LAWPRJ	mgm037	RR
Journal Name	Art. No.	CE Code

NOT FOR PUBLIC RELEASE

Law, Probability and Risk Page 1 of 20 doi:10.1093/lpr/mgm037

The role of epidemiology in the law: a toxic tort litigation case

FRANCESCA DOMINICI†

Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA

AND

SHIRA KRAMER AND APRIL ZAMBELLI-WEINER

Epidemiology International, Owings Mills, MD 21117, USA

[Received on 20 April 2007; revised on 20 August 2007; accepted on 23 August 2007]

Toxic tort cases provide a natural framework for an in-depth illustration and an application of statistical methods for small-scale studies of putative sources of hazard. In this paper, we describe the aspects of a toxic tort case that focussed on quantifying the strength of evidence concerning the hypothesis that carcinogenic substances emitted from an industry source were associated with a statistically significant higher than expected incidence rate of neuroblastoma in children. We first present the epidemiological analysis carried out by the plaintiffs' experts (Drs P1, P2 and P3). We then summarize the key critiques by the defense experts (Drs D1, D2 and D3) followed by the plaintiff's reply. In the context of toxic torts, the plaintiff must demonstrate that the exposure resulting from the defendants' conduct is more likely than not causally related to the injury. We use this toxic tort case to identify common criticisms of the defense experts that take advantage of the complexity in evaluating causation in toxic torts. In the discussion, we summarize the common defense positions and question whether such questions are scientifically appropriate.

1. Introduction

1.1 *Plant operations*

Beginning in 1954 and continuing through 2001, an industry source (hereinafter referred to as 'the industry source') located in a small community near Los Angeles, CA (hereinafter referred to as 'the exposed community') disposed of explosive contaminated waste through open detonation (OD), subsurface detonation, burn barrel, open burning (OB), popper furnace activities and drop-box activities. Unlined ponds for disposal of chemical agents were utilized from the mid-1960s until 1968 when the ponds were lined. OB of wastes began in burn pits around 1971 and OD began around 1979. Popping ovens and incinerators were employed beginning in the early 1970s but were required to close in 1989 in accordance with the Resource Conservation and Recovery Act. At this time, OB/OD activities continued as the main waste disposal method until the major areas were required to close in 1992; however, an additional burn area was opened in 1992 and continued through 2001. OB/OD remediation began in 1994 and off-site shipping of the OB/OD soil began in 1997, which continued

† Email: fdominic@jhsph.edu

until mid-2002. Experts who evaluated the industry source's waste disposal practices and procedures showed that they resulted in the release of hazardous and carcinogenic substances into the environment, which resulted in multiple-exposure pathways for the population of this community. Permit violations were specifically identified from 1989 forward, due largely to the closing of the popper ovens and incinerators.

1.2 *Cases of neuroblastoma*

Neuroblastoma (NB) is a rare embryonal tumour that affects primarily infants and children. The average annual incidence of NB in the United States is approximately nine cases per 1 000 000 people (all races and genders, age < 15) (Goodman, 1999). Rates are slightly higher in males versus females and in Caucasians versus persons of African descent. NB is virtually non-existent in adults, with recent estimates placing the average annual U.S. incidence in adults around 0.12 cases per 1 000 000 people (Esiashvili *et al*., 2007). NB is characterized as an embryonal cancer which can occur very early in life and has no minimum latency period that can be supported (Brodeur and Castleberry, 1997; Goodman *et al*., 1999; Brodeur, 1991; Kramer *et al*., 1987; Schwartzbaum, 1992).

Four Caucasian cases of NB in children younger than 15 were identified in the exposed community during the period 1995–1999, and a fifth Asian case was identified in the year 2000 (see Table 1). Four cases in a 5-year period (1995–1999) are substantially greater than expected. Although the industry source had been operating since the early 1950s, two notable factors changed during the 1990s: (a) significant growth of the population in the community, with residential housing developments established within close proximity to the site and (b) the composition of the waste burned prior to July 1990 was markedly different from the waste burned after that date. Environmental modelling experts involved in the litigation provided evidence and expert testimony that the air emission plumes emanating from the industry source post-1990 posed a much higher risk to human health in the community than prior to that time period.

Members of the community expressed serious concerns about exposure to the carcinogenic emissions emanating from the industry source. The litigation against the industry source was initiated in 2001 by the families in the community whose children were diagnosed with NB. These diagnoses occurred subsequent to high-level emissions of a number of carcinogenic chemicals from the industry source which could have impacted parental germ cells, the developing foetus and the child.

The plaintiffs' experts (P1 and P2) were asked to conduct an analysis to evaluate whether the children in the community were experiencing a higher rate of NB, higher than expected, based on national and regional reference rates. The statistical analysis conducted by the plaintiff was challenged by the defense experts (D1, D2 and D3) who presented alternative approaches for the analyses. The

TABLE 1 *Birth, death and diagnosis dates for five of the exposed community NB cases*

Case no.	Year of diagnosis	Age at diagnosis	Year of death
1	1995	4.5 years	1998
2	1995	4 months	1997
3	1996	13 years	N/A
4	1998	4.5 years	2000
5	2000	2 years	N/A

plaintiff then contacted Dr Dominici (P3), a biostatistician, to have an independent review of their work and also an independent evaluation of the critiques written by defense experts.

1.3 *Exposures*

Routes of exposure to the carcinogenic emissions emanating from the industry source included inhalation of vapours and particulates, dermal contact, soil ingestion and garden ingestion. The carcinogens that were allegedly emitted from the site into the environment included (a) *N*-nitroso compounds (NOCs) which were produced by low-temperature decomposition of explosives and released in the carcinogenic emissions emanating from the industry source; (b) volatile organic compounds (VOCs), including benzene, toluene, dichlorobenzenes, trichlorobenzenes, ethylbenzene, chloromethane and dichloromethane (methylene chloride) and 1,3-butadiene; (c) naphthalene; (d) polycyclic organic matter compounds and polycyclic aromatic hydrocarbons; (e) dioxins; (f) metals, including lead and chromium, and (g) vinyl chloride. Prior research has shown that chemicals emitted from the site, or agents in the same class of chemicals, are associated with an increased risk of adult and childhood nervous systems cancers, including NB, in humans and animals, and are multi-potential carcinogens in humans and animals. Exposure to the same NOCs emitted from the industry source occurs in industries that have shown an association with an increased risk of NB in offspring (Olshan *et al*., 1999; Fajen *et al*., 1982; Bunin *et al*., 1990), including hairdressers, landscapers or groundskeepers, in farming or in florist/garden shops. Paternal exposure to rubber dust containing NOCs has also been associated with elevated risk of NB in offspring (Bunin *et al*., 1990). Epidemiological studies have also observed an association between hair dye use in mothers and increased risk of NB in offspring (Kramer *et al*., 1987; Olshan *et al*., 1999), as well as an association between pesticide exposure and increased risk of NB (Kerr *et al*., 2000; Daniels *et al*., 2001; Michaelis *et al*., 1996), both of which involve exposure to NOCs. Animal studies have also shown that subcutaneous injection exposure of NOCs, such as *N*-nitrosopiperidine, can result in tumours, including NBs (U.S. Department of Health and Human Services, 2002). With regard to VOC emissions from the industry source, several studies have reported that parental exposures to benzene and 1,3-butadiene polymer are associated with NB risk in offspring (McKinney *et al*., 2003; De Roos *et al*., 2001a; Spitz and Johnson, 1985). Animal studies have also shown that subchronic inhalation exposure to 1,3-butadiene causes NB (National Toxicology Program, 1993).

