A standardisation approach to the control of socioeconomic confounding in small area studies of environment and health

H Dolk, B Mertens, I Kleinschmidt, P Walls, G Shaddick, P Elliott

Abstract

Objective **– To assess how effectively a routine adjustment can be made for socioeconomic confounding in small area studies of environment and health using indirect standardisation and small area deprivation indices, including analysis of the appropriate size of population unit on which to base the deprivation index and the importance of region and urban/rural status as axes of stratification.**

Method **– Standardised morbidity ratios were calculated for cancers in Great Britain for 1981 and standardised mortality ratios for all cause mortality in Great Britain betweeen 1982 and 1985. Deprivation indices were calculated for enumeration districts and wards from 1981 small area census statistics. Cancers and deaths were allocated to enumeration districts via their postcode. Standardised morbidity and mortality ratios were calculated by quintile of enumeration district according to the deprivation index. Standardised mortality ratios were further analysed by deprivation of ward, region, and urban/rural status.**

Results **– Strong relationships were found between all cause mortality and the incidence of selected cancers and deprivation quintile – there was up to a twofold difference in lung cancer incidence between the highest and lowest quintile. The deprivation index can be used to measure gradients of deprivation according to the distance from industrial sites. The deprivation index for enumeration districts showed similar discrimination of mortality as the index for wards. There is some interaction between deprivation and region in their effect on the standardised mortality ratios, leading to a small bias in the estimation of expected numbers if this is not taken into account. The relationship between deprivation, urban/rural status, and mortality is complex and confounded by region, but mortality tends to be higher in urban than in rural areas within quintiles of deprivation.**

Discussion **– Whether calculated for enumeration districts or wards, the main problems in the interpretation of the deprivation index may be its limited correlation with the risk factors of interest and its concentration on present rather than past socioeconomic status. Indirect standardisation based on stratification for deprivation and other variables involves a trade off between bias and precision in determining the fineness and the number of axes of stratification. Some bias may occur due to interaction between region and deprivation and the effects of urban/rural status. Complementary approaches including modelling and proportional mortality or morbidity analyses may be needed and the possibility of residual socioeconomic confounding must always be considered.**

Conclusion **– There is potential for important socioeconomic confounding in small area studies of environmental pollution and health where the health outcome under examination has a strong relationship to socioeconomic status and where the putative excess risk due to pollution may be small. One method of controlling for confounding is to use an ecological measurement of deprivation in small areas, and to adjust for deprivation by indirect standardisation. However, residual socioeconomic confounding can be expected, which may seriously complicate the interpretation of small area studies.**

(*J Epidemiol Comm Health* 1995;**49**(Suppl 2):S9–S14)

Small Area Health Statistics Unit, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, London
H Dolk
B Mertens
I Kleinschmidt
P Walls
G Shaddick
P Elliott

Correspondence to:
Dr H Dolk.

It is now well established that deprivation as measured from census variables for small areas, is strongly related to a variety of health outcomes, including mortality and cancer incidence.[1] Historically, it tends to be the more deprived groups in the community who live near "dirty" industry. This suggests that when we examine health statistics around a source of pollution, we may find more ill health closer to the source than farther away, regardless of whether there are adverse health effects related to the toxic emissions themselves. This socioeconomic confounding is a major problem in the design and interpretation of small area studies of environmental pollution and health, especially as any risk related to typical levels of environmental pollution will tend to be small and may be swamped by the effects of deprivation.

The Small Area Health Statistics Unit (SAHSU) was established to assess from routinely collected health data whether there is evidence of adverse health effects among the local population near an industrial source.[2] SAHSU has adopted a common methodology for these studies which incorporates control of socioeconomic confounding using indirect standardisation.[2-6]

The SAHSU methodology is founded on the comparison of observed and expected numbers of a given health outcome in small areas defined by distance from a source of pollution. These numbers are calculated using a postcoded database of all registered cancers, deaths and births occurring in Great Britain (GB). This database is linked to GB small area census statistics. "Observed" numbers are counts of health events among persons with postcodes falling within the study area. "Expected" numbers are calculated using national incidence and mortality rates, stratified by year, age, sex and, deprivation quintile. The deprivation quintile is a regrouping of the deprivation index calculated for small areas (enumeration districts) following the method described by Carstairs.[1 7] A further adjustment is made using a "regional adjustment factor", which measures any excess or deficit in the region as a whole following the above standardisation procedure.

