An Analysis of Contaminated Well Water and Health Effects in Woburn, Massachusetts

S. W. LAGAKOS, B. J. WESSEN, and M. ZELEN*

In 1979, two of the eight municipal wells servicing Woburn, Massachusetts, were discovered to be contaminated with several chlorinated organics. Shortly afterwards, the town was found to have an elevated rate of childhood leukemia. Using recent information about the space–time distribution of water from the two contaminated wells, we find positive statistical associations between access to this water and the incidence rates of childhood leukemia, perinatal deaths (1970–1982), two of five categories of congenital anomalies, and two of nine categories of childhood disorders. We find no associations with spontaneous abortions, low birth weight, or the other categories of congenital anomalies and childhood disorders. This article discussed these results and other features of the data relevant to their interpretation.

KEY WORDS: Environmental exposure; Health survey; Observational study; Proportional hazards model; Time-dependent covariate.

1. INTRODUCTION

Woburn, Massachusetts, is a community of 37,000 residents located 12 miles from Boston. It has been an industrial site for more than 130 years. The town was a major chemical and leather processing center, a producer of arsenic compounds for insect control, and a producer of textiles, paper, TNT, and animal glues.

The town's drinking water was supplied by eight municipal wells, two of which (designated by G and H) are contiguous and operated as a single source in eastern Woburn. The chance discovery of some toxic wastes near wells G and H led to testing of them in May 1979 for 32 volatile organics in the Environmental Protection Agency's (EPA) list of 129 priority pollutants. Trichloroethylene (267 ppb), tetrachloroethylene (21 ppb), and chloroform (12 ppb) were detected, whereupon the wells were shut down (Massachusetts Department of Environmental Quality and Engineering 1979). Trichlorotrifluoroethane (23 ppb) and dichloroethylene (28 ppb) were also detected (U.S. EPA 1979) in water samples taken several months later. The types and amounts of contamination in wells G and H prior to 1979 are not known. The remaining town wells, located near Horn Pond in southwest Woburn (Fig. 1), were tested and found to meet both state and federal drinking-water standards.

Independently, during site excavations in July 1979 for an industrial complex located north of wells G and H (Fig. 1), large pits of buried animal hides and chemical wastes were discovered. A nearby abandoned lagoon was found to be heavily contaminated with lead, arsenic, and other metals. Subsequently, the groundwater under eastern Woburn was sampled at 61 test wells and found to contain 48 EPA priority pollutants and raised levels of 22 metals (Ecology and Environment, Inc. 1982).

The closing of the two municipal wells and the discovery of the abandoned waste sites occurred at about the same time as the Love Canal incident and alerted Woburn residents to possible health hazards. One resident contacted the Centers for Disease Control (CDC), asking if cancer rates were elevated in Woburn. A citizens' group formed and, in late 1979, produced a list of children diagnosed with leukemia.

These events led to a series of studies in 1980–1981 by the Massachusetts Department of Public Health (MDPH) and CDC. A review of mortality statistics for 1969–1979 showed that the overall cancer mortality rate in Woburn was significantly higher than those of the state and six adjacent communities (Kotelchuck and Parker 1979). Another investigation (Parker and Rosen 1981) found a potential concentration of adult renal cancer in the Horn Pond area. The study also concluded that there was a significantly elevated rate of childhood leukemia in Woburn between 1969 and 1979, with 12 cases diagnosed when only 5.3 were expected ($O/E = 2.3$, $P = .008$). The study noted that the leukemia excess was accounted for primarily by six cases occurring in one (3334) of the town's six census tracts (Fig. 1). As part of the same report, a case-control study did not identify any etiologic factors for the leukemias. Because of a lack of information on exposure to the contaminated wells, however, it was not possible to assess the association between access to this water and the risk of leukemia.

We attempted to build on these earlier results in two ways. First, by using recent information on the space–time distribution of water from wells G and H, we were able to assess the association between access to this water and the incidence rate of childhood leukemia. Second, we undertook a sample survey of adverse pregnancy outcomes and childhood disorders to determine whether any of these were correlated with exposure to water from the contaminated wells. Such disorders are much more common and

* S. W. Lagakos and M. Zelen are Professors and B. J. Wessen is Research Assistant, Department of Biostatistics, Harvard School of Public Health, and Division of Biostatistics and Epidemiology, Dana-Farber Cancer Institute, Boston, MA 02115. The data management and computing costs for this project were partially funded by a Biomedical Research Support Grant from the Harvard School of Public Health and grants from the Alfred P. Sloane Foundation, the National Cancer Institute (CA23415, CA39640), and the SIAM Institute for Mathematics and Society. The authors are grateful to the Massachusetts Department of Environmental Quality and Engineering for their water distribution report, to A. M. Walker, who was involved in the planning of the health survey, to their colleagues and others who reviewed earlier drafts of this article, to the volunteer interviewers, and to the citizen's group FACE, who helped in the design and implementation of the survey.

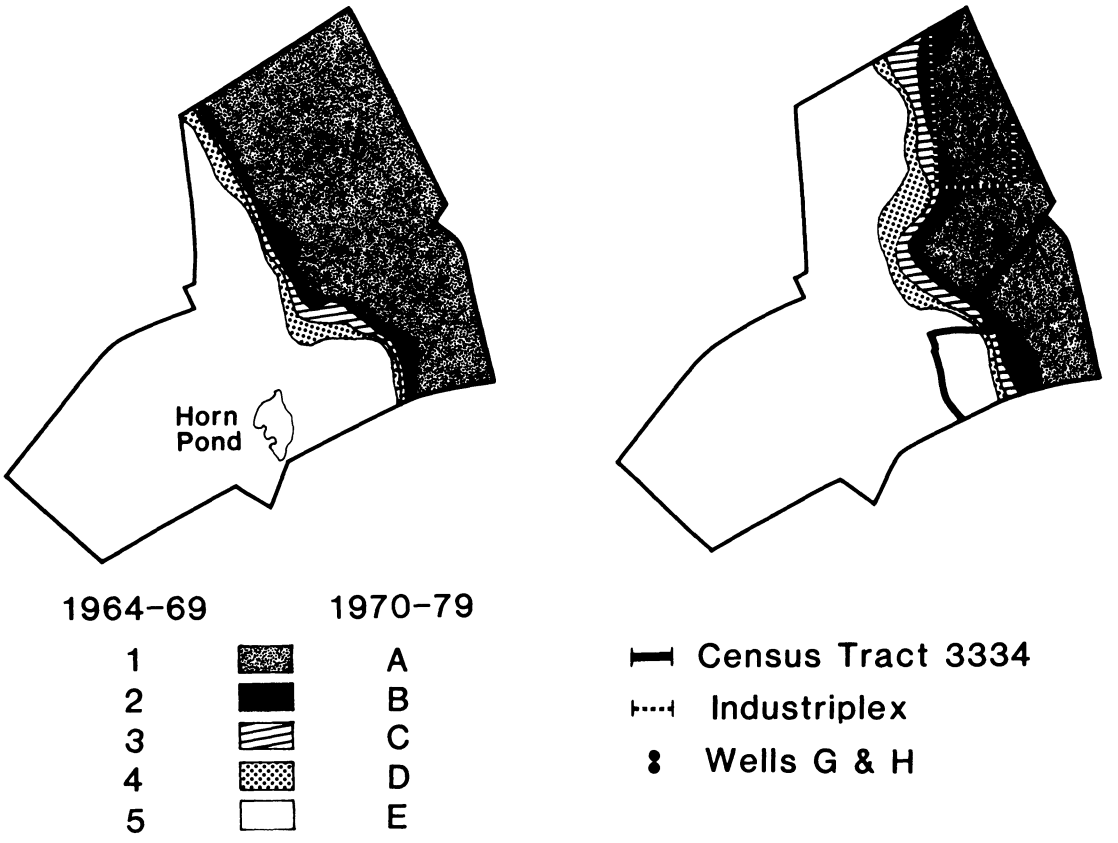

Figure 1. Outline of Woburn, Massachusetts, Showing 1960–1969 and 1970–1982 Zones of Graduated Exposure to Wells G and H. Zones 1 and A represent the greatest G and H exposure, and Zones 5 and E represent the least (no) G and H exposure.

presumably have a shorter latency period than leukemia and adult cancers and, therefore, may be more sensitive indicators of adverse health effects. In this article we present and discuss our findings. Because of limitations of space, a number of details could not be included in this article but are contained in a technical report (Lagakos, Wessen, and Zelen 1985) available upon request.

2. SOURCES OF DATA

2.1 Childhood Leukemia

We obtained information on 20 cases of childhood leukemia (ages 19 and under) diagnosed in Woburn between 1964, the year wells G and H began pumping, and 1983. The 20 cases are all those identified by the state and Dana-Farber Cancer Institute/Children's Hospital tumor registries. These include the 12 cases used in the 1969–1979 MDPH study, 1 case identified by the MDPH but not used because it was diagnosed prior to 1969, and 7 new cases diagnosed since 1980. Overall, 15 of the 20 cases were males, 17 had acute lymphocytic leukemia, and 15 were born in Woburn. The median age at diagnosis was 7 years.

2.2 The Health Survey

In 1982, a telephone sample survey of Woburn households gathered information on adverse pregnancy outcomes and childhood disorders occurring to former and current family members between 1960 and 1982. The survey is discussed here. Potential biases and results of efforts to corroborate survey findings are discussed in Section 5.

