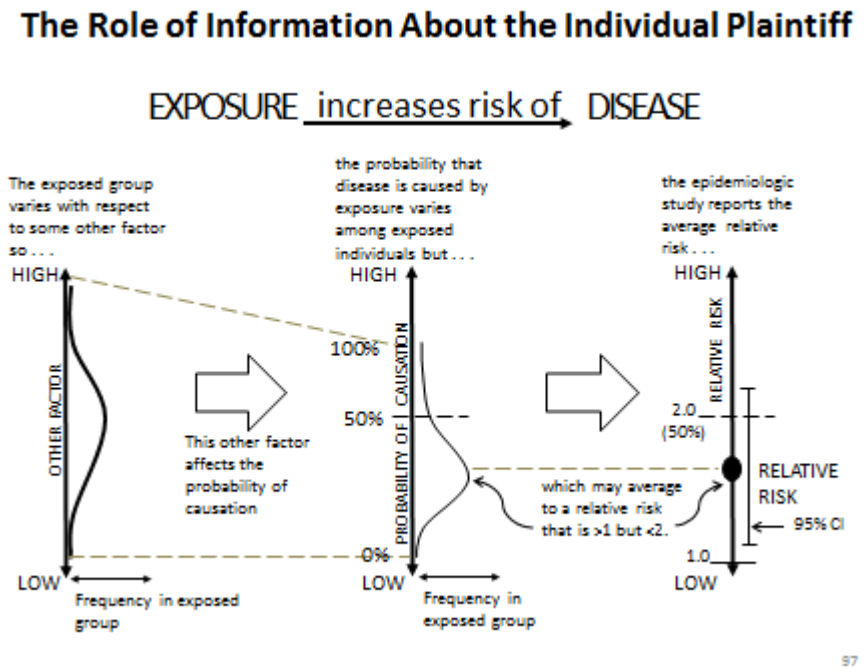
more clearly when it discussed Dr. LeMaster's quantitative meta-analysis: the meta-analysis reported a "summary risk estimate for NHL [of] 1.51, again a value less than 2.0." The dissent, however, criticized the trial court and the majority for seeming to insist that every epidemiologic study (almost every study? a majority of studies?) find a relative risk greater than 2.0.

In light of the majority's conclusions, can its description of doubling of risk as a "helpful benchmark" be distinguished from treating more-than-doubling of risk as a bright-line threshold? See, e.g.*, Merck & Co. v. Garza*, 347 S.W.3d 256, 265 (Tex. 2011) (stating that when a plaintiff seeks to use epidemiologic evidence to prove causation, doubling of the risk is a "threshold requirement" for admissible expert testimony). Even the dissenting opinion in *Estate of George* accepted risk doubling as an appropriate "standard."

The dissent contended, though, that the trial court should have admitted the claimant's experts' testimony because (1) it was based, in part, on two epidemiologic studies that satisfied the risk-doubling standard and (2) even if the epidemiologic studies did not prove more than a doubling of risk in general, the claimant's experts relied on "specific knowledge about claimant" to fill the "analytical gap" and permit a factual finding that the claimant's NHL more likely than not was caused by his firefighting occupation.

The latter argument condenses several logical steps that the instructor should generalize and explore with students. Step 1: The epidemiologic evidence shows some causal association between the alleged exposure and the disease. Step 2: Some other factor affects the degree of risk conferred by a toxic exposure in a way that is not accounted for by the point-risk estimates of the epidemiologic evidence. Step 3: The other factor identified in Step 2 affects the plaintiff or claimant in a way that would increase the plaintiff's risk conferred by exposure, as compared to the average risk increase identified by the epidemiologic evidence. Step 4: There is some reason to infer that the plaintiff's or claimant's risk is elevated enough to exceed the "more likely than not" threshold. Slides 97 and 98 present this reasoning in a conceptual and generic way. In the PowerPoint presentations, these slides have been animated so the instructor can click through them in a step-by-step way.

Slide 97 The Role of Information about the Individual Plaintiff



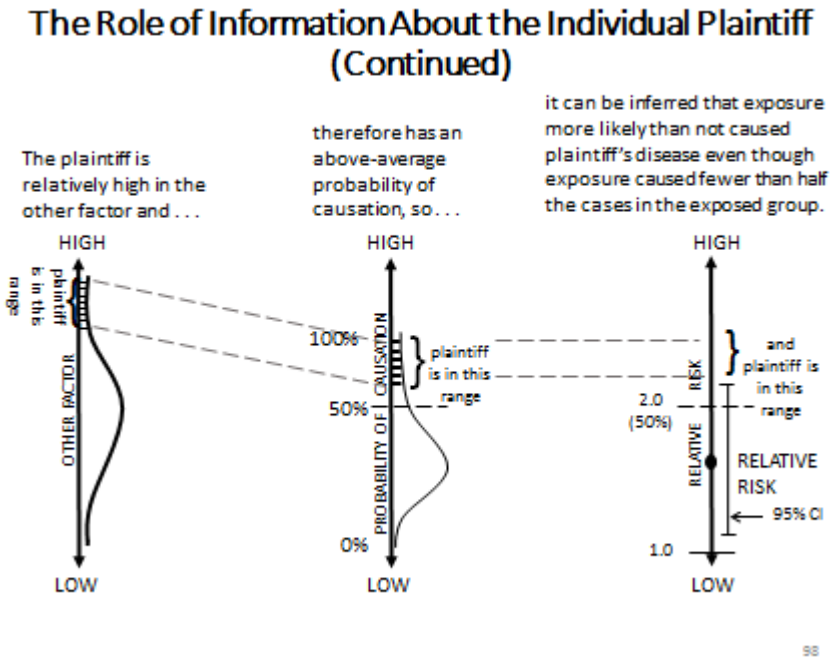
Step 1 appears across the top of Slide 97: the toxic exposure in question increases the risk of a plaintiff's disease. Below Step 1, a series of graphs illustrate Step 2.

The left graph (on click) is a frequency distribution (rotated 90 degrees) for the "other factor" that influences the extent to which the toxic exposure increases an individual study subject's risk. Along the axis, the values of the other factor range from low (at the bottom) to high (at the top). The curve shows the number of study subjects who experienced each level of the other factor. The curve is a normal distribution, coinciding with the intuition that in most circumstances, individual values of the other factor will cluster near the middle and extreme high and low values will be relatively rare.

The center graph (on click) links the range of values of the other factor to a range of individual probabilities that the toxic exposure caused the plaintiff's disease. This axis also is labeled "low" to "high," but because this axis represents a probability, the lowest possible value is 0 (if there is some value of the other factor such that the toxic exposure does not increase disease risk—a threshold effect) and the highest possible value is 1 (if there is some value of the other factor such that the toxic exposure causes every case of disease—a signature effect). Again, most individuals are clustered in the center of the frequency distribution. Note that the center of the distribution is below the 50% probability mark: this will correspond with an overall relative risk less than 2.0. But a part of the curve, for those with high values of the "other factor," is above the 50% mark.

The right graph (on click) portrays the results of the epidemiologic study of the group depicted in the center graph: the range of individual probabilities becomes a point estimate of relative risk that is greater than 1.0 but less than 2.0, corresponding with the peak of the frequency distribution in the center graph. The point estimate of relative risk has a 95% confidence interval marked around it. Note that the 95% confidence interval (which is a statistical function of the strength of the association and the number of people in the study, as described in the epidemiology section, Section III.E.1., supra) is *not* the same as the range of individual risks portrayed in the center graph (which could not be derived directly from the epidemiologic study).

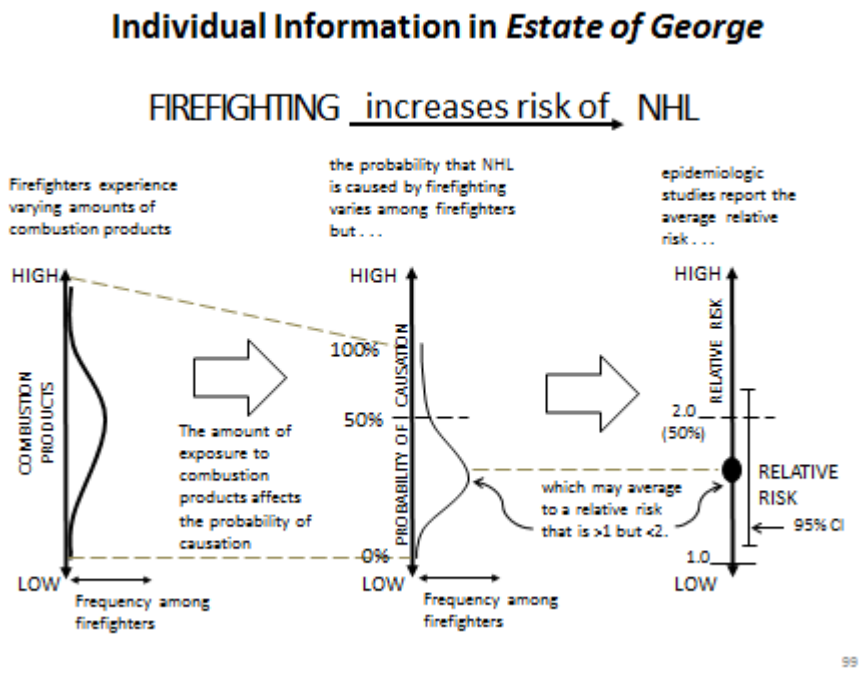
Slide 98 The Role of Information about the Individual Plaintiff (continued)



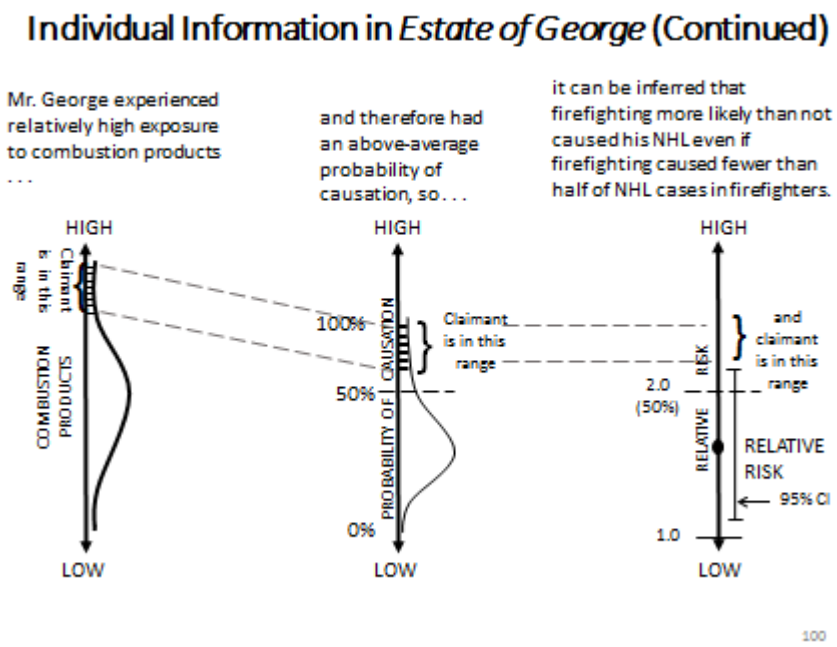
Slide 98 presents Steps 3 and 4, which position the individual plaintiff on the graphs presented in Step 2. On the left graph (on click), the plaintiff's value for the "other factor" is presented as falling within a range near the high end of the distribution. The center graph (on click) shows this range near the high end of the distribution for probability of causation, above the 50% mark. The right graph (on click) shows Step 4, the plaintiff's ultimate argument: although the average relative risk is below 2.0, equivalent to less than a 50% probability of causation, the toxic exposure more than doubled the plaintiff's individual risk of disease, making it more likely than not that the toxic exposure caused plaintiff's illness.

Slides 99 and 100 apply the generic conceptual approach of Slides 97 and 98 to the argument made in *Estate of George*. (An instructor concerned that Slides 85 and 86 are too abstract could proceed directly to Slides 99 and 100.) As the slides show, the argument progressed as follows. Step 1: Working as a firefighter increases the risk of NHL. Thus, a firefighting occupation is the "exposure" studied by the epidemiologic studies. Step 2: However, firefighters actually experience a range of on-the-job amounts of the toxic combustion products that cause NHL. These combustion products are the "other factor" that an epidemiologic study of occupation cannot measure. The greater the amount of these compounds inhaled, ingested, or absorbed by a firefighter, the greater the degree to which firefighting increases that firefighter's risk of NHL. Step 3: Compared to other firefighters, Mr. George experienced relatively high amounts of carcinogenic combustion products. Step 4: From Mr. George's high levels of exposure and other known facts about him, one can infer that it is more likely than not that his NHL was job-related, even if the firefighting occupation does not double the average firefighter's risk of NHL. The *Estate of George* dissent concluded, explicitly or implicitly, that the claimant's experts' opinions included a reliable basis for each of these steps and therefore should have been admitted. The instructor may wish to explore these steps, and the scientific evidence available to support them, in more detail.

Slide 99 Individual Information in *Estate of George*



Slide 100 Individual Information in *Estate of George* (continued)



For Step 1, the dissent noted the four studies that showed that firefighters faced statistically significant (even if the risk in two studies less than doubled) increased risk of NHL. To reinforce this point, the dissent referred to Dr. Guidotti's testimony that the claimant's particular subtype of NHL (small-cell lymphoma) had been associated with environmental exposures "including some of the same chemicals that are 'released during firefighting.'"

For Step 2, the dissent identified variations in the actual degree of toxic exposure as the primary "other factor" that might cause an individual firefighter to experience a higher relative risk than the average point estimate of relative risk reported in an epidemiologic study. None of the studies (most of which were based on death certificates) had any way to measure the study subjects' actual consumption of toxic constituents while fighting fires, which might depend on (for example) the length of their firefighting careers, the number of fires they responded to, the nature of the fires they fought, their roles at those fires, the quality of any protective breathing equipment provided to them, and whether they used such equipment. The claimant's experts assumed a type of dose-response relationship—that greater exposure to toxins during firefighting meant a larger increase in risk. The dissent's adoption of this assumption was mostly implicit, in the Step 3 discussion of the testimony about why Mr. George's exposure was higher than average.

How strongly does the scientific evidence support the implicitly assumed dose-response relationship? This question explains the emphasis that both the majority and the dissent put on the Baris study. Here is the study's abstract:

Cohort Mortality Study of Philadelphia Firefighters

Dalsu Baris, MD, PhD,[1]* Thomas J. Garrity,[2] Joel Leon Telles, PhD,[3] Ellen F. Heineman, PhD,[1] Andrew Olshan, PhD,[4] and Shelia Hoar Zahm, ScD[1]

Background *Fire fighters are exposed to a wide variety of toxic chemicals. Previous studies have reported excess risk of some cancers but have been limited by small numbers or little information on employment characteristics.*

Methods *We conducted a retrospective cohort mortality study among 7,789 Philadelphia firefighters employed between 1925 and 1986. For each cause of death, the standardized mortality ratios (SMRs) and 95% confidence intervals were estimated. We also compared mortality among groups of firefighters defined by the estimated number of career runs and potential for diesel exposure.*

Results *In comparison with U.S. white men, the firefighters had similar mortality from all causes of death combined (SMR = 0.96) and all cancers (SMR = 1.10). There were statistically significant deficits of deaths from nervous system diseases (SMR = 0.47), cerebrovascular diseases (SMR = 0.83), respiratory diseases (SMR = 0.67), genitourinary diseases (SMR = 0.54), all accidents (SMR = 0.72), and suicide (SMR = 0.66). Statistically significant excess risks were observed for colon cancer (SMR = 1.51) and ischemic heart disease (SMR = 1.09). The risks of mortality from colon cancer (SMR = 1.68), kidney cancer (SMR = 2.20), non-Hodgkin's lymphoma (SMR = 1.72), multiple myeloma (SMR = 2.31), and benign neoplasms (SMR = 2.54) were increased among firefighters with at least 20 years of service.*

Conclusions *Our study found no significant increase in overall mortality among Philadelphia firefighters. However, we observed increased mortality for cancers of the colon and kidney, non-Hodgkin's lymphoma and multiple myeloma. There was insufficient follow-up since the introduction of diesel equipment to adequately assess risk.* Am. J. Ind. Med. 39:463–476, 2001. *Published 2001 Wiley-Liss, Inc.*†

KEY WORDS: firefighters; mortality; cohort study

Abstract from Dalsu Baris et al., "Cohort Mortality Study of Philadelphia Firefighters," 39 *Am. J. Indus. Med.* 463, 463 (2001). The article is a U.S. Government work and, as such, is in the public domain in the United States of America.

The Baris study used at least two proxies for degree of exposure: the number of firefighting runs and the number of years spent as a firefighter. For each proxy, the researchers divided the firefighters into three groups: a low, medium, and high number of runs (Table VI) and a short, medium, and high duration of employment (Table III). These tables are reproduced in Slides 101 and 102. The majority focused on Table VI (Slide 102), showing that the relative risk for NHL was highest for firefighters in the group with the smallest number of runs and was lowest for the group with the largest number of runs. (In fact for the "high" group, the relative risk was below 1.0, which—if the association were causal—would suggest that going on more firefighting runs has a protective effect against NHL risk.) The dissent called this finding "strange" and argued that plaintiff's experts gave good reasons for disregarding it and for focusing instead on the Table III (Slide 101), finding that firefighters with more than 20 years of service had the highest relative risk of NHL. But questions arise even from Table III. Firefighters with 10 to 19 years of service had a lower relative risk than firefighters with either shorter *or* longer service.

Slide 101 Part of Table III from Baris Study

Part of Table III from Baris study.

TABLE III. Observed (Obs) Deaths, Standardized Mortality Ratios (SMR), and 95% Confidence Intervals (CI) Among Philadelphia Firefighters by Duration of Employment as a Firefighter (1925–1986) (N = 7,789)

	Duration of employment								
	≤ 9 years			10–19 years			≥ 20 years		
Cause of death (ICD-9)	Obs	SMR	95% CI	Obs	SMR	95% CI	Obs	SMR	95% CI
All causes (001–999)	586	**0.88**	0.81–0.95	862	**1.07**	1.00–1.14	772	**0.91**	0.85–0.98
All cancers (140–209)	143	**1.26**	1.07–1.49	170	**1.10**	0.94–1.27	187	**0.99**	0.86–1.15
Buccal cavity and pharynx (140–149)	4	**1.15**	0.43–3.07	9	**1.83**	0.95–3.51	6	**1.09**	0.50–2.43
Esophagus (150)	0	—	—	3	**0.82**	0.26–2.52	3	**0.65**	0.21–2.02
Stomach (151)	4	**0.55**	0.21–1.48	14	**1.39**	0.83–2.35	6	**0.65**	0.29–1.44
Colon (153)	18	**1.78**	1.12–2.82	16	**1.11**	0.68–1.81	30	**1.68**	1.17–2.40
Rectum (154)	3	**0.86**	0.28–2.66	6	**1.16**	0.52–2.58	5	**0.92**	0.38–2.22
Liver (155–156)	1	**0.40**	0.06–2.83	3	**0.84**	0.27–2.60	4	**1.09**	0.41–2.91
Pancreas (157)	13	**2.33**	1.36–4.02	5	**0.60**	0.25–1.45	5	**0.49**	0.21–1.19
Larynx (161)	1	**0.66**	0.09–4.59	1	**0.43**	0.06–3.05	3	**1.08**	0.35–3.36
Lung (162)	50	**1.52**	1.16–2.01	56	**1.20**	0.92–1.56	56	**0.89**	0.68–1.15
Skin (172–173)	2	**0.75**	0.19–3.01	5	**1.70**	0.71–4.09	3	**1.05**	0.34–3.26
Prostate (185)	15	**2.36**	1.42–3.91	5	**0.47**	0.19–1.12	11	**0.72**	0.40–1.31
Bladder (188)	4	**1.36**	0.51–3.61	7	**1.48**	0.70–3.09	6	**1.01**	0.45–2.25
Kidney (189)	2	**0.72**	0.18–2.87	0	—	—	10	**2.20**	1.18–4.08
Brain (191–192)	2	**0.47**	0.12–1.89	2	**0.44**	0.11–1.75	4	**0.94**	0.35–2.49
Non-Hodgkin's lymphoma (200,202)	6	**1.47**	0.66–3.26	5	**1.03**	0.43–2.47	9	**1.72**	0.90–3.31
Multiple myeloma (203)	1	**0.73**	0.10–5.17	3	**1.50**	0.48–4.66	6	**2.31**	1.04–5.16
Leukemia (204–207)	5	**0.94**	0.39–2.25	7	**1.14**	0.54–2.38	3	**0.45**	0.15–1.40

101

SOURCE: Dalsu Baris et al., "Cohort Mortality Study of Philadelphia Firefighters," 39 *Am. J. Indus. Med.* 463, 463 (2001). The article is a U.S. Government work and, as such, is in the public domain in the United States of America.

Slide 102 Part of VI from Baris Study

Part of Table VI from Baris study.

TABLE VI. Observed (Obs) Deaths, Standardized Mortality Ratios (SMR), and 95% Confidence Intervals (CI) Among Philadelphia Firefighters by Cumulative Number of Runs[a] in all Positions (1935–1986) (N = 6,477)

	Low (< 3,323 runs)			Medium (≥ 3,323 & < 5,099 runs)			High (≥ 5,099 runs)		
Cause of death (ICD–9)	**Obs**	**SMR**	**95% CI**	**Obs**	**SMR**	**95% CI**	**Obs**	**SMR**	**95% CI**
All causes (001–999)	644	**0.94**	0.88–1.02	327	**0.93**	0.84–1.04	310	**0.90**	0.81–1.00
All cancers (140–209)	155	**1.14**	0.98–1.34	89	**1.09**	0.88–1.34	96	**1.12**	0.92–1.38
Buccal cavity and pharynx (140–149)	7	**1.72**	0.82–3.61	0	—	—	2	**0.77**	0.19–3.09
Esophagus (150)	2	**0.66**	0.17–2.64	1	**0.50**	0.07–3.54	3	**1.40**	0.45–4.33
Stomach (151)	4	**0.66**	0.25–1.75	1	**0.31**	0.05–2.22	2	**0.66**	0.16–2.63
Colon (153)	23	**1.93**	1.29–2.91	16	**2.22**	1.36–3.62	9	**1.22**	0.64–2.35
Rectum (154)	5	**1.37**	0.51–3.29	1	**0.51**	0.07–3.59	1	**0.54**	0.08–3.85
Liver (155–156)	2	**0.80**	0.20–3.21	1	**0.73**	0.10–5.22	1	**0.76**	0.11–5.38
Pancreas (157)	7	**1.02**	0.48–2.13	5	**1.17**	0.49–2.80	7	**1.61**	0.77–5.74
Larynx cancer (161)	1	**0.53**	0.07–3.76	1	**0.83**	0.11–5.87	1	**0.80**	0.11–5.74
Lung (162)	47	**1.06**	0.79–1.41	30	**1.00**	0.70–1.44	38	**1.18**	0.86–1.63
Skin cancer (172–173)	1	**0.36**	0.05–2.50	5	**3.10**	1.29–7.46	1	**0.52**	0.07–3.75
Prostate (185)	10	**1.33**	0.72–2.48	3	**0.65**	0.21–2.03	6	**1.42**	0.64–3.16
Bladder (188)	4	**1.20**	0.44–3.18	1	**0.50**	0.07–3.56	1	**0.54**	0.08–3.81
Kidney (189)	4	**1.18**	0.44–3.15	4	**1.90**	0.71–5.07	2	**0.89**	0.22–3.55
Brain (191–192)	3	**0.60**	0.19–1.85	2	**0.78**	0.20–3.11	2	**0.73**	0.18–2.93
Non-Hodgkin's lymphoma (200,202)	11	**2.36**	1.31–4.26	4	**1.55**	0.58–4.13	2	**0.73**	0.18–2.94
Multiple myeloma (203)	1	**0.57**	0.08–4.06	3	**2.69**	0.87–8.35	2	**1.73**	0.43–6.90
Leukemia (204–207)	5	**0.84**	0.35–2.02	4	**1.35**	0.51–3.59	4	**1.33**	0.50–3.55

102

SOURCE: Dalsu Baris et al., "Cohort Mortality Study of Philadelphia Firefighters," 39 *Am. J. Indus. Med.* 463, 463 (2001). The article is a U.S. Government work and, as such, is in the public domain in the United States of America.

How much do the data in these tables mean? An instructor might provide the students with the study, the tables, or just the data on NHL from these tables. Students should observe that in both tables, the number of NHL cases in each subgroup is very small. The 95% confidence interval around the estimated relative risk for each subgroup is, conversely, large. None of the subgroup effects is statistically significant (even though the study's abstract listed the standardized mortality ratio for NHL for firefighters with less than 20 years of employment). In fact, the NHL relative risk for the *entire* study group was not statistically significant (SMR 1.41, 95% CI 0.91–2.19). It is noteworthy that the *Estate of George* majority treated the Baris study, on the one hand, as just one of a group of studies that failed to find a statistically significant association between firefighting and NHL overall, but on the other hand, concluded that the claimant's experts needed to "account specifically" for the relative risks the Baris study reported for the subgroups with different numbers of firefighting runs.

For Step 3, the dissent—in the absence of any actual measurements of either Mr. George's exposure or the exposure of the firefighters included in the study—relied on the claimant's experts' testimony about various proxies suggesting that Mr. George endured unusually high exposure. The instructor might ask students to identify these proxies. They are the length of Mr. George's firefighting career (40 years) and the relative lack of use of adequate protective equipment for much of that time.

For Step 4, the dissent cited the claimant's experts' overall conclusions in light of the epidemiologic studies and their knowledge of Mr. George's individual circumstances. The quantitative results of the epidemiologic studies cannot themselves demonstrate that Mr. George's above-average exposure to toxins put him in a subgroup of firefighters for which the risk of NHL more than doubled. This should not be surprising: if the studies had provided such results, there would have been no need to make the "individual characteristics" argument at all.

Notes and Questions after *Estate of George*:

1. Assess the persuasiveness of the majority and dissenting opinions.

   Both opinions have their weaknesses. The majority opinion seems to have imposed a rigid "relative risk greater than 2.0" rule, seems to have required a high level of consistency in epidemiologic results to allow a witness to testify, and was fairly dismissive of the claimant's experts' attempts to rely on factors unique to the claimant (as shown, for example, by the reliance on the Baris study, discussed above). The dissenting opinion seems to have taken little account of the amount of epidemiologic study that failed to find a statistically significant association between firefighting and NHL, readily accepted the claimant's experts' conclusion that the claimant's unique factors supported a finding of causation more likely than not, and applied a narrow view of the court's evidentiary gatekeeping role. Different students are apt to find one or the other opinion more persuasive, which should lead to lively class discussion.

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 [him]." 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 IV.G.3., infra.

   This is discussed in Study Question 3 below.

3. 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. However, 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).

This is discussed in Study Question 4 below. The *Milward* case involved a plaintiff who claimed that occupational exposure to products containing benzene caused his Acute Promyelocytic Leukemia (APL), a subtype of a class of malignancies called Acute Myeloid Leukemia (AML). In the cited district court opinion, the issue was whether the plaintiff's expert could give a general-causation opinion. Epidemiologic studies of APL specifically did not find statistically significant associations with benzene exposure, although benzene was known to cause other subtypes of AML. Citing various pieces of research, the plaintiff's expert contended that from the weight of the evidence, it could be inferred that benzene exposure could cause APL as well as other AML subtypes. The district court excluded the testimony and granted summary judgment to the defendants. The court of appeals reversed, holding that the expert's "weight of the evidence" opinion was based on a reliable methodology and should have been admitted. On remand, a different district judge excluded the plaintiff's proffered testimony on specific causation and again entered summary judgment for the defendants. *Milward*, 969 F. Supp. 2d at 101.

4. 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 that the expert considered. The exclusion was based on Vermont Rule of Evidence 702, which is patterned after Federal Rule of Evidence 702, and 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.

This Note is self-explanatory.

5. 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?"

   This is discussed in Study Question 7, below, and in the introduction to the subsection that follows.

Study Questions:

1. Can you formulate hypotheses that might explain the varying results of the eight epidemiologic studies the claimant's experts relied on? Could those hypotheses be tested?

   A relatively obvious hypothesis is that random chance accounts for the variations. It is unlikely that even repeated samples of the same population using the same methodology would produce identical results. But of course, the eight studies were not repeated samples of the same population using the same methodology. Some were cohort studies and some were case-control studies. They studied firefighters in different places (metropolitan Toronto, Canada; Philadelphia; 27 U.S. states; 3 cities in the northwestern United States; 24 U.S. states; Melbourne, Australia; Florida; Massachusetts) over different periods of time (ranging from just a few years to several decades).

   Another hypothesis is that some undetected difference among the firefighters included in the various studies accounts for the different results. For example, perhaps firefighters in one study had greater toxic exposures because of more frequent runs, less use of protective equipment, etc. Or perhaps firefighters in one study happened to be less susceptible to NHL—or to toxin-induced NHL—than firefighters in another study. In principle, such hypotheses could be tested, but to do so would require data that apparently were not available to the researchers.

   Students might also offer other hypotheses based on the methodological difficulties of conducting retrospective cohort or case-control studies. These difficulties are discussed in Sections III.B.2.a., III.C., and III.E.2., supra.

This question could make a good in-class exercise. The instructor could divide the class into groups to brainstorm about possible hypotheses. A prize could be offered for the best (most creative?) hypothesis.

2. The claimant argued that his firefighting career placed him at higher personal risk of NHL than the average risk for firefighters observed in the studies for two reasons: (1) because his firefighting career was relatively long; and (2) because his firefighting career included a relatively long period of time during which no protective gear was provided. What assumptions underlay this argument? How well were those assumptions established by the evidence in the case?

   This study question is addressed in detail in the materials above.

3. How persuasive do you find the testimony referred to in Note 2 after the case excerpt? Might additional information help you decide? What would that be?

   This question refers to Dr. Lockey's testimony 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." Dr. Lockey noted particularly that immune deficiency is a known risk factor for NHL and Mr. George did not have an immune deficiency.

   This testimony ruled out a possible competing cause and therefore had at least some probative value. As with any differential etiology testimony, however, its power depends on how much ruling out the competing cause (immune deficiency) increased the likelihood of the alleged cause (firefighting). If it were known that virtually every case of NHL was caused either by immune deficiency or by exposure to toxic combustion products, this testimony would be extremely powerful. If, on the other hand, there were many risk factors for NHL, ruling out one particular risk factor would not seem particularly helpful to the claimant's case. A large proportion of idiopathic cases (cases of unexplained etiology) would also make it less important that a witness could rule out any given competing cause. See Section V.H.4., infra.

4. What significance did the majority and dissent assign to the fact (mentioned in Note 3 after the case excerpt) that the medical community considers NHL to be a label for a category of different yet related diseases, rather than the name of a single disease? What additional information would be relevant to deciding the significance of this fact?

   Both the majority and the dissent expressed concern that epidemiologic study of "NHL" generally might not provide accurate information about the role of firefighting in causing Mr. George's particular type of lymphoma. But the two

opinions addressed this concern from opposite directions.

The majority appeared to approve the trial judge's reasoning that because NHL comprises many distinct cancers, epidemiologic studies of "NHL" could find false associations. Why? The problem cannot be simply that the "overinclusive" disease definition would include too many firefighters with NHL, because the same broad disease category would also apply to the non-firefighters against whom the firefighters were compared. The real concern must be, then, that a claimant's particular type of lymphoma might not be associated with firefighting even if a study found an association for NHL generally. Taken to its logical conclusion, this reasoning would suggest that a claimant must find an epidemiologic study that shows more than a doubling of risk of the claimant's particular type of NHL. Is that realistic? Dr. Guidotti testified that epidemiologic studies do not ordinarily separate the many types of NHL. Why not? A look at the studies suggests a likely reason. The Figgs case-control study included some 25,000 cases of NHL, of which 12 were among firefighters. The Sama surveillance case-control study included 321 firefighters with cancer, of whom 14 had NHL. The Baris study included nearly 8,000 firefighters, of whom 20 died of NHL. With 30 or more NHL types, attempting to analyze each type separately would likely produce extremely small numbers.

The dissent, by contrast, focused on the possibility that firefighting might more than double the risk of a claimant's particular type of lymphoma, even if a study found less than a doubling of risk, or even no association, for NHL generally. Of course, the studies provided no data specifically about Mr. George's type of NHL, small-cell lymphoma. But Dr. Guidotti testified that small-cell lymphoma had been associated with exposure to solvents (presumably in other studies?), including some chemicals "released during firefighting." For the dissenters, this testimony tended to support an inference that firefighting more than doubled Mr. George's risk of NHL, even if firefighting does not double the risk of NHL in general. In other words, the dissent treated Mr. George's type of leukemia as an additional "other factor" increasing the probability of causation.

5. The dissent contended that the trial court and the majority, by focusing on whether the claimant's experts relied on epidemiologic studies showing more than a doubling of risk, conflated two distinct issues: whether the expert testimony was admissible (a legal issue for the court), and whether the expert testimony was sufficient to prove claimant's case (a factual issue for the factfinder). Is this contention well taken? Is it ever appropriate for the court to rule on the sufficiency of a plaintiff's evidence?

There is little doubt that when trial judges became gatekeepers of expert testimony, the line between admissibility and sufficiency blurred. Many critics of *Daubert* (and of the vast majority of the federal and state case law that *Daubert*

spawned) have contended that overzealous judicial gatekeeping improperly impinges on the jury's traditional function. See, e.g., Wendy Wagner, "The Perils of Relying on Interested Parties to Evaluate Scientific Quality," 95 *Am. J. Pub. Health* S99, S101–03 (2005). Nonetheless, courts have long been empowered to take cases from the jury, through summary judgment, directed verdict, or judgment notwithstanding the verdict if the evidence is insufficient to permit a reasonable inference, as opposed to only supporting speculation by the jury. Two of the authors of this module have argued that, especially with respect to toxic-tort-causation issues, determinations about the admissibility of expert testimony are increasingly indistinguishable from determinations about whether certain scientific evidence is sufficient to support an expert's ultimate conclusions—and that this is not necessarily a bad thing, although it should perhaps have some procedural implications. See Michael D. Green & Joseph Sanders, "Admissibility Versus Sufficiency: Controlling the Quality of Expert Witness Testimony," 50 *Wake Forest L. Rev.* 1057 (2015).

6. The dissenting opinion criticized the trial court for its "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." Assuming that the dissent accurately characterized the trial court's reasoning, is the criticism well taken?

   Yes. Strength of association and statistical significance are different concepts. Relative risk is a measure of the strength of a relationship. Statistical significance measures the probability that a sample result could have occurred by chance if there is no relationship between two variables (that is, if the null hypothesis is true). For any particular epidemiologic study with a sample of a given size, if the null hypothesis of no association is true, a larger relative risk is less likely to be observed by chance than is a smaller relative risk. But a large relative risk observed in one study might not be statistically significant, and a statistically significant relative risk observed in another study might not be large. See epidemiology section, Section III.E.1., supra. In *Estate of George*, the Aronson, Burnett, and Ma studies depicted on Slide 96 illustrate this. Aronson found more than a doubling of risk that was not statistically significant at the 95% level; Burnett and Ma found relative risks smaller than 2.0 that were (barely) statistically significant at the 95% level.

7. Note 5 after the case excerpt discusses the healthy survivor effect, which one of the claimant's experts posited as an explanation for the Baris study's finding that firefighters with the shortest time on the job appeared to have the highest risk of NHL. In effect, the expert argued that the healthy survivor effect is a source of error

that led the study to overestimate the relative risk for firefighters with shorter careers and to overestimate the relative risk for firefighters with longer careers. Epidemiologic studies confront several possible types of error; into which category of potential error would you place the healthy survivor effect?

The healthy survivor effect as described in *Estate of George* is a form of bias or systematic error. See the discussion in the epidemiology section, Section III.G., supra, in connection with the Demers study and *Lindquist*.

Section V.F.2. Using Genetics to Refine the Probability of Causation: Individual Susceptibility to Toxic Effects

The last sentence of Note 5 after *Estate of George* sets up the transition to this section. In *Estate of George*, Dr. Lockey suggested that the epidemiologic data were skewed because firefighters who were likely to contract NHL because of exposure to on-the-job toxins were apt to develop the disease relatively early in their careers, leaving only those who were "naturally less susceptible" to have longer careers. By asking the students to ponder why some firefighters might be "naturally less susceptible," Note 5 prompts them to think about the possibility that something intrinsic to the individual—genetics—might affect risk. A person's genes are not the only endogenous feature that might affect susceptibility to disease or to the effects of exposure to toxins, but genetic variation is an increasingly central focus of biomedical research that has implications for toxic-tort-causation problems.

a. The Basics of Genetics

This section begins with a very elementary introduction to genetics. The material is designed to be largely self-explanatory, but the instructor should gauge the level of background knowledge among students to determine whether to devote class time to the basic science. For those interested in obtaining more detailed background, a wealth of sources is available. We mention just two examples. Online, an excellent source is the U.S. National Library of Medicine's Genetics Home Reference; a good entree for the novice is *Genetics Home Reference: Help Me Understand Genetics* (2016), https://ghr.nlm.nih.gov/handbook. In print, Israel Rosenfield et al., *DNA: A Graphic Guide to the Molecule that Shook the World* (2011) is a relatively recent treatment that includes a historical perspective and transmits knowledge through an accessible pictorial format that belies the book's thoroughness. For those who do not feel a need to pursue other sources, we provide here just a few details that are omitted from the module.

