Figure 7.3. PROBABILITY MODELLING: Two Examples

EM8601:

SCIENCE AND THE CITIZEN

The Heart of the Matter

In its final report, the presidential commission charged with examining the explosion of the space shuttle *Challenger* identifies the design flaw that caused the accident and describes in detail the events leading up to the tragedy. It does not describe the underlying causes, within the organization of the National Aeronautics and Space Administration, that made it possible for serious dangers to be ignored. One of the commissioners, Richard P. Feynman of the California Institute of Technology, has addressed this question in a separate document, in which he treats the errors in judgement and execution discovered by the commission as symptoms by which to diagnose larger problems within the space agency. He concludes that NASA's effectiveness in selling its projects to Congress has interfered with its effectiveness as a science and engineering agency.

It is well known by now that the immediate cause of the accident was found to be a faulty seal in one of the joints between sections of the shuttle's right-hand solid-rocket booster. Hot gases eroded a rubber O-ring in the seal and "blew by" it, creating a leak that eventually allowed a plume of flame to escape through the joint and pierce the shuttle's external fuel tank.

The finding came as no surprise. Testimony before the Rogers commission (named for its chairman, William P. Rogers, a former secretary of state) revealed that O-rings in the solid-rocket boosters had been a matter of concern for nearly a decade. Seals are an essential part of the booster, because like all large solid rockets, they are built in sections. There are several reasons for such a design. One is that the fuel is first cast as a liquid, and it might not dry and cure correctly if it were deposited in a single container as large as the shuttle booster. Another reason is that an intact booster rocket would be too large to be transported by rail from the manufacturer to the launch site; because the boosters were made in landlocked Utah, no other means of transportation was available. The particular method of joining the sections that was proposed by the manufacturer, Morton Thiokol, Inc., had been criticized by NASA, however. That was in 1977, when tests first indicated that Thiokol's method of sealing the joints between sections might lead to erosion and leaks.

During the second flight of the space shuttle, in November, 1981, one O-ring in the right-hand solid-rocket booster was eroded, although no gases blew by it. O-rings were eroded during 11 subsequent flights – often in more than one joint – and in nine flights hot gases blew by the "primary" O-ring in at least one joint but did not pass completely through the rest of the seal.

Engineers at Thiokol were alarmed by the unexpected frailty of the seals. In July, 1985, Roger M. Boisjoly, a Thiokol engineer, wrote a memorandum to Robert K. Lund, the vice-president of engineering, "to insure that management is fully aware of the seriousness of the current O-ring erosion problem .... It is my honest and very real fear that if we do not take immediate action to dedicate a team to solve the problem ... we stand in jeopardy of losing a flight along with all the launch pad facilities." A later memorandum, written in October, 1985, by the head of the task force eventually created to solve the problem, begins with the word "HELP!" and ends with "This is a red flag." The engineers' concern came to a head the night before the *Challenger* launch, when, in a teleconference, they tried to convince both the NASA and Thiokol managements not to launch because of the extremely cold temperatures at the launch pad.

Officials tried to understand the erosion by making a mathematical model, based on data from flights on which the O-rings eroded, to predict the amount of damage to be expected under various conditions.

Why were shuttles allowed to fly when critical parts were being damaged in unexpected ways? According to Feynman, managers at NASA and at Thiokol came to regard O-ring erosion as an acceptable risk because O-rings had eroded on previous flights without causing the boosters to fail. Officials noted that in the earlier flights the rings had been eroded by no more than one-third of their radius. Experiments had indicated that an O-ring would have to be eroded by one full radius before it would fail, and so the officials asserted that there was a "safety factor of three." Feynman observes, "This is a strange use of the engineer's term 'safety factor'.... Erosion was a clue that something was wrong. Erosion was not something from which safety can be inferred."

Officials tried to understand the erosion by making a mathematical model, based on data from flights on which the O-rings eroded, to to predict the amount of damage to be expected under various conditions. Feynman discusses the way the model was developed and the final form it took and then adds: "There is nothing much so wrong with this as believing the answer! Uncertainties appear everywhere .... The empirical formula was known to be uncertain, for it did not go through the very data points by which it was determined." NASA used this mathematical model to rationalize flying with ever greater risks. Feynman also discusses the design, testing and certification of the shuttle's main liquid-fuel engine and concludes that here too there was a "slow shift toward decreasing safety factor." In these and other cases, subtly, and often with apparently logical arguments, the criteria are altered so that flights may still be certified in time."