To carry out the survey, we recruited 301 volunteer interviewers, of whom 235 actually participated in the interview process. About half were from Woburn and the remaining were from nearby towns or the university community. Interviewers attended a training session that included an explanation of the items in the interview and two practice interviews, one as interviewer and one as respondent. They were then randomly assigned a packet of interview forms bearing 25 consecutive telephone numbers selected from the 1982 town directory. The interviewers knew neither the names nor the addresses corresponding to their assigned phone numbers and were instructed to keep the entire interview anonymous. The telephone numbers themselves are noninformative with respect to location within Woburn. After the interviewers returned their packets (some finished several packets), we reviewed each questionnaire and recontacted the households of any with missing or ambiguous information. In a follow-up phase we also attempted to contact and complete interviews for households that had never been called or had been called but never reached.

A total of 8,109 telephone numbers were called at least once, and 7,134 (88%) of these were completed. The remaining 975 numbers were called an average of 3.9 times. Of the completed calls, 915 were businesses, second phones, or disconnected numbers, leaving 6,219 distinct residences

that were reached. Of these 6,219, 1,149 (18%) refused to be interviewed and 60 were excluded from analysis because the respondent was not sufficiently fluent in English (32 households) or because the data were too incomplete (28 households). We estimate that the sample of 5,010 completed interviews represents 57% of the town's residences with listed telephones.

The telephone interview itself followed a scripted self-coding questionnaire. Households were first asked if any members resided in Woburn prior to 1979, and the interview was terminated for those with no such members. For the remaining households, the interviewer obtained data on all pregnancies ending between 1960 and 1982 of women born since 1920, including when the pregnancy ended, maternal age and smoking status during pregnancy, offspring vital status at delivery, and offspring weight, gender, and congenital anomalies. Information was also collected on chronic and recurrent health problems in children and on the residence history of each current or former family member, excluding the current address (which was kept in a separate file and only later linked to the interview data).

For the purposes of the survey we defined a spontaneous abortion as a pregnancy in which the embryo or fetus is prematurely expelled in the first six months and a perinatal death as a pregnancy lasting over 6 months and resulting in a stillbirth or a livebirth that survives fewer than 7 days. The reported data on pregnancy outcomes exclude an unknown number of spontaneous abortions that occur early in pregnancy and go unnoticed, as well as information on elective abortions. We defined the cutoff for low birth weight to be 6 pounds (2,722 g) instead of the customary 2,500 grams because we felt that respondents would be more familiar with the English system and might sometimes remember only the number of pounds (but not ounces) that their children weighed at birth.

Medically diagnosed congenital anomalies were grouped according to the involved organ or system using the International Classification of Disease (ICD) codes. Apart from musculoskeletal, cardiovascular, and eye/ear defects (Heinonen, Slone, and Shapiro 1977), the remaining major system-grouped categories of congenital anomalies contained too few cases to be analyzed separately. Rather than analyze these remaining anomalies as one large "catch-all" category, two subcategories were formed: (a) central nervous system (CNS), chromosomal, and oral cleft anomalies, for which we could find assertions in the literature of potential links with chemicals, pesticides, or trace elements (see Bennett and Aborns 1979; Dorsch, Scragg, McMichael, Baghurst, and Dyer 1984; Gordon and Shy 1981; Holmberg 1979; Klingberg, Papier, and Hart 1983; Poladnek and Janerich 1983) and (b) "other" defects, for which we could find no specific assertions of a potential environmental link. All groupings of congenital anomalies and childhood disorders (described subsequently) were made prior to examination of any residence, risk factor, or G and H exposure data.

The survey questionnaire grouped medically diagnosed childhood disorders into 14 categories. For the purpose of analysis, these categories were collapsed into 9 broader categories of disorders, each with at least 20 cases: anemia and other blood disorders, heart and blood pressure disorders, kidney and other urinary tract disorders, lung and other respiratory disorders, neurologic and sensory organ disorders, allergy and skin disorders, learning disabilities, diabetes and glandular disorders, and "other" disorders.

Information on socioeconomic status (SES) was obtained from census data for 1960, 1970, and 1980 (Bureau of the Census 1960, 1970, 1980). On the basis of four socioeconomic indicators (median family income, median home value, percentage of high school graduates, and percentage of families above poverty level), two of Woburn's six census tracts (3333 and 3331) consistently ranked lowest and highest, respectively, and the remaining four were similar. We defined, therefore, a three-point SES scale, where tract 3333 was low, tract 3331 was high, and the remaining four tracts were middle SES.

2.3 Exposure to Water From Contaminated Wells

Wells G and H operated as a single source in eastern Woburn. They are 88 feet and 84 feet deep, respectively, and tap a different aquifer than the six other municipal wells, which are located in southwest Woburn. Since the town's water pipes are interconnected, however, each Woburn residence received a "blend" of water from several wells. The specific blend varied with the location of the residence within Woburn and with chronologic time.

In the 15-year period (October 1964–May 1979) of their operation, wells G and H pumped for 3,162 days, or 59% of the time, and supplied some water to an average of 23% of the town's households per year. A report by the Massachusetts Department of Environmental Quality and Engineering (DEQE) estimated the regional and temporal distribution of water from wells G and H during this period (Waldorf and Cleary 1983). The DEQE study relied on a detailed model of the Woburn water distribution system and partitioned the town into five zones of graduated exposure to wells G and H. Because of a substantial change in industrial demand in 1970, a different set of five zones was used before and after 1970 (Fig. 1). The irregular shapes and sizes of the zones of graduated exposure reflect the uneven population density of Woburn. The DEQE study estimated, on a monthly basis, which zones received none, some, or all of their water from wells G and H. These results, obtained in August 1983, were combined to estimate the percentage of each household's annual water supply that arose from wells G and H (Table 1). For example, in 1970 a residence located in Zone B received an estimated 27% of its water from wells G and H, the remainder coming from the town's other municipal wells. Note that wells G and H are estimated never to have serviced residences in "West" Woburn (the intersection of Zones 5 and E), which contains about $\frac{2}{3}$ of the town's population. The G and H water distribution in the rest of Woburn ("East" Woburn) varied markedly with both region and chronologic time.

We merged G and H exposure information with our

Table 1. Annual G and H Exposure Scores by Zone

Year	1960–1969 Zones 1	2	3	4
1960–1963	0	0	0	0
1964	.23	.05	0	0
1965	.08	.08	.08	.08
1966	.95	.70	.25	.12
1967	.51	.47	.32	.17
1968	.72	.34	0	0
1969	.75	.40	0	0
Year	**1970–1982 Zones A**	**B**	**C**	**D**
1970	.54	.27	.03	0
1971	.46	.38	.08	0
1972	.29	.29	.14	0
1973	0	0	0	0
1974	.37	.32	.25	.14
1975	.55	.44	.13	.01
1976	.49	.38	.09	0
1977	.94	.52	.04	.01
1978	1.00	.88	.61	.05
1979	.39	.39	.29	0
1980–1982	0	0	0	0

NOTE: Table entries give estimated fraction of residential water supply derived from wells G and H by year and residential zone. Refer to Figure 1 for zonal definitions. Exposure scores are zero for Zones 5 and E in all years.

other data and assigned to each pregnancy the annual exposure score corresponding to the mother's residence in the year the pregnancy ended. We also determined for each child an exposure "history," consisting of his or her set of annual exposure scores, beginning from the first year of Woburn residency. For example, an exposure of .49 would be assigned to a pregnancy ending in 1976 for a mother residing in Zone A. Similarly, a child born in 1967 and residing in the intersection of Zones 1 and B for the first 4 years of life would generate cumulative exposures of .51, 1.23, 1.98, and 2.25 during this period. If a child changed residences, we arbitrarily defined his or her exposure score for that year to be the score corresponding to the former residence.

3. STATISTICAL METHODS

3.1 Childhood Leukemia

Based on national rates (SEER 1981), the 20 childhood leukemia cases observed between 1964 and 1983 are significantly higher than expected ($E = 9.1$, $P = .001$). To determine whether the space–time distribution of these cases within Woburn is correlated with water from wells G and H, we used the failure time regression model (Cox 1972)

$$h\{t \mid x(t), y\} = h_y(t)\exp\{\alpha x(t)\}, \tag{1}$$

where $x(t)$ is some expression of G and H exposure history from birth to age t, y is the year of birth, $h\{t \mid x(t), y\}$ is the leukemia risk (hazard function) at age t for an individual born in year y and with exposure $x(t)$, and $h_y(t)$ is the baseline Woburn risk at age t for an unexposed person born in year y. With this model, the relative risk at age t for someone with exposure $x(t)$, relative to an unexposed individual born in the same year, is $\exp\{\alpha x(t)\}$. The hypothesis of no association is given by $\alpha = 0$.

We used two exposure metrics $x(t)$: (a) cumulative G and H exposure from birth until age t and (b) a binary indicator of whether there had been any G and H exposure by age t. With the first of these measures, risk increases steadily with cumulative exposure. With the latter, an individual's hazard function jumps from $h_y(t)$ to $h_y(t)\exp(\alpha)$ upon exposure. Partial-likelihood-based tests of $\alpha = 0$ from this "none versus some" exposure model are closely related to the variation of the log-rank test proposed by Mantel and Byar (1974) (see also Aitkin, Laird, and Francis 1983; Crowley and Hu 1977). Misspecification of the form of $x(t)$ with either test results in a loss of efficiency but not in a distortion of size.