The DNA and chromosomes described in this section are found in the nucleus of human cells. Other subcellular structures called mitochondria also contain a small

amount of DNA that is different from the nuclear DNA, but we need not address nuclear DNA in this module.

The body cells that ordinarily contain 46 chromosomes in 23 pairs are called somatic cells, which are different from germ cells (sperm and egg) that combine during reproduction. The germ cells ordinarily contain 23 chromosomes—one member of each pair—so that when a sperm cell fertilizes an egg, the resulting zygote will have 46 chromosomes in 23 pairs.

A chromosome does not consist of pure DNA. The DNA in chromosomes is bundled with associated proteins called histones, which play an important role in exposing the DNA for replication and translation (described below).

Figure V-4 in the module presents a schematic view of DNA copied from the U.S. National Library of Medicine website. Each base in the DNA molecule is depicted as a colored oblong; the four different chemical bases are represented by four different colors. In the depiction, the bases are strung together by a ribbon-like "backbone" connected to the oblongs. The part of each DNA nucleotide that forms the "backbone" is composed of a sugar and phosphate unit. In a molecule of DNA, these sugar-phosphate units are chemically bonded to each other.

As noted in the module and depicted in Figure V-4, in the DNA double helix each base projecting from one strand of the helix is paired with a specific complementary base in the other strand—adenine and thymine always complement each other, as do cytosine and guanine. This is how DNA can replicate to form new cells: the double helix "unzips" into two distinct strands of DNA bases, each of which is joined by a new strand of complementary bases, resulting in two molecules of DNA that are (ordinarily) identical to the original DNA molecule from which they formed.

The process of protein synthesis, through which the "genetic code" is translated into amino acids that are joined together to create proteins, is considerably more complex than described in the module. For our purposes, additional detail is probably unnecessary. The instructor might wish to know, however, that when a segment of DNA is to be translated into a chain of amino acids, the DNA segment is first unzipped in a way similar to what happens during DNA replication. Instead of generating a new complementary DNA strand, however, the DNA is "transcribed," as stated in the module, into a molecule of messenger RNA (ribonucleic acid).

RNA is a single-stranded molecule that, like DNA, is composed of four nucleotide bases (adenine, cytosine, and guanine, as in DNA, and uracil, in place of thymine). Each RNA base complements a particular DNA base (uracil in RNA complements adenine in DNA; thymine in RNA complements adenine in RNA; cytosine and guanine complement each other). Thus the process of transcription, when it works correctly, preserves in messenger RNA the information that was encoded in the DNA segment's sequence of bases.

The messenger RNA's three-base-long genetic "words" (codons) are "translated" into amino acids by another form of RNA: transfer RNA. Each transfer RNA molecule

includes a set of three bases complementary to a particular codon of messenger RNA and also carries with it a particular corresponding amino acid. Sub-cellular structures called ribosomes facilitate the assembly of the amino acids brought by transfer RNA into chains that, sometimes after additional biochemical processing, become proteins. For more detail about transcription and translation, see U.S. Nat'l Library of Med., *Genetics Home Reference: How Do Genes Direct the Production of Proteins?* (2016), https://ghr.nlm.nih.gov/handbook/howgeneswork/makingprotein, which includes links to animated renditions of these processes.

The correspondence between "genes" and proteins was proposed even before the structure of DNA was understood. Today's understanding is more complex, but for our purposes it will suffice to understand a gene as a unit of DNA that codes for a protein or a part of a protein. Among the more interesting findings to emerge from the Human Genome Project is the understanding that only a small fraction—about 2%—of human DNA consists of coding genes; the rest has a variety of functions that are still being identified. The number of human genes, interestingly enough, is not particularly greater than the number of genes in other species, even in much simpler organisms. Estimates of the number of coding genes in the human genome have varied. The rough estimate of "on the order of 20,000" given in the module is based on Iakes Ezkurdia et al., "Multiple Evidence Strands Suggest that There May Be as Few as 19,000 Human Protein-Coding Genes," 23 *Hum. Molecular Genetics* 5866 (2014).

We make no effort to provide details of the regulation of gene expression, a topic that is the subject of intense research and frequent discoveries. Altered gene expression, however, plays an important role in many human diseases, including diseases frequently involved in toxic tort claims. As discussed infra, x-ref, the profile of gene expression in affected tissues may be a potential biomarker that is relevant to toxic-causation issues.

b. The Basics of Genetic Variability

Our discussion of genetic variability focuses on variations in the sequences of nucleotide bases making up a gene, which alter the DNA "words" in the genetic code and therefore can alter the amino acid sequence and function of the protein the gene encodes. Variability arises from mutations—either imperfect copying of DNA during replication or subsequent damage to DNA. To be inherited by the next generation, the mutation must exist in a germ cell, that is, a sperm or egg cell that will form a new person.

Over time, mutations may produce many forms (polymorphisms) of a given gene. The easiest form of polymorphism to envision is the replacement of a single nucleotide base with another. Such "single nucleotide polymorphisms (SNPs)" are analogous to replacing a single letter in a word: *ball* might be altered to *fall*, *gall*, *hall*, or *wall*, or perhaps to *bill*, *bell*, or *bull* (and then maybe altered again to *fill*, *fell*, or *full*). The

instructor should be aware that these nucleotide replacements are not the only way in which genes may vary from person to person. For instance, one nucleotide or an entire three-nucleotide group might be deleted (like converting *blather* to *bather* or *father* to *fat*) or inserted (*mother* to *smother* or *northern* to *northeastern*) or inverted (*sore* to *eros* or *top* into *pot*). The same types of alterations sometimes occur at larger scales in the genome, involving substantial parts of chromosomes or even whole chromosomes.

Students will, of course, understand that genetic variability affects how people look, how their bodies function, and sometimes whether they are healthy or ill. We provide a couple of examples of how different genotypes produce different phenotypes—the observable physical characteristics of an organism. We start with the ABO blood types.

Slide 103 shows how each parent contributes one blood type allele: A, or B, or O. With three possible alleles from each parent, there are nine possible combinations. But notice that some of them are identical: AO in the upper right corner is indistinguishable from OA in the lower left corner. For blood groups, as for most (but not all) traits, which allele is inherited from which parent makes no difference to the phenotype. Slide 104 makes this clear. The upper chart on Slide 104 shows how the genotypes for this gene are translated into blood group phenotypes. The lower chart links each of the nine possible genotypes to its corresponding phenotype. Students should note that the genotype is more variable than the phenotype: a person with blood type A might have one or two A alleles; DNA analysis, but not blood type testing, could distinguish between these genotypes.

Slide 103 Inheritance of ABO Blood Groups: Genotype

Inheritance of ABO Blood Groups: Genotype

		Gene inherited from father		
		A	B	O
Gene inherited from mother	A	AA	AB	AO
	B	BA	BB	BO
	O	OA	OB	OO

103

Slide 104 Inheritance of ABO Blood Groups: Genotype to Phenotype

Inheritance of ABO Blood Groups: Genotype to Phenotype

If genotype includes:	Phenotype will be:
At least one A, no B	Blood Type A
At least one B, no A	Blood Type B
One A and one B	Blood Type AB
Neither an A nor a B	Blood Type O

		Gene inherited from father		
		A	B	O
Gene inherited from mother	A	Type A	Type AB	Type A
	B	Type AB	Type B	Type B
	O	Type A	Type B	Type O

104

It is worth cautioning here against a possible misinterpretation of the charts in Slides 103 and 104. Students might get the impression that the number of cells for each phenotype indicates how common that phenotype is, for example, that because blood type A can result from three different combinations of alleles, it must be more common than blood type O, which results from only one combination of alleles. But this would be true only if everybody had an equal chance of inheriting any of the three alleles from each parent. In fact, the O allele occurs more frequently in the population than either the A or B allele (although the exact mix of allele frequencies is different in populations of different ancestries). As a result, type O blood is the most common phenotype (at least in the United States). See American Red Cross, Blood Types, http://www.redcrossblood.org/learn-about-blood/blood-types (last visited Mar. 11, 2016).

The concept of penetrance, introduced at the end of the paragraph on ABO blood types, is important to understanding much of the discussion about the role of genetics in toxic-tort-causation disputes. Therefore, we include a few slides to illustrate the concept.

Slide 105 introduces penetrance by raising the possibility that the same genotype might produce different phenotypes in different people. The slide then defines penetrance as the proportion of people with a given allele or genotype who exhibit the phenotype in which we are interested. The line in the slide is to emphasize that this is calculated as a fraction. Hence, penetrance is a group property, not an individual property. In any individual with a certain genotype, the phenotype of interest is either present or absent, but for a group, we can compute the percentage of individuals who display the phenotype. Slide 106 illustrates this for a hypothetical genotype and phenotype. The genotype is either "@" or "*"; of the 100 individuals represented by the cells in the grid, 80 have the @ genotype and 20 have the * genotype. The frequency of * can be computed as 20/100 = 0.20 = 20%. The phenotype for this gene's trait is either beige or purple. An @ genotype ensures a beige phenotype, but only 10 of the 20 * genotypes have the beige phenotype; the other 10 have the purple phenotype. The penetrance of the * gene, with respect to the purple phenotype, is 10/20 = 0.50 = 50%.

Slide 105 Penetrance

Penetrance

What if some people with the same genotype had different phenotypes? To put it another way: what if a change (mutation) in a gene only sometimes resulted in a change in the phenotype (e.g. a disease)?

- **Penetrance defined: The Proportion of individuals with a particular allele or genotype who exhibit the associated phenotype of interest**

- **Calculating penetrance:**

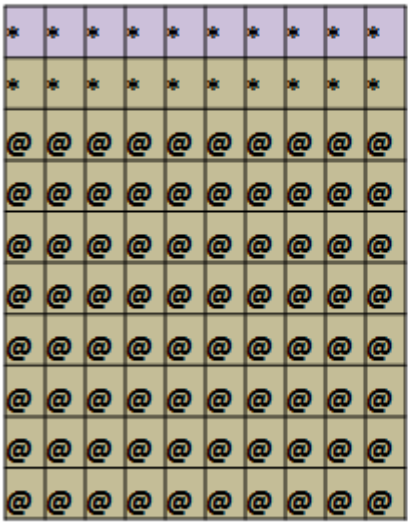
$$\frac{\text{\# of individuals with genotype who exhibit phenotype}}{\text{divided by}}$$

Total # of individuals with genotype

105

Slide 106 Penetrance, Illustrated

Penetrance, illustrated

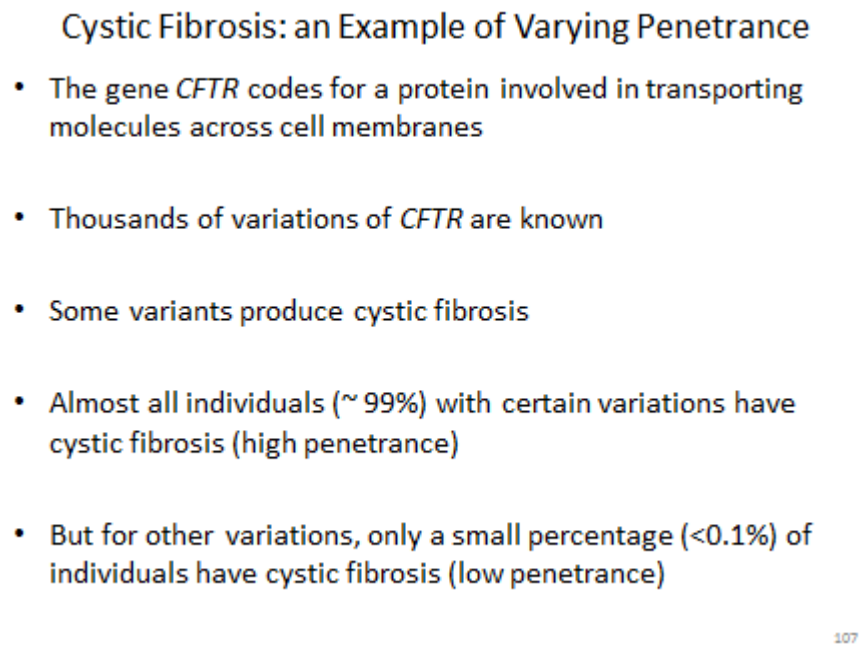


- Most people have the "@" genotype but some have the "*" genotype
- Frequency of the * genotype is 20/100 = 0.2 = 20%
- Everyone with the @ genotype has the beige phenotype
- Some with the * genotype have the purple phenotype
- Penetrance of the * genotype with respect to the purple phenotype is 10/20 = 0.5 = 50%

106

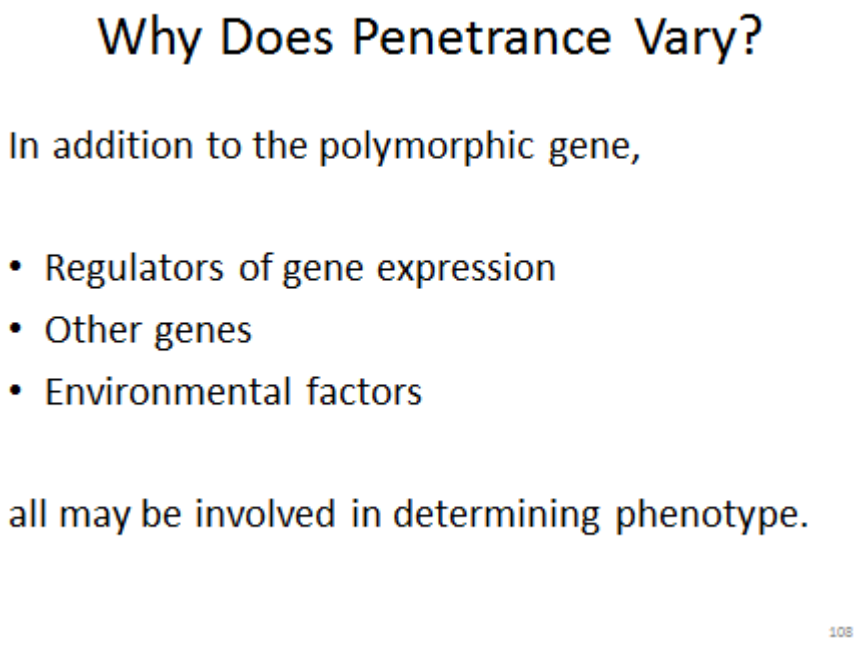
Slide 107 uses cystic fibrosis as an example of penetrance in the real world. We discuss cystic fibrosis later in the module, see infra, Section V.F.3., so the instructor may choose to defer this slide until then. The *CFTR* gene, sometimes called the "cystic fibrosis gene," codes for a protein (Cystic Fibrosis Transmembrane conductance Regulator) involved in transporting molecules across cell membranes. The gene is highly polymorphic, and many of the polymorphisms cause cystic fibrosis, a serious lung disease, in at least some people with the polymorphism. The penetrance of these alleles differs substantially, however, ranging from around 99% to less than 0.1%. For more information on the genetics of cystic fibrosis, see Sherman Elias et al., "Carrier Screening for Cystic Fibrosis: A Case Study in Setting Standards for Medical Practice," in George J. Annas &Sherman Ellis (eds.), *Gene Mapping* 186 (1992); Steven M. Rowe et al., "A Breath of Fresh Air: New Hope for Cystic Fibrosis Treatment," *Sci. Am.,* Aug. 1, 2011, at 69; C. Thauvin-Robinet et al., "The Very Low Penetrance of Cystic Fibrosis for the R117H Mutation: A Reappraisal for Genetic Counseling and Newborn Screening," 46 *J. Med. Genetics* 752 (2009); and Austl. Law Reform Comm., *Essentially Yours: The Protection of Human Genetic Information in Australia* (ALRC Report 96) Chapter 2 (2003), http://www.alrc.gov.au/publications/2-genetics-and-human-health-primer/importance-penetrance.

Slide 107 Cystic Fibrosis: An Example of Varying Penetrance

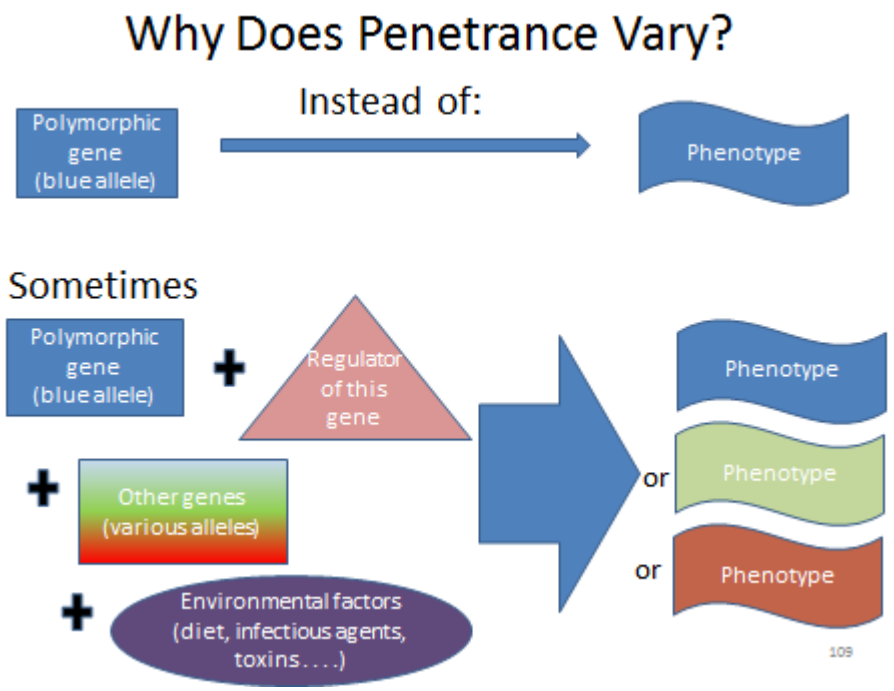


Students accustomed to thinking of genes in deterministic ways may be puzzled by the existence of genes that are not uniformly converted into a particular phenotype. If a gene that codes for a protein that would affect a particular trait does not always result in the same phenotypic manifestation of that trait, something besides that gene must affect the phenotype. Slides 108 and 109 present, in verbal and diagrammatic form, some of the major contributors to variable penetrance. For scientific reviews, see David N. Cooper et al., "Where Genotype Is Not Predictive of Phenotype: Towards an Understanding of the Molecular Basis of Reduced Penetrance in Human Inherited Disease," 132 *Hum. Genetics* 1077 (2013); and Rabah M. Shawky, "Reduced Penetrance in Human Inherited Disease," 15 *Egyptian J. Med. Hum. Genetics* 103 (2014).

Slide 108 Why Does Penetrance Vary?

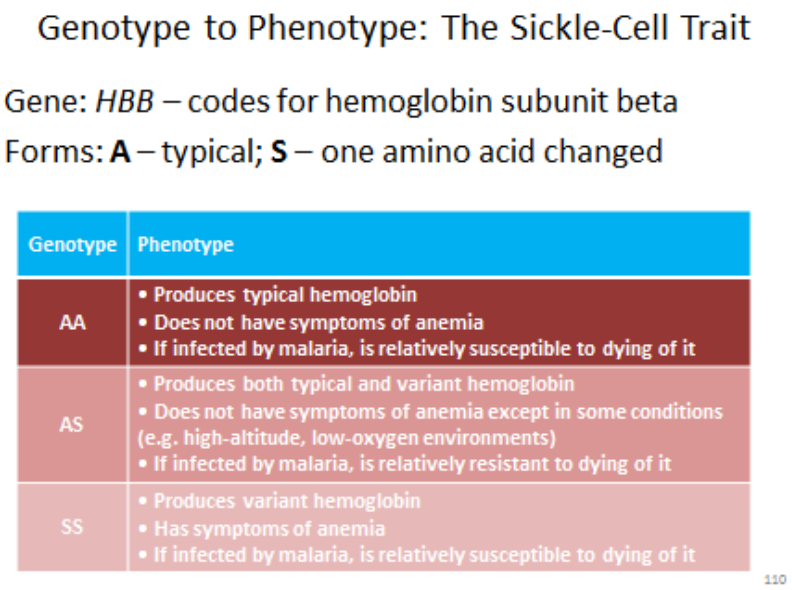


Slide 109 Why Does Penetrance Vary?



The sickle-cell example shows how genes and the environment interact to determine health. The discussion relates to a specific mutation (others exist) from the typical "A" form of the *HBB* gene, which codes for a chain of amino acids that forms one of the subunits of hemoglobin. In this "S" mutation, an adenine DNA nucleotide is replaced by a thymine nucleotide, changing the genetic code "word" so that the amino acid valine replaces the amino acid glutamic acid at a particular point in amino acid sequence. Slide 110 summarizes the phenotypic effects for persons with no, one, or two copies of the S allele (without regard to the possibility of other variations). Those interested in pursuing the subject more deeply will find innumerable discussions of sickle cell trait and sickle cell disease both online and in print. For a few examples, see Michael Aidoo et al., "Protective Effects of the Sickle Cell Gene Against Malaria Morbidity and Mortality," 359 *Lancet* 1311 (2002); Lucio Luzzatto, "Sickle Cell Anaemia and Malaria," 4 *Mediterranean J. Hematology & Infectious Diseases* e2012065 (2012), http://www.mjhid.org/article/view/10928; U.S. Nat'l Library of Med., *Genetics Home Reference: HBB—Hemoglobin Subunit Beta* (2016), https://ghr.nlm.nih.gov/gene/HBB; and Nat'l Heart, Lung & Blood Inst., *What is Sickle Cell Disease?*, http://www.nhlbi.nih.gov/health/health-topics/topics/sca (last updated June 12, 2015).

Slide 110 Genotype to Phenotype: The Sickle-Cell Trait



c. Introducing Genetics in Toxic-Tort Causation

At this point in the materials, students have learned (or remembered) enough information about genetics to grasp two important concepts: first, that genetic variations can cause (or protect against) disease on their own, and second, that genetic variations can cause (or protect against) diseases by interacting with environmental factors. These twin concepts frame the remainder of our discussion of the role of genetics in toxic-tort-causation issues, although—because our focus is on claims that toxic exposure caused disease—we take them up in the reverse order, starting with the way genetic variability affects susceptibility to toxic exposures.

Before plunging in, the instructor may wish to preview the subject by exploring with students the different roles that genetics might play in causing disease and linking them to claims of toxic causation. Slide 111 poses and partly answers this question: as students should recall from the basic presentation of causation (see Section II.A., supra), to be a but-for cause of a plaintiff's disease, a genetic variation must be a necessary element of a sufficient causal set. We illustrate this graphically on Slide 112, presenting a genetic variant (G0) as a link in the causal chain that ends in disease.

Slide 111 What Role Might a Genetic Variation Play in Causing Disease?

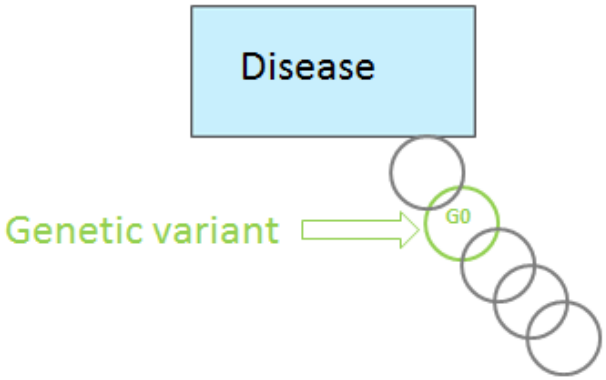
What role might a genetic variation play in causing disease?

To be a but-for cause of a disease, a genetic variation must be a necessary element of a sufficient set that brings about the disease.

111

Slide 112 Necessary Element of a Sufficient Set (Link in the Chain)

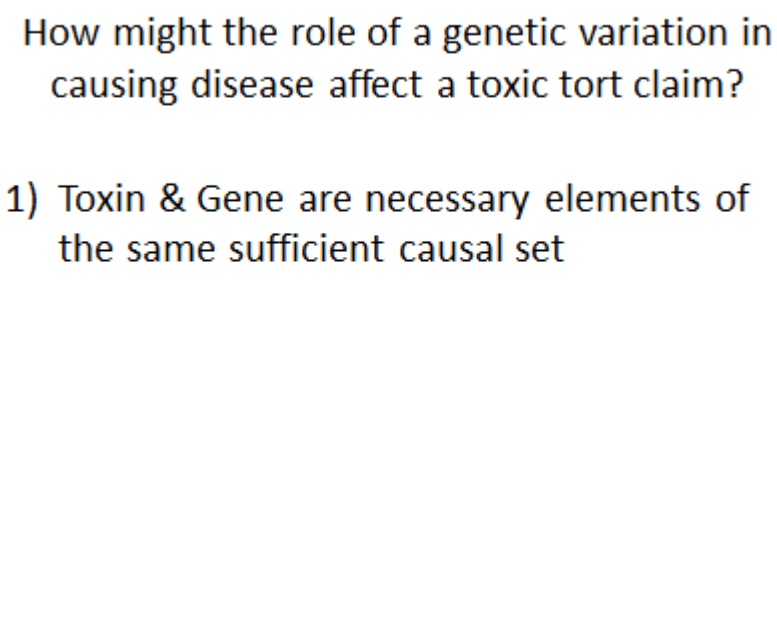
Necessary element of a sufficient set (Link in the chain)



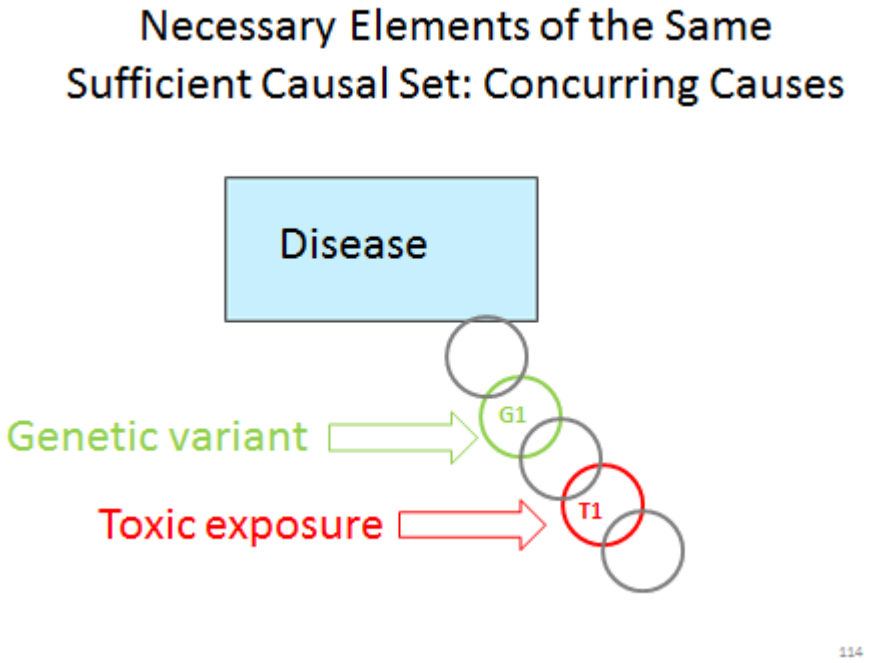
112

The instructor might ask: if a genetic variant were a necessary element in the causal chain leading to a plaintiff's disease, how would that affect a plaintiff's claim that exposure to a defendant's allegedly toxic substance caused the disease? The correct answer: it depends. Slide 113 asks this question and begins a list of possible answers. If the toxic exposure is *also* a necessary element of the causal set, then both the toxic exposure (T1) and the genetic variant (G1) would be but-for causes of the disease, as depicted in Slide 114. Or, as Slide 115 continues, the toxic exposure T1 and the genetic variant G1 might be necessary elements of different causal sets. If—as we implicitly assume here—only one or the other of these causal sets could have brought about the plaintiff's disease, the exposure and the gene would be competing causes, as shown in Slide 116. (We address the possibility of multiple sufficient causes in Slide 119, below.)

Slide 113 How Might the Role of a Genetic Variation in Causing Disease Affect a Toxic Tort Claim?



Slide 114 Necessary Elements of the Same Sufficient Causal Set: Concurring Causes



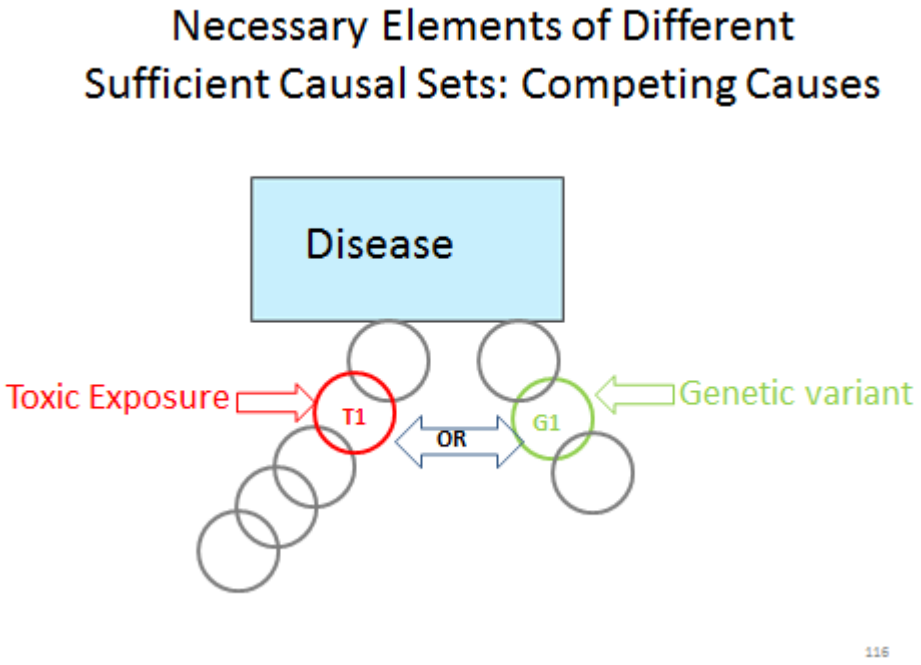
Slide 115 How Might the Role of a Genetic Variation in Causing Disease Affect a Toxic Tort Claim

How might the role of a genetic variation in causing disease affect a toxic tort claim?

1) Toxin & Gene are necessary elements of the same sufficient causal set
2) Toxin & Gene are necessary elements of different causal sets

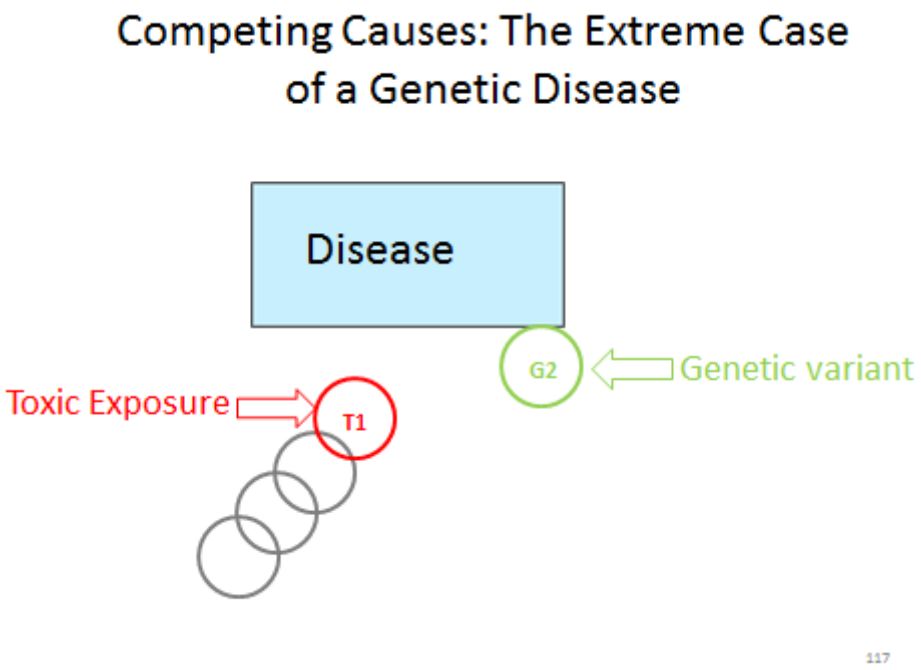
115

Slide 116 Necessary Elements of Different Sufficient Causal Sets: Competing Causes



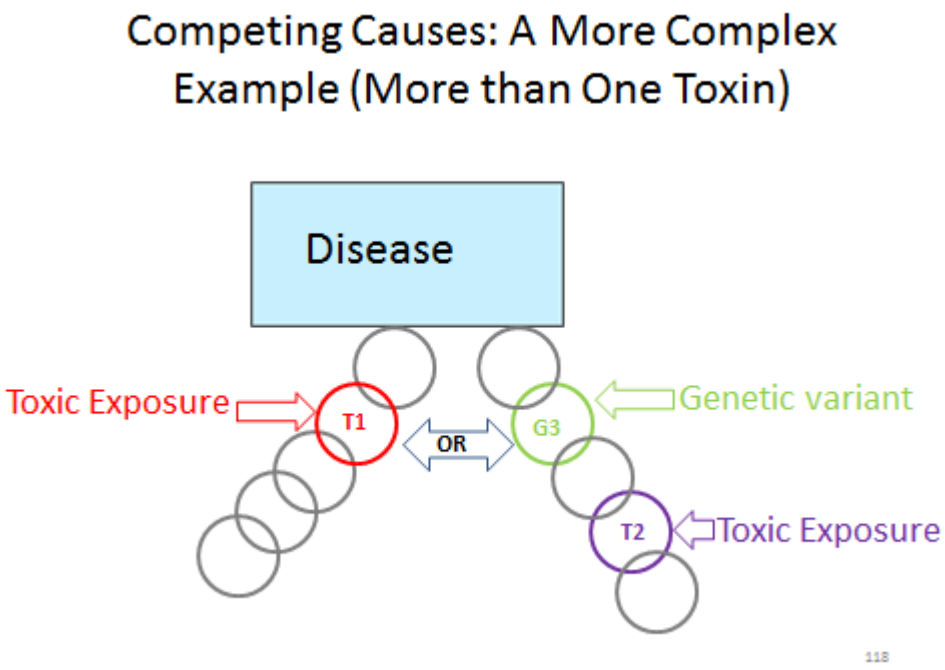
The extreme competing-cause scenario would be a genetically determined disease, such as sickle cell disease. In this situation, a person with the genetic variant (e.g., two copies of the S allele of the *HBB* gene) will have the disease, and a person without the genetic variant will not. Slide 117 portrays this situation as a causal "chain" consisting of one link (the genetic variant G2) and a putatively causal chain that includes the toxic exposure T1, but does not reach the supposed effect (the plaintiff's disease).

Slide 117 Competing Causes: The Extreme Case of a Genetic Disease



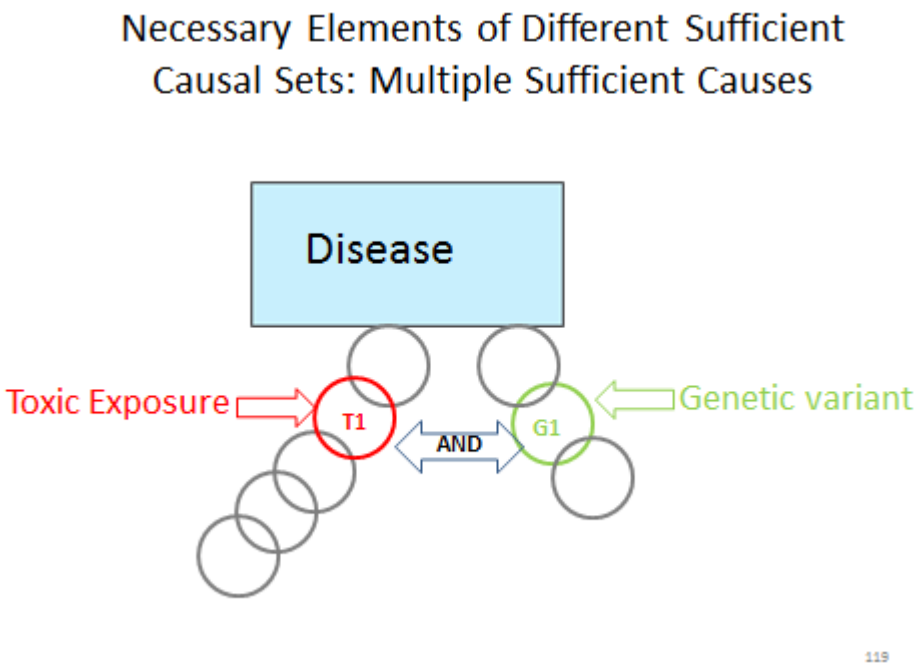
Slide 118 shows a more complex example: the alleged exposure T1 and the genetic variant G3 are elements of competing causal sets, but a toxic exposure T2 is also a necessary element of the latter causal set. If the plaintiff were exposed to both toxins and we knew nothing about a genetic component of the causal sets, a situation like this could present a very difficult competing-cause problem. Knowledge of the genetic component, however, opens up at least the possibility of a resolution: if it were proven that the plaintiff did not have genetic variant G3, the causal chain including T2 could be ruled out. T1 would be the cause (assuming, of course, that there were no other competing causes and that there were no other genetic variants that might have completed the causal chain including T2). The issues here are similar to those explored in Slides 65 and 66, supra, in connection with *Lindquist*.

