Journal of Epidemiology and Community Health 1992; **46**: 305–310

Problems in the interpretation of small area analysis of epidemiological data: the case of cancer incidence in the West of Scotland

David J Hole, Douglas W Lamont

Abstract

***Study objective*—The aim was to examine the extent to which random variation alone will produce differences in observed incidence rates between small areas which will affect measures of spatial clustering and estimates of relative risk.**

***Design*—This was a study of changes in the pattern of spatial concentration of cancer incidence over a five year time period. A comparison was made of observed incidence rates for 34 tumour sites with randomly generated values and, where possible, with expected values derived from known relative risks.**

***Setting*—Twenty six local government districts in the west of Scotland.**

***Main results*—A statistically significant relationship was observed between sample size and the stability of a summary measure of spatial concentration. Almost all observed highest:mean rate ratios were within the 95% confidence interval of the simulated distribution of these values. In three cases examined, both observed and simulated highest:lowest rate ratios were larger than those expected on the basis of known exposures to risk.**

***Conclusions*—In the absence of a prior hypothesis, small area analysis of epidemiological data for periods of less than 10 years will almost always give misleading results for all but the most common diseases.**

Small area analysis of epidemiological data in the absence of a prior hypothesis has become common in recent years. The advent of sophisticated computer systems and graphics packages and the computerisation of census data down to enumeration district level has made possible the production of cancer incidence and mortality rates for ever smaller geographical units.[1–4] It is our concern that the ease with which these data can be generated will encourage fruitless attempts to formulate hypotheses based on differences for which random variation may well be the major component.

Where small area analysis has been of use in suggesting possible risk factors, two conditions have been met. Firstly, widely differing incidence rates have occurred in areas which are close together geographically, and secondly the areas used have been of sufficient population size for observed differences in rates to be more than simply a function of random variation.

Populations are at risk from two types of hazard: lifestyle factors, which vary gradually across different parts of the country and which are best described in terms of a general pattern of concentration or dispersion (by mapping incidence rates); and localised environmental hazards, which affect only those living in the immediate vicinity and which are more likely to be identified by examining differences in incidence rates between individual areas. This paper examines the extent to which random variation alone can produce substantial differences in observed rates between small areas which will affect measures of spatial clustering and estimates of excess risk based on those areas with the highest rates.

Methods

Cancer incidence data for the 26 local government districts of the west of Scotland (median population 79 000) were obtained from the Scottish cancer atlas[3] for the period 1975–80 and from the West of Scotland Cancer Registry for 1980–85. In order to ensure the same degree of completeness of registration in both data sets (ie, allowing the same length of time from the latest date of diagnosis to the time at which analysis was undertaken) a small overlap in the two time periods was necessary.

Three approaches were used to examine the extent to which results from small area analysis can be influenced by random error alone. (1) Comparison was made of the stability of a summary measure of spatial concentration between the two time periods, over which little change would be expected. (2) Comparison was made of observed highest/mean ratios with simulated values derived from a distribution randomly generated on the assumption of no difference in rates between districts. Mean values were taken as being the rate for the area as a whole. The use of minimum values throughout would have made the ratios unstable for areas with very low rates and impossible to calculate for those with zero rates. (3) Comparison was made of observed and simulated highest/lowest ratios, with expected values calculated on the basis of different levels of exposure to risk between areas.

The summary measure of spatial concentration used was the statistic "D", developed for use in the Scottish cancer atlas. [3] This is defined as the average absolute difference in rank order of incidence between all possible pairs of spatially adjacent districts.[3] Its sampling distribution is approximately normal about a mean of $(N+1)/3$ where N is the total number of districts. Low values of D are indicative of some degree of spatial clustering.

West of Scotland Cancer Surveillance Unit, Ruchill Hospital, Glasgow G20 9NB, United Kingdom
D J Hole
D W Lamont

Correspondence to:
Mr Hole

Accepted for publication
July 1991

The sampling distribution of D for the west of Scotland was derived from 100 000 random assignments of ranks to each of the 26 local government districts in the region. This yielded a distribution with a mean of 9·01, in which 95% of the observations were within the range 7·55 to 10·32 (fig 1). Values less than 6·0 were encountered only twice in the course of 100 000 generations of the D statistic, and no value occurred which was greater than 12·0.

