Objectivity and Ethics in Environmental Health Science

Steve Wing

Department of Epidemiology, School of Public Health, University of North Carolina, Chapel Hill, North Carolina, USA

During the past several decades, philosophers of science and scientists themselves have become increasingly aware of the complex ways in which scientific knowledge is shaped by its social context. This awareness has called into question traditional notions of objectivity. Working scientists need an understanding of their own practice that avoids the naïve myth that science can become objective by avoiding social influences as well as the reductionist view that its content is determined simply by economic interests. A nuanced perspective on this process can improve research ethics and increase the capacity of science to contribute to equitable public policy, especially in areas such as environmental and occupational health, which have direct implications for profits, regulation, legal responsibility, and social justice. I discuss research into health effects of the 1979 accident at Three Mile Island near Harrisburg, Pennsylvania, USA, as an example of how scientific explanations are shaped by social concepts, norms, and preconceptions. I describe how a scientific practice that developed under the influence of medical and nuclear physics interacted with observations made by exposed community members to affect research questions, the interpretation of evidence, inferences about biological mechanisms in disease causation, and the use of evidence in litigation. By considering the history and philosophy of their disciplines, practicing researchers can increase the rigor, objectivity, and social responsibility of environmental health science. *Key words:* cancer, chance, dose reconstruction, environmental justice, epidemiology, ionizing radiation, research ethics, significance testing, Three Mile Island. *Environ Health Perspect* 111:1809–1818 (2003). doi:10.1289/ehp.6200 available via *http://dx.doi.org/* [Online 19 June 2003]

In politics, policy, and law, science has emerged as an alternative to folk traditions, religion, and superstition as a way to understand and manipulate the material world. The value and prestige of the sciences derive not only from their erudition, explanatory power, and applied technology but from their perceived objectivity. A common view among scientists and lay persons alike is that scientific objectivity is a consequence of standardized methods of quantitative observation and experimentation. The scientific method, by removing subjectivity and social influence, yields knowledge that is ostensibly trustworthy and objective.

Despite the persistence of this view, historians, philosophers, and scientists themselves have shown that it does not provide an adequate account of the production of scientific knowledge (Harding 1991; Holtzman 1981; Hubbard 1990; Kuhn 1970). There are several reasons why method cannot remove social influences from science. First, the content and methods of science are formed in relation to answering questions or testing hypotheses that are socially embedded. Second, scientific explanation requires language, concepts, and models that are cultural products. Although all sciences expend considerable effort to rationalize concepts and terminology, these tools of inquiry are inevitably shaped and transformed by historical forces. Therefore, scientists cannot even see the world, much less provide explanations of its workings, without a socially formed perspective. Ironically, the belief that science could attain objectivity through independence from social forces places science in the role of a religion's omniscient God (Harding 1991). The illogic of the naïve view of scientific objectivity has been described in physics, genetics, and epidemiology, as well as in mathematics and statistics (Armstrong 1999; Hubbard 1990; Keller 1992, 1995; Kuhn 1970; Levins 1979; Levins and Lewontin 1985).

The reluctance of scientists to acknowledge the shaping of their work by social forces and their ongoing avowal of science as value-free can be viewed as a self-serving argument against public oversight (Keller 1995). However, even among scientists who accept an ethic of social responsibility, attempts to salvage naïve objectivity persist because the alternative is perceived to be a judgmental relativism in which there is no basis for adjudicating competing claims. Such relativism is anathema to the most basic assumptions of science: that a real world exists independent of human cognition, and that theories and hypotheses about that world can be tested by controlled methods of observation and experimentation. A logical alternative to both judgmental relativism and naïve objectivity is "strong objectivity," an objectivity attained through revealing, rather than concealing, the cultural content and social forces that are embedded in science (Harding 1991). To practice strong objectivity, scientists must consider not only the technical aspects of their discipline, but must also take into account its history, conceptual foundations, preconceptions, taboos, and the social forces that shape its content and application. This requires scientists to distinguish truth, in the form of statements about the world made by people, from reality, the world itself (Hubbard 1990; Rorty 1989), and to be self-critical about the ways scientists create truths and facts in relation to the real world that they study.

This article explores how a contextual research practice can improve the rigor, ethics, and social responsibility of environmental health science. I use as a case example research on cancer incidence after the 1979 nuclear accident at Three Mile Island (TMI) near Harrisburg, Pennsylvania, USA. I consider how unarticulated cultural views about the reliability of assumptions and evidence shaped the framing of questions, the design of research, and the interpretation of findings in the scientific literature as well as in the courtroom.

Accident at Three Mile Island

Three Mile Island is in the Susquehanna River about 16 km (10 miles) from Harrisburg, the capital of Pennsylvania, where the Susquehanna cuts across parallel ranges of hills that rise hundreds of meters above the river. Several smaller towns are located along the river near to TMI (Figure 1). Dairy and other farms are common in the area, which has a strong agricultural tradition (Figure 2).

At 4:00 A.M. on 28 March 1979, a series of events began that led to a loss of control of the nuclear chain reaction in the TMI Unit 2 reactor. For several days it was not clear how or when the reactor could be shut down. Uncontrolled releases of radioactivity to the environment began shortly after 4:00 A.M. on 28 March. Within hours, radiation monitors at the plant went off-scale because radiation

levels exceeded the instruments' measurement capacity (Macleod 1981). Although thermoluminescent dosimeters were placed off-site on 30 March, there were large angular gaps between the monitors (Beyea 1985). As a result, there was little information about early releases and poor capacity to detect narrow plumes with low dispersion. At the time the region was experiencing unusually balmy temperatures and low winds as an upper-level cold air mass kept lower-level warm air from rising—ideal conditions for trapping radioactive emissions (Steinacker and Vergeiner 2002). Xenon-133 from TMI was detected in Albany, New York (Wahlen et al. 1980).

The possibility that a hydrogen explosion in the reactor containment or a meltdown of the reactor core would result in high-level radiation exposures generated great fear and anxiety among officials and the public (Del Tredici 1980; Gray and Rosen 1982). Lack of knowledge about details of the plant's condition, lack of experience with this type of situation, and nonfunctional radiation monitors gave rise to conflicting reports about the severity of the accident and its threat to the public. The Nuclear Regulatory Commission (NRC) reported a reading of 3,000 millirads per hour taken above the plant on 29 March (NRC 1979b). About 5–6% of people within 5 miles of the plant left during the first 2 days of the accident. After Governor Thornburgh's 30 March order to evacuate pregnant women and children from the 5-mile area, nearly 50% of residents left (Houts and Goldhaber 1981).

On 1 April, the hydrogen gas bubble began to dissipate, and concerns of imminent danger diminished. Industry and government representatives assured the public that only small quantities of radiation had been released and that exposures were far below levels that could affect health. The NRC and a presidential commission released reports indicating that the maximum possible off-site radiation dose was less than average annual background levels (NRC 1979a; President's Commission on the Accident at Three Mile Island 1979).

The official position that high-level radiation exposures were impossible was questioned by hundreds of local residents who reported metallic taste, erythema, nausea, vomiting, diarrhea, hair loss, deaths of pets and farm and wild animals, and damage to plants (Del Tredici 1980; Molholt 1985; Osborn 1996; Three Mile Island Alert 1999). Many of these phenomena could be caused by radiation; however, the maximum possible dose was officially reported to be orders of magnitude less than the dose needed to produce acute symptoms. Residents were told that their symptoms were due to stress. People who pressed their concerns about radiation were treated as though they had psychologic problems.

Epidemiology of the Accident

Health studies at TMI began to be planned soon after the immediate danger had ended. In June 1979, the Pennsylvania Department of Health, working with the Centers for Disease Control and the U.S. Census Bureau, conducted a special census of residents living within 5 miles of TMI (Goldhaber et al. 1983c). The University of Pittsburgh's Department of Radiation Health provided estimates of radiation doses for 5-mile area residents "for educational, public relations and defensive epidemiology purposes" (Gur et al. 1983). The demographics of evacuation (Goldhaber et al. 1983a), medical care use, and spontaneous abortion (Goldhaber et al. 1983b) were studied. Many studies of stress have been published, making the 1979 accident at TMI one of the best-studied cases of psychologic response to disaster and evacuation (Baum 1990; Baum et al. 1983, 1993; Cleary and Houts 1984; Cornely and Bromet 1986; Davidson et al. 1987; Dew and Bromet 1993; Dew et al. 1987a, 1987b; Fabrikant 1983; Gatchel et al. 1985; Houts et al. 1991; Houts and Goldhaber 1981; McKinnon et al. 1989; Prince-Embury and Rooney 1988; Schaeffer and Baum 1984).

