

ADJUSTING MORBIDITY RATIOS IN TWO COMMUNITIES USING RISK FACTOR PREVALENCE IN CASES¹

ANDREW G. DEAN,^{2,3} HARRIET H. IMREY,⁴ KARI DUSICH,³ AND WILLIAM N. HALL⁵

Dean, A. G. (CDC, Atlanta, GA 30333), H. H. Imrey, K. Dusich, and W. N. Hall. Adjusting morbidity ratios in two communities using risk factor prevalence in cases. *Am J Epidemiol* 1988;127:654-62.

Available data on cancer incidence for 1969-1971 showed statistically elevated rates for breast cancer in St. Louis Park, Minnesota, a community with creosote contamination of the water supply, when compared with the rest of the Minneapolis-St. Paul area taken as the reference population. In order to assess the effect of other known risk factors for breast cancer, 75 persons with breast cancer in each of the two populations were interviewed to obtain frequencies of known risk factors. An adjusted morbidity ratio in the two populations and an expected case rate in the exposed community were calculated from these frequencies, using relative risk values from the medical literature. The adjusted morbidity ratio was less than 1.0, and the observed rate was almost identical to the new expectation, although the age-adjusted rates alone had suggested a significant difference in incidence. This method makes use of relative risks from published studies rather than those associated with local cases and controls. It allows more refined evaluation of differences in cancer rates between communities than can be provided by age- and sex-specific calculations alone, and may allow use of available statistics in situations where cost, temporal considerations, or population size do not favor large new studies.

breast neoplasms; environmental exposure; epidemiologic methods; polycyclic hydrocarbons; statistics

The first step in evaluating the effect of environmental contamination or other geographically related risk factors is often a comparison of disease rates in communities with and without the putative risk factor—

the so-called "ecologic study". Information on disease is obtained from cancer registries, hospital discharge records, or other sources; denominators are obtained from census reports; and age- and sex-specific rates are calculated to perform the comparison. Disease rates are influenced by many risk factors other than the one being considered, however, and community-wide prevalence of risk factors other than age, sex, and race is often not available. A method of adjusting for such risk factors, using data from case interviews alone, is illustrated by its use in a community with a contaminated water supply.

St. Louis Park, Minnesota is a community in which creosote contamination of the water supply and an elevated breast cancer

Received for publication June 30, 1986, and in final form April 28, 1987.

¹ From the Division of Disease Prevention and Control, Minnesota Department of Health, Minneapolis, MN.

² Division of Surveillance and Epidemiologic Studies, Epidemiology Program Office, Bldg. 1, Room 5127, Centers for Disease Control, Atlanta, GA 30333.

³ (Formerly) Division of Disease Prevention and Control, Minnesota Department of Health, Minneapolis, MN.

⁴ Regional Health Resource Center, Urbana, IL.

⁵ (Formerly) Centers for Disease Control, Atlanta, GA.

rate had raised questions about a causal association (1). Polynuclear aromatic hydrocarbons were first detected in the water supply of the city in November 1978, but might have been present for decades. Because data on cancer incidence for the three-year period 1969-1971 were available, rates for St. Louis Park were compared with those for the surrounding Minneapolis-St. Paul metropolitan area, which had a different water supply. The results showed statistically significantly elevated rates for breast cancer, although the rate in St. Louis Park was only 1.45-fold that for the metropolitan area.

Many polynuclear aromatic hydrocarbon compounds are carcinogenic for laboratory animals by a variety of routes of administration, and creosote and coal tar are known to produce human skin cancer (1). Cancer rates in human populations after oral ingestion of polynuclear aromatic hydrocarbons have been much less thoroughly studied (1). Several polynuclear aromatic hydrocarbon compounds such as 3-methylcholanthrene produce mammary cancer when fed to rats; almost all such rats that develop cancer are females (2-4). The parallel between St. Louis Park cancer patterns and laboratory data made it impossible to rule out a connection between the water contamination and the elevated rates of breast cancer. Press coverage and public discussion led to a great deal of concern in the community over the issue of "cancer in the water".

