Scand J Work Environ Health 1984;10(4):219-223

<https://doi.org/10.5271/sjweh.2338>

Issue date: Aug 1984

Extensions to a technique for relating mortality and environment--exemplified by nasal cancer and industry.

by [Gardner MJ](#), [Winter PD](#)

This article in PubMed: www.ncbi.nlm.nih.gov/pubmed/6494841

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Extensions to a technique for relating mortality and environment — Exemplified by nasal cancer and industry

by Martin J Gardner, PhD, Paul D Winter, MSc¹

GARDNER MJ, WINTER PD. Extensions to a technique for relating mortality and environment — Exemplified by nasal cancer and industry. *Scand J Work Environ Health* 10 (1984) 219–223. The use of geographic data to interrelate mortality rates and environmental variables is an important approach to determining reasons for the variations in disease frequency. An alternative methodology to standard regression analysis was recently suggested, which is particularly useful when there are a large number of geographic areas and relatively few deaths. An extension of this technique is described, which includes a modified regression approach using areas aggregated according to their level of the environmental factor. An example is given relating nasal cancer mortality to the distribution of industry in England and Wales. The known hazards in the furniture and leather industries are indicated and other potential risk industries, including tailoring, are suggested.

Key terms: geographic studies, industrial risks, regression analysis.

A recent paper by Selvin and his colleagues (10) suggested a novel technique for looking at the relationship between mortality rates and environmental variables in different geographic areas. It was proposed by the authors as an alternative to classical regression analysis. The new approach has the attraction that it is simple to use and to interpret, but it retains the essence of regression analysis.

The relationship investigated by Selvin et al (10) was that between mortality rates from acute lymphocytic leukemia in white females under five years of age during 1969–1977 in 3,075 counties in the United States and various socioenvironmental variables. Rather than use standard regression analyses, they collapsed the 3,075 counties into a small number (actually 10) of “equal-risk categories” in terms of the socioenvironmental variable — and then examined the statistical significance of the variation in mortality between these categories. Specifically, for an income variable, the 3,075 counties were first ranked in increasing order of the percentage of their population with an annual income of USD 15,000 or more. The counties were then sequentially amalgamated into categories so that each of the 10 categories had an equal population-at-risk of white females under the age of five years. The mortality rate from acute lymphocytic leukemia in each category was

then calculated, and a (chi-square) statistic presented for testing the variation between the 10 categories.

We have been interested in relating cancer mortality to environmental factors in a similar way among the 1,366 local authority areas of England and Wales. For reasons mentioned by Selvin et al (10) — such as the small numbers of deaths involved in estimating small-area mortality rates and the large differences in population size between administrative areas — we were unsure of the wisdom of using conventional regression techniques. We have found the proposed method a helpful one in analyzing mortality/environmental data, and it has become clear that certain extensions of the technique make it even more useful in some ways. The purpose of this paper is to describe an investigation into nasal cancer and industry using this approach with modifications.

Materials and methods

We have used the “equal-risk categories” technique to search for associations between nasal cancer mortality among men during 1968–1978 and the distribution of industry among the 1,366 local authority areas. The data we have used are the observed and expected numbers of deaths from nasal cancer among men during the 11 years in each area (the expected number being calculated by the indirect method of standardization based on the age-specific rates for men in England and Wales) and the percentage of the local male population who were employed in different industries. The mortality data were available through the compilation of an atlas of

¹ MRC Environmental Epidemiology Unit (University of Southampton), Southampton General Hospital, Southampton SO9 4XY, England.

Address for reprints: Dr MJ Gardner, MRC Environmental Epidemiology Unit (University of Southampton), Southampton General Hospital, Southampton SO9 4XY, England.

cancer mortality (5). The industrial data were obtained from the 1971 census and contained information on employment in 207 different industrial units (8).

The 1,366 areas were first ranked from lowest to highest in terms of the percentage of men working in a particular industrial unit. In Selvin et al (10) the United States counties were then amalgamated into equal (color-, sex-, and age-specific) population-size categories — in our case, since we were dealing with deaths of men at all ages, the amalgamation was carried out with the use of the expected numbers of deaths in each area. This procedure acts as a reflection of the size of the “population-at-risk” for nasal cancer, in the same manner as the age-specific population size did in Selvin et al’s example. Clearly, the total population would not have been suitable as a measure of “at-risk” in the 1,366 areas because of their differing age structures.