Naphthalene was also a component of the carcinogenic emissions emanating from the community facility. It is produced from coal tar and petroleum, and parental occupational exposure to coal soot, coal tar and creosote has been associated with increased risk of NB in offspring (Kerr *et al*., 2000). Animal studies have shown that chronic inhalation exposure to naphthalene resulted in NBs of the olfactory epithelium in male and female rats (National Toxicology Program, 2000). Exposure to hydrocarbons in general is common in occupations for which associations with NB have been identified, including electrical work, farming, mechanics and painting (Olshan *et al*., 1999; Michaelis *et al*., 1996; Bunin *et al*., 1990; Spitz and Johnson, 1985; Wilkins and Hundley, 1990; Smulevich *et al*., 1999). In addition, parental occupational exposure to both non-volatile hydrocarbons, such as petroleum, and volatile hydrocarbons, such as lacquer thinner and turpentine, has been shown to be associated with increased risk of NB in offspring (Kerr *et al*., 2000; De Roos *et al*., 2001a,b; Spitz and Johnson, 1985). An association between parental exposure to dioxin and increased risk of NB in offspring has also been reported (Kerr *et al*., 2000). Inhalation exposure of animals to vinyl chloride resulted in NB of the brain as well as tumours at many other organ sites.

2. Plaintiff's statistical analysis

During the discovery phase of the litigation, plaintiff's experts were asked to conduct a standardized incidence rate (SIR) analysis of the exposed community. The main objective of the statistical analysis was to determine whether the exposed community experienced a rate of childhood NB that is statistically significant higher than the expected rate of childhood NB for the same community absent exposure.

To accomplish this goal, the plaintiffs' experts performed a standardized incidence ratio (SIR) analysis (Rothman and Boice, 1982; Breslow and Day, 1987; Sahai and Khurshid, 1996; Gordis, 2004). The main analysis focussed on the four cases of paediatric NB (age at diagnosis $\leqslant 14$) of Caucasian ethnicity for the 5-year period 1995–1999. The SIR is defined as the number of cases observed in the study population divided by the number of cases that would be expected to occur in this population if it experienced the same disease rates as the comparison population. Ninety-five percent confidence limits were computed for each SIR.

Specifically, let y_j be the number of paediatric NB cases for the subpopulation j (younger than 14 years old and Caucasian), for the period 1995–1999 in the exposed community. For a rare disease, we can assume that y_j follow the Poisson distribution with average rate equal to SIR $\times$ E_j, where E_j is the expected number of cases in the control (or unexposed) population. From this assumption, it follows that the SIR can be estimated by taking y_j divided by E_j. Note that the Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate, and are independent of the time since the last event. The expected number of cases in the control population can be defined as $E_j = N_j \times q_j$, where N_j is the size of the exposed population in the strata j (younger than 14 years old and Caucasian) and q_j is the probability of disease in the control population in the same strata j.

Below is a description of how the data were gathered to perform the incidence analysis.

- Observed number of cases (y_j): The number of paediatric cases of NB in the exposed community was identified through community sources and verified by medical records. A case was defined as a child (aged 0–14) of Caucasian ethnicity with a medically confirmed diagnosis of NB between the years 1995 and 1999. To be eligible, cases in the exposed community had to have had a minimum period of residence in the community of one continuous year (including the period of gestation) prior to diagnosis. Four Caucasian cases were identified that met the inclusion criteria. They had periods of residence in the community (beginning at birth) that ranged from 4 months to 13 years (plus gestational period). The main analysis was carried out for these four Caucasian NB cases for the period 1995–1999. Two additional analyses were carried out to address criticisms. The first is for the period 1995–2000 which includes a fifth Asian case of NB who was identified in the year 2000. The second is for an extended 10-year period 1990–1999 for the four Caucasian cases only.

- Size of the exposed community (N_j): Total population estimates for all age groups for the exposed community were obtained for the years 1990–2000 from the local Department of Finance. Total population estimates were also available from the U.S. Census Bureau and were comparable to those provided by the local Department of Finance. Neither source of data provided population estimates that were age and race specific. Let N be the size of the exposed population for the periods 1990–1999. Let r_j be the proportion of individuals that were less than 15 years old and

Caucasian, obtained from the 1990 U.S. census. The plaintiffs' experts estimated the size of the exposed population as $N_j = r_j N$. However, this method did not take into account any changes in the age and racial structure of the community, post-1990. Temporal trends in the age and racial structure of a population can underlie observed differences in incidence, if not accounted for. For example, assume a community has an annual incidence of two cases of NB in each of 4 years for a total of eight cases across the 4-year time period and that the population of Caucasian individuals less than age 15 in the community is known for Year 1 to be 1000 persons. If we assume the population at risk to be constant across the 4 years, then the incidence rate for NB in Caucasian children for that time period in that community would be 8/4000 or 20.00×10^{-4}. However, now assume that the population at risk (Caucasian, aged $\leqslant$14) is actually decreasing over the time period so that there are 1000 persons in Year 1, 800 in Year 2, 600 in Year 3 and 400 in Year 4 (a total of 2800 for the 4-year period). In the latter ('actual') case, the incidence rate would be 8/2800 or 28.57×10^{-4}—higher than the rate calculated when we did not account for temporal changes in the population at risk. Therefore, plaintiff's experts also obtained the age and racial composition of the exposed community from both the 1990 and the 2000 census to account for changes in the age and racial structure of the community over time. To do so, a constant rate of change was assumed across the 10-year interval and a second set of estimates for the size of the exposed population of the community for the years 1995–1999 was obtained that was reflective of the changes in the age and racial composition of the city in the 1990s. While the proportion of individuals aged $\leqslant$14 years changed very little across the 10-year time period (0.4% decrease), the change in the racial structure of the community between 1990 and 2000 was more significant. There was a 17.8% decrease in the proportion of individuals who were Caucasian between the years 1990 and 2000. Corresponding increases were seen in the Asian and non-black 'other' racial groups. The proportion of black individuals remained constant over the 10 years. In summary, to estimate the size of the exposed population, the size of the total population in the exposed community was obtained for each year of the analysis and then adjusted based on the estimated proportion of individuals $\leqslant$14 and Caucasian ethnicity. These estimates were then aggregated over the total analysis period.

- Unexposed populations: the following populations were considered as control (or unexposed) populations:
  - U.S. counties with available incidence data from the surveillance, epidemiology and end results (SEER) cancer registry that were demographically similar to the exposed community according to a set of predefined criteria. There were approximately 431 U.S. counties that were part of SEER. Five unexposed populations were extracted from these 431 counties matched to the exposed community on the following demographic variables: gender, age, race, educational attainment, income above the poverty level and median household income.
  - California counties with available SEER incidence data that were demographically similar to the exposed community according to a same criterion used for the U.S. counties.
  - The State of California (average annual rates for the years 1990–1994 published by the California Cancer Registry), not matched on demographic characteristics.
  - A 10% sampling of the United States as represented by the SEER database not matched on demographic characteristics.

Each control population group described above was obtained by aggregating the units it included (e.g. counties from the SEER registries) to form one population. The demographically similar comparison groups do not include the exposed community or the county of San Bernardino, California. However, the California rates were for the state as a whole and no such exclusions were made.

- Matching criteria: Initially, the matching criteria were such that the racial composition (specifically the proportion that were Caucasian and Asian), the age distribution (specifically the proportion of subjects younger than 15 years of age) and the median household income in the control population were similar to the distribution in the exposed community within a range of ±3%. None of the SEER 431 U.S. counties with available incidence data met these criteria. We found that the median household income was a limiting factor. Because of its proximity to the Los Angeles area, the median income of the exposed community was much higher than other areas of the United States that were otherwise similar to the exposed community. To produce a range of matched control populations, specific matching criteria were relaxed—specifically, the proportion of people of Caucasian and Asian ethnicity and median household income because these were the two characteristics in which the exposed community was an outlier compared to the overall United States. The matching criteria are described in Table 2.