The temporal sequence of the polynuclear aromatic hydrocarbon contamination in St. Louis Park is not known precisely. A plant that used creosote and coal tar operated there in the period 1917-1972, and a 100 acre (40.5 hectare) plot of creosote-containing soil is the apparent source of the contamination of the municipal water supply. The Third National Cancer Survey included St. Louis Park, the Twin Cities of Minneapolis and St. Paul, and the surrounding five-county area, for the three-year period 1969-1971. Since data on cancer incidence for other years had not been

collected at any central point, our study was confined to these three years. Computer tapes from the Third National Cancer Survey were used to calculate age-adjusted rates for 45 body sites and types of cancer for male and female cancer cases who resided in St. Louis Park or elsewhere in the Minneapolis-St. Paul metropolitan area.

Comparisons of the 90 pairs of rates, age-adjusted to the metropolitan area population, showed statistically significant differences for breast cancer in women (1). The average annual age-adjusted rates for breast cancer per 100,000 white females were 113 in St. Louis Park and 78 in the metropolitan area population. Mantel-Haenszel chi-square tests (5) gave p values of <0.0005 for the difference. Two other suburbs of higher socioeconomic level and 45 artificial subdivisions of the metropolitan area all had lower rates for breast cancer than St. Louis Park. The difference between the 95 cases observed and the 65 cases expected amounted to 10 excess cases per year among 25,000 women. The only other "significant" difference in rates between the two communities was for gastrointestinal cancer in women ($p = 0.05$). Since one or more such observations would be expected in making 90 comparisons, and since breast cancer and colon cancer share several common risk factors, the remainder of the study focused on the rates of breast cancer.

It seemed likely that differences in known risk factors for breast cancer in the two communities would explain the findings (6), but information on major factors such as age at first childbirth, family history, menarche, and menopause was not available either for persons with breast cancer or for the general population of either community. Classical case-control or cohort studies could have been conducted, but only at considerable expense. Ascertaining risk factor prevalence in two mobile and heterogeneous communities 10 years earlier would have been a difficult task.

The medical literature contains a large

number of studies which compare rates for breast cancer in populations of women with and without specific risk factors, and relative risk values are quite well defined (6). The frequencies of known risk factors in cases for the Minneapolis-St. Paul metropolitan area, and for St. Louis Park cases were determined through an interview study. The risk factor frequencies for cases were combined with relative risks from the literature to produce an adjusted morbidity ratio and an estimate of the "expected" rate for residents of St. Louis Park.

MATERIALS AND METHODS

Attempts were made to locate all 95 female residents of St. Louis Park included in the Third National Cancer Survey files with a diagnosis of breast cancer made in 1969-1971. With the permission of the attending physician, his or her successor, or the chief of service of the hospital, the patient or closest surviving relative was contacted by mail and telephone, and an interview was arranged with either of two interviewers. For each of the 95 cases in St. Louis Park, a breast cancer case in the remainder of the five-county metropolitan area was selected. Random numbers were used to choose comparison cases from the several thousand persons in the Third National Cancer Survey files so that the number of interviews in each five-year age group was the same in the two communities. Sixty-three per cent of the metropolitan area interviews and 67 per cent of the St. Louis Park interviews were conducted face-to-face; the others were done by telephone. The proportion of patients who had been identified in the Third National Cancer Survey and who were still alive in the two groups at the time of our study was 44 per cent in the metropolitan area and 51 per cent in St. Louis Park. The interviews averaged 30 and 31 minutes each, respectively, for the two groups.