The partial likelihood score test for $\alpha = 0$ can be expressed in the form (see Kalbfleisch and Prentice 1980)

$$\sum (X_i - E_i)/(\sum V_i)^{1/2}, \tag{2}$$

where the sum is over the leukemia cases and X_i is the observed value of $x(t)$ for the ith case at t_i, the age of diagnosis. The quantities E_i and V_i are the average and variance of the $x(t_i)$ for the "risk set" of children born in the same year as the ith case and not diagnosed with leukemia before age t_i and represent the null mean and variance of X_i, conditional on this risk set. When $\alpha = 0$, the distribution of (2) is approximately $N(0, 1)$. A closed-form approximation to the maximum likelihood estimator (MLE) of α is given by $\Sigma (X_i - E_i)/\Sigma V_i$. This will closely approximate the MLE for α close to 0 but may be conservative for large $|\alpha|$.

In our situation we had the exposure histories for all of the leukemia cases. We had only the exposures, however, for those noncases identified in the sample survey. Accordingly, E_i and V_i could not be computed directly and were estimated. One approach is to adopt a form of risk-set sampling and estimate E_i and V_i from the survey data (see Breslow, Lubin, Marek, and Langholtz 1983; Cox and Oakes 1984; Liddell, McDonald, and Thomas 1977; Prentice 1985); that is, for each case we identified all surveyed children who were born in the same year and were residents at the same time as the case and then computed the average and variance of their $x(t)$ values for the period of residency of the case. The results of this approach are presented in detail in Section 4.1. Alternatively, since each individual's exposure history is uniquely determined by his or her residence history, E_i and V_i can be estimated from the population distributions of the G and H exposure zones. Details of this approach are given in the Appendix.

When $x(t)$ is positively associated with the risk of leukemia, the number of leukemia cases that are statistically "explained" by the association is

$$\Sigma_1 [h(t_i \mid X_i, y) - h_y(t_i)]/h(t_i \mid X_i, y) = \Sigma_1 [1 - \exp(-\alpha X_i)],$$

where the sum is over cases (see National Research Council 1985).

3.2 Adverse Pregnancy Outcomes

Adverse pregnancy outcomes were analyzed using the method of maximum likelihood applied to a logistic regression model (see Cox 1970); that is,

$$\log [p/(1 - p)] = \mu + \alpha x + \beta_1 z_1 + \beta_2 z_2 + \cdots + \beta_k z_k, \quad (3)$$

where p is the probability of an event, $z_1, z_2, \ldots, z_k$ are risk factors, and x is the mother's G and H exposure score for the year her pregnancy ended. Note that $\exp(\alpha x)$ is the odds-ratio of an event for a pregnant woman who receives $100x\%$ of her residential water from wells G and H relative to one with the same risk factors who receives none of her residential water from wells G and H.

Analyses of spontaneous abortions and perinatal deaths were based on all reported pregnancies. Analyses of low birth weight and birth anomalies were based on 7-day survivors rather than all live births, since these events were not well documented in some of the surveyed children who died before 7 days.

The risk factors used in the analysis were maternal age and smoking status during pregnancy, year of pregnancy, SES, sex, and mother's pregnancy history. We first identified important risk factors by fitting Equation (3) with $\alpha = 0$ and then controlled for these in evaluating G and H exposure by fitting Equation (3) and testing $\alpha = 0$ versus $\alpha > 0$. Risk factors involving a mother's pregnancy history were controlled by the inclusion of indicator variables of pregnancy number into the regression models. If there was no statistically significant association between G and H exposure and an adverse pregnancy outcome, we also tested for an interaction between G and H exposure and time period (1960–1969 vs. 1970–1982). The approach of fitting x only after screening for important risk factors can, in theory, fail to identify a risk factor that is strongly correlated with x. Each of the risk factors that we examined, however, was essentially uncorrelated with x ($r^2 < .01$), so this is unlikely to have occurred.

3.3 Childhood Disorders

Childhood disorders were analyzed using a "survival time" model, where time was the age of diagnosis. Ideally, we would like to have used Cox's (1972) model with covariates $x(t), z_1, z_2, \ldots, z_p$, where $x(t)$ is cumulative exposure to age t and the z_i are risk factors. Given the large number of observations and the time-dependent covariate $x(t)$, however, such an approach was not computationally feasible. Instead, we first identified important risk factors by fitting Cox's model without $x(t)$ and then controlled for these in assessing G and H exposure by applying model (1), with y indexing strata corresponding to combinations of important risk factors.

4. RESULTS

4.1 Childhood Leukemia

Using either the cumulative ($P = .03$, $\hat{\alpha} = .33$) or none versus some ($P = .02$, $\hat{\alpha} = 1.11$) exposure metrics, there is a positive association between G and H exposure and the incidence rate of childhood leukemia. Table 2 gives information on the 20 cases and their contributions to the score tests of $\alpha = 0$. For example, case 13 was born in

Table 2. Observed and Expected Exposures to Wells G and H for 20 Childhood Leukemia Cases

Case	Year of diagnosis	Year of birth	Period of residency	Observed cumulative exposure	Size of risk set sample	Expected cumulative exposure (var)	Proportion of risk set exposed
1	1966	1959	1959–1966	1.26	218	.31 (.26)	.33
2	1969	1957	1968–1969	0	290	.34 (.36)	.26
3	1969	1964	1969	.75	265	.17 (.10)	.25
4	1972	1965	1965–1972	4.30	182	.90 (2.23)	.36
5	1972	1968	1968–1972	2.76	183	.58 (.88)	.32
6	1973	1970	1970–1973	.94	170	.20 (.20)	.19
7	1974	1965	1968–1974	0	213	.56 (1.04)	.29
8	1975	1964	1965–1975	0	239	.99 (2.78)	.38
9	1975	1975	1975	0	115	.09 (.03)	.25
10	1976	1963	1963–1976	.37	219	1.18 (3.87)	.40
11	1976	1972	1972–1976	0	132	.24 (.32)	.18
12	1978	1963	1963–1978	7.88	219	1.41 (6.23)	.40
13	1979	1969	1969–1979	2.41	164	.73 (2.56)	.31
14	1980	1966	1966–1980	0	199	1.38 (6.00)	.39
15	1981	1968	1968–1981	0	187	1.14 (4.20)	.35
16	1982	1979	1979–1982	.39	154	.08 (.02)	.23
17	1983	1974	1974–77, 1980–83	0	84	.25 (.45)	.23
18	1982	1981	1981–1983	0	—	0 (0)	0
19	1983	1980	1980–1982	0	—	0 (0)	0
20	1983	1980	1981–1983	0	—	0 (0)	0
Totals				21.06		10.55 (31.52)	5.12
Score test statistic:						1.87	2.08
Significance level:						$P = .03$	$P = .02$

NOTE: Risk set for a case consists of children born in the same year as the case and who were residents of Woburn when the case was. Variance of proportion, say p, of risk set exposed equals $p(1 - p)$. Cases 18–20 do not contribute to the test statistic because birth occurred after closure of wells G and H.

1969 and resided in Woburn until being diagnosed in 1979, at which time his cumulative G and H exposure was 2.41. The corresponding 164 surveyed children who were born in Woburn in 1969 and resided there until 1979 had an average cumulative G and H exposure of .73, and 31% of these were exposed. Overall, given the years of birth and periods of Woburn residence of the 20 cases, the expected number exposed to wells G and H when $\alpha = 0$ is $\sum_i E_i = 5.1$, compared with 9 observed, and the sum of their expected cumulative exposures is 10.6, compared with an observed number of 21.1. The alternative approach (see the Appendix) based on estimating E_i and V_i from the regional distribution of the population gave very similar results for both the cumulative ($\sum_i E_i = 11.0$, $P = .04$, $\hat{\alpha} = .29$) and none versus some ($\sum_i E_i = 4.9$, $P = .02$, $\hat{\alpha} = 1.22$) exposure metrics.

Although both G and H exposure metrics are associated with leukemia risk, there are two few cases to be confident of which, if either, best describes this relationship. Furthermore, it does not logically follow that Woburn's entire leukemia excess, based on national rates, is explainable by these associations. Indeed, the cumulative and none versus some metrics of G and H exposure statistically explain about 4 and 6 leukemia cases, respectively, whereas national rates suggest a townwide excess of about 11 cases between 1964 and 1983 (Sec. 3.1). We return to this point in Section 6.

4.2 Adverse Pregnancy Outcomes

The surveyed households provided information on 4,396 pregnancies that terminated in Woburn between 1960 and 1982. The median maternal age at pregnancy was 29 years, and 40% of the mothers smoked during pregnancy. Only 16% of mothers received any water from wells G and H in the year their pregnancy ended. The age distribution of the population and the age-specific smoking rates of pregnant women exhibited no unusual regional variations within Woburn.