Slide 118 Competing Causes: A More Complex Example (More than One Toxin)



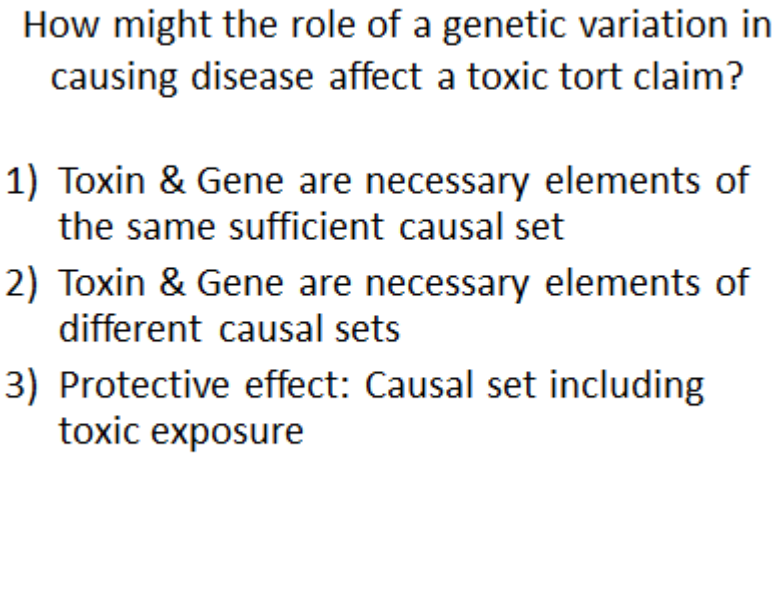
Slide 119 shows a different situation in which a toxic exposure T1 and a genetic variant G1 are necessary elements of different sufficient causal sets. Suppose both causal chains operate at the same time, and either would be sufficient to cause the disease? Then instead of competing causes, the toxic exposure and the genetic variant would be multiple sufficient causes, and each would be considered a cause-in-fact of the plaintiff's disease. See Section II.A., supra. A competing-cause model dominates toxic-tort litigation and jurisprudence, however. Court decisions frequently reflect an assumption that among potential causes (e.g., exposure to the defendant's toxin, exposure to the same toxin from some other source, exposure to some other toxin, or genetics), only one was the true cause of the plaintiff's disease.

Slide 119 Necessary Elements of Different Sufficient Causal Sets: Multiple Sufficient Causes



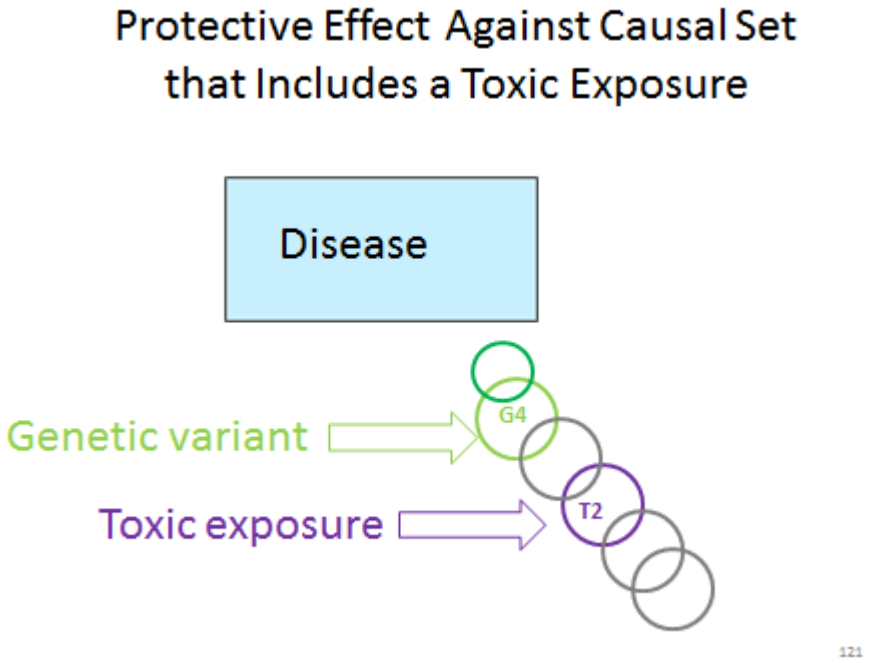
Variant genes do not always make matters worse. Slide 120 returns to our list of possible genetic effects and adds a third possibility: perhaps a genetic variation protects against an exposure's toxic effects.

Slide 120 How Might the Role of a Genetic Variation in Causing Disease Affect a Toxic Tort Claim?



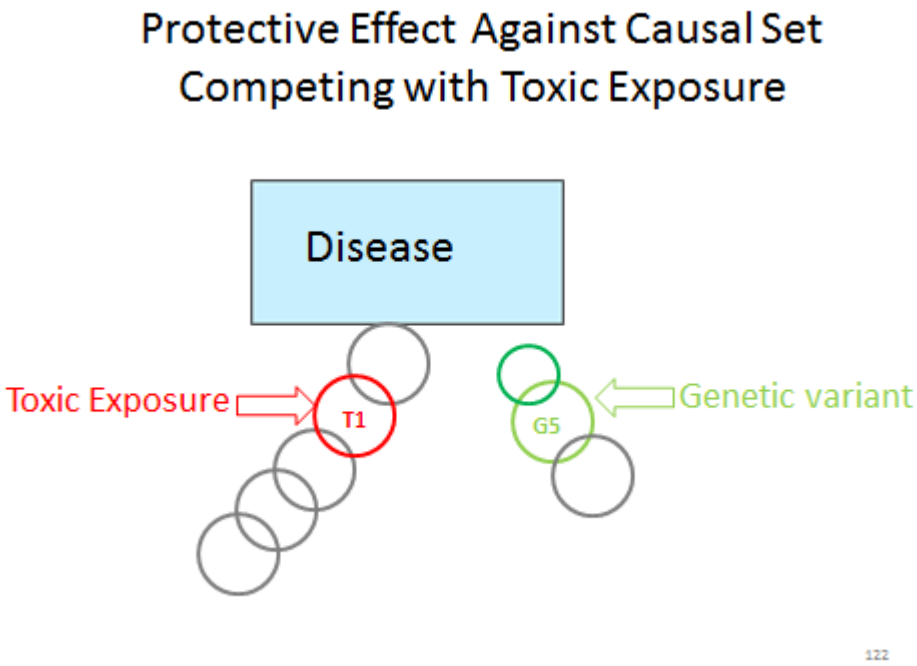
Slide 121 presents protective genetic variant G4 in the causal chain that includes toxic exposure T2. With G4 present, the chain no longer ends in disease. (Savvy students may recognize that to say one genetic variant is "protective" is to imply that other genetic variants are effectively causal. The presence of G4 in Slide 121 is really the same as the absence of G1 in Slide 114).

Slide 121 Protective Effect Against Causal Set that Includes a Toxic Exposure



Slide 122 presents protective genetic variant G5 in a causal chain that potentially competes with the causal set that includes toxic exposure T1 (and similarly, the presence of G5 in Slide 122 is really the same as the absence of G3 in Slide 118). G5's protective effect rules out a genetic cause for the disease and so (again assuming that no other possible causal sets exist that do not include T1), we can conclude that T1 was a but-for cause of plaintiff's disease.

Slide 122 Protective Effect Against Causal Set Competing with Toxic Exposure



In toxic-tort cases, however, the causal chains are not always so clearly defined. Epidemiologic research into the causes of disease often identifies a variety of risk factors. Those risk factors may include a person's genes, as Slide 123 states in completing our list of possible roles for genetics in toxic-tort-causation disputes. Slide 124 presents this idea graphically. Numerous risk factors (arrows), any of which may modify risk upward or downward, are shown influencing the risk of disease. They are grouped by type: genetics; nongenetic factors that nevertheless are endogenous to the individual (age being the easiest to grasp; epigenetics, the last on the list of examples, involves the regulation of gene expression and is discussed below); environmental factors (including both voluntary and involuntary exposures); and unknown or background factors.

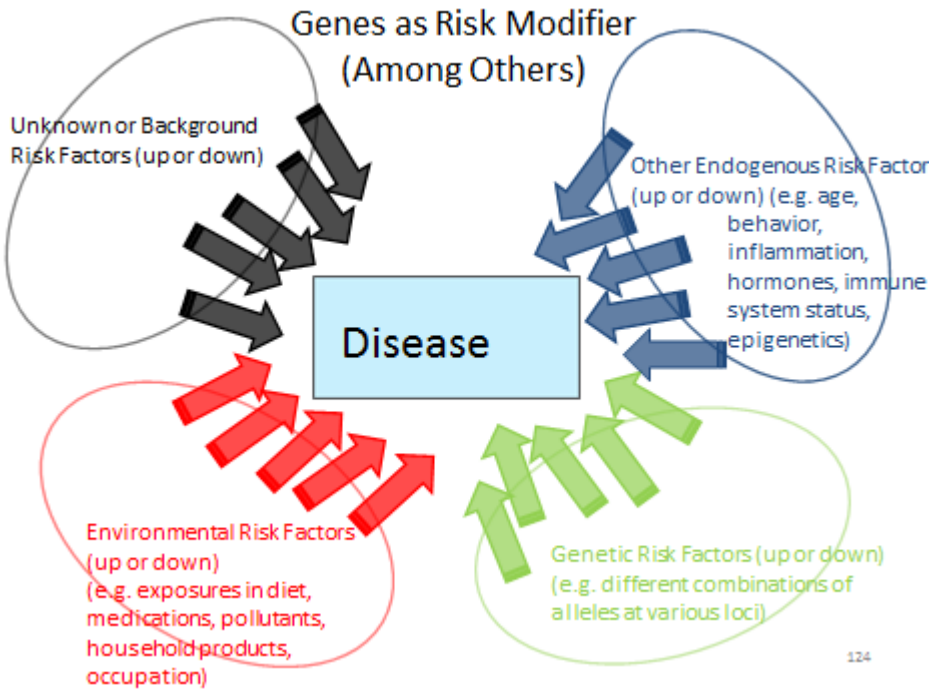
Slide 123 How Might the Role of a Genetic Variation in Causing Disease Affect a Toxic Tort Claim?

How might the role of a genetic variation in causing disease affect a toxic tort claim?

1) Toxin & Gene are necessary elements of the same sufficient causal set
2) Toxin & Gene are necessary elements of different causal sets
3) Protective effect: Causal set including toxic exposure
4) Risk modifier (causal or protective)

123

Slide 124 Genes as Risk Modifiers (Among Others)



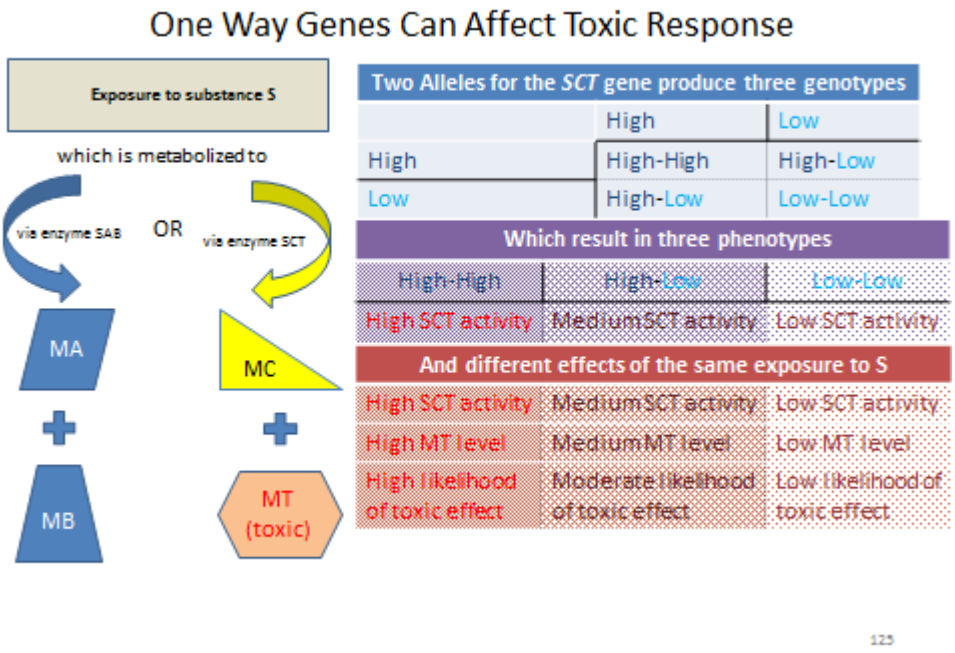
d. Genes as Modifiers of Toxic Susceptibility

With the foregoing conceptual introduction in hand, the instructor and students are ready to tackle the material on how genetic variability may modulate variable susceptibility to the effects of toxic exposures. We begin with a conceptual discussion of two mechanisms by which a person's genetic endowment may affect the outcome of a particular exposure to a potentially toxic substance.

Slide 125 illustrates the first general mechanism. A substance, "S," after entering the body, may be metabolized in either of two ways. The first pathway results in two metabolites, MA and MB, which are harmless. The second pathway results in two different metabolites, MC and MT. MC is harmless, but MT is toxic in some way—say, MT is carcinogenic or MT passes through the placenta and can cause birth defects. Each pathway depends on the action of a different enzyme: a protein synthesized by the body that catalyzes one or the other chemical reaction. The conversion of S to MC and MT is catalyzed by the enzyme SCT. The *SCT* gene, which codes for the SCT protein, has two alleles: High and Low. (By convention, the gene name is italicized and the protein name is not.) Persons with two High alleles have high-SCT enzyme activity in their body; persons with two Low alleles have low-SCT activity; persons with one High and one Low allele have an intermediate level of SCT activity. Higher levels of SCT activity result

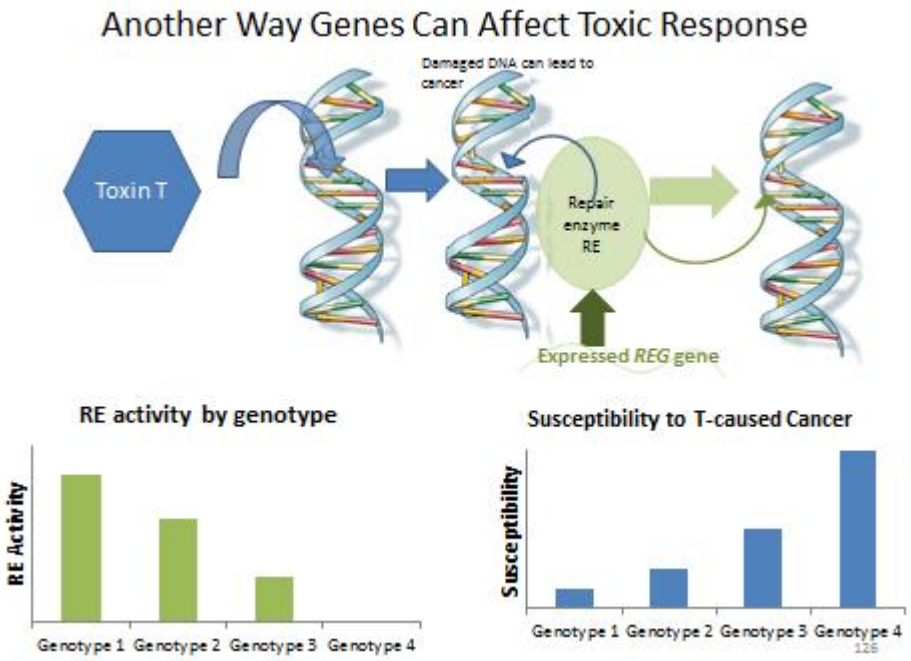
in more of any S consumed being converted to the toxic metabolite MT and, correspondingly, a higher likelihood of experiencing MT's toxic effect. This general mechanism is similar to the suspected interaction among exposure to trichloroethylene, polymorphic genes, and the risk of renal cell carcinoma, discussed infra in the Student Materials. The instructor may prefer to use Slide 125 to illustrate that example.

Slide 125 One Way Genes Can Affect Toxic Response



Slide 126 illustrates the second general mechanism. A toxin, "T," causes a change in DNA (exemplified by the deletion of one base pair) that is a step along the path from a normal cell to a cancer cell. The pathway can be interrupted by a DNA repair enzyme, RE. The gene, *REG*, that codes for the enzyme is polymorphic. Depending on the form of the gene inherited, functional RE may be absent from a person's cells, present but of limited activity, or present and of high activity. The less RE activity a person has, the more likely that a given exposure to T will cause cancer. This general mechanism is similar to the suspected interaction among exposure to tobacco smoke, the polymorphic *NAT2* gene, and the risk of breast cancer, discussed infra Section IV.F.2. of themodule. The instructor may prefer to use Slide 126 to illustrate that example.

Slide 126 Another Way Genes Can Affect Toxic Response



SOURCE: Image of DNA Molecule courtesy U.S. National Library of Medicine. Image in the public domain.[19] The remaining material in the slide is courtesy of the authors.

Several points made a bit later in the module are implicit in Slides 125 and 126. In each of these conceptual examples, the polymorphic gene does not absolutely determine whether exposure to T inevitably causes or never causes the disease outcome. Instead, having each form of the gene presents an individual with a different likelihood of developing disease after exposure. The variations in genotype modify risk. Genotype in this sense is a risk factor just as exposure is—after all, even for known toxins, exposure generally does not lead to certain disease, but only to an increased probability of disease. Thus, the response to the question "do you see why," following Figure 8 infra Section IV.F.2. of the module, is that even though a person's genotype is an individual property of that person, the degree of risk that genotype confers on an exposed person is determined by reference to a sample group of people who have that genotype and have been exposed to the toxin. We provide two examples drawn from the medical literature.

The *NAT2*–smoking–breast cancer example is from Christine B. Ambrosone et al., "Cigarette Smoking, *N-Acetyltransferase 2* Genotypes, and Breast Cancer Risk: Pooled Analysis

[19] http://ghr.nlm.nih.gov/handbook/basics/dna, visited Feb. 4, 2016.

and Meta-analysis," 17 *Cancer Epidemiology, Biomarkers & Prevention* 15 (2008). As the module points out, despite the finding that exposure to cigarette smoking and certain forms of the *NAT2* gene interact to increase a woman's risk of breast cancer, women with both of these factors may not develop breast cancer, and women with neither of these factors may develop breast cancer. What this shows (which the module asks the students) is that the *NAT2* genotype and cigarette smoking are, at most, risk factors for breast cancer—or to put it another way, that there are other necessary elements in the causal chains that include both of these factors. The abstract of this article follows. It demonstrates that the fundamental nature of the research is epidemiologic, albeit with molecular-scale risk factors considered.

Cigarette Smoking, *N-Acetyltransferase 2* Genotypes, and Breast Cancer Risk: Pooled Analysis and Meta-analysis

Christine B. Ambrosone,[1] Silke Kropp,[2] Jun Yang,[1] Song Yao,[1]
Peter G. Shields,[3] and Jenny Chang-Claude[2]

[1]Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York; [2]Unit of Genetic Epidemiology, Division of Cancer Epidemiology, Deutsches Krebsforschungszentrum, Heidelberg, Germany; and [3]Cancer Genetics and Epidemiology, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Northwest, Washington, District of Columbia

Abstract

Approximately 10 years ago, it was noted that smoking increased risk of breast cancer among women with *N-acetyltransferase* 2 (*NAT2*) slow acetylation genotypes. This report was followed by a number of studies to address this question. We pooled data from 10 existing studies and also conducted a meta-analysis of 13 studies published from 1996 to October 2006 that were conducted among women, were published in English, and had adequate information on smoking and *NAT2* genotyping. Raw data were requested from authors. Unconditional logistic regression was done for pooled analysis, and random effect models was done for meta-analysis. Study heterogeneity was assessed, and sensitivity tests were done when subgroups were excluded from the analysis. In the pooled analysis, there was a significant interaction between smoking, *NAT2* genotype, and risk of breast cancer [pack-years (continuous variable, $P_{interaction}$ = 0.03)], with higher pack-years significantly associated with an increased risk of breast cancer among women with *NAT2* slow genotypes (pooled analysis relative risk, 1.49; 95% confidence interval, 1.08-2.04). These findings were supported by the meta-analysis including all studies; pack-years were significantly associated with risk among slow acetylators in a dose-dependent fashion (meta-analysis relative risk, 1.44; 95% confidence interval, 1.23-1.68 for ≥20 pack-years versus never smokers), but not among rapid acetylators. Similar relationships were noted for smoking status (ever, never) and duration of smoking. Our results show that cigarette smoking is associated with an increase in breast cancer risk among women with *NAT2* slow acetylation genotypes. Because slow *NAT2* genotypes are present in 50% to 60% of Caucasian populations, smoking is likely to play an important role in breast cancer etiology. (Cancer Epidemiol Biomarkers Prev 2008;17(1):15–26)

SOURCE: Republished with permission of American Association for Cancer Research, from Christine B. Ambrosone et al., "Cigarette Smoking, *N-Acetyltransferase 2* Genotypes, and Breast Cancer Risk: Pooled Analysis and Meta-analysis," 17 *Cancer Epidemiology, Biomarkers & Prevention* 15 (2008); permission conveyed through Copyright Clearance Center, Inc.

The example involving polymorphic genes, TCE, and renal cell carcinoma (RCC) comes from several sources. The description of relative risks of RCC associated with TCE exposure is derived from a review article, Cheryl Siegel Scott & Weihsueh A. Chiu, "Trichloroethylene Cancer Epidemiology: A Consideration of Select Issues," 114 *Envtl. Health Persp.* 1471, 1477

fig.3 (2006). The first German study mentioned in the Student Materials is Thomas Brüning et al., "Influence of Polymorphisms of GSTM1and GSTT1for Risk of Renal Cell Cancer in Workers with Long-Term High Occupational Exposure to Trichloroethene," 71 *Archives Toxicology* 596 (1997). The study design, as described in the module, compared two groups of TCE-exposed workers, one consisting of people with RCC and one consisting of people without RCC, stratified by genotype. In response to the questions in the module: this is a case-control study design; the reported odds ratios were both statistically significant at the 95% level, because the lower bound of the 95% confidence interval was in each instance greater than 1.0.

The hospital-based study by the same research group is Thomas Brüning et al., "Renal Cell Cancer Risk and Occupational Exposure to Trichloroethylene: Results of a Consecutive Case-Control Study in Arnsberg, Germany," 43 *Am. J. Indus. Med.* 274 (2003). The follow-up study that could not confirm the hypothesis that the studied genotypes affected the relative risk for TCE of persons occupationally exposed to TCE is, as footnoted in the module, 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 (2007).

At the conclusion of the discussion of these studies, the module asks: "How would you interpret the results of these studies overall?" The answer, of course, is that it is difficult to say. The study published in 1997 was small: it involved 45 cases and 48 controls. Nearly 3 times as many cases and more than 8 times as many controls were included in the study published in 2003, but only 20 of the cases had occupational TCE exposure. It is not clear that occupational exposure was determined the same way in the two studies, or that all the exposures were equivalent. For the molecular epidemiology study published in 2007, the researchers obtained genotypes for only about 75% of the subjects in the 2003 study (a fact stated in the 2007 article's text but not its abstract). The results of the studies were inconsistent, although the 2007 article noted that the result for one genotype was fairly close to statistical significance at the P=0.5 level, particularly if additional controls were added to the sample. The results of the 2007 study casted doubt on the reality of the effect of genotype on TCE-RCC toxicity found in the 1997 study, but the failure to find such an effect in the 2007 study does not disprove that the effect exists.

One lesson that students can take from consideration of the genotype TCE-RCC studies is the importance of replication. A result obtained once may not be repeated, and even for studies that delve into the genome, few scientists would, in general, reach definitive conclusions from a single study. A second lesson is that adding a molecular dimension to an epidemiologic study does not make the difficulties of observational epidemiologic research disappear. A third lesson is that drawing conclusions from a collection of research results such as these requires the exercise of informed scientific judgment. Students get the chance to apply all these lessons in their assessment of the *Expert Report of T. Toxicologist* just a couple of paragraphs below.

Before presenting the expert report, however, we mention that the examples of variations in a single gene affecting susceptibility to the effects of toxic exposure are simpler

than the likely reality, because varying toxic susceptibility may arise from interactions among the studied genetic variation, variations in other polymorphic genes, and other variable environmental factors. The module discusses a review of studies of genetic susceptibility of air pollution, Cosetta Minelli et al., "Interactive Effects of Antioxidant Genes and Air Pollution on Respiratory Function and Airway Disease: A HuGE Review," 173 *Am. J. Epidemiology* 603 (2011). The statement about 90,000 variations in more than 600 genes is from Mark J. Rieder, "The Environmental Genome Project: Reference Polymorphisms for Drug Metabolism Genes and Genome-Wide Association Studies," 40 *Drug Metabolism Rev.* 241, 244 (2008).

The possibility of interaction is implicit in materials already studied. For example, in Slide 125 we focused on variations in the hypothetical *SCT* gene that coded for the enzyme that catalyzed one of two possible metabolic pathways for substance S. But what if the gene for the other enzyme, SAB, were also polymorphic in a way that affected the level of enzyme activity in the body? Those variations could also affect the risk of experiencing the toxic effects of exposure to S. The studies of TCE and RCC also evaluated the effects of multiple genes. Students should also understand, however, that interaction means more than multiple polymorphic genes acting independently to affect toxic risk. It is possible that having risk-enhancing alleles for two different genes will increase risk more than the sum of the risk increases produced by each of these alleles alone (i.e., a synergistic interaction). The expert report that follows includes an example of such an interaction.

Plaintiff v. Oil Co.: Expert Report by T. Toxicologist

State Superior Court.
Docket # YY-NNNN.

We include an excerpt from an expert report rather than from a court opinion in this section for two reasons. First, expert testimony about genetic susceptibility to toxic exposures has, at the time of this writing, figured in more trials than in reported court opinions. Second, we feel it is useful for students, for many of whom an expert report is only an abstraction, to see one—particularly in a module more focused on scientific than legal principles.

In the module, we use fictitious generic names for the case, the expert, and the court. As the modules states, however, the excerpt is adapted from real expert witness reports filed in a real case. "Adapted," rather than "excerpted," is accurate. The excerpt is a mash-up of portions of two reports that, in the actual case, were prepared by different experts: ArrayXpress, Inc. (by Len van Zyl, Ph.D.) and John B. Sullivan, Jr., M.D. At the time these materials were prepared, the litigation in which these reports were filed was still pending. The materials are used with the permission of the experts and of counsel with a request, which we honor, that we not identify the case by name.

Portions of the *Expert Report by T. Toxicologist* are taken verbatim from the ArrayXpress and Sullivan reports, but these excerpts have been rearranged and combined. The *Expert Report by T. Toxicologist* is also significantly different from the original reports from which it is derived. For pedagogical purposes, we paraphrased rather than quoted certain parts of the

original reports, added some material that was not in the original reports, and deleted large portions of the original reports (including methodological details and some studies that were supportive of the experts' conclusions). We also made a number of simplifications in service of our educational objectives. The original expert reports reached conclusions based on the plaintiff's genotype with respect to genes said to confer independent hereditary risk of plaintiff's disease as well as genes said to affect the plaintiff's susceptibility to toxic exposure. *Expert Report by T. Toxicologist* focuses only on the latter. We address genetic susceptibility as an independent competing cause in Section V.F.3., below.

The opening paragraph of the expert report sets the stage: plaintiff claims that on-the-job exposure to benzene in gasoline caused him to develop Acute Myeloid Leukemia, AML. These claims are brought fairly often by those who work with gasoline, for example, refinery employees and gasoline tanker drivers. Gasoline inevitably contains benzene, and benzene is a known leukemogen. Refiners and other defendants, however, note that the concentration of benzene in gasoline is low and contend that the actual exposures to benzene that workers receive have not been shown to cause leukemia. The expert report immediately identifies that dispute as a general-causation issue which is outside its scope. Instead, the expert report focuses only on specific causation, with reference to the plaintiff's genetic susceptibility to benzene's leukemogenic effects. The instructor might wish to take the opportunity to review the distinction between general causation and specific causation. The module notes that certain "other issues" have been omitted; these are primarily issues related to inherited genetic susceptibility to AML, as mentioned above.

The instructor should pause on the second paragraph to note the importance of a source of tissue from the plaintiff. The plaintiff's DNA could not have been analyzed without a sample of it! Obtaining a sample from a living plaintiff may involve a physical invasion of the plaintiff's person, albeit a minimal one. A plaintiff may voluntarily consent to testing or a defendant may obtain a court order requiring testing. In this case, the plaintiff was deceased, but the defendant located and obtained a tissue sample that had been preserved as a result of a medical procedure entirely unrelated to the plaintiff's claim. The stability of DNA and today's sequencing technologies allow analysis of old preserved samples and of small amounts of tissue.

Why must the tested DNA be obtained from healthy, noncancerous tissue? The DNA in cancer cells typically bears many mutations, at least some of which are consequences of the malignant state of the cells themselves. If the expert had analyzed DNA from the plaintiff's tumor, it would have been impossible to tell whether any mutations found were related to the causes or the effects of the cancer. Using healthy tissue allows the expert to be confident that any genetic variants found were inherited from the plaintiff's parents and did not arise as part of the development of the plaintiff's cancer. The same distinction, in a different way, is important to the Brauch study presented in Section V.G. on biomarkers, infra.

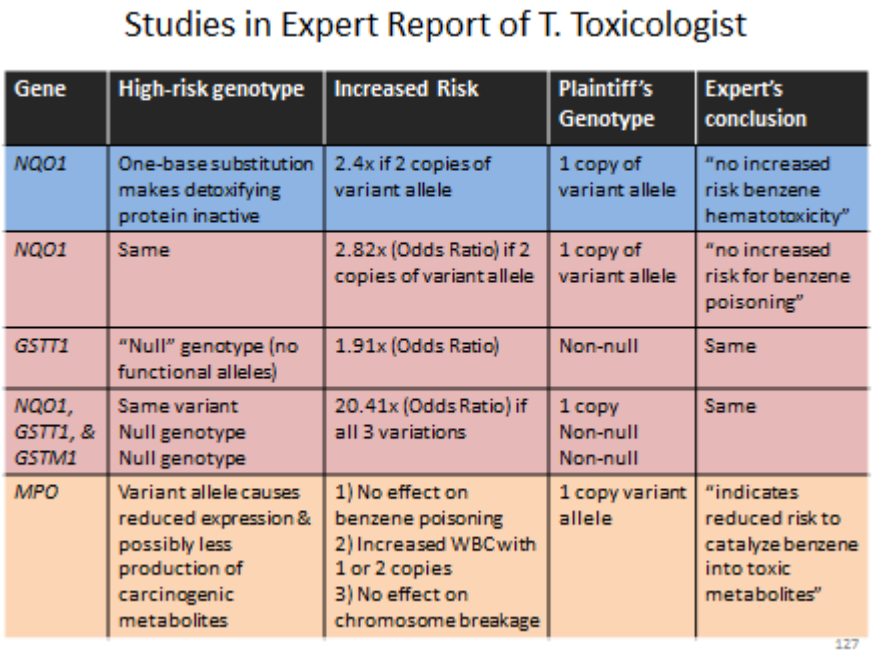
Students should appreciate that although the technology for sequencing DNA is highly automated and highly accurate, the machine does not simply spit out answers to toxic-tort-causation questions. As noted, the sequencing data go through a "curation" process and are

evaluated for degree of confidence. Even if a plaintiff's genotype for a particular gene is determined with a high degree of confidence, more important is the assessment of a genotype's significance. Making that assessment requires knowledge of: the gene's function (what protein does the gene code for, or what other gene does the gene regulate?), the metabolic role of the gene's product (how is the protein or regulated gene involved in the toxic substance's mode of action?), the effect of the plaintiff's variation on that metabolic role, and the alteration of risk that results from that effect.

T. Toxicologist focused on six genes that are "involved in benzene metabolism, detoxification, and repair pathways." In these six genes, the plaintiff's DNA contains seven variants that the expert compares to the variants described in six published studies of the genetics of susceptibility to benzene hematotoxicity. "Hematotoxicity" refers to toxicity to blood cells; leukemia is a cancer of white blood cells. The original expert reports in the *Harvey* case described each of these six studies; we limit our adapted excerpt to three of them.

We address T. Toxicologist's description of and inferences from the three studies in the Study Questions that follow. To set up the class discussion, the instructor might present a summary of the studies' findings, how the plaintiff's genotype compared to the genotypes discussed in the studies, and the conclusions the expert witness drew from them. We provide such a summary in Slide 127. Each colored band represents one of the three studies or groups of studies discussed in the expert report. The top band (blue) gives one study's results for one variant of one gene. The central band (pink) gives another study's results for three genotypes or groups of genotypes. The bottom band (orange) shows the results of three different studies of one genotype with respect to three different phenotypic endpoints.

Slide 127 Studies in Expert Report of T. Toxicologist



Studies in Expert Report of T. Toxicologist

Gene	High-risk genotype	Increased Risk	Plaintiff's Genotype	Expert's conclusion
NQO1	One-base substitution makes detoxifying protein inactive	2.4x if 2 copies of variant allele	1 copy of variant allele	"no increased risk benzene hematotoxicity"
NQO1	Same	2.82x (Odds Ratio) if 2 copies of variant allele	1 copy of variant allele	"no increased risk for benzene poisoning"
GSTT1	"Null" genotype (no functional alleles)	1.91x (Odds Ratio)	Non-null	Same
NQO1, GSTT1, & GSTM1	Same variant Null genotype Null genotype	20.41x (Odds Ratio) if all 3 variations	1 copy Non-null Non-null	Same
MPO	Variant allele causes reduced expression & possibly less production of carcinogenic metabolites	1) No effect on benzene poisoning 2) Increased WBC with 1 or 2 copies 3) No effect on chromosome breakage	1 copy variant allele	"indicates reduced risk to catalyze benzene into toxic metabolites"

127

This slide was prepared by the authors. The sources of the data summarized in the slide are the Len van Zyl report and references cited therein, as follows:

> Rothman, N. et al. "Benzene Poisoning, a Risk Factor for Hematological Malignancy, Is Associated with the NQO1 ^{609}C->T Mutation and Rapid Fractional Excretion of Chlorzoxazone." 57 *Cancer Research* 2839-2842 (1997), p. 2841 Table 3; Chen, Y. et al. "Genetic polymorphisms involved in toxicant-metabolizing enzymes and the risk of chronic benzene poisoning in Chinese occupationally exposed populations." 37 *Xenobiotica* 103-112 (2007), p. 103 (abstract); Wan, J. et al. "Association of Genetic Polymorphisms in CYP2E1, MPO, NQO1, GSTM1, and GSTT1 Genes with Benzene Poisoning." 110 *Envtl. Health Perspectives* 1213-1218 (2002), p. 1216 Table 4; Lan, Q. et al. "Hematotoxicity in Workers Exposed to Low Levels of Benzene." 306 *Science* 1774-1776, p. 1775; and Garte, S. et al. "Genetic Susceptibility to Benzene Toxicity in Humans." 71 *J. Toxicology Envtl. Health Part A* 1482-1489 (2008), p. 1485.

In addition to individually considering the studies or study groups, the instructor might bring out a general point about replication of studies, which is no less important in the genomic or molecular epidemiology context than in the context of classical epidemiology or toxicology. T. Toxicologist relied on two studies of the *NQO1* gene that had fairly consistent results, on one study of the *GSTM1* and *GSTT1* genes, and on three studies of the *MPO* gene that studied different endpoints and gave somewhat inconsistent results. If instead of addressing the role of genotypes, these studies had addressed the role of various toxic exposures or occupations, would the expert report have satisfied the admissibility criteria of the *Estate of George* majority?