The simulation of rates for different areas was based on the assumption that the "true" incidence rate was the same for each district, equal to that of the west of Scotland as a whole. The expected number of cases for each district was then calculated. This provided the mean value for a Poisson distribution conditioning on area population size, similar to the approach used by Alexander *et al*[5] whereby a site specific distribution of the D statistic was derived by allocating cases to districts according to population size. A value was chosen at random from each of these distributions from which a simulated rate could then be calculated, allowing highest/mean and highest/lowest ratios to be derived. This process was repeated 1000 times to derive a sampling distribution for these ratios.

Real differences between areas in the proportions of the population exposed to a known risk factor should be reflected in the incidence rates observed and thus in the ratio of the highest to the lowest of these rates. The age standardised incidence rate is a function of the proportion of the population exposed to a particular risk factor and the relative risk of contracting a given disease if thus exposed. This can be calculated from the formula

$$r = (r_n \times P_n) + (RR \times r_n \times P_e)$$

where RR = relative risk for disease; r = incidence rate in area; r_n = incidence rate among those not exposed to the risk factor; RR × r_n = incidence rate among those exposed to the risk factor (r_e); P_e = proportion of population of area exposed to risk factor; and P_n = proportion of population of area not exposed to risk factor.

Assuming that the "true" incidence rate of the disease in the non-exposed population is the same for all areas, the top/bottom ratio would be given by

$$\frac{P_{nt} + RR \times P_{et}}{P_{nb} + RR \times P_{eb}}$$

where the suffixes t and b refer to the areas with the top and bottom rates. This expected top/bottom ratio was compared with the observed ratio, and with that which would occur on the basis of chance alone (as produced by the simulation described above).

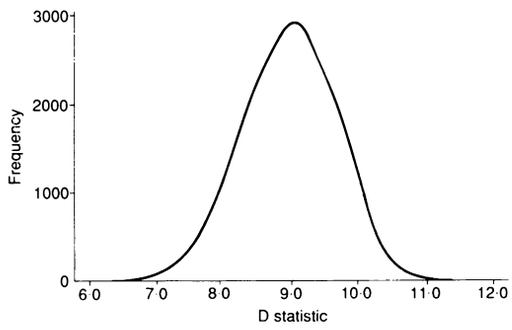

Figure 1 Sampling distribution of D statistic for west of Scotland based on 100 000 random permutations of district ranks

Results

STABILITY OF D STATISTIC OVER TIME

Figures 2 and 3 illustrate the contrasting spatial patterns produced by mapping incidence rates for two of the most and least common tumours in Scotland. Figure 2 shows the spatial distribution of areas of high and low incidence of lung cancer in men for 1980–85 compared with the situation five years earlier. The pattern is highly concentrated, with most cases per 100 000 population occurring in the Central Clydeside conurbation. This is reflected in a D statistic which is signifi-