To estimate the chances of a space shuttle's failing, NASA managers substituted what they termed "engineering judgement" for the standard methods of probability. They set the probability of failure at about one chance in 100,000. Working engineers thought the chances were closer to one in 100. "If we are to replace standard numerical probability usage with engineering judgment," Feynman asks, "why do we find such an enormous disparity between the management estimate and the judgment of the engineers?"

Feynman hypothesizes that the fundamental cause of NASA's systemic overconfidence was that a major role of NASA management was to get funding from Congress. To do so, he says, they painted too rosy a picture of what could be accomplished with current technology. At a press conference held when he released his independent remarks, Feynman speculated that "by exaggerating what they said they could do, they got in a position where they didn't want to hear too much about the truth .... The *Challenger* mission was the final accident of a sequence of things in which there was warning after warning after warning that something was wrong .... For 10 years they discussed this problem and didn't do anything about it ... because it was hard for information to come up. But we know the information was there at the lowest levels. Why the engineers are at the lowest levels I have no idea, but the guys who know something about what the world is really like are at the lowest levels of these organizations and the ones who know how to influence other people by telling them how

(*continued overleaf*)

the world would be nice ... they're at the top."

Although Feynman judges NASA management more harshly than the official report, the latter does suggest that NASA's original plans for the shuttle were overambitious: the commitment to provide routine and economical access to space locked the agency into a schedule too tight to be met with the available resources. For example, the inventory of spare parts was not large enough to accommodate the launch schedule, and so each orbiter was made ready for launch by cannibalizing parts from other orbiters. The commission suggests that NASA's desire to make the shuttle the only major U.S. launch system put too much pressure on the program to meet tight schedules and to be able to handle any payload. NASA's can-do attitude, its willingness to undertake challenging tasks at the last minute, also strained the resources of the ground crews and forced NASA officials to focus on the near term at the expense of long-term safety and economy.

Yet the report does not recommend any major changes in the overall structure of the space program, nor does it hold the highest levels of management responsible for the accident; it reserves its strongest criticism for management at Thiokol and at NASA's Marshall Space Flight Center, the division of NASA responsible for the boosters. The report concludes by urging the Administration and the country to continue supporting NASA.

Feynman's report goes on to draw the connection between the over-optimistic attitude of top management and the accident. He concludes by admonishing NASA to be realistic in estimating costs and setting schedules. "If in this way the Government would not support them, then so be it. NASA owes it to the citizens from whom it asks support to be frank, honest and communicative, so that these citizens can make the wisest decisions for the use of their limited resources." His final remark is that of a physicist who is galled to see what he calls "fantasy" enter the realm of science and engineering: "For a successful technology, reality must take precedence over public relations, for nature cannot be fooled."

REFERENCE: *Scientific American* **255**(#2): 62-64 (August, 1986). [DC Library call number: PER T1.S5]

1 In the paragraph beginning near the top of the right-hand column of the article EM8601 reprinted overleaf on page 7.11, there is discussion of *mathematical modelling*, a topic that should bring immediately to mind a maxim of the late Dr. George E. P. Box, a respected U.S. statistician: *All model are wrong, some are useful.*

- Describe briefly the *reason(s)* why this modelling was undertaken.
- Explain briefly why mathematical modelling might seem preferable in this sitution to *experimental* work on the O-ring erosion problem.
  - Describe briefly the potential *danger* associated with mathematical modelling in this situation.
- Explain briefly what you infer in the context of the article from the statement: *The empirical formula was known to be uncertain, for it did not go through the very data points by which it was determined*.
  - Explain briefly whether this statement *necessarily* means that the mathematical model would be of little or no use.

2 In the middle paragraph of the right-hand column overleaf on page 7.11, the article EM8601 states that NASA managers set the probability of failure at about one chance in 100,000 by 'engineering judgement' whereas *Working engineers thought the chances were closer to one in 100*. What reason is suggested in the article for the difference of about three orders of magnitude between these two (personal) probabilities?

