LUNG CANCER

Does living near heavy industry cause lung cancer in women? A case-control study using life grid interviews

R Edwards, T Pless-Mulloli, D Howel, T Chadwick, R Bhopal, R Harrison, H Gribbin

....

See end of article for authors' affiliations
....

Correspondence to:
Dr R Edwards, Department of Public Health, Wellington School of Medicine and Health Sciences, University of Otago, P O Box 7343, Wellington, New Zealand; Richard.Edwards@otago.ac.nz

Background: The incidence of lung cancer among women is high in the highly industrialised area of Teesside in north-east England. Previous research has implicated industrial pollution as a possible cause. A study was undertaken to investigate whether prolonged residence close to heavy industry is associated with lung cancer among women in Teesside.

Methods: Two hundred and four women aged <80 years with incident primary lung cancer and 339 age matched community controls were recruited to a population based case-control study. Life course residential, occupational, and active and passive smoking histories were obtained using an interviewer administered questionnaire.

Results: The age adjusted odds ratio (OR) for lung cancer among people living >25 years v 0 years near (within 0–5 km) heavy industry in Teesside was 2.13 (95% CI 1.34 to 3.38). After adjustment for confounding factors the OR was 1.83 (95% CI 0.82 to 4.08) for >25 years or 1.10 (95% CI 0.96 to 1.26) for an additional 10 years living near industry. ORs were similar after residence near heavy industry outside Teesside was also included, and when latency was allowed for by disregarding residential exposures within the last 20 years. Adjustment for active smoking had the greatest effect on the OR.

Conclusions: This population based study using life grid interviews for life course exposure assessment has addressed many deficiencies in the design of previous studies. The findings support those in most of the international literature of a modestly raised risk of lung cancer with prolonged residence close to heavy industry, although the confidence intervals were wide. The effect of air pollution on the incidence of lung cancer merits continued study.

While tobacco smoking is accepted as the main causal factor for lung cancer, there is a long standing debate about the contribution of ambient air pollution from industrial, domestic, transport related, or other sources.

Teesside is an area of 360 square miles in the north-east of England with a population of about 464 000. The iron and steel, chemical, and heavy engineering industries expanded rapidly through the 19th and early 20th centuries.[1] By 1945 Billingham on Teesside was the largest single chemical production complex in the world. Houses for the workforce were built as close as possible to the industrial sites,[2] leading to long standing concerns about the effects of industrial air pollution on health.[3–6]

Local ecological studies in the 1960s and 1970s partially supported the hypothesis that industrial air pollution contributed to the causation of lung cancer on Teesside by finding higher lung cancer mortality in urban areas than rural areas.[7 8] However, exposure to air pollution and living close to heavy industry was poorly characterised—for example, exposure status was defined only by place of residence at time of death.

Teesside also has a long standing history of poverty and deprivation. Recent studies[9–11] found that some urban wards (a UK administrative area with a population of 5000–10 000) of Middlesbrough were among the most deprived with the poorest health in the north-east region, an area with many highly deprived communities. Lung cancer mortality among women from deprived wards in Teesside was particularly high, and the observed excess of deaths from cancers and respiratory diseases was hypothesised as compatible with environmental pollution compounding the effects of deprivation.[10]

The stimulus for the current study was the Teesside Environmental Epidemiology Study which compared morbidity and all-cause and cause specific mortality rates for populations from similarly socioeconomically deprived localities.[6 12 13] This study found little evidence that living close to industry was associated with morbidity, birth outcomes, or most measures of mortality.[12] However, it found increased lung cancer mortality in women from communities living close to heavy industry compared with women from similar communities living further from industry on Teesside (directly standardised rate ratio 1.5).[12 13] Community surveys suggested that this increase was not the result of differences in occupational exposure, smoking history, or socioeconomic circumstances. The authors concluded that industrial pollution was a plausible causative factor and recommended further investigation.

Views differ about the prospects for determining whether residential exposure to industrial air pollution causes lung cancer.[14 15] Methodological difficulties include the weak to moderate strength of observed associations, low and declining levels of exposures, difficulties in measuring individual exposures retrospectively, and adjusting for multiple confounding factors. Pending the development of adequate biomarkers of exposure, some authors have argued that well designed case-control studies with improved methods for retrospective assessment of exposure to industrial pollution and potential confounding factors should be used.[14 16]

This paper describes the Teesside Lung Cancer Case Control Study (TLCCCS) which investigated whether lung cancer among women on Teesside is associated with prolonged residence close to heavy industry.

METHODS

Study design and subjects

A population based case-control study was undertaken with concurrent sampling of controls. Eligible cases were women aged under 80 years with incident primary malignant lung

cancer (adenocarcinoma, squamous cell carcinoma, small cell cancer, large cell cancer, or other malignant lung tumours) and currently resident in the Teesside district. Exclusions included benign lung tumours or tumours of pleural origin; clinical uncertainty over the diagnosis of lung cancer; suspected secondary lung cancer; and if the clinical team considered a woman too unwell to take part.

