Using Geographic Information Systems To Assess Individual Historical Exposure to Air Pollution from Traffic and House Heating in Stockholm

Tom Bellander,1,2 Niklas Berglind,1 Per Gustavsson,3 Tage Jonson,4 Fredrik Nyberg,2,5 Göran Pershagen,1,2 and Lars Järup1,2,6

1Department of Environmental Health, Stockholm County Council, Stockholm, Sweden; 2Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 3Department of Occupational Health, Stockholm County Council, Stockholm, Sweden; 4Environment Health Protection Administration, Slö-analys, Stockholm, Sweden; 5International Agency for Research on Cancer, Lyon, France; 6Department of Epidemiology and Public Health, Imperial College, London, UK

Recent developments of geographic information systems (GIS) have provided environmental epidemiologists with new tools to study associations between environmental exposures and disease. A GIS is a powerful computer mapping and analysis tool, which permits spatial linking of different types of data (e.g., residential addresses, environmental exposure levels, demographic information) (1,2). GIS can also be used for data management, such as automated address matching to geographic coordinates (geocoding).

A specific aim of a population-based case-control study of lung cancer in Stockholm, Sweden, was to use emission data, dispersion models, and geographic information systems (GIS) to assess historical exposure to several components of ambient air pollution. Data collected for 1,042 lung cancer cases and 2,364 population controls included information on residence from 1955 to the end of follow-up for each individual, 1990-1995. We assessed ambient air concentrations of pollutants from road traffic and heating throughout the study area for three points in time (1960, 1970, and 1980) using reconstructed emission data for the index pollutants nitrogen oxides (NOx, NO2) and sulfur dioxide with dispersion modeling. NOx estimates for 1980 compared well with actual measurements, but no independently measured (study-external) data were available for SO2, precluding similar validation. Subsequently, we used linear intra- and extrapolation to obtain estimates for all other years 1955-1990. Eleven thousand individual addresses were transformed into geographic coordinates through automatic and manual procedures, with an estimated error of <0.100 m for 90% of the addresses. Finally, we linked annual air pollution estimates to annual residence coordinates, yielding long-term residential exposure indices for each individual. There was a wide range of individual long-term average exposure, with an 11-fold interindividual difference in NO2 and an 18-fold difference in SO2. The 30-year average for all study subjects was 20 µg/m³ NO2 from traffic and 53 µg/m³ SO2 from heating. The results indicate that GIS can be useful for exposure assessment in environmental epidemiology studies, provided that detailed geographically related exposure data are available for relevant time periods.

Key words: air pollution, combustion, epidemiology, exposure assessment, GIS, lung cancer, traffic.

Environ Health Perspect 109:633-639 (2001). [Online 15 June 2001]
http://ehpnet1.niehs.nih.gov/docs/2001/109p633-639bellander/abstract.html

Materials and Methods

The source population consists of all men 40 to 75 years of age who lived in Stockholm County at any time between 1985 and 1990, and who had lived in the County since 1950, with a maximum of five years of residence outside the County. From this population we recruited three groups, which form the study population for this study: 1,042 cancer cases reported to the Stockholm County cancer registry 1985-1990, and two population-based control groups comprising 2,364 individuals, age-stratified to the case group (18,19).

We collected information on individual exposures using a postal questionnaire sent to living respondents or to next of kin when the study subject had died. The questionnaire included questions on occupations, smoking habits, and food consumption. To assess exposure to air pollution, we also collected information on all dwellings where the case or control had lived for more than one year from 1955 onward. In case of incomplete data in form of gaps, overlaps, or other

Address correspondence to T. Bellander, Department of Environmental Health, Norrbacka 3rd floor, Karolinska Hospital, S-176 76 Stockholm, Sweden. Telephone: +46 8 5177 7906. Fax: +46 8 5177 7900. E-mail: tom.bellander@smd.sll.se

The work was performed at Department of Environmental Health, Karolinska Hospital, Norrbacka, S-171 76 Stockholm, Sweden.

The study was supported by the Swedish Environmental Protection Agency, the Swedish Council for Work Life Research, and the Stockholm County Council.

We are indebted to C. Bengtsson, A. Boberg, E-B. Gustavsson, M. Hugosson, L-G. Hörte, R. Jakobsson, E. Junghahn, U. Klinga, A. Lauber, S. Norberg, M. Pettersson, C. Rudengren, P. Schéele, E. Skärke, A. Wibom, and K. Aström for valuable contributions to various aspects of the study.

Received 8 February 2000; accepted 11 December 2000.
inconsistencies in the residential histories, we sought supplemental information from the local offices for demography (parishes) and/or the tax authority. Respondents who had lived in apartment houses were also asked whether they had lived on the ground floor.