Two metropolitan area cases were excluded because of prolonged previous residence in St. Louis Park. One metropolitan

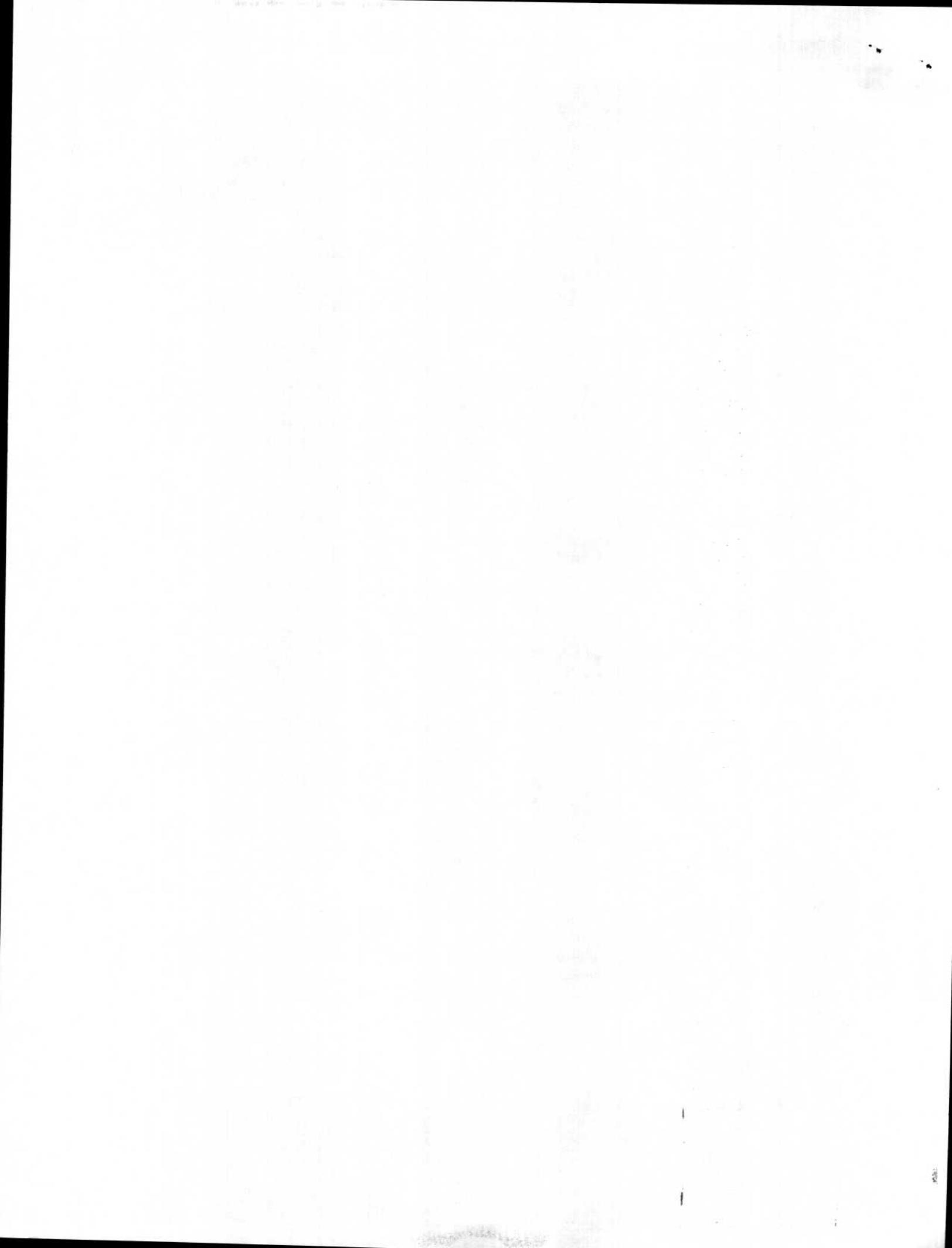
area patient who had resided in St. Louis Park for two years was allowed to remain. If a St. Louis Park patient could not be interviewed, the age-matched metropolitan area patient was removed from the study. Interviews were conducted in groups to allow those for both communities to be completed at nearly the same time. The final distribution of interviews was such that one interviewer did two more St. Louis Park cases than metropolitan area cases, and the other the reverse. A total of 75 cancer patients in St. Louis Park were located and interviewed; these and 75 matched metropolitan area controls comprised the study groups.