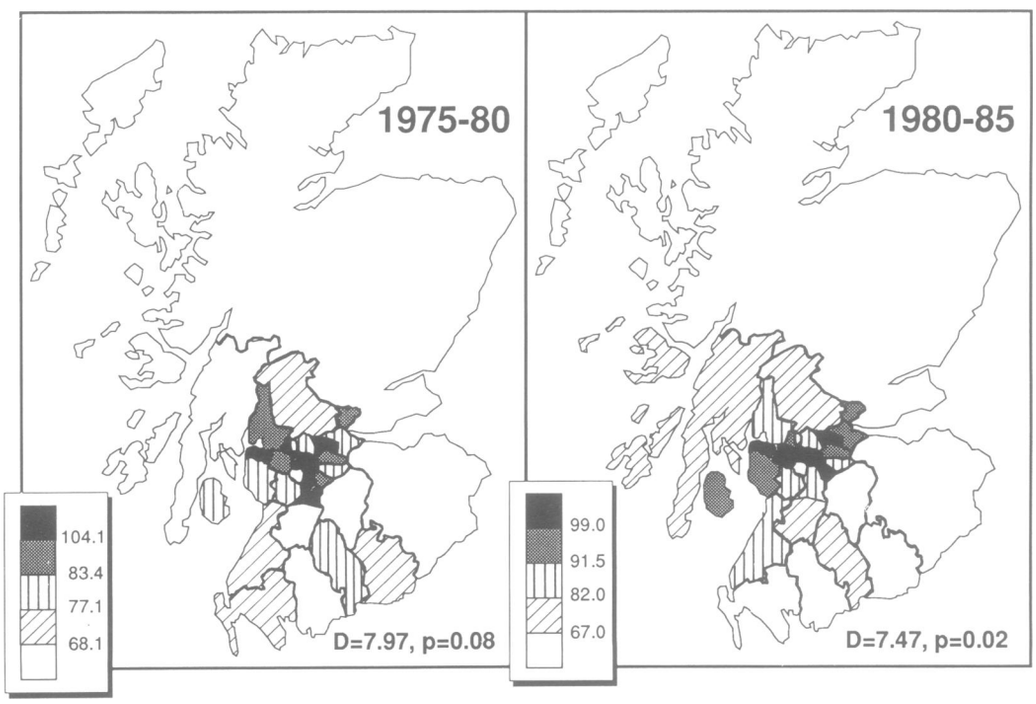

Figure 2 World age standardised incidence of lung cancer in males—west of Scotland 1975–85 (quintiles)

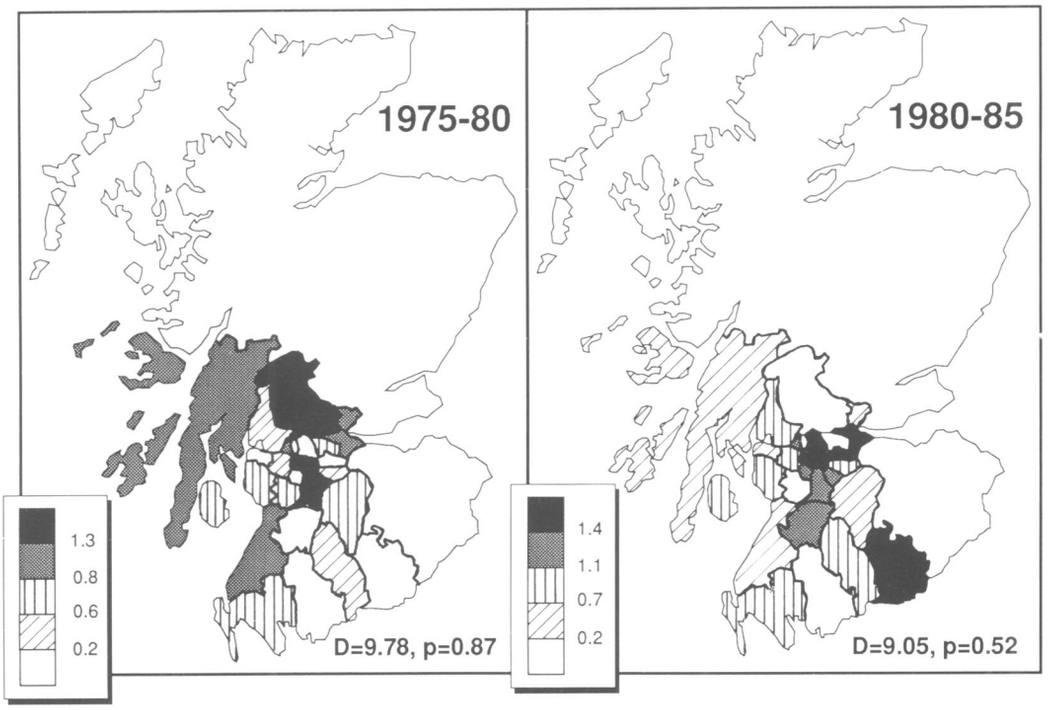

Figure 3 World age standardised incidence of laryngeal cancer in females—west of Scotland 1975–85 (quintiles)

cantly lower than would have been expected by chance and which also shows little change over time. In contrast, two of the five highest ranking districts for cancer of the larynx in women between 1980 and 1985 (fig 3) were among the five lowest ranking districts five years earlier. The patterns produced by mapping the incidence of this tumour over time show no evidence of concentration, and are quite different from each other and from the distribution of lung cancer, despite a common risk factor (cigarette smoking).

Table I shows the values of D obtained from analysis of the spatial distribution of incidence rates at each of the two time periods for 16 tumour sites in men and 18 in women. Six D statistics achieve statistical significance at the 5% level for 1975–80 but only in one case (large bowel in women) is there any support for a similar degree of clustering five years later. A statistically significant relationship exists between stability of the D statistic over time and the total number of cases available for analysis (fig 4). The largest difference in D statistic (2·05 for lip cancer in females) was produced by comparison of patterns based on fewer than 50 cases recorded for the whole of the west of Scotland. All of the nine sites for which 1500 cases or more were available recorded differences in the value of D of 0·73 or less.