In this paper, we illustrate the potential for socioeconomic confounding and discuss how effectively a routine adjustment can be made for this confounding using deprivation indices and indirect standardisation. Specifically, we analyse all cause mortality data to assess how well the deprivation index performs when applied to small area "building blocks" (enumeration districts) compared with larger areas such as electoral wards, and to estimate the magnitude of the interaction between deprivation, region, and urban/rural status.

Methods

Analyses presented here use a subset of the postcoded SAHSU database of all registered cancers and deaths in GB. Regional health authority (here referred to as "region") of residence refers to regions existing before the reorganisation of boundaries in 1994.

A deprivation index was calculated for each enumeration district, using 1981 small area census statistics. Enumeration districts contain on average around 400 persons. The index has four components, as described by Carstairs[7]: access to a car, overcrowding, unemployment and social class. Each component is standardised across GB to have zero mean and unit variance. The index is the sum of the four scores. Enumeration districts were further regrouped into quintiles of deprivation. A sixth "unclassifiable" category contains all those enumeration districts with fewer than 50 people, comprising 0·8% of the GB population. This category tends to include enumeration districts containing institutional populations. The social class component was calculated for wards and applied to all enumeration districts within the ward, since social class was coded for only a 10% sample of the GB population in 1981 and enumeration districts were therefore too small for the calculation of this component. Due to a difference between Scotland and England and Wales in the classification of kitchens in the 1981 census, affecting the "overcrowding" component of the index, the overcrowding component was multiplied by a factor of 0·72 in Scottish enumeration districts, following the analysis of this question by Bajekal (personal communication).

A deprivation index was also calculated for all wards in GB, using the same methodology. The proportion of the GB population living in unclassifiable wards was 0·1%. Except where stated, the enumeration district deprivation index was used.

For England and Wales, an Office of Population Censuses and Surveys (OPCS) urban/rural flag for wards[8 9] was regrouped into two categories, grouping wholly and mostly urban areas as one category, and mixed urban/rural to wholly rural areas as a second category. Enumeration districts were classified as urban or rural according to the ward in which they were situated. Scottish enumeration districts were classified in the same way, using the urban/rural classification available in the census small area statistics. For GB, the rural category thus obtained encompassed 14% of the 1981 population.

Standardised morbidity ratios and standardised mortality ratios (SMR) were calculated for cancers (1981) and deaths (1982–85) respectively. GB rates stratified by age (five year age group) and sex were used in this standardisation. A longer period (1982–85) was analysed for deaths in order to increase the sample size to a level where statistical significance would cease to be an issue in most comparisons of SMR by deprivation, region, and urban/rural status. The number of deaths is shown in the table as an indication of sample size. The total number of deaths was 2 564 453.

Interaction between region and deprivation was analysed in a way which could be directly related to current SAHSU methodology. The regional adjustment factor was calculated as the ratio of observed to expected deaths in the region between 1982 and 1985, on the basis of national rates stratified by age, sex, and deprivation quintile. This is directly equivalent to the regional adjustment factor used in SAHSU studies except that it is not year and sex specific, and the sixth "unclassifiable" category was excluded from the calculation. Regional "multipliers" in each deprivation quintile for the same period were calculated as the ratio of observed to expected deaths, based on national deprivation quintile specific age and sex stratified rates.

Results

The potential for socioeconomic confounding in small area analyses of environment and health is illustrated in figures 1 and 2. Relationships between the deprivation index and 1982–85 all cause mortality and the 1981 incidence of specific cancers are shown in figure 1. For lung cancer, there was a twofold difference in risk between the lowest and highest quintile of deprivation (there was a higher risk in more deprived areas). In contrast, skin melanoma shows the well known inverse relationship between deprivation and incidence, while the leukaemia incidence shows little, if