Because of time, money, and personnel constraints and the 10 years which had elapsed since the study period, a traditional case-control study to measure the residual difference in rates after accounting for known risk factors was out of the question. Instead, we conducted a literature review of other breast cancer studies and selected the relative risks reported in Helmrich et al. (7) for standardizing the two populations. The relative risks in this study were derived from multivariate analysis of a large multinational study and were therefore presumed to be more stable than those from smaller studies. In addition, this study included all of the variables which had been collected for the current study.

The calculation of the adjusted morbidity ratio derives from the formula for population attributable risk proportion (PARP) or etiologic fraction (5, p. 163), which can be stated as

$$\text{PARP} = 1 - \frac{1}{\sum_{i=0}^k p_i \text{RR}_i}, \quad (1)$$

where p_i is the proportion of the total population in risk stratum i , RR_i is the relative risk in that stratum, and k is the number of strata. Bruzzi et al. (8) recently reported an analogous formula for calculating population attributable risk proportion using



prevalence data from cases only. If X represents the number of cases, then

$$\text{PARP} = 1 - \frac{1}{X} \sum_{i=1}^X \frac{1}{\text{RR}_i} \quad (2)$$

In this formula, only case information is needed; prevalence of risk factors in the entire population is not required. The adjusted morbidity ratio is the baseline case rate in the exposed community (St. Louis Park) divided by the baseline case rate in the control community (the metropolitan area), i.e., the ratio of cases which would have occurred if every woman in each community had reference levels ($\text{RR} = 1.0$) of all risk factors known to be associated with breast cancer and included in the study. The proportion of disease not due to the defined risk factors may be called the non-attributable risk proportion (NARP). Since

$$\text{PARP} + \text{NARP} = 1 \quad (3)$$

then

$$\text{NARP} = \frac{1}{X} \sum_{i=1}^X \frac{1}{\text{RR}_i} \quad (4)$$

from equation 2. Note that the non-attributable risk is the reciprocal of the harmonic mean of the relative risks for each case in the population.

The formula for the adjusted morbidity ratio (AMR) is the ratio of the non-attributable risk proportion in one community to that in the other:

$$\text{AMR} = \frac{I_1 \frac{1}{X_1} \sum_{i=1}^X \frac{1}{\text{RR}_i}}{I_0 \frac{1}{X_0} \sum_{i=1}^X \frac{1}{\text{RR}_i}} \quad (5)$$

where I_1 is the incidence rate in the exposed community, St. Louis Park, and I_0 is the incidence rate in the control community (metropolitan area). The composite relative risk for each case is the product of the relative risks for each factor in the model. The relative risks reported by Helmrich et al. (7) are univariate, but are reported not to be substantially different from the mul-

tivariate relative risks, which suggests that confounding is not a problem. We therefore obtained the relative risk for each case as the product of the relative risks for the case's individual risk factors. No interactions were reported; if they had been, the adjusted mortality ratio formula could accommodate the relative risks for effect modification.

RESULTS

Breast cancer cases in St. Louis Park had a higher prevalence of risk factors than did cases in the metropolitan area. The non-attributable risk proportion (the proportion of the case rate which would have occurred in the absence of the risk factors accounted for) was 0.402 in St. Louis Park and 0.689 in the metropolitan area. The adjusted morbidity ratio, from equation 4, was therefore

$$\text{AMR} = \frac{113 \times 0.402}{78 \times 0.689} = 0.85.$$

The unadjusted morbidity ratio was 1.45.

The expected case rate in St. Louis Park, under the null hypothesis that the adjusted morbidity ratio = 1.0 (i.e., the distribution of risk factors among cases is identical in the two populations) is

$$EI_1 = 78 \frac{0.689}{0.402} = 134.$$

The observed case rate was 113 per 100,000 white women.