Table 3. Frequency and Exposure Status of Specific Birth Anomalies

Anomaly (ICD 9 code)	*Frequency*	*Anomaly (ICD 9 code)*	*Frequency*
Musculoskeletal anomalies	55 (7, 48)	CNS, chromosomal, oral cleft anomalies	27 (8, 19)
Congenital hip defect (754.3)	13 (2, 11)	Central nervous system	15 (5, 10)
Clubfoot (754.7)	10 (4, 6)	Spina bifida (741.9)	3 (2, 1)
Scoliosis (754.2)	6 (0, 6)	Hydrocephalus (742.3)	1 (0, 1)
Various foot deformities (754.5)	5 (0, 5)	Cerebral gigantism (742.4)	1 (1, 0)
Lower limb anomalies (755.6)	4 (0, 4)	Mental retardation (319.0)	4 (0, 4)
Diaphragmatic hernias (756.6)	4 (0, 4)	Cerebral palsy (343.9)	6 (2, 4)
Unspecified limb (755.8)	3 (0, 3)	Chromosomal anomalies	9 (3, 6)
Reduction deformities, lower limb (755.3)	2 (0, 2)	Down's syndrome (758.0)	7 (3, 4)
Polydactyly (755.0)	2 (0, 2)	Other chromosomal (758.9)	2 (0, 2)
Syndactyly (755.1)	2 (0, 2)	Oral cleft	3 (0, 3)
Anomalies of skull (756.0)	2 (0, 2)	Cleft palate (749.0)	3 (0, 3)
Other muscular anomalies (756.8)	2 (1, 1)	Other congenital anomalies	45 (5, 40)
Cardiovascular anomalies	43 (5, 38)	Genitourinary anomalies	13 (3, 10)
Hole in heart (745.9)	8 (1, 7)	Anomaly of ureter (753.4)	3 (2, 1)
Ventricular septal defect (745.4)	2 (0, 2)	Anomaly of kidney (753.3)	1 (0, 1)
Tetralogy of Fallot (745.2)	2 (1, 1)	Extroversion of bladder (753.5)	1 (0, 1)
Mitral insufficiency (746.6)	1 (0, 1)	Atresia of bladder (753.6)	2 (1, 1)
Patent ductus arteriosus (747.0)	1 (0, 1)	Other bladder anomaly (753.8)	3 (0, 3)
Endocardiofibroelastosis (425.3)	1 (0, 1)	Hypospadius (752.6)	1 (0, 1)
Unspecified heart defect (746.9)	9 (1, 8)	Undescended testicle (752.5)	2 (0, 2)
Coarctation of aorta (747.1)	3 (0, 3)	Digestive anomalies	7 (0, 7)
Aortic atresia (747.2)	1 (0, 1)	Pyloric stenosis (750.5)	4 (0, 4)
Heart murmur (785.2)	15 (2, 13)	Hiatal hernia (750.6)	1 (0, 1)
Eye, ear anomalies	18 (9, 9)	Malrotation of small intestine (751.4)	1 (0, 1)
Eye anomalies	16 (8, 8)	Tongue tie (750.0)	1 (0, 1)
Amblyopia (368.0)	3 (1, 2)	Integument anomalies	7 (0, 7)
Strabismus (378.9)	3 (3, 0)	Vascular hematomas (757.3)	5 (0, 5)
Blindness since birth (369.6)	5 (3, 2)	Breast enlargement (778.7)	1 (0, 1)
Eyes severely crossed (378.0)	1 (0, 1)	Congenital lymphedema (757.0)	1 (0, 1)
Coloboma (743.4)	1 (1, 0)	Other anomalies	18 (2, 16)
Congenital cataract (743.3)	1 (0, 1)	Hyperkinesis (314.9)	4 (1, 3)
Hereditary retinal dystrophy (362.7)	1 (0, 1)	Hemangioma (228.0)	2 (0, 2)
Enlarged eye (743.8)	1 (0, 1)	Hypertrophy of adenoids (474.1)	1 (0, 1)
Ear anomalies	2 (1, 1)	Inguinal hernia (550.9)	1 (0, 1)
Born deaf (744.0)	2 (1, 1)	Omphalocele (553.1)	1 (1, 0)
		Pilonidal cyst (685.1)	1 (0, 1)
		"Genetic" Legg Perthes (732.1)	1 (0, 1)
		Anomaly of larynx (748.3)	1 (0, 1)
		Thyroglossal cyst (759.2)	1 (0, 1)
		Paraciticus (759.4)	1 (0, 1)
		Rubenstein Tabi (759.8)	1 (0, 1)
		Deformities of head, fingers, toes (759.8)	1 (0, 1)
		Abnormal liver, kidney (751.6)	1 (0, 1)
		Unspecified defect (759.9)	1 (0, 1)

NOTE: In parentheses are numbers of exposed and unexposed cases, respectively.

Overall, 12% (520) of the 4,396 pregnancies resulted in a spontaneous abortion, and 1.5% (67) resulted in a perinatal death (46 stillbirths and 21 deaths before 7 days). Of the 7-day survivors with known birth weights, 6.4% (220/3,462) were under 6 pounds at birth. Among 177 (4.6%) with congenital anomalies, 167 had a single anomaly, 9 had 2 anomalies, and 1 had 3 anomalies. Table 3 lists the ICD code and G and H exposure status corresponding to each anomaly. Of the 3,809 7-day survivors, 18 died before 1 year of age. These 18 infants had no unusual exposure to wells G and H.

Logistic regression analyses identified several important risk factors for the adverse pregnancy outcomes (Table 4). The importance of maternal age, smoking status, and pregnancy history that we found for spontaneous abortions, perinatal deaths, and low birth weight have all been noted elsewhere (Butler, Goldstein, and Ross 1972; Kline 1977; Kline, Stein, Susser, and Warburton 1977; Naylor and Warburton 1979; Niswander, Gordon, and Berendes 1972; Shapiro, Jones, and Densen 1962; Taffel 1976; U.S. Dept. of Health, Education and Welfare 1979). Similarly, others have noted the increased risk of musculoskeletal defects in children with an older sib who had this anomaly (Heionen et al. 1977), the increasing rate in cardiovascular defects over time (CDC 1981), the declining rate of perinatal deaths (Reed and Stanley 1977), and the association between eye/ear defects and maternal smoking (Christianson, 1980; Heionen et al. 1977; Rantakallio, Krause, and Krause 1979).

Controlling for the important risk factors, there were no statistically significant associations between access to water from wells G and H in the year of pregnancy and the incidence rate of spontaneous abortions ($P = .66$), low birth weight ($P = .77$), or perinatal deaths before 1970 ($P = .55$), or with musculoskeletal ($P = .78$), cardiovascular ($P = .91$), or "other" ($P = .62$) birth anomalies. Positive associations were found, however, with perinatal deaths since 1970 $\{\exp(\hat{\alpha}) = 10.0, P = .003\}$, eye/ear anomalies $\{\exp(\hat{\alpha}) = 14.9, P < .0001\}$, and CNS/chromosomal/oral cleft anomalies $\{\exp(\hat{\alpha}) = 4.5, P = .01\}$. Similar results were obtained when the analyses of low birth weight and congenital anomalies were based on all live births rather than on 7-day survivors and when x was taken to be the average of the G and H exposure scores for the year the pregnancy ended and the previous year.

Table 5 summarizes the unadjusted data, with exposure scores grouped into four intervals. For perinatal deaths, the positive association with G and H exposure is primarily due to the three events in the ".51–1.00" exposure interval; these were stillbirths and occurred in 1977–1978 to women with G and H exposure scores of .94, .94, and 1.00. All but two of the eye/ear anomalies were eye anomalies, primarily blindness, amblyopia, and strabismus. We examined the individual eye anomalies to see if the associ-

Table 4. Risk Factors for Adverse Pregnancy Outcomes and Childhood Disorders

Adverse pregnancy outcome	*Risk factor (reference group)*	*Estimated odds-ratio (95% CI)*
Spontaneous abortion	Smoking during pregnancy (no)	1.4 (1.1–1.7)
	Age at pregnancy	1.2 (1.1–1.4)
	Prior spontaneous abortion (no)	2.2 (1.7–2.9)
	Prior perinatal death (no)	1.5 (.9–2.4)
Perinatal death	Smoking during pregnancy (no)	2.2 (1.3–3.6)
	Year pregnancy ended (1970–1982)	2.8 (1.5–5.0)
	Prior perinatal death (no)	6.4 (2.8–13.9)
Low birth weight	Smoking during pregnancy (no)	1.9 (1.4–2.5)
	Age at pregnancy	1.2 (1.1–1.4)
	Prior low birth weight (no)	10.1 (6.2–16.5)
	Prior perinatal death (no)	2.5 (1.0, 6.1)
Musculoskeletal anomaly	Prior musculoskeletal anomaly (no)	83 (36, 192)
Cardiovascular anomaly	Age at pregnancy	.7 (.5–.9)
	Year pregnancy ended (1960–1969)	2.4 (1.2–4.6)
Eye, ear anomaly	Smoking during pregnancy (no)	2.4 (1.0–5.6)
	Age at pregnancy	1.4 (1.0–2.2)
CNS, chromosomal, oral cleft anomaly	Age at pregnancy	2.3 (1.5–3.3)

Childhood disorder	*Risk factor (reference group)*	*Estimated relative risk*
Anemia/Blood	Year pregnancy ended (1960–1969)	2.4 (1.3–4.3)
	Age at pregnancy	.7 (.6–1.0)
Allergy/Skin	SES (middle:low or high:middle)	1.3 (1.1–1.6)
	Year pregnancy ended (1960–1969)	1.7 (1.4–2.2)
Kidney/Urinary tract	Females (males)	4.6 (1.9–11.1)
Lung/Respiratory tract	Males (females)	1.4 (1.0–1.8)
	Year pregnancy ended (1960–1969)	2.0 (1.5–2.8)
Neurologic/Sensory	Year pregnancy ended (1960–1969)	2.3 (1.5–3.6)
Learning disability	Males (females)	3.1 (2.1–4.4)
	Year pregnancy ended (1960–1969)	1.6 (1.2–2.3)
Other disorders	Year pregnancy ended (1960–1969)	3.2 (1.8–5.9)

NOTE: Odds-ratios and relative risks for age based on age intervals <26, 26–30, 31–35, >35, referring to comparisons of successive intervals. Approximate 95% confidence intervals based on normal approximation for the MLE of regression coefficient β.