EM1201: The Sydney Morning Herald, February 14, 2012

Space engineer's warnings brushed aside before *Challenger* exploded

ROGER BOISJOLY
US ROCKET ENGINEER
25-4-1938 – 6-1-2012

ROGER Boisjoly, an engineer whose warnings of catastrophy were ignored on the eve of the 1986 *Challenger* space shuttle disaster, which killed seven crew members and plunged the US space program into crisis, has died aged 73.

Boisjoly worked at Morton Thiocol, the company that made shuttle booster rockets. In his time he had watched several successful shuttle launches, including *Discovery* on January 24, 1985.

Examiming *Discovery's* discarded boosters, he was horrified that seals in the rockets had been burned through. Only a secondary seal had prevented *Discovery* becoming a fireball.

By the middle of that year, Boisjoly thought he had identified the fault, and wrote a report to alert the company and NASA.

The problem was that in cold temperatures, the seals' rubber stiffened and became more likely to fail. Morton Thiocol formed a study group but, by January 27, 1986, as temperatures dropped to zero on the eve of *Challenger's* blast-off, little progress had been made. In a pre-launch conference, he and four other engineers demanded the launch be postponed.

But their testimony was brushed aside. The next day, unable to watch, his worries seemed unfounded. He had expected the seals to fail on the launch pad. As *Challenger* cleared the launch tower, a colleague whispered to him: "We just dodged a bullet." Seconds later, the spacecraft exploded.

Roger Mark Boisjoly was born in Lowell, Massachusetts, and studied mechanical engineering at the city's university. He then worked for aerospace companies in California on projects including NASA's lunar module and the moon vehicle.

After the *Challenger* disaster, Boisjoly experienced intense feelings of guilt and depression not helped by the fact that many in the business he loved rejected him as an unwelcome whistleblower.

Boisjoly is survived by his wife, the former Roberta Malcolm, and two daughters.

TELEGRAPH

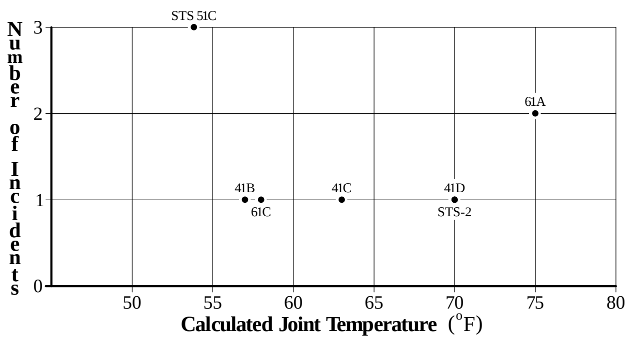
(*continued*)

Figure 7.3. PROBABILITY MODELLING: Two Examples (continued 1)

The excerpt EM8911 given below is reprinted from the *Journal of the American Statistical Association* **84**(#408), pages 945-957, December, 1989 [DC Library call number: PER QA276.A599x]; the introduction of the article, by Siddartha R. Dalal, Edward B. Fowlkes and Bruce Hoadley and entitled *Risk Analysis of the Space Shuttle: Pre-*Challenger *Prediction of Failure*, is the basis of the information below. It is of interest as an illustration of the use of scatter diagrams in the Analysis stage of the FDEAC cycle and of the possible consequence of assessing a trend from only *part* of a set of bivariate data. [Author Bruce Hoadley is interviewed in Program 16 of *Against All Odds: Inside Statistics* in the segment on the *Challenger* disaster.]

EM8911: On the night of January 27, 1986, the night before the space shuttle *Challenger* accident, there was a three-hour teleconference among people at Morton Thiokol (manufacturer of the solid rocket motor), Marshall Space Flight Center [NASA (National Aeronautics and Space Administration) center for motor design control], and Kennedy Space Center. The discussion focused on the forecast of a 31°F temperature for launch time the next morning, and the effect of low temperature on O-ring performance. A data set, Figure 1a below at the left, played an important role in the discussion. Each plotted point represents a shuttle flight that experienced thermal distress in the field-joint O-rings; the x axis shows the joint temperature at launch and the y axis shows the number of O-rings that experienced some thermal distress. The O-rings seal the field joints of the solid rocket motors, which boost the shuttle into orbit. Based on the U-shaped configuration of points (identified by the flight number), it was concluded that there was no evidence from the historical data about a temperature effect.