The relative risk for Jewish origin in the study by Helmrich et al. (7) was 2.8, which is somewhat higher than the relative risks of 1.3-1.6 reported elsewhere (9-11). Since 19 St. Louis Park patients and only two metropolitan area patients were Jewish, we repeated the analysis, ignoring religion, and calculated an adjusted morbidity ratio of 0.99, and an expected rate of 115—almost identical to the observed rate of 113.

The adjusted morbidity ratio of 0.85 suggests that the observed breast cancer incidence rate in St. Louis Park is, if anything,

TABLE 1
Risk factor prevalence in breast cancer cases, St. Louis Park and remaining Minneapolis-St. Paul metropolitan area, 1969-1971

| Risk factor | Relative risk assigned* | Frequencies among 75 cases in: | |
|---------------------------------------------------|-------------------------|--------------------------------|-------------------|
| | | St. Louis Park | Metropolitan area |
| 1. Age (years) at menarche | | | |
| Premenopausal | | | |
| ≥15 | 1.0 | 4 | 1 |
| <15 | 2.0 | 17 | 18 |
| Postmenopausal | | | |
| ≥15 | 1.0 | 6 | 5 |
| <15 | 1.1 | 43 | 38 |
| Unknown | 1.0 | 5 | 13 |
| Mean relative risk | | 1.30 | 1.10 |
| 2. Age (years) at first birth | | | |
| Nulliparous | | | |
| <20 | 1.0 | 14 | 15 |
| Parity 1-2 | 0.8 | 0 | 1 |
| Parity >2 | 0.6 | 1 | 3 |
| 20-24 | | | |
| Parity 1-2 | 0.9 | 5 | 13 |
| Parity >2 | 0.7 | 14 | 13 |
| 25-29 | | | |
| Parity 1-2 | 1.5 | 7 | 8 |
| Parity >2 | 1.0 | 13 | 11 |
| ≥30 | | | |
| Parity 1-2 | 1.5 | 15 | 6 |
| Parity >2 | 1.2 | 5 | 4 |
| Unknown | 1.0 | 1 | 1 |
| Mean relative risk | | 1.10 | 1.03 |
| 3. Age (years) at menopause (postmenopausal only) | | | |
| <40 | | | |
| Bilateral oophorectomy | 0.2 | 0 | 2 |
| Other | 0.3 | 5 | 4 |
| 40-44 | | | |
| Bilateral oophorectomy | 0.4 | 0 | 1 |
| Other | 0.8 | 4 | 3 |

| | | 1.10 | 1.03 |
|---------------------------------------------------|-----|------|------|
| Mean relative risk | | | |
| 3. Age (years) at menopause (postmenopausal only) | | | |
| <40 | | 0 | 2 |
| Bilateral oophorectomy | 0.2 | 5 | 4 |
| Other | 0.3 | | |
| 40-44 | | 0 | 1 |
| Bilateral oophorectomy | 0.4 | 4 | 3 |
| Other | 0.8 | | |
| 45-49 | | | |
| Bilateral oophorectomy | 0.8 | 1 | 1 |
| Other | 1.0 | 11 | 13 |
| >49 | | | |
| Bilateral oophorectomy | 1.6 | 4 | 0 |
| Other | 1.3 | 22 | 18 |
| Age unknown or premenopausal | 1.0 | 28 | 33 |
| Mean relative risk | | 1.07 | 1.01 |
| 4. Body mass index (lb/in ² × 1,000) | | | |
| Premenopausal | | | |
| <30 | 1.0 | 10 | 4 |
| 30-34 | 0.9 | 10 | 8 |
| 35-39 | 0.7 | 0 | 4 |
| >39 | 0.5 | 1 | 2 |
| Unknown | 0.9 | | 1 |
| Postmenopausal | | | |
| <30 | 1.0 | 15 | 13 |
| 30-34 | 1.5 | 23 | 23 |
| 35-39 | 1.6 | 12 | 12 |
| >39 | 1.3 | 2 | 8 |
| Unknown | 1.5 | 2 | |
| Mean relative risk | | 1.25 | 1.40 |
| 5. History of benign breast disease | | | |
| No | 1.0 | 55 | 65 |
| Yes | 2.7 | 17 | 10 |
| Not sure | 1.4 | 3 | |
| Mean relative risk | 1.2 | 1.39 | 1.30 |
| 6. Family history (Mother or sister) | | | |
| No | 1.0 | 56 | 65 |
| Yes | 2.9 | 16 | 10 |
| Not sure | 1.4 | 3 | 0 |
| Mean relative risk | | 1.44 | 1.19 |
| 7. Religion | | | |
| Jewish | 2.8 | 56 | 2 |
| Other | 1.0 | 19 | 73 |
| Mean relative risk | | 1.48 | 1.0 |