TABLE 2 *Standardized incidence ratios (SIRs) for NB, age* $\leqslant$ 14*, for the exposed community, for the years* 1995–1999*, Caucasian only (no. of observed cases* = 4*)*

Total US	Counties	No. of cases	Person-time	SIR	95% CI	*p*-value (one sided)
Total US (SEER)	N/A	171	20 749 294	10.81	2.80–30.09	0.0006
DS counties in the US (1)	6	9	758 958	7.48	1.68–26.79	0.0023
DS counties in the US (2)	9	25	2 872 001	10.19	2.58–29.50	0.0007
DS counties in the US (3)	16	42	4 085 668	8.62	2.25–23.77	0.0013
DS counties in the US (4)	10	42	3 707 325	7.83	2.04–21.57	0.0019
DS counties in the US (5)	19	41	5 791 807	12.52	3.26–34.58	0.0003
Total CA	N/A	N/A	N/A	9.05	2.47–23.44	0.0011
DS counties in CA (6)	4	24	1 678 105	6.20	1.56–18.04	0.0057
DS counties in CA (7)	7	85	4 289 404	4.47	1.19–11.88	0.0143
DS counties in CA (8)	10	94	5 093 287	4.80	1.28–12.71	0.0011

Definitions of control populations: (1) Relaxed race, tight income: race (proportion Caucasian plus Asian) within upper quartile of distribution containing the exposed community; age ±5%, gender ±5%, median income \$ $\geqslant$ 45 000. (2) Tight race, tight income: race (proportion Caucasian plus Asian); age ±5%, gender ±5%, median income \$ $\geqslant$ 40 000. (3) Relaxed race, relaxed income: race (proportion Caucasian plus Asian) within upper quartile of distribution containing the exposed community; age ±5%, gender ±5%, median income \$ $\geqslant$ 40 000. (4) Relaxed race, relaxed income minus Alaska: race (proportion Caucasian plus Asian) within upper quartile of distribution containing the exposed community; age ±5%, gender ±5%, median income \$ $\geqslant$ 40 000, minus Alaskan counties that matched. (5) Within tertile of exposed community: with regard to proportion of residents who were male, of Caucasian or Asian ethnicity, between the ages of 0 and 14, had high school diploma (or equivalent), incomes above the poverty level (age 15–64) and lived in a rural (non-farm) area. (6) California, TIER 1 (similar but for income): race (proportion of Caucasian or Asian) within upper quartile of distribution containing the exposed community; age ±5%, gender ±5%. (7) California, TIER 2: Tier 1 counties plus three counties that either (1) met median income and age criteria but not the race criteria or (2) were largely Caucasian/Asian but missed the age and income criteria. (8) California, TIER 3: all 10 California counties with available SEER incidence data.

TABLE 3 *Example of SIR calculation* $\left(SIR = \frac{No.\ of\ observed\ cases\ in\ the\ exposed\ community}{No.\ of\ cases\ expected\ in\ the\ exposed\ community\ based\ on\ rate\ in\ the\ control\ population}\right)$ *in Table* 2

Sample calculation	
Observed no. of cases (Caucasian, age $\leqslant$ 14) in exposed community during the time period 1995–1999	4
Population at risk in the exposed community (Caucasian, age $\leqslant$ 14) during the time period 1995–1999	45 115
Incidence rate of NB (Caucasian, age $\leqslant$ 14) in the control community (DS counties in the US (1)) during the time period 1995–1999	Nine cases of NB/758 958 population at risk $= 11.86 \times 10^{-6}$
Expected no. of cases in exposed community during the time period 1995–1999 based on incidence rates in control community	= (Rate of NB in control community, 1995–1999) × (population at risk in the exposed community, 1995–1999) $= (11.86 \times 10^{-6}) \times (45\,115)$ = 0.535
SIR	= (Observed no. of cases in exposed community, 1995–1999)/(expected no. of cases in exposed community, 1995–1999) = 4/0.535 = 7.48

- Probability of disease in the unexposed population (q_j): For each control population, the probability of disease was calculated by dividing the number of cases that occurred in the control population by the size of the population in strata j. For each control population, the number of NB cases $\leqslant$14 years of age was obtained from the SEER program data for the years 1995–1999. Case counts for study and comparison groups were all generated using identical disease definitions (i.e. ICD-0-2 codes) which were provided by diagnosing physicians for the NB cases in the exposed community, and were part of the SEER public-use data as provided on CD-ROM. The age- and race-specific population estimates were obtained, at the county level, from the U.S. Census Bureau. SEER registries cover approximately 10% of the U.S. population.

Table 2 shows the control populations, the number of counties, the number of NB cases, the person-time in the control populations (i.e. the number of people in the control population over the course of the specified time period), the estimated SIRs and their 95% confidence intervals (CIs). The results of Table 2 show that, regardless of the control population, the observed number of NB cases in the exposed community exceeds the expected value for all 10 comparison groups, and these increased incidence rates were all statistically significant (the 95% CIs did not include 1). SIRs ranged from 4.47 to 9.05 using California control populations and from 7.48 to 12.52 using U.S. control populations.

The plaintiffs' experts carried out several sensitivity analyses to explore the sensitivity of the estimated SIR to the prior decisions listed above. Carrying out extensive sensitivity analyses is a sound practice to assess whether criticisms will have an important impact on the inferences (see, e.g. Gastwirth, 2005). Specifically, analyses were carried out for the period 1995–2000 which includes a fifth Asian case, and extending the study period from 1990 to 1999 for the four Caucasian cases. Analyses for five cases for the period 1995–2000 and for four cases for the extended 10-year period 1990–1999 also showed statistically significant SIR (see Tables 4 and 5).

TABLE 4 *Standardized incidence ratios (SIRs) for NB, age of diagnosis* $\leqslant$ 14*, for the exposed community, for the years* 1995–2000*, Caucasian and Asian subgroup*

Control population	Counties	No. of cases	Person-time	SIR	95% CI
Total US (SEER)	N/A	233	28 727 642	8.33	2.61–20.55
DS counties in US (1)	6	10	1 013 447	6.86	1.84–22.01
DS counties in US (2)	9	33	3 747 324	7.68	2.34–19.80
DS counties in US (3)	16	60	6 017 255	6.78	2.13–16.73
DS counties in US (4)	10	60	5 522 040	6.23	1.95–15.35
DS counties in US (5)	19	56	7 394 077	8.93	2.79–22.11
Total CA (CA Cancer Registry)	N/A	N/A	N/A	7.43	2.42–17.50

TABLE 5 *Standardized incidence ratios (SIRs) for NB, age of diagnosis* $\leqslant$ 14*, for the exposed, for* 10*-year interval* (1990–1999)*, Caucasians only*

Control population	Counties	No. of cases	Person-time	SIR	95% CI
DS counties in US (4)	10	78	7 318 926	4.28[†]	1.14–11.40

[†]Compares to SIR = 7.83 for same control population, years 1995–1999.

3. The critiques of the defense experts

The statistical analyses conducted by the plaintiff were challenged by three defense experts (D1, D2 and D3). Their opinions are summarized below.