Table I Spatial pattern of cancer incidence rates 1975–1985: index of non-randomness (D)

Site (males)	West Scotland (1975–1980) No of cases	D_1	p	West Scotland (1980–1985) No of cases	D_2	p
			Males			
Lung	11 110	7·97	0·08	11 505	7·47	0·02
Large bowel	3682	9·10	0·55	3878	9·83	0·88
Stomach	2405	9·22	0·62	2408	8·59	0·28
Prostate	2238	8·43	0·21	2922	8·60	0·29
Bladder	1970	8·79	0·38	2388	9·17	0·59
Pancreas	958	8·90	0·44	971	9·31	0·67
Oesophagus	876	7·64	0·03	1050	8·59	0·28
Leukaemia	647	9·00	0·49	780	9·05	0·52
NH lymphoma	560	7·97	0·08	784	8·57	0·27
Larynx	456	10·16	0·96	612	8·43	0·21
Oral/pharynx	431	8·88	0·43	595	9·19	0·60
Testis	263	7·72	0·04	372	9·09	0·54
Hodgkin's disease	241	9·52	0·77	239	9·07	0·53
Melanoma	191	9·28	0·65	352	8·50	0·24
Lip	164	6·71	0·00	183	8·45	0·22
Thyroid	53	7·57	0·03	59	8·69	0·33
			Females			
Breast	7264	8·47	0·23	7760	8·48	0·23
Large bowel	4417	7·43	0·02	4451	7·62	0·03
Lung	3524	8·52	0·25	4690	8·83	0·40
Stomach	2114	8·31	0·17	1915	8·05	0·10
Ovary	1357	8·24	0·15	1552	7·57	0·03
Cervix	1208	8·88	0·43	1390	9·31	0·67
Pancreas	982	9·55	0·78	1066	8·29	0·16
Bladder	884	9·79	0·87	1168	9·67	0·83
Oesophagus	740	8·07	0·10	817	8·48	0·23
Endometrium	736	8·36	0·19	880	7·47	0·02
NH lymphoma	576	9·64	0·82	813	8·90	0·44
Leukaemia	566	9·38	0·70	650	7·79	0·05
Melanoma	401	8·43	0·21	652	8·91	0·45
Oral/pharynx	274	8·66	0·31	344	8·78	0·37
Hodgkin's disease	215	8·74	0·35	204	7·81	0·05
Thyroid	193	7·95	0·08	193	8·28	0·16
Larynx	141	9·78	0·87	171	9·05	0·52
Lip	20	6·78	0·00	33	8·83	0·40

COMPARISON OF OBSERVED AND SIMULATED RATE RATIOS

Table II compares observed highest to mean rate ratios for 1980–85 with simulated values based on 1000 randomly generated distributions of cases. For 28 of the 34 sites, observed highest/mean ratios were within the 95% confidence limits of those achieved by simulation, indicating that random variation would more than account for the range observed. Two sites (larynx and stomach in females) showed less variation than would be expected by chance, observed highest/mean ratios being below the lower 95% confidence limit. In the four remaining sites the observed value was significantly higher than the simulated figure. These comprised two of the most common (male lung and female breast) and least common (male lip and female melanoma) and the ratios were not correlated with sample size, indicating a mixture of random variation and genuine differences in incidence.

COMPARISON OF OBSERVED RATE RATIO WITH THAT CALCULATED ON THE BASIS OF DIFFERENTIAL EXPOSURE TO RISK

Table III shows expected rate ratios for lung, bladder and breast cancer, calculated on the basis

of differences between areas in the prevalence of a history of cigarette smoking,[6 7] employment in a pigments factory,[8] and obesity.[7 9 10] The final two columns present the ratios for the highest and lowest rates when only random variation is present (simulated ratio) and when random variation and genuine differences are combined (observed ratio). In all cases the expected ratios based on known risks and exposures are considerably smaller than those observed, and are also lower than simulated values.