Few studies took on the topic of radiation exposures. Ionizing radiation is considered to be one of the best-understood carcinogens, and scientists asserted that doses at TMI had been too low to produce any observable effects on cancer. Population dose estimates and quantitative cancer risk estimates based on studies of the survivors of the atomic bombing of Hiroshima and Nagasaki yielded a prediction of, at most, one accident-related cancer death in the lifetimes of the population in the 50-mile area (Hatch et al. 1990). Therefore, there was no scientific reason to study health effects of radiation.

Yet, concerns persisted that some areas near TMI had been exposed to high radiation levels during the accident. In 1984 Carl Johnson, former Director of Health in Jefferson County, Colorado, site of the Rocky Flats nuclear weapons plant, spoke at a public meeting at the Pennsylvania State University–Harrisburg campus in Middletown, Pennsylvania. He described reports of symptoms consistent with high-level radiation exposure that had been experienced during the accident in several hilltop neighborhoods near TMI (Johnson C. Unpublished data). Johnson's description piqued the interest of Marjorie Aamodt, an experimental psychologist by training, who, with her husband, Norman Aamodt, an engineer, had been participating in hearings regarding the restart of TMI Unit 1.

In the spring of 1984, Marjorie Aamodt initiated a household survey in three hilltop communities with a total population of about 450 people. All three neighborhoods had unobstructed views of TMI at distances of between 3 and 8 miles (Aamodt and Aamodt 1984). Two of the communities were areas

Figure 1. View of TMI from the roof of an apartment building in Middletown, Pennsylvania, 19 July 1979. Photograph copyright of Robert Del Tredici (Del Tredici 1980) and used with permission.

Figure 2. Farm in Londonderry Township near TMI, 4 July 1979. Photograph copyright Robert Del Tredici (Del Tredici 1980) and used with permission.

that Dr. Johnson had identified. Using a structured questionnaire designed with input from Dr. Johnson, Ms. Aamodt and several volunteers interviewed residents about symptoms and diseases experienced during and after the accident. They obtained descriptions of metallic taste, nausea, vomiting, hair loss, and erythema, almost all from people who had been out-of-doors, as well as information on the occurrences of cancers, cardiovascular diseases, reproductive problems, dermatologic conditions, and ruptured/collapsed organs. Residents reported 19 cancer deaths during 1980–1984, compared with an expected number of 2.6. In June 1984, the Aamodts submitted a report to the NRC proceeding on the competency of the utility to conduct surveillance of radiation releases (Aamodt and Aamodt 1984).

The Aamodt survey soon came to the attention of the TMI Public Health Fund. The Fund, financed by the nuclear industry and administered by the Federal District Court in Harrisburg, had been created in 1981 as part of a settlement for economic losses from the accident. Scientific advisors to the Fund verified several aspects of the Aamodt study, including the ascertainment of cancer deaths and calculation of expected deaths, and recommended a more comprehensive study of cancer in the TMI area. The Fund chose a team led by Mervin Susser, a highly renowned epidemiologist from Columbia University, New York, New York, to design and conduct the cancer study. The Columbia investigators proposed an innovative design that avoided several common problems that can lead to ambiguous results in environmental epidemiology (Hatch et al. 1990).

Because concerns about cancer among both patients and physicians could have resulted in earlier detection of cancer and cause higher incidence rates as an artifact of publicity, the Columbia group did not compare TMI area residents to an unexposed control group from another area. Rather, they divided the 10-mile area into small study blocks, each of which was assigned an accident dose based on a state-of-the-art dispersion model that considered release estimates, meteorologic, and topographic data (Beyea and Hatch 1999). Dose estimates for the 69 study blocks varied by more than three orders of magnitude. This permitted a comparison of cancer rates along a continuum from low to high exposure areas, all of which had a similar potential for early detection of cancer. Although the accident occurred in 1979, incident cancer cases were identified for the period 1975–1985, making it possible to evaluate the variation in cancer rates that existed in the area both before and after the accident. This design feature was important because cancer rates show significant geographic variability, and it

would have been a mistake to attribute high cancer rates in a more exposed area to accident emissions if rates there were already higher prior to the accident. Population counts according to age and gender for each year from 1975 to 1985 were derived from census data. The Columbia design permitted an evaluation of the relationship between estimated accident dose and cancer incidence for the population within 10 miles of TMI, with adjustment (control) for preaccident variation in cancer incidence. Estimation of dose–response relationships with adjustment for differences in risk between the exposure groups prior to exposure is rarely possible in environmental epidemiology.

Two publications by Columbia investigators describe cancer incidence in relation to the TMI accident (Hatch et al. 1990, 1991). Hatch et al. (1990) reported positive associations between accident doses and non-Hodgkin's lymphoma, lung cancer, and all cancers combined. Leukemia, analyzed separately for children and adults, was also positively associated with accident dose. However, these estimates lacked statistical precision because of small numbers (54 cases at all ages combined). The authors reasoned that results did not "provide convincing evidence" that TMI radiation releases had influenced cancer in the area. Among the considerations weighing against a causal interpretation were

> the lack of effects on the cancers believed to be most radiosensitive and the indeterminate effects on children . . . the low estimates of radiation exposure and the brief interval since exposure occurred.

They continued,

> Pending a demonstration that very low dose gamma radiation can act as a tumor promoter or the identification of another late-stage carcinogen in the effluent stream, an effect of plant emissions in producing the unusual patterns of lung cancer and non-Hodgkin's lymphoma appears unlikely, and alternative explanations need to be considered.

One alternative explanation was considered in a second report (Hatch et al. 1991). Using distance from TMI as the measure of psychologic stress, Hatch et al. considered the hypothesis that increased cancer rates after the accident were caused by stress. Findings were equivocal because of the speculative nature of the mechanism of cancer promotion by stress-induced neuroendocrine dysfunction as well as lack of a specific measure of stress.

Reanalysis of the Cancer Incidence Study

By the time the Columbia studies were published, a lawsuit alleging health damages from radiation released in the TMI accident had been under way for several years. Approximately 2,000 plaintiffs argued that emissions of radioactive gases during the

accident were much larger than had been stated by industry and government officials; meteorologic conditions and hilly terrain had caused the radioactive gases to disperse in narrow plumes; and these intense plumes had exposed small areas of the surrounding countryside to high radiation doses, resulting in health impacts including cancer (Merwin et al. 2001). Marjorie and Norman Aamodt were consulting for plaintiffs' attorneys and asked to meet with me to discuss the litigation. They provided documents including their health survey, sworn affidavits from TMI neighbors, analyses of local mortality records, scientific articles, government reports, and letters and memoranda from scientists and government officials suggesting that radiation releases and doses from the accident had been substantial. They asked me to provide epidemiologic support for the plaintiffs in the suit.

I was wary of becoming involved in the lawsuit. Although I had not thoroughly studied the TMI accident, I knew that allegations of high radiation doses at TMI were considered by mainstream radiation scientists to be a product of radiation phobia or efforts to extort money from a blameless industry. Years of collaboration with epidemiologists and health physicists aligned with the U.S. Department of Energy (DOE) Health and Mortality Study had familiarized me with a culture in which concerned workers and community members, as well as scientists who claimed there was evidence of health effects of low-level radiation, were viewed as threats to the nuclear industry (Morgan 1992). Taking the plaintiffs' allegations seriously enough to become involved in any professional capacity might expose me to ostracism and loss of scientific credibility.

However, I was impressed with the intelligence and humanity of the Aamodts and with their thoughtful compilation of evidence. As in all research of this type, important measurements of interest had not been made, case reports, statistical observations, and related records were not entirely consistent, and mechanisms for some putative effects were uncertain. Nevertheless, their scenario of higher-than-reported doses did not seem implausible. My reaction to the Aamodts' work was not only a function of evidence suggestive of high releases; it was also a function of a willingness to consider the possibility that official conclusions might be in need of revision.