Figure 1a. Frequency of O-Ring Distress vs. Temperature

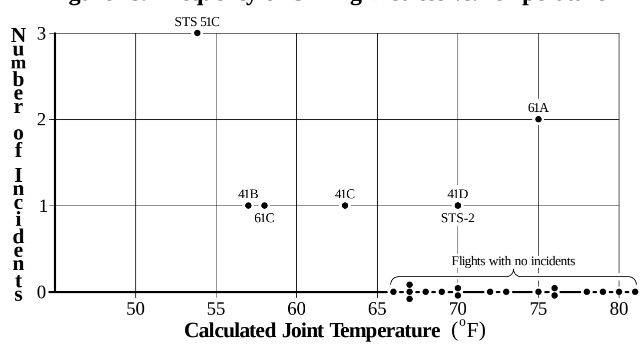


Nevertheless, there was a debate on this issue, and some participants recommended that the launch be postponed until the temperature rose above 53°F – the lowest temperature experienced in previous launches – because the corresponding flight had the highest number of distressed O-rings. Some participants believed, based on the physical evidence, that there *was* a temperature effect on O-ring performance; for example, one of the participants, Roger Boisjoly, stated: *temperature was indeed a discriminator.* In spite of this, the final recommendation of Morton Thiokol was to launch the *Challenger* on schedule. The recommendation transmitted to NASA stated that *Temperature data [are] not conclusive on predicting primary O-ring blowby.* The same telefax stated that *Colder O-rings will have an increased effective durometer ('harder'), and 'Harder' O-rings will take longer to 'seat'* [Presidential Commission Report, Vol. 1 (PC1), p. 97 (Presidential Commission on the Space Shuttle *Challenger* Accident 1986)].

After the accident, a commission was appointed by President R. Reagan to find the cause. The commission was headed by former Secretary of State William Rogers and included some of the most respected names in the scientific and space communities. The commission determined the cause of the accident to be the following: *A combustion gas leak through the right Solid Rocket Motor aft field joint initiated at or shortly after ignition eventually weakened and/or penetrated the External Tank initiating vehicle structural breakup and loss of the Space Shuttle* Challenger *during mission 51-L* (PC1, p. 70). This is the type of failure that was debated the night before the *Challenger* accident.

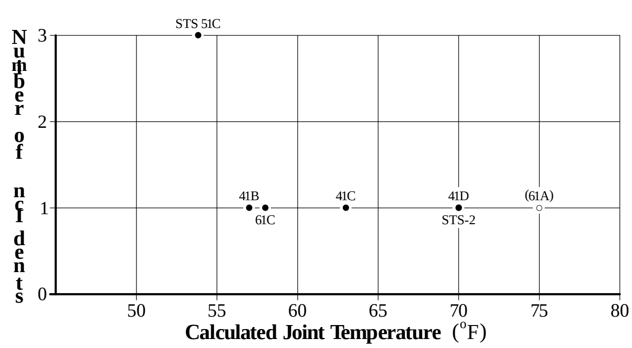
The Rogers Commission (PC1, p. 145) noted that a mistake in the analysis of the thermal distress data – Figure 1a at the left – was that flights with zero incidents were left off the plot because it was felt that these flights did not contribute any information about the temperature effect (see Figure 1b below). The Rogers Commission concluded that *A careful analysis of the flight history of O-ring performance would have revealed the correlation of O-ring damage in low temperature* (PC1, p. 148).

Figure 1b. Frequency of O-Ring Distress vs. Temperature



This article aims to give more substance to this quote and show how statistical science could have provided valuable input to the launch decision process. Clearly, the key question was: *What would have constituted proof that it was unsafe to launch?* Since our model of the phenomeon is stochastic, our answer is necessarily probabilistic. As in the teleconference, a good start would have been an examination of the thermal distress data – Figure 1b – for the presence of a temperature effect. Nevertheless, the most important question was: *What is the probability of catastrophic field-joint failure if we launch tomorrow morning at 31°F?* Both these issues are addressed in the article.