Discussion

The collection and mapping of cancer incidence data for geographical areas has proved useful in generating hypotheses of aetiological factors when the areas have contained sufficiently large populations and the lifestyles or other risk factors involved have been quite distinct.[11 12] The transfer of this approach to smaller areas still provides wide ranges in incidence rates but the role of random variation is considerably greater. Computationally sophisticated measures of spatial clustering have been devised in recognition of the fact that risk factors may cross administrative boundaries and to make full use of the data available.[13] Our concern is that a proliferation of inductive studies at this small area level will detract from deductive investigations of localised environmental hazards, such as those which have been associated with nuclear installations.[14 15] To this end, using data from the west of Scotland, we have attempted to show the shortcomings of small area analysis with no prior hypothesis.

Substantial change in the spatial pattern of incidence of different types of cancer due to changes in lifestyle or environment was not expected over such a short period as five years. Most changes which did occur in the chosen

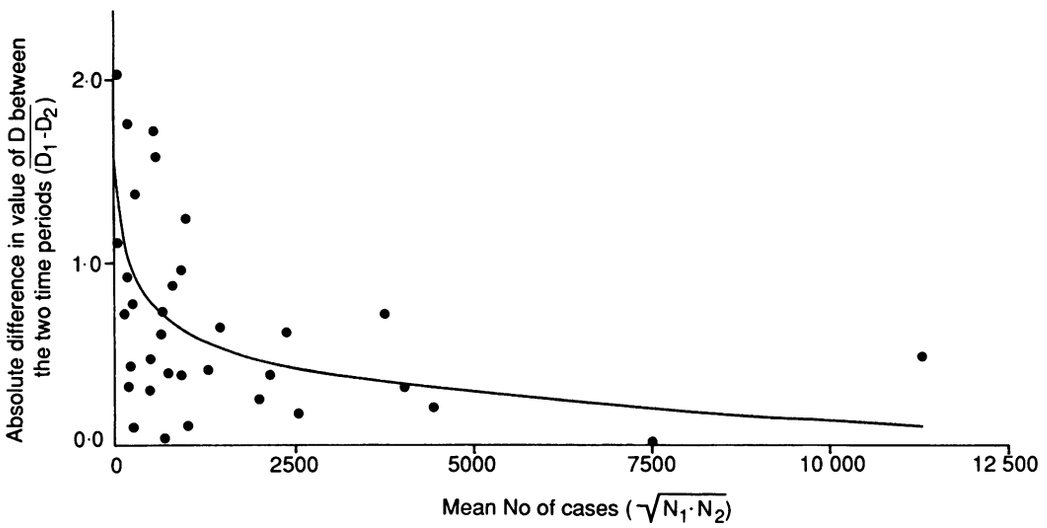

Figure 4 Relationship between stability of D statistic and sample size by tumour site (males and females), west of Scotland 1975–85

Table II Observed and expected ratios between highest district rates and those for the west of Scotland as a whole, 1980–1985

Tumour site	Observed incidence rates: Highest	Mean	Lowest	Observed H/M ratio	Simulated H/M ratio (95% CI)
		Males			
Thyroid	1·8	0·6	0·0	3·00	(2·00–6·40)
Lip	7·2	1·6	0·0	4·50*	(1·60–3·77)
Hodgkin's disease	6·3	2·7	0·0	2·33	(1·43–3·20)
Melanoma	6·1	3·5	1·2	1·74	(1·37–2·82)
Testis	10·8	4·4	0·8	2·45	(1·37–2·59)
Oral/pharynx	8·1	5·5	1·3	1·47	(1·30–2·44)
Larynx	10·6	5·6	1·6	1·89	(1·31–2·42)
NH lymphoma	12·0	7·5	4·7	1·60	(1·26–2·20)
Leukaemia	11·5	7·9	3·3	1·46	(1·27–2·11)
Pancreas	12·0	8·4	4·7	1·43	(1·25–2·15)
Oesophagus	14·0	9·3	2·9	1·51	(1·23–2·06)
Bladder	30·2	20·8	12·4	1·45	(1·16–1·66)
Stomach	27·0	21·1	13·5	1·28	(1·15–1·65)
Prostate	36·4	24·2	18·3	1·50	(1·14–1·62)
Large bowel	45·2	34·0	23·8	1·33	(1·12–1·51)
Lung	133·4	100·1	57·2	1·33*	(1·07–1·29)
		Females			
Lip	1·5	0·2	0·0	7·50	(2·87–11·0)
Larynx	2·0	1·2	0·0	1·67*	(1·69–4·42)
Thyroid	3·5	1·5	0·4	2·33	(1·57–3·83)
Hodgkin's disease	4·5	1·9	0·0	2·37	(1·51–3·51)
Oral/pharynx	5·0	2·4	0·9	2·08	(1·46–3·09)
Oesophagus	8·5	4·6	2·6	1·85	(1·33–2·49)
Leukaemia	8·8	4·9	2·6	1·80	(1·31–2·42)
Melanoma	18·4	5·6	3·1	3·29*	(1·31–2·30)
NH lymphoma	8·1	5·6	2·4	1·45	(1·31–2·30)
Pancreas	8·5	6·1	3·7	1·39	(1·30–2·31)
Endometrium	9·9	6·3	3·0	1·57	(1·29–2·28)
Bladder	10·2	7·1	4·1	1·44	(1·27–2·22)
Stomach	12·1	10·5	5·2	1·15*	(1·22–1·91)
Ovary	16·8	11·5	8·4	1·46	(1·22–1·87)
Cervix	19·6	12·4	5·6	1·58	(1·20–1·85)
Large bowel	36·7	25·9	16·9	1·42	(1·14–1·58)
Lung	43·9	31·2	18·9	1·41	(1·12–1·51)
Breast	82·7	59·7	44·5	1·39*	(1·09–1·37)