My personal experience, as well as study of the history and current practices of radiation epidemiology, had led me to adopt a skeptical attitude toward official assumptions and logic. On a small scale, our collaborative team engaged in the DOE Health and Mortality Studies had been assured repeatedly by industry and federal officials that no records existed that could account for a gap

we had found in annual radiation dose records. However, shortly after publication of our report that the existing workers' badge readings were related to their cancer mortality (Wing et al. 1991), over 14,000 radiation records that we had been seeking for more than 2 years were provided to us (Wing et al. 1994). On a larger scale, major radioactive releases from nuclear weapons sites had been concealed from the public for decades (Thomas 2001); government-funded scientists had conducted human radiation experimentation without informed consent (Advisory Committee on Human Radiation Experiments 1995; McCally et al. 1994); risks to workers and the public had been withheld because of concerns about litigation and loss of public support for the nuclear industry (Advisory Committee on Human Radiation Experiments 1995; Makhijani et al. 1995; Office of Technology Assessment 1991; Sterling 1980); and epidemiologists who deviated from status quo views about radiation and health had lost funding and suffered professionally (Greenberg 1991; Lyon 1999; Morgan 1992; Stewart and Kneale 1991; Thomas 2001; Wilkinson 1999). At the scale of the scientific culture itself, there had long been a lack of nuanced scientific logic in the deference given by radiation epidemiologists to quantitative estimates of radiation risks based on the world's most studied radiation-exposed population, the Japanese A-bomb survivors (Stewart 2000; Wing et al. 1999).

Given these facts, it seemed possible that the full story about radiation releases, doses, and health effects from the TMI accident had yet to emerge. Although I had never previously participated in litigation, I was aware that some of my colleagues had provided testimony for the defendants in radiation cases. Knowing that the defendants in the TMI case had many experienced experts at their disposal and that it was difficult for plaintiffs to find help, I agreed to examine epidemiologic data related to the case.

One of the first tasks I undertook was a review of published studies on "mass hysteria," a medical term for outbreaks of psychogenic illness with no physical etiology (Brodsky 1988; Donnell et al. 1989; Faust and Brilliant 1981; Hefez 1985; Simon et al. 1990; Small and Borus 1987). There is no doubt that the TMI accident had created high stress, even panic nor is there any doubt that stress can produce physical symptoms. Although some symptoms reported by TMI area residents—nausea and vomiting, for instance—are commonly reported in the literature on outbreaks of psychogenic symptoms, other symptoms, most notably metallic taste and hair loss, are not. Furthermore, one hallmark of mass psychogenic illness is that it occurs in public places where people witness the symptoms of

others. At TMI most symptoms were reported by people who had not been in public, some of whom said they had not even heard about the accident at the time of their symptoms. Thus, although it seemed plausible that some psychosomatic illness would have occurred at TMI, most reports did not fit the classic scenario of mass psychogenic illness. Nonhuman occurrences were obviously not psychogenic, although reports could have resulted from increased vigilance or altered perception due to the accident. I did not attempt to validate independently cases of unusual mortalities or abnormalities in animals and plants, and lack of ongoing systematic surveillance precluded comparisons with baseline (preaccident) rates; however, the detail and quality of observation left me unable to dismiss these reports.

The Aamodts were interested in further examination of cancer incidence data assembled by Columbia. The Columbia investigators had not acknowledged the possibility that community members' symptoms might have been a sign of significant amounts of radiation. They assumed from the beginning that accident doses were below average background levels, and they did not analyze data for the hilltop communities in the 1984 health survey. The plaintiffs' legal team asked me to reanalyze the Columbia cancer incidence data to check whether study block boundaries had been constructed in a way that might have obscured clusters and to evaluate whether excess cancers may have occurred in areas where there had been reports of acute symptoms and other unusual phenomena.

Although skepticism about the Columbia study was understandable, the study was, in principle, well designed, and I knew that the investigators were highly respected. However, recognizing that there were several potentially interesting omissions from the published results, I agreed to conduct reanalyses if the data could be released through the court. The court decided not to disclose the locations of residence of cancer cases on the grounds that this would violate patient confidentiality; however, the court did direct that we be given, for each study block, *a*) dose estimates, *b*) numbers of cases of specific types of cancer, *c*) population counts, and *d*) average levels of education, income, and population density. Estimates of doses from gamma radiation for each study block were given in relative units that ranged from 0 to 1,666. The values were not assigned a unit of measurement (e.g., Sv or rem) but were calculated on a ratio scale such that a value of 10 was twice that of 5, and a value of 1,000 was twice that of 500 (Hatch et al. 1990).

We considered different primary hypotheses and used a primary analytical method different from the one used by the original investigators (Wing et al. 1997c). Because ionizing radiation

is a mutagen related to most if not all malignancies, and because higher doses of radiation can lead to immune suppression and promotion of initiated cancers, we considered all cancer as one primary outcome. We examined lung cancer because beta-emitting radioactive gases in the accident plumes could have produced higher doses to the lung than to other organs. We considered leukemia, including chronic lymphocytic leukemia (which had been omitted in the Columbia reports) because studies of high-dose radiation have shown that leukemia is more radiation sensitive and appears sooner after exposures than solid tumors. Although leukemia was a primary outcome of the Columbia study, incidence among children and adults had been analyzed separately, reducing an already small sample size. More important, the childhood cancer analyses considered children conceived after the accident as exposed, potentially diluting any differences in cancer incidence between exposed and unexposed children. Finally, we used baseline (preaccident) cancer rates rather than socioeconomic status variables (education, income, and population density) as the primary method to control for potentially confounding differences in other cancer risk factors between more- and less-exposed populations within the 10-mile area (Wing et al. 1997c).

Our analyses corrected for two problems that had affected the Columbia results. One of their published analyses included duplicate case counts that we were able to eliminate from the reanalysis. In addition, the original preaccident period was defined as 1975–1979. However, case ascertainment for 1975 was incomplete. This led to an underestimation of cancer incidence in the preaccident period and an overestimate of the increase in cancer following the accident. To correct this problem, we redefined the preaccident period as 1976–1979 (Wing et al. 1997b, 1997c).

We found positive relationships between accident dose estimates and cancer rates for all three categories of cancer. The slope of the dose–response estimates was largest for leukemia, intermediate for lung cancer, and smallest for all cancers. Estimates were larger for cancers that occurred in 1984–1985 (a 5-year lag) than for cancers that occurred in 1981–1985 (a 2-year lag), and they were larger when statistical adjustments were made for differences in socioeconomic status between areas of low and high dose. Lung cancer showed the most consistent dose–response relationship across levels of dose. Figure 3 shows dose estimates in relation to lung cancer rates, based on the 440 cases diagnosed in the 10-mile area during 1981–1985, adjusted for preaccident variation in lung cancer incidence, but not for socioeconomic status. The height of the bars represents the difference between the observed numbers of cases at each dose

level and the number that would have occurred if each area had experienced the average lung cancer rates of the 10-mile area population as a whole.

Epidemiologic Evidence in Court

Scientific rigor and objectivity have a special value in courts of law, which specify rules for deciding what constitutes scientific evidence and whether scientific evidence is reliable and therefore admissible at trial. Since 1993, the standards for admissibility of scientific evidence in federal courts have been based on the U.S. Supreme Court's ruling in Daubert *v.* Merrell Dow Pharmaceuticals, Inc. (1993). Judges decide whether expert testimony can be considered in litigation based on criteria expounded in Daubert, other cases, and in the Federal Rules of Evidence:

1) there is a testable hypothesis
2) the methodology has been subjected to peer review
3) the results have an acceptable "known or potential rate of error"
4) standards are used to control the reliability of the methodology
5) there is adequate support and acceptance of the methodology by a scientific community
6) the testimony relies on facts or data used by other experts in the discipline
7) the expert's professional qualifications are adequate
8) the methods have the potential to be used for purposes outside the courtroom (Merwin et al. 2001)

Exclusion of evidence that does not meet these standards is intended to protect juries from being exposed to untrustworthy scientific testimony.