We aimed to maximise identification and recruitment of eligible cases from the source population. Local clinicians suggested that almost all incident lung cancer cases among Teesside residents would present or be referred to chest physicians at Teesside's two main hospitals where the study was based. Cross checking of cancer registry data with clinical databases and records during a 1 year period confirmed this, with 94% (125/133) of incident female lung cancer cases aged <80 years on the cancer registry lists under the care of chest physicians at the two hospitals.

The diagnosis of lung cancer was based on symptoms and signs, radiographic evidence, and bronchoscopic findings, supported wherever possible by cytological or histological evidence. Chest physicians and lung cancer nurse specialists at the two hospitals identified incident cases, informed patients about the study, and carried out a preliminary assessment of eligibility and willingness to participate. We informed the family doctor (GP) and, provided their GP did not object, the study's research nurse asked eligible women to take part.

We aimed to recruit two population based controls per case, drawn at random from the Teesside district population as each new case was diagnosed. The sampling frame was people registered with GPs on Teesside (almost all people in the UK are registered with GPs). Exclusion criteria mirrored those for the cases, with the addition that controls with a history of lung cancer were ineligible.

We randomly selected two index controls per case matched by 5 year age group and reserve controls from the same age group. We used reserve controls from the same electoral ward as the index controls to prevent potential selection bias due to differential participation rates between wards. Controls were invited to take part provided their GPs did not object. If the index control was deemed ineligible by her GP, did not want to participate, or did not reply, successive reserve controls were approached until one agreed to participate. We aimed to identify and interview matched controls shortly after each case. However, for practical reasons, towards the end of the study some controls were recruited on the basis of the observed age distribution and later matched to an appropriate case.

Data collection

A research nurse collected data from participants on the exposure of interest and potential confounding factors by a structured interview incorporating a life grid approach—a method for enhancing recall of relevant life events and behaviours by using temporal reference line(s) of significant personal, family, or external dates and events to prompt and structure the recall of the information of interest.[17 18] There is a theoretical justification and some empirical evidence[17 19–22] to suggest that life grid methods aid the process of memory construction, particularly for dating and recall of duration of temporally distant exposures. Such methods have been used in epidemiological studies requiring retrospective recall over long time periods.[23–26]

The life grid interview began with the construction of a timeline of important personal and external events; full residential, occupational, and smoking histories were then constructed with cross checking to dates within the existing timelines. Additional data collection occurred using a structured questionnaire. Data collection began in January 2000 and ended in February 2004.

Exposure assessment

For residential exposure assessment we divided Teesside into three zones based on residential proximity to current and historical distributions of heavy industry (fig 1). This was guided by a validation study using data from historical records, current or recent routine monitoring data, and air quality surveys carried out for the Teesside Environment Epidemiological Study.[6] Zone A was close to (<5 km) current and historical locations of heavy industry. It encompassed an area close to the river Tees in central and east Middlesbrough, Stockton-on-Tees, the chemical plant complexes at Billingham, Seal Sands and Wilton, and the iron and steel works at Lackenby and Dormanstown. Zone B was intermediate (5–10 km) from heavy industry. Zone C was >10 km from heavy industry and included suburban areas of Middlesbrough and Stockton-on-Tees, and more rural settlements and outlying towns across Teesside. The zone boundaries also incorporated proximity based zones around a large chromium works to the south of Stockton-on-Tees that had been operational since the 1920s.

For all Teesside residences of all subjects we identified a six figure grid reference from current or historical maps (for residences in streets which had been demolished). The grid reference was used to assign each Teesside residence to an exposure zone using a 1:50 000 scale map. Residences outside Teesside were categorised as close to (zone D) or not close to (zone E) heavy industry according to subjects' self-report of whether the house was within a mile (1.6 km) of heavy industrial plant. To check for selection bias, we also identified a grid reference and assigned an exposure zone to current addresses for non-included cases and controls.

Information on lifetime exposure to second hand smoke (SHS) was collected for domestic, workplace, social, and other settings. For domestic exposure we used an approach similar to that suggested by Cummings et al.[27] Information was obtained about duration of co-residence, smoking, and mean numbers smoked for each regular smoker who lived in the same house as the subject and smoked indoors. For workplace exposure to SHS, subjects were asked if they were regularly exposed (defined as ≥2 hours per week most weeks for ≥6 months) to people smoking in the workplace throughout their occupational history. If the response was positive, further information was sought on the frequency, intensity, typical daily duration of exposure, and years of exposure. This method has been validated against nicotine levels measured in personal breathing zones.[28] Exposures to SHS through social activities and in other public places were assessed similarly.

Socioeconomic status was assessed using life course measures based on occupation and housing tenure. Occupations were coded to the 1990 Registrar General's classification using standard coding books.[29 30] We also used an area based deprivation index—the Townsend score[11]—which is a composite score based on four census variables using 1991 census data (unemployment, housing tenure, access to car, overcrowding). A higher score indicates a more deprived area.