Selvin et al (10) justified testing for statistical variation among the categories by the standard chi-square test ie, $\Sigma(0_i - E)^2/E$, where the 0_i s are the numbers of observed deaths and E is the number of expected deaths in each category. An extension to Selvin et al’s proposal would be to subdivide the degrees of freedom to test specific hypotheses about contrasts between the categories. A common interest — as in a standard regression situation — will be a test for trend. This procedure can be carried out with the method described by, among others, Darby & Reissland (3) and Hakulinen (6). The appropriate statistic is

$$\frac{\{\Sigma z_i (0_i - E)\}^2}{E\{\Sigma(z_i - \bar{z})^2\}},$$

where z_i is the value of the environmental variable

for category i (in this example, the percentage in the particular industrial unit) and $\bar{z}$ is the average of the z_i s. This statistic is distributed as a chi-squared variable with one degree of freedom.

This test for linear trend is associated with the increasing ratio of the observed : expected numbers of nasal cancer deaths as the proportion employed in industry increases. The actual value of the linear slope can be found as

$$\frac{\Sigma z_i (0_i - E)}{E\{\Sigma(z_i - \bar{z})^2\}},$$

with approximate standard error given by

$$\frac{1}{\sqrt{E\{\Sigma(z_i - \bar{z})^2\}}}.$$

Note that the chi-squared test can equivalently be carried out by calculating (slope/se slope)². The intercept on the vertical (observed : expected) axis is found as

$$1 - \text{slope} \times \bar{z}.$$

Other contrasts can be tested if wanted by further subdivision of the overall chi-square into its component degrees of freedom. These subdivisions can be used, for example, to test quadratic or other terms as might be considered a priori, or as suggested from a plot of the data.

Results

The relationship between nasal cancer in men and employment in the furniture and upholstery industry

Table 1. “Equal-risk categories” based on the percentage of the local male population employed in the furniture and upholstery industry and the number of deaths from nasal cancer in each category.^a

Category	Percent in industry	Men employed in industry ^b	Total of employed men ^b	Number of areas	Number of observed deaths	Observed: Expected ^c
1	0.00	0	73,488	75.5	76.5	0.98
2	0.00	0	65,245	130.6	63.5	0.82
3	0.00	0	58,166	323.2	71.0	0.91
4	0.06	47	73,268	51.4	64.1	0.82
5	0.12	85	71,684	44.3	67.9	0.87
6	0.16	111	70,414	43.0	82.0	1.05
7	0.19	134	70,781	34.8	90.8	1.17
8	0.21	147	68,675	48.9	75.9	0.98
9	0.26	177	68,690	48.2	57.9	0.74
10	0.30	191	64,456	39.4	70.6	0.91
11	0.34	228	67,382	64.6	84.5	1.09
12	0.39	265	68,832	54.3	67.0	0.86
13	0.44	305	69,480	47.9	75.6	0.97
14	0.50	336	67,943	50.6	67.0	0.86
15	0.59	392	66,500	51.8	75.3	0.97
16	0.72	474	66,176	28.8	91.4	1.17
17	0.82	557	68,219	54.9	94.6	1.22
18	0.98	680	69,571	45.7	100.2	1.29
19	1.40	963	68,940	63.1	68.2	0.88
20	3.12	2,153	68,958	65.0	111.9	1.44

^a Furniture and upholstery — Industrial unit number 472 (8); nasal cancer — International Classification of Diseases, eighth revision, code 160 (11).

^b Ten percent sample of men aged 15–64 years in the 1971 census.

^c The number of expected deaths in each category is 77.8 (see text).

is described first. Table 1 sets out the basic information. Overall there were 1,556 deaths among men from cancer of the nose during 1968–1978 in England and Wales. With 20 equal-risk categories in use, each category will include $77.8 (= 1,556 : 20)$ expected deaths. Sometimes this procedure involved splitting the expected number of deaths for an area on the boundary between two equal-risk categories — when this situation occurred the observed number of deaths in that area was split in the same proportion among the two adjacent categories. The ratios of the observed to expected numbers of deaths in each equal-risk category are shown in the final column of table 1, and an inconsistent upward trend is noticeable (results of a formal statistical test are given below) and demonstrated in figure 1.

With 20 equal-risk categories the overall chi-square statistic for testing variation in the observed : expected ratios has 19 degrees of freedom and, for these data, gives a value of 47.4 ($p < 0.01$). The trend statistic chi-square value is 20.02 for which, with one degree of freedom, $p < 0.001$. The associated value of the slope is 0.1640, with a standard error of 0.0366. The intercept was found to be 0.9133.

In table 1 for each of the first three equal-risk categories the percentage of men employed in the furniture and upholstery industry is zero. The industrial data used were collected in the 1971 census on a 10 % sample basis, and as a consequence only an estimate of the percentage of men employed for each area is available. The sampling error involved will be smaller, however, for larger population areas. Therefore we have ranked areas with zero (or equal) estimated percentages in decreasing numerical size of their male population — so that areas with larger populations appear in the first of the initial three categories in table 1. Combining these categories gives 211 observed deaths compared to 233.4 expected — a deficit of some 22 deaths in the areas where there was an (estimated) absence of any furniture and upholstery industry. If these three separate categories were to be collated into one overall category with zero percentage in the industry, the chi-square test for trend and the slope and intercept estimates would be unaltered — this result can be seen from the general formulas given in the appendix.