H/M = highest/mean
*Significant at $p < 0·05$

measure of spatial concentration and in the incidence rates themselves could therefore be attributed to random variation. Even within one time period (1980–85) random error accounted for a large proportion of the variation observed (table II).

Ten of the 34 cancer sites presented in the Scottish cancer atlas yielded a D statistic significant at $p<0.01$. This level of significance was reached only twice when analysis was confined to the west of Scotland (table I). While some of the explanation for this disparity may be due to a greater variation in lifestyle or environment at a national rather than regional level, the influence of a greater proportion of very small local government districts in the national data set should not be underestimated. One additional case of cancer in an area of less than 25 000 population would result on average in an increase of at least 8·0 in the sex specific incidence rate per 100 000. Eleven of Scotland's 56 local government districts (20%) contained fewer than 25 000 people at the 1981 census (table IV). These same areas recorded cancer incidence rates which were the highest in the country for 18 (53%) out of the 34 tumour sites presented.

Comparison of the spatial patterns produced by mapping on a relative (quintile) scale incidence rates for two of the largest and smallest tumour sites (cancer of the lung in males and of the larynx in females) illustrates clearly the problems involved in achieving a balance between a sufficiently detailed spatial framework (such as that provided by local government districts) and a sufficiently large number of cases in each (figs 2 and 3). The statistical reliability of any spatial analysis is dependent on the number and size of the spatial units used. Clayton and Kaldor[16] have used a Bayesian approach to produce smoothed estimates of relative risk where the extent of smoothing is determined by the magnitude of the observed rate, its precision, and (optionally) the estimated underlying relative risk distribution. Our analysis has shown that unless study is confined either to the larger cancer sites, or to areas with sizeable populations, random variation alone can explain most differences in incidence. The relationship between sample size and random variation for selected confidence intervals is shown in fig 5. Even at the 95% level, at least 60 cases would be required in any one district to prevent random error exceeding 50% of the observed rate. Account must also be taken of the undue influence on rates of very small populations—in order to achieve a maximum rate per case of 1·0 in every local government district in the west of Scotland, nine years' data would be required.

On the basis of 1980–85 cancer incidence data, the following number of years' cases would be required for each site in order to meet a requirement of an average (though not a minimum) of 60 cases per area for an analysis at the level of local government district in the west of Scotland:

1 year	Lung (males)
2–5 years	Breast and lung (females), large bowel, stomach, bladder (males), prostate
6–10 years	Ovary, cervix, bladder (females), oesophagus (males), pancreas
11–20 years	Oesophagus (females), endometrium, leukaemia, non-Hodgkin's lymphoma, melanoma (females), larynx (males), oral cavity and pharynx (males)
20 years or more	Melanoma (males), testis, oral cavity and pharynx (females), thyroid (females), lip (males), Hodgkin's disease, larynx (females), thyroid (males), lip (females)

Table IV Population distribution of local government districts in the west of Scotland and Scotland as a whole (1981 census)

Population	*Scotland*	*West of Scotland*
<10 000	1	0
<25 000	10* (18%)	1 (4%)
<50 000	21 (38%)	6 (23%)
<100 000	40 (71%)	17 (65%)
Total	56 (100%)	26 (100%)