Our reanalysis of the TMI cancer incidence data was submitted to the District Court for the Middle District of Pennsylvania in support of the plaintiff's allegations. My testimony about relationships between estimated doses and cancer incidence in the 10-mile area was being used as evidence that significant exposure had occurred, not as evidence that a particular plaintiff's cancer was caused by radiation. The defendants argued that our study was unreliable and that the judge should exclude it from trial on grounds that it did not satisfy the Daubert criteria.
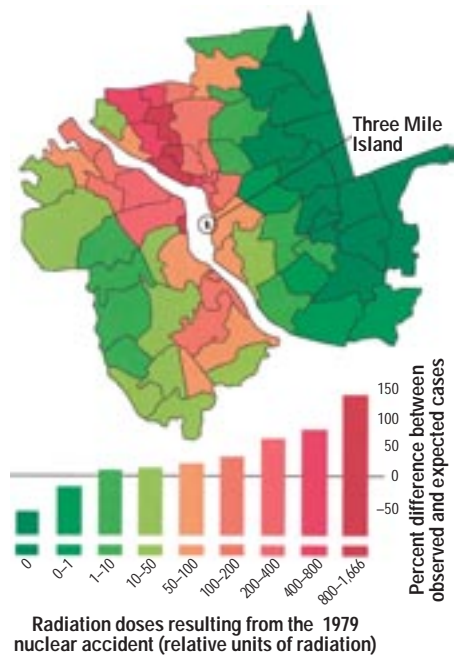
The court considered the admissibility of our reanalysis based on our report describing the study's hypotheses, design, materials, analytical methods, results, and conclusions; transcripts of depositions conducted by defense attorneys; the defendants' motions *in limine* (to suppress); plaintiffs' responses to defendants; reports on our study prepared by defendants' experts; my affidavits responding to questions from the judge; and testimony by the defense experts and me at an *in limine*

hearing. The court's evaluations of each of the criteria are summarized in Table 1. Defendants did not challenge the first Daubert criterion, and the judge found that "the cancer incidence study methodology consists of a testable hypothesis" (TMI Litigation Cases Consolidated II 1996a).

Under the second criterion the court considered whether the methodology used to produce the results had been peer reviewed. The Columbia team's papers, and therefore the basic study design and data collection, had been peer reviewed and was recognized by the court. Part of our study was a replication of the original analyses showing that we could apply the same statistical methods to produce the same results. In other analyses we applied the same statistical method to corrected data using different controlling (or adjustment) factors corresponding to the before–after design of the study. Despite these facts, the judge, noting that "Aside from Wing's 'unadorned assertions' that the methodologies have been subject to peer review, Plaintiffs have presented no evidence of this fact," ruled, "This factor will weigh against the admission of the proffered testimony" (TMI Litigation Cases Consolidated II 1996a).

The court made two rulings on the issue of the third criterion, the known or potential rate of error (TMI Litigation Cases Consolidated II 1996a, 1996b). Because of a lack of understanding of how to interpret the standard errors of the dose–response regression coefficients provided in our report, the court initially deferred its decision on the "rate of error." However, the first ruling cited defendants' argument that the findings could have been "ascribable to chance as well as to any real association with accident emissions" on the basis of 95% confidence limits around our estimates of dose response for all cancer (in the models that omitted socioeconomic status) and leukemia (in models with and without socioeconomic status) (Wing et al. 1997c), which, defendants noted, included the null value of no effect (TMI Litigation Cases Consolidated II 1996b). The court then cited defendants' experts as claiming that lung cancer "has never been identified with radiation exposure as an isolated effect," implying that lung cancer was the only type of cancer that showed a dose effect, and also cited their claim that there was not sufficient latency for lung cancer to appear after the accident. In its second ruling on the rate of error, the court accepted the defendants' claims that minimum latency for lung cancer was known to be 10 years (TMI Litigation Cases Consolidated II 1996b). Below I discuss this juxtaposition of argument about statistical testing and the state of knowledge about cancer latency and radiosensitivity. However, in its second ruling, the court did not exclude evidence from our study on rate of error grounds.

The court next considered whether there were standards controlling the operation of our technique. Although seemingly impressed with our quality control procedures, which led us to identify biases associated with double-counting of cancer cases and an undercount in 1975, the court found fault with what the defendants cited as our failure to conduct analyses of incidence of "the types of cancers known to be radiogenic" and of cancer deaths. Our failure to conduct these analyses balanced against the quality control procedures in the court's evaluation of our standards of control. The court ruled, "This factor will not weigh against the admission of the proffered testimony" (TMI Litigation Cases Consolidated II 1996a).

Three Mile Island

Percent difference between observed and expected cases

150
100
50
0
−50

Radiation doses (0, 0–1, 1–10, 10–50, 50–100, 100–200, 200–400, 400–800, 800–1,666)

Radiation doses resulting from the 1979 nuclear accident (relative units of radiation)

Figure 3. Radiation emissions and incidence of lung cancer, 1981–1985, in the TMI 10-mile area. Figure adapted from Dalrymple (1997).

Table 1. Consideration of admissibility of evidence from the reanalyses of cancer incidence in the TMI 10-mile area, District Court for the Middle District of Pennsylvania.

Criterion for admission	Weight in consideration of admission
1. Testable hypothesis	For
2. Peer-reviewed methodology	Against
3. Known or potential rate of error	Neither for nor against
4. Standards are used to control reliability	Neither for nor against
5. Support and acceptance of methodology	Neither for nor against
6. Relies on facts or data used by other experts	Against
7. Professional qualifications	For
8. Potential for use outside the courtroom	For

On the fifth criterion, the court found that the principle of reanalysis was generally accepted, and it attributed to defendants the claim that "the statistical technique employed by Wing is generally accepted in the scientific community" (TMI Litigation Cases Consolidated II 1996a). The latter point is ironic given the court's decision that lack of peer review of the methodology weighed against admissibility. However, the court agreed with defendants that the methodology was problematic "because it produces conclusions at odds with what is generally known and accepted about cancer latency periods" (TMI Litigation Cases Consolidated II 1996a). This factor, like the standards of control, did not weigh against admission of the testimony.

Regarding our reliance on facts or data used by other experts, the court found, "In many ways Wing's cancer incidence study closely resembles a standard and reliable epidemiologic reanalysis. Yet in one important way it does not. Wing's reanalysis produces no conclusive findings" (TMI Litigation Cases Consolidated II 1996a). This lack of conclusiveness, according to the court, weighed against admission of our study. The court found that criteria seven and eight, professional qualifications and nonjudicial uses, weighed in favor of admissibility.

Deferring only the final decision on rate of error, which eventually was decided in favor of admissibility, the court weighed all the Daubert criteria to decide on admissibility of the reanalysis (TMI Litigation Cases Consolidated II 1996a). The ruling found our results on all cancer and leukemia to be admissible at trial. The lung cancer findings, including the regression analysis summarizing the dose response shown in Figure 3, were excluded based on defendants' arguments that radiation-induced lung cancer has a minimum latency of 10 years, and that the analyses were therefore irrelevant to the case.

Discussion

The TMI cancer incidence studies were conducted in the context of conflict between residents who believed they had been injured and officials who denied that such injuries were possible, as well as conflicts between scientists over the magnitude of radiation releases, the state of knowledge about radiation-induced cancer, and the meaning of evidence produced by the TMI cancer incidence studies. In the next section, I intend to show how the objectivity, rigor, and ethics of science can be increased by analyzing the influence of these conflicts on research assumptions, methods, and conclusions. I begin with a discussion of the court's reasoning on the admissibility of the reanalysis, focusing on Daubert criteria 1–3 and 5.

Testable hypotheses: collision of evidence and assumptions. The court found that our methodology did involve a testable hypothesis. Ironically, we argued that although the Columbia investigators designed a study with a testable hypothesis, they were unable to test it because of their assumptions. A testable hypothesis requires that evidence of the effect be interpretable as supporting the hypothesis. Support should be strengthened to the extent that the design and conduct of the study help rule out alternative explanations for the findings. However, the Columbia investigators clearly stated that the accident doses were known to be too low to produce the effects being hypothesized; the increased risk of cancer at the assumed maximum dose to a member of the public would have been less than one-half of 1% according to standard assumptions, clearly an excess too small to be detectable by epidemiologic methods (Wing et al. 1997c). We argued that a follow-up study of cancer mortality among adults in the TMI 5-mile area (Talbott et al. 2000, 2003) suffered from the same logical flaw: assumptions of low doses clearly precluded an interpretation of the positive dose–response relationships as supportive of the hypothesis under investigation (Wing and Richardson 2001).