Table 5. Unadjusted Rates ($\times 10^3$) of Adverse Pregnancy Outcomes by Fraction of Residential Water Obtained From Wells G and H in Year Pregnancy Ended

End point	G and H exposure score			
	0	.01–.20	.21–.50	.51–1.0
Spontaneous abortion	117 (436/3697)	153 (27/176)	111 (25/226)	108 (32/297)
Low birth weight	66 (193/2908)	38 (5/132)	33 (6/183)	67 (16/239)
Perinatal death (1960–69)	22 (42/1932)	27 (3/113)	10 (1/96)	19 (4/209)
Perinatal death (1970–82)	7 (13/1765)	0 (0/63)	8 (1/130)	34 (3/88)
Congenital anomalies				
Musculoskeletal	15 (48/3206)	14 (2/146)	15 (3/199)	8 (2/258)
Cardiovascular	12 (38/3206)	21 (3/146)	10 (2/199)	0 (0/258)
Eye/Ear	3 (9/3206)	7 (1/146)	15 (3/199)	19 (5/258)
Eye	2 (8/3206)	7 (1/146)	10 (2/199)	19 (5/258)
Ear	0 (1/3206)	0 (0/146)	5 (1/199)	0 (0/258)
CNS/Chromosomal/Oral cleft	6 (19/3206)	7 (1/146)	20 (4/199)	12 (3/258)
CNS	3 (10/3206)	7 (1/146)	20 (4/199)	0 (0/258)
Chromosomal	2 (6/3206)	0 (0/146)	0 (0/199)	12 (3/258)
Oral cleft	1 (3/3206)	0 (0/146)	0 (0/199)	0 (0/258)
Other anomalies	10 (32/3206)	14 (2/146)	5 (1/199)	12 (3/258)

NOTE: Regression analyses based on actual G and H exposure scores. Parentheses contain events/pregnancies for spontaneous abortions and perinatal deaths and events/7-day survivors for other disorders. Low birth weight results exclude 347 pregnancies with unknown birth weights.

ation with G and H was confined to specific subcategories (see Table 3), but no unusual pattern was evident. For the CNS/chromosomal/oral cleft group, the association with G and H exposure is evident in the CNS and chromosomal subcategories but not in the oral cleft category. The CNS anomalies are mainly cerebral palsy, mental retardation, and spina bifida, and all but two of the chromosomal anomalies are Down's syndrome.

4.3 Childhood Disorders

The survey collected information on 5,018 past and current household members who were born since 1960 and resided in Woburn sometime before age 19. We excluded from analysis the 39 infant deaths described earlier and one other child whose records were uninterpretable. Of the 4,978 evaluable children, 3,775 were born in Woburn and 1,203 became residents at some later age. The average period of Woburn residency for the evaluable children is 11.9 years, and 27% had some G and H exposure.

Table 4 lists the important risk factors for childhood disorders that were used as stratification variables in the analysis of G and H exposure. The higher rates of allergies in children from higher SES families (Sultz, Schlesinger, Mosher, and Feldman 1972), of kidney disorders in females (Smallpiece 1969), and of lung (Gordis 1973) and learning disabilities (Kinsbourne and Caplan 1979) in males have been noted previously by others.

Table 6 summarizes the analyses for association with cumulative G and H exposure. For each disorder, the table gives the number of cases, their median age at diagnosis, and the sum of their observed (X_i) and expected (E_i) cumulative G and H exposures at diagnosis. There is evidence

Table 6. Observed and Expected Cumulative Exposures for Children With Disorders

Disorder	Number of cases	Median age of diagnosis	Cumulative G and H exposure	
			Observed	Expected (standard error)
Allergies	476	7	276.5	268.8 (26.8)
Anemia	77	4	33.7	36.5 (10.8)
Diabetes	19	10	10.4	14.3 (6.8)
Heart/Blood pressure	21	13	2.0	14.3 (6.6)
Kidney/Urinary tract	43	5	31.9	18.5 (6.5)
Females	32	5	21.7	11.7 (5.1)
Males	11	8	10.2	6.8 (4.1)
Learning disability	218	8	122.7	135.3 (17.4)
Lung/Respiratory tract	192	3	83.9	63.1 (12.5)
Females, born 1960–1969	40	6	25.8	20.5 (8.2)
Females, born 1970–1982	36	2	7.4	4.3 (2.3)
Males, born 1960–1969	55	5	32.4	25.9 (8.1)
Males, born 1970–1982	61	2	18.3	12.4 (4.9)
Neurologic/Sensory	113	4	47.7	49.6 (11.1)
Other disorders	77	7	26.2	37.7 (10.8)

NOTE: Expected cumulative exposure and corresponding variance obtained by summing, over cases, averages and variances of cumulative exposures of children of the same age and in the same stratum as the case. Stratification factors are given in Table 4.

of a positive association between cumulative G and H exposure and the risk of kidney/urinary tract ($\hat{\alpha} = .30$, $P = .02$) and lung/respiratory disorders ($\hat{\alpha} = .15$, $P = .05$). The table also decomposes the statistics for lung and kidney disorders by the stratification variables used in their calculation. Higher-than-expected G and H exposure is evident within each stratum. Note also that if a two-sided test for association between G and H exposure and heart/blood pressure disorders had been used, we would have obtained $P = .06$, suggesting a possible negative association.

The 43 kidney/urinary tract disorders are primarily kidney or urinary tract infections. These children had a combined cumulative G and H exposure of 31.9, compared with an expectation of 18.5. The 192 children with lung/respiratory tract disorders (mostly asthma, chronic bronchitis, or pneumonia) had a cumulative G and H exposure of 83.9 versus an expectation of 63.1. The exposed cases did not appear to be confined or clustered in any subcategories of lung/respiratory or kidney/urinary tract disorders.

4.4 Rate Changes Since the Closure of Wells G and H

Table 7 indicates how the numbers of adverse health events in areas formerly serviced by wells G and H have changed since their closure in 1979. For the high exposure area consisting of Zones A–C, the table gives the frequency of the three adverse pregnancy outcomes for the 3-year periods immediately preceding (1977–1979) and following (1980–1982) the closing of wells G and H. There is a clear decline ($P = .02$) in the rates of these events after the wells were shut down. The table also gives the corresponding data for the remainder of Woburn (Zones D–E), where virtually none of the water came from wells G and H during 1977–1979. The rates for Zones A–C are higher ($P = .004$) than those for the rest of Woburn between 1977 and 1979, but comparable ($P = .34$) thereafter. This pattern is what we might expect if wells G and H had generated increased risk. It could also be a regression-to-the-mean phenomenon, though this is unlikely. Similar patterns were not evident with the other adverse pregnancy outcomes.

Table 7. Adverse Pregnancy Outcomes in Areas of High and Low Exposure Before (1977–1979) and After (1980–1982) Closure of Wells G and H

	1977–1979	1980–1982
Zones A–C (high exposure)		
Average G and H exposure	58%	0%
Number of pregnancies	107	78
Perinatal deaths	4	0
Eye/Ear anomalies	1	0
CNS/Chromosomal/Oral cleft	2	0
Zones D–E (low exposure)		
Average G and H exposure	.01%	0%
Number of pregnancies	436	397
Perinatal deaths	2	3
Eye/Ear anomalies	0	0
CNS/Chromosomal/Oral cleft	3	3

NOTE: Average G and H exposure refers to average exposure scores corresponding to pregnancies.

None of the three cases of childhood leukemia among children born since 1980 (Table 2) resided in Zones A–C. In addition, from the survey data, there have been no kidney disorders and only one lung disorder among children from Zones A–C born since 1980. There is still insufficient follow-up time, however, to determine whether these results reflect a decline in risk in Zones A–C since the closing of the wells.

5. ADDITIONAL ANALYSES, POTENTIAL BIASES

5.1 Baseline Rates in East and West Woburn

Recall that West Woburn never received any water from wells G and H. Even though the amount of water supplied by wells G and H to East Woburn varied markedly over time, the regional nature of the distribution raises the possibility that a difference in the baseline rates of adverse health events for East and West Woburn might have caused some of the positive statistical associations between wells G and H and these events.

We assessed this question by comparing the rates of adverse health effects in East and West Woburn among unexposed individuals. For the three adverse pregnancy outcomes, we compared the East and West Woburn rates for the years wells G and H were not operating (1960–1963, 1973, 1980–1982). For leukemia and childhood disorders, we used the stratified log-rank test (see Kalbfleisch and Prentice 1980), censoring each child's period of observation at the point of first exposure to wells G and H. Stratification was by year of birth for leukemia, by gender for kidney/urinary tract disorders, and by gender and year of birth (1960–1964, 1965–1969, 1970–1975, 1976–1982) for lung/respiratory tract disorders. Table 8 summarizes the results and indicates that the risks to unexposed individuals are similar in East and West Woburn. Thus it does not appear as though the positive associations with G and H exposure were caused by a difference in baseline rates between East and West Woburn. There were sufficient data to compare also the baseline rate of the adverse pregnancy outcomes and lung disorders in Zones A–C (easternmost Woburn) to the rest of East Woburn. No evidence of a difference was found.

The similarity of baseline rates between East and West Woburn implies that the elevated event rates among exposed individuals are also evident when focusing only on East Woburn. This can be seen for the adverse pregnancy outcomes by combining the zero exposure results for East Woburn with the nonzero exposure data in Table 5. Alternatively, Table 9 expands the results in Table 7 for Zones A–C to include the entire study period and indicates how the rates of perinatal deaths, eye/ear anomalies, and CNS/chromosomal/oral cleft anomalies are higher in years in which wells G and H were pumping than in years in which they were not pumping.