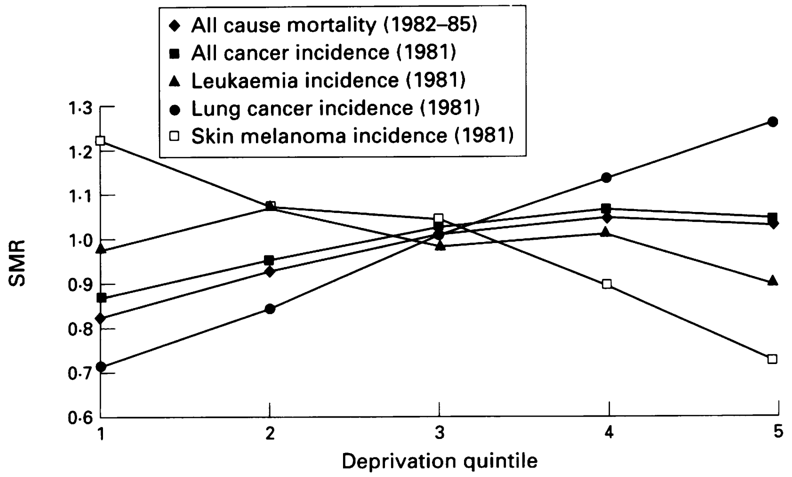

Figure 1 Relationship between quintile of deprivation index and standardised mortality or morbidity ratio (SMR) for selected outcomes, Great Britain.

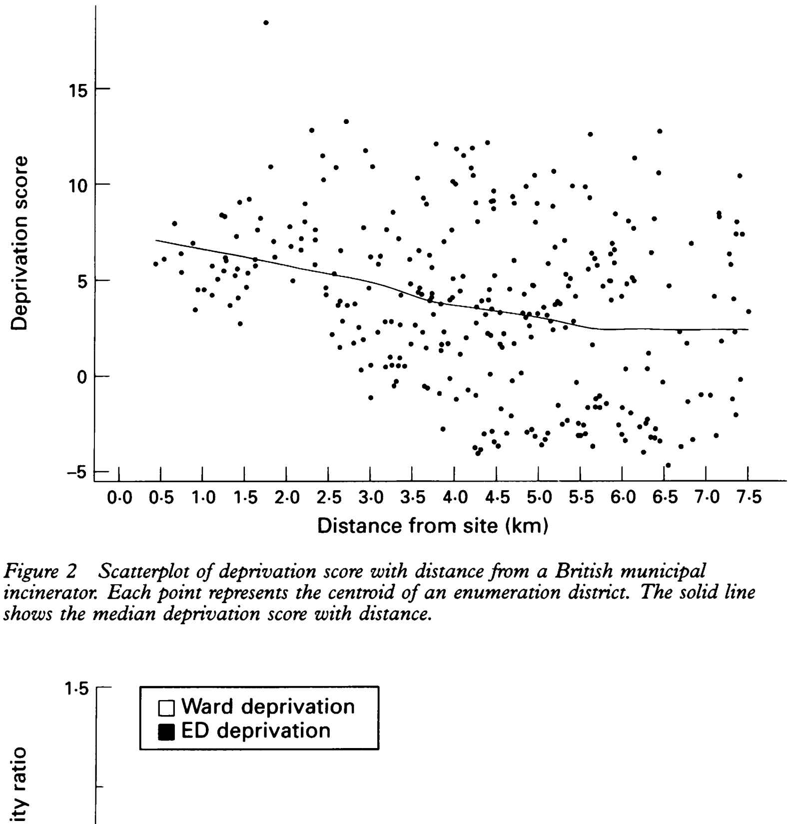

Figure 2 Scatterplot of deprivation score with distance from a British municipal incinerator. Each point represents the centroid of an enumeration district. The solid line shows the median deprivation score with distance.