3.1 *D1*

3.1.1 *Public health methods (#1 and #2).* D1 critiqued the report of the plaintiff by introducing the 'public health methods #1 and #2' as methods designed to establish causation. D1 wrote that the choice of method #1 or #2 should be guided by how the concerns about a particular rare disease came to light. The two methods are described below:

1. *The sequence of events begins when neighbours express concern about the predictable health effects of an environmental exposure to a known toxin. In this case, positive conclusions about the relationship between an exposure and a condition can be accepted as scientifically valid only if several facts are verified, including the fact that a causal relationship has been established already.*
2. *The first alert occurs when the local frequency of cases is perceived to be excessive, whether or not suspicion subsequently falls upon a local exposure. In this case, method* #1 *is inappropriate and the most useful public health procedure starts with an assessment of the degree to which the observed number of cases is truly unusual.*

Under D1's opinion, because the initial observation was the excess number of cases, we should use the public health method #2 and not #1. According to D1, the results of public health method #1 cannot be accepted as scientifically valid because there is no prior documentation of a causal relationship between exposure to carcinogens from the industry source and NB. Specifically, D1 wrote: *If positive conclusions about the relationship between an exposure and a condition based on this method, such as those advanced in the plaintiff's expert report, are to be accepted as scientifically valid and useful, several facts must be verified.* The first fact mentioned in D1's report is: *That studies have already established a causal relationship between the exposure and the condition; i.e., that the condition is already known to result from the observed exposure.*

3.1.2 *95% versus 99% CIs.* D1 criticized the plaintiffs' calculation of 95% CIs and advocated the calculation of 99% CIs as a more conventional approach.

3.1.3 *The study period.* Under D1's opinion, the period at risk should be 1988–2003. D1 wrote: *since the emissions from the industry source are presumed by the plaintiff to have begun well before 1988, the beginning of population-based cancer ascertainment in San Bernardino County, the period at risk should be presumed to last from 1988 until the end of 2003, the last complete year of cancer ascertainment.*

3.1.4 *The exposed population.* Under D1's opinion, the exposed population should include the total population, including adults and all races. He wrote: *it makes little difference, whether adults, at lower risk, are considered along with the children, since the use of appropriate age-adjusted rates from the reference standard are designed to maintain comparability.*

3.1.5 *Prevalence of similar circumstances.* D1 also wrote: *to establish whether there exists a truly extreme concentration of cases, the investigator must therefore necessarily compute the number of comparable communities of comparable size that might be similarly affected by chance*. D1 uses the Poisson statistical distribution to demonstrate that the observed number of three, four or five cases of NB occurred in the exposed community could be attributed to chance. Specifically, D1 suggested taking the population of California (approximately 32 million), dividing it by the population of the exposed community (approximately 53 000) and obtaining 599 communities (32 million divided by 53 000) of the same population size as the exposed community. He focussed on a 16-year study period 1988–2003, and by linearly interpolating the total population of the exposed community for the period 1990 and 2000, he estimated a population at risk for a period of 16 years to be equal to 800 731 (more simply, the person-years could also be calculated as $53\,000 \times 16 = 848\,000$). Then, D1 used the Los Angeles NB incidence rate for all age groups (0.2 per 100 000 people) to calculate the expected number of cases in the exposed community which is equal to 0.2×8.007. Finally, by sampling 599 independent and identically distributed Poisson random variables with a rate equal to 0.2×8.007, D1 found that in 113, 48 and 15 out of the 599 communities, of the same size as the exposed population, we would expect to see at least 3, 4 and 5 NB cases by chance alone.

D1 also used the national SEER rate of NB of 0.5 per 100 000. In this case, the number of cases to be expected would be equal to $0.5 \times 8.007 = 4$, and he found that in 456, 339 and 222 communities out of the 599 California communities, we would expect to see at least 3, 4 and 5 cases by chance alone.

3.2 *D2*

3.2.1 *Selection bias.* D2 considers the plaintiff's incidence analysis flawed because it is affected by 'selection bias'. D2 wrote: *I understand that plaintiff's analysis of the incidence of neuroblastoma in the exposed community arose in this case not through an a priori choice by her of the exposed community as a research site, independent of any a priori knowledge of the incidence of neuroblastoma there, but because she was retained by counsel for plaintiff, prompted specifically by their knowledge of cases of neuroblastoma in the exposed community.* He concluded: *As a result of her failure to recognize or to remedy this 'Texas sharpshooter' issue, plaintiff's statistical inferences in this case are fatally flawed.*

3.2.2 *Conditional probability.* D2 criticizes the plaintiff's assumption that the observed NB cases follow a Poisson sampling distribution. He/she believes that this assumption is inappropriate when an unusually high number of cases has been brought to light prior to conducting the analysis. D2 wrote: *The statistical tests performed by the plaintiff to establish the statistical significance of the apparently 'elevated incidence rates' of neuroblastoma in the exposed community are flawed, and the conclusions that she draws from those tests are unreliable.* D2 added: *plaintiff's finding that elevated incidence rates of neuroblastoma in the exposed community are statistically significant is sensitive to the precise specification of her tests, and evaporates when the specifications of the tests are adjusted to account for selection bias in the data.*

Based upon D2's opinion, because the plaintiff knew that more than two NB cases occurred in the exposed community prior to starting their data collection, the sampling distribution should be a conditional probability law that takes into account this information (when the counsel for plaintiffs approached the experts, they had already begun case ascertainment, so they were aware of some

number of NB cases, but the exact number was unclear). More specifically, the calculation of the CIs should be based upon a Poisson distribution conditional on an outcome of two or more incident cases, instead of the naive, unconditional Poisson model as a benchmark distribution for statistical testing.

3.3 *D3*

3.3.1 *Clustering methods*. D3 suggested using statistical methodology for cluster investigation, and in particular to use the method of Cuzick–Edwards (1990). D3 wrote: *in the various reports of the plaintiff's expert witnesses, none discuss the assessment of the alleged the exposed community neuroblastoma cluster using any new and appropriate scientific methodology, such as the Cuzick-Edwards tests. Instead, they rely on the improper and outdated methods that are currently widely recognized as being scientifically flawed. All rely in lockstep upon the 'Texas Sharpshooter', 'usual suspects' and 'raindancers' fallacies.*

4. The plaintiff's reply

4.1 *Reply to D1*

4.1.1 *Public health methods*. Public health methods #1 and #2 for establishing causation are not standard practice in epidemiology. For example, in Gordis (2004), the epidemiological approach is defined as follows: *Epidemiology is a multi-step process. The first step is to determine whether an association exists between a factor (e.g., an environmental exposure), and the development of a disease. We do this by studying the characteristics of groups and the characteristics of individuals. If we find that there is indeed an association between an exposure and a disease, is it necessarily a causal relationship? No, not all associations are causal. The second step is therefore to try to derive appropriate inferences regarding a possible causal relationship from the patterns of the associations that have been found.*

In fact, epidemiologists first gather evidence for an association and then try to establish causation by doing sensitivity analyses, exploring alternative explanations for the observed associations, replicating their findings and carrying out animal studies when appropriate. The plaintiff's incidence analysis was not aimed at establishing causation. The plaintiff conducted a standardized incidence ratio analysis to calculate incidence rates. This weight of evidence analysis was performed in response to a public concern of an excess number of NB cases that could be attributed to potential exposures to toxic substances released from the industry source. In this case, there was a specific concern inherent to the potential health effects of the toxic emissions of the industry source. Because we know the putative source of hazard, its location and when the contamination occurred, conducting a SIR analysis is the standard and commonly accepted approach.

4.1.2 *99% versus 95% CIs*. The plaintiff uses the standard approach of reporting the 95% CI which corresponds to a 5% level of significance level. This provides strong evidence of an excess of NB cases in the exposed community to justify a public health action.

4.1.3 *Study period*. The plaintiff's choice of a 5-year study period 1995–1999 was dictated by the fact that the toxicity of the emissions from the industry source changed substantially after 1990.

Therefore, they hypothesized that there was a spike in NB incidence after 1990. They also repeated the analyses including the extended 10-year period 1990–1999 and found similar results (Table 5).