If the testability of a hypothesis depends on assumptions as well as methodology, then an evaluation of the quality of a study must address the logic of key assumptions as well as the methodology. The assumption that the cancer risk of the maximally exposed person would increase by only 0.5% was supported by official reports. We addressed a testable hypothesis only because we considered the possibility that these reports could be wrong. Without that possibility, there would be no testable hypothesis. Our association with plaintiffs in the litigation introduced us to critical reevaluations of radiation monitoring, detailed case reports of symptoms (Aamodt and Aamodt 1984; Molholt 1985), biodosimetric studies of persons who reported symptoms at the time of the accident (Shevchenko 1996; Shevchenko and Snigiryova 1996), and meteorologic and environmental analyses (Field et al. 1981; Steinacker and Vergeiner 2002; Wahlen et al. 1980), as well as the court order that directed calculation of radiation doses for the Columbia study. The order prohibited "upper limit or worst case estimates of releases of radioactivity or population doses . . . [unless] such estimates would lead to a mathematical projection of less than 0.01 health effects," and further specified that "a technical analyst . . . designated by counsel for the Pools [nuclear industry insurers] concur on the nature and scope of the [dosimetry] projects" (Three Mile Island Litigation 1986). These court-imposed restrictions, which conditioned the input of the investigation (release estimates) on projections

of its outcome (health effects), constitute a manipulation of research that was possible, in part, because of years of investigative complacency brought on by entrenched assumptions that precluded even consideration of the possibility of high releases. The requirement of prior concurrence by lawyers for the industry suggests that the industry's image and liability were more important than accuracy and full disclosure.

Peer review: normal science. Under Daubert, the court considers peer review of the methodology to be a factor in the admissibility of evidence. The court's decision that lack of peer review weighed against admissibility of our study was curious and internally contradictory for reasons noted above. In principle, peer review is one of several Daubert criteria that gives preference to normative scientific views under the debatable assumption that widely held beliefs are more reliable. Although peer review may catch obvious flaws or poor writing, it cannot ensure that findings are correct, or even that research is not fraudulent (Broad and Wade 1982). In areas where scientific research, professional meetings, fellowships, and journals are funded through organizations with interests in an established perspective, peer review by orthodox scientists may lead to rejection of studies whose results challenge established assumptions, even if their methodology is appropriate (Nussbaum 1998; Nussbaum and Köhnlein 1994).

Known or potential rate of error. This criterion of admissibility is intended to recognize that scientific studies make measurements to quantify phenomena, and that the accuracy of these measurements is an important criterion. For example, upon repeated measurements of identical samples, variability in a scale, assay, or other measurement device produces a distribution of results similar to the patterns of card combinations produced by well-shuffled decks. In the case of epidemiologic studies of exposure–disease relationships, the courts have taken statistical parameters such as standard errors, confidence limits, test statistics, and p-values to be indicators of this rate of error.

Our report provided information on sample size, goodness of fit, and standard errors of regression coefficients rather than p-values or 95% confidence intervals. Although confidence limits and p-values can be easily calculated from standard errors and likelihood ratio tests, they are commonly misinterpreted as reflecting a process of randomization in which there is an a priori probability distribution of results from repeated unbiased, well-controlled experiments distributed around the true parameter (Greenland 1990). In the absence of randomization or random sampling in the TMI cancer incidence study

of a total population, I argued that statistical evidence should be evaluated

> in the context of the sensitivity of the findings to changes in assumptions and data, coherence of the evidence with other knowledge, the magnitude of associations, their consistency across groups, and temporal relationships of exposure and effect. (Wing 1995)

Although the court recognized literature recommending that the admissibility of evidence should not be determined solely on the basis of a significant p-value or a confidence interval excluding the null value of no statistical association, defendants specifically made that argument, and the court was uncertain about how to judge my less mathematical approach. The problem, according to the judge, was that, "To the extent that the results are more likely the product of random error than a true causal relationship, the probative value of the study necessarily diminishes" (TMI Litigation Cases Consolidated II 1996a). I argued that random error was not the issue because randomization had not been employed.

The court's interpretation reflects confusion between the use of chance as a tool and the idea of chance as a force of nature. Chance as a tool is familiar in card games, random sampling, and random allocation, and can be created only by complete control over the materials being manipulated: cards, sampled units, or patients assigned to treatment. Under these conditions, probabilities of particular occurrences are defined because the materials can be ordered or mixed through a process (shuffling, randomization) that has been constructed to eliminate systematic influence on the order or assignment. Following the use of chance as a tool, long-run probabilities are determined if there is no bias in the conduct of the game or research.

Statistical testing in observational settings, including most epidemiology, occurs in studies of aspects of the real world, such as dose–response relationships, in which there have been no randomization and no random sampling. Under these conditions there is no a priori probability distribution because chance has not been used as a tool. Therefore, chance introduced through randomization is not a possible explanation of a result. What, then, do researchers (or courts) mean when they conclude from statistical tests that results are likely due to chance? They should not mean that some unknown factor is the cause, because other causes that create (or mask) the appearance of an association between exposures and outcomes are referred to as confounders (whether or not they have been measured). Rather, the concept of chance in observational research has been confused with its original interpretation in randomized studies: it is treated as a force of nature that functions as an alternative explanation to specific causes.

In its second ruling on the rate of error issue, the court quoted from the affidavit I prepared on this issue in which I attempted to explain why the strict application of statistical tests being advocated by defendants was increasingly recognized as inappropriate:

> Abuse of significance testing in epidemiological research is now widely appreciated and discussed. The problem is clearly recognized by one of the defense's experts, K. Rothman, who wrote in an editorial introducing a paper on logical problems with statistical significance testing, "In a century in which science has revealed that molecules and atoms are mostly unoccupied space and that matter is energy anyway, we should be accustomed to having the substance of our scientific foundations dissolve into emptiness. In this issue [of the journal], Greenland perforates the foundations of statistical interpretations used by epidemiologists in just this way" (Rothman 1990). The problems with significance testing noted in the paper by Greenland (Greenland 1990) have been recognized for some time, but have been ignored in the interests of preserving a simple if fallacious method that was believed to result in the separation of conclusive, causal associations from those that are inconclusive or spurious. That method is the determination of the statistical significance of a finding. As noted by Rothman, "conventional statistics [p-values and confidence limits] have a strict interpretability in experiments with random assignment of exposures. The results of those experiments are nearly always dressed in the same statistical garb that was developed for and is applied to experimental studies. Greenland shows us that, despite our reliance on conventional statistics for interpreting our nonexperimental results, *there is no basis for the interpretations usually given to these statistics in nonexperimental settings*" [emphasis added]. This is because, in the absence of randomization of exposure in a fair experiment, it is not possible to distinguish the extent to which test statistics reflect bias or the exposure under investigation. (TMI Litigation Cases Consolidated II 1996b)

That a normative scientific practice used to distinguish causal from noncausal relationships—central not only to science but to its social and legal applications—could "dissolve into emptiness" should be disconcerting to those who count on science for objective and rational knowledge. Critical analysis of normal scientific practice can help to identify such emptiness before serious mistakes are made. In the TMI reanalysis we chose a contextual set of criteria to evaluate quantitative evidence because we did not believe that results could be a product of random error unless randomization had been introduced by design. If our primary quest had been small p-values, we would have used one-tailed tests (because our hypothesis was one-directional), emphasized results adjusted for socioeconomic status, and analyzed log-transformed doses, which, as noted by Mangano (Mangano 1997), would have better fit the data and produced smaller p-values.

Invocation of chance as a force of nature—a cause incapable of being further

analyzed—can discourage scientists from humility about the scientific enterprise as well as deeper mechanistic analysis (Gigerenzer et al. 1989). In the case of environmental health research, this may discourage the testing of hypotheses or use of methods with the greatest potential to implicate institutions that permit or produce pollution. An understanding of how chance as a tool has been conflated with chance as an explanatory force of nature could help to improve the practice and applications of science. However, even without a detailed understanding of that process, it is clear that current practice, although normative science, is inconsistent and illogical. Chance may have created an empire in the world of science (Hacking 1990), but its emperor has no clothes.