Notes and Questions after *Plaintiff v. Oil Co.: Expert Report by T. Toxicologist*:

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 his allegation that his exposure to the 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. 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." Elizabeth A. Ward et al., "Research Recommendations for Selected IARC-Classified Agents," 118 *Envt'l Health Persp.* 1355, 1356 (2010).

   Note 1 helps explain an apparent weakness in T. Toxicologist's report. The goal of the study of the plaintiff's genes, as stated in T. Toxicologist's report, was simply to determine whether the plaintiff had "an inherited susceptibility" to benzene hematotoxicity. So what if he didn't? Does that mean he was invulnerable to benzene's leukogemic effects? As we develop further in the study questions that follow, in this context, the plaintiff's lack of "an inherited susceptibility" means only that the plaintiff did not have a genetically determined above-average probability of experiencing toxic effects from exposure to benzene. Did T. Toxicologist seem to suggest that a stronger conclusion can be drawn?

   Viewed against the defendant's trial objectives, however, even the narrow conclusion—that the plaintiff was no more susceptible to benzene toxicity than an average person—may be helpful. First, it may be useful as an attempt to refute the opinions of the plaintiff's experts. The original expert reports from which T. Toxicologist's report was derived also implied that the reports of the plaintiff's experts (which were not available to the authors of this module) raised the possibility that the plaintiff was more susceptible than average to benzene's hematotoxic effects. If so, then T. Toxicologist's report may be understood simply as an attempt to refute any implication that the plaintiff was at an unusually high risk of experiencing benzene's toxic effects (as opposed to an attempt to prove that the plaintiff was not at elevated risk of benzene toxicity).

   On the other hand, how strong is the scientific support even for that limited purpose? It depends on how much of the genetic basis of benzene susceptibility T. Toxicologist could rule out, just as the strength of a plaintiff's expert's differential etiology depends on how much of the universe of risk factors that expert can rule out. The quote from the Ward paper is intended to get students thinking about this. If most of the genetic basis for toxic susceptibility is still unexplained, excluding the

possibility that the plaintiff has a few known high-risk variations may not mean much.

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.

   The conclusions about the plaintiff's genetic susceptibility to benzene toxicity were accompanied, in the actual experts' reports, by a further conclusion that we did not include in the adapted T. Toxicologist excerpt: that the plaintiff had a genetic predisposition to AML. We mention the additional conclusion in Note 2 and address the topic of genetic predisposition to disease in subsection 3, infra.

Study Questions for *Plaintiff v. Oil Co.: Expert Report of T. Toxicologist*:

1. How persuasive is this expert report on the issue of specific causation?

   Questions can be raised with respect to the studies in each band of Slide 127; we discuss these issues in Study Questions 3 through 5. This Study Question may be used to explore the expert report more generally. At most, this report shows that the plaintiff did not have certain specific genotypes that are associated with a higher risk of cancer upon exposure to benzene. Does this mean that benzene probably did not cause the plaintiff's AML? Might other genes, not yet identified, mediate a causal chain to AML that includes benzene exposure? For people with the plaintiff's genotypes for the four genes discussed in the *Expert Report by T. Toxicologist*, what is the relative risk of AML with the plaintiff's occupational benzene exposure? The studies discussed in the expert report do not tell us.

   On the other hand, from the defendant's perspective, this expert report served purposes analogous to those for which the claimant in *Estate of George* tried to use the testimony of Dr. Guidotti and Dr. Lockey. If we conceive of these genotypes as "other factors" that correlate with the AML risk presented by benzene exposure, T. Toxicologist's report ruled out some factors that might have elevated the plaintiff's risk to the high end of the distribution. How strongly does this support the inference that the plaintiff was not, in fact, in the high end of the distribution—or, to track the legal issue more precisely, that the plaintiff was not in the "more likely toxic causation than not" range of the distribution? It depends on how fully these particular genes accounted for the high end of the distribution, much as the strength of a differential etiology depends on how thoroughly alternative causes can be ruled out. See also the discussion of Study Question 1 to the *Expert Report by T.*

Toxicologist, supra.

2. The expert report distinguished the genomic approach, which "uses the molecular person of Plaintiff and compares to the group," with the epidemiologic approach, which "use the group to extrapolate to the person." Is this a meaningful distinction? Does anything in the report suggest commonalities between epidemiology and the expert's method of reasoning or the scientific research upon which the expert relies?

   The sequence of the plaintiff's DNA is, of course, an individual property of the plaintiff rather than a property of a group. But the plaintiff's individual DNA sequence is not really the subject of interest, as it might be in a criminal case where a crime-scene DNA sample is used to establish the identity of a person who was there. The relative risk of toxic effect that varies with genotype is derived from study, not of the plaintiff or any other one individual, but of the group. Students should observe that the basic method of the studies the report relied on was epidemiologic; indeed, this research is usually called "molecular epidemiology." One could just as easily say that a classical epidemiology study "uses the unique exposure profile of the person and compares to the group," or that a molecular epidemiology study "uses the group risk profile to extrapolate to the person" who has a particular genotype in common with a study group. The very design of the studies discussed in the expert report, as well as the way in which the results are reported, show that these are epidemiologic studies even though they analyze risk factors that are genetic. Moreover, it is clear that some difficulties in using epidemiologic studies to support toxic-tort-causation claims—such as comparing a plaintiff's exposure levels to the exposure categories used in the studies—apply to the molecular-epidemiology studies that T. Toxicologist relied on no less than they apply to classical epidemiologic research.

3. Consider the first study the expert report discussed, of the *NQO1* gene. What conclusions might be drawn from the study? What, if any, additional information would be helpful to assess the weight this study should be given with regard to the plaintiff's claim of specific causation, that is, that the plaintiff's exposure to benzene caused the plaintiff's case of AML?

   This study was identified, in the original expert report on which the adapted excerpt is based, as Nathaniel Rothman et al., "Benzene Poisoning, a Risk Factor for Hematological Malignancy, Is Associated with the NQO1 609CT Mutation and Rapid Fractional Excretion of Chlorzoxazone," 57 *Cancer Res.* 2839 (1997). We have included a copy of the study in the Appendix.

The expert report relied on the study's finding that persons with two copies of the altered allele have a 2.4-fold risk of developing benzene poisoning compared to persons who have one or no copy of the altered allele. Students should understand the narrowness of this finding. Translated into legal terms, the result supports the conclusion that if a worker occupationally exposed to benzene developed benzene poisoning and had two copies of the variant *NQO1* allele, it is more likely than not that the variant allele was a cause of the benzene poisoning.

Students may notice that the endpoint studied was benzene poisoning, not AML. This may make a difference, although a different part of the same research paper reported that benzene poisoning is a powerful risk factor for the class of leukemias that includes AML: subjects with benzene poisoning had a relative risk of 70.6 (95% CI 11.4 – 439.3) for such leukemias and associated syndromes.

The study grouped together those who, like the plaintiff, had one variant allele and those who had no variant alleles. It is possible that the genetics of *NQO1* justify this (one "normal" allele may result in as much function of the coded enzyme as do two alleles), but neither the study nor the expert report said so explicitly. The next study discussed in the expert report (see Study Question 4, below) did distinguish between subjects with one and two variant genes and found no difference in risk between these two genotypes.

Most fundamentally, the Rothman study was not designed to investigate, and did not report on, the question that would bear most directly on the plaintiff's case: how much did exposure to benzene increase the risk of developing AML for persons with the plaintiff's genotype? Suppose, for example, that unexposed workers with the plaintiff's genotype faced a 1/100,000 risk of AML, exposed workers with the plaintiff's genotype faced a 3/100,000 risk, and exposed workers with two variant alleles faced a 7.2/100,000 risk (2.4 times the risk of exposed workers with the plaintiff's genotype). Using risk-doubling logic, it would be more likely than not that benzene exposure caused the plaintiff's AML, even though the plaintiff did not have the highest risk genotype.

Students might find it easier to grasp the nature of the study's conclusion if they see it as it was presented in the study itself. On Slide 128, we present the pertinent data extracted from a table in the study. The "cases" in the study were workers occupationally exposed to benzene who had been diagnosed with benzene poisoning. The "controls" were demographically matched healthy workers in industries without benzene exposure. The odds that a case had two variant *NQO1* alleles were 2.4 times higher than the odds that a control did. Why is no confidence interval reported for the odds ratio of 1.0 for the group with one or no variant allele? Because that is the group against which the other group is being compared.

(N.B.: Computing the odds ratio from the table in the manner described in the epidemiology section, see supra x-ref, yields an odds ratio of 2.32. The slight difference between this figure and the study's reported 2.4 figure is explained by the

adjustments the researchers made to match cases and controls by age and sex. The instructor might also ask the students if the reported odds ratio of 2.4 is statistically significant. The lower bound of the 95% confidence interval is 1.0. The authors reported this as evidence to reject the null hypothesis, but some would say that because the 95% confidence interval includes 1.0, albeit just barely, the result is not statistically significant at the P=.05 level).

Slide 128 Effect of *NQ01* Genotype on Risk for Benzene poisoning in Shanghai, China, 1992

Effect of *NQO1* Genotype on risk for Benzene Poisoning in Shanghai, China, 1992

Attribute	Number (%) of Cases (workers exposed to benzene with benzene poisoning)	Number (%) of Controls (workers not exposed to benzene without benzene poisoning)	Odds Ratio (95% Confidence Interval)
Two copies of variant *NQO1* allele	20 (41%)	11 (23%)	2.4 (1.0 – 5.7)
One or no copy of variant *NQO1* allele	29 (59%)	37 (77%)	1.0

128

SOURCE: Courtesy of the authors based upon data from Nathaniel Rothman et al., "Benzene Poisoning, a Risk Factor for Hematological Malignancy, Is Associated with the NQO1 ^{609}C→T Mutation and Rapid Fractional Excretion of Chlorzoxazone," 57 *Cancer Research* 2839, 2841 Table 2 (1997).

It is worth noting in passing that the researchers faced a methodological difficulty that traditional epidemiologic studies also face: how to assess and categorize levels of exposure, particularly because of the study's retrospective design. The cases in this study were workers who faced much higher benzene exposures than the plaintiff gasoline worker; as is the case for the studies on which experts for such plaintiffs typically rely.

4. Consider the second study the expert discusses, of the three genes *NQO1*, *GSTT1*, and *GSTM1*. Why did the study consider three genes? Does the study indicate anything about how the three genotypes studied combined to affect the risk of benzene poisoning in people like the plaintiff? Can you explain why one of the 95% confidence intervals reported in this study is so much larger than the others?

   This study was identified, in the original expert report on which the adapted excerpt is based, as Y. Chen et al., "Genetic Polymorphisms Involved in Toxicant-Metabolizing Enzymes and the Risk of Chronic Benzene Poisoning in Chinese Occupationally Exposed Populations," 37 *Xenobiotica* 103 (2007). We provide the study's abstract[20] here:

[20] Reproduced with permission. The full article is available at: http://www.tandfonline.com.

Genetic polymorphisms involved in toxicant-metabolizing enzymes and the risk of chronic benzene poisoning in Chinese occupationally exposed populations

Y. CHEN[1], G. LI[2], S. YIN[2], J. XU[2], Z. JI[2], X. XIU[3], L. LIU[4], & D. MA[5]

[1]*College of Public Health, XinJiang Medical University, Urumqi, China,* [2]*Institute of Occupational Health and Poison Control, Chinese Center for Disease Control and Prevention, China,* [3]*Hospitals of Prevention and Treatment for Occupational Diseases in Zhengzhou, China,* [4]*Hospitals of Prevention and Treatment for Occupational Diseases in Kaifeng, China, and* [5]*Hospitals of Prevention and Treatment for Occupational Diseases in Tianjin, China*

(Received 4 July 2006; revised 18 August 2006; accepted 9 September 2006)

Abstract

Benzene is a recognized haematotoxin and leukaemogen, but its mechanism of action and the role of genetic susceptibility are still unclear. Cytochrome P450 2E1 (*CYP2E1*) and myeloperoxidase (*MPO*) are involved in benzene activation; and NAD (P)H:quinine oxidoreductase 1 (*NQO1*), glutathione *S*-transferase theta 1 (*GSTT1*) and glutathione *S*-transferase mu 1 (*GSTM1*) participate in benzene detoxification. The common, well-studied single-nucleotide polymorphisms (SNPs) were analysed in these genes drawn from the toxicant-metabolizing pathways. A total of 100 workers with chronic benzene poisoning (CBP) and 90 controls were enrolled in China. There was a 2.82-fold (95% CI = 1.42–5.58) increased risk of CBP in the subjects with the *NQO1* 609C > T mutation genotype (T/T) compared with those carrying heterozygous (C/T) and wild-type (C/C). The subjects with the *GSTT1* null genotype had a 1.91-fold (95% CI = 1.05–3.45) increased risk of CBP compared with those with *GSTT1* non-null genotype. There was no association of *CYP2E1* and *MPO* genotype with CBP. A three genes' interaction showed that there was a 20.41-fold (95% CI = 3.79–111.11) increased risk of CBP in subjects with the *NQO1* 609C > T T/T genotype and with the *GSTT1* null genotype and the *GSTM1* null genotype compared with those carrying the *NQO1* 609C > T C/T and C/C genotype, *GSTT1* non-null genotype, and *GSTM1* non-null genotype. The study provides evidence of an association of a gene–gene interaction with the risk of CBP.

Chen's group, like Rothman's group, assessed the effects of various genotypes on the risk of benzene poisoning in occupationally exposed workers. T. Toxicologist's expert report relied on three findings from this study: the 2.82-fold increased risk associated with two copies of the same *NQO1* allele that Rothman studied; the 1.91-fold increased risk associated with the *GSTT1* null genotype; and the 20.41-fold increased risk associated with having all three genotypes (*NQO1* two variant alleles, *GSTT1* null, and *GSTM1* null).

The study evaluated three genotypes because: (1) multiple genes were suspected of affecting susceptibility to benzene poisoning, and (2) the researchers

wanted to evaluate possible interactions of the multiple genes involved. In fact, the study investigated five genes (and for one of the genes, investigated mutations at two different bases in the DNA sequence). But some genetic variations or combinations of genetic variations showed no statistically significant association with benzene poisoning; T. Toxicologist's report focused on three statistically significant findings in the Chen report.

Like the Rothman study, the Chen study supported conclusions about the probability that a particular genotype or genotype combination contributed to causing benzene poisoning in a worker occupationally exposed to benzene. It did not study the effects of exposure versus non-exposure for the different genotypes. In fact, unlike the Rothman study, the Chen study recruited "controls" who *had* occupational exposure to benzene but who *did not* have benzene poisoning—a more appropriate choice of control group given the objectives of the study. See the discussion in Study Question 3.

On Slide 129, we present the data underlying the findings of the Chen study that were cited in T. Toxicologist's report. The "adjusted" odds ratio is computed using statistics that adjust for variations in the study subjects' age, sex, exposure duration, alcohol consumption, and smoking (all of which are potential confounding variables). The authors of the Chen study, as well as the authors of the expert reports from which T. Toxicologist's report is derived, used the adjusted odds ratios. The unadjusted odds ratios, which may be readily computed from the raw data given on Slide 129, can be found in Table III of the study.

Slide 129 Effect of Genetic Polymorphisms on the Risk of Chronic Benzene Poisoning

Effect of Genetic Polymorphisms on the Risk of Chronic Benzene Poisoning

Polymorphism	No. (%) Cases (benzene exposure + poisoning)	No. (%) controls (benzene exposure, no poisoning)	Adjusted Odds Ratio (95% confidence interval)
NQO1 **no variant allele**	22 (22.00%)	25 (27.78%)	1.00
NQO1 **1 variant allele**	40(40.00%)	49 (54.44%)	1.00 (0.48 – 2.09)
NQO1 **2 variant alleles**	38 (38.00%)	16 (17.78%)	2.94 (1.25 – 6.90)*
GSTT1 **non-null**	45 (45.00%)	53 (58.89%)	1.00
GSTT1 **null**	55 (55.00%)	37 (41.11/5)	1.91 (1.05 – 3.45)*
GSTM1 **non-null**	37 (37.00%)	43 (47.78%)	1.00
GSTM1 **null**	63 (63.00%)	47 (52.23%)	1.67 (0.92 – 3.05)

Red print with * indicates statistically significant results.

129

SOURCE: Y. Chen et al., "Genetic Polymorphisms Involved in Toxicant-Metabolizing Enzymes and the Risk of Chronic Benzene Poisoning in Chinese Occupationally Exposed Populations," 37 *Xenobiotica* 103, 108 Table III (2007). Reprinted by permission of Taylor & Francis Ltd, www.tandfonline.com.

The three purple rows in the table on Slide 129 present data for the same *NQO1* variant that Rothman's group studied. Unlike Rothman et al., Chen et al. produced separate data for those who had one or no copy of the variant allele. The odds ratio for the 1-variant group is 1.00 (which, of course, is not a statistically significant difference), meaning this study found no increased risk from having just one copy of the variant *NQO1* allele (compared to having no copies of the variant allele). But for individuals with two variant alleles, the odds ratio (compared to those with no variant copy) was 2.94—somewhat higher than, but not very different from the Rothman result.

The two blue rows in the table on Slide 129 present data for the null and non-null genotypes of *GSTT1*. This gene codes for an enzyme that detoxifies benzene; people with null genotypes do not synthesize any active enzyme. The Chen study found that for workers with occupational exposure who developed benzene poisoning, the odds of having a null *GSTT1* genotype were 1.91 times higher than the

odds that a similarly exposed worker without benzene poisoning had a null genotype.

Finally, the two orange rows in the table on Slide 129 present data for the null and non-null genotypes of *GSTM1*, another gene that codes for an enzyme that detoxifies benzene. This portion of the study was not mentioned in T. Toxicologist's report. What did the Chen study's data tell us about the effect of a *GSTM1* null genotype on the risk of benzene poisoning after occupational exposure to benzene? Students should, by now, recognize that although the study found an association (odds ratio 1.67), the association was not statistically significant (because the lower end of the 95% confidence interval is less than 1.0, albeit only slightly less). Thus, if one applies alpha = 0.05 as the significance level, the Chen study did not produce evidence to reject the null hypothesis that there is no difference between *GSTM1* null and *GSTM1* non-null genotypes with respect to benzene-poisoning risk after occupational exposure.

After considering the effect of each of the five genes studied individually, Chen's group analyzed the effect on benzene poisoning of various combinations of genotypes for groups of these genes. The result for one combination—two variant *NQO1* alleles and *GSTT1* null genotype and *GSTM1* null genotype as compared to individuals with none of these genotypes—is discussed in T. Toxicologist's report. Slide 130 presents the data and the odds ratios for this combination.

Slide 130 *NQO1*, *GSTT1*, *GSTM1* and Genetic Susceptibility to Chronic Benzene Poisoning

NQO1, GSTT1, GSTM1 and Genetic Susceptibility to Chronic Benzene Poisoning

Genotype	No. Cases (benzene exposure + poisoning)	No. (%) controls (benzene exposure, no poisoning)	Adjusted Odds Ratio (95% confidence interval)
NQO1 **no or 1 variant allele +** ***GSTT1*** **non-null +** ***GSTM1*** **non-null**	11	21	1.00
NQO1 **2 variant alleles + GSTT1 null + GSTM1 null**	17	2	20.41 (3.79 – 111.11)*

Red print with * indicates statistically significant results.

130

SOURCE: Y. Chen et al., "Genetic Polymorphisms Involved inToxicant-Metabolizing Enzymes and the Risk of Chronic Benzene Poisoning in Chinese Occupationally Exposed Populations," 37 *Xenobiotica* 103, 109 Table V (2007). Reprinted by permission of Taylor & Francis Ltd, www.tandfonline.com.

What does this study show about how these genes combine to affect risk of benzene poisoning in occupationally exposed workers? Students, recalling the epidemiology material (Section V.E., supra), should see that the data are a clear example of interaction. The genotype with two variant *NQO1* genes, measured alone, increased risk 2.94-fold; the *GSTT1* null genotype increased risk 1.91-fold; the *GSTM1* null genotype did not increase risk at all (if only statistically significant results are considered) or increased risk 1.67-fold. The effect of all three genotypes far exceeds the sum of each effect independently. Slide 131 shows this. In their conclusion, "the authors suggest that the interaction of multi-genes may be an important factor in chronic benzene poisoning. Chen et al., supra, at 110. Those interested in pursuing further the distinction between additive risk factors and interacting risk factors may consult Susan R. Poulter, "Genetic Testing in Toxic Injury Litigation: The Path to Scientific Certainty or Blind Alley?" 41 *Jurimetrics* 211, 221–229 (2001) (contrasting

different ways multiple factors can affect risk in the context of discussing gene environment rather than gene-gene interaction).

Slide 131 Effect of Genetic Polymorphisms on the Risk of Chronic Benzene Poisoning

Effect of Genetic Polymorphisms on the Risk of Chronic Benzene Poisoning

Polymorphism	Adjusted Odds Ratio (95% confidence interval)
NQO1 2 variant alleles (compared to 1 or none)	2.94 (1.25 – 6.90)*
GSTT1 null (compared to non-null)	1.91 (1.05 – 3.45)*
GSSTM1 null (compared to non-null)	1.67 (0.92 – 3.05) (not sig.)
NQO1 2 variant alleles + *GSTT1* null + *GSTM1* null (compared to 1 or non, non-null, and non-null)	**20.41 (3.79 – 111.11)***

Red print with * indicates statistically significant results.

131

SOURCE: Y. Chen et al., "Genetic Polymorphisms Involved inToxicant-Metabolizing Enzymes and the Risk of Chronic Benzene Poisoning in Chinese Occupationally Exposed Populations," 37 *Xenobiotica* 103, 109 Table V (2007). Reprinted by permission of Taylor & Francis Ltd, www.tandfonline.com.

Finally, the study question invites students to contemplate the very wide confidence interval for the odds ratio for the three-genotype combination presented in Slides 130 and 131. The point estimate of the odds ratio is 20.41, but the 95% confidence interval stretches from 3.79 to 111.11. Why is that? The answer is in plain sight on Slides 140 and 141. It has to do with the size of the study sample. In the studies of individual genotypes, the comparison involved all of the cases and all of the controls. For example, in the study of *GSTT1*, every case and every control could be categorized either as a null or non-null genotype. Thus, the statistical comparison was between the odds of a null genotype among 100 cases and 90 controls (Slide 129). But the number of individuals in the study with the suspected high-risk genotype for

all three genes was much smaller: only 11 cases and 21 controls, who had to be compared to just 17 cases and 2 controls who had no suspected high-risk genotypes for any of the three genes. Everything else being equal, a smaller sample size will produce a wider confidence interval. Note, however, how few controls—benzene-exposed individuals without benzene poisoning—had all three suspected high-risk genotypes. This relative rarity resulted in a very high ratio, so that even though the 95% confidence interval was very high, the estimated odds ratio of 20.42 was statistically significant not only at the P=.05 level, but even at the P=.01 level.

The Chen study appears to provide evidence that if a person with two variant *NQO1* alleles and a null genotype for both *GSTT1* and *GSTM1*, who was occupationally exposed to benzene, develops benzene poisoning, it is highly likely that the combination of these three genotypes played a causal role in the development of the illness. The study did not directly address how much exposure to benzene would increase the risk of benzene poisoning or AML in someone with none of those three genotypes.

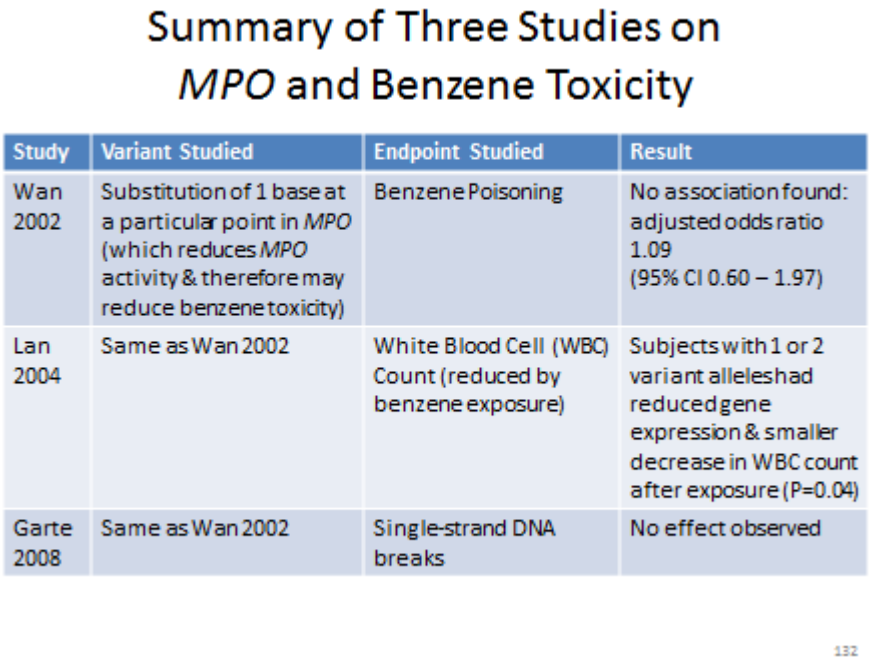
5. Consider the third group of studies the expert discusses. What are the strengths and weaknesses of these studies? How probative is this group of studies on the specific-causation issue?

The three studies cited in the expert reports on which this portion of T. Toxicologist's report is based were, respectively: Junxiang Wan et al., "Association of Genetic Polymorphisms in CYP2E1, MPO, NQO1, GSTM1, and GSTT1 Genes with Benzene Poisoning," 110 *Envtl. Health Persp.* 1213 (2002); Q. Lan et al., "Hematotoxicity in Workers Exposed to Low Levels of Benzene," 306 *Sci.* 1774 (2004); and S. Garte et al., "Genetic Susceptibility to Benzene Toxicity in Humans," 71 J. *Toxicology Envtl. Health A.* 1482 (2008). These papers all involved studies of the same variation at one base in the *MPO* gene. This group of three studies is sufficiently complex, and their results sufficiently equivocal, such that we choose not to present them in any detail. The original expert reports involved even more complexity, as they also discussed research involving other polymorphisms in *MPO*.

Slide 132 briefly summarizes the results of the Wan, Lan, and Garte studies. Students should note that the three studies investigated three different health outcomes: benzene poisoning, reduced white-blood-cell count, and chromosome breakage. All of these are effects that have been associated with benzene exposure or are risk factors for AML or both, but none is AML itself. Also noteworthy is that the only one of these studies (Lan) that found an association between the outcome and the variant genotype is the study that involved workers exposed to low concentrations of benzene. On one hand, this might suggest that the Lan study involved subjects with exposures similar to the plaintiff's. On the other hand, it seems odd that the gene would have an effect at low exposure concentrations but

no effect (albeit different effects) at higher exposure concentrations. At the very least, Lan's results suggest a need for further investigation.

Slide 132 Summary of Three Studies on *MPO* and Benzene Toxicity



Summary of Three Studies on *MPO* and Benzene Toxicity

Study	Variant Studied	Endpoint Studied	Result
Wan 2002	Substitution of 1 base at a particular point in *MPO* (which reduces *MPO* activity & therefore may reduce benzene toxicity)	Benzene Poisoning	No association found: adjusted odds ratio 1.09 (95% CI 0.60 – 1.97)
Lan 2004	Same as Wan 2002	White Blood Cell (WBC) Count (reduced by benzene exposure)	Subjects with 1 or 2 variant alleles had reduced gene expression & smaller decrease in WBC count after exposure (P=0.04)
Garte 2008	Same as Wan 2002	Single-strand DNA breaks	No effect observed

132

SOURCE: Courtesy of the authors based upon information from Junxiang Wan et al., "Association of Genetic Polymorphisms in CYP2E1, MPO, NQO1, GSTM1, and GSTT1 Genes with Benzene Poisoning," 110 *Envtl. Health Persp.* 1213 (2002); Q. Lan et al., "Hematotoxicity in Workers Exposed to Low Levels of Benzene," 306 *Sci.* 1774 (2004); and S. Garte et al., "Genetic Susceptibility to Benzene Toxicity in Humans," 71 J. *Toxicology Envtl. Health A.* 1482 (2008).

T. Toxicologist's report concluded that "Plaintiff has one copy of the variant *MPO* allele, indicating a reduced risk to catalyze pre-carcinogens such as benzene into toxic metabolites." The studies, however, do not strongly support this conclusion. The Wan and Garte studies found no effect associated with this particular genetic variation in *MPO*. The Lan study found an effect on white-blood-cell counts, but T. Toxicologist's report said nothing about how strong this effect was. The report inferred from this observation about white blood cells and the plaintiff's genotype that the plaintiff was at reduced risk of catalyzing benzene into toxic metabolites, leaving implicit the conclusion that the plaintiff was at a reduced risk of AML. Slide 132 traces these issues.

All this may well be relevant evidence, and may be a basis for an expert opinion.

But how probative is it? How much protection from benzene-caused AML does the plaintiff's genotype confer? The studies cannot answer that question and T. Toxicologist's report did not attempt to.

Section V.F.3. Using Genetics to Refine the Probability of Causation: Individual Inherited Susceptibility to Disease as a Competing Cause

This section introduces the concept of genetic risk factors for disease that are independent of toxic exposures. The idea that genetic "defects" can cause disease directly will probably be more familiar to students than the idea that genes can interact with the environment (e.g., toxins) to cause disease.

In the opening paragraph, we refer back to our discussion in the introduction to genetics of the genetic basis for sickle cell disease. See Section V.F.2., supra. This type of high-penetrance genetic disease, in which having a particular genotype for a single gene leads to disease, is what most people think of when they think of genetic disease; this image propagates through the popular media's frequent reporting that researchers have discovered the "such-and-such gene." As the cystic-fibrosis example shows, however, even for single-gene disorders, the situation can be much more complex. The instructor may want to return to, or use for the first time, Slide 107 here. One point worth noting is that researchers and clinicians distinguish between penetrance of a genotype—which is defined as the proportion of persons with the genotype who exhibit the phenotype, disorder, or disease of interest—and variable expressivity (the severity of symptoms experienced by different individuals who have the disorder or disease). See U.S. Nat'l Library of Med., *Genetics Home Reference: What Are Reduced Penetrance and Variable Expressivity?* (2016), https://ghr.nlm.nih.gov/handbook/inheritance/penetranceexpressivity.

The module touches only briefly on high-penetrance genetic diseases, but the instructor might wish to pause to consider such diseases as the limiting case of genetics as a competing cause. We presented this situation schematically in Slide 117. Essentially, if a genetic variation (plus background causes) always, or nearly always, results in disease regardless of exposure, that is strong evidence that exposure is not a *necessary* element of any sufficient causal set.

As a practical matter, epidemiologic data may not exist that compare the incidence of disease across groups of people who all have the genetic variation but who vary with respect to whether they were exposed to the alleged toxin. Instead, the irrelevance of exposure will be inferred from genomic research that explains the disease mechanism, identifies the altered protein responsible, and traces the alteration in the protein to an inherited genetic variation—thus elucidating a complete pathway from gene to disease that does not depend in any way on exposure. Inherited genetic disease as a competing cause has figured prominently in claims for compensation under the Childhood Vaccine Injury Act of 1986, 42 U.S.C. Sections 300aa-1 to -34. We mention this in the module immediately before the *Bowen* case. For examples, see

Snyder v. Secretary of Health and Human Services, 553 Fed. App'x 994 (Fed. Cir. 2014) (court accepted testimony that mutation was sole cause of claimant's condition, rather than a cause together with vaccine); *Deribeaux v. Secretary of Health and Human Services*, 717 F.3d 1363 (Fed. Cir. 2013) (compensation denied where genetic cause of symptoms was discovered after claim was filed); and *Hopkins v. Secretary of Health and Human Services*, 84 Fed. Cl. 530 (2008) (court concluded claimant's condition was hereditary based on family history even though plaintiff had no known disease-causing mutation in the relevant gene).

As the module notes, most toxic-tort claims involve complex diseases that are not an invariable consequence of a particular genotype of a particular gene. Nevertheless, researchers have identified numerous genotypes that are associated with a wide range of diseases and conditions; such discoveries are reported in the media on a daily basis and students will have heard of them. What students need to understand is that almost always, such "susceptibility genes are neither necessary nor sufficient to cause disease. They modify risk." Kenneth Olden & Janet Guthrie, "Genomics: Implications for Toxicology," 473 *Mutation Res.* 1, 5 (2001).

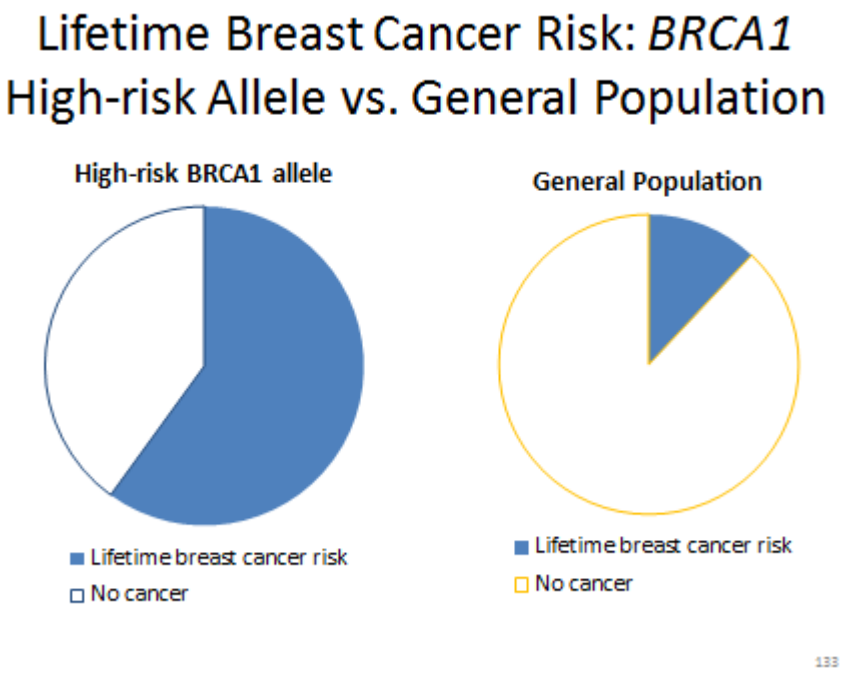
Even among risk-modifying alleles, penetrance varies substantially. We begin our discussion with a relatively high-penetrance example: the "breast cancer genes" (more precisely, breast cancer susceptibility alleles of these genes) *BRCA1* and *BRCA2*. The instructor may wish to remind students that an understanding of these genotypes is already affecting important personal medical decisions, at least for some people. The actress Angelina Jolie, knowing of her family history of early death from breast cancer, had her genotype tested and upon learning the results, opted in a very public way to have preventive surgery removing the organs at risk. See Angelina Jolie, "My Medical Choice," *New York Times*, May 14, 2013, at A25; Angelina Jolie Pitt, "Diary of a Surgery," *New York Times*, Mar. 24, 2015, at A23.

Published estimates of the penetrance of the major *BRCA1* and *BRCA2* susceptibility alleles have varied widely. The meta-analysis referred to in the Student Materials is Sining Chen and Giovanni Parmigiani, "Meta-Analysis of BRCA1 and BRCA2 Penetrance," 25 *J. Clinical Oncolog*y 1329 (2007). Published estimates of the overall lifetime risk of breast cancer also vary substantially. The 12.3% figure in the module is from Jennifer Scalia-Wilbur et al., "Breast Cancer Risk Assessment: Moving Beyond BRCA1 and 2," 26 *Seminars Radiation Oncology* 3 (2016). Many scientific research papers make the last point in this paragraph—that despite the increased risk associated with certain *BRCA1* and *BRCA2* variants, individual breast-cancer risk depends on many other factors, including the genotype of other genes and (as discussed further below) environmental factors.