Table 8. Comparison of Baseline Rates in East and West Woburn

Adverse pregnancy outcomes in years 1960–1963, 1973, 1980–1982 (wells G and H were not pumping)			
	West Woburn rate × 10^3	*East Woburn rate × 10^3*	*Significance level*
Perinatal deaths	17.3 (18/1,042)	16.7 (10/599)	.55
Eye/Ear anomalies	4.4 (4/914)	2.0 (1/507)	.58
CNS/Chromosomal/Oral cleft anomalies	5.5 (5/914)	7.9 (4/507)	.41

Childhood disorders and leukemia in unexposed children				
		No. residing in East Woburn		
	Total no. of cases	*Observed*	*Expected (var.)*	*Significance level*
Kidney/Urinary tract	28	4	3.7 (3.1)	.33
Lung/Respiratory tract	136	26	23.3 (17.8)	.26
Leukemia	11	3	3.7 (2.3)	.68

NOTE: Significance levels for adverse pregnancy outcomes and childhood disorders based on one-sided exact tests and one-sided stratified log-rank tests, respectively. Fractions in parentheses are events/pregnancies for perinatal deaths and events/7-day survivors for congenital anomalies.

5.2 Potential Biases From the Sample Survey

Information about adverse pregnancy outcomes and childhood disorders was obtained from a sample survey. In the following sections we identify and discuss several features of the survey that could have biased the tests of association between G and H exposure and adverse health events.

5.2.1. Sampling Frame. The sampling frame for the survey was Woburn households with a listed telephone in 1982. The target population was pregnancies ending between 1960 and 1982 in women born since 1920. Families that moved from Woburn prior to 1982 are not in the sample. The omission of these families could distort the association between G and H exposure and an adverse health event if both (a) emigration rates varied with exposure to wells G and H and (b) the adjusted (for risk factors) event rates for families that moved from Woburn differed from those of families that remained. We checked the first condition by examining U.S. census records and found the transiency rates for East and West Woburn to be nearly identical. The second condition was more difficult to check because of limited published data and differences in the ways in which adverse health events are reported. The rates, however, of spontaneous abortion, low birth weight, and musculoskeletal and cardiovascular birth anomalies in the survey data are similar to those reported elsewhere (Heionen et al. 1977; National Center for Health Statistics 1972–1976, 1980; Shapiro et al. 1962).

Table 9. Rates (× 10^3) of Adverse Pregnancy Outcomes in Zones A–C by Pumping Status of Wells G and H

Adverse pregnancy outcomes	*Years wells G and H not pumping*	*Years wells G and H pumping*
Perinatal deaths, 1970–1982	0 (0/103)	16 (4/258)
Eye, ear anomalies	4 (1/228)	19 (8/414)
CNS, chromosomal, oral cleft anomalies	9 (2/228)	19 (8/414)

NOTE: Parentheses contain events per pregnancies for perinatal deaths and events per 7-day survivors for congenital anomalies.

5.2.2. Incompleted Calls and Refusals. Of the batches of 25 telephone numbers distributed to the participating interviewers, an average of 2 numbers were never called, 3 were called at least once but never completed, and 20 were completed. Phone numbers that were never distributed are, by the survey design, a random sample of the numbers in the directory. Since telephone numbers are noninformative with respect to location within Woburn, those that were distributed but never called are also a random sample. On the other hand, the households among the 975 numbers that were called but never completed might not be representative of those in the directory (e.g., more likely to be single-parent families). The proportions of telephone numbers that were not completed, as well as the average number of attempts to complete these, were virtually identical in East and West Woburn. It still is possible that the omission of these households caused bias, since the reasons they were not reached could have been different in East and West Woburn in ways related to the risks of adverse health events.

Among completed calls, the overall interview refusal rate was 18%, which is quite good for a telephone or mail survey (see Marks, Hogelin, Jones, Gentry, and Trowbridge 1983), and the rates for East (17.3%) and West (19.0%) Woburn were similar. The size of tests of association with G and H exposure, however, could still be distorted if the rates of adverse health events among refusing households in East Woburn differed from those in West Woburn (see Schlesselman 1982).

5.2.3. Interviewers and Respondents. Random errors by respondents in the reporting of adverse health events, or by interviewers in ascertaining and recording this information, cause a loss of sensitivity in tests for association between G and H exposure and adverse health events.

When error rates vary with exposure, the sizes of these tests can also be distorted. Of particular concern are higher rates of overreporting among exposed respondents and higher rates of underreporting among unexposed respondents, as these would increase the chances of a false-positive association with G and H exposure.

Recall that interviewers were not given the location of their assigned households and were instructed to keep the interview anonymous. Provided this anonymity was not compromised, the error rates of interviewers should not depend on the exposure status of the respondents. We checked the accuracy of interviewers in ascertaining information by recontacting certain respondents. We first recontacted 154 women who were reported to have a pregnancy loss to recheck information about this loss and about their smoking status during pregnancy. In only three of the callbacks was the initial information recorded by the interviewers not confirmed. We also attempted to recontact all households in which a perinatal death (post-1970), eye/ear, or CNS/chromosomal/oral cleft anomaly was reported, and the households corresponding to a random sample of 30 and 10 of the lung/respiratory and kidney/urinary disorders, respectively. We were able to reach the households for 96 of these 102 adverse health events. All 96 events were confirmed by the respondents, although in three instances the reported date of the disorder was different from that obtained in the original interview. These results suggest that the rates of overreporting events by interviewers are negligible and not dependent on the exposure status of the respondent.

We also checked whether the survey results depended on the residence (East Woburn, West Woburn, outside Woburn) of the volunteer interviewer to which the survey forms were assigned. Of the nine categories of childhood disorders and eight categories of adverse pregnancy outcomes, the only evidence of a difference in rates was with CNS/chromosomal/oral cleft anomalies, where the results for interviewers from outside Woburn were lower than those for interviewers from East or West Woburn. In addition, since six interviewers were from the families of the 20 leukemia cases, and some of these are involved in litigation concerning the contaminated water, we checked to see if their survey results differed from those of the other interviewers. There was no evidence of a difference, and virtually identical study results would have been obtained if the data from these six interviewers were omitted.

To check the accuracy of respondents in reporting health information, we attempted to obtain medical verification for the 96 confirmed disorders discussed previously. For 14 of these, the parents either refused to sign and witness the required consent form that we sent them or they never returned it to us. Of the remaining 82 events for which consent was obtained, the physicians for 16 could not be located, did not respond to our request for verification, or were not the individual's attending physician at the time of the event. Thus we obtained responses for 66 events, of which 62 were verified (37 unexposed, 25 exposed), 2 were equivocal (1 unexposed, 1 exposed), and 2 were not verified (both unexposed). The nonverified conditions were a perinatal death (to a woman diagnosed as having a false pregnancy) and a kidney infection (symptoms but negative culture). These small false-positive rates for unexposed (2/37) and exposed (0/25) respondents suggest that overreporting is infrequent and not more common among exposed respondents.

Distortions to the size of tests for association with wells G and H can also be caused by differential underreporting of adverse health events by exposed and unexposed respondents. It is well known that the ability to recall certain adverse health events diminishes with time, especially with less serious conditions. If respondents from East Woburn had greater recall than those from West Woburn, perhaps due to anxiety from a perception that they were exposed to wells G and H, the probability of a false-positive association with water from wells G and H would be increased. We did not have the resources to check for this potential bias directly by examining the medical records of individuals who did not report an adverse health effect. Nor did we ask about perception of exposure to wells G and H in the survey. The data provide indirect evidence, however, that differential recall did not play an important role in the positive associations that were found. Specifically, if perception of G and H exposure among East Woburn respondents caused them to have greater recall than respondents from West Woburn, we would expect unexposed individuals from East Woburn to have higher event rates than those from West Woburn. This was not evident, however, with any of the disorders that were associated with wells G and H (Sec. 5.1), nor with any of the other adverse pregnancy outcomes. In addition, differential recall between respondents of East and West Woburn would not explain the results in Table 9, where rates of adverse health events in East Woburn depended on G and H exposure status.

5.3 Inexact Measurement of G and H Exposure

Since the monthly amounts of water pumped by each of Woburn's municipal wells are routinely recorded, the proportion of the town's total water supply arising from wells G and H is known. The estimation of this proportion for individual residences, however, required a number of assumptions and approximations and did not account for water consumed by individuals at locations other than the residence (e.g., at school). It is, therefore, important to consider how errors in the estimation of G and H exposure might have affected the analyses of association with adverse health effects.

The DEQE estimates of G and H exposure did not depend on the locations of any adverse health events. Thus if there were no association between actual G and H exposure and the risk of some disorder, a positive association between estimated exposure and risk would necessarily be a chance occurrence; that is, errors in the measurement of exposure do not distort the size of tests. They do, however, reduce the sensitivity of these tests.

To determine whether our results depended critically on the specific exposure estimates that were used (i.e., those in Table 1), we repeated the analyses of adverse pregnancy

outcomes using two coarser measures of G and H exposure. First, keeping the same zones as before, we replaced each original exposure score by 0, 1, 2, or 3 corresponding to the intervals used in Table 5. Second, we partitioned Woburn into two zones based on the city engineer's 1980 determination of the area of coverage of wells G and H under average pumping conditions (Parker and Rosen 1981). We then assigned a yearly score of 0 to the region of Woburn that was estimated to have received no water from wells G and H, and for the other region, a yearly score equal to the percentage of all town water supplied by wells G and H. Both analyses resulted in the same significant associations as the original analyses.

6. DISCUSSION

The motivation for this study was the MDPH's finding of a high rate of childhood leukemia in Woburn in the wake of the discovery of contaminated drinking water and their inability to assess the relationship between these events because of inadequate G and H exposure information. There is little doubt that leukemia rates in Woburn are high. Between 1964 and 1983, 20 cases were diagnosed when only 9.1 were expected. Our objective was not to compare these rates with those elsewhere but to determine whether the space–time distribution within Woburn of the 20 cases was correlated with access to water from wells G and H. Using a test derived from Cox's (1972) regression model, we found that both a "cumulative" and "none–some" metric of G and H exposure were positively associated with leukemia rates.