We assessed occupational exposure by asking subjects if they were heavily exposed to dusts, fumes, or chemicals in any job. Where this was the case or the job was known to have a high risk of exposure to asbestos or other lung carcinogen, the interviewer enquired about the nature, intensity, frequency, and duration of possible exposures and job tasks. In the case of asbestos exposure, the interviewer used a list of local employers/factories with a high risk of exposure to asbestos. Two members of the study
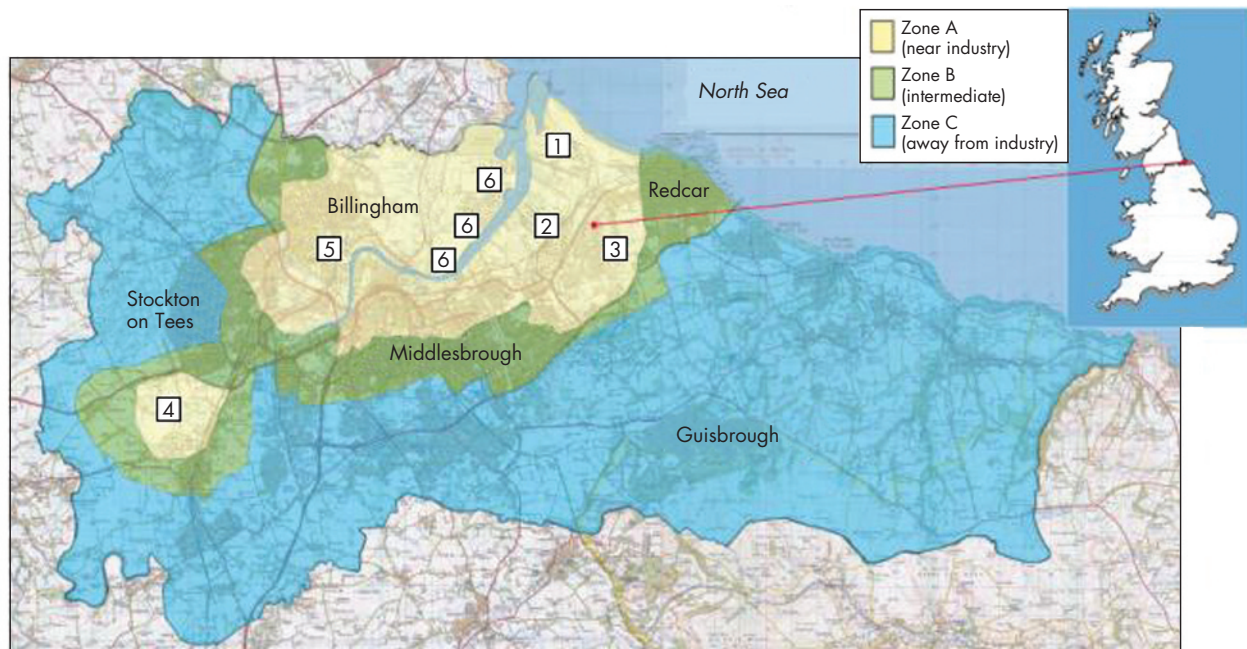
Figure 1 Teesside Lung Cancer Case Control Study exposure zones and current distribution of heavy industry. Teesside is the shaded area. 1, Dormanstown steelworks; 2, Lackenby steelworks; 3, Wilton chemical plant complex; 4, Chromium works; 5, Billingham chemical plant complex; 6, Seal Sands chemical and refinery complex.

team (RE, TPM) blind to the case-control status of the subject independently coded exposures as: no/minimal exposure; possible exposure; probable exposure; or definite exposure to asbestos or other lung cancer carcinogens. Where there were disagreements or uncertainty, these were resolved by discussion or further review of the literature. Exposure measures used in the analysis are available in the online appendix available at http://www.thoraxjnl.com/supplemental.

Power calculation

Assuming 30% of controls were exposed to prolonged residence close to heavy industry, 183 cases with two matched controls per case would provide 80% power to detect an exposure odds ratio (OR) of 1.7.

Analysis of data

Analyses were conducted using Stata version 8.0 (Stata Corporation, College Station, TX, USA). Conditional logistic regression was used to examine the relationship between case-control status and years of residence close to heavy industry after controlling for potential confounding factors. We built up a model for the confounding factors in stages. Stepwise regression was used to find the best choice of variables and functional forms to describe the active smoking

exposure, then occupational exposures (in addition to terms already in the model), and so on to include terms describing socioeconomic variables and finally passive smoking. We considered various functional forms for continuous covariates using fractional polynomial regression.[31]

Age was entered as a continuous variable in all models. The best fitting model for smoking included: ever smoker (Y/N), pack years smoked (square root and linear term), and years stopped smoking (= 0 for current and never smokers). In addition, current marital status, early adult occupational status, and number of years of asbestos exposure were retained in the full models. No other socioeconomic or occupational or SHS exposure variables improved the fit of the model. We then added years lived close to heavy industry to the models, first as a continuous variable and then as categories of approximate tertiles of exposure for residence in zone A (0 years, 1–25 years, >25 years), and the same cut off points for years of residence in zones A or D. To allow for the effect of latency, we refitted models after omitting residence close to industry in the previous 20 years. We tested for interaction between residence close to industry and smoking by adding interaction terms for the categorical variables for years lived close to heavy industry by ever smokers (Y/N). The analysis was repeated using the same models after excluding 21 cases without a confirmed histological or cytological diagnosis.