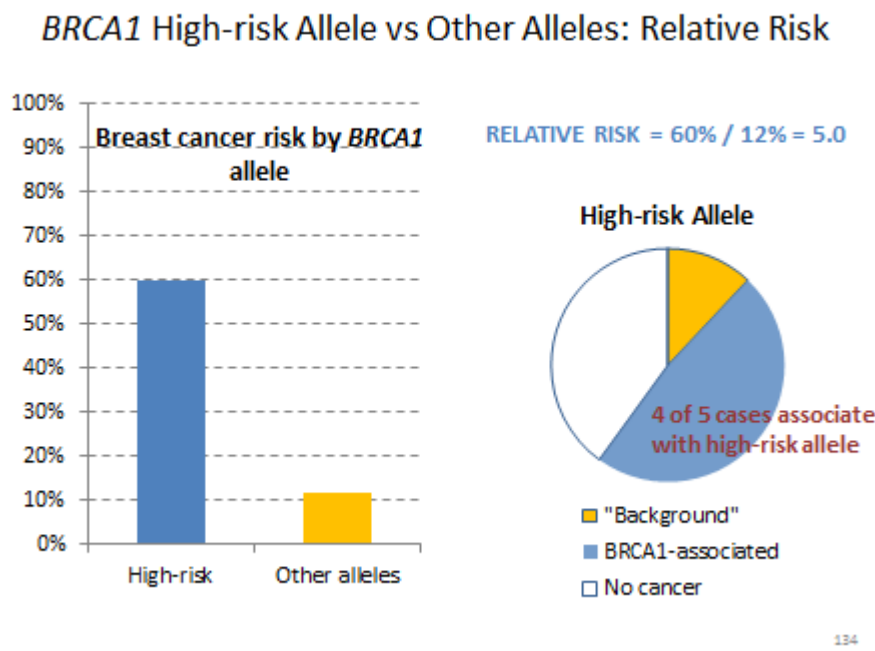
The instructor might take a moment here to review the mechanics of relative risk and attributable fraction, which will help students analyze the hypothetical toxic-tort case that follows. Consider a high-risk *BRCA1* allele. The module states that women with high-risk *BRCA1* alleles have a 57% lifetime risk of breast cancer compared to a 12.3% lifetime risk for all women. For teaching purposes, we can round these numbers to 60% and 12%. Slide 133 presents this situation graphically. Can we calculate a relative risk? No, because the "all women" risk includes some women who have the high-risk allele. To calculate a relative risk, we would need to know the frequency of the high-risk allele in the population. For teaching

purposes, we will make a further simplifying assumption (without asserting that it is true) that the high-risk allele is rare enough that we can disregard its effect on the overall population's lifetime risk of breast cancer. With that assumption, we could compute a relative risk of 5 (60% / 12%) for the high-risk *BRCA1* allele, implying that in the subpopulation of women with the high-risk allele, the genetic variation accounts for four out of five cases of breast cancer. Slide 134 shows this.

Slide 133 Lifetime Breast Cancer Risk: *BRCA1* High-risk Allele vs. General Population



Slide 134 BCA1 High-risk Allele vs Other Alleles: Relative Risk



Now consider a hypothetical plaintiff who has a high-risk *BRCA1* allele and who also has been exposed to a toxin that—we will assume for purposes of analysis—also increases the risk of breast cancer. Certainly it would appear relevant to know the plaintiff's *BRCA1* genotype. As the module says, the defense would argue that the plaintiff's unfortunate allele and not the toxic exposure was the real cause of the cancer. In fact, the defendant might argue that because this plaintiff was born with a 57% chance of developing breast cancer sometime during her life (or, equivalently, that the gene confers a risk five times higher than the general population's), it is impossible that the exposure "more likely than not" caused the disease.

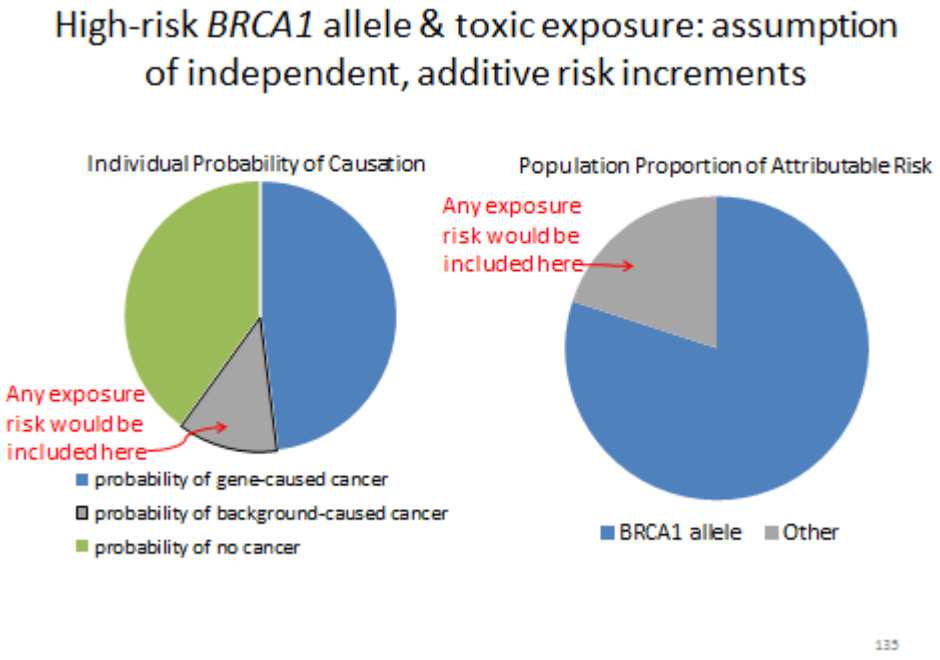
The argument depends on an essential assumption: that the *BRCA1* allele and the exposure independently confer additive risks of breast cancer. If this assumption is false, then some of the incidence of breast cancer that appears to be caused by the high-risk *BRCA1* allele is also caused by the exposure. Stated another way: if the genetic and environmental risks are not independent, then at least some of the time, the high-risk allele—instead of being a competing cause in a separate causal chain that does not include the exposure—would be a concurring cause in the same causal chain that includes the exposure. See Slides 114 and 116.

The treatment of *BRCA1* variants as independent sources of differential risk of disease stands in contrast to the studies of *NQO1* (referred to in the *Expert Report of T. Toxicologist*), which study how *NQO1* variants interact with benzene exposure to confer differential risk of disease. A look back at *Lindquist* (Section III.H., supra) will show that the court treated

hereditary deficiency of alpha-1-antitrypsin (AAT) as a competing cause of emphysema, in contrast to environmental risk factors, such as tobacco smoking and exposure to air pollution at work. Was that treatment justified? AAT is a protein; it is coded by the *SERPINA1* gene. Mutations in the gene reduce the amount or effectiveness of AAT synthesized. Even in the *Lindquist* opinion itself, there is a clue that genetic and environmental factors may interact to cause emphysema: the defendant's expert testified that the risk of emphysema is especially high for smokers with close relatives (the genetic factor) who were smokers (the environmental factor) and had emphysema. According to the U.S. National Library of Medicine, "Environmental factors, such as exposure to tobacco smoke, chemicals, and dust, likely impact the severity of [AAT] deficiency." U.S. Nat'l Library of Med., *Genetics Home Reference: Alpha-1 Antitrypsin Deficiency* (2016), https://ghr.nlm.nih.gov/condition/alpha-1-antitrypsin-deficiency. So the mutations, by affecting the amount or activity of AAT, may confer risk of emphysema on their own, but (some of?) the mutations also may interact with toxic exposure(s) to create additional risk of emphysema or more severe disease.

The instructor may choose to explore these issues at this point, or to defer them for discussion in connection with genome-wide association studies and lower-penetrance alleles, infra. For those inclined to pursue the issue in connection with high-penetrance breast cancer alleles, Slide 135 illustrates graphically the assumption of independent risks.

Slide 135 High-risk *BRCA1* Allele and Toxic Exposure: Assumption of Independent, Additive Risk Increments

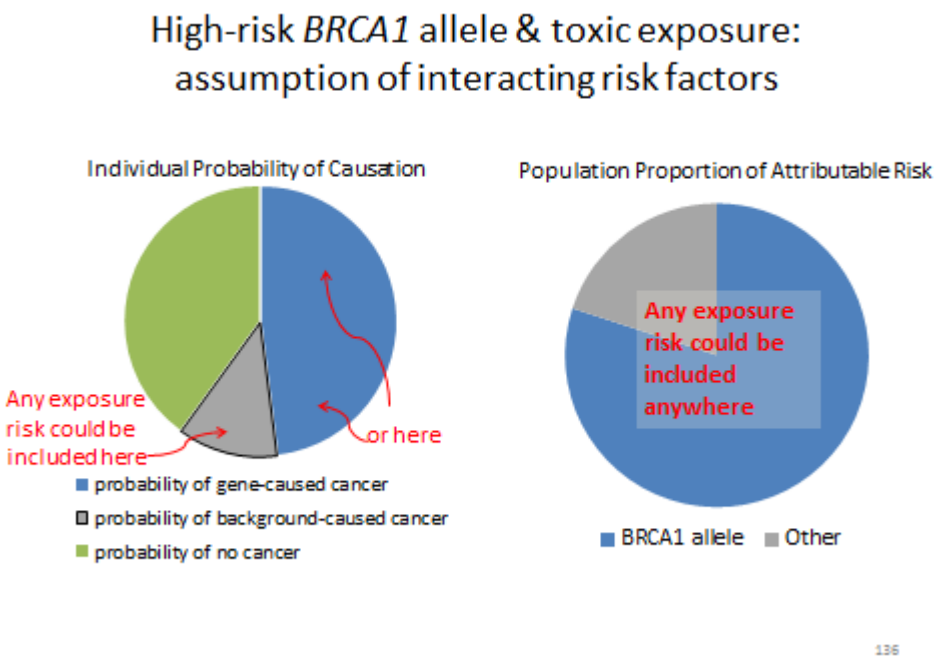


On Slide 135, the pie chart on the left shows the risks faced by an individual woman with the high-risk *BRCA1* allele: a 48% probability of gene-caused breast cancer, a 12% "background" probability of breast cancer caused by some other set of causes, and a 40% probability of not developing breast cancer at all. If the gene-associated risk is independent of any exposure-associated risk, then any risk added by the exposure must be included in the 12% "background" probability—and epidemiologic studies of exposure are really about reducing the share of risk attributed to unknown background causes.

The pie chart on the right on Slide 135 shows the attributable fraction of risk for the subpopulation that has the high-risk allele: the *BRCA1* genotype accounts for 80% of the incidence of breast cancer, and other risk factors account for 20%. If the gene-associated risk is independent of any exposure-associated risk, then any risk added by the exposure must be included in the 20% "other" fraction. Viewed either way, the implication is that epidemiologic studies of breast-cancer risk associated with exposure is an exercise in explaining a portion of the "background" or "other" risk, and it is more probable than not that a randomly selected case of breast cancer, even in someone exposed to the toxin, was caused by the genetic variant rather than by the exposure.

Slide 136 illustrates the possibilities if the risk factors are not independent. If the *BRCA1* gene interacts with the toxic exposure to cause breast cancer, then the added risk from exposure and the added risk from the high-risk allele would overlap. The exposure might add some incremental risk previously characterized as "background" or "other," but some of the risk included in the "*BRCA1* allele" category would *also* properly be characterized as exposure risk. Without knowing the joint effect on risk of exposure and genotype, and comparing that joint effect to the effect of genotype alone, it is impossible to know whether exposure more than doubles the risk of breast cancer for someone who has a high-risk allele. The combination of risk factors could be analogous to the example of tobacco smoking, asbestos exposure, and lung cancer. See Sections IV.A.7.c., V.E., supra.

Slide 136 High-risk *BRCA1* Allele and Toxic Exposure: Assumption of Interacting Risk Factors



Despite the high penetrance of some *BRCA1* and *BRCA2* alleles, many researchers believe that a combination of genetic and environmental factors accounts for most breast cancer, and there is some evidence that the risk associated with these genes may be affected by carcinogens that figure in toxic-tort cases. For example, at least one study found that tobacco smoking is associated with a higher risk of breast cancer in women who have certain *BRCA1* and *BRCA2* genotypes as compared to women with the same genotype who do not smoke. David H. Phillips & Seymour Garte, "Smoking and Breast Cancer: Is There Really a Link?, 17 Cancer Epidemiology," *Biomarkers & Prevention* 1, 2 (2008). On the other hand, there is also some evidence of a purely genetic component to breast cancer incidence. For a review of the complexities, see Nasim Mavaddat et al., "Genetic Susceptibility to Breast Cancer," 4 *Molecular Oncology* 174 (2010).

Among these complexities is the existence of *BRCA1* and *BRCA2* variants, not to mention variants of other genes, that also appear to confer breast-cancer risk, but in much smaller increments. For a review, see Mavaddat et al., supra. For an example, see Paul D. Pharoah et al., "Polygenes, Risk Prediction, and Targeted Prevention of Breast Cancer," 358 *New Eng. J. Med.* 2796 (2008). Genome-wide association studies (GWAS), now using very fast technology called "next generation sequencing," are the usual way that such lower-penetrance genetic variants are discovered. After our discussion of the high-penetrance *BRCA1* alleles, we turn our attention to GWAS.

Students should readily appreciate that the fundamental approach of GWAS is similar to epidemiology: searching for associations between a variable set of attributes and a specified disease. In a GWAS, however, the variable attributes are genetic polymorphisms rather than environmental exposures. There is also an enormous difference in scale, with (as them odule states) literally millions of polymorphisms being examined for association with disease in a GWAS.

Figure V-5 in the module presents a highly conceptual and much simplified image of how the process works: relatively common single nucleotide polymorphisms (SNPs) are identified throughout the genome and statistically tested for association with the incidence of disease. We omit several steps; for more detail, the instructor might want to examine our source, Roelof Koster & Stephen J. Chanock, "Hard Work Ahead: Fine Mapping and Functional Follow-up of Susceptibility Alleles in Cancer GWAS," 2 *Current Epidemiology Rep.* 205 (2015).

The "Manhattan plot" shown in Figure V-5 is a conventional way that researchers display the results of a GWAS. The basic interpretation of such a graph is succinctly described in the module. One detail worth considering is the scale of the *y* axis. As the module says, the *y* axis plots the *statistical* strength of the association between an allele and the disease of interest, rather than directly plotting the magnitude of the relative risk. This ties in to the question we ask in the module about statistical significance standards for GWAS. Recall that the purpose of statistical significance testing is to minimize the probability of a false-positive result. The conventional choice of alpha = 0.05 means that an effect as large as the observed effect would be expected to occur as a result of sampling error no more than 5% of the time. In a GWAS testing of, say, 1,000,000 polymorphisms for association with a disease, a 5% rate of false-positive associations would mean 50,000 "statistically significant" false-positive results! So, as described in the module, GWAS researchers use much smaller values of alpha, on the order of 1 in 1,000,000 (10^{-6}) to 1 in 100,000,000 (10^{-8}). The *y* axis shows the p-value of the statistical test for each polymorphism on a negative-logarithmic scale. Thus, a polymorphism with a 1-in-10 probability of arising by chance ($P=0.1=10^{-1}$) would be plotted at 1 on the *y* axis; a 1-in-100 probability ($P=0.01=10^{-2}$) would be plotted at 2; and a 1-in-100,000,000 chance ($P=10^{-8}$) would be plotted at 8. The dashed horizontal line corresponds to the arbitrarily chosen statistical-significance level—which in our example, is set at $5 * 10^{-8}$, or 5 chances in 10,000,000, a commonly chosen alpha in GWAS research. Even with such stringent significance levels, false positives occur. Samuel P. Dickson et al., "Rare Variants Create Synthetic Genome-Wide Associations," *PLOS Biology*, Jan. 26, 2010, at 5–7 (demonstrating that 179 SNPs could be statistically significantly associated with sickle cell anemia, which is caused by a known variant of a known gene).

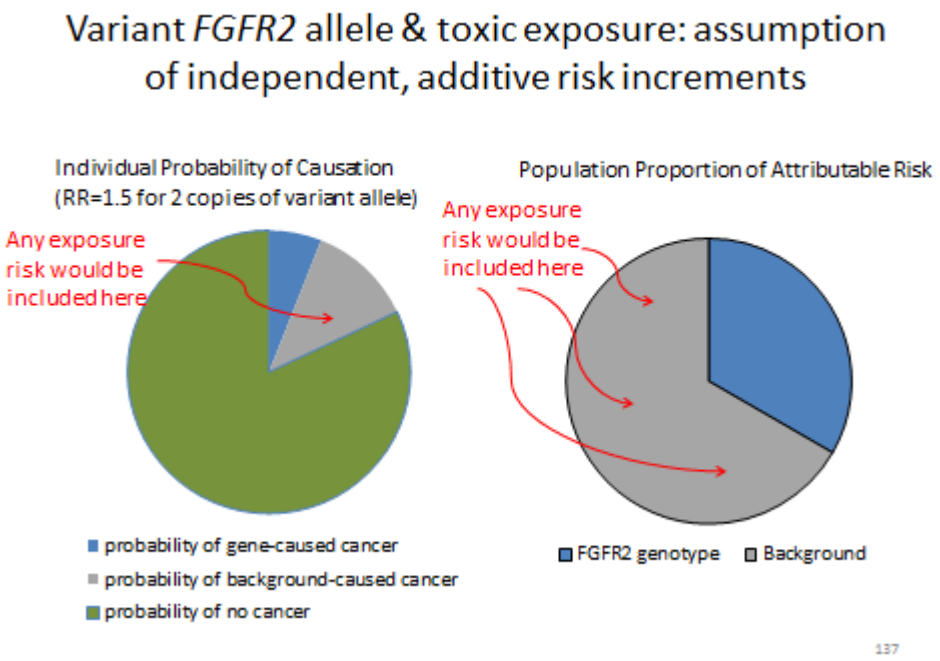
Thus, the assessment of associations for causality is no less important in genomic research than in classical epidemiology. After the initial screening of millions of genes in a GWAS, the subsequent "fine mapping" of risk increments associated with specific genotypes produces odds ratios or other epidemiologic data that are tested using conventional statistical-significance levels. For further discussion, see Koster & Chanock, supra. The review of low-

penetrance breast-cancer-susceptibility alleles discussed in the module is from Pharoah et al., supra.

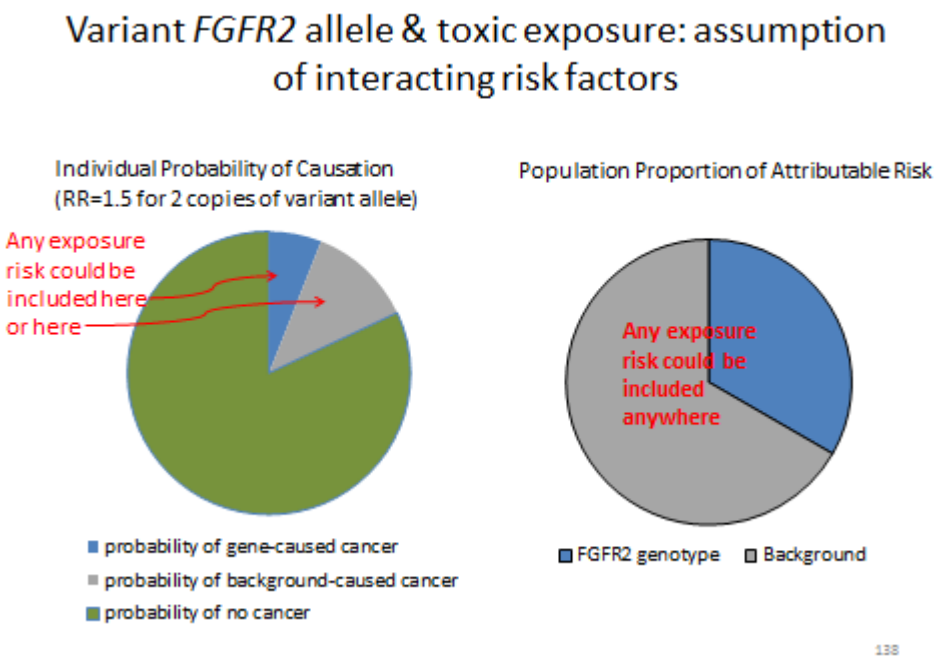
We conclude our discussion of GWAS and low-penetrance risk alleles by imagining that a sick plaintiff who was exposed to a known toxin that more than doubled the plaintiff's risk of disease also was found to carry several alleles that conferred modest increased risk for the same disease. The issues are very similar to those discussed above in connection with the *BRCA1* and *BRCA2* genes. The primary difference is that the lower the penetrance, the more likely that some other genetic or environmental factor explains a significant amount of the incidence of disease.

Slides 137 and 138 illustrate the point in much the same way as did Slides 135 and 136 for the high-penetrance *BRCA1* allele. The slides are based on the discussion of a rare variant allele of the *FGFR2* gene in Pharoah et al., supra, at 2789–2790, with a slight simplification and alteration in the relative risk to make the arithmetic easier. (*FGFR2* codes for fibroblast growth factor receptor 2, a protein involved in regulating cell growth and division.) If the genotype with two copies of the variant allele confers a relative risk of 1.5 and the lifetime risk to those without this genotype is 12%, then an individual with the genotype would have a lifetime risk of 18% and in the sub-population with the genotype, one-third of the risk would be attributable to the genotype. Even if the risk of exposure were independent and additive to the genetic risk, as in Slide 137, the exposure risk could account for more than half of the total risk. This is even more clearly true if the possibility of interaction between genotype and exposure is considered, as in Slide 138.

Slide 137 Variant *FGFR2* Allele and Toxic Exposure: Assumption of Independent, Additive Risk Increments



Slide 138 Variant *FGFR2* Allele and Toxic Exposure: Assumption of Interacting Risk Factors

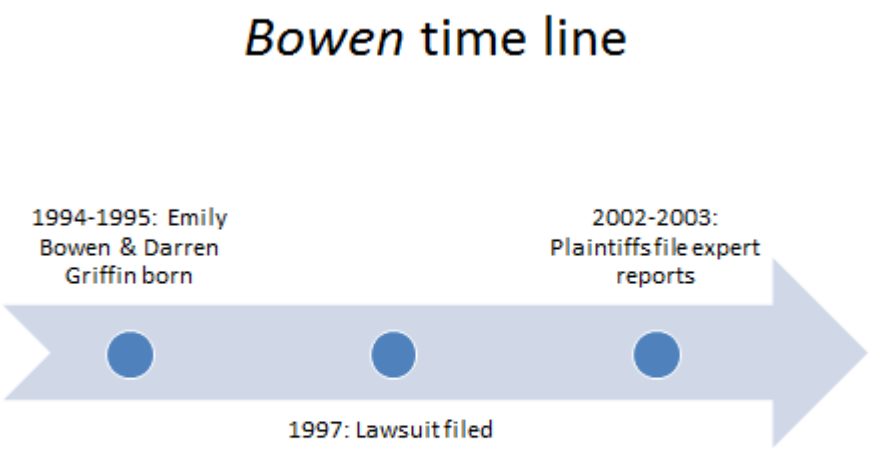


Bowen v. E.I. Du Pont de Nemours & Co., Inc., 906 A.2d 787 (Del. 2006):

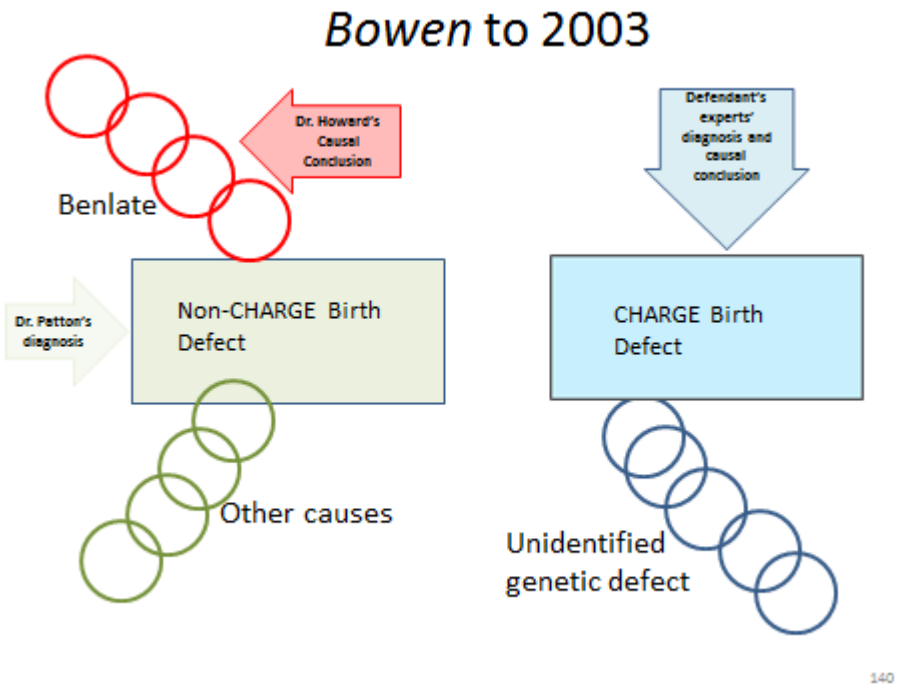
Bowen provides an opportunity for students to consider a decision in which genetics figured as a competing cause. It is a bit different than the hypotheticals we have been considering for two reasons: (1) the evidence for general causation by the allegedly toxic exposure was weak; and (2) initially, the nature and cause of Emily Bowen's condition was disputed.

We find it helpful to work through the way that the opinions of the plaintiffs' experts evolved in tandem with the scientific understanding of CHARGE syndrome. Slide 139 brings us from the birth of Emily Bowen and Darren Griffin to the plaintiffs' initial expert reports. Slide 140 summarizes the competing views of the case at that time. The plaintiffs contended that Emily Bowen did not have CHARGE syndrome and that Benlate was a necessary link in a causal chain that led to her condition. The defendant contended that Emily Bowen had CHARGE syndrome that was caused by some then-unknown causal chain including one or more necessary genetic factors.

Slide 139 *Bowen* Time Line

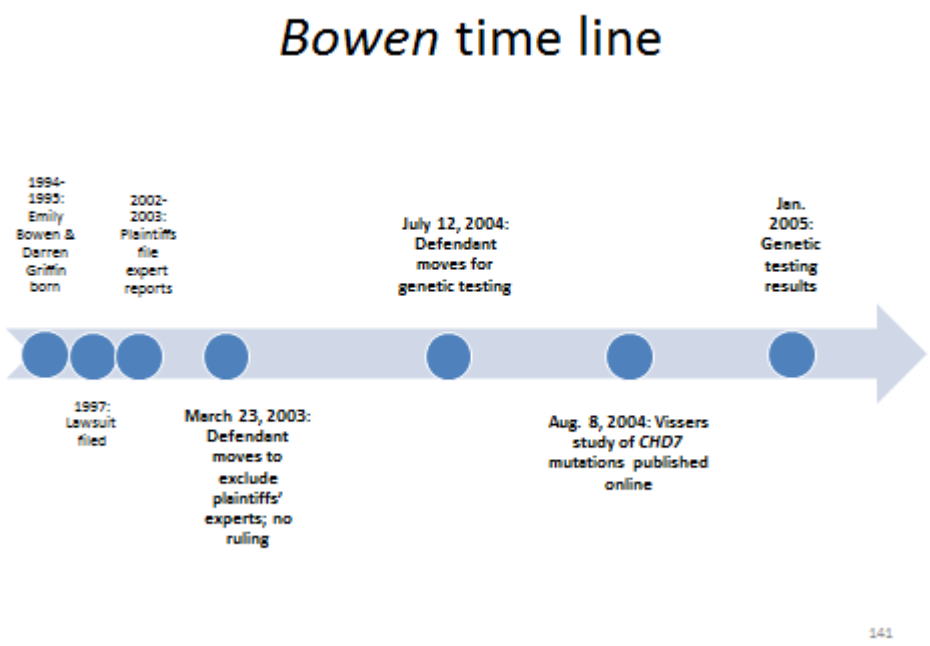


Slide 140 *Bowen* to 2003

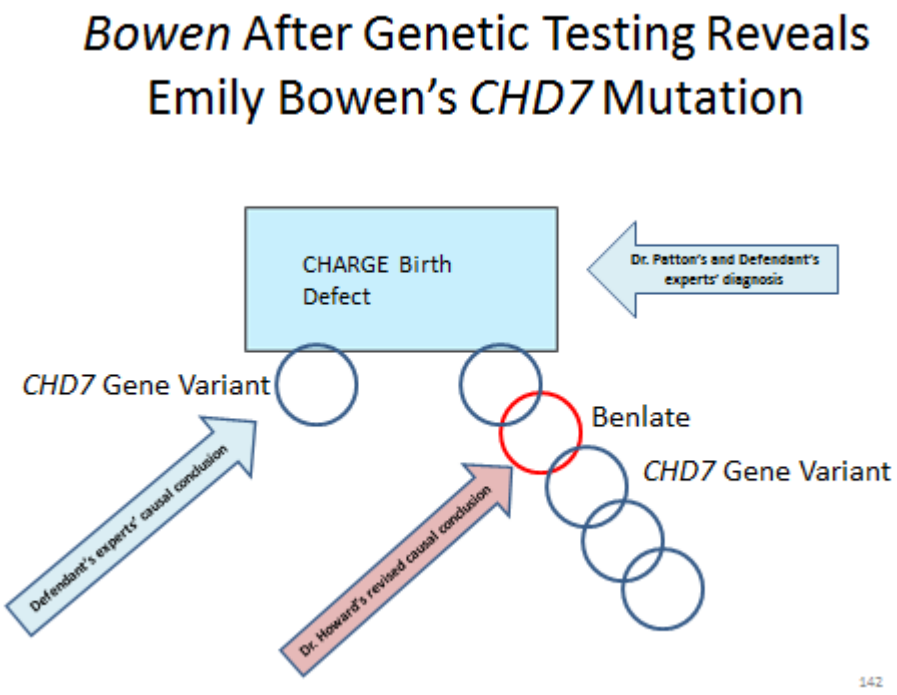


Slide 141 advances the time line through 2005. The timing of the defendant's motion for genetic testing suggests that the defendant (or its experts) knew about the Vissers research, which identified *CHD7* mutations as a cause of CHARGE syndrome, even before the paper was first published online. Slide 142 shows how the experts reacted to the news that Emily Bowen had a *CHD7* mutation linked to CHARGE syndrome. The defendant's experts contended that the genetic variation determined the condition, and that Benlate exposure was irrelevant. Dr. Howard, for the plaintiffs, contended that Benlate exposure and the *CHD7* gene interacted—as part of the same causal chain—to cause Emily Bowen's CHARGE syndrome. Slide 143 presents another possible way of understanding Dr. Howard's amended opinion. Even if Benlate exposure and the *CHD7* mutation were in different causal chains, if each causal chain was sufficient to produce CHARGE syndrome, then both the exposure and the gene would be considered factual causes of the syndrome.

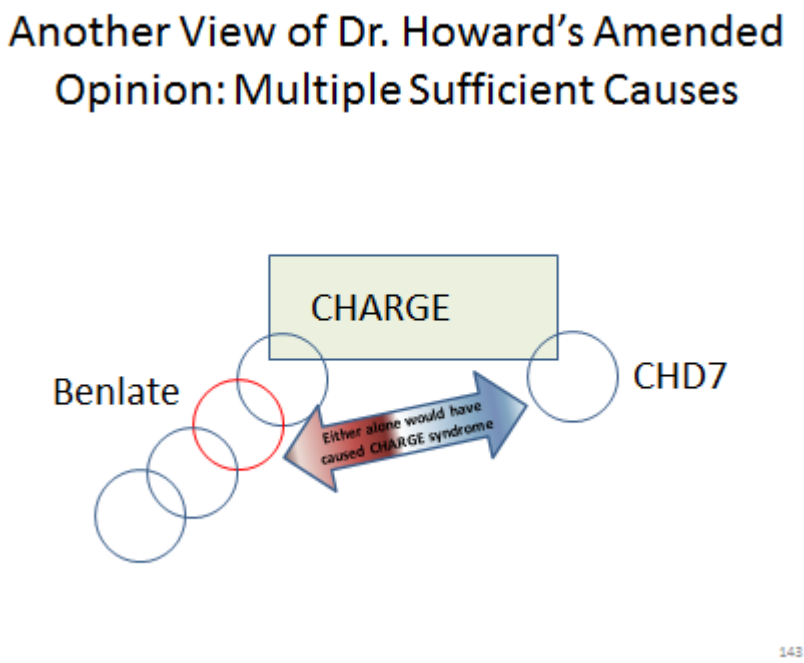
Slide 141 *Bowen* Time Line



Slide 142 *Bowen* after Genetic Testing Reveals Emily Bowen's *CHD7* Mutation



Slide 143 Another View of Dr. Howard's Amended Opinion: Multiple Sufficient Causes



The instructor should challenge students to assess whether the court's exclusion of Dr. Patton's and Dr. Howard's testimony was justified. Dr. Patton conceded that Emily Bowen had CHARGE syndrome, but stated that he could not rule out the possibility that a teratogen, or Benlate specifically, might have been involved in causing the condition. This did not advance the plaintiffs' case, however: Dr. Patton equally could not "rule in" a teratogenic cause. Thus, plaintiffs' causation claim depended on Dr. Howard's opinion that Emily Bowen's *CHD7* mutation and her mother's exposure to Benlate combined to cause Emily's CHARGE syndrome. The problem with that opinion, as succinctly stated by the court, is that Dr. Howard could not "state how or in what percentage or proportion Benlate and the *CHD7* mutation act together to produce CHARGE Syndrome in Emily Bowen" or "explain how, why, or where the *CHD7*/Benlate combination works . . . ." Thus, Dr. Howard had neither mechanistic evidence nor epidemiologic evidence to rely on to support his conclusion that an interaction occurred between the variant *CHD7* allele and the Benlate exposure. This lack of evidence of general causation, combined with a thoroughly convincing demonstration of specific causation by a competing cause, proved fatal to the plaintiffs' case.

The instructor might ask students: which fact most strongly supports the view that the variant-*CHD7* allele was a competing rather than a concurring cause of Emily Bowen's CHARGE syndrome? It is that every person studied to date who had the *CHD7* mutation was diagnosed with CHARGE syndrome. It is rather unlikely that all of them were also exposed to Benlate.

The *CHD7* gene codes for a protein that is believed to help regulate the expression of other genes. Many mutations in *CHD7* result in a truncated and nonfunctional protein, leading to incorrect gene expression profiles during fetal development and the symptoms of CHARGE syndrome. U.S. Nat'l Library of Med., *Genetics Home Reference: CHD7—Chromodomain Helicase DNA Binding Protein* 7 (2016), https://ghr.nlm.nih.gov/gene/CHD7. Nevertheless, "[a]bout one-third of individuals with CHARGE syndrome do not have an identified mutation in the *CHD7* gene. Researchers suspect that other genetic and environmental factors may be involved in these individuals." U.S. Nat'l Library of Med., Genetics Home Reference: CHARGE Syndrome (2016), https://ghr.nlm.nih.gov/condition/charge-syndrome. Of course, because Emily Bowen did have a *CHD7* mutation, there was no need to invoke other genes or environmental factors in her case.

Notes and Questions after *Bowen v. E.I. DuPont de Nemours & Co.*:

1. The Delaware Supreme Court affirmed the Superior Court's grant of summary judgment for the defendant, 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., 906 A.2d 787, 797–98 (Del. 2006).

   The instructor might ask what the outcome should be if Dr. MacIntosh's testimony on exposure were admissible. Would that affect the claim of causation with respect to Emily Bowen? Probably not, because of her high-penetrance genotype that appeared to explain fully what caused her CHARGE syndrome. But what about Darren Griffin? The trial court noted (in footnote 20, in Section IV.F.3. of the module) that his genetic tests did not identify a *CHD7* mutation. So for Darren Griffin's claim, the question would not focus on whether a competing cause had been established, but rather on the scientific evidence underlying Dr. Howard's original opinion that Benlate can cause birth defects similar to CHARGE syndrome and that Benlate did so in these individual cases. Recall from the procedural history that Du Pont moved to exclude Dr. Howard's opinion as to both Darren Griffin and Emily Bowen even before the connection between CHARGE syndrome and *CHD7* was established. That discovery and the results of Emily Bowen's genetic testing made it easier for the court to exclude Dr. Howard's testimony as to her, but the testimony might well have been excluded in any event.

Study Questions for *Bowen v. E.I. Du Pont de Nemours & Co.*:

1. The plaintiffs claimed that Emily Bowen suffered a birth defect because of in utero exposure to Benlate, and genetic testing showed that she had a mutation in the *CHD7* gene. The court did not expressly consider, and plaintiffs' experts did not advocate, the possibility that Benlate caused the *CHD7* mutation itself. What evidence might help exclude that possibility?

   One way would be to test Emily's parents to see if she could have inherited the variant allele from one or both of them. Another would be to compare the timing of the exposures with the timing of Emily's conception and development. A somatic mutation caused by a mutagen after conception would be present only in cells derived from the original mutated cell, not in all the cells of Emily's body. Presumably, there were good reasons that plaintiffs' experts eschewed this theory.