4.1.4 *Population at risk*. In the plaintiff's report, the population at risk was restricted to younger than 15 years of age because this form of cancer is almost non-existent in the adult population (Goodman *et al*., 1999). The accepted and conventional practice for calculating incidence rates of childhood cancers is to restrict the age range of children (Ries *et al*., 2006). Would the evidence change by including the adult population? We have calculated the p-value in this circumstance. The p-value is the probability of observing four or more cases under the null distribution (i.e. under a Poisson distribution with a relative risk (RR) equal to 1) for a 5-year period. In the exposed community, the total population in 1995 is 53 523; 28% are younger than 14 and 72% are adult, so the sizes of the two populations at risk are $N_1 = 53\,523 \times 0.28 = 14\,986.44$ and $N_2 = 53\,523 \times 0.72 = 38\,536.56$, respectively. The 5-year NB incidence rate for younger than 14 in the control population (p_1) can be calculated by dividing the number of cases by person-time. Using the data summarized in Table 2, the 5-year NB incidence rate for the younger than 14 with a control population equal to the nation is equal to 171 divided by 20 749 298, which is equal to 8.24 per 1 million. We assume that the 5-year NB incidence rate (p_2) for the adult population is 2 per 1 million (this is a conservative assumption because this is the annual age-adjusted rate of NB in Los Angeles County for all ages combined for the period 1972–1999 used by D1). Thus, the 5-year expected number of cases is equal to $E = N_1 p_1 + N_2 p_2 = (14\,986.44 \times 8.24/1\,000\,000 + 38\,536.56 \times 2/1\,000\,000) = 1.002807$. The p-value, here defined as the probability of observing four or more cases under the null (SIR $= 1$), is equal to $p(y \geqslant 4|\text{SIR} = 1, E = 1.002807) = 0.019$. In summary, if we include the entire population (younger and older than 15 years old) as the population at risk, we still find a p-value lower than 0.05.

4.1.5 *Prevalence of similar circumstances*. D1's calculation is a tutorial example for illustrating how to calculate the p-value and it is summarized in Table 6. The p-value is the exact significance probability of obtaining a value of statistics at least as extreme, in relation to the null hypothesis, as that observed (Gibbons and Pratt, 1975). In this case, D1 calculates the probability of observing four or more cases under the null distribution (i.e. under a Poisson distribution with a SIR equal to 1) for a prespecified time period and for a prespecified expected number of cases. Thus, with rates equal to $1.6 = 0.2 \times 8.007$, the p-values is equal to $p(y \geqslant 4|\text{SIR} = 1) = 1 - p(y \leqslant 3|\text{SIR} = 1) = 0.078$. Therefore, we would expect to see $599 \times 0.078 = 48$ communities with at least four NB cases by chance alone.

D1's calculations are based upon assumptions that merit further discussion. First, D1 assumes that because the 599 California communities have the same number of people at risk as the exposed community, then they are 'comparable' to the exposed community. However, simply dividing California into 599 sections with the same population size as the exposed community does not make these units similar to the exposed community because they are likely to be very different in demographic and socio-economic status and not truly contiguous 'communities'. A community must be defined by exposure, geographical borders and density. Second, the choice of a 16-year study period (1988–2003) used by D1 is inappropriate because the toxic emissions were much different after 1990 than before. Third, there was no justification given for choosing the Los Angeles population as the unexposed population. Los Angeles is close to the study location and therefore it is not a good

TABLE 6 *Summary of the calculations conducted by D1: calculation of the number of communities of the same size as the exposed community where at least four NB cases would be expected by chance*

	Study period (years)	Person-years	Expected number of cases using the Los Angeles incidence rate for all age groups combined (0.2 per million) (E)	p-value $p(y \leqslant 3 \mid E, \text{SIR} = 1)$	Number of communities of the same size of the exposed community where at least four cases would be expected by chance
D1	16	800 731	1.601	0.078	48
Plaintiff's expert	5	267 615	0.534	0.022	1

control population. In addition, it is important to point out that the rate of NB in Los Angeles County was three times higher than the national average during the study period.

In the second row of Table 6, we repeat the calculations conducted by D1 but for a 5-year period instead of 16-year period, and under the very conservative (and inappropriate) assumption of using Los Angeles as a unexposed population. We found that, with the population at risk equal to (53 523 × 5) = 267 614, the expected number of cases is equal to 0.534 which leads to a p-value equal to 0.002. Thus, in this case, we found that in only 1 in 599 California communities, would we expect to see at least 4 cases by chance alone. In summary, the result—that in less than 2% of the communities of the same size of the exposed community, we would see at least four cases by chance alone—suggests that four is an unusually high number.

4.2 *Reply to D2*

4.2.1 *Selection bias*. In epidemiological studies for putative sources of hazard, the goal is to estimate the risk of disease among people living in an area near a possible source of environmental pollution. Therefore, the choice of a geographical area to study is determined by the exposure. The usual statistical approach is to choose a time period and an area around the source and to classify the disease cases and person-years by age and sex. The statistical and epidemiological literature have pointed that these types of studies are affected by the following weakness: '*the hypothesis about a source may have been formed, and some of the boundaries of the study area chosen, because of the informal knowledge about the number of cases nearby* (Hills *et al*., 1989)'. However, Lawson (1993) observed that if a study region is noted *a priori* to be of interest because it includes a pollution source, one does not suffer from *post hoc* analysis problems because the internal spatial structure of disease incidence did not influence the choice of the region.

The plaintiff defined the geographical boundaries of the exposed community based upon the location of the hazardous waste. In addition, the plaintiff chose as study periods 5- and 10-year intervals (1995–1999 and 1990–1999), as is typically reported by cancer registries. The 1990–1999 period was chosen because the composition of waste burned after July 1990 posed a much higher risk to human health in the exposed community than prior to 1990. This was due to the closing of the popper ovens and incinerators and more uncontrolled waste disposal activities. In addition, because NB is an embryonal cancer, analyses were also performed for the 'latent' period 1995–1999 to allow for some latency with regard to exposure of parental germ cells and developing foetuses. In summary, the study area and the time period were chosen because they include the pollution source, a heightened period of exposure and a heavy influx of people, including young children.

Certain populations and cohorts have been studied intensively over a period of decades because of their proximity to a point source of hazardous exposure. A well-known example is the population of Seveso, Italy, which was exposed to TCDD (dioxin) through an explosion at a trichlorophenol plant in 1976. Studies of this cohort exposed to dioxin have demonstrated an excess risk of breast cancer (Warner *et al*., 2002), lung cancer, non-Hodgkin's lymphoma, Hodgkin's disease, rectal cancer, leukemia, lymphohemopoietic cancers, lymphoreticulosarcoma myeloma and melanoma (Bertazzi *et al*., 1993, 1997, 2001). In children, increases were seen for all cancers, including brain cancer, Hodgkin's disease, lymphatic leukemia and myeloid leukemia and thyroid cancer (Pesatori *et al*., 1993).

Close residential proximity to municipal incinerators and hospital incinerators has also been shown to be causally related to the incidence of childhood cancers, including leukemia (Knox, 2000), and to the increased risk of soft tissue sarcomas and non-Hodgkin's lymphoma (Floret *et al*., 2003; Viel *et al*., 2000).

4.2.2 *Conditional probability.* It is important to point out that whether one or two additional NB cases are an unusually high *number with respect to having already reported more than two cases* is not the scientific question under study. The scientific question is whether the total number of NB cases diagnosed during the study period is unusually high *with respect to a background number in a comparable unexposed population*. The standard Poisson analysis is the most widely acceptable scientific methodology for an SIR analysis. It is a commonly used method in cancer surveillance (Hills and Alexander, 1989; Brookmeyer and Stroup, 2004) even when this calculation is performed in response to a concern from the public triggered by an unusually high number of cases.