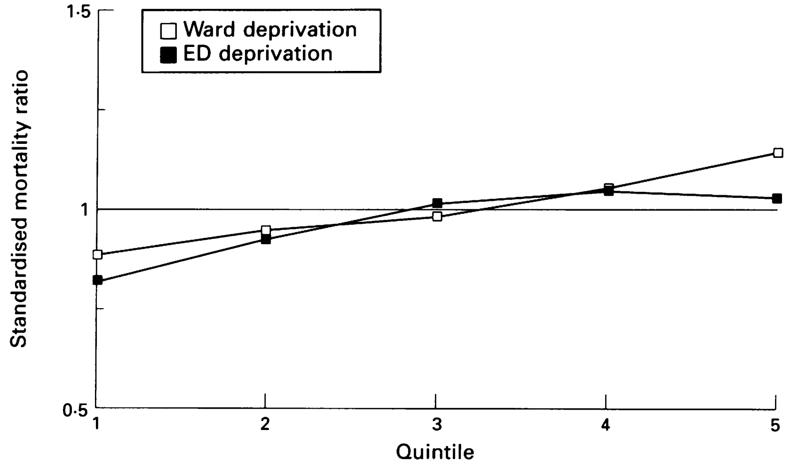

Figure 3 All cause standardised mortality ratio (SMR) by quintile of enumeration district (ED) deprivation and ward deprivation, Great Britain 1982–85.

any, relationship. The strength of the relationship can vary according to age. For example, for all cause mortality, the deprivation gradient is steepest for the 0–4 years age group, with twice as many deaths in the most deprived quintile than in the most affluent quintile, and most shallow for the 75+ age group, where the most deprived quintile has 20% more deaths than the most affluent.

Figure 2 illustrates the use of a deprivation index in measuring a gradient of deprivation with distance from a municipal incinerator, in this case with a clear decrease in median deprivation level with increasing distance. The existence, shape, size, and direction of a gradient depends on the characteristics and history of a site, and of the local area. In a study of 51 municipal incinerators,[4] the median deprivation score in the first 2 km was in a more deprived quintile than the median from 2–7·5 km in 35%, in the same quintile in 47% and in a more affluent quintile in 18%. Only one site was in a rural ward. A current study of radiotransmitters suggests a quite different potential for socioeconomic confounding, with the possibility of strong reverse gradients and more frequent rural siting.

Figure 3 shows how the SMR between 1982 and 1985 varies according to quintile of population as measured by enumeration district deprivation or ward deprivation. The ratio of SMR in the highest to lowest quintile (or the slope of the deprivation gradient) is slightly higher for wards than for enumeration districts.

The table gives the distribution of people and deaths in relation to region and enumeration district deprivation. The proportion of the population in each region that lives in an electoral district in the most deprived quintile varies from 2·9% (SW Thames) to 41·9% (Scotland), while the proportion in the most affluent quintile varies from 9·3% (Scotland) to 43·6% (SW Thames), generally demonstrating the north-south divide.

The regional adjustment factor is compared with regional multipliers in the most deprived and most affluent quintiles in the table. Differences in these multipliers between quintiles within a region reflect interaction of deprivation and region in their effect on all cause mortality. Except in Scotland, the region multiplier was lower in the most deprived quintile than in the most affluent quintile, such that in the more affluent regions, the positive regional effect was greater in deprived areas, and in more deprived regions, the negative regional effect was greater in affluent areas. The regional adjustment factor was always within 7% of the multiplier in either of the extreme quintiles.

Overall, rural communities had 7% lower mortality than the national average (not shown). However, there was considerable interaction between region, deprivation, and urban/rural status in their effects on mortality, and the deprivation profile of rural populations differed considerably between regions. Within regions, the excess mortality in urban areas rarely exceeded 10%, although quintile specific differences of up to 21% were observed. The general exception to the pattern of lower rural mortality was Wales, where mortality was 4–6% higher in rural areas in all but the most affluent quintile.

Percentage of region population and regional multipliers in deprived and affluent quintiles of enumeration districts and regional adjustment factor for all cause mortality, Great Britain, 1982–85.*