* Relative risks from Helmrich et al. (7) with slight alteration to decrease the number of categories. If the category was unknown, the relative risk of the group mean or the mean relative risk of the group was assigned.

lower than would be expected, after known risk factors are removed from consideration. The observed case rate of 113 per 100,000 is lower than the expected rate of 134, based on the same considerations.

Simple inspection of risk factor data, shown in table 1, may lead to similar qualitative conclusions. Calculation of the adjusted morbidity ratio depends on the relative risks that are chosen from the literature and requires the assumption that the study community, the comparison community, and the population reported in the literature are comparable. In the St. Louis Park example, however, none of the results support the hypothesis that other risk factors, such as water supply contamination, are responsible for the elevated cancer rates found in the Third National Cancer Survey study. The elevation is accounted for by the known factors observed and the relative risks derived from other studies.

DISCUSSION

The present study is a practical application of case-derived population attributable risk techniques to a common situation in environmental epidemiology. The rationale for calculating the adjusted mortality ratio is somewhat similar to direct and indirect standardization of disease rates using a reference population; it uses relative risks and results in a ratio of rates. The major advantage of the adjusted mortality ratio is that it does not require knowledge of the distribution of risk factors in the general populations of either of the two communities being compared and can therefore be obtained from case information alone. A conventional approach to the problem would have ascertained the frequency of known risk factors in each community's general population by interviewing randomly selected subjects. Indirect standardization of breast cancer risks using published relative risks would have allowed comparison of the incidence in the two communities. Alternative approaches would include determination of relative risks from a case-control study conducted within the two communi-

ties or following the entire cohort of "exposed" women living in St. Louis Park on a chosen date together with an unexposed comparison group.

Environmentally related cancer studies of this type, however, have several features that make conventional case-control studies difficult or impossible. In dealing with exposure to a water supply, product, or occupational factor, the number of cases in the exposed population may be too small to allow relative risks for other risk factors to be determined. In St. Louis Park, for example, with 25,424 "exposed" women, there were 95 cases of breast cancer in a three-year period. To determine relative risks for various factors, typical studies have involved hundreds or thousands of cases and similar numbers of controls.

There are advantages to comparing cases with other cases rather than with controls, particularly if the cases were diagnosed long before the study—as in St. Louis Park. The choice of community controls in such situations is complicated by migration into and out of the community, and controls and cases are likely to have different rates of mortality in the intervening years. Hospital record controls may be used, but their illnesses may introduce unwanted associations with risk factors for other diseases. Two groups in which all members are breast cancer patients, however, may be expected to have comparable levels of knowledge and concern about cancer, roughly comparable mortality levels, and the same difficulties associated with tracing both groups through their physicians, relatives, and neighbors.

For single risk factors, the non-attributable risk proportion ($1 - \text{population attributable risk proportion}$) of a group of cases is the mean value of the reciprocal of relative risk ($1/RR_i$) for each case i . When the composite relative risk of an individual represents the combined effect of several risk factors, the problem becomes one of combining the effects of the several factors. In practice, relative risks or odds ratios derived from multivariate analysis are not

always available from the literature. One must also assume that the populations studied have relative risks similar to those described in case-control studies in the literature that supplied the relative risks. If relative risks from multivariate studies are not available, an estimate may have to be made from univariate relative risks, after reaching a judgment concerning which risk factors are independent of each other. The alternative—examination of the direction of the individual population risks—may be preferable when the degree of interaction among risk factors is unknown.