4.3 *Reply to D3*

Clustering methods determine when an observed collection of cases represents an anomaly unlikely to have arisen by chance. These methods are used for public health surveillance when the source of contamination is unknown (Cuzick and Edwards, 2000; Brookmeyer *et al*., 2004). If there is evidence of clustering, then additional epidemiological studies are carried out to identify the source of the hazard. A cluster can be defined as a 'geographically bounded group of occurrences of sufficient size and concentration to be unlikely to have occurred by chance'.

A disease clustering analysis, which ignores the location of the source of hazard, is not appropriate in this case because the choice of the exposed community made by the plaintiff was determined by the location of the source of the hazardous waste. However, statistical methods for cluster analyses are useful because the infrequency of a cluster is generally a critical piece of evidence that might lead to a study of association between the disease and the putative source of hazard (Fienberg and Kaye, 1991). Interestingly, D3 did not conduct the Cuzick–Edwards analysis.

5. Discussion

This paper illustrates a toxic tort case where injury is alleged to have resulted from exposure to toxic substances. The current status of the case is in the court records.

Toxic tort cases commonly rely on epidemiological evidence. Epidemiology seeks to describe and explain disease occurrence and prevention on a population level (Gordis, 2004). In the context

of litigation, including actions to recover damages for injuries alleged to have arisen from exposure to specific substances, the law is concerned with the facts at the individual level, and seeks a determination as to whether the disease in a specific individual is attributable to a specific cause or factor, to a reasonable degree of scientific certainty. The dissonance in both purpose and perspective (the search for the true versus the search for justice) creates a tension between the two disciplines that is perhaps most apparent in their approach to an evaluation of causation. This topic has been discussed extensively in the literature (Gold, 1986; Loue, 2000; Gastwirth, 2003; Kennedy, 2004). Therefore, it is natural that the defense experts shared similar criticisms that take advantage of the complexity in evaluating causation in toxic torts. Some of these common defense positions (including some points that were critical to the case but beyond the scope of this paper) are briefly discussed below:

- *Any result that does not reach statistical significance is not relevant*. The defense experts generally do not acknowledge the several factors that can influence the failure to reach statistical significance. The actual risk estimates, their associated statistical uncertainty and their sensitivity to untestable assumptions are the most important aspects of the study findings and not the *p*-values.
- *Negative studies prove that there is no evidence of increased risk, and the lack of consistency in the literature indicates that there is no real causal association*. Often in the courtroom, negative studies are misinterpreted as representing proof of the null hypothesis. It is important to note that a 'negative' finding (e.g. an SIR with a non-statistically significant *p*-value or a CI that includes the null value of 1.0) does not prove the null hypothesis of no association (Axelson, 2004). Negative studies should be interpreted in the context of the potential limitations of the analysis. Rothman and Greenland (1998) wrote: *When a single study forms the sole basis for a choice between two alternative actions… a decision-making mode of analysis may be justifiable. Even then, however, a rational recommendation about which of two actions is preferable will require consideration of the costs and benefits of each action, and these considerations are rarely incorporated into statistical tests… Such decisions are inevitably based on results from a collection of studies, and proper combination of the information from the studies requires more than just a classification of each study into 'significant' or 'not significant'* [emphasis added].
- '*Need for specificity, i.e. it is necessary to have a large database of human epidemiological studies that show statistically significantly elevated risks for the SPECIFIC cancer and the SPECIFIC chemical being investigated. Evidence from animal studies is not relevant*'. Virtually, all diseases are multifactorial and many, if not most, aetiologic agents have a multiplicity of effects on a range of organ systems (Rothman, 1976). An example would be the numerous diseases caused by smoking (e.g. lung cancer and heart diseases), as documented in the many reports to the Surgeon General in the 1980s. While it is widely accepted that the observational nature of epidemiologic studies, if well conducted, provides the best basis for estimating human risk (Stayner and Smith, 1992), animal studies provide important and complementary data on toxicology and risk assessment. There is general acceptance for at least a qualitative relationship between experimental animals and humans with regard to the pathophysiology of disease and qualitative inter-species extrapolation is widely accepted and applied in the field of epidemiology and risk assessment (Goodman and Wilson, 1991). Simply put, if a substance causes disease in test animals, then that same substance is considered harmful to humans. This is evidenced by the classification scheme used by the International Agency for Research on Cancer, in which evaluations of the strength of evidence for carcinogenicity are made using both human and experimental animal data.

- '*Causality can be determined only if all the Bradford Hill guidelines are met*'. Hill (1971) himself asserted that the postulates should be used as controls and that none could be required as essential (*a sine qua non*) (see also Morabia, 1991; Phillips and Goodman, 2004). We believe that a weight of evidence analysis that takes into consideration all toxicological, human epidemiological and *in vitro* data is also appropriate. Components of the Hill criteria which are considered in a weight of evidence analysis include the following: strength of the association, consistency of findings across epidemiological and toxicological studies, evidence for a dose response, biologic plausibility, temporality and coherence of the scientific evidence across disciplines (e.g. epidemiology and toxicology). However, it should be noted that these criteria were meant to constitute '*a sensible or reasonable standard for imputing causation in practice or for policy reasons where knowledge of a disease may be incomplete*' (Charlton, 1996) and are clearly multidisciplinary, requiring both human population studies and support from basic, experimental science.

- '*The cancer is caused by other factors*' (such as smoking, obesity and other environmental exposures) even though these other factors have even less causal basis than the putative source of hazard. It is widely accepted that most, if not all, chronic diseases have a multifactorial aetiology and arise through multiple mechanisms; i.e. the 'cause' actually consists of multiple components that act in concert (MacMahon and Pugh, 1970; Rothman and Boice, 1982; Rothman and Greenland, 1998). Rothman defines a cause of a disease as 'an event, condition or characteristic that plays an essential role in producing an occurrence of the disease'. Multiple non-mutually exclusive 'constellations' of component causes can and do exist (Kelsey *et al*., 1996; MacMahon and Pugh, 1970; Rothman and Boice, 1982; Rothman and Greenland, 1998). As such, the existence of any particular cause does not negate the causal role of any other cause (e.g. hormonal status, age and genetic make-up).

- '*The estimated RR must be larger than two to support causality from an observed association*'. Specifically, the question arises whether epidemiological data demonstrating a RR greater than two are required to meet the standard for proof ('more likely than not') (Carruth and Goldstein, 2000). This approach is problematic for several reasons. There are many factors that influence the magnitude of the RR in an epidemiological study, including the prevalence of the risk factor in the population, the background rate of the disease and the sample size. The adoption of such a rule has the undesired limitation that it would preclude the compensation of individuals harmed by an exposure to a hazardous waste, where a RR less than two has been estimated. In summary, a RR of 1.5 may have greater public health significance than a RR of 5, depending on the prevalence of the risk factor. The discussion of a threshold requirement based upon the magnitude of the RR can be analogized to the definition of sensitivity and specificity of an epidemiological test. Loue (1999) wrote: '*in the context of litigation, a high threshold requirement, such as a RR of 3.0, would result in a lesser proportion of unharmed individuals being erroneously classified as harmed, but would also result in a larger proportion of truly harmed individuals being classified as unharmed, there by precluding the recovery of these injuries'*.

- '*A two-sided test of significance must be conducted*' instead of one-sided even if the scientific context of the study clearly supports the unidirectional nature of the association (as in this case). The choice of a one-sided versus a two-sided test is determined by prior knowledge from other studies about the direction of the disease/exposure relationship and the purpose of the study. One-tailed tests are normally used when the association between exposure and disease is hypothesized

to be in one direction only. A two-sided test is one in which the values for which we can reject the null hypothesis are located in both tails of the probability distribution and therefore allows for the possibility that exposure can either increase or decrease the risk of disease.