Regional health authority	% of region population in quintile of enumeration districts which is . . .		Total no of deaths in region	Total (ie regional adjustment factor)	Regional multipliers	
	Most deprived	Most affluent			Most affluent	Most deprived
West Midlands	25·2	17·1	228 320	1·01	1·05	0·96
Northern	33·1	13·0	153 225	1·05	1·08	1·02
Yorkshire	24·0	15·9	175 715	1·03	1·07	1·00
Trent	17·4	17·5	210 804	1·00	1·06	0·96
East Anglia	4·8	23·8	86 152	0·94	0·97	0·88
NW Thames	11·9	27·4	139 110	0·96	0·95	0·92
NE Thames	20·3	21·1	166 287	0·94	0·96	0·92
SE Thames	13·4	23·9	178 897	0·95	0·96	0·94
SW Thames	2·9	43·6	138 355	0·96	0·95	0·92
Wessex	4·8	27·5	129 908	0·95	0·96	0·88
Oxford	7·7	38·3	87 792	0·98	1·00	0·92
South Western	5·6	24·9	153 926	0·96	0·95	0·93
Mersey	34·0	19·6	116 561	1·04	1·04	1·01
North Western	27·1	16·1	207 207	1·07	1·09	1·04
Wales	17·7	14·0	139 549	1·03	1·06	0·97
Scotland	41·9	9·3	252 645	1·07	1·03	1·11

* The ratio of observed to expected deaths in the region, based on GB age-sex-deprivation specific rates.

Discussion

It is clear that there is a great potential for socioeconomic confounding in small area studies of environment and health, although the extent of potential confounding will depend on the disease(s) and type(s) of industrial site under consideration. Current SAHSU methodology for control of socioeconomic confounding involves the ecological measurement of deprivation, and indirect standardisation using national rates stratified for this deprivation index. Both these elements need to be evaluated in order to assess the effectiveness of this methodology.

For our purposes, the social or biological meaning of "deprivation" is not really of crucial importance. The aim is to estimate as closely as possible how much of a given health outcome we would expect in an area in view of its socioeconomic characteristics, irrespective of any causal mechanisms underlying socioeconomic variation in health. Nevertheless, to ignore causal mechanisms might potentially lead us to "over control" for socioeconomic confounding. This might occur if the relationship between deprivation and health were at least partly explained by the tendency for more deprived groups to have higher exposure to the types of industrial emissions being investigated. This becomes less problematic, however, the more specific the exposure-disease association under study, the stronger that association, and the rarer the exposure.

One of the issues in the ecological measurement of deprivation is the choice of the size of population unit on which to base this measurement. Most small area deprivation indices have been developed for classifying census wards or postcode sectors.[1 6 7] In SAHSU studies the index has been adapted for enumeration districts, some 10 times smaller in population, and closer to the postcode "building blocks" reaggregated to define exposure areas. As discussed also by Carr-Hill in this volume,[10] it is not obvious that reducing the population unit will automatically lead to a better measure of deprivation. We might expect the smaller unit to contain a more socioeconomically homogeneous population, and the deprivation index to give a better reflection of the socioeconomic status of most people in the area. There are, however, potential disadvantages of using very small population units. The characteristics of the larger surrounding area may be important in assessing the likely health experience of the smaller population (whether because of access to services or other factors). Applied to the smaller area, the index may be undesirably sensitive to local variations in age structure. For example, in an area of predominantly elderly people, two of the components of the Carstairs deprivation index, access to a car and overcrowding, might take on a different meaning in relation to deprivation than in an area of predominantly young families. Enumeration districts also vary in size, with the index becoming unstable when referring to a very small number of households, and with a tendency for the component variables to be skewed in their distributions. In our comparison of the level of discrimination in mortality which could be obtained from the enumeration district and ward measures, we found that the enumeration district measure performs slightly worse. Use of an enumeration district based deprivation index nevertheless seems reasonable, especially as in studies of point source exposures, aggregation will take place over many enumeration districts.

The deprivation index measures the relative socioeconomic level of the study population in the census nearest in time to the deaths or cancer diagnoses being counted. However, mortality and morbidity are influenced by past as well as present socioeconomic status. For cancer, it is well established that the latent period before clinical diagnosis may be long, although socioeconomically related risk factors could act at any stage of the multistage carcinogenic process. There is evidence that early life factors may be important for other disorders such as cardiovascular disease.[11] We would tend to undercontrol for socioeconomic confounding using this deprivation measure if the relative deprivation of the population near the industrial source had been greater in the past. Historic tendencies for areas to remain stable in their relative deprivation level would often reduce this problem, but the issue requires consideration for the particular small areas chosen for study.