It does not appear that Woburn's entire leukemia excess, based on national rates, is statistically explainable by wells G and H. The cumulative and none–some metrics statistically explain about four and six cases, which still leaves an excess of five or six cases. This difference is partially due to recent results in West Woburn, which is estimated never to have received any water from wells G and H. Here four cases have been diagnosed since 1980, compared with only four in the preceding 16 years (based on national rates, only .9 cases were expected since 1980). Thus even if wells G and H were responsible for some leukemia cases, it does not appear that all of Woburn's problems have ended with their closure.

The sample survey collected information on 4,396 pregnancies, and in 16% of these there was some exposure to wells G and H. For spontaneous abortions and low birth weight, there was a relatively large number of events (520 and 220, respectively) and no evidence of an increasing risk with increasing G and H exposure. For perinatal deaths, there was a positive association [$\exp(\hat{\alpha}) = 10.0$, $P = .003$] with G and H exposure after 1970, primarily due to 3 stillbirths occurring among the 39 pregnancies with very high (over 90%) G and H exposure. It is not clear why an association was evident after 1970, but not earlier. If contaminants in the water did elevate the risk of a perinatal death, these may have been present only in recent years or may have elevated risk only at higher exposure levels.

The analysis of congenital anomalies in this or any study is complicated by the many different types of anomalies that can occur and by the infrequency of each type (see Table 3). If anomalies are grouped too finely in analyses, there are problems of insensitivity due to the small numbers of events per category and problems of interpretation due to the multiplicity of analyses that must be carried out. The use of broader categories reduces these problems but has less interpretative value, since a factor that is associated with a category need not be associated with each type of anomaly in that category. Moreover, associations confined to only a few types of anomalies in the category could go unnoticed. For example, Saxen (1969) noted that thalidomide would not have been detected as a cause of anomalies of the extremities if all birth anomalies had been grouped into a single category.

Our grouping of anomalies into five categories, including two (CNS/chromosomal/oral cleft and "other") with different types of anomalies, was an attempt to strike a balance between these two extremes and to use prior knowledge about which anomalies might be more susceptible to environmental insults. We found no evidence of an association between G and H exposure and the categories of musculoskeletal, cardiovascular, and "other" anomalies, but found positive associations with eye/ear [$\exp(\hat{\alpha}) = 14.9$, $P < .0001$] and CNS/chromosomal/oral cleft [$\exp(\hat{\alpha}) = 4.5$, $P = .01$] anomalies. For the latter, exposed women had higher rates of CNS and chromosomal defects, but not of oral cleft defects. Of the nine chromosomal anomalies, seven were Down's syndrome, including the three with nonzero (.54, .95, .95) G and H exposure scores. This is noteworthy because of the association between Down's syndrome and leukemia in sibships (Miller 1963) and a previously reported double population cluster of leukemia and Down's syndrome (Dowsett 1966).

The survey also collected health information about 4,978 children, of whom 27% had some exposure to wells G and H. Five categories of childhood disorders (allergies, anemia, learning disabilities, neurologic, and other) gave no indication of an association with wells G and H and involved a sufficiently large number of events to make us confident that a moderate or large effect is unlikely to have gone undetected. There were far fewer events in the diabetes (19) and heart/blood pressure (21) categories, although neither gave any indication of a positive association. On the other hand, positive associations were found between cumulative G and H exposure and lung/respiratory tract and kidney/urinary tract disorders. We are not aware of any reported associations between these two types of disorders and the contaminants found in wells G and H. The plausibility of such an effect, however, cannot be ruled out because, once blood-borne, contaminants can be circulated throughout the body. Indeed, both of these types of disorders have been associated with other contaminants (see CDC 1984; Ehrenreich 1977; Nakanishi et al. 1985).

Because this was an observational study, some or all of the six positive associations with wells G and H could be purely statistical, having occurred because the true causal factor (confounder) was not controlled for and has a distribution that is correlated with that of water from wells G and H. U.S. Census data show East and West Woburn to be very similar with respect to ethnicity, socioeconomic status, and job types (Bureau of the Census 1960, 1970,

1980), and hence it does not seem likely that these would be confounders. In addition, the MDPH leukemia study examined information on family and occupational history, previous medical treatment, church membership, and certain environmental factors, and none of these distinguished their cases and controls. More generally, the similarity of baseline rates in East and West Woburn and the higher rates within East Woburn for exposed individuals suggest that a potential confounding factor would not be a purely "regional" variable, but one whose distribution also varied over time similar to the distribution of water from wells G and H.

A difficulty in trying to determine whether wells G and H caused any adverse health events is the lack of knowledge about the contaminants in wells G and H and the effects of these contaminants on human health. Several of the chemicals detected in 1979 have been shown to cause cancer in laboratory animals (National Cancer Institute 1976, 1977; Page and Safiotti 1976) when given in high concentrations, and others found in the surrounding groundwater are suspected of causing cancer in humans (see Davies 1978; Fraumeni 1975; White, Infante, and Chu 1984). The effects on humans, however, of exposure to these chemicals at lower concentrations and in combination are not well understood. Furthermore, the types and levels of contaminants in wells G and H prior to 1979 are not known. Thus wells G and H might have become contaminated shortly before they were tested in 1979, or might have been contaminated for many years, at much higher levels and with chemicals other than those found in 1979. The EPA is currently conducting geological studies to determine the sources of the contaminants found in wells G and H and the surrounding groundwater. These studies might make it possible to predict the types and levels of contamination in earlier years. Without this information, it is unlikely that the role of wells G and H as a cause of adverse health outcomes will ever be fully understood.

A key question among citizens is whether individuals who were formerly exposed to wells G and H might still be at elevated risk. One way this can be investigated is by monitoring adverse health outcomes in Woburn over the next few years. In the 3 years following the closure of wells G and H, the rates of perinatal deaths, eye/ear, and CNS/chromosomal/oral cleft anomalies in Zones A–C have declined to levels comparable with the rest of Woburn (Table 7). If these rates, as well as those of leukemia, lung, and kidney disorders, continue to remain low in these areas, individuals who were formerly exposed to wells G and H will have reason to believe that any earlier elevation in risk that might have been due to wells G and H is no longer present.

APPENDIX: ESTIMATES OF E_I AND V_I FOR LEUKEMIA SCORE TEST OF $\alpha = 0$ BASED ON REGIONAL DISTRIBUTION OF THE POPULATION

Suppose that the town is partitioned into R exposure regions and that $e(r, u)$ denotes the annual G and H exposure score for individuals who reside in region r in year u. Assume that a child can live in at most one of the R regions of the town and that the age-specific transition rates into (i.e., birth and immigration) and out of (i.e., death and emigration) the town do not vary by region.

Let $\delta(u)$ equal 1 if a child lived in the town in year u, and equal 0 otherwise, and let $\boldsymbol{\Delta}$ be the vector whose rth component is 1 if the child lived in region r, and 0 otherwise. Then under the null hypothesis, $\boldsymbol{\Delta}$ has the multinomial distribution with parameter $\boldsymbol{\Pi} = (\pi_1, \pi_2, \ldots, \pi_R)^T$, where π_r is the fraction of children in the town that live in region r.

The vector of possible cumulative exposure scores for the child is given by

$$\mathbf{e} = \sum_u \delta(u)\mathbf{e}(u),$$

where $\mathbf{e}(u) = (e(1, u), e(2, u), \ldots, e(R, u))^T$. The observed cumulative exposure score is $X = \mathbf{e}^T\boldsymbol{\Delta}$. Conditional on the $\delta(u)$, the mean and variance of X under the null hypothesis are, therefore,

$$\mathbf{e}^T\boldsymbol{\Pi} = \sum_u \delta(u)\mathbf{e}(u)^T\boldsymbol{\Pi} \tag{A.1}$$

and

$$\mathbf{e}^T[\text{diag}(\boldsymbol{\Pi}) - \boldsymbol{\Pi}\boldsymbol{\Pi}^T]\mathbf{e} = \sum_{r=1}^{R}\left\{\sum_u e(r, u)\delta(u)\right\}^2 - (\mathbf{e}^T\boldsymbol{\Pi})^2. \tag{A.2}$$

Modifications of this formula for other exposure metrics are straightforward. In addition, it can be shown that the right side of (A.2) is greater than the variance of X when children can move between different regions of the town. In these circumstances, a test based on (A.2) will be conservative.

[*Received July 1984. Revised January 1986.*]

REFERENCES

Aitkin, M., Laird, N., and Francis, B. (1983), "A Reanalysis of the Stanford Heart Transplant Data," *Journal of the American Statistical Association,* 78, 264–292.

Bennett, J. W., and Aborns, K. I. (1979), "Gametogenesis and Incidence of Down's Syndrome," *Lancet,* 2, 913.

Breslow, N. E., Lubin, J. H., Marek, P., and Langholtz, B. (1983), "Multiplicative Models Cohort Analysis," *Journal of the American Statistical Association,* 78, 1–12.

Bureau of the Census, U.S. Department of Commerce (1960, 1970, 1980), *Massachusetts: Census of Population and Housing* (Census Tract Reports), Washington, DC: U.S. Government Printing Office.

Butler, N. R., Goldstein, H., Ross, E. M. (1972), "Cigarette Smoking in Pregnancy: Its Influence on Birthweight and Perinatal Mortality," *British Medical Journal,* 2, 127–130.