In conclusion, epidemiologists and biostatisticians play key roles in establishing causation in toxic tort litigation. Studies of the incidence of cancer or other conditions in a community with a known or putative source of hazardous emissions can serve as a critical piece of evidence in the causal argument. Despite the application of standard and widely accepted methodological practices in epidemiology to causal questions in a toxic tort case, such as the calculation of standardized incidence ratios and the choice of defensible periods of analysis, defense experts more often than not rely on an arsenal of epidemiological fallacies (e.g. a non-significant study proves the null hypothesis), unconventional points of view (e.g. a RR must exceed two and be statistically significant at the 0.05 level to be meaningful), anecdotal jargon (e.g. 'Texas Sharpshooter' and 'Raindancer' fallacy) and hand-waving to dismiss positive findings. It is also important to note that often defense experts did not reanalyse the data, they only criticized it. In this regard, Judge Posner in a employment discrimination case (*Allen and Battle* v. *Seidman* 881 F. 2d (7th Cir. 1987)) said: '*This is especially true in the present case since the defendant, while taking pot shots—none fatal—at the plaintiffs' statistical comparison, did not bother to conduct its own regression analysis, which for all we know would have confirmed and strengthened the plaintiffs' simpler study'*. In summary, for a complex process such as toxic tort litigation—and with practical matters of consequence at stake—awareness of the important role of epidemiology in litigation, including the appropriate utilization of standard epidemiological tools and practices, is essential for good practice in the administration of justice.

Acknowledgments

We thank the editors and the referees for their insightful comments on this manuscript.

References

Axelson, O. Negative and non-positive epidemiological studies. *Int J Occup Med Environ Health* 2004; **17**:115–121.

Bertazzi, P. A., Pesatori, A. C., Consonni, D., Tironi, A., Landi, M. T. & Zocchetti, C. Cancer incidence in a population accidentally exposed to 2,3,7,8-tetrachlorodibenzo-para-dioxin. *Epidemiology* 1993; **4**:398–406.

Bertazzi, P. A., Zocchetti, C., Guercilena, S., Consonni, D., Tironi A., Landi, M. T. & Pesatori, A. C. Dioxin exposure and cancer risk: a 15-year mortality study after the "Seveso accident". *Epidemiology* 1997; **8**:646–652.

Bertazzi, P. A., Consonni, D., Bachetti, S., Rubagotti, M., Baccarelli, A., Zocchetti, C. & Pesatori, A. C. Health effects of dioxin exposure: a 20-year mortality study. *Am J Epidemiol* 2001; **153**:1031–1044.

Breslow, N. E. & Day, N. E. *Statistical Methods in Cancer Research. Vol. II, The Design and Analysis of Cohort Studies* (IARC Scientific Publication No. 82). Lyon, France: International Agency for Research on Cancer, 1987.

Brodeur, G. M. Neuroblastoma and other peripheral neuroectodermal tumors. Chapter 24. In: Fernbach, D. J. & Vietti, T. J. (Eds). *Clinical Pediatric Oncology*, 4th ed. St. Louis, MO: Mosby Yearbook, 1991.

BRODEUR, G. M. & CASTLEBERRY, R. P. Neuroblastoma. In: Pizzo, P. A. & Poplack, D. G. (Eds). *Principles and Practices of Pediatric Oncology*, 3rd ed. Phildadelphia, PA: Lippencott-Raven, 1997:761–797.

BROOKMEYER, R. & STROUP, D. *Monitoring the Health of Populations: Statistical Methods for Public Health Surveillance*. Oxford University Press, 2004.

BUNIN, G. R., WARD, E., KRAMER, S., RHEE, C. A. & MEADOWS, A. T. Neuroblastoma and parental occupation. *Am J Epidemiol* 1990; **131**(5):776–780.

CARRUTH, R. S. & GOLDSTEIN, B. D. Relative risk greater than two in proof of causation in toxic tort litigation. *Jurimetrics J* 2000; **41**:195–209.

CHARLTON, B. G. Attribution of causation in epidemiology: chain or mosaic? *J Clin Epidemiol* 1996; **49**:105–107.

CUZICK, J. & EDWARDS, R. Spatial clustering for inhomogenous populations. *J R Stat Soc B* 1990; **52**:73–104.

DANIELS, J. L., OLSHAN, A. F., TESCHKE, K., HERTZ-PICCIOTTO, I., SAVITZ, D. A., BLATT, J., BONDY, M. L., NEGLIA, J. P., POLLOCK, B. H., COHN, S. L., LOOK, A. T., SEEGER, R. C. & CASTLEBERRY, R. P. Residential pesticide exposure and neuroblastoma. *Epidemiology* 2001; **12**(1):20–27.

DE ROOS, A. J., OLSHAN, A. F., TESCHKE, K., POOLE, C., SAVITZ, D. A., BLATT, J., BONDY, M. L. & POLLOCK, B. H. Parental occupational exposures to chemicals and incidence of neuroblastoma in offspring. *Am J Epidemiol* 2001a; **154**(2):106–114.

DE ROOS, A. J., POOLE, C., TESCHKE, K. & OLSHAN, A. F. An application of hierarchical regression in the investigation of multiple paternal occupational exposures and neuroblastoma in offspring. *Am J Ind Med* 2001b; **39**(5):477–486.

ESIASHVILI, N., GOODMAN, M., WARD, K., MARCUS, R. B. & JOHNSTONE, P. A. Neuroblastoma in adults: incidence and survival analysis based on SEER data. *Pediatr Blood Cancer* 2007; **49**(1):41–46.

FAJEN, J. M., ROUNBEHLER, D. P. & FINE, D. H. Summary report on N-nitrosamines in the factory environment. *IARC Sci Publ* 1982; **41**:223–229.

FIENBERG, S. E. & KAYE, D. H. Legal and statistical aspects of some mysterious clusters. *J R Stat Soc* 1991;**154**(1):61–74.

FLORET, N., MAUNY, F., CHALLIER, B., ARVEUX, P., CAHN, J. Y. & VIEL, J. F. Dioxin emissions from a solid waste incinerator and risk of non-Hodgkin lymphoma. *Epidemiology* 2003; **14**:392–398.

GASTWIRTH, J. L. The need for careful evaluation of epidemiological evidence in product liability case: a reexamination of Wells v. Ortho and Key Pharmaceuticals. *Law Probab Risk* 2003; **2**:151–189.

GASTWIRTH, J. L. Some issues arising in the presentation of statistical testimony. *Law Probab Risk* 2005; **4**:5–20.

GIBBONS, J. D. & PRATT, J. W. P-values: interpretation and methodology. *Am Stat* 1975; **29**:20–25.

GOLD, S. Causation in toxic torts: the burdens of proof, standards of persuasion, and statistical evidence. *Yale Law J* 1986; **96**:376–402.

GOODMAN, G. & WILSON, R. Predicting the carcinogenicity of chemicals in humans from rodent bioassay data. *Environ Health Perspect* 1991; **94**:195–218.

GOODMAN, M. T., GURNEY, J. G., SMITH, M. A. & OLSHAN, A. F. Sympathetic Nervous System Tumors. In: Ries, L. A. G., Smith, M. A., Gurney, J. G., Linet, M., Tamra, T., Young, J. L., Bunin, G. R. (Eds). *Cancer Incidence and Survival among Children and Adolescents: United States SEER Program 1975–1995*. Bethesda, MD: National Cancer Institute, SEER Program. NIH Pub. No. 99-4649, 1999.

GORDIS, L. *Epidemiology*. W.B, Saunders Industry source, 2004.

HILL, A. B. *Principles of Medical Statistics*, 9th ed. New York: Oxford University Press, 1971.

HILLS, M. & ALEXANDER, F. Statistical methods used in assessing the risk of disease near a source of possible environmental pollution: a review. *J R Stat Soc A* 1989; **152**:353–363.

KENNEDY, D. Science, law, and the IBM case. *Science* 2004; **305**:340–341.