Even as a measure of present deprivation, the index may be sensitive to local conditions such as rapid changes in unemployment level, although the use of multiple indicators within the deprivation index tends to prevent severe fluctuation.

Most importantly, the deprivation index is only a proxy for the risk factors of interest, and limited correlation between the deprivation index and underlying risk factors is to be expected. For example, much of the geographical variation in lung cancer incidence may be explained by differences in smoking rates, and a direct measure of smoking, were it available, might be more powerful than its proxy, deprivation.[12]

The method of indirect standardisation using deprivation stratification has the advantage of being quite transparent and computationally simple, and of allowing for full interaction with the other axes of stratification (age, sex, and year). However, in a routine system such as SAHSU, stratification involves a trade off between bias and precision. The finer the strata (for example, deciles rather than quintiles of deprivation), and the greater the number of axes of stratification (such as region and urban/rural status), the less bias in the estimation of the expected numbers will result. However, reduction in bias comes at the expense of greater imprecision in the calculation of national stratified rates, particularly for rare diseases where stratified numbers become very small. In order to examine the issue of bias resulting from simple stratification, we have looked at the effect of region and urban/rural status, and their interaction with deprivation, on mortality. We have concentrated on the magnitude of the effects, rather than any nominal statistical significance, because when dealing with national mortality data even small differences are highly statistically significant, yet of little importance in this context.

As shown by the regional adjustment factor, considerable regional differences in mortality remain after adjustment for measured deprivation. This was analysed in more detail by Eames *et al*,[13] who found that regional differences within England in all cause mortality for the under 65 years age group persisted at any one level of deprivation. Carstairs and Morris,[14] on the other hand, found that the mortality differential between Scotland and England and Wales 1980–82 was almost entirely explained by the greater level of deprivation in Scotland. Three implications of using an adjustment factor for region rather than full stratification for region should be noted. Firstly, region may be to a small extent a confounder in the relationship between deprivation and mortality, since different regions are represented unequally in the deprivation spectrum. Secondly, use of the same regional adjustment factor might not be appropriate where studies are restricted to young people, since regional differences have been shown to be smaller in younger age groups.[15] Thirdly, this method does not take into account any interaction between deprivation and region. We have shown here that some interaction between region and deprivation does occur, in agreement with Eames *et al.*[13] Any bias that would result from this interaction would generally be small in multiple site studies covering areas of mixed deprivation level, but could reach a maximum of 7% in some regions and for areas with extreme deprivation profiles. Thus, use of regional rates directly may be preferred, particularly when not dealing with rare diseases, or when dealing with diseases where the interaction between region and deprivation is greater than that found for all cause mortality.

The effect of urban/rural status and its interaction with deprivation may be complex. Firstly, rural areas may have a consistently better or worse health experience at any one level of deprivation, perhaps because of other aspects of rural life which influence mortality or morbidity. Secondly, the impact of deprivation on health may differ in rural and urban areas i.e. the deprivation gradient may be more or less steep in rural areas. Thirdly, the deprivation indicators themselves (social class of head of household, access to a car, overcrowding, and unemployment) may take on a different meaning in relation to deprivation in rural areas. Thus discrimination between different levels of deprivation may differ from that in urban areas, also affecting the observed gradient. It is difficult to unravel this complexity by analysis of available data, but in the present context it is simply necessary to determine whether information on urban/rural status may be an important additional consideration in the control of socioeconomic confounding. It is clear from our analysis of mortality that it is not possible to take into account urban/rural status without also taking into account region. When taking the full interaction between deprivation, urban/rural status and region into account, most of the effects of urban/rural status on mortality were less than 10% but some were more important. Full stratification for deprivation, region, and urban/rural status would seem impractical given the small size of many rural cells. It is perhaps fortunate that most studies of industrial sources involve almost entirely urban populations, but it would seem that the urban/rural character of the population and the potential importance of urban/rural status to the health outcome of interest should be considered for each study.

Since small area environmental studies are often concerned with the detection of low levels of excess risk, the persistence of any bias that may result from too simple stratification may require the development of modelling as an additional tool, as discussed by Bithell in this volume,[16] allowing more flexibility in dealing with a range of confounders and their interaction.