Basis of assumptions about radiogenic cancers and cancer latency. The court's ruling that our lung cancer findings were inadmissible at trial was based on defendants' claims that *a*) only lung cancer was statistically significantly related to accident dose estimates, *b*) lung cancer has never been found as the sole effect of radiation, and *c*) radiogenic lung cancer is known to have a minimum latency of 10 years. These arguments, repeated in the court's rulings under several Daubert criteria, especially the fifth, imply that dose–response relationships for leukemia and all cancer resulted from random error, and that the results for lung cancer must be due to some other error because they could not occur as a result of the hypothesized cause. Although the leukemia dose–response coefficients, based on 75 cases, were less precise than the lung cancer estimates, they were roughly 40% larger in magnitude, which would be consistent with studies showing steeper relationships for leukemia than for solid tumors after high-dose radiation. Dose–response coefficients for all cancers were more precise but smaller in magnitude, which would be consistent with lower doses to organs other than the lung (Wing et al. 1997a, 1997c).

Our interest in lung cancer was based on the presence of radioactive gases, primarily xenon-133 and krypton-85, in the accident plumes. In addition to penetrating gamma radiation, these gases emit beta radiation, which has low penetration and therefore would have delivered direct doses selectively to exposed skin and respiratory tissues, which would be consistent with reports of erythema and putative impacts on plants as well as elevations in lung cancer. However, even if lung doses were substantial, no cancer effect would be seen in the incidence study if radiogenic lung cancer has a minimum latency of 10 years.

There have been no epidemiologic studies of the exposure of human populations to radioactive xenon and krypton gases.

Consequently, assumptions about the types and timing of cancers that could result must be based on inference from studies of other types of ionizing radiation. Defendants argued that the study of Japanese A-bomb survivors, upon which official estimates of radiation risks and latency have been based, proved that lung cancer has a 10-year minimum latency. In response, we noted that ionizing radiation can act as a promoter as well as an initiator of cancer (Doll 1978); that high doses can suppress immune function, which is associated with the appearance of secondary tumors within 2 years of radiotherapy (Appelbaum 1993); and that latencies of less than 5 years have been observed for miners exposed to radon (Hornung and Meinhardt 1987). A recent study of lung cancer among uranium miners found the best estimate of minimum latency to be less than 1 year (Langholz et al. 1999).

Epidemiologists often remain skeptical of estimates of dose–response relationships because of questions about possible confounding and measurement error. In the area of radiation health effects, however, the A-bomb survivor studies are widely used for risk estimation and have long functioned as a "gold standard" for judging other epidemiologic evidence (BEIR V 1990). The A-bomb studies have this status despite being based on a select group of survivors that had to resist radiation, blast, and the aftermath of war to enter the study, and whose radiation doses were not measured but calculated based on *a*) estimates of radiation releases that have been repeatedly revised, and *b*) interviews whose accuracy depended on the survivors' memories and their trust of researchers connected with the U.S. military occupation forces (Lindee 1994; Stewart 2000; Wing et al. 1999). The status of the A-bomb studies as a gold standard has shaped the normative scientific culture, including peer review, research funding, and dismissal of conflicting evidence from other populations (Nussbaum and Köhnlein 1994; Wing et al. 1999). Critical analysis of the evolution of radiation epidemiology within the context of the military, medical, and industrial uses of ionizing radiation can help scientists reevaluate the A-bomb studies more objectively. Such an approach would help to increase the explanatory capacity of radiation health science and improve its applications in the courts, compensation programs, and public education (Wing and Richardson 2002).

Radiation versus stress as causal explanations. From their inception, epidemiologic studies of the TMI accident focused on stress. Although this was not specifically an issue in the court, it is central to the epidemiology and public relations of the accident. Citizens were told that symptoms similar to those which are caused by radiation exposure were due to stress. Studies were designed to quantify health effects of stress in general, and specifically as an alternative explanation to radiation-induced cancer increases following the accident. An editorial commenting on the paper by Hatch et al. (1991) on stress and cancer suggested that the stress–cancer link might be used as grounds for not disclosing accidents to the public because the resulting stress would injure people (Janerich 1991). When radiation exposures were studied, discussion of findings focused on reasons why radiation effects may have been overestimated or spurious, ignoring plausible reasons why they may well have been underestimated (Hatch et al. 1990; Talbott et al. 2000; Wing and Richardson 2001). The Columbia investigators chose not to discuss dose misclassification and migration, for example, as reasons to expect underestimation of radiation effects (Hatch et al. 1990). In fact, they planned to consider the possibility of confounding bias in estimates of dose response only in the event that they found a positive radiation–cancer relationship (Susser 1997), despite the fact that such bias could also mask a true effect. The lack of attention to these standard interpretive issues can be understood in terms of the key role that assumptions play in evaluating the meaning of results (Wing et al. 1997b).

Despite differences in results between the Columbia studies and ours, both found evidence of impacts of the accident on cancer incidence. However, the evidence led us to different conclusions regarding both cause and biological mechanism. The Columbia group concluded that the evidence suggested stress as a cause, and stress-induced immune system depression as a mechanism. We concluded that the evidence suggested radiation as a cause, and promotion of cancer through late-stage "hits" in a multistage process of carcinogenesis, as well as radiation-induced immune system depression, as mechanisms. Keller described an analogous situation in which experimental observations on genetic mutations were redescribed and reinterpreted to produce different conclusions about causes and mechanisms (Keller 1992).

Conclusion: The Ethics of Strong Objectivity

Conflicts over responsibility for damage to health and the environment are increasingly common. They often involve disputes between actors, such as industries and governments, with the ability to make large impacts as well as sponsor research on those impacts, and communities that are most directly affected but that have little political power or capacity to conduct research to document their exposures or health conditions (Wing 2002). Affected communities may experience these situations as examples of environmental injustice. Many rural people living near TMI had modest levels of formal schooling and little experience in being assertive with government and industry officials. Those that spoke out about their experiences of physical problems from the accident endured ridicule. The Aamodts were able to influence the TMI Public Health Fund's sponsored research on physical impacts of the accident by initiating their own survey, researching government records, and petitioning the NRC. Other residents who lived within the 10-mile area also conducted surveys, constructed disease maps, and documented damage to plants and animals (Osborn 1996; Three Mile Island Alert 1999). However, when health studies were undertaken through official channels, citizens who believed they had been affected by accident emissions and their supporters were not included in the framing of questions, study design, analysis, interpretation, or communication of results. The studies themselves were funded by the nuclear industry and conducted under court-ordered constraints, and a priori assumptions precluded interpretation of observations as support for the hypothesis under investigation.

The naïve approach to objectivity, represented in the Daubert criteria, contends that scientists can produce unbiased evidence by standing apart from legal conflicts and adhering to normative science. The problem with this position is that scientific questions and the details of specific working hypotheses emerge from conflicts, which also influence the assumptions that frame methodologies used to produce evidence and interpretations of the meaning of evidence. This process occurs at various scales, from decisions about how much to trust conflicting assertions regarding a specific event like the TMI accident, to the role of the A-bomb studies as a gold standard for evaluating evidence, to widespread conventions such as the confusion between chance as a tool and chance as a force of nature. Although science has strong rationalist traditions, it has also been shaped by perspectives of dominant gender, race, and class groups, excluding perspectives of groups with less power (Harding 1991; Holtzman 1981; Hubbard 1990; Levins 1979; Levins and Lewontin 1985). Pretending that there are no assumptions embedded in scientific methodology conceals and reinforces existing inequalities.

Strong objectivity demands that scientists critically evaluate how the knowledge they create is shaped at every point by historical social forces. Strong objectivity is therefore not a static feature of scientific knowledge that, once attained, becomes a property of that knowledge. It is an evolving process that is never finished, like scientific inquiry itself. Scientists should be trained to engage in careful reflection

about how the history of their discipline has affected their hypotheses, assumptions, and tools, and how their work, like the work of others before them, is shaped by contemporary forces (Armstrong 1999). This is essential as careful measurement and analysis for producing an objective science that will be maximally rigorous, rational, reliable in courts of law, and useful for improving the world. Strong objectivity is needed, not only for good science, but for ethical conduct of research.

Postscript

In December 1999, after summary dismissal of the TMI case by the District Court for the Middle District of Pennsylvania, the U.S. Third Circuit Court of Appeals found that the district court had erred in excluding our lung cancer findings (although the appeals court also ruled that the lung cancer findings would not change the outcome of the case). In December 2002, the circuit court declined to hear an appeal of Judge Rambo's second ruling granting summary dismissal. Attorneys representing 1,990 remaining plaintiffs in the TMI case declared they would take no further legal action (Associated Press 2002).