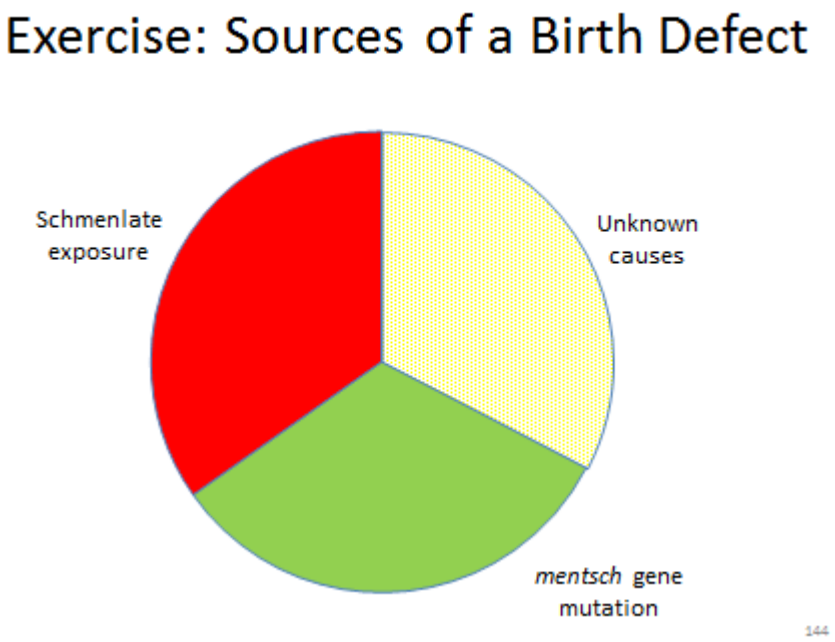
2. The court asserted that Dr. Howard was not qualified to give an opinion about the interaction of the *CHD7* gene and Benlate exposure because of his "lack of expertise and/or qualification as a geneticist." Does it follow that only a geneticist could give such an opinion?

   The types of studies that find interactions among genotypes, exposures, and diseases require methods of genetics or genomics, toxicology, and epidemiology. A testifying expert need not necessarily be an expert in all of those fields. In a case-control study of genotypes and disease risk, for example, an epidemiologist would be fully qualified to interpret the odds ratio, even if the epidemiologist had to rely on a geneticist for the results of the genetic analysis.

3. Exercise: Suppose that repeated epidemiologic studies showed that for the children of mothers exposed to the teratogenic substance Shmenlate, the attributable proportion of risk for a certain birth defect was as follows: one-third attributable to Shmenlate exposure, one-third attributable to inherited mutations in the *mentsch* gene, and one-third attributable to unknown causes. Assume a child is born with the birth defect in question, and it can be proven that the mother was exposed to a dose of Shmenlate sufficient to cause the defect. How would knowing the child's genotype affect your assessment of the probability of specific causation?

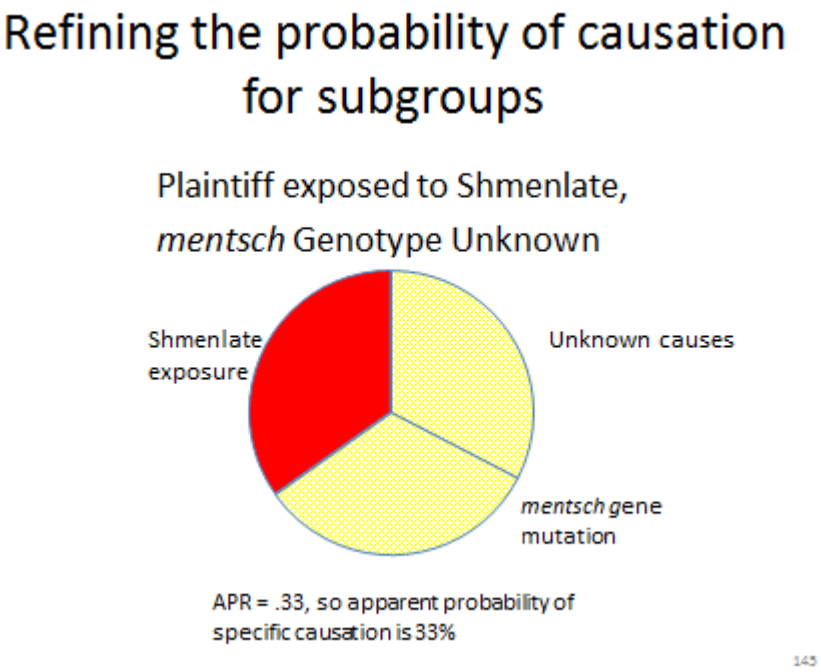
   We work through this exercise in Slides 144 through 147. The pie chart on Slide 144 sets up the problem by showing the attributable proportion of risk for Shmenlate exposure, the *mentsch* gene mutation, and unknown causes: one-third for each.

Slide 144 Exercise: Sources of a Birth Defect



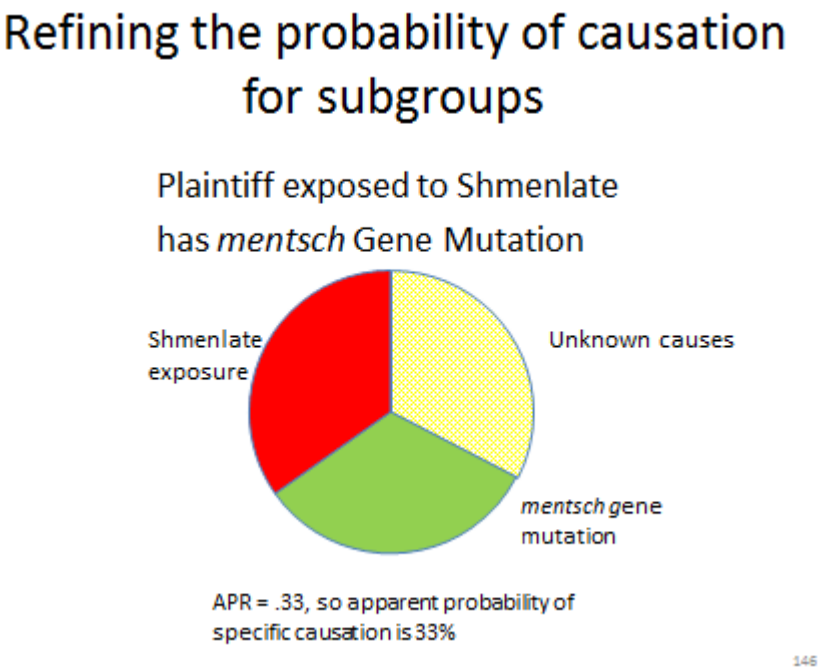
If we do not know the child's genotype for the *mentsch* gene, we cannot do any better than this: it is as if the genetic source of the birth defect is lumped in with the unknown causes, as depicted by the use of color on Slide 145. If we infer from the epidemiologic data to specific causation, we would conclude that the probability of causation by Shmenlate exposure is 0.33.

Slide 145 Refining the Probability of Causation for Subgroups



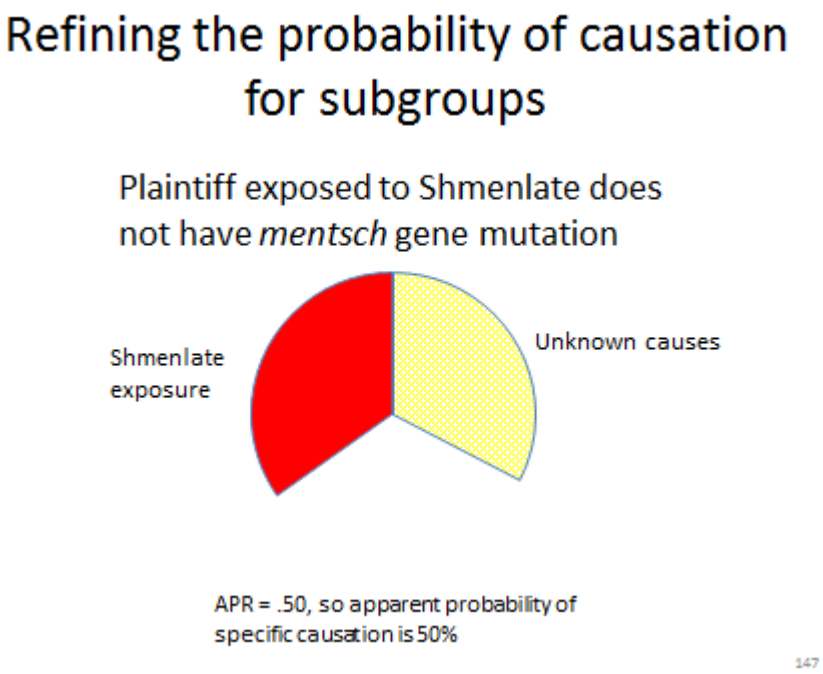
Slide 146 depicts the situation if we know that the child has a *mentsch* gene mutation. Again, there are three possible sources of the child's birth defect (Shmenlate, *mentsch*, and unknown), each of which contributes an equal amount of risk. And once again, the inferred probability of specific causation would be 0.33.

Slide 146 Refining the Probability of Causation for Subgroups



Slide 147 shows how the situation changes if we know that the child does not have a *mentsch* gene mutation. With the genetic risk factor ruled out, we can deduce that the child's birth defect could have been caused only by Shmenlate or by unknown causes. These two risk factors confer equal amounts of risk, so in the subpopulation with this genotype, the attributable proportion of risk for Shmenlate is 0.50, not 0.33. Inferring from the population risk data to specific causation would put the claim on the cusp of the "more likely than not" standard. Obviously, the hypothetical data could have been tweaked so as to satisfy the standard.

Slide 147 Refining the Probability of Causation for Subgroups



Another valuable exercise would be to ask students to "reverse engineer" what the epidemiologic studies might have looked like that produced the risk profile presented in the hypothetical. The key is for students to realize that the pie charts show that the risks of Shmenlate exposure and the *mentsch* gene mutation are independent and additive: these two risk factors do not interact, but rather, each confer one-third of the total population risk.

Imagine a cohort study comparing children whose mothers were exposed to Shmenlate and children whose mothers had no exposure. Assuming that the *mentsch* mutation is distributed randomly across exposed and unexposed people, we would expect the study to show a relative risk of 2.0 for Shmenlate exposure (equivalent to a 50% APR). Similarly, assuming that Shmenlate exposure is distributed randomly across those with and without *mentsch* mutations, we would expect a study to show a relative risk of 2.0 for children with *mentsch* mutations as compared to children without *mentsch* mutations.

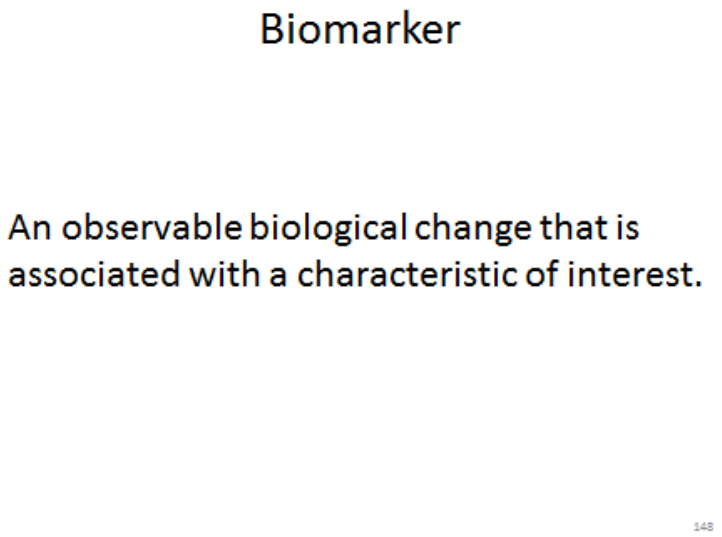
What if an epidemiologist had studied Shmenlate exposure and *mentsch* mutations simultaneously? As compared to people with neither Shmenlate exposure nor *mentsch* mutations, we would expect that study to show relative risks of 2.0 for those with Shmenlate exposure only, 2.0 for those with *mentsch* mutations only, and 3.0 for those with both Shmenlate exposure and *mentsch* mutations. These relative risks are equivalent to the APRs shown in the pie charts. Among those with Shmenlate exposure and no *mentsch* mutations, half of the birth defects would be attributed to the exposure and half to unknown causes; among those with *mentsch* mutations and no Shmenlate exposure, half of the birth defects would be

attributed to the mutation and half to unknown causes; among those with both Shmenlate exposure and *mentsch* mutations, two-thirds of the birth defects would be attributed to the exposure or the mutation and one-third to unknown causes. Instructors may wish to remind students of the formula for APR given in Section III.D.3., supra.

Section V.G. Using Toxicogenomics and Biomarkers as Proof of Specific Causation

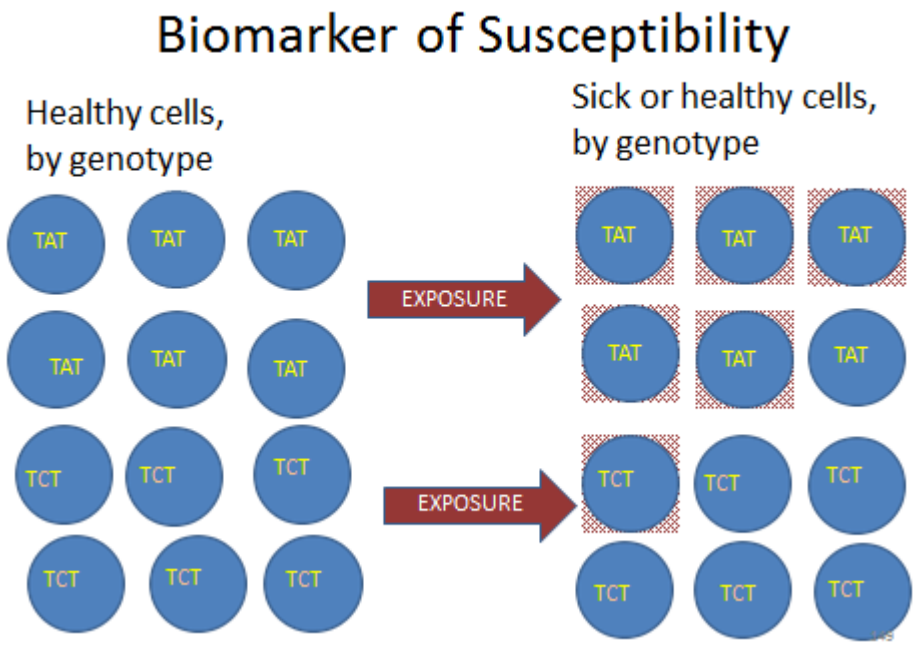
The introductory material describing toxicogenomics and the search for biomarkers is technical, but fairly self-explanatory. We provide some slides to aid any exposition the instructor chooses to give, beginning with Slide 148, which defines "biomarker" for our purposes.

Slide 148 Biomarker



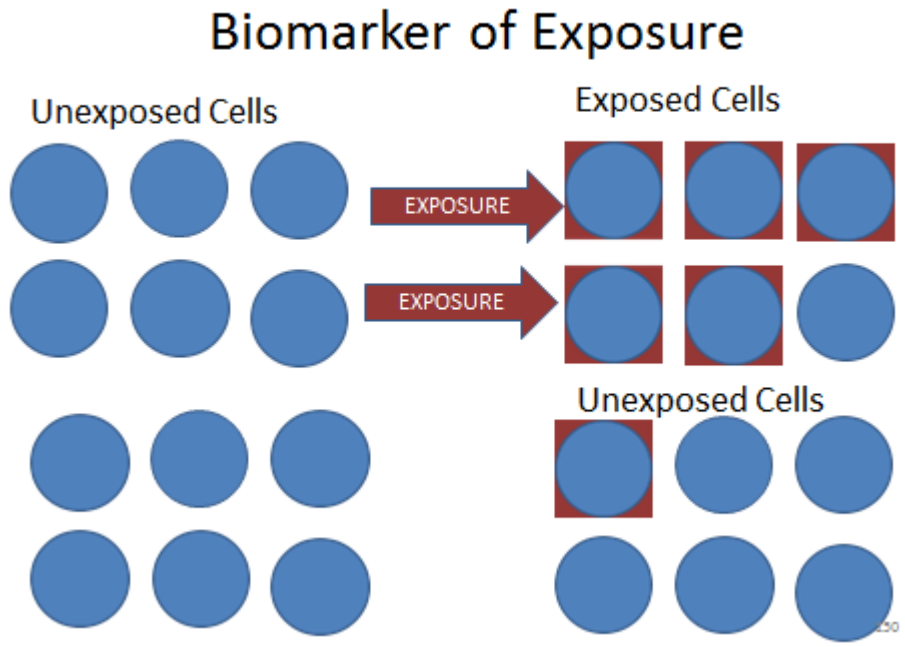
Slide 149 illustrates a biomarker of susceptibility. Sets of six cells of each of two genotypes (TAT or TCT, representing a single nucleotide polymorphism in which an adenine base is replaced by a cytosine base) are exposed to a toxin. Five of the six TAT cells become sick (indicated by a patterned border), but only one of the six TCT cells does. Exposed cells with the TAT genotype are much more likely to develop disease after exposure. The TAT genotype is a biomarker of susceptibility to the toxin.

Slide 149 Biomarker of Susceptibility



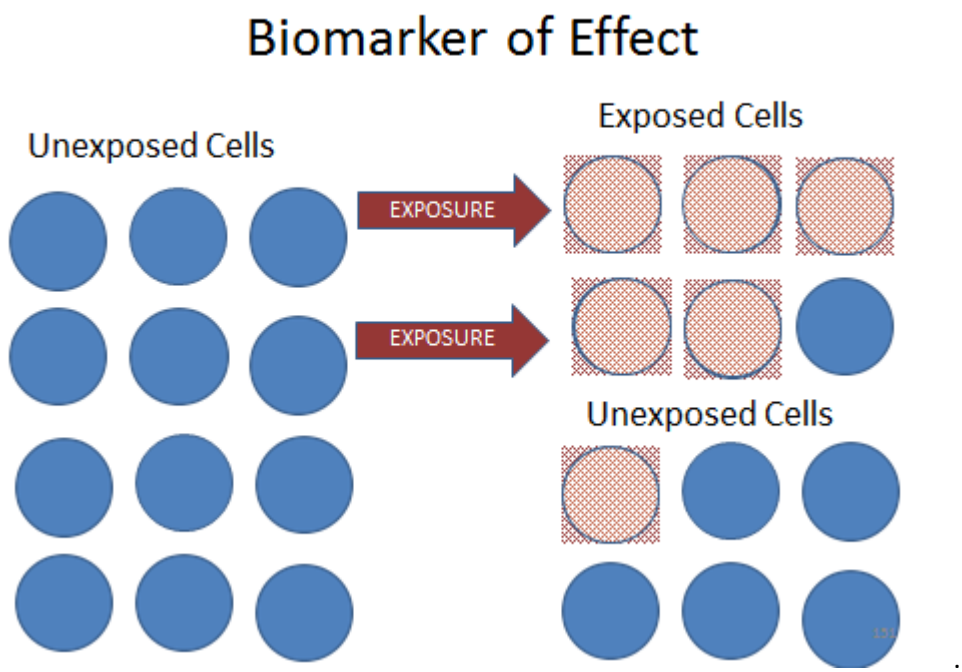
Slide 150 illustrates a biomarker of exposure. Six cells are exposed to a toxin. Afterward, five of the cells display a change (indicated by a solid border). Meanwhile, of the six cells that remain unexposed (the control group), five do not display the change. The change represented by a solid border is much more likely to occur in cells that have been exposed to the toxin. The change is a biomarker of exposure.

Slide 150 Biomarker of Exposure



Slide 151 illustrates a biomarker of effect. Six cells are exposed to a toxin. Afterward, five of the six are not just changed, but harmed (indicated by a patterned border and patterned fill). Meanwhile, five of the six cells that remain unexposed do not display the harm. (The harm might take a range of forms: the cell might become malignant, might have altered gene expression or other biochemical changes that impede the cell's proper function, might have DNA damage that is a step on a pathway toward malignancy, etc.). The harmful change indicated by the patterned border and fill is much more likely to occur in exposed cells than in unexposed cells. The harmful change is a biomarker of effect.

Slide 151 Biomarker of Effect

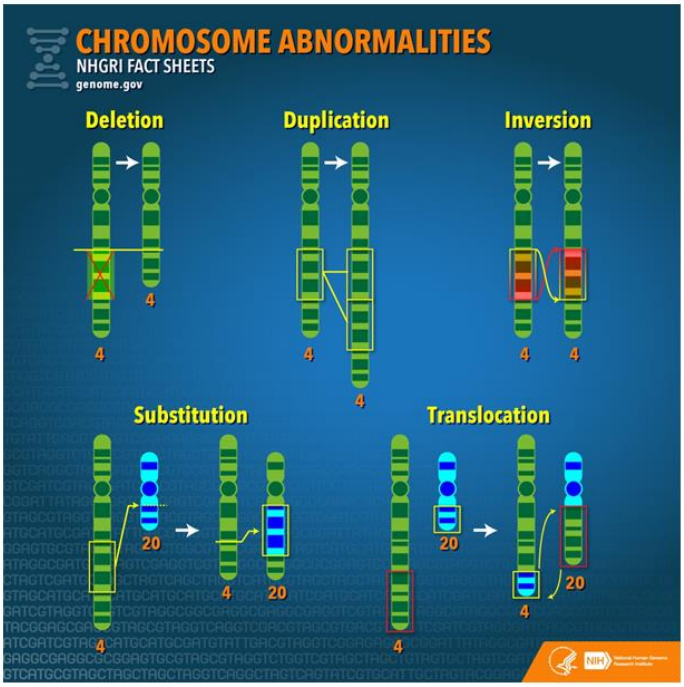


Of course, Slides 149 through 151 assume implicitly that the biomarkers are valid and the results are statistically significant. The module then gives some examples of potential biomarkers of exposure or effect. To make these examples less abstract (although still schematic), Slides 152 through 156 illustrate some of the ways these biomarkers can manifest.

Slide 152, using an illustration from the National Human Genome Institute (which can be found at http://www.genome.gov/images/content/Chromosome_abnormalities_factsheet.jpg (last visited Mar. 22, 2016)), provides examples of chromosome aberrations, using standard representations of chromosome 4 and chromosome 20. In the top row, only chromosome 4 is altered: a chunk of it is missing (deletion), repeated (duplication), or turned upside down (inversion). In the bottom row, a piece of chromosome 4 is removed and inserted into chromosome 20 (substitution) or a piece of each chromosome is switched to the other (translocation). Other chromosome aberrations may involve having more or fewer than the usual two copies of one or more chromosomes. For more information, see U.S. Nat'l Library of Med., *Genetics Home Reference: Can Changes in the Number of Chromosomes Affect Health and Development?* (2016), https://ghr.nlm.nih.gov/handbook/mutationsanddisorders/chromosomalconditions; and Nat'l Human Genome Research Inst., Chromosome Abnormalities, http://www.genome.gov/11508982#6 (last updated Jan. 6, 2016).

Slide 152 Potential Biomarkers of Exposure or Effect: Chromosome Abnormalities

Potential biomarkers of exposure or effect: chromosome abnormalities

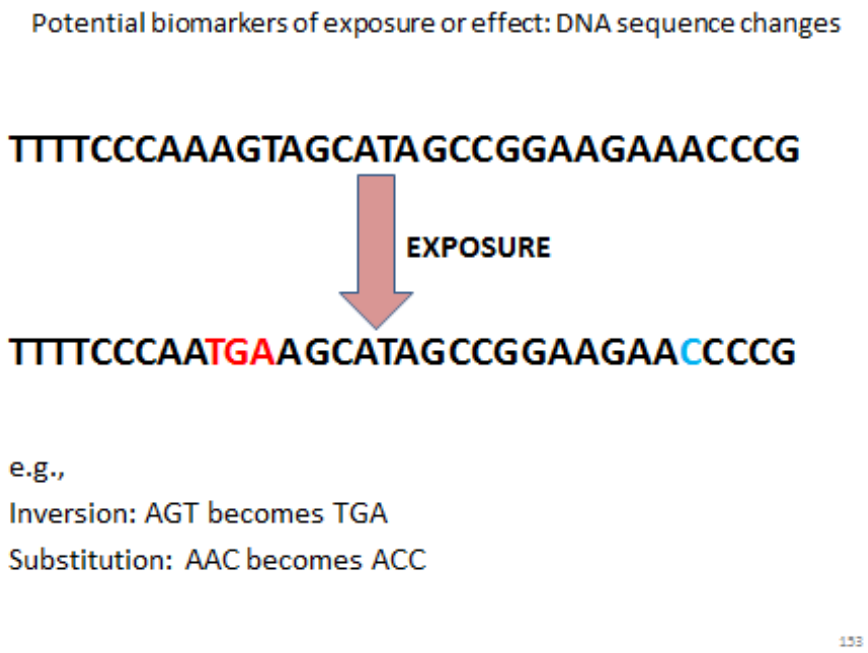


152

SOURCE: Image courtesy National Human Genome Institute, www.genome.gov. Image in the public domain.

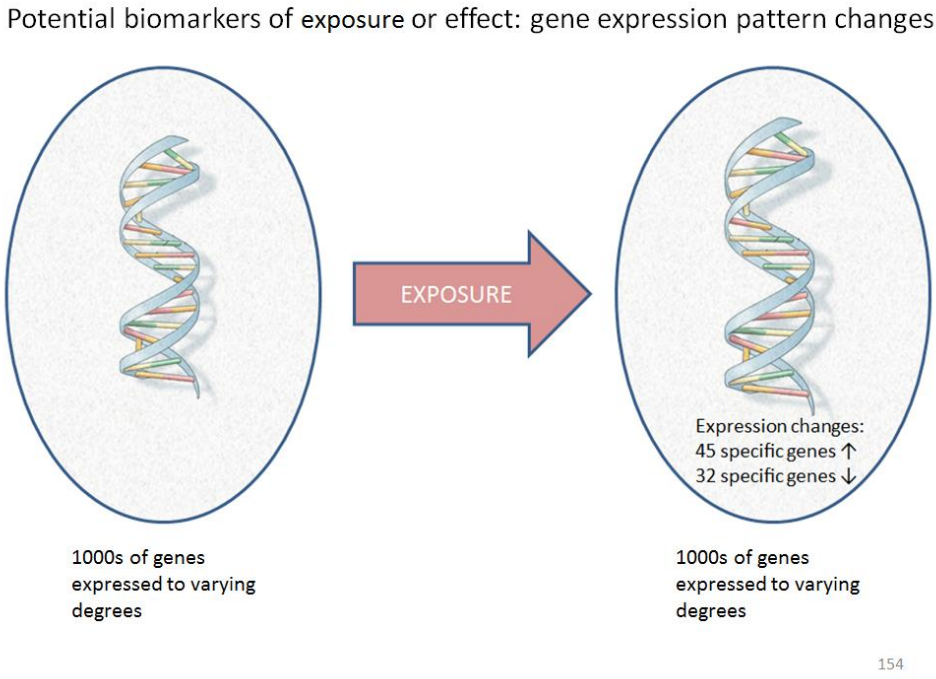
Slide 153 illustrates two of the many types of alterations of DNA sequence that might be caused by a toxic agent. Reading from left to right, the top chain represents 33 DNA bases constituting 11 genetic code "words" (codons). Exposure alters this bit of DNA in two ways. In an inversion, shown in red, an AGT codon is flipped to read TGA. This particular change would replace a codon for the amino acid serine with a "stop" codon that ends transcription instead of adding an amino acid. The resulting protein would be truncated and probably nonfunctional. In a substitution, a single adenine base is replaced with cytosine. The codon changes from AAC to ACC, and the coded amino acid changes from asparagine to threonine, with possible effects on the protein's structure or function. Many other ways of altering DNA on a subchromosomal scale are possible.

Slide 153 Potential Biomarkers of Exposure or Effect: DNA Sequence Changes



Slide 154 addresses changes in gene-expression pattern. Researchers are now able to measure simultaneously the degree to which thousands of genes are expressed in a cell or tissue sample, as suggested by the schematic cell on the left. Exposure to a toxic agent may cause some genes to be expressed more or less than they otherwise would be. In our example, the exposed cell on the right has 45 specific genes that are expressed statistically significantly more than in the unexposed cell, and 32 specific genes that are expressed statistically significantly less. The numbers chosen, of course, are arbitrary. Students should understand that the potential biomarker is not simply the number of genes with increased or decreased expression, but the pattern of *which* groups of genes have increased or decreased expression.

Slide 154 Potential Biomarkers of Exposure or Effect: Gene Expression Pattern Changes



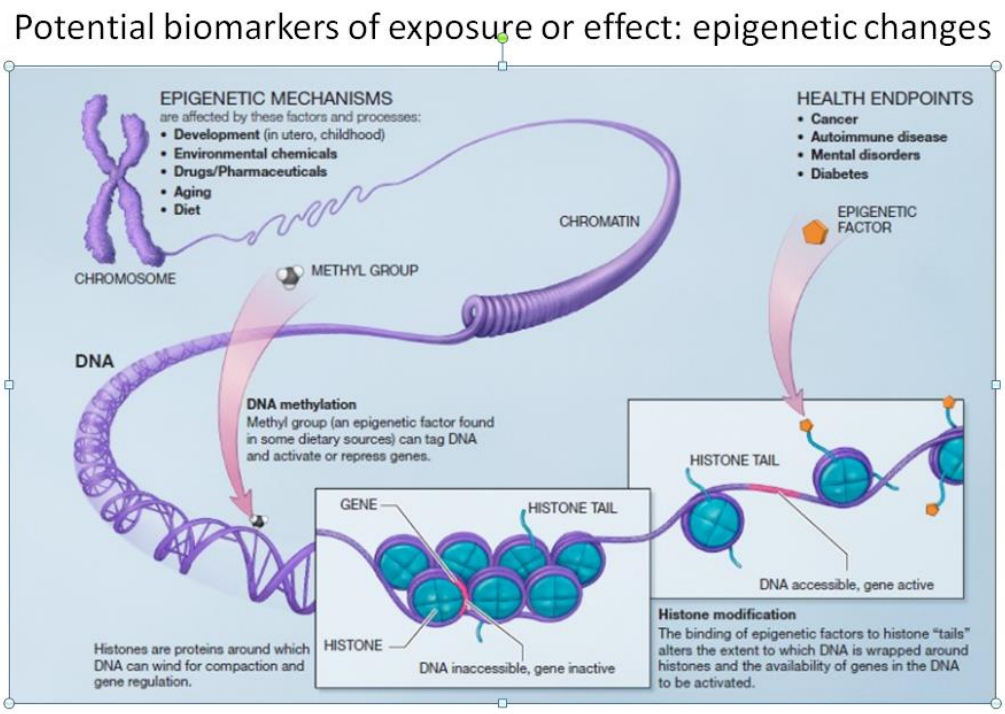
SOURCE: Image of DNA Molecule courtesy U.S. National Library of Medicine. Image in the public domain.[21] The remaining material in the slide is courtesy of the authors.

Slides 155 and 156 illustrate some examples of the "other biochemical constituents" that might serve as biomarkers of exposure or effect, focusing on one subset of biochemical constituents: epigenetic factors. Slide 155, from the Nat'l Ctr. for Biotech. Info., Epigenomics Scientific Background: Figure 1, http://www.ncbi.nlm.nih.gov/books/NBK45788/pdf/Bookshelf_NBK45788.pdf (last updated Jan. 20, 2011) shows two of a number of known epigenetic mechanisms: (1) DNA methylation (in which a small organic chemical structure is added to the "backbone" portion of one or more cytosine bases), and (2) modification of the histone proteins with which DNA is associated in chromosomes. Either of these types of changes can affect the degree to which a specific segment of DNA is physically accessible to be transcribed by messenger RNA and therefore can affect the degree to which the gene is expressed (translated into protein products). Slide 156, parts of which are adapted from Figure 2 in the same source, illustrates how exposure can change the degree of methylation. Before exposure to the toxic agent, some methyl groups are

[21] http://ghr.nlm.nih.gov/handbook/basics/dna, visited Feb. 4, 2016.

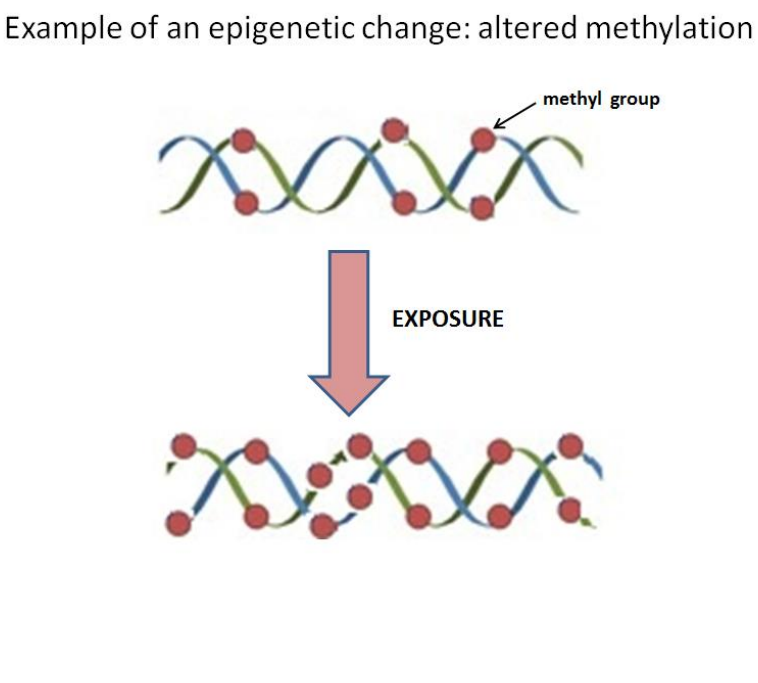
attached to this segment of DNA. After exposure, many more methyl groups are attached. The effect on gene expression will depend on the normal function of this segment of DNA. Note that an exposure may increase or decrease methylation in different parts of the genome.

Slide 155 Potential Biomarkers of Exposure or Effect: Epigenetic Changes



SOURCE: Image courtesy National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, www.ncbi.nlm.nih.gov. Image in public domain.

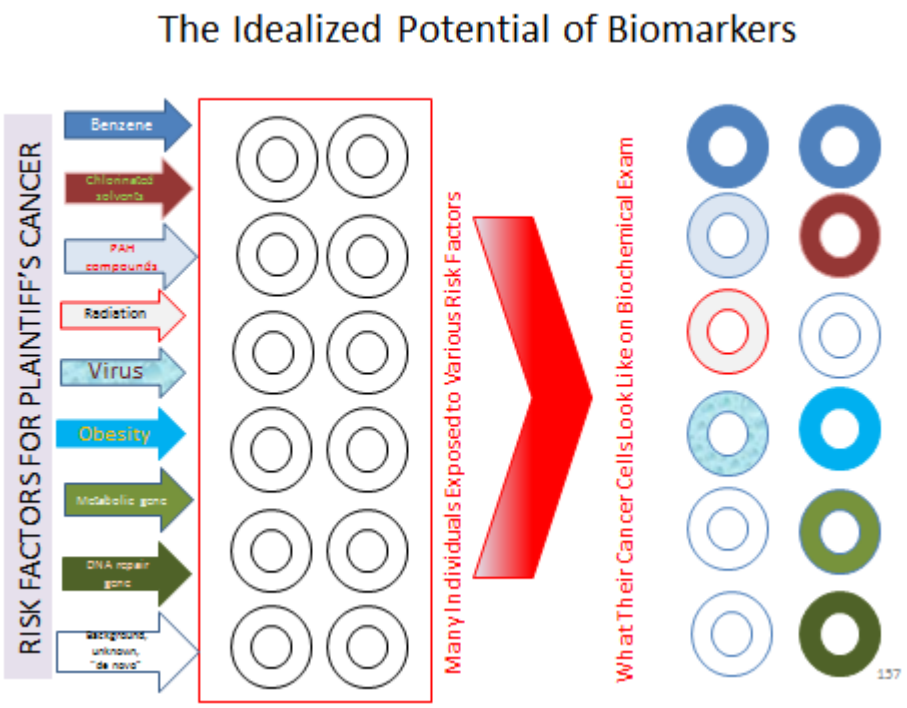
Slide 156 Example of an Epigenetic Change: Altered Methylation



SOURCE: Portions of adapted image courtesy National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, www.ncbi.nlm.nih.gov. Image in public domain.

The idealized end point of toxicogenomic research is the end of specific-causation disputes. This ideal is depicted in Slide 157. After presenting it, we explore the reasons to believe that this ideal may seldom be realized, even though toxicogenomics and molecular epidemiology are likely to provide important evidence relevant to specific causation.