The prevalence of smoking in our control group was relatively low. We therefore conducted a sensitivity analysis to investigate the effect of varying the prevalence of smoking in the control group on the adjusted OR for living near industry. We increased the prevalence of smoking in the control group by deleting some non-smoking and ex-smoker controls. In order to preserve matched sets this was done in a controlled manner (for example, not deleting controls if there was only one per case) but with a random element. We then refitted the fully adjusted model and calculated the OR for an additional 10 years of residence in zone A. This was repeated many times with varying numbers of deletions to allow the prevalence of smoking among controls to vary between 15% and 25%.

Table 1 Response among cases and controls

	Cases	Controls
Potentially eligible subjects identified*	461	817
Too ill/died (%)	213 (46.2)	108 (13.2)
Refused/non-responders (%)	44 (9.5)	370 (45.3)
Included	204	339
Response (%)		
Response among potentially eligible subjects	44.3	41.5
Response among eligible subjects†	82.3	47.8

*This excludes three cases whose diagnosis was uncertain, and five cases and 24 controls who lived outside the study area.
†Too ill/died (cases and controls not included in denominator).

Table 2 Characteristics and representativeness of included cases

Characteristic	Included cases (n = 204)	Excluded (too ill/died) (n = 213)	Refusers (n = 44)
Mean (SD) age	64.8 (9.6)	67.8 (9.5)	67.2 (11.1)
Median (IQR) Townsend score of	1.20	1.66	3.84
enumeration district of current residence	(−2.17 to 4.12)	(2.25 to 4.63)	(−1.23 to 5.62)
Basis of diagnosis (%)			
Clinical diagnosis only	10.3	28.2	22.7
Histology/cytology	89.7	72.8	77.3
Histological type where known (%)			
Small cell carcinoma	28.4	28.8	14.7
Non-small cell carcinoma	64.5	66.0	79.4
Other	7.1	5.2	5.9
Smoking status before diagnosis (%)			
Current smoker	59.3	67.1	58.3
Ex-smoker	37.8	28.7	33.3
Never smoker	3.0	4.2	8.3
Exposure zone of current residence (%)			
Zone A	27.0	22.6	25.0
Zone B	39.2	44.3	54.6
Zone C	33.8	33.0	20.5

Ethical issues

Ethical approval for the study was granted by the North and South Tees Research ethics committees. The study protocol was presented to key local stakeholders and all Teesside GPs were informed in writing in advance about the study. Cases and controls were approached only if their GPs (and hospital clinical team for cases) did not object to their participation. All subjects gave their informed consent to taking part in the study.

RESULTS

Numbers and response

We recruited and interviewed 204 cases and 339 controls. Recruitment ended before we reached the target of two controls per case. Table 1 shows the response rates. Refusal among cases was rare (9.5%). The most common reason for not recruiting potentially eligible cases was that the clinical team judged them too weak, frail, or ill to take part. Many died soon after diagnosis.

One control subsequently developed lung cancer. She was included as a case and control in line with the concurrent sampling strategy. The median time between the first and last interview in matched sets was 50 days (maximum 317).

Characteristics and representativeness of cases

Table 2 shows the characteristics of potentially eligible cases grouped by recruitment outcome. Included cases were more likely to have histologically or cytologically confirmed lung cancer and were slightly younger than non-included cases. The distribution of current zone of residence and current

Table 3 Characteristics and representativeness of included controls

Characteristic	Included controls (n = 339)	Index controls (n = 435)
Mean (SD) age	65.5 (9.4)	65.0 (9.6)
Median (IQR) Townsend score of enumeration district of current residence	−1.86 (−3.57 to 2.39)	−1.14 (−3.25 to 3.57)
Exposure zone of current residence (%)		
Zone A	17	17
Zone B	44	43
Zone C	39	38

smoking status was similar between the groups, with no statistically significant differences between the included and non-included cases. "Refusers" tended to be from more deprived areas.

Characteristics and representativeness of controls

The representativeness of included controls was assessed by comparing them with index controls who, as a random sample, should be representative of the study population weighted to the age distribution of the cases. The included and index controls had similar age distributions (table 3). The median Townsend scores were slightly lower among the included controls. The distribution of current zone of residence was almost identical.