REFERENCES

Aamodt M, Aamodt N. 1984. Petitioners v. U.S. Nuclear Regulatory Commission. Aamodt motions for investigation of licensee's reports of radioactive releases during the initial days of the TMI-2 accident and postponement of restart decision pending resolution of this investigation. Docket Number 50-289. Administrative Court, Washington, DC, 21 June.

Advisory Committee on Human Radiation Experiments. 1995. Final Report, Advisory Committee on Human Radiation Experiments. Washington, DC:U.S. Government Printing Office.

Appelbaum F. 1993. The influence of total dose, fractionation, dose rate, and distribution of total body irradiation on bone marrow transplantation. Semin Oncol 20:3–10.

Armstrong D. 1999. "Controversies in epidemiology", teaching causality in context at the University at Albany, School of Public Health. Scand J Public Health 2:81–84.

Associated Press 2002. Plaintiffs end action in TMI lawsuit. Available: http://www.philly.com/mld/inquirer/news/local/4826524.htm [accessed 16 May 2003].

Baum A. 1990. Stress, intrusive imagery, and chronic distress. Health Psychol 9:653–675.

Baum A, Cohen L, Hall M. 1993. Control and intrusive memories as possible determinants of chronic stress. Psychosom Med 55:274–286.

Baum A, Gatchel RJ, Schaeffer MA. 1983. Emotional, behavioral, and physiological effects of chronic stress at Three Mile Island. J Consult Clin Psychol 51:565–572.

BEIR V. 1990. National Research Council, Committee on the Biological Effects of Ionizing Radiation (BEIR V), Health Effects of Exposure to Low Levels of Ionizing Radiation. Washington, DC:National Academy Press.

Beyea J. 1985. Three Mile Island—six years later. J Nucl Med 26:1345–1346.

Beyea J, Hatch M. 1999. Geographic exposure modeling: a valuable extension of geographic information systems for use in environmental epidemiology. Environ Health Perspect 107:181–190.

Broad W, Wade N. 1982. Betrayers of the Truth: Fraud and Deceit in the Halls of Science. New York:Simon and Schuster.

Brodsky CM. 1988. The psychiatric epidemic in the American workplace. Occup Med 3:653–662.

Cleary PD, Houts PS. 1984. The psychological impact of the Three Mile Island incident. J Hum Stress 10:28–34.

Cornely P, Bromet E. 1986. Prevalence of behavior problems in three-year-old children living near Three Mile Island: a comparative analysis. J Child Psychol Psychiatry 27:489–498.

Dalrymple M. 1997. Science on the firing line. Endeavors:The University of North Carolina at Chapel Hill 14(Fall):12–13.

Daubert v. Merrell Dow Pharmaceuticals, Inc. 1993. Case No. 509 U.S. 579, U.S. Supreme Court, Washington, DC.

Davidson LM, Fleming R, Baum A. 1987. Chronic stress, catecholamines, and sleep disturbance at Three Mile Island. J Hum Stress 13:75–83.

Del Tredici R. 1980. The People of Three Mile Island. San Francisco, CA:Sierra Club.

Dew MA, Bromet EJ. 1993. Predictors of temporal patterns of psychiatric distress during 10 years following the nuclear accident at Three Mile Island. Soc Psychiatry Psychiatr Epidemiol 28:49–55.

Dew MA, Bromet EJ, Schulberg HC. 1987a. A comparative analysis of two community stressors' long-term mental health effects. Am J Community Psychol 15:167–184.

Dew MA, Bromet EJ, Schulberg HC, Dunn LO, Parkinson DK. 1987b. Mental health effects of the Three Mile Island nuclear reactor restart. Am J Psychiatry 144:1074–1077.

Doll R. 1978. An epidemiological perspective of the biology of cancer. Cancer Res 38:3573–3583.

Donnell HD, Bagby JR, Harmon RG, Crellin JR, Chaski HC, Bright MF, et al. 1989. Report of an illness outbreak at the Harry S. Truman state office building. Am J Epidemiol 129:550–558.

Fabrikant JI. 1983. The effects of the accident at Three Mile Island on the mental health and behavioral responses of the general population and nuclear workers. Health Phys 45:579–586.

Faust HS, Brilliant LB. 1981. Is the diagnosis of "mass hysteria" an excuse for incomplete investigation of low-level environmental contamination? J Occup Med 23:22–26.

Field R, Field E, Zegers D, Steucek G. 1981. Iodine-131 in thyroids of the meadow vole (Microtus Pennsylvanicus) in the vicinity of the Three Mile Island nuclear generating plant. Health Phys 41:297–301.

Gatchel RJ, Schaeffer MA, Baum A. 1985. A psychophysiological field study of stress at Three Mile Island. Psychophysiology 22:175–181.

Gigerenzer G, Swijtink Z, Porter T, Daston L, Beatty J, Kruger L. 1989. The Empire of Chance: How Probability Changed Science and Everyday Life. New York:Cambridge University Press.

Goldhaber MK, Houts PS, DiSabella R. 1983a. Moving after the crisis: a prospective study of Three Mile Island area population mobility. Environ Behav 15:93–120.

Goldhaber MK, Staub SL, Tokuhata GK. 1983b. Spontaneous abortions after the Three Mile Island nuclear accident: a life table analysis. Am J Public Health 73:752–759.

Goldhaber MK, Tokuhata GK, Digon E, Caldwell GG, Stein GF, Lutz G, et al. 1983c. The Three Mile Island Population Registry. Public Health Rep 98:603–609.

Gray M, Rosen I. 1982. The Warning: Accident at Three Mile Island. New York:W.W. Norton.

Greenberg M. 1991. The evolution of attitudes to the human hazards of ionizing radiation and to its investigators. Am J Ind Med 20:717–721.

Greenland S. 1990. Randomization, statistics, and causal inference. Epidemiology 1:421–429.

Gur D, Good WF, Tokuhata GK, Goldhaber MK, Rosen JC, Rao GR, et al. 1983. Radiation dose assignment to individuals residing near the Three Mile Island nuclear station. Proc PA Acad Sci 57:99–102.

Hacking I. 1990. The Taming of Chance. New York:Cambridge University Press.

Harding S. 1991. Whose Science? Whose Knowledge? Ithaca, NY:Cornell University Press.

Hatch MC, Beyea J, Nieves JW, Susser M. 1990. Cancer near the Three Mile Island nuclear plant: radiation emissions. Am J Epidemiol. 132:397–412; discussion 413–417.

Hatch MC, Wallenstein S, Beyea J, Nieves JW, Susser M. 1991. Cancer rates after the Three Mile Island nuclear accident and proximity of residence to the plant. Am J Public Health 81:719–724.

Hefez A. 1985. The role of the press and the medical community in the epidemic of "mysterious gas poisoning" in the Jordan West Bank. Am J Psychiatry 142:833–837.

Holtzman E. 1981. Science, philosophy and society: Some recent books. Int J Health Serv 11:123–148.

Hornung RW, Meinhardt TJ. 1987. Quantitative risk assessment of lung cancer in U.S. uranium miners. Health Phys 52:417–430.

Houts PS, Goldhaber MK. 1981. Psychological and social effects on the population surrounding Three Mile Island after the nuclear accident on March 28, 1979. In: Energy, Environment and the Economy (Majumdar S, ed). Philadelphia:Pennsylvania Academy of Sciences, 151–164.

Houts PS, Tokuhata GK, Bratz J, Bartholomew MJ, Sheffer KW. 1991. Effect of pregnancy during TMI crisis on mothers' mental health and their child's development. Am J Public Health 81:384–386.

Hubbard R. 1990. The Politics of Women's Biology. New Brunswick, NJ:Rutgers University Press.

Janerich DT. 1991. Can stress cause cancer? [Editorial]. Am J Public Health 81:687–688.

Keller E. 1992. Between language and science: the question of directed mutation in molecular genetics. Perspect Biol Med 35:292–306.

———. 1995. Science and its critics. Academe 81:10–15.

Kuhn T. 1970. The Structure of Scientific Revolutions. Chicago:University of Chicago Press.

Langholz B, Thomas D, Xiang A, Stram D. 1999. Latency analysis in epidemiologic studies of occupational exposures: application to the Colorado Plateau uranium miners cohort. Am J Ind Med 35:246–256.

Levins R 1979. Class Science and Scientific Truth. The New York Marxist School Available: http://csf.colorado.edu/gimenez/papers/Class-Science.html [accessed 20 May 2003].