NOTE: It is interesting to speculate, if (hypothetically) the right-most point in Figure 1a above had involved *one* (instead of two) distressed O-rings, whether the modified Figure (shown at the right) would have been interpreted to yield the *correct* Answer (as in Figure 1b at the right above) about the temperature-O-ring distress relationship.



(*continued overleaf*)

In the 1992 article EM9204 from *Newsweek* [March 16, page 62 (Vol. 119, Issue 11)] reprinted below, the matter of primary statistical interest in the context of this Figure 7.3 is the dispute arising over the form of a statistical model appropriate for analyzing data on lead levels in teeth and IQ; the work was part of data-based investigating over many years by Dr. Herbert Needleman and his colleagues of possible effects of low-level lead exposure on children's intellectual development. The article describes a claim that age should have been included in the statistical model as an explanatory variate.

EM9204:

Lead, Lies and Data Tape

A nasty academic fight becomes a federal case

By Sharon Begley

Scientific misconduct is exceedingly rare and extremely serious. Charges have been brought alleging plagiarism or faking data or falsifying results. The latest case, however, involves the manner in which a researcher strung together a set of equations in order to find a message hidden in a stack of raw data. To reach for a metaphor, this is like bringing a felony indictment for jaywalking.

Two psychologists, both of whom have testified for the lead industry and one of whom has received tens of thousands of dollars in research grants from the industry, have filed misconduct charges against the scientist who first linked "low" levels of lead to cognitive problems in children. They don't suspect that Herbert Needleman of the University of Pittsburgh stole, faked or fabricated data. Rather, they say, he selected the data and the statistical model – the equations for analyzing those data – that show lead in the worst possible light. That's a dispute usually aired in research journals. Now it's become a federal case – and even those scientists most diligent about pursuing misconduct are uneasy. "If it gets to which statistical model is appropriate", a leading government fraud-buster told *Newsweek*, "it gets real hard to believe a misconduct charge."

The case began last year. Psychologists Sandra Scarr of the University of Virginia and Claire Ernhart of Case Western Reserve University filed charges of scientific misconduct against Needleman with the National Institutes of Health. The allegations centre on a 1979 paper. It describes how Needleman and colleagues measured the lead in baby teeth, looking for a link between lead and intelligence. NIH told Pittsburgh to convene a panel of inquiry. The panel's report, submitted in December and obtained by *Newsweek*, found that Needleman didn't fabricate, falsify or plagiarize." It did have problems with how he decided whether or not to include particular children in his analysis, but called this "a result of lack of scientific rigour rather than the presence of scientific misconduct." The panel found Needleman's statistical model "questionable", though. On that basis, the university launched an investigation.

Scarr, Ernhart and the Pittsburgh panel all condemn Needleman for not using a different model – one that, say, factored in the age of each child. If he had, they say, lead would not have had an impact on IQ. But last year Environmental Protection Agency scientist (and recipient of a MacArthur Foundation "genius" award) Joel Schwartz re-analyzed Needleman's data. He factored in age explicitly. "I found essentially identical results", he says.

Flawed sampling? Another criticism addresses whether Needleman ignored data he didn't like. Scarr alleges that he looked at the children's lead levels and IQ score, and only then "decided in or out for each child." In fact, "the reasons for exclusion can be found in the protocol", says econometrician Hugh Pitcher of Battelle Memorial Institute, who analyzed the 1979 data when he was at EPA. They include such things as the child's having a head injury. The selection, says Pitcher, was done before the researchers knew the kids' IQs.

This is not to say Needleman's work was perfect, just that any lapses did not change the outcome. Ernhart insists this is not good enough. "He doesn't feel it's necessary to do things the way you're supposed to", she says. "You have the sense that he was going to demonstrate the effects of lead no matter what."

The panel ... did have problems with how he decided whether or not to include particular children in his analysis, but called this 'a result of lack of scientific rigour rather than the presence of scientific misconduct.' The panel found Needleman's statistical model 'questionable', though. On that basis, the university launched an investigation.