Centers for Disease Control (1981), "Annual Summary, 1981: Reported Morbidity and Mortality in the United States," *Morbidity and Mortality Weekly Report,* 30, No. 54, 107–109.

——— (1984), "Results of a Pilot Study of Health Effects Due to 2,3,7,8-Tetrachlorodibenzodioxin Contamination—Missouri," *Morbidity and Mortality Weekly Report,* 33, No. 5, 54–61.

Christianson, R. E. (1980), "The Relationship Between Maternal Smoking and the Incidence of Congenital Anomalies," *American Journal of Epidemiology,* 112, 684–695.

Cox, D. R. (1970), *The Analysis of Binary Data,* London: Methuen.

——— (1972), "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society,* Ser. B, 34, 269–276.

Cox, D. R., and Oakes, D. (1984), *Analysis of Survival Data,* London: Chapman & Hall.

Crowley, J., and Hu, M. (1977), "Covariance Analysis of Heart Transplant Survival Data," *Journal of the American Statistical Association,* 72, 27–36.

Davies, J. M. (1978), "Lung Cancer Mortality of Workers Making Chrome Pigments," *Lancet,* 1, 384.

Dorsch, M. M., Scragg, R. K. R., McMichael, A. J., Baghurst, P. A., and Dyer, K. F. (1984), "Congenital Malformations and Maternal Drinking Water Supply in Rural South Australia: A Case-Control Study," *American Journal of Epidemiology,* 119, No. 4, 473–486.

Dowsett, E. G. (1966), "Leukemia in Kingston, Surrey, 1955–1964: An Epidemiological Study," *British Journal of Cancer,* 20, 16–31.

Ecology and Environment, Inc. (1982), *Evaluation of the Hydrogeology and Groundwater Quality of East and North Woburn, MA* (Vol. 1, Final Report).

Ehrenreich, T. (1977), "Renal Disease From Exposure to Solvents," *Annals of Clinical Laboratory Science,* 7, No. 1, 6–16.

Fraumeni, J. F. (1975), "Respiratory Carcinogenesis: An Epidemiologic Appraisal," *Journal of the National Cancer Institute,* 55, 1039.

Gordis, L. (1973), *Epidemiology and Chronic Lung Disease in Children,* Baltimore: John Hopkins University Press.

Gordon, J. E., and Shy, C. M. (1981), "Agricultural Chemical Use and Congenital Cleft Lip and/or Palate," *Archives of Environmental Health,* 36, 213–220.

Heionen, O. P., Slone, D., and Shapiro, S. (1977), *Birth Defects and Drugs in Pregnancy,* Littleton, MA: Publishing Sciences Group.

Holmberg, P. C. (1979), "Central Nervous System Defects in Children Born to Mothers Exposed to Organic Solvents During Pregnancy," *Lancet,* 2, 177–179.

Kalbfleisch, J. D., and Prentice, R. L. (1980), *The Statistical Analysis of Failure Time Data,* New York: John Wiley.

Kinsbourne, M., and Caplan, P. J. (1979), *Children's Learning and Attention Problems,* Boston: Little, Brown.

Kline, J. (1977), "An Epidemiological Review of the Role of Gravidity in Spontaneous Abortion," *Early Human Development,* 1, 337–344.

Kline, J., Stein, Z. A., Susser, M. W., and Warburton, D. (1977), "Smoking: A Risk Factor for Spontaneous Abortion," *New England Journal of Medicine,* 297, 793–796.

Klingberg, M. A., Papier, C. M., and Hart, J. (1983), "Birth Defects Monitoring," *American Journal of Industrial Medicine,* 4, 309–328.

Kotelchuck, M., and Parker, G. (1979), *Woburn Health Data Analysis, 1969–1978,* Massachusetts Department of Health.

Lagakos, S. W., Wessen, B. W., and Zelen, M. (1985), "Contaminated Well Water and Health Effects in Woburn, Massachusetts: Addendum," technical report, Harvard School of Public Health, Dept. of Biostatistics.

Liddell, F. D. K., McDonald, J. C., and Thomas, D. C. (1977), "Methods of Cohort Analysis: Appraisal by Application to Asbestos Mining" (with discussion), *Journal of the Royal Statistical Society,* Ser. A, 140, 469–491.

Mantel, N., and Byar, D. P. (1974), "Evaluation of Response Time Data Involving Transient States: An Illustration Using Heart Transplant Data," *Journal of the American Statistical Association,* 69, 81–86.

Marks, J. S., Hogelin, G. C., Jones, J. T., Gentry, E. M., and Trowbridge, F. L. (1983), "Telephone Factor Prevalence Surveys," in *Proceedings of the 19th National Meeting of the Public Conference on Records and Statistics,* Hyattsville, MD: National Center for Health Statistics, pp. 93–95.

Massachusetts Department of Environmental Quality and Engineering (1979), Special Analysis: Woburn, water samples taken from wells G and H by McCall, DEQE, May 14.

Miller, R. W. (1963), "Down's Syndrome (Mongolism), Other Congenital Malformations and Cancer Among Sibs of Leukemic Children," *New England Journal of Medicine,* 268, No. 8, 393–401.

Nakanishi, Y., Shigematsu, N., Kurita, Y., Matsuta, K., Kanagae, H., Ishimaru, S., and Kawazoe, Y. (1985), "Respiratory Involvement and Immune Status in Yusho Patients," *Environmental Health Perspectives,* 59, 31–36.

National Cancer Institute (1976), *Carcinogenesis Bioassay of Trichloroethylene,* CAS No. 79-01-6, NCI-CG-TR-2.

——— (1977), *Bioassay of Tetrachloroethylene for Possible Carcinogenesis,* DHEW Publication (NIH)77-813.

National Center for Health Statistics (1972–1976), *Vital Statistics of the U.S., Volume II—Mortality, 1972–1976,* Rockville, MD: U.S. Department of Health, Education and Welfare, Public Health Service.

National Center for Health Statistics (1980), "Factors Associated With Low Birthweight," *Vital and Health Statistics* (Ser. 21, No. 37), Hyattsville, MD: U.S. Department of Health, Education and Welfare, Public Health Service.

National Research Council (1985), "Assigned Share for Radiation as a Cause of Cancer," Report of the Oversight Committee for Radioepidemiologic Tables, Washington, DC: National Academy Press.

Naylor, A. F., and Warburton, D. (1979), "Sequential Analysis of Spontaneous Abortion II: Collaborative Study Data Show That Gravidity Determines a Very Substantial Rise in Risk," *Fertility and Sterility,* 31, 282–286.

Niswander, K., Gordon, M., and Berendes, H. (1972), *The Women and Their Pregnancies,* Philadelphia: W. B. Saunders.

Page, N. P., and Safiotti, U. (1976), "Report on Carcinogenesis Bioassay of Chloroform," Bethesda, MD: National Cancer Institute.

Parker, G. S., and Rosen, S. L. (1981), *Cancer Incidence and Environmental Hazards 1960–1978,* Massachusetts Department of Public Health.

Poladnek, A. P., and Janerich, D. T. (1983), "Uses of Available Record Systems in Epidemiologic Studies of Reproductive Toxicology," *American Journal of Industrial Medicine,* 4, 205–244.

Prentice, R. L. (1985), "A Case-Cohort Design for Epidemiologic Cohort Studies and Disease Prevention Trials," *Biometrika,* 73, 1–11.

Rantakallio, P., Krause, U., and Krause, K. (1979), "The Use of Ophthalmological Services During the Preschool Age, Ocular Findings and Family Background," *Journal of Pediatric Ophthalmology and Strabismus,* 15, 253–258.

Reed, D. M., and Stanley, F. J. (1977), *The Epidemiology of Prematurity,* Baltimore: Urban and Schwartzenberg.

Saxen, L. (1969), *Congenital Defects,* New York: Holt, Rinehart & Winston, pp. 10–11.

Schlesselman, J. J. (1982), *Case-Control Studies,* New York: Oxford University Press.

Shapiro, S., Jones, E. W., and Densen, P. M. (1962), "A Life-Table of Pregnancy Termination and Correlates of Fetal Loss," *Millbank Memorial Fund Quarterly,* 40, 1–27.

Smallpiece, V. (1969), *Urinary Tract Infection in Childhood and Its Relevance to Disease in Adult Life,* St. Louis: C. V. Mosby.

Sultz, H. A., Schlesinger, E. R., Mosher, W. E., and Feldman, J. G. (1972), *Longterm Childhood Illness,* Pittsburgh: University of Pittsburgh Press.

Surveillance, Epidemiology and End-Results (SEER) (1981), Incidence and Mortality Data, 1973–1977, National Institute of Public Health Publication 81–2330, Washington, DC: U.S. Government Printing Office.

Taffel, S. (1976), *Factors Associated With Low Birthweight, United States,* Washington, DC: National Center for Health Statistics, Ser. 21, No. 37.

U.S. Department of Health, Education and Welfare (1979), *Smoking and Health: A Report of the Surgeon General,* Publication 79-50066, Washington, DC: Author.

U.S. Environmental Protection Agency (1979), "Analysis for Volatile Organics on Woburn P.W.S. Wells G&H," Results contained in a memorandum from T. Siscanaw, EPA, to E. Taylor, EPA, July 24.

Waldorf, H., and Cleary, R. (1983), *Water Distribution System, Woburn Massachusetts, 1964–1979,* Engineering Draft Report, Massachusetts Department of Environmental Quality.

White, M. C., Infante, P. F., and Chu, K. C. (1984), "A Quantitative Estimation of Leukemia Mortality Associated With Occupational Exposure to Benzene," *Risk Analysis,* 2, 195–204.