*Sutherland, Skye and Lochalsh, Lochaber, Nairn, Badenoch and Strathspey, Tweeddale, Berwickshire, Stewartry, Orkney, and Shetland

Table III Comparison of observed, simulated and expected highest to lowest incidence rate ratios for local government districts with highest and lowest exposure to selected risks

Tumour site	*Risk factor*	*% Exposed*		*Relative risk*	*Rate ratio*		
		High	*low*		*Exp*	*Sim*	*Obs*
Lung (m)	Ever smoked cigarettes	83	70	9·1	1·2	1·3	2·3
Breast (f)	Body mass index ⩾ 29 kg/m^2	20	13	2·7	1·1	1·4	1·9
Bladder (m)	Employed in pigments factory	3	0	5·3	1·1	1·9	2·4

Exp = expected; Sim = simulated; Obs = observed
Exposure data refer to a sample of west of Scotland districts.[7] Relative risk and threshold value for body mass index in relation to breast cancer were derived from unpublished data for Renfrew and Paisley. Estimates of the proportion of women with body mass index greater than or equal to 29 kg/m^2 were calculated by assuming the same frequency distribution about the mean in the highest and lowest sample district as for Scotland as a whole.[10] Observed and simulated rate ratios are for 1980–85 and are for the west of Scotland.

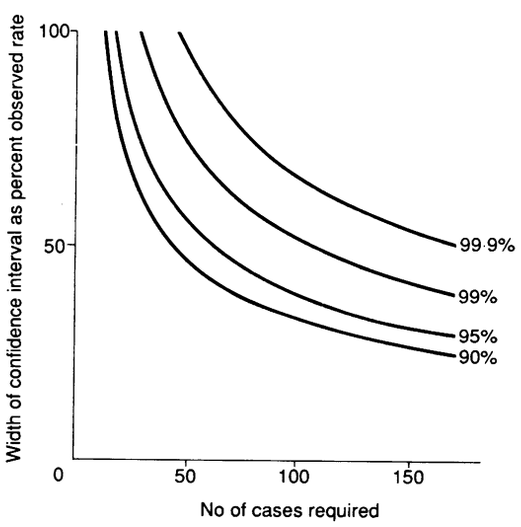

Figure 5 Relationship between random variation and sample size for selected confidence intervals

Obviously for most purposes such amounts of data would be difficult to obtain. There is also a greater chance of confounding changes taking place within too long a time period, although this is likely to be much less serious a problem than the production of meaningless statistics calculated over too short a time scale. Twenty years' data were deemed necessary for the analysis of cancer mortality at county level in the United States.[17]

Observed differences between highest and lowest recorded rates at a local government district level were far in excess of estimated ratios based on actual patterns of exposure to risk (table III). While such excesses were to be expected in the cases of breast and bladder cancer, where the proportions of the total risk attributable to the factors selected were low (18–25 and 0–11% respectively), a substantial difference between observed and expected ratios also occurred in relation to lung cancer, where cigarette smoking accounted for 85–87% of the total risk in all districts for which exposure data were available. The greater the number of cases, the greater was the difference between simulated and observed values, ie, the lower the proportion of random variation present in the observed pattern.

For a given relative risk, levels of exposure would have to be extremely high and the hazard highly localised before the effect of the population attributable risk on incidence rates in small areas was sufficient to exceed random error. Examples of these would include the identification 20 years ago of nasal cancer among woodworkers in High Wycombe[18] and present excesses of mesothelioma in former shipbuilding workers in Clydebank.[19] Employment in a west of Scotland pigments factory the work force of which comprised 3% of the local population carried a relative risk of 5·3 for carcinoma of the bladder in men (table III). This level of risk and exposure would indicate an expected ratio between highest and lowest incidence rates of 1·13, considerably below the simulated ratio for the period 1980–1985 of 1·9, which in turn accounted for almost 80% of the observed ratio of 2·4. In the case of lung cancer, the expected ratio is still slightly below the simulated figure, although random variation accounts for much less of the observed rate ratio.

The conclusions drawn in this paper do not in themselves invalidate the small area approach. It is however essential that the effects of random variation are taken into account. Case-control studies, by considering groups of individuals with a common characteristic rather than examining heterogeneous groups on a geographical basis, will almost always offer a more effective way of testing hypotheses and identifying risks.