How does this case affect lead policy? "We don't even use Needleman's study any more", says EPA's Schwartz: it has been superseded by research showing effects of lead at even lower levels (*Newsweek*, July 15, 1991). The politicization of misconduct may be just starting, though. Crying fraud, says an NIH scientist, "can be used to railroad people you don't like."

EM1701:

Dr. Herbert Needleman, Who Saw Lead's Wider Harm to Children, Dies at 89

By Benedict Carey
July 27, 2017

Dr. Herbert Needleman, whose studies of children exposed to low levels of lead prompted regulations that limited or banned the metal in a range of common products, like gasoline and paint, and set a standard for the modern study of environmental toxins, died on July 18 in Pittsburgh. He was 89.

His son, Dr. Joshua Needleman, said the cause was lung failure resulting from edema, an excess of fluid.

Dr. Needleman was working at a community psychiatric clinic in North Philadelphia after medical school when he met a young man who would become a touchstone for a crusading career. The boy approached Dr. Needleman and explained his ambitions, which were large, even as the boy struggled with words. He was bright and open; nonetheless he had deficits that struck Dr. Needleman as similar to those found in children with lead poisoning.

"I thought, how many of these kids who are coming to the clinic are in fact a missed case of lead poisoning?" he said in a later interview. His clinic office overlooked a school playground; the view gave him an idea.

Doctors had long known that exposure to high doses of lead caused mental lapses,

(*continued*)

Figure 7.3. PROBABILITY MODELLING: Two Examples (continued 2)

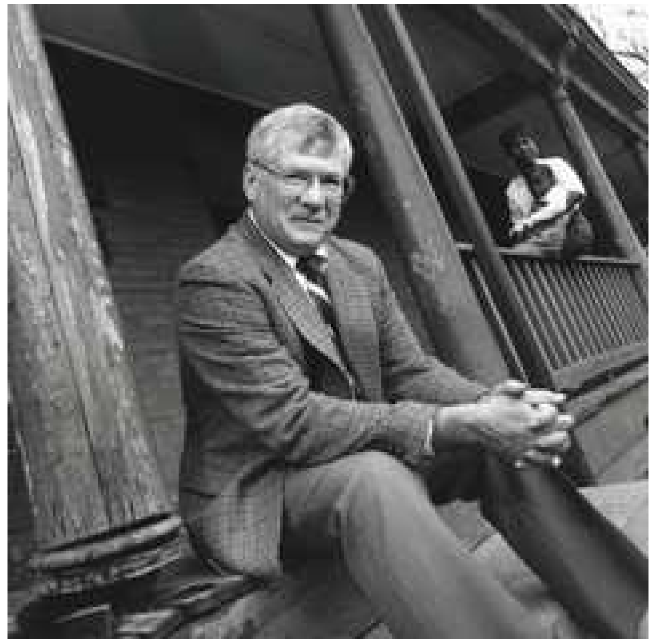
even permanent brain damage and death. But what about the low-level exposure that many children, like the ones playing in the yard, absorbed every day – merely by living in older urban neighborhoods thick with lead paint and industrial contamination?

No one knew. No one could study the effects carefully, because the available tests for lead exposure were of hair, blood, or fingernails – each flawed in its own way. Bone is the most accurate long-term repository: Once absorbed into the body, lead circulates in the blood and accumulates in the skeleton. But taking bone samples – biopsies – is painful and hardly justifiable for the sake of a hypothesis, especially in young children.

Yet Dr. Needleman had seen an earlier study of lead poisoning, a small one, which measured accumulated lead exposure in teeth. Teeth are a part of the human skeleton. And young children shed them.

"That was the insight that changed everything," said Dr. Benard Goldstein, former dean of the University of Pittsburgh's graduate school of public health. "Herb became the Tooth Fairy."

In a series of studies – small ones in Philadelphia and a much larger project in the Boston area – Dr. Needleman offered children aged 6 and 7 small rewards for their loose teeth, once they had fallen out. Those teeth told a story: Children living in poor urban neighborhoods had lead levels five times higher, on average, than those of their peers in the suburbs.