Smoking data were not available for non-included controls. However, a Health and Lifestyle Survey from Teesside in 2000[32] suggested that smoking prevalence was lower among the TLCCCS included controls than in the source population.

Table 4 Distribution of potential confounding variables across cases and controls

Characteristic	Cases (n = 204)	Controls (n = 339)
Mean (SD) age	64.8 (9.6)	65.7 (9.4)
Median (IQR) Townsend score (current address)	1.2 (−2.2 to 4.7)	−1.9 (−3.6 to −2.4)
Currently married or co-habiting (%)	54.4	68.7
Last main job = manual social class (%)	64.2	41.9
Current tenure = rented (%)	44.1	18.6
Current car owner (%)	42.2	66.3
Active smoking status (%)		
Current	59.3	14.7
Ex-smoker	37.8	36.5
Never smoker	3.0	48.7
Regular exposure to passive smoking in any setting for >30 years over lifetime (%)	96.1	87.9
Possible or probable exposure to asbestos ever (%)	17.2	9.1
Possible or probable exposure to other lung cancer occupational exposures ever (%)	27.0	14.9

Table 5 Distribution of residential exposure to heavy industry in cases and controls

	None n (%)	1–25 years n (%)	>25 years n (%)	Median (IQR) years
Years in zone A				
Cases (n = 204)	58 (28.4)	58 (28.4)	88 (43.1)	20 (0–41)
Controls (n = 339)	126 (37.2)	130 (38.4)	83 (24.5)	11 (0–25)
Years in zone A or D				
Cases (n = 204)	45 (22.1)	60 (29.4)	99 (48.5)	24 (4.5–42)
Controls (n = 339)	86 (25.4)	156 (46.0)	97 (28.6)	16 (0–29)

The prevalence of smoking in the 2000 Teesside survey and among TLCCS included controls was 28% v 21% in women aged 45–54 years, 22% v 18% in women aged 55–64 years, and 21% v 13% in those aged 65–74 years.

Comparison of cases and controls

Table 4 compares the distribution of potential confounding variables across cases and controls. Cases were of lower socioeconomic status, were more heavily exposed to passive smoking, asbestos, and other occupational risk factors, and were much more likely to be smokers and less likely to be never smokers.

We investigated evidence of recall bias in residential exposure by asking whether subjects believed that living close to heavy industry caused lung cancer, and whether those who believed this reported more years of residence in zone A than other subjects. While cases (62%) were more likely than controls (52%) to believe that living close to heavy industry probably or definitely caused lung cancer, the mean number of years lived in zone A was slightly lower among cases (23.8 v 24.5 years) and controls (16.2 v 18.3 years) with this belief. Recall bias of this type was therefore unlikely.

Table 5 shows the distribution of exposure to living close to heavy industry for cases and controls. Teesside has a very stable population: the mean years lived in Teesside was 56.4 and 55.3 years in cases and controls, respectively, and the mean length of time at their current address was 21.0 years for cases and 23.9 years for controls.

Table 6 shows the results of the conditional logistic regression analysis. The age adjusted ORs for residence close to heavy industry on Teesside were raised, either by comparing those living in zone A for >25 years with those

Table 6 Results of conditional logistic regression analysis for years lived in zone A (close to heavy industry on Teesside) and in zones A/D (close to heavy industry on Teesside or outside Teesside)

	Age adjusted* OR (95% CI)	Age and smoking adjusted† OR (95% CI)	Fully adjusted‡ OR (95% CI)
Zone A			
No adjustment for latency			
Categorical			
0 years (reference)	1.0	1.0	1.0
1–25 years	0.95 (0.60 to 1.51)	0.93 (0.46 to 1.91)	0.92 (0.42 to 2.02)
>25 years	2.13 (1.34 to 3.38)	1.85 (0.91 to 3.79)	1.83 (0.82 to 4.08)
Continuous variable (10 years additional exposure)	1.14 (1.05 to 1.25)	1.11 (0.99 to 1.26)	1.10 (0.96 to 1.26)
Latency adjusted			
Categorical			
0 years (reference)	1.0	1.0	1.0
1–25 years	1.50 (0.98 to 2.31)	1.19 (0.63 to 2.24)	1.06 (0.55 to 2.08)
>25 years	1.82 (1.03 to 3.19)	1.56 (0.71 to 3.44)	1.38 (0.59 to 3.26)
Continuous variable (10 years additional exposure)	1.22 (1.07 to 1.38)	1.17 (0.98 to 1.40)	1.16 (0.95 to 1.40)
Zone A/D			
No adjustment for latency			
Categorical			
0 years (reference)	1.0	1.0	1.0
1–25 years	0.70 (0.42 to 1.16)	0.53 (0.23 to 1.24)	0.42 (0.17 to 1.07)
>25 years	1.75 (1.07 to 2.86)	1.54 (0.69 to 3.45)	1.21 (0.48 to 3.02)
Continuous variable (10 years additional exposure)	1.15 (1.05 to 1.25)	1.16 (1.02 to 1.32)	1.14 (0.99 to 1.31)
Latency adjusted			
Categorical			
0 years (reference)	1.0	1.0	1.0
1–25 years	1.19 (0.78 to 1.81)	0.92 (0.50 to 1.70)	0.83 (0.43 to 1.60)
>25 years	1.85 (1.10 to 3.09)	2.01 (0.93 to 4.36)	1.85 (0.80 to 4.24)
Continuous variable (10 years additional exposure)	1.20 (1.06 to 1.35)	1.23 (1.02 to 1.47)	1.21 (0.99 to 1.47)