To compare the relationships between nasal cancer and the percentage of men employed in different industries, it is important to modify the slope estimate already given, because the variation between areas in the percentage employed will differ from industry to industry. One appropriate procedure is to use effectively a standardized slope — obtained by multiplying the slope already given by the standard deviation of the distribution of the percentage of men in the particular industry. For the furniture and upholstery industry this standard deviation is 0.71, giving a standardized slope of 0.1164. Multiplying this

figure by 100 implies an 11.64 % increase in the observed : expected ratio for a unit standard deviation increase in the proportion of men employed in the industry.

In this manner table 2 shows the industrial units which have a statistically significant positive trend, listed in decreasing rank order of their standardized slope. As would be expected from previous epidemiology (1, 4, 7), the furniture and upholstery industry tops the list. The leather goods industry — in which boot and shoe manufacture is known to be associated with nasal cancer (2) — is one of the highest. A number of sections of the tailoring industry appear in table 2, and a relationship between nasal cancer and this industry has been suggested earlier (4).

Discussion

The relationship described between nasal cancer and employment in various industries is part of a much larger study. Our intention is to relate a large number of cancers to these same industrial units (and simultaneously to a list of occupational units) in a hypothesis-generating exercise, using basically the same method. The hope is that, by carrying out a systematic search of the data, we will find new leads to possible causes of cancer — for both men and women. It will be possible to compare significant findings with this approach to the relationships between occupation-industry and cancer found in the occupational mortality decennial supplement for England and Wales (9) for consistency or otherwise. Any replicated associations would be of more interest to follow up further because of the difference in approach — the latter is based on occupational data taken from individual death certificates. There is, of course, a statistical problem of which we are aware in this approach — that of testing multiple compari-

Figure 1. Relationship between mortality from nasal cancer in men and employment in the furniture and upholstery industry.

Table 2. Relationships between mortality (100 × observed deaths : expected deaths) from nasal cancer in men during 1968–1978 and the percentage of males employed in different industries in the 1,366 local authority areas.

Industrial unit ^a	Slope of relationship ^b	Chi-square test ^c	
		Trend	Remainder
Furniture and upholstery	11.64	20.02**	27.36
Glass	7.44	8.17**	24.78
Men's and boys' tailored outerwear	7.42	8.15**	19.83
Copper, brass and other copper alloys	7.30	7.87**	31.02*
Universities	7.16	7.58**	17.52
Leather goods	7.09	7.43**	22.41
Brushes and brooms	6.66	6.55*	25.17
Brewing and malting	6.38	6.02*	22.14
Dresses, lingerie, infant's wear	6.35	5.96*	25.61
Women's and girls' tailored outerwear	6.31	5.88*	34.98**
Metal industries (not elsewhere specified)	6.25	5.77*	15.84
Iron castings	6.15	5.59*	25.33
Motorcycle, tricycle and pedal cycles	6.03	5.38*	26.77
Miscellaneous manufacturing	6.02	5.35*	27.08
Telegraph and telephone apparatus	5.77	4.91*	19.17
Public houses	5.66	4.74*	13.67
Mechanical handling equipment	5.33	4.20*	22.19
Locomotive and railway track equipment	5.21	4.01*	27.97
Domestic electrical appliances	5.19	3.99*	31.04*

^a Standard industrial classification (8). Only the 19, out of a total of 207, industries with statistically significant positive trends are shown — in ranked order.

^b Standardized slope × 100 (see text).

^c Trend, 1 degree of freedom: * $p < 0.05$ ($\chi^2 > 3.84$), ** $p < 0.01$ ($\chi^2 > 6.63$); remainder, 18 degrees of freedom: * $p < 0.05$ ($\chi^2 > 28.87$), ** $p < 0.01$ ($\chi^2 > 34.81$).

sons. But the intention is to provide hypotheses for further study, rather than to use statistical significance in a formal sense.

Other socioenvironmental data will be used in a way similar to that used for the industrial and occupational data. By stratification of the areas on the basis of, for example, social class it will be possible also to study the effect of more than one variable at a time. Some variables will not be subject to the restriction of being collected from a 10 % sample, which will remove the potential for misclassifying areas in terms of the environmental indices of interest. Where such a potential occurs, as in the present example, and where the sampling error is unbiased, there will be a tendency for any real relationship to be underestimated.