Dr. Herbert Needleman in an undated photo. Credit Jim Harrison/Heinz Awards

In a landmark 1979 paper in *The New England Journal of Medicine*, which included more than 2,000 children, he and his co-authors – he was then at Harvard – explained the associated consequences in devastating detail. Children whose accumulated exposure to lead was highest in the group scored four points lower on an I.Q. test than youngsters whose exposure was at the lowest end.

Teachers rated the high-exposure children as having a host of classroom issues, including attention deficits and behavior problems. A follow-up of the same children a decade later found a correlation between high lead levels and reading delays.

"It's not like you can look at one kid and spot a four-point difference in I.Q., and say, 'O.K., we know lead caused this,'" said Linda Birnbaum, director of the National Institute of Environmental Health Sciences, in Durham, N.C. "It's a population effect; you have to have a population of the right kids and ask the right questions. That's what Dr. Needleman did, and it has become a model" for subsequent research.

In the decade that followed – and over strenuous opposition from the lead industry – the findings of Dr. Needleman and others spurred stiffer regulation of lead in gas, tin cans, paint, household pipes and other products. The federal health authorities now consider lead at any level unsafe for children.

The story did not end there, however; it had an important political coda, one that itself set a template for commercial backlash against research into possible environmental hazards. The lead industry's opposition culminated in the late 1980s in an attack on Dr. Needleman's research and character. A pair of psychologists approached him for his data from the 1979 study, as part of a court case in which they were testifying on behalf of a lead smelting company. The psychologists proceeded to accuse him of scientific misconduct, an accusation taken up by the newly formed federal Office for Scientific Integrity.

In testimony, Dr. Needleman acknowledged that he had made some math mistakes in his analysis but that those errors were minor and did not change the findings. Investigators eventually agreed and dismissed all charges. But not before the University of Pittsburgh, where he was then on the faculty, conducted its own investigation and locked him out of his own files, putting bars on his file cabinets. He was cleared in that investigation as well.

"You have no idea what he went through," said Dr. Philip Landrigan, the dean for global health at the Icahn School of Medicine at Mount Sinai in New York. "He swung in the wind for those years, and he never backed down. I don't use this word often, but hero is appropriate in Herb's case."

Herbert Leroy Needleman was born on Dec. 13, 1927, in Philadelphia, one of two sons of Joseph and the former Sonia Shupak. His father sold furniture; his mother, whose family owned a pickle business, ran the household.

He graduated from Muhlenberg College in Allentown, Pa., in 1948, and received a medical degree in 1952 from the University of Pennsylvania. After serving in the Army, at Fort Meade, Md., he completed residencies in both pediatrics and psychiatry in Philadelphia.

In the 1960s, while teaching at Temple University, he became an active opponent of the Vietnam War. He went to jail at least once for participating in a protest; and he was chairman of a group called the Committee of Responsibility to Save War-Burned and War-Injured Vietnamese Children, which brought injured youngsters to the United States for medical care.

"One of those children lived with us, at our house," said his son, Joshua. "I was only 4 years old, but I remember."

Dr. Needleman's first marriage, to Shirley Weinstein, ended in divorce. He is survived by his wife of 54 years, the former Roberta Pizor; a son from his first marriage, Samuel; two children from his second marriage, Joshua and Sara Needleman Kline; as well as seven grandchildren and three great-grandchildren.

In a 2005 interview, Dr. Needleman was asked whether the attack on his credibility was meant to scare off other researchers looking into environmental toxins. "If this is what happens to me, what is going to happen to someone who doesn't have tenure?" he replied.

"I'm worried that people who are trying to get a niche and don't have tenure are asked to do things they question the ethics of," he continued. "They are intimidated. It's a real force."

A version of this article appears in print on July 28, 2017, on Page A25 of the New York edition of **The New York Times** with the headline: *Dr. Herbert Needleman, Who Saw Lead's Harm To Children, Dies at 89.*

(*continued overleaf*)

A third example of pobability modelling, involving the normal distribution, is given in Figure 5.8 on pages 5.19 and 5.20 in Part 5 of these Course Materials.

- The context is predictions, in 1990, of the number of new AIDS cases each year in the U.S. over the following decade. On the overleaf side (page 5.20) of Figure 5.8, more recent data are given which indicate the level of inaccuracy of these predictions and, in doing so, show the usefulness of the normal model in this context.