*Adjusted for age at interview entered as a linear term.
†Adjusted for age at interview, current smoking status (ever/never), pack years smoked, pack years smoked square root, years stopped smoking.
‡As for age and smoking adjusted, with years of possible/probable asbestos exposure, current marital status, and early adult main job manual/non-manual.

who had never lived there (OR 2.13, 95% CI 1.34 to 3.38), or by comparing women who had lived an extra 10 years in zone A (OR 1.14, 95% CI 1.05 to 1.25). These ORs were reduced when terms were added to the model to include smoking and other confounding factors: adjusted OR for living >25 years (v 0 years) in zone A 1.83 (95% CI 0.82 to 4.08) or 1.10 (95% CI 0.96 to 1.26) for an additional 10 years lived in zone A when entered as a continuous variable (linear term). Adjustment for active smoking had the greatest effect on OR estimates. The models using years lived close to heavy industry on Teesside or outside Teesside (zone A/D) and including latency gave similar results (table 6). An indication of the goodness of fit of the fully adjusted models is the pseudo r^2 value which was 0.61–0.63. The interaction terms for smoking status and residential exposure categories were not retained in the models.

The ORs for lung cancer among people living >25 years in zone A in the analysis restricted to the 183 cases with a histologically or cytologically confirmed diagnosis were slightly lower: 1.88 (95% CI 1.17 to 3.01) for the age adjusted model, 1.64 (95% CI 0.83 to 3.22) for the age and smoking adjusted model, and 1.47 (95% CI 0.69 to 3.10) for the fully adjusted model. There were similar reductions in ORs in the models with years lived in zone A entered as a continuous variable, and in the latency adjusted models and models using years lived in zone A or D as the exposure measure.

Varying the smoking prevalence in the control group for the sensitivity analysis had little effect on the OR. The mean OR across runs was close to the value obtained in our analysis of the full dataset, although the width of the confidence interval increased slightly when the prevalence of smoking was reduced by control deletions.

DISCUSSION
An earlier study had shown an increased risk of lung cancer in those who lived near to heavy industry in Teesside.[12][13] In the current study the age adjusted ORs also suggested a positive association but, after adjusting for life long smoking history, occupational and socioeconomic factors, the strength of the association was reduced and the confidence intervals overlapped the null. The association with living >25 years close to heavy industry was similar if exposure to heavy industry was expanded to that outside Teesside (zones A and D) or if latency was addressed by excluding residential exposure in the last 20 years.

The findings add to existing data and are broadly consistent with most studies summarised in two reviews of the role of industrial air pollution in causing lung cancer. Pershagen[14] in 1990 reported that most ecological studies showed an increased risk of lung cancer among populations living close to non-ferrous smelters and a variety of other heavy industry types. However, he noted that few studies controlled for potential confounders like smoking or employment within heavy industry. Bendetti et al[16] reviewed 10 case-control studies in 2001, seven of which found that lung cancer was associated with residential proximity to smelters, complex industrial areas, or localised sources of industrial emissions, and three found little evidence for such an association.

A detailed appraisal of 19 case-control studies in 2004 found nine which broadly supported the hypothesis that there is a weak association (OR generally <2.0) between lung cancer and living close to heavy industry.[33] There were great variations in study quality. Common deficiencies were: inadequate processes for the identification and validation of cases; inappropriate controls; reliance on exposure data from routine data sources or from proxies among a significant proportion of cases; crude assessment of the exposure of interest (for example, characterising exposure based on place

of residence at time of death or diagnosis only); and failure to control for one or more of the main potential confounding factors.

The TLCCS had important strengths and innovations and addressed many of the best practice recommendations proposed in the 2002 WHO/HEI report for studies investigating the long term effects of air pollution.[34] The findings and novel methodological aspects are an important addition to the literature in this area of research.

Firstly, our study design is a population based study using incident cases and community controls drawn from a population based sampling frame, representative of the source population for cases. The setting was highly suitable to investigate the research question owing to the long history of heavy industry and the variation in exposure to heavy industry within the study area. We collected data directly from subjects, without the use of proxy informants, using life grid interviews—an innovative and rigorous methodology. Participants were blind to the study hypothesis and equivalent methods of data collection were used for cases and controls. We sought evidence for recall bias on residential exposure and found none.