Slide 157 The Idealized Potential of Biomarkers



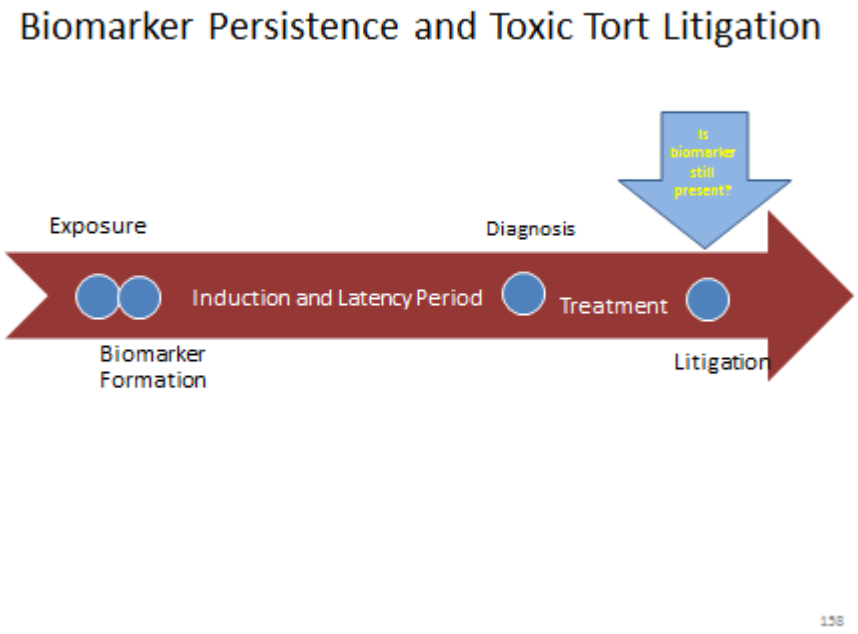
The module invites students to ponder: "What questions would you want answered about these types of studies if a party tried to use them in a toxic-tort case?" With a little reflection, students should see that the types of questions that must be asked are very familiar. How accurate and precise were the measurements of the exposures and the putative biomarkers? Is there a biological basis for assuming that the biomarker and the exposure are causally related to each other? Are the results of in vivo animal tests (often done on genetically identical animal models) applicable to humans (often highly genetically diverse with respect to potential toxic modes of action)? Were the dosages applied in in vivo or in vitro studies similar to doses experienced by exposed people, and if not, can the effect of different doses be validly extrapolated? What is the dose-response relationship? Were the individuals in the exposed and unexposed groups correctly identified? Were there any biases in the study? Were potential confounding factors accounted for? Can the results of the study of exposed individuals be validly extrapolated to the plaintiff? Have the study's results been replicated, or at least the study's conclusions been supported, by other studies? In other words, students should see that although toxicogenomics and molecular epidemiology operate at a genomic and molecular scale, they are still toxicology and epidemiology—and the same methodological issues still apply.

Students might generalize their questions by asking: "are the studies that identify the biomarker internally and externally valid with respect to the use of the biomarker to help

establish specific causation in this case?" We discuss internal and external validity in sections IV.A.7.d. and V.C., supra, and neither we nor the instructor need rehash that discussion in detail here. For the instructor interested in pursuing the topic of validity with particular reference to biomarkers, two good sources are Paul A. Schulte & Frederica P. Perera, *Validation, Molecular Epidemiology* 79 (Paul A. Schulte & Frederica P. Perera eds., 1st ed. 1998); and Nada Majkic-Singh, "Biomarkers: From Standardization to Performance," 30 *J. Med. Biochemistry* 183 (2011). Sources indicating that it has been difficult to validate biomarkers include Patrick M.M. Bossuyt, "Defining Biomarker Performance and Clinical Validity," 30 *J. Med. Biochemistry* 193 (2011); and Martin Latterich & Jan E. Schnitzer, "Streamlining Biomarker Discovery," 29 *Nature Biotech.* 600 (2011).

Students should, however, focus on the particular validation issues discussed in the module that may affect not just the clinical use of biomarkers, but especially their use as evidence of specific causation in toxic-tort cases. The first of these is persistence and durability. Recall that the legal inquiry is backward-looking: a plaintiff with disease presents himself or herself to the court and seeks to prove that a past event or series of past events caused the disease. The biochemical changes that may serve as biomarkers of effect or exposure may arise contemporaneously with, or soon after the exposure. In toxic torts, considerable time may elapse between exposure and disease manifestation and litigation. Will the biomarker, even if valid, still be detectable in the plaintiff's tissue when the legal system needs it, years after exposure? Slide 158 presents this issue by showing exposure, biomarker creation, disease manifestation, and litigation on a time line.

Slide 158 Biomarker Persistence and Toxic Tort Litigation



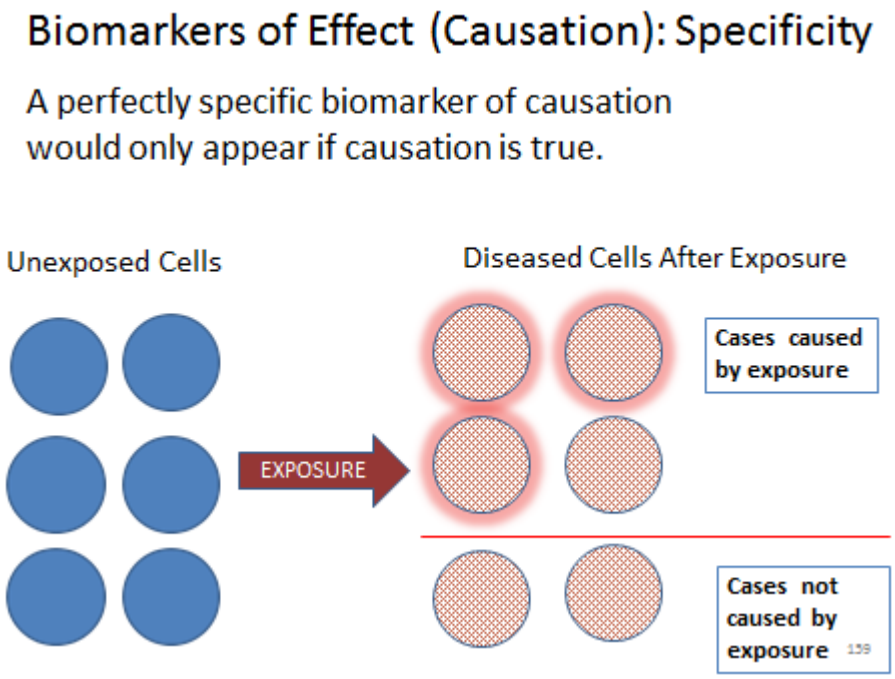
This issue may be significant, particularly for certain kinds of biomarkers. For example, some potential carcinogens and teratogens react chemically with DNA to form DNA "adducts" (shorthand for "addition products"). In principle, the presence or absence of DNA adducts known to be formed by a particular toxin could provide important evidence of whether or not that toxin caused the DNA damage that led to a plaintiff's disease. At least some studies, however, have shown that adducts have short lifespans. See, for example, Kimiko Hori et al., "Stability of Acetaldehyde-derived DNA Adduct in Vitro," 423 *Biochemical & Biophysical Res. Comm.* 642 (2012).

Even if a biomarker were long-lived enough to be available to litigants, questions would remain. Does the marker's presence prove causation (i.e., effect), or merely exposure (with something else having caused the disease)? Does the marker's absence prove the lack of causation? The concepts of specificity and sensitivity help address these questions.

Specificity measures how often the biomarker appears in the absence of the phenomenon it is supposed to mark. In the context of toxic causation, a perfectly specific biomarker would appear only in cases of disease that were in fact caused by the exposure of interest. The less specific the biomarker, the more likely it is that the presence of the biomarker will indicate a false-positive causation result. Slide 159 presents the idea graphically. After exposure to a toxic substance, the cells of six individuals display disease (red patterned fill). The cells are examined for the presence of a biomarker (red glow). Four of the six cases were

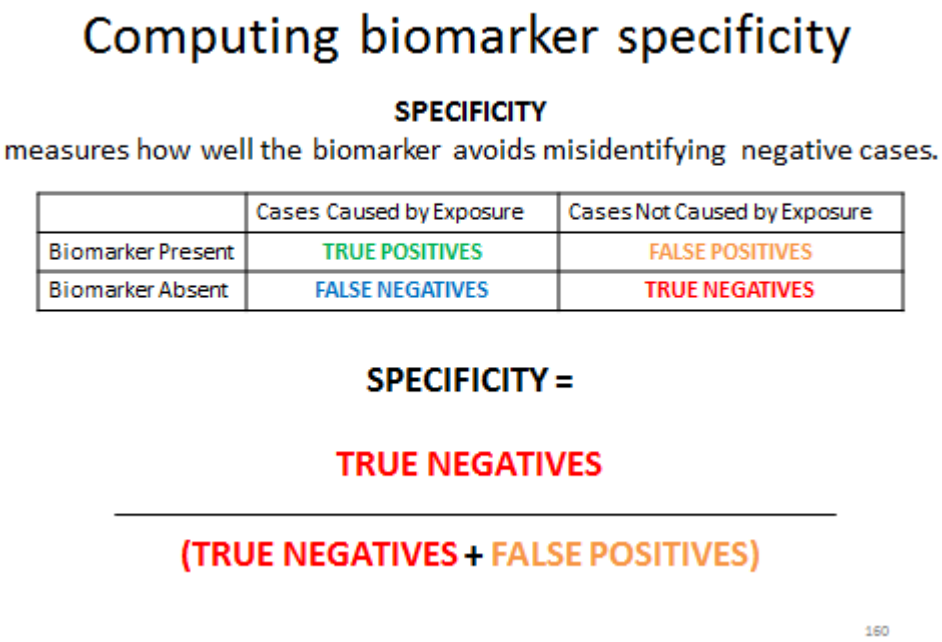
caused by the exposure (top four cells). The biomarker is not found in either of the two cells that have disease *not* caused by the exposure; therefore, the biomarker has perfect (1.00) specificity. Students might notice that the biomarker is *not* perfectly sensitive: only three of the four cases of causation display the biomarker. Nevertheless, the biomarker is perfectly sensitive: it gives no false positives.

Slide 159 Biomarkers of Effect (Causation): Specificity



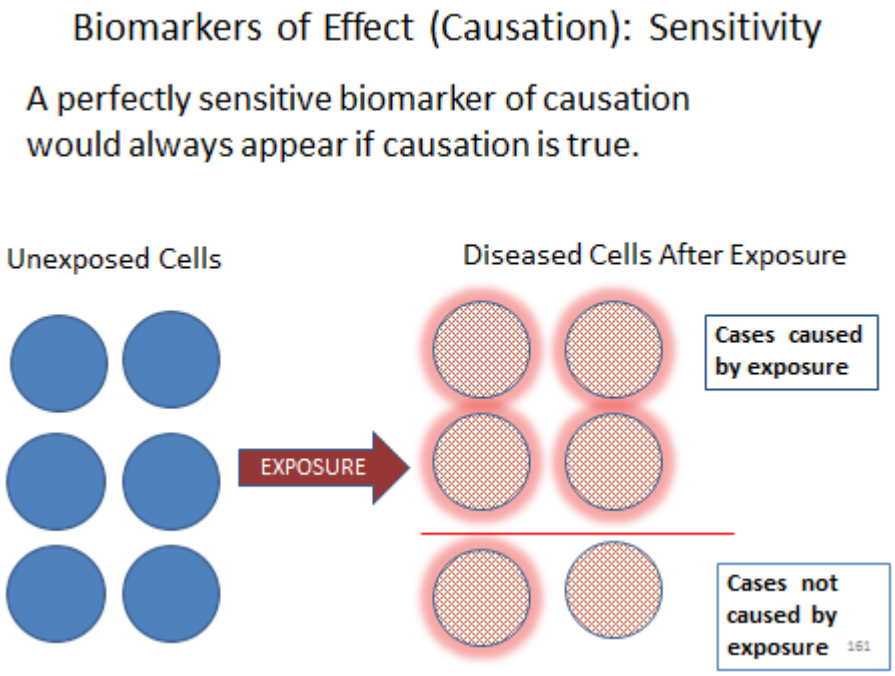
Slide 160 presents the formula for computing biomarker specificity. Specificity is the fraction of all negative cases that are marked by the absence of the biomarker. One way to express this mathematically is: specificity = [(true negatives) / (true negatives + false positives)].

Slide 160 Computing Biomarker Specificity



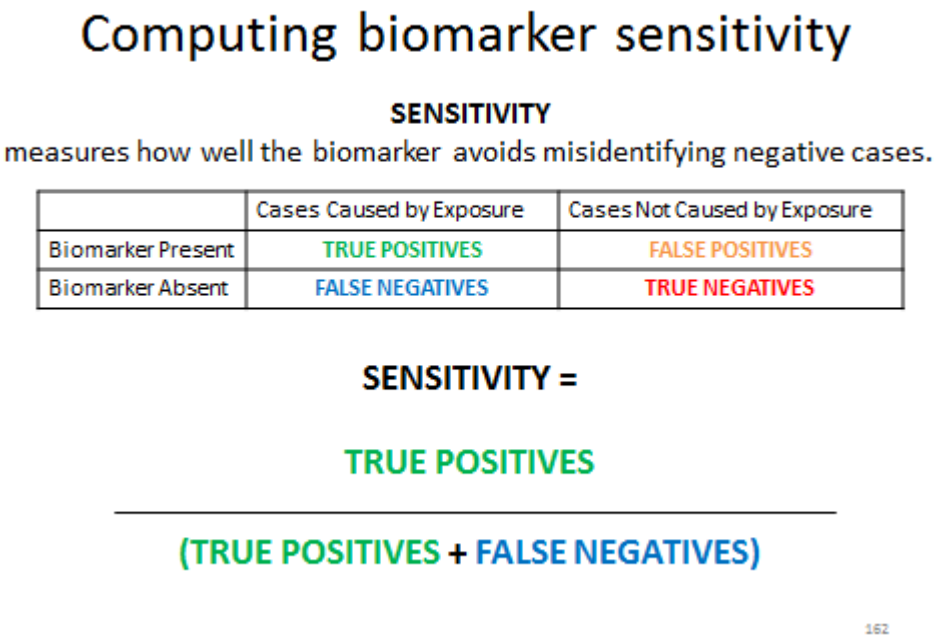
Sensitivity measures how often the biomarker fails to appear in the presence of the phenomenon it is supposed to mark. In the context of toxic causation, a perfectly sensitive biomarker would appear in every case of disease that was in fact caused by the exposure of interest. The less sensitive the biomarker, the more likely it is that the absence of the biomarker will indicate a false-negative causation result. Slide 161 presents the idea graphically. After exposure to a toxic substance, the cells of six individuals display disease (red patterned fill). The cells are examined for the presence of a biomarker (red glow). Four of the six cases were caused by the exposure (top four cells). The biomarker is found in all four cells that have diseased *caused* by the exposure; therefore, the biomarker has perfect (1.00) sensitivity. Students might notice that the biomarker is *not* perfectly specific: one of the two cases of non-causation displays the biomarker. Nevertheless, the biomarker is perfectly sensitive: it gives no false negatives.

Slide 161 Biomarkers of Effect (Causation): Sensitivity



Slide 162 presents the formula for computing biomarker sensitivity. Sensitivity is the fraction of all positive cases that are marked by the presence of the biomarker. One way to express this mathematically is: sensitivity = [(true positives) / (true positives + false negatives)].

Slide 162 Computing Biomarker Sensitivity



To give students some practice computing specificity and sensitivity, the instructor might ask students calculate the sensitivity of the biomarker illustrated on Slide 159, the specificity of the biomarker illustrated on Slide 161, and both for the biomarkers illustrated on Slide 150 or 151. On Slide 159, the biomarker gives one false negative out of four true cases of causation; the sensitivity is three-fourths, or 0.75. On Slide 161, the biomarker gives one false positive out of two cases of no causation; the specificity is one-half, or 0.50. On Slide 150 or 151, the students could complete the results table below and compute the specificity and sensitivity.

	Positive for Result of Interest (Exposure or Effect)	Negative for Result of Interest (Exposure or Effect)
Biomarker Present	5 True positives	5 True negatives
Biomarker Absent	1 False negative	1 False positive

SOURCE: Courtesy of the authors.

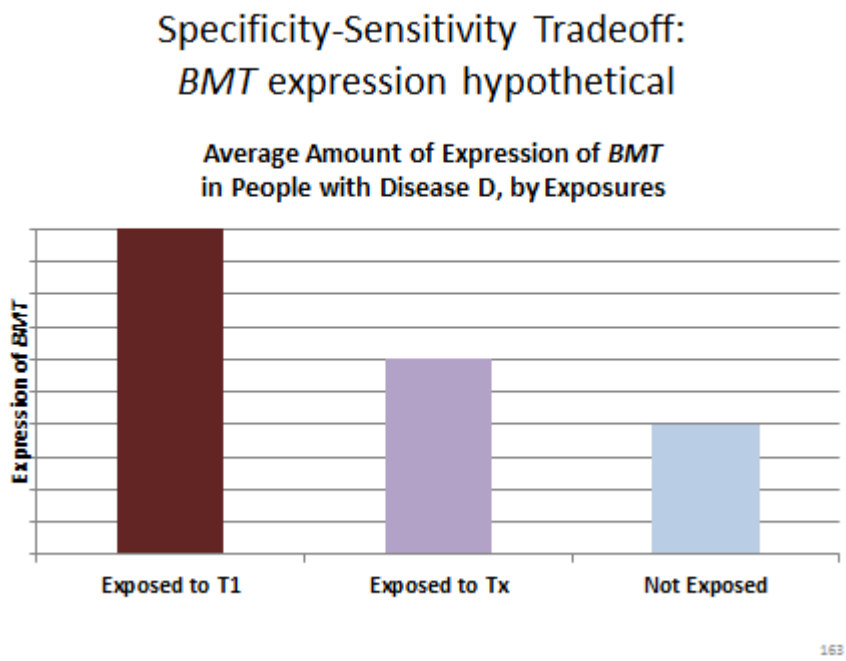
The specificity is [(5 true negatives) / (5 true negatives + 1 false positive)] = 5/6 = 0.83. The sensitivity is [(5 true positives) / (5 true positives + 1 false negative)] = 5/6 = 0.83.

To best aid judicial decision making, a biomarker should be both highly specific and highly sensitive. The module notes that in practice, researchers often must accept a tradeoff

between these two characteristics of a biomarker. Students may find that this makes intuitive sense: by choosing a biomarker that is highly likely to appear if causation is true, a researcher increases the risk that the biomarker may also sometimes appear if causation is false. The tradeoff makes particular sense in the context of biomarkers that are ambiguous to determine or quantitative in nature. To take an example that does not (usually!) involve toxic-tort claims, the instructor might discuss fasting-blood-sugar concentration. Elevated fasting-blood-sugar level is sometimes used as a diagnostic biomarker for diabetes. But fasting-blood-sugar levels vary, even in people without diabetes. Doctors need a cutoff to distinguish between "elevated" levels (the marker) and "normal" levels (the absence of the marker). The higher that cutoff level is set, the less likely that someone without diabetes will be diagnosed as having the disease (i.e., fewer false positives), but the more likely that someone with diabetes will go undiagnosed (i.e., more false negatives). Choosing a relatively high cutoff level gives priority to specificity at the expense of sensitivity. Choosing a relatively low cutoff level gives priority to sensitivity at the expense of specificity.

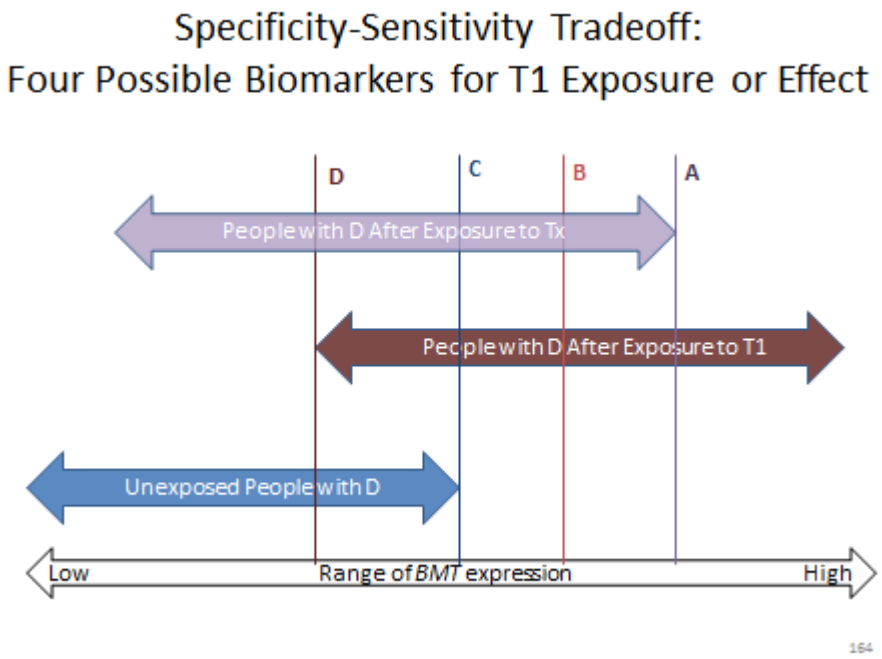
Slides 163 and 164 present this idea in a toxic-tort context, using a grossly oversimplified example of change in the expression of a single gene as a biomarker. Slide 163 sets up the hypothetical. Exposure to agent *T1* is believed, based on toxicological and epidemiologic studies, to cause disease *D*. But another family of toxins, *Tx*, is also believed to cause *D*, and some people develop *D* without any known toxic exposure. Researchers have observed a potential biomarker of effect for *D* after exposure to *T1*: on average, the *BMT* gene is overexpressed in diseased tissue from people with *D* who were exposed to *T1*. That is, more of the protein product for which *BMT* codes is found in sick people with *T1* exposure than in unexposed sick people. *BMT* is also overexpressed, but not as strongly, in people with *D* who were exposed to *Tx*. Unexposed people with *D* who were not exposed to any toxin also express *BMT*, but their average level of gene expression is lower.

Slide 163 Specificity-Sensitivity Trade-off: *BMT* Expression Hypothetical



Slide 164 graphically illustrates the trade-off. The horizontal bands represent ranges of *BMT* expression, from low expression on the left to high expression on the right, found in people with *D* who have no toxic exposure (bottom band), *T1* exposure (middle band), and *Tx* exposure (top band). The vertical lines marked A through D represent different levels of *BMT* expression that might be selected as the biomarker for *T1* causation. Students should readily grasp that the more sensitive biomarkers are less specific, and vice versa. For example, biomarker A would never be present in the absence of *T1* exposure. Because it would yield no false positives, it would be perfectly specific. Biomarker A would yield many false negatives, however (all cases in the central band to the left of line A), so its sensitivity would be relatively low. Biomarker C, by contrast, is much more sensitive (yielding many fewer false negatives), but at the expense of reduced specificity (yielding false positives exposed to *Tx* in the top band to the right of line C, although not yielding any false positives among the unexposed). Instructors interested in pursuing these issues further could consult, among many other sources, Schulte & Perera, supra, at 84–86 or Sandhya Pruthi et al., "Evaluation of Serum Estrogen-DNA Adducts as Potential Biomarkers for Breast Cancer Risk," 132 *J. Steroid Biochemistry & Molecular Biology* 73, 75–76 (2012).

Slide 164 Specificity-Sensitivity Trade-off: Four Possible Biomarkers for T1 Exposure or Effect

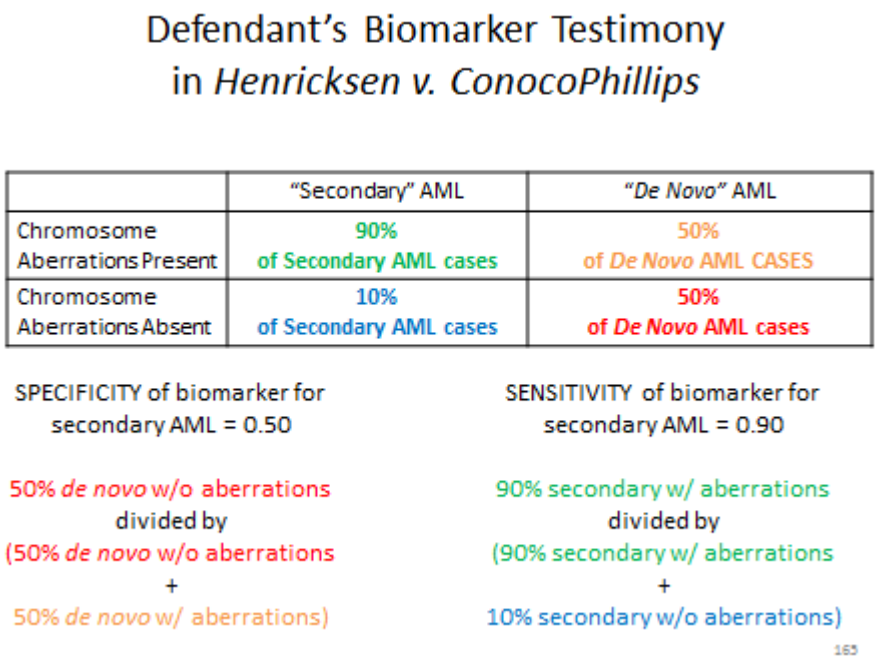


As the module explains, to achieve the ideal of determinative proof of causation or lack of causation, a biomarker (or, conceivably, a group of biomarkers) would have to achieve perfect specificity and sensitivity: always appearing when causation is true and never appearing when causation is false. The module explains why such biomarker "signatures" are unlikely to be common. But even without signature biomarkers, the presence or absence of valid biomarkers of reasonably high specificity or sensitivity could provide relevant evidence that would help refine the probability of specific causation. The module gives a couple of examples.

The example about chromosome aberrations and benzene exposure is from *Henricksen v. ConocoPhillips Co.*, 605 F. Supp. 2d 1142 (E.D. Wash. 2009). The plaintiff, who had been employed as a driver of gasoline tanker trucks, alleged that exposure to benzene in the gasoline caused his AML. We have seen similar allegations earlier in the module. The defendant successfully moved to exclude the proffered testimony of the plaintiff's causation experts and was therefore granted summary judgment, dismissing the complaint. The court relied, in part, on testimony of a defense expert concerning the distribution of certain chromosomal abnormalities in "*de novo*" (idiopathic) AML cases as compared to "secondary" AML cases (caused by environmental exposure). The expert testified that these abnormalities were observed in "nearly ninety percent" of all secondary AML and "approximately fifty percent" of *de novo* AML. A possible exercise for students would be: assuming that these chromosome

abnormalities are in fact a valid biomarker for secondary (as opposed to *de novo*) AML, what are the sensitivity and specificity of the biomarker? Slide 165 presents the calculations.

Slide 165 Defendant's Biomarker Testimony in *Henricksen v. ConocoPhillips*



For examples of cases discussing hormone receptors and genetic susceptibility to breast cancer, see *In re Prempro Products Liability Litigation*, 586 F.3d 547, 566 (8th Cir. 2009); and *Kaufman v. Wyeth*, LLC, No. 1:02-CV-22692, 2011 WL 10483576 (S.D. Fla. Aug. 15, 2011). For a discussion of the observed relation between *BRCA1* mutations and hormone receptors, see William D. Foulkes, "BRCA1—Sowing the Seeds Crooked in the Furrow," 40 *Nature Genetics* 8, 8 (2008).

To further assess student understanding, the instructor could ask the student to evaluate the following trial testimony, which was described in *Tompkin v. Philip Morris USA*, Inc., 362 F.3d 882, 890 n.5 (6th Cir. 2004). The plaintiff, who had lung cancer, had a history of cigarette smoking and of asbestos exposure. The defendant cigarette manufacturer argued that the asbestos exposure, rather than the smoking, caused the cancer. Among the supporting evidence for this contention was the testimony of one of the defendant's experts "that P53 and K-Ras studies . . ., which test for genetic changes associated with smoking, were negative." Students should understand that the import of the testimony is that the genetic changes are biomarkers that validly distinguish smoking-caused lung cancers from other lung cancers. The instructor might ask: What additional information would be needed to assess the probative

value of the negative genetic tests? Among other possible answers, one would want to know the strength of the association between the missing genetic changes and smoking, whether the association truly reflected causation, and how sensitive the biomarkers were (that is, how likely is it that the biomarkers would be missing in a case of true causation).

The discussion of sensitivity and specificity in the module focuses on biomarkers of effect and exposure because these are the easiest to understand in the context of backward-looking tort litigation: What characteristics of the sick plaintiff's tissues or cells tell us why the plaintiff became sick? But even for biomarkers of exposure, which measure a plaintiff's ex ante genetic risk of toxic effect, sensitivity and specificity are important. A signature marker of susceptibility would have perfect sensitivity (everyone with the susceptible genotype would get the disease) and perfect specificity (everyone with the disease would have the susceptible genotype). If the biomarker genotype is less than perfectly sensitive or specific (as was the case in the studies discussed in the *Expert Report of T. Toxicologist*), the issues for biomarkers of susceptibility are quite similar to the issues for biomarkers of effect or exposure.

Trichloroethylene Exposure and Specific Somatic Mutations in Patients with Renal Cell Carcinoma

Hiltrud Brauch et al.

Journal of the National Cancer Institute, Vol. 91, no. 10, pp. 854-861 (May 19, 1999)

This section concludes with the abstract of an early study that sought to identify a "signature" for renal cell carcinoma caused by trichloroethylene exposure (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). Students should be able to identify the putative biomarker and assess its evidentiary utility.

Notes and Questions after Brauch et al. abstract:

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, 18 (2007), http://www.occup-med.com/content/pdf/1745-6673-2-13.pdf.

   This note emphasizes, again, the importance of replication. The authors of the study excerpted above concluded that they had found "the first molecular evidence for a relationship between exposure to a defined carcinogen, gene damage, and

kidney cancer." Yet, the fact that a second study failed to replicate the result would weigh against scientific or legal reliance on the study's conclusions.

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, sometimes 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 *Lohrmann v. Pittsburgh Corning Corp.*, 782 F.2d 1156 (4th Cir. 1986); *Borel v. Fibreboard Paper Products Corp*., 493 F.2d 1076 (5th Cir. 1973); *Rutherford v. Owens-Illinois, Inc.*, 941 P.2d 1203 (Cal. 1997); *Thacker v. UNR Industries, Inc.*, 603 N.E.2d 449 (Ill. 1992); *Gregg v. V-J Auto Parts, Co.*, 943 A.2d 216 (Pa. 2007); *Bostic v. Georgia-Pacific Corp.*, 439 S.W.3d 332 (Tex. 2014); and *Sienkiewicz v. Greif (UK), Ltd.* [2011] 2 AC 229 (appeal taken from Eng.). These issues are discussed in, for example: Steve C. Gold, "Drywall Mud and Muddy Doctrine: How Not to Decide a Multiple-Exposure Mesothelioma Case," 49 *Ind. L. Rev.* 117 (2015); Michael D. Green, "Second Thoughts About Apportionment in Asbestos Litigation," 37 *Sw. U. L. Rev.* 531 (2008); and Joseph Sanders, "The 'Every Exposure' Cases and the Beginning of the Asbestos Endgame," 88 *Tul. L. Rev.* 1153 (2014).

   The type of biomarker information discussed in this section cannot distinguish among different sources of chemically identical exposures. Thus, even if biomarker data begin to help assign specific causation to a particular agent, it will not solve the challenging problem of the indeterminate defendant.

3. For discussions of the potential role of biomarkers in toxic-tort litigation, see 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); Jamie A. Grodsky, "Genomics and Toxic Torts: Dismantling the Risk-Injury Divide," 159 *Stan. L. Rev*. 1671 (2007); Andrew R. Klein, "Causation and Uncertainty: Making Connections in a Time of Change," 49 *Jurimetrics J.* 5 (2008); and Gary E. Marchant, "Genetic Data in Toxic Tort Litigation," 45 *Brief* 22 (Winter 2016).

Study Questions for Brauch et al. abstract:

1. How would you describe the design of this study?

   This is a case-control design. Because the researchers were not interested in assessing whether TCE exposure was associated with RCC, but only whether specific mutations were associated with RCC in individuals exposed to TCE, "cases" in this study are people with both occupational TCE exposure and RCC. "Controls" in this study include both people without RCC and people with RCC, but without exposure.

2. What type of biomarker were the researchers investigating?

   The researchers were looking for a biomarker of effect, to wit, an observable change that occurred uniquely or disproportionately in renal cell carcinomas that resulted from TCE exposure. Note that the researchers had no independent way of knowing whether TCE was the "true" cause of a patient's cancer, even if the patient had been exposed to TCE. The researchers used occupational exposure to TCE as a proxy for cancer causation by TCE. To refine their proxy, they classified the exposed individuals into subgroups based on the amount (duration, frequency, and mode) of exposure.

3. Why did the researchers compare the DNA from the tumors of kidney cancer patients who had been exposed to TCE to the DNA from: (a) kidney cancer cells from patients without TCE exposure; (b) white blood cells of healthy subjects; and (c) noncancerous kidney cells adjacent to the tumors of exposed patients whose tumors had a particular mutation?

   Each of these comparisons provided a control group for the study. Comparison to cancer cells from unexposed kidney cancer patients enabled the researchers to test the hypothesis that certain mutations were associated with TCE exposure, rather than simply being associated with having RCC. Comparison to white blood cells from healthy subjects enabled the researchers to test the hypothesis that certain mutations were associated with having RCC, rather than being randomly distributed. Comparison to noncancerous kidney cells adjacent to the tumors of exposed patients enabled the researchers to test the hypothesis that the observed mutations were somatic mutations that resulted from the carcinogenic process, rather than being inherited germline mutations that made the patients more likely to develop RCC. Having this control was particularly important because both inherited mutations and somatic (potentially carcinogen-induced) mutations in the *VHL* gene are known to be associated with RCC.

4. What conclusion did the researchers draw from their observations? How strong was the evidence for that conclusion?

   The abstract stated the researchers' conclusion explicitly: "Our results suggest that RCC in patients with high, cumulative TRI [TCE] exposure is associated with a unique mutation pattern in the VHL gene." The abstract cited three pieces of evidence in support of the conclusion: (1) the high percentage (75%) of exposed patients who had somatic *VHL* mutations; (2) the association between the number of somatic mutations and the intensity of the exposure; and (3) the presence of a mutation at nucleotide 454 of the *VHL* protein in the tumors of 39% of the exposed patients with *VHL* mutations, but in none of the patients without exposure and in none of the healthy subjects. There are some weaknesses in the evidence for their conclusion, at least insofar as one can tell from the abstract. The abstract does not report how frequently the tumors of unexposed patients had *VHL* mutations or whether the difference between the frequency of *VHL* mutations in exposed versus unexposed patients was statistically significant; the abstract does not report the strength or statistical significance of the association between the number of mutations and the intensity of the exposure; and the abstract reports that the nucleotide 454 mutation was absent from a majority of the tumors of exposed patients. Furthermore, the apparent dose-response relationship (more mutations in those with more exposure) depends on the correctness of the retrospective estimates of the degree of exposure.

5. Suppose a person with kidney cancer who had been occupationally exposed to high levels of TCE sued the manufacturer of the TCE used on the job. Based on the Brauch study excerpted above, what weight would you give to evidence that:

   a. cells from the plaintiff's RCC had a mutation in nucleotide 454 of the *VHL* gene?

      This evidence would have limited probative value because it could not exclude the possibility that the plaintiff had an inherited mutation in *VHL* nucleotide 454. Four of the 13 patients in the Brauch study who had this mutation in their RCC cells also had it in adjacent noncancerous tissue. The plaintiff could argue that more often than not (9/13 or 69% of the time), an observed nucleotide 454 mutation is a somatic mutation, and that therefore (as explained in the response to part (b) that follows), a factfinder could infer that TCE caused the plaintiff's cancer. But a reasonable factfinder might wonder why only the genotype of the cancerous tissue was obtained, when it presumably would have been easy to test adjacent noncancerous tissue as well.

   b. cells from the plaintiff's RCC had a mutation in nucleotide 454 of the *VHL* gene,

but cells of the plaintiff's adjacent noncancerous kidney tissue did not have that mutation?

This evidence would show that the nucleotide 454 mutation was a somatic mutation that arose in the cancerous kidney tissue. The Brauch study found that the nucleotide 454 mutation is a highly specific biomarker: it did not appear at all in patients without TCE exposure. This could support an inference that the presence of that biomarker indicates that the RCC was caused by TCE exposure. Based on that inference, a factfinder could further infer that the mutation's presence in the plaintiff's tumor established specific causation. To distinguish whether the nucleotide 454 mutation indicates causation by TCE rather than simply exposure to TCE, it would be helpful to know whether and how often the mutation appeared in individuals with similar levels of TCE exposure who did not develop RCC or to understand the molecular mechanism of TCE-caused kidney carcinogenesis.

c. Noncancerous cells from other tissues in the plaintiff's body had mutations in the *VHL* gene that were also present in the plaintiff's RCC?

This evidence would show that the plaintiff had germline mutations in the *VHL* gene, not only somatic mutations. It would tend to suggest that the plaintiff had an inherited predisposition to RCC, but more information would be needed before a factfinder could reach that conclusion. What if researchers found an association between the plaintiff's germline mutations and a higher risk of RCC? Was the overall pattern of *VHL* mutations different in the tumor than in the non-cancerous tissue? If the nucleotide 454 mutation is a valid biomarker for TCE-caused RCC, did the plaintiff's tumor have this mutation? If it did, was the nucleotide 454 mutation also present in plaintiff's noncancerous tissue?

d. Noncancerous cells from other tissues in the plaintiff's body had numerous mutations in the *VHL* gene, but no mutation in nucleotide 454 of the *VHL* gene, while cells from the plaintiff's kidney cancer had the same mutations, but also had a mutation in nucleotide 454 of the *VHL* gene?