We are grateful to Alison Stewart of the Department of Audiovisual Services, Stobhill Hospital, Glasgow for technical assistance in the production of the illustrations. A preliminary report on this study was presented to the 8th Annual Meeting of the Scottish Epidemiology Group, Dundee, June 1988.

The views expressed in this paper are those of the authors alone and are not necessarily those of the Greater Glasgow Health Board.

1 Levine PH, McKay FW, Connelly RR. Patterns of nasopharyngeal cancer mortality in the United States. *Int J Cancer* 1987; **39**: 133–7.
2 Abel U, Becker N. Geographical clusters and common patterns in cancer mortality of the Federal Republic of Germany. *Arch Environ Health* 1987; **42**: 51–7.
3 Kemp I, Boyle P, Smans M, Muir C. *Atlas of cancer in Scotland 1975–80: incidence and epidemiological perspective.* Lyon: IARC, 1985.
4 Moller TR, Anderson H, Olsson H, Jogreus C, Manuswin C, Pretki L. *Cancer incidence in southern Sweden 1983–1987.* Lund: Onkologiskt Centrum, 1989.
5 Alexander FE, Cartwright RA, McKinney PM. A comparison of recent statistical techniques of testing for spatial clustering. In: Elliott P, ed. *Methodology of enquiries into disease clustering.* London: London School of Hygiene and Tropical Medicine, 1989: 23–31.
6 Gillis CR, Hole DJ, Hawthorne VM. Cigarette smoking and male lung cancer in an area of very high incidence II Report of a general population cohort study in the West of Scotland. *J Epidemiol Community Health* 1988; **42**: 44–8.
7 Tunstall-Pedoe H, Smith WCS, Crombie IK, Tavendale R. Coronary risk factor and lifestyle variation across Scotland: results from the Scottish Heart Health Study. *Scot Med J* 1989; **34**: 556–60.
8 Gillis CR, Boyle P, MacIntyre I. Bladder cancer in a pigments factory: environment or occupation? In: *Prevention of occupational cancer—International Symposium.* Occupational Safety and Health Series No 46. Geneva: International Labour Office, 1982.
9 Rohan TE, Bain CJ. Diet in the etiology of breast cancer. *Epidemiol Rev* 1987; **9**: 120–45.
10 Smith WCS, Tunstall-Pedoe H, Crombie IK, Tavendale R. Concomitants of excess coronary deaths—major risk factor and lifestyle findings from 10,359 men and women in the Scottish Heart Health Study. *Scot Med J* 1989; **34**: 550–5.
11 Muir CS, Waterhouse J, Mack T, Powell J, Whelan S. Cancer incidence in five continents. Vol. V Lyon: IARC, 1987.
12 Burkitt DP. Epidemiology of some human tumours. (4) Burkitt's Lymphoma. In: Symington T, Carter RL, eds. *Scientific foundations of oncology.* London: Heinemann, 1976: 232–8.
13 Openshaw S, Charlton M, Craft AW, Birch JM. Investigation of leukaemia clusters by use of a geographical analysis machine. *Lancet* 1988; i: 272–3.
14 Cook-Mozzaffari PJ, Darby SC, Doll R, *et al.* Geographical variation in mortality from leukaemia and other cancers in England and Wales in relation to proximity to nuclear installations, 1969–78. *Br J Cancer* 1989; **59**: 476–8.
15 Roman E, Beral V, Carpenter L, *et al.* Childhood leukaemia in the West Berkshire and Basingstoke and North Hampshire District health authorities in relation to nuclear establishments in the vicinity. *BMJ* 1987; **294**: 597–602.
16 Clayton D, Kaldor J. Empirical Bayes estimates of age-standardised relative risk for use in disease mapping. *Biometrics* 1987; **43**: 671–81.
17 Blot WJ, Fraumeni JF. Geographic epidemiology of cancer in the United States. In: Schottenfeld D, Fraumeni JF, eds. *Cancer epidemiology and prevention.* Philadelphia: WB Saunders, 1982.
18 Acheson ED, Cowdell RH, Hadfield E, MacBeth RG. Nasal cancer in woodworkers in the furniture industry. *BMJ* 1968; ii: 587–96.
19 Gillis CR, Hole DJ, Lamont DW. Incidence of mesothelioma in Glasgow 1981–1984. *J Soc Occup Med* 1990; **40**: 5–10.