Whatever approach is used to measure socioeconomic status and stratify the national rates or model the data, it is likely that at least some residual confounding may remain which may seriously complicate the interpretation of a small area study. Other approaches are also needed. These could include proportional mortality/morbidity[4] or case-control approaches comparing health outcomes that have similar relationships to deprivation and to underlying risk factors such as smoking, and comparison of preoperational and postoperational periods for industrial sites.[4] In a recent study looking at municipal incinerators,[4] it was found that control for socioeconomic confounding reduced the size of the excess risk near municipal incinerators for selected cancers. After control for confounding, a gradient in risk of selected cancers with distance from the sites persisted, but this was found both before and after the start of operation of the incinerators, indicating the extent of residual socioeconomic confounding. Thus, where the potential for socioeconomic confounding is likely to be important, control for and interpretation may be best effected by a variety of approaches, and the persistence of residual confounding must always be considered.

1 Carstairs V. The use and interpretation of deprivation indices in relation to health. *J Epidemiol Community Health* 1995; **49**(Suppl 2):S3–8.
2 Elliott P, Westlake AJ, Kleinschmidt I, *et al.* The Small Area Health Statistics Unit: a national facility for investigating health around point sources of environmental pollution in the United Kingdom. *J Epidemiol Community Health* 1992;**46**:345–9.
3 Elliott P, Hills M, Beresford J *et al.* Incidence of cancer of the larynx and lung near incinerators of waste solvents and oils in Great Britain. *Lancet* 1992;**339**:854–8.
4 Elliott P, Shaddick G, Kleinschmidt I, Jolley D, Walls P, Beresford J, Grundy C. Cancer incidence near municipal solid waste incinerators in Great Britain. *Br J Cancer* (in press).
5 Elliott P. Small area studies. In: *Proceedings of an international workshop: setting priorities in environmental epidemiology.* Rome, 28–30 January, 1993 (in press).
6 Jolley DJ, Jarman B, Elliott P. Socioeconomic confounding. In: Elliott P, Cuzick J, English D, Stern R, eds. *Geographical and environmental epidemiology: methods for small area studies.* Oxford: Oxford University Press, 1992;115–24.
7 Carstairs V, Morris R. *Deprivation and health in Scotland.* Aberdeen: Aberdeen University Press. 1991.
8 Denham C. Urban Britain. *Population trends* 1984;**36**:10–18.
9 Craig J. An urban-rural categorisation for wards and local authorities. *Population Trends* 1987;**47**:6–11.
10 Carr-Hill R, Rice N. Is enumeration district level an improvement on ward level analysis in studies of deprivation and health? *J Epidemiol Community Health.* 1995;**49**(Suppl 2):S28–S29.
11 Barker DJP ed. Fetal and infant origins of adult disease. London: *BMJ*,1992.
12 Kleinschmidt I, Hills M, Elliott P. Smoking behaviour can be predicted by neighbourhood deprivation measures. *J Epidemiol Community Health.* 1995;**49**(Suppl 2):S00–00.
13 Eames M, Ben-Shlomo Y, Marmot MG. Social deprivation and premature mortality: regional comparison across England. *BMJ* 1993;**307**:1097–102.
14 Carstairs V, Morris R. Deprivation: explaining differences in mortality between Scotland and England and Wales. *BMJ* 1989;**299**:886–9.
15 Illsley R, Le Grand J. Regional inequalities in mortality. *J Epidemiol Community Health* 1993;**47**:444–9.
16 Bithell JF, Dutton SJ, Neary NM, Vincent TJ. Use of regression methods for control of socioeconomic confounding. *J Epidemiol Community Health* 1995;**49**(Suppl 2):S15–19.

Open discussion

WILKINSON – Dr Dolk, you mention there may be some circumstances in which there could be "over control" for socioeconomic confounding in point source studies. How can you easily judge what is the appropriate level of control in those circumstances?

DOLK – It is not easy to judge this but you have to think about it in terms of the particular outcome and pollution source that you are interested in. It depends how much of the national relationship between the outcome or disease of interest and deprivation is explained by pollution. Then it depends if those close to the pollution source experience more of the relevant pollution than average in their deprivation group. Probably potentially the worst situation will come with very common sources of pollution like roads.