Even a crude estimate of the population attributable risk proportions in the two case communities, however, will allow adjustment of the crude disease ratio first obtained for the influence of other risk factors. Although the result may not be mathematically precise due to interactions among risk factors or differences between literature populations and the ones under study, the probability of reflecting the true situation in nature is better than the alternative of ignoring risk factors other than age and sex. Although it would be desirable to have a test of statistical significance for the adjusted morbidity ratio, we have not developed one, and this must be left to future work.

The study reported here, performed in two stages using Third National Cancer Survey data files (comparable with cancer registry data), cost approximately \$10,000 for the initial comparison of cancer rates in St. Louis Park and the metropolitan area. Literature review, questionnaire design, and tracing and interviewing of 150 cases required a year of a Master's-degree-level epidemiologist's time, 1-2 weeks of a physician-epidemiologist's time, and a few hundred dollars in data processing costs. With modern techniques, processing of the second study could have been done with a microcomputer and a spreadsheet program. Full-scale epidemiologic studies would have cost many times as much if the biases mentioned above could have been surmounted at all. Had we had previous experience with

the technique described and adequate funds, the question of an association between breast cancer and the water supply could have been put in perspective within a month or two, thus shortening the period of anxiety on the part of St. Louis Park residents after breast cancer rates were initially found to be elevated.

This technique allows differences in rates to be evaluated through case interviews to ascertain known risk factors for the disease in question. It has a number of pitfalls, primarily in the lack of multivariate relative risks for most diseases in the literature, but it allows one to use, rather than ignore, current knowledge about risk factors for the disease in evaluating hypotheses. In the case of St. Louis Park, further studies covering other time periods, and hopefully evaluating individual polynuclear aromatic hydrocarbon exposure, would be desirable. In the interim, the differences in rates observed in Third National Cancer Survey data that first suggested an association between breast cancer in 1969-1971 and residence in St. Louis Park appear to be attributable to known breast cancer risk factors.

Since chemical measurements may be far more sensitive indicators of potential low-level chemical hazard than epidemiologic studies, this study contributes only one piece of evidence in attempts to solve a complex environmental problem. The adjusted mortality ratio technique may be useful in similar situations that occur fairly frequently in environmental epidemiologic work.

REFERENCES

1. Dusich K, Sigurdson E, Hall WN, et al. Cancer rates in a community exposed to low levels of creosote components in municipal water. *Minn Med* 1980;63:803-6.
2. Dao TL, Sunderland H. Mammary carcinogenesis by 3-methylcholanthrene. I. Hormonal aspects in tumor induction and growth. *JNCI* 1959;23:567-85.
3. Andervont HB, Dunn TB. Responses of strain DBA/2 mice without the mammary tumor agent to administration of 3-methylcholanthrene. *JNCI* 1953;14:329-39.
4. Dao TL. Carcinogenesis of mammary gland in rats. *Progr Exp Tumor Res* 1964;5:157-216.

ire cohort of "ex-
St. Louis Park on
with an unexposed

ed cancer studies
ve several features
case-control stud-
e. In dealing with
apply, product, or
number of cases in
may be too small to
ther risk factors to
ouis Park, for ex-
sed" women, there
cancer in a three-
e relative risks for
studies have in-
sands of cases and
ols.

o comparing cases
han with controls,
s were diagnosed
s in St. Louis Park.
y controls in such
by migration into
nity, and controls
ave different rates
vening years. Hos-
be used, but their
unwanted associa-
or other diseases.
all members are
however, may be
comparable levels of
n about cancer,
mortality levels, and
ciated with tracing
ir physicians, rel-

the non-attribut-
population attrib-
f a group of cases
reciprocal of rela-
case i . When the
an individual rep-
ect of several risk
omes one of com-
several factors. In
r odds ratios de-
analysis are not