The results from amalgamating the 1,366 areas into 20 equal-risk categories have been given in this paper. There is no particular justification for using this number, but replicate analyses using from 2 to 200 categories produced similar findings. Thus, for the furniture and upholstery industry, standardized slopes of 10.50, 11.64, 8.91, 12.02, 13.42, and 15.65 were found when 2, 5, 10, 50, 100, and 200 categories were used. These figures are somewhat similar to the slope for 20 categories (11.64), although there is a noticeable small upward trend among the final values. In this case, the nasal cancer rate in the last of the categories is influenced strongly by two areas where the furniture industry is concentrated. These areas (High Wycombe Municipal Borough and Wycombe Rural District) are known to have particularly high mortality rates from nasal cancer (1, 4), and, as the number of equal-risk categories increases, they have a progressively greater effect on the

observed : expected ratio in the last category. However, no matter how many equal-risk categories (other than only two) are used, the industries associated with nasal cancer are ranked similarly to the classification used in table 2. However, the desirable number of categories to use in any particular application would be of interest to study further.

An alternative to grouping areas together on an "equal-risk" basis would be to amalgamate them on some geographic basis. The next largest units to local authority areas in England and Wales are counties, of which there are 59. This kind of aggregation would, of course, result in the combination of areas less homogeneous in terms of the proportion of men working in industry. However the amalgamation would be possibly more homogeneous in relation to any confounding variables. In this particular case a slope of 11.92 is found — not dissimilar to the values already presented.

It is encouraging that the method has detected known risk industries for nasal cancer — furniture and upholstery and leather goods. Other industries, in particular, the tailoring industry, are also highlighted as possibly containing a hazard. We have underway a case-referent study of nasal cancer in east London, an area of high mortality (4); this study will be looking at occupational and industrial factors, including tailoring.

Acknowledgment

We would like to acknowledge the help and advice of Dr S Selvin from the University of California at Berkeley in discussions and program development during a sabbatical period he spent with us.

References

1. Acheson ED, Cowdell RH, Hadfield EH, MacBeth RG. Nasal cancer in woodworkers in the furniture industry. *Br Med J* 2 (1968) 587—596.
2. Acheson ED, Cowdell RH, Jolles B. Nasal cancer in the Northamptonshire boot and shoe industry. *Br Med J* 1 (1970) 385—393.
3. Darby SC, Reissland JA. Low levels of ionising radiation and cancer — Are we underestimating the risk? *J R Stat Soc A* 144 (1981) 298—331.
4. Gardner MJ, Winter PD, Acheson ED. Variations in cancer mortality among local authority areas of England and Wales: Relations with environmental factors and search for causes. *Br Med J* 284 (1982) 784—787.
5. Gardner MJ, Winter PD, Taylor CP, Acheson ED. Atlas of cancer mortality in England and Wales, 1968—78. Wiley, Chichester 1983.
6. Hakulinen T. A Mantel-Haenszel statistic for testing the association between a polychotomous exposure and a rare outcome. *Am J Epidemiol* 113 (1981) 192—197.
7. Hernberg S, Westerholm P, Schultz-Larsen K, Degerth R, Kuosma E, Englund A, Engzell U, Sand Hansen H, Mutanen P. Nasal and sinonasal cancer: Connection with occupational exposures in Denmark, Finland and Sweden. *Scand J Work Environ Health* 9 (1983) 315—326.
8. Office of Population Censuses and Surveys. Standard industrial classification. Her Majesty's Stationery Office, London 1968.
9. Office of Population Censuses and Surveys. Occupational mortality decennial supplement 1970—72, England and Wales. Her Majesty's Stationery Office, London 1978. (Series DS, no 1).
10. Selvin S, Merrill D, Sacks ST. An alternative to ecologic regression analysis of mortality rates. *Am J Epidemiol* 115 (1982) 617—623.
11. World Health Organization. International classification of diseases. Eighth revision. Geneva 1967.

Received for publication: 20 March 1984

Appendix

The formulas given in the main text of this paper are specific results which apply to the situation of "equal-risk categories." If the expected numbers vary from group to group, the more general formulas below are appropriate for trend testing.

$$\chi^2 = \frac{\{\sum z_i(0_i - E_i)\}^2}{\sum z_i^2 E_i - (\sum z_i E_i)^2 / \sum E_i},$$

$$\text{slope} = \frac{\sum z_i(0_i - E_i)}{\sum z_i^2 E_i - (\sum z_i E_i)^2 / \sum E_i},$$

$$\text{se (slope)} = \frac{1}{\sqrt{\sum z_i^2 E_i - (\sum z_i E_i)^2 / \sum E_i}},$$

and the

$$\text{intercept} = 1 - \text{slope} \times \frac{\sum E z_i}{\sum E_i}.$$

Finally, for the specific case of equal E s, there is a direct relationship between the standardized slope and the trend chi-square statistic, with the standardized slope given as

$$\sqrt{\frac{\chi^2}{E(K-1)}},$$

where K is the number of equal-risk categories. This relationship can be seen in table 2 of the Results section.