Levins R, Lewontin RC. 1985. The Dialectical Biologist. Cambridge, MA:Harvard University Press.

Lindee MS. 1994. Suffering made real: American science and the survivors at Hiroshima. Chicago:University of Chicago Press.

Lyon JL. 1999. Nuclear weapons testing and research efforts to evaluate health effects on exposed populations in the United States. Epidemiology 10:557–560.

Macleod GK. 1981. Some public health lessons from Three Mile Island: a case study in chaos. Ambio 10:18–23.

Makhijani A, Hu H, Yih K. 1995. Nuclear Wastelands: A Global Guide to Nuclear Weapons Production and Its Health and Environmental Effects. Cambridge, MA:MIT Press.

Mangano JJ. 1997. Low-level radiation harmed humans near Three Mile Island. Environ Health Perspect 105:786–787.

McCally M, Cassel C, Kimball DG. 1994. U.S. government-sponsored radiation research on humans, 1945–1975. Med Glob Surviv 1:4–17.

McKinnon W, Weisse CS, Reynolds CP, Bowles CA, Baum A. 1989. Chronic stress, leukocyte subpopulations, and humoral response to latent viruses. Health Psychol 8:389–402.

Merwin S, Moeller D, Kennedy W, Moeller M. 2001. Application of the Supreme Court's Daubert criteria in radiation litigation. Health Phys 81:670–677.

Molholt B. 1985. Summary of acute symptoms by TMI area residents during accident. In: Proceedings of the Workshop on Three Mile Island Dosimetry, Vol 2. Academy of Natural Sciences, 12–13 November 1984 Philadelphia, PA:TMI Public Health Fund, A109–111.

Morgan KZ. 1992. Health physics: its development, successes, failures, and eccentricities. Am J Ind Med 22:125–133.

NRC. 1979a. Population Dose and Health Impact of the Accident at the Three Mile Island Nuclear Station. Washington, DC:U.S. Nuclear Regulatory Commission.

———. 1979b. Investigation into the March 28,1979 Three Mile Island Accident by Office of Inspection and Enforcement. Washington, DC:U.S. Nuclear Regulatory Commission.

Nussbaum RH. 1998. The linear no-threshold dose-effect relation: Is it relevant to radiation protection? Med Phys 25:291–299.

Nussbaum R, Köhnlein W. 1994. Inconsistencies and open questions regarding low-dose health effects of ionizing radiation. Environ Health Perspect 102:656–667.

Office of Technology Assessment. 1991. Complex Cleanup: The Environmental Legacy of Nuclear Weapons Production. Washington, DC:U.S. Congress, Office of Technology Assessment.

Osborn, MS. 1996. Three Mile Island: then, now and next time. Synthesis/Regeneration 10(Spring):11–14.

President's Commission on the Accident at Three Mile Island. 1979. The Need for Change: The Legacy of TMI. Washington, DC:U.S. Government Printing Office.

Prince-Embury S, Rooney JF. 1988. Psychological symptoms of residents in the aftermath of the Three Mile Island nuclear accident and restart. J Soc Psychol 128:779–790.

Rorty R. 1989. Contingency, Irony, and Solidarity. New York:Cambridge University Press.

Rothman KJ. 1990. Statistics in nonrandomized studies. Epidemiology 1:417–418.

Schaeffer MA, Baum A. 1984. Adrenal cortical response to stress at Three Mile Island. Psychosom Med 46:227–237.

Shevchenko V. 1996. Assessment of genetic risk from exposure of human populations to radiation. In: Consequences of the Chernobyl Catastrophe: Human Health (Burlakova E, ed). Moscow:Center for Russian Environmental Policy and Scientific Council on Radiobiology Russian Academy of Sciences, 46–61.

Shevchenko V, Snigiryova G. 1996. Cytogenetic Effects of the Action of Ionizing Radiations on Human Populations. In: Consequences of the Chernobyl Catastrophe: Human Health (Burlakova E, ed). Moscow:Center for Russian Environmental Policy and Scientific Council on Radiobiology, Russian Academy of Sciences, 23–45.

Simon GE, Katon WJ, Sparks PJ. 1990. Allergic to life: psychological factors in environmental illness. Am J Psychiatry 147:901–906.

Small G, Borus J. 1987. The influence of newspaper reports on outbreaks of mass hysteria. Psychiatr Q 58:269–278.

Steinacker R, Vergeiner I. 2002. The meterological setting of the 'TMI-2' nuclear accident on 28 March 1979. Weather 57:341–351.

Sterling TD. 1980. The health effects of low-dose radiation on atomic workers. A case study of employer-directed research. Int J Health Serv 10:37–46.

Stewart A. 2000. The role of epidemiology in the detection of harmful effects of radiation. Environ Health Perspect 108:93–96.

Stewart AM, Kneale GW. 1991. An overview of the Hanford controversy. Occup Med 6:641–663.

Susser M. 1997. Consequences of the 1979 Three Mile Island accident continued: further comment. Environ Health Perspect 105:566–570.

Talbott EO, Youk AO, McHugh KP, Shire JD, Zhang A, Murphy BP, et al. 2000. Mortality among the residents of the Three Mile Island accident area: 1979–1992. Environ Health Perspect 108:545–552.

Talbott E, Youk A, McHugh-Pemu K, Zborowski J. 2003. Long-term follow-up of the residents of the Three Mile Island accident area: 1979–1998. Environ Health Perspect 111:341–348.

Thomas JP. 2001. Rebuilding trust in established institutions. Health Phys 80:379–383.

Three Mile Island Alert 1999. 20th Anniversary of the TMI Accident, 3/28/79–3/28/99, Press Packet. Available: http://www.tmia.com/PressP2.html#health [accessed 20 May 2003].

Three Mile Island Litigation. 1986. Civil Action 79-0432, U.S. District Court for the Middle District of Pennsylvania, Harrisburg, PA, 15 December.

TMI Litigation Cases Consolidated II. 1996a. Civil Action No. 1:CV-88-1452, U.S. District Court for the Middle District of Pennsylvania, Harrisburg, PA, 5 January.

———. 1996b. Civil Action No. 1:CV-88-1452, U.S. District Court for the Middle District of Pennsylvania, Harrisburg, PA, 2 April.

Wahlen M, Kunz CO, Matuszek JM, Mahoney WE, Thompson RC. 1980. Radioactive plume from the Three Mile Island accident: xenon-133 in air at a distance of 375 kilometers. Science 207:639–640.

Wilkinson G. 1999. Seven years in search of alpha: the best of times, the worst of times. Epidemiology 10:340–344.

Wing S. 1995. Affidavit in TMI Litigation Cases Consolidated II, Civil Action No. 1:CV-88-1452, U.S. District Court for the Middle District of Pennsylvania, Harrisburg, PA, 19 October.

———. 2002. Social responsibility and research ethics in community-driven studies of industrialized hog production. Environ Health Perspect 110:437–444.

Wing S, Richardson D. 2000. Collision of evidence and assumptions: TMI deja view [Letter]. Environ Health Perspect 109:A546–A547.

———. 2002. Use of A-bomb survivor studies as a basis for nuclear worker compensation [Letter]. Environ Health Perspect 110:A739.

Wing S, Richardson D, Armstrong D. 1997a. Reply to comments on "A reevaluation of cancer incidence near the Three Mile Island." Environ Health Perspect 105:266–268.

———. 1997b. Response: Science, public health and objectivity: research into the accident at Three Mile Island. Environ Health Perspect 105:567–570.

Wing S, Richardson D, Armstrong D, Crawford-Brown D. 1997c. A reevaluation of cancer incidence near the Three Mile Island nuclear plant: the collision of evidence and assumptions. Environ Health Perspect 105:52–57.

Wing S, Richardson DB, Stewart A. 1999. The relevance of occupational epidemiology to radiation protection standards. New Solutions 9:133–151.

Wing S, Shy CM, Wood JL, Wolf S, Cragle DL, Frome EL. 1991. Mortality among workers at Oak Ridge National Laboratory. Evidence of radiation effects in follow-up through 1984. JAMA 265:1397–1402.

Wing S, West CM, Wood JL, Tankersley W. 1994. Recording of external radiation exposures at Oak Ridge National Laboratory: implications for epidemiological studies. J Expo Anal Environ Epidemiol 4:83–93.