KERR, M. A., NASCA, P. C., MUNDT, K. A., MICHALEK, A. M., BAPTISTE, M. S. & MAHONEY, M. C. Parental occupational exposures and risk of neuroblastoma: a case-control study (United States). *Cancer Causes Control* 2000; **11**(7):635–643.

KNOX, E. Childhood cancers, birthplaces, incinerators and landfill sites. *Int J Epidemiol* 2000; **29**:391–397.

KRAMER, S., WARD, E., MEADOWS, A. T. & MALONE, K. E. Medical and drug risk factors associated with neuroblastoma: a case-control study. *J Natl Cancer Inst* 1987; **78**(5):797–804.

LAWSON, A. B. On the analysis of mortality events around a pre-specified fixed point. *J R Stat Soc A* 1993; **156**:363–377.

LOUE, S. *Forensic Epidemiology: A Comprehensive Guide for Legal and Epidemiology Professionals*. Carbondale, IL: Southern Illinois University Press, 1999.

LOUE, S. Epidemiological causation in the legal context: substance and procedures. Chapter 13. In: Gastwirth, J. L. (Ed). *Statistical Science in the Courtroom*. New York: Springer-Verlag, 2000.

MCKINNEY, P. A., FEAR, N. T. & STOCKTON, D. Parental occupation at periconception: findings from the United Kingdom Childhood Cancer Study. *Occup Environ Med* 2003; **60**(12):901–909.

MACMAHON, B. & PUGH, T. F. *Epidemiology: Principles and Methods*. Boston, MA: Little, Brown, and Industry source, 1970.

MICHAELIS, J., HAAF, H. G., ZOLLNER, J., KAATSCH, P., KRUMMENAUER, F. & BERTHOLD, F. Case control study of neuroblastoma in west-Germany after the Chernobyl accident. *Klin Padiatr* 1996; **208**(4): 172–178.

MORABIA, A. On the origin of Hill's criteria. *Epidemiology* 1991;**2**:367–369.

NATIONAL TOXICOLOGY PROGRAM. *TR-434: Toxicology and Carcinogenesis Studies of 1,3-Butadiene (CAS No. 106-99-0) in B6C3F$_1$Mice (Inhalation Studies)*, 1993.

NATIONAL TOXICOLOGY PROGRAM. *TR-500: Toxicology and Carcinogenesis Studies of Naphthalene (CAS No. 91-20-3) in F344/N Rats (Inhalation Studies)*, 2000.

OLSHAN, A. F., DE ROOS, A. J., TESCHKE, K., NEGLIA, J. P., STRAM, D. O., POLLOCK, B. H. & CASTLEBERRY, R. P. Neuroblastoma and parental occupation. *Cancer Causes Control* 1999; **10**(6): 539–549.

PESATORI, A. C., CONSONNI, D., TIRONI, A., ZOCCHETTI, C., FINI, A. & BERTAZZI, P. A. Cancer in a young population in a dioxin-contaminated area. *Int J Epidemiol* 1993; **22**: 1010–1013.

PHILLIPS, C. V. & GOODMAN, K. J. The missed lessons of Sir Austin Bradford Hill. *Epidemiol Perspect Innov* 2004:1–3.

RIES, L. A. G., HARKINS, D., KRAPCHO, M., MARIOTTO, A., MILLER, B. A., FEUER, E. J., CLEGG, L., EISNER, M. P., HORNER, M. J., HOWLADER, N., HAYAT, M., HANKEY, B. F. & EDWARDS, B. K. (Eds). *SEER Cancer Statistics Review, 1975-2003*, Bethesda, MD: National Cancer Institute. http://seer.cancer.gov/csr/1975_2003/ based on November 2005 SEER data submission, posted to the SEER web site, 2006.

ROTHMAN, K. J. Causes. *Am J Epidemiol* 1976; **104**: 587–592.

ROTHMAN, K. J. & BOICE, J. D., JR. *Epidemiologic Analysis with a Programmable Calculator*, New Edition. Boston, MA: Epidemiology Resources, Inc., 1982.

ROTHMAN, K. J. *Modern Epidemiology*. Boston, MA: Little, Brown and Industry source, 1986.

ROTHMAN, K. J. & GREENLAND, S. *Modern Epidemiology*. Lippincott Williams & Wilkins Publishers, 1998.

SAHAI, H. & KHURSHID, A. *Statistics in Epidemiology: Methods, Techniques, and Applications*. Boca Raton, FL: CRC Press, Inc., 1996.

SCHWARTZBAUM, J. A. Influence of mother's prenatal drug consumption on risk of neuroblastoma in the child. *Am J Epidemiol* 1992; **135**: 1358–1367.

SMULEVICH, V. B., SOLIONOVA, L. G. & BELYAKOVA, S. V. Parental occupation and other factors and cancer risk in children: II. Occupational factors. *Int J Cancer* 1999; **83**(6):718–722.

SPITZ, M. R. & JOHNSON, C. C. Neuroblastoma and paternal occupation. a case-control analysis. *Am J Epidemiol* 1985; **121**(6):924–929.

STAYNER, L. T. & SMITH, R. J. Methodologic issues in using epidemiologic studies of occupational cohorts for cancer risk assessment. *Epidemiol Prev* 1992; **14**:32–39.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES. *PHSNTP. Report on Carcinogens*, 10th ed, 2002.

VIEL, J. F., ARVEUX, P., BAVEREL, J. & CAHN, J. Y. Soft-tissue sarcoma and non-Hodgkin's lymphoma clusters around a municipal solid waste incinerator with high dioxin emission levels. *Am J Epidemiol* 2000; **152**:13–19.

WARNER, M., ESKENAZI, B., MOCARELLI, M., GERTHOUX, P., SAMUELS, S., NEEDHAM, L., PATTERSON, D. & BRAMBILA, P. Serum dioxin concentrations and breast cancer risk in the Seveso Women's Health Study. *Environ Health Perspect* 2002; **110**:625.

WILKINS, J. R., III & HUNDLEY, V. D. Paternal occupational exposure to electromagnetic fields and neuroblastoma in offspring. *Am J Epidemiol* 1990; **131**(6):995–1008.

Appendix A

A.1 *Glossary of terms*

Confidence interval: A CI is a measure of the precision of an estimated value. Instead of estimating the parameter by a single value, a whole interval of likely estimates is given. The more likely it is for the interval to contain the parameter, the wider the interval will be. More precisely, a CI for a population parameter is an interval with an associated probability p that is generated from a random sample of an underlying population such that if the sampling was repeated numerous times and the CI recalculated from each sample according to the same method, a proportion p of the CIs would contain the population parameter in question. CIs are the most prevalent form of interval estimation.

***p*-value:** In statistical hypothesis testing, the p-value is the probability of obtaining a result at least as extreme as a given data point, assuming the data point was the result of chance alone. The fact that p-values are based on this assumption is crucial to their correct interpretation.

More technically, the p-value of an observed value t of some random variable T used as a test statistic is the probability that, given that the null hypothesis is true, T will assume a value as or more unfavourable to the null hypothesis as the observed value t . 'More unfavourable to the null hypothesis' can in some cases mean greater than, in some cases less than and in some cases further away from a specified centre.

Poisson distribution: Poisson distribution is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and are independent of the time since the last event.

Quartile: In descriptive statistics, a quartile is any of the three values which divides the sorted data set into four equal parts so that each part represents one quarter of the sampled population.

Standardized incidence ratio (SIR): The ratio of the observed to the expected disease cases; the expected number is based on the age-specific rates in a control population designated as 'normal', average or 'unexposed'.

Tertile: In descriptive statistics, a tertile is either of the two values which divides the sorted data set into three equal parts so that each part represents one-third of the sampled population.