Our exposure assessment included collection of a life long residential history. We defined exposure zones with reference to a historical review of exposure by place over time from a variety of data sources.[6] Detailed life course data were collected on all main potential confounding factors including smoking. Adjustment for potential confounding factors should therefore be less susceptible to measurement imprecision causing non-differential misclassification and hence the possibility of observed associations being attributable to residual confounding. We investigated for dose response and the effect of allowing for latency.

A potential limitation was the large number of ineligible cases and the low response rate among controls. This can introduce selection bias. The former is common in lung cancer case-control studies not using proxy respondents, and the latter in case-control studies recruiting population controls. For example, a recent report from a case-control study of leukaemia in the UK using the same sampling frame and an almost identical method of control recruitment to the TLCCS also reported a response rate of 47% among controls.[35] The risk of selection bias affecting the estimate of an association through differential participation by place of residence among controls was minimised by using index and reserve controls drawn from the same electoral ward. Furthermore, tables 2 and 3 suggest important selection bias was unlikely as cases and controls were representative of potentially eligible subjects for exposure zone of current residence. It remains possible that participating cases or controls had longer or shorter mean duration of residence in their current zone than non-participating subjects, although there is no reason to believe this should be the case.

Participating cases and controls had slightly lower mean deprivation scores than non-participating cases and all index controls, respectively, and participating controls had a lower smoking prevalence than equivalent age groups in a recent Teesside community survey. Given the association between smoking and both lung cancer and residential history, the latter could introduce bias. However, in the sensitivity analysis, increasing the prevalence of smoking among the controls had little effect on the OR, which suggests that the estimate from the full dataset was not sensitive to plausible variations in the underlying prevalence of current smokers among the controls.

We measured and adjusted for a range of potential confounding factors, although not for co-morbidities such as pulmonary tuberculosis and chronic obstructive airways disease which have been associated with an increased risk of

lung cancer in some previous studies. We did not do so, firstly, because of the inconsistency of the results and also because of doubts over the validity of associations found between chronic lung conditions and lung cancer in some studies. Secondly, in the previous Teesside Environmental Epidemiology Study there was no evidence of increased occurrence of respiratory symptoms or death from chronic bronchitis, asthma or tuberculosis in people living in communities closest to heavy industry.[12] Another confounder not assessed was domestic radon gas. This is because Teesside is not a high radon exposure area so it was unlikely to be an important cause of lung cancer among this population.

Finally, some misclassification for the exposure of interest as a result of measurement imprecision is inevitable. Given that participants were blind to the study hypothesis and the lack of evidence for recall bias, this was likely to be non-differential exposure misclassification and would have resulted in an underestimate of the association between living close to heavy industry and lung cancer.

In conclusion, this population based study using life grid interviews for life course exposure assessment addressed many deficiencies in the design of previous studies. The exposure OR for living close to heavy industry for 25 years or more on Teesside was modestly raised (1.83), although the confidence intervals were wide. These findings support those in much of the international literature of an increased risk of lung cancer with prolonged residence close to heavy industry. The effect of air pollution on the incidence of lung cancer merits continued study.

Exposure measures used in the analysis are shown in the online appendix available at http://www.thoraxjnl.com/supplemental.

....

Authors' affiliations

R Edwards, Department of Public Health, Wellington School of Medicine and Health Sciences, University of Otago, Wellington, New Zealand

T Pless-Mulloli, Department of Environmental Epidemiology, Institute of Health and Society, University of Newcastle upon Tyne, Newcastle, UK

D Howel, T Chadwick, School of Population and Health Sciences, University of Newcastle upon Tyne, Newcastle, UK

R Bhopal, Public Health Sciences, University of Edinburgh, Edinburgh, UK

R Harrison, Department of Respiratory Medicine, University Hospital of North Tees, Stockton, UK

H Gribbin, Department of Respiratory Medicine, James Cook University Hospital, Middlesbrough, UK