This evidence would point in more than one direction. The evidence of germline mutations might suggest (subject to the caveats in (c) above) that the plaintiff had an inherited predisposition that caused the plaintiff's RCC. The evidence of a somatic nucleotide 454 mutation in only the cancer cells might suggest (subject to the caveats in (b) above) that TCE caused plaintiff's RCC.

It would be possible for both the above conclusions to be true if the genetic predisposition interacted with exposure to TCE to increase the plaintiff's RCC

risk. To assess that possibility, it would be helpful to know how exposure to TCE alters RCC risk for individuals with the plaintiff's genotype.

It would also be possible that the plaintiff had a genetic predisposition to RCC, but that TCE caused the plaintiff's RCC, independent of the genetic predisposition. To assess that possibility, it would be helpful to know the penetrance of plaintiff's germline mutations in the absence of any exposure to TCE or other kidney carcinogens.

e. Noncancerous cells from other tissues in the plaintiff's body did not have any mutations in the VHL gene, while cells from the plaintiff's kidney cancer had numerous mutations in the VHL gene, but did not have a mutation in nucleotide 454 of the VHL gene?

This evidence would also point in more than one direction, but not equally strongly. The absence of mutations in the *VHL* gene would imply that the plaintiff does not have a genetic predisposition to RCC—at least because of the *VHL* gene. Unknown genetic contributors to RCC risk might exist, but at the very least, a factfinder could conclude that the evidence made a genetic cause less probable. The absence of somatic mutations in nucleotide 454 in the tumor cells does not necessarily prove that TCE did not cause the plaintiff's cancer, because the Brauch study found that the nucleotide 454 mutation was not a very sensitive biomarker.

Brauch's team studied 44 TCE-exposed patients. Of those 44, 33 (75%) had somatic mutations in the *VHL* gene. Of those 33, 13 (39%) had somatic mutations in nucleotide 454. Thus, among TCE-exposed patients who, like the hypothetical plaintiff in this example, had somatic *VHL* mutations, the sensitivity of the nucleotide 454 mutation biomarker as a biomarker of exposure was 13/33 = 0.39 = 39%. Among all TCE-exposed patients, the sensitivity of the nucleotide 454 mutation biomarker as a biomarker of exposure was even lower, 13/44 = 0.295, or 30%. Thus, the absence of the biomarker does not correlate well with the absence of TCE exposure and, presumably, TCE causation.

Conceivably, however, the nucleotide 454 mutation is more sensitive as a biomarker of effect than as a biomarker of exposure. What if—as is plausible or even likely—for some patients, TCE exposure was coincident with RCC and not causal of RCC?

Suppose all TCE-exposed patients with the nucleotide 454 mutation were true cases of TCE causation and those without the mutation were background cases whose RCC did not depend on exposure. In that case, the nucleotide 454 mutation would be a perfectly sensitive biomarker of effect, despite being a relatively insensitive biomarker of exposure. Alternatively, suppose that only *some* TCE-exposed patients without the mutation were background cases. In

that situation, the true sensitivity of the biomarker of effect would be unknown, but would have to be greater than 29.5% (13/44). As a final alternative, suppose that the absence of *any* somatic *VHL* mutations were found conclusively to rule out TCE causation, and that some but not all of the TCE-caused tumors had the nucleotide 454 mutation. In that situation, the true sensitivity of the biomarker of effect would be unknown, but would have to be greater than 39% (13/33).

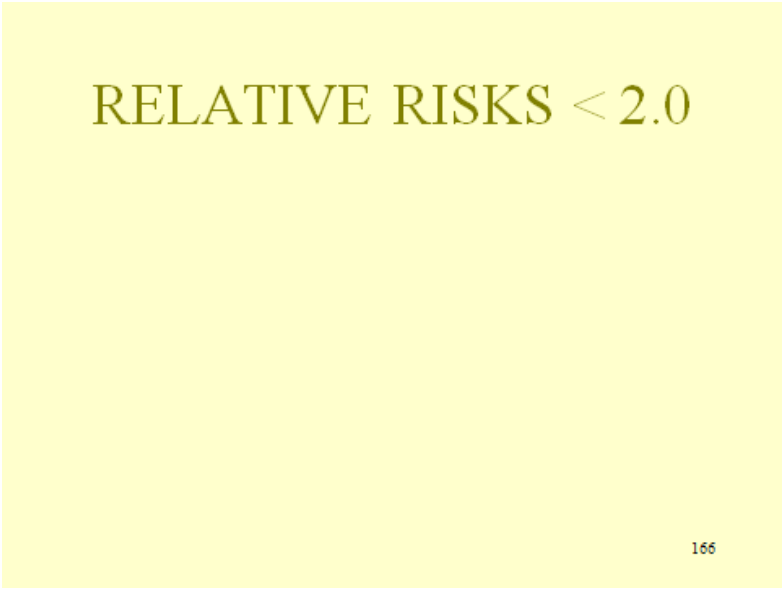
The Brauch study did not produce data that could distinguish the above possibilities. To assess the meaning of the absence of the putative biomarker of effect, it would be helpful to know the overall relative risk of RCC for individuals with this level of exposure to TCE as compared to unexposed individuals.

Finally, the instructor and the student should recall that all of the foregoing examples, a through e, are based on the assumption that the somatic nucleotide 454 mutation is, in fact, a valid biomarker at least of exposure and possibly of effect. The Brauch study alone may not be sufficient to support the conclusion that the biomarker is valid. The failure of the Charbotel study to replicate the Brauch study's result reminds us of that limitation.

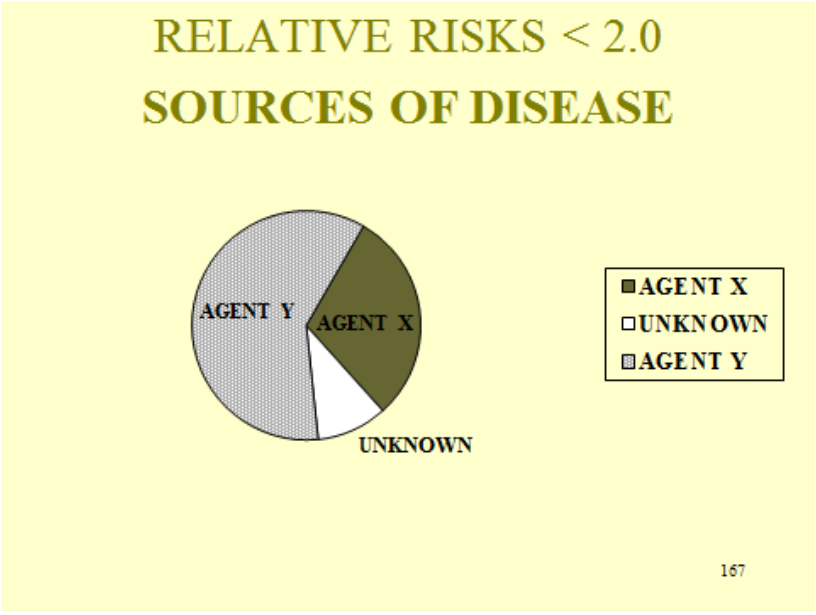
Section V.H. The Sufficiency of Relative Risks ≤ 2.0

Slide 166 introduces this topic. To illustrate the point made in the module about differential etiologies, Slide 167 shows the attributable proportion of risk for the three risk factors described in the module in a pie chart. Slide 168 then removes the contribution of Agent Y, after its elimination in a differential etiology, and helps students visualize that the denominator has changed with Agent Y's removal. When the denominator is reduced (numerator remains the same), the fraction is increased.

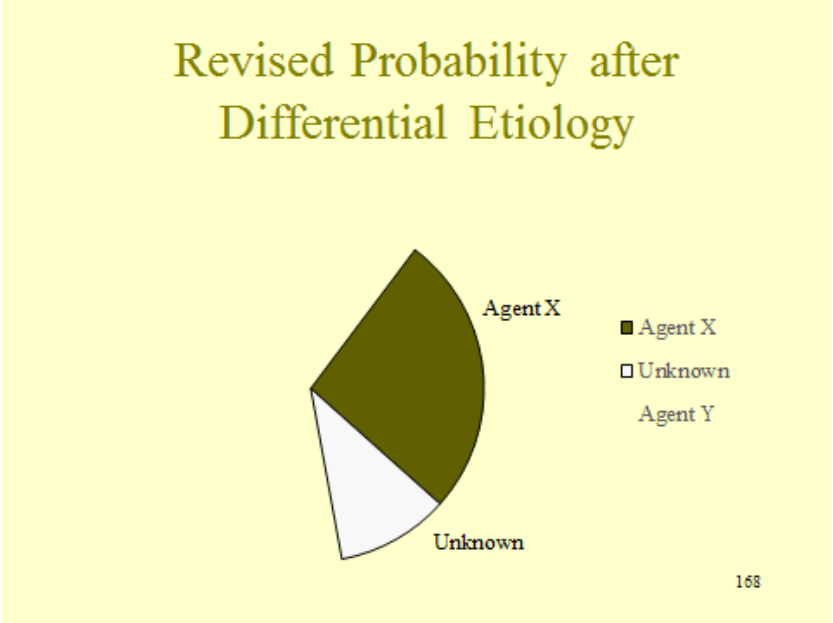
Slide 166 Relative Risks < 2.0



Slide 167 Relative Risks < 2.0



Slide 168 Revised Probability after Differential Etiology



With the removal of Agent Y, the probability for Agent X for this individual is now 75%, rather than the 30% when Agent Y was a risk factor. With Agent Y removed, there are only two risk factors remaining, X and Unknown. The answer to the question posed by the module for the formula is straightforward. The denominator is the sum of the APRs of each non-eliminated risk factor, here .3 and .1, and the numerator is the APR for the risk factor of interest, here .3. Thus, the formula is: APR (agent in question) / APRs (agent in question and other non-eliminated risk factors).

Study Question:

1. Can a differential etiology be used to establish that an agent caused the disease in the absence of evidence of general causation?

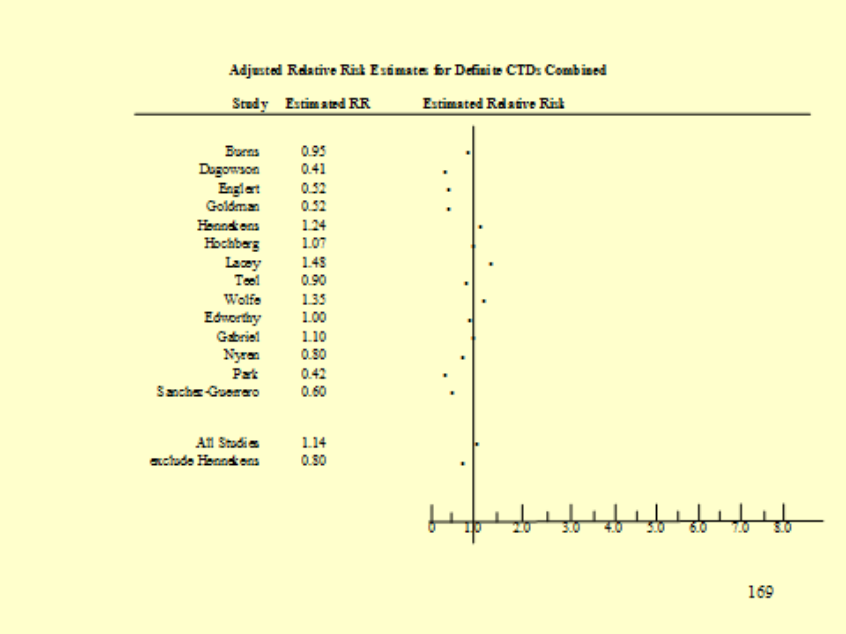
   No, although experts in a number of court cases have attempted to do so. Even if a differential etiology could eliminate all known causes of the disease, that does not establish that the agent is a cause, either generally or specifically. Any other environmental agent or genetics may be the cause, and without general causation evidence, there is no reason to believe that the agent is any more likely than all of the multitude of possible explanations for the plaintiff's disease.

   Unknown causes cannot be eliminated, which means that they always remain as a possible cause. Even if all other known competing causes have been eliminated, when the risk (APR) of unknown causes is equal to or greater than the risk of the

alleged agent, the probability of the alleged agent's being the cause will not exceed 50%, as required by the preponderance standard.

The argument against using studies finding relative risks under 2.0 for *general causation* is made by one of the authors of this module, in Michael D. Green, "The Future of Proportional Liability," in Stuart Madden (ed.), *Exploring Tort Law* (2005). The graph of study outcomes for silicone-gel breast implants and connective-tissue disease (Slide 169) almost certainly reveals noise rather than signal. Because numerous studies were conducted based on the concern that breast implants were causing chronic disease, we have available not only studies that found a small increased risk but also ones that found a small protective effect. Of course, suits are not brought when studies find a negative association; in most cases, the author above contends they should not be brought based on small-positive relative risks in observational studies.

Slide 169 Adjusted Relative Risk Estimates for Definite CTDs Combined



SOURCE: Michael D. Green, *The Future of Proportional Liability*, in "Exploring Tort Law," Stuart Madden, ed. (Cambridge: Cambridge University Press, 2005). Reproduced courtesy of the author.

SECTION VI. THE ROLE OF CONSENSUS ORGANIZATIONS

We feel that this material does not require further exposition for instructors.

GOLD RIVER HYPOTHETICAL

Mrs. Wynona Smith, a resident of Jefferson, Mississippi, filed this diversity action in the United States District Court for the District of Mississippi as Administratrix of her husband's estate, against QualChem Manufacturing Co. for wrongful death and for personal injuries suffered during his lifetime (survival action). The claim results from Mr. Smith's exposure to Dreion 43, a chemical released into the Gold River by QualChem Manufacturing Co. as a by-product of its manufacturing operations. Recovery is sought on theories of negligence, strict liability for abnormally dangerous activities, private nuisance, and trespass.

The complaint states that Mr. Smith contracted liver cancer and died as a result of this cancer, caused by exposure to Dreion 43. Mr. Smith died 12 months ago at age 69 after a 4-year battle with cancer. Mrs. Smith alleges that her husband was exposed to Dreion 43 in the drinking water of the family home, located on the banks of the Gold River. Mrs. Smith and her husband resided on the banks of the Gold River for 45 years. Throughout that time, the family drew water for drinking and bathing from a well on their property.

In its answer, QualChem Manufacturing Co. admits that it began manufacturing operations in Mississippi 85 years ago. The plant is 10 miles upriver from plaintiff's property. The answer also admits that Dreion 43 is a by-product of QualChem's manufacturing process and that various amounts of Dreion 43 have been discharged into the Gold River for at least 75 years. The answer denies that Mr. Smith's liver cancer was caused by Dreion 43.

The parties have admitted or stipulated to certain undisputed facts. As a result of flooding that occurs in Jefferson, a number of wells on riverfront property are contaminated with Dreion 43 at a level of approximately 20 micrograms per liter. That level of contamination has been stable over the past 10 years. Prior to that, contamination levels were stable at 50 micrograms per liter from the time when QualChem Manufacturing Co. began its operations in Mississippi. Testing of the Smith well revealed current and historical contamination levels consistent with those above. In addition, fish in the Gold River have been found to contain Dreion 43, at a level varying from 20 to 100 micrograms per kilogram of body weight. The contamination of the Gold River and its fish with Dreion 43 has been highly publicized in Jefferson and other communities in the vicinity of the river for at least a decade. No other source of Dreion 43 contamination of the Gold River is known. However, like most rivers, the Gold River contains low levels of many other contaminants.

Discovery also revealed that QualChem discharged Dreion 43 for a number of years before any law required a permit for that discharge. Pursuant to the federal Clean Water Act (known as the Federal Water Pollution Act from 1948 to 1972), those releasing pollutants into navigable waters were required to obtain a permit. QualChem ignored that requirement until 10 years ago when it obtained a discharge permit that required it to treat its waste water to reduce the pollutants it contained. The permit required QualChem to reduce the amount of Dreion 43 that it discharged into the Gold River by sixty percent. Since obtaining the permit, QualChem has filed Discharge Monitoring Reports stating that QualChem has complied with the permit limit on Dreion 43 discharges.

Following expert discovery, defendant filed a motion in limine to exclude the testimony of all four of plaintiff's experts: Dr. Teresa Toxicologist, Dr. Ellen Epidemiologist, Dr. Tom Treating, and Dr. Gina Geneticist. The motion alleged that the experts' testimony is inadmissible because it fails to satisfy the legal requirement for scientific reliability. Accompanying the motion in limine was a motion for summary judgment claiming that no genuine issue of material fact exists with regard to the question of Dreion 43 causing Mr. Smith's cancer. Defendant's motion was supported by affidavits deemed adequate by the court. The court held an evidentiary hearing on the motion in limine and each of the four challenged experts testified. A summary of the testimony of each of those witnesses follows.

PLAINTIFF'S FIRST WITNESS - DR. TERESA TOXICOLOGIST

Mrs. Smith's first witness is Dr. Teresa Toxicologist. The court reviewed Dr. Toxicologist's curriculum vitae and determined that she is qualified to testify. Dr. Toxicologist has identified two published studies on the carcinogenic effects of Dreion 43 on animals.

Exhibit A (attached at the end of this essay question) shows the results of a published study conducted on mice. In the study, one group of mice was exposed to Dreion 43 in its drinking water, and one group was not exposed to Dreion 43 in its drinking water. Both groups were given 5 milliliters of water per kilogram of body weight on a daily basis. The exposed group's water contained 5 milligrams of Dreion 43 per liter of water. Among the exposed mice, 14 out of 92 developed liver cancer within six months for an incidence of .15, and among the unexposed mice, 2 out of 85 developed liver cancer within the same period, for an incidence of .02. Dr. Toxicologist testifies that the difference in the rate of liver cancer in the two groups is statistically significant with a *p* value of .014 and confidence limits of .97 to 19.6. She also testifies that the dose (based on milligrams of Dreion 43 to total body weight multiplied by the time of exposure) of Dreion 43 to which the mice were exposed is approximately 125 times the dose a human weighing 175 pounds would receive over a lifetime if their drinking water contained 45 micrograms per liter of Dreion 43 (approximately the dose to which Mr. Smith, who weighed 175 pounds, was exposed over his lifetime).

Exhibit B (also attached) shows the results of a published study conducted on Sprague Dawley rats. This study was conducted using three separate groups of exposed rats and a group of unexposed rats. The three exposed groups of rats were subjected to a low dose of Dreion 43 (80 micrograms of Dreion 43 per liter of water), a medium dose (500 micrograms of Dreion 43 per liter of water), and a high dose (25 milligrams[*] of Dreion 43 per liter of water). Dr. Toxicologist explains that these dosages roughly conform to a lifetime exposure in humans of two times, 10 times, and 500 times, respectively, the dose to which Mr. Smith was exposed. The study investigated kidney cancer as the only endpoint, as it was funded by the National Kidney Foundation.

Among the unexposed group of rodents, 4 out of 200 developed kidney cancer within a year, an incidence of .02. In the low dose group, 3 out of 170 developed kidney cancer, an incidence of .018. In the medium exposed group, 16 out of 194 rats developed kidney cancer, an incidence of .08, and in the high dose group, 44 out of 210 developed kidney cancer, an incidence of .21, all within 1 year of onset of exposure. The difference in incidence between the control group and the low exposure cohort was not statistically significant ($p > .05$), the differences between each of the medium and high dose groups and the control group were both statistically significant at the .05 level ($p = .039$). Because there is evidence that both liver and kidney cancer are initiated by perturbations in cell differentiation, Dr. Toxicologist believes that

The results of this study are relevant to the question of whether Dreion 43 causes liver

[*] A milligram constitutes 1,000 micrograms.

cancer.

A dose-response curve, reflecting the results of the rat study, is contained in Exhibit C (attached).

Based on these two studies, Dr. Toxicologist testifies that in her opinion Dreion 43 is clearly carcinogenic in two different species of rodents and that it is likely to be carcinogenic in the human species.

Questions:

1. What is the import of the .15 incidence of liver cancer in mice exposed to Dreion 43 for assessing whether Dreion 43 causes liver cancer in mice?
2. What is the relative risk of liver cancer for mice exposed to Dreion 43 in the study discussed by Dr. Toxicologist?
3. What is of concern about Dr. Toxicologist's testimony about the mice study, aside from the problem of extrapolating animal studies to human toxicity?
4. What is the significance of the dose-response relationship reflected in Exhibit C?
5. Is the reliability of evidence based on the results of animal studies affected by the higher doses to which the animals were exposed?
6. Should issues of biological plausibility play a role in consideration of the reliability of evidence regarding causation in humans based on results of animal studies? If so, in what way?
7. Under what circumstances might evidence based on the results of animal studies be sufficiently reliable to be admitted to prove causation in humans?
8. What if these toxicology findings are inconsistent with epidemiologic studies of humans?
9. Aside from the admissibility of the experts' opinions, what affirmative defense did the defendant's lawyer leave on the table by not asserting it? What outcome if she had?

PLAINTIFF'S SECOND WITNESS - DR. ELLEN EPIDEMIOLOGIST

Mrs. Smith's second expert witness is Dr. Ellen Epidemiologist who has degrees in epidemiology and medicine. The court has determined that this witness is qualified to testify as an epidemiologist and as a physician.

On direct examination, the witness testifies that there is one epidemiologic study of the effects of Dreion 43 on humans, a case-control study. The study was performed recently, has been submitted to an epidemiologic journal for publication, and is currently undergoing peer review as part of the submission process. She testifies that the study was conducted because of anecdotal reports of an increased incidence of cancer, particularly liver cancer, among residents living near the river in and around Jefferson.

The study was conducted by epidemiologists employed by the Mississippi Department of Public Health. In the study, the epidemiologists selected a sample of individuals with liver cancer from a state solid tumor registry. Controls were matched for age, gender, geographic proximity to the Gold River area, and occupation and obtained from a state registry of organ donors. None of the controls had liver cancer.

Both cases and controls were personally interviewed and their medical records obtained. A variety of factors were assessed, including other known causes of liver cancer, which include alcoholism and hepatitis B virus. Cases and controls were asked about the source of their drinking water, and unless the initial answer clearly included or excluded drinking water from the Gold River, the interviewer asked specifically about whether the subject drank water from the Gold River. Unless the subject obtained virtually all of his or her water from the Gold River or a well situated close to it, the individual was treated as unexposed.

Among the cases, 9 were classified as exposed to Gold River water (and therefore Dreion 43) and 3 were unexposed. In the control group, 44 were exposed to Gold River water and 44 were not. Using standard statistical methodology, Dr. Epidemiologist testifies that the odds ratio for the association between exposure to Gold River water and liver cancer is 3.2. A nested confidence interval, representing two different statistical significance levels was constructed by the authors of the study and is displayed in Exhibit D (attached).

Based on the association between liver cancer and exposure to Dreion 43 in drinking water found in this study, Dr. Epidemiologist testifies that, to a reasonable degree of scientific certainty, Dreion 43 causes liver cancers in humans.

On cross-examination, Dr. Epidemiologist testifies:

- The association between liver cancer and exposure to Dreion 43 found in the study is not statistically significant at the .05 level. However, Dr. Epidemiologist testifies that the association is statistically significant at the .10 level.
- The authors of the epidemiological study stated in the study that their conclusion was that the study was suggestive that Dreion 43 caused liver cancer but that further research was required before they would be prepared to state that Dreion 43 causes liver cancer in humans.

Questions:

1. What is the difference between the "cohort" studies of mice and rats and the "case-control" study of humans (aside from the different species involved)?
2. Assuming the epidemiologic study reflects a causal relationship, what is the attributable percentage of risk ("APR") of liver cancer for those exposed to Dreion 43? Explain your answer.
3. What assumption did you have to make in calculating an APR from the study results?
4. What step(s) did Dr. Epidemiologist fail to address in concluding that the association found in the study means that Dreion 43 causes liver cancer?
5. What potential biases exist based on the study design and what effects would those biases have on the outcome found in the study?
6. What effect should the lack of statistical significance at the .05 level have on the admissibility of Dr. Epidemiologist's opinion?
7. What is the relationship between a .05 level of statistical significance and the civil law burden of proof of a preponderance of the evidence?
8. What do scientists mean when they use "reasonable degree of scientific certainty" to explain the strength of their opinions and how does that level of conviction relate to the legal "preponderance of the evidence" standard?
9. What impact should the epidemiologic study authors' opinion that further research was required to prove the causal connection between Dreion 43 and liver cancer have on the admissibility of Dr. Epidemiologist's testimony?

PLAINTIFF'S THIRD WITNESS – DR. TOM TREATING

Dr. Treating testifies that he diagnosed and treated Mr. Smith's liver cancer. He testifies that Mr. Smith died of liver cancer and that Mr. Smith's liver cancer was most likely caused by exposure to Dreion 43 both in drinking water and from eating fish caught in the Gold River.

Dr. Treating determined that Mr. Smith had been exposed to Dreion 43 in his drinking water and consuming fish through inquiries to him while he was being treated for his cancer. Dr. Treating learned from Mr. Smith that he caught and ate fish from the river approximately twice a week for most of his adult life. When at home, decedent drank water that was drawn from the well on the property. Based on this information, Dr. Treating concluded Mr. Smith was exposed to Dreion 43.

Dr. Treating explains that, in addition to Mr. Smith's exposure, he bases his opinion on the epidemiologic study presented by Dr. Epidemiologist, the toxicology studies, and the French Genetic study (see testimony of Dr. Gina Geneticist below). Dr. Treating obtained copies of the four studies and reviewed them during his investigation into the cause of Mr. Smith's death, which was conducted at the request of plaintiff's counsel.

Dr. Treating also found from his research that Dreion 43 is structurally similar to QualChem 29, a chemical known to cause colon cancer in humans. Based on the toxicologic principle that chemicals that are structurally similar have a similar effect, this structure activity evidence also contributed to Dr. Treating's opinion. Dr. Treating testifies that governmental agencies look at the structure of substances when they are determining their safety and performing risk assessments. Based on this evidence, Dr. Treating concluded that Dreion 43 causes liver cancer in humans.

In addition, Dr. Treating reviewed the other known causes of liver cancer. While roughly 50 percent of liver cancers are due to unknown causes, Dr. Treating ascertained that two known causes of liver are alcoholism and the hepatitis B virus. Based on Mr. Smith's medical records and the medical history that Treating conducted of the decedent, Dr. Treating was able to rule out both of these risk factors. Based on the epidemiologic, toxicologic, and genetic evidence, Mr. Smith's exposure to Dreion 43, the structural similarity between Dreion 43 and QualChem 29, and ruling out alcoholism and hepatitis B as possible causes of Mr. Smith's cancer, Dr. Treating concluded that Mr. Smith's cancer was caused by exposure to Dreion 43 to a reasonable degree of medical probability.

On cross-examination, defense counsel elicits the following testimony from Dr. Treating:

- Dr. Treating is aware that there are issues concerning the methodology of the epidemiologic study, but he believes these issues are inherent to the nature of observational epidemiology, and therefore they do not change his opinion.

Questions:

1. What significance does the magnitude of the odds ratio of the epidemiology study (3.2) have with respect to Dr. Treating's testimony about whether Dreion 43 caused Mr. Smith's liver cancer? What difficulties are there in using an odds ratio or relative risk for purposes of determining what caused an individual's cancer?
2. What effect does the elimination of alcoholism and hepatitis B as potential causes of Mr. Smith's cancer have on the plausibility that Dreion 43 was a cause of his liver cancer?
3. What effect does the fact that half of liver cancers arise from unknown causes have on the plausibility that Dreion 43 was the cause of Mr. Smith's liver cancer?

PLAINTIFF'S FOURTH WITNESS - DR. GINA GENETICIST

Dr. Gina Geneticist obtained several preserved pathology samples of tissue taken from Mr. Smith before he died: a sample of Mr. Smith's liver tumor, a sample of nearby unaffected liver tissue, and a sample of skin tissue that was removed in a biopsy of what turned out to be a benign mole. Dr. Geneticist subjected those tissues to genetic testing. She testifies about the results of the genetic tests and about three peer-reviewed, published studies that the plaintiff seeks to introduce into evidence.

The first study involved 50 liver cancer patients in France. The results showed that among 25 patients with known Dreion 43 exposure, 20 (80%) exhibited a particular inversion mutation (in which a segment of DNA is reversed) in the *uhoo* gene in the tumor cells, but not in surrounding normal tissue. By contrast, among 25 patients without known exposure, only 5 (20%) tumors exhibited that *uhoo* inversion mutation. The difference in frequency of the *uhoo* inversion is statistically significant at $p<.05$. Mr. Smith's tumor does not display the *uhoo* inversion.

The second study was a genome-wide association study (GWAS) study of liver cancer. The GWAS identified a locus on chromosome 12 that was associated ($p<10^{-7}$) with risk of liver cancer. Further research, published in the same report, identified the gene involved as the *J0M1* gene and quantified the relative risk of two particular single nucleotide polymorphisms (SNPs) at different locations in the gene. In the first SNP an adenine base is replaced with a thymine base. The odds ratio was 1.31 (95% confidence interval 1.06–1.63) for a person with one or two copies of the variant allele containing thymine. In the second SNP, a cytosine base is replaced with a guanine base. The odds ratio was 0.79 (95% confidence interval 0.79–0.92) for a person with two copies of the variant allele containing guanine. Mr. Smith's genome includes one copy of each of these variant alleles.

The third study was a molecular epidemiology study of the *D43d* gene that Dr. Geneticist herself conducted and published. Dr. Geneticist explains that *D43d* codes for a protein called D43-deactivase. D43-deactivase is an enzyme produced in the liver that catalyzes the conversion of Dreion 43 into smaller units that the body rapidly eliminates. Thus, the amount of D43-deactivase activity affects the length of time that liver cells are exposed to any Dreion 43 that has been consumed. Most people have high D43-deactivase activity, but a significant minority of individuals have much lower D43-deactivase activity because their genome includes at least one copy of a variant *D43d* allele. Researchers therefore hypothesized that exposure to Dreion 43 would pose a greater risk of liver cancer to individuals with at least one copy of the variant allele. To study this, Dr. Geneticist conducted a case-control study in Turkey, where Dreion 43 exposure is common. For both cases and controls, she determined whether the individual had been exposed to Dreion 43 in drinking water and whether the individual had at least on copy of the variant *D43d* allele. She matched cases and controls for age and ancestry. The results of her study are summarized in Exhibit E. Based on those results and on genetic testing that showed Mr. Smith had two copies of the variant *D43d* allele, Dr. Genetecist testified that in her opinion, to a reasonable degree of medical certainty, exposure

to Dreion 43 had caused Mr. Smith's liver cancer and death.

Questions:

1. Is the result of the France study relevant to the causation issue in this case? Explain why or why not.
2. In what way does the France study support the testimony of Dr. Treating about Dreion 43 having caused the deceased liver's cancer?
3. Are the results of the GWAS study relevant to the causation issue in this case? Explain any assumption(s) necessary to support your answer.
4. What do the results of the GWAS study indicate about Mr. Smith's inherited susceptibility to liver cancer?
5. Does the study Dr. Geneticist conducted in Turkey support her conclusion about causation? Explain why or why not.

EXHIBIT A

SUMMARY OF MOUSE DATA

UNEXPOSED GROUP		
# with Liver Cancer	# in Group	Incidence
2	85	.02

EXPOSED GROUP		
# with Liver Cancer	# in Group	Incidence
14	92	.15

SOURCE: Courtesy of the authors

EXHIBIT B

SUMMARY OF RAT DATA

UNEXPOSED GROUP		
# with Kidney Cancer	# in Group	Incidence
4	200	.02

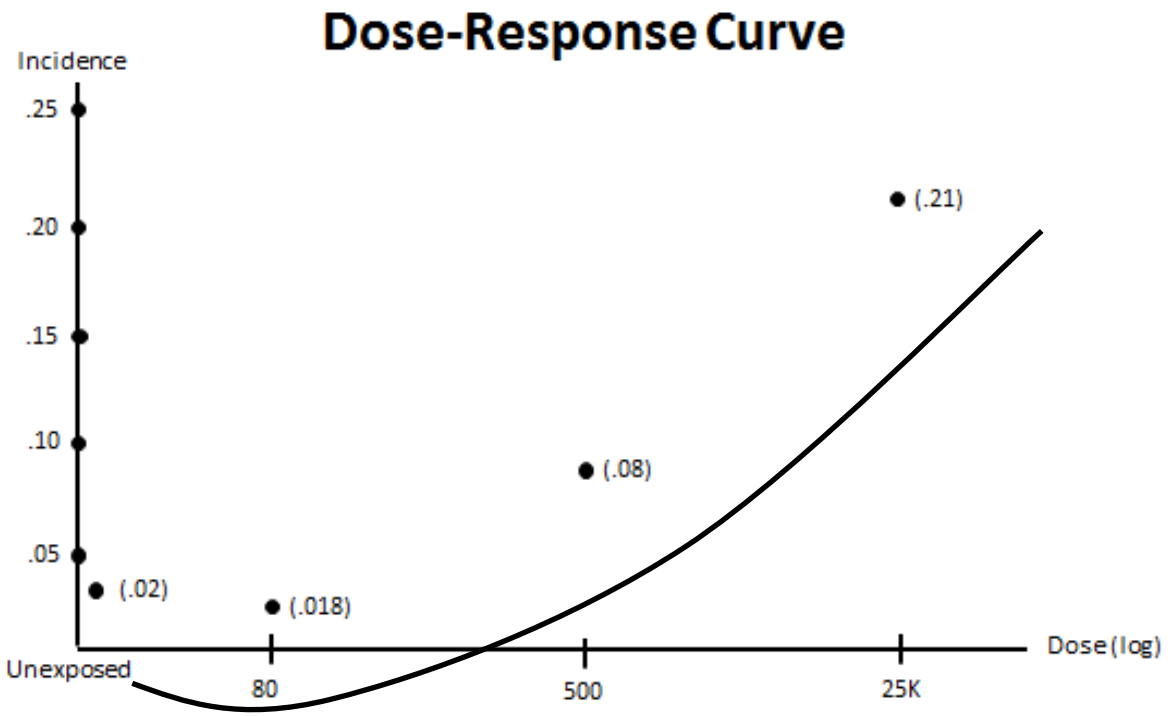
LOW EXPOSED GROUP		
# with Kidney Cancer	# in Group	Incidence
3	170	.018

MEDIUM EXPOSED GROUP		
# with Kidney Cancer	# in Group	Incidence
16	194	.08

HIGH EXPOSED GROUP		
# with Kidney Cancer	# in Group	Incidence
44	210	.21

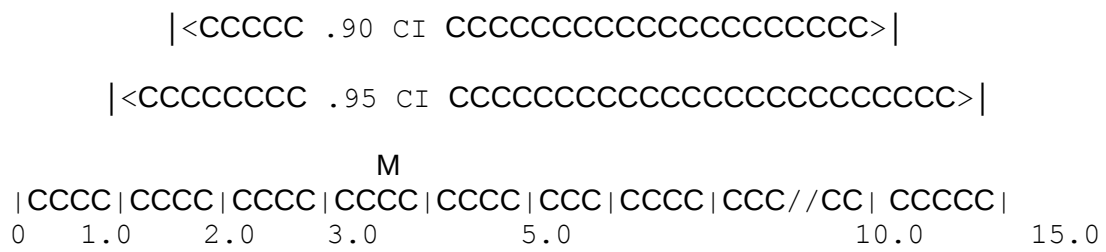
SOURCE: Courtesy of the authors.

EXHIBIT C



SOURCE: Courtesy of the authors.

EXHIBIT D



ODDS RATIO

SOURCE: Courtesy of the authors

EXHIBIT E

SUMMARY OF DR. GINA GENETICIST'S CASE- CONTROL STUDY

Table 1: Individuals with no variant alleles

	Number of cases	Number of controls	Odds Ratio (95% confidence interval)
Exposed	28	110	1.08 (0.66 – 1.78)
Nonexposed	74	314	1.0

SOURCE: Courtesy of the authors.

Table 2: Individuals with one or two variant alleles

	Number of cases	Number of controls	Odds Ratio (95% confidence interval)
Exposed	24	29	3.91 (1.87 – 8.17)
Nonexposed	22	104	1.0

SOURCE: Courtesy of the authors.

The following table was not in Dr. Geneticist's study, but under questioning from the judge she created it from the data in her study. She was unable to do a statistical analysis while testifying.

Table 3: Individuals with Dreion 43 exposure

	Number of cases	Number of controls	Odds Ratio (95% confidence interval)
One or two variant alleles	24	29	3.25 (not calculated)
No variant alleles	28	110	1.0

SOURCE: Courtesy of the authors.