REFERENCES

1 **Beynon H**, Hudson R, Sadlere D. *A place called Teesside: a locality in global economy.* Edinburgh: Edinburgh University Press, 1994.
2 **Gladstone F**. *The politics of planning.* London: Maurice Temple Smith Ltd, 1976.
3 **Medical Officer of Health**. *Annual report.* Middlesbrough: Eston Urban District Council, 1925.
4 **Medical Officer of Health**. *Annual report.* Middlesbrough: Eston Urban District Council, 1956.
5 **Medical Officer of Health**. *Annual report.* Stockton: Borough of Stockton, 1964.
6 **Pless-Mulloli T**, Dunn CE, Bhopal R, *et al.* Is it feasible to construct a community profile of exposure to industrial air pollution? *Occup Environ Med* 2000;**57**:542–9.
7 **Dean G**, Lee PN, Todd GF, *et al. Report on a second retrospective mortality study in North-East England.* London: Tobacco Research Council, 1977.
8 **Wicken AJ**, Buck SF. *Report on a study of environmental factors associated with lung cancer and bronchitis mortality in areas of North East England.* London: Tobacco Research Council, 1964.
9 **Phillimore P**, Beattie A. *Health and inequality: the Northern region 1981–1991.* Newcastle upon Tyne: University of Newcastle upon Tyne, 1994.
10 **Phillimore PR**, Morris D. Discrepant legacies: premature mortality in two industrial towns. *Soc Sci Med* 1991;**33**:139–52.
11 **Townsend P**, Phillimore P, Beattie A. *Health and deprivation: inequality and the north.* London: Croom Helm, 1988.
12 **Bhopal RS**, Moffatt S, Pless-Mulloli T, *et al.* Does living near a constellation of petrochemical, steel, and other industries impair health? *Occup Environ Health* 1998;**55**:812–22.
13 **Pless-Mulloli T**, Phillimore P, Moffatt S, *et al.* Lung cancer, proximity to industry and poverty in north east England. *Environ Health Perspect* 1998;**106**:189–96.
14 **Pershagen G**. Air pollution and cancer. In: Vaino H, Sorsa M, Michael AJ, eds. *Complex mixtures and cancer risk.* Lyon: IARC, 1990.
15 **Speizer FE**, Samet JM. Air pollution and lung cancer. In: Samet JM, eds. *Epidemiology of lung cancer.* New York: Marcel Dekker, 1994.
16 **Benedetti M**, Lavarone I, Comba P. Cancer risk associated with residential proximity to industrial sites: a review. *Arch Environ Health* 2001;**56**:342–9.
17 **Blane DB**. Collecting retrospective data: development of a reliable method and a pilot study of its use. *Soc Sci Med* 1996;**42**:751–7.
18 **Holland P**, Berney L, Blane D, *et al.* The lifegrid method in health inequalities research. *Health Variations*, 2000;8–9..
19 **Berney LR**, Blane DB. Collecting retrospective data: accuracy of recall after 50 years judged against historical records. *Soc Sci Med* 1997;**45**:1519–25.
20 **Elias P**. Methodological, statistical and practical issues arising from the collection and analysis of work history information by survey techniques. *Bull Methodol Sociol* 1991;**31**:3–31.
21 **Engel LS**, Keifer MC, Zahm SH. Comparison of a traditional questionnaire with an icon/calendar-based questionnaire to assess occupational history. *Am J Ind Med* 2001;**40**:502–11.
22 **Engel LS**, Keifer MC, Thompson ML, *et al.* Test-retest reliability of an icon/calendar-based questionnaire used to assess occupational history. *Am J Ind Med* 2001;**40**:512–22.
23 **Blane D**, Berney L, Davey Smith G, *et al.* Reconstructing the life course: health during early old age in a follow-up study based on the Boyd Orr cohort. *Public Health* 1999;**113**:117–24.
24 **Dawson J**, Thorogood M, Marks SA, *et al.* The prevalence of foot problems in older women: a cause for concern. *J Public Health Med* 2002;**24**:77–84.
25 **Parry O**, Thompson C, Fowkes G. Life course data collection: qualitative interviewing using the life grid. *Sociological research online.* 1999;4: Issue 2, Available at http://www.socresonline.org.uk/4/2/parry.html (accessed 23 July 2005).
26 **Zahm SH**, Colt JS, Engel LS, *et al.* Development of a life events/icon calendar questionnaire to ascertain occupational histories and other characteristics of migrant farm workers. *Am J Ind Med* 2001;**40**:490–501.
27 **Cummings K**, Markello S, Mahoney M, *et al.* Measurement of lifetime exposure to passive smoking. *Am J Epidemiol* 1989;**130**:122–31.
28 **Coghlin J**, Hammond SK, Gann PH. Development of epidemiologic tools for measuring environmental tobacco smoke exposure. *Am J Epidemiol* 1989;**130**:696–704.
29 **Office of Population Censuses and Surveys**. *Standard occupational classification. Volume 3: Social classifications and coding methodology.* London: HMSO, 1991.
30 **Office of Population Censuses and Surveys**. *Standard occupational classification. Volume 2,* 2nd ed. London: HMSO, 1995.
31 **Royston P**, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 1999;**28**:964–74.
32 **Reilly M**, Ashton V, Eynon C, *et al. Health and lifestyle, Tees area 2000.* Middlesbrough: Tees Health Authority, 2001.
33 **Edwards R**. *Does living close to heavy industry cause lung cancer? Investigation through a case control study employing the lifegrid approach to collect life course exposure data,* MD Thesis, Cambridge University, 2005.
34 **World Health Organisation**. *Exposure assessment in studies on the chronic effects of long-term exposure to air pollution,* Report of a WHO/HEI Workshop held in Bonn, Germany. Geneva: WHO, 2002.
35 **Smith AG**, Fear NT, Law GR, *et al.* Representativeness of samples from general practice lists in epidemiological studies: case control studies. *BMJ* 2004;**328**:932.