We transformed the addresses into geographic coordinates using the MAPINFO computer software (20) in conjunction with a regional geographic address database (21). The address database covers all urban areas in Stockholm County, defined as areas with more than 500 inhabitants and no more than 200 meters between houses (22). The database contains the names of all streets in the urban area. Each street is divided into segments (generally the length of a block), with geographic coordinates on both ends of the street segment and with street numbers on each side of the street at both endpoints. When an address with street and street number is matched to the address database, coordinates for street numbers between endpoints of the street segment are calculated through interpolation.

We geocoded the 10,800 addresses in the study in a three-step procedure. The first step used the automatic geocoding function in MAPINFO. Addresses that could not be coded directly by the program were coded interactively in a second step, with an operator adjusting spelling or street number. In the third step, addresses not matched interactively were coded manually.

We tested the reliability of the automatic coding function using two external suppliers of geocoding services (Telemedia (Stockholm, Sweden), a private company making the maps for the Swedish telephone directory, and CFD (Centrala fastighetsdata; Gavle, Sweden), central authority for real estate data). Telemedia uses a technique similar to that of MAPINFO, whereas CFD sets the coordinates for an address to the center of the property.

We assessed exposure to traffic-related air pollution (NO\textsubscript{x} and NO\textsubscript{y}) for 1955–1990 for the whole territory of Stockholm County, approximately 6,500 km\textsuperscript{2}. The exposure assessment was based on geographic modeling of three points in time: 1960, 1970, and 1980. We used the AIRVIRO dispersion modeling system in conjunction with retrospectively constructed emission databases for these years. We used a detailed regional emission database for 1993, collected by the Stockholm-Uppsala Air Quality Management Association (Stockholm), as a basis (23). It contains approximately 4,300 line sources related to road traffic, covering all roads with more than 1,000 vehicles/24 hr, which corresponds to 90% of the estimated total emission from road traffic. It also contains information on over 500 point sources, including major industries and energy plants as well as small industry and ferries in ports. Limited diffuse emission sources, e.g., air traffic and merchant vessels in commercial routes, are treated as area sources, and several population-density related sources, such as local heating, are mapped as grid-sources (250 m or 1,000 m grids). Work machine emissions are treated as grid sources, distributed with one-quarter over industrial areas, and the remainder according to population density. The total emission of NO\textsubscript{x}, in the region in 1993 is estimated at 18,900 tons, of which 79% originated from road traffic.

To reconstruct historical emission databases, we collected retrospectively data from 1955 to 1993 on the growth of the urban areas in the Stockholm county region, the development of the district heating (1970–1980), and the growth and redistribution of the road traffic (Table 1). We constructed separate emission databases for regional emissions of NO\textsubscript{x} for 1960, 1970, and 1980. The two former databases contain only contributions from road traffic, whereas the 1980 database was constructed in two versions: with road traffic only and with all sources. The total emissions of NO\textsubscript{x} from road traffic in these databases are 13,400, 18,100, and 20,800 tons/year.

Emissions from heating were assessed as SO\textsubscript{2} in a similar way as those from NO\textsubscript{x}, from road traffic. We reconstructed three historical regional emission databases for the same years as for the NO\textsubscript{x} emissions, with information again based on the 1993 database. Detailed data on the development of district heating (point sources) and other energy plants, and local restrictions of sulfur content in oil were used for the reconstruction (Table 2), as well as SO\textsubscript{2} measurement data. We used data from five monitoring stations (annual and decade mean levels) to calibrate the model for average emission levels from grid-type sources (mainly local oil-fueled house heating) for 1970 and 1980, through iterative model runs. Measured levels from the two central monitoring stations were allowed to influence the relative level of emissions from grid sources in the 1960 and 1970 databases, respectively.

We performed dispersion calculations from the retroactively constructed emission databases using the Gaussian model in the AIRVIRO system (8). We used the 1990s average distribution of climate in 180 different conditions over a year for all periods, and we used no correction factors. We calculated SO\textsubscript{2} and NO\textsubscript{x} annual mean concentrations for the relevant period, using only the contributions from road traffic and heating sources, respectively. The NO\textsubscript{x} concentrations were transformed to NO\textsubscript{2} data using the equation

\[ [\text{NO}_2] = [\text{NO}_x]^{A+B[C]} \]

in which the parameters A, B, and C were estimated from measurements in Stockholm County in the early 1980s (hourly means from 10 fixed and five mobile stations). The parameters A and B were allowed to vary according to location, and the estimates were in the ranges 0.60–0.70 and 30–40, respectively. The parameter C was estimated at 100 throughout the county. The performance of these conversions was tested in 220,418 paired 1-hr measurements of NO\textsubscript{x} and NO\textsubscript{2} from the 1990s. The correlation between

---

**Table 1. Examples of input data for reconstruction of NO\textsubscript{x} emission data in Stockholm County for 1960, 1970, and 1980, and local interpolation of NO\textsubscript{x} levels 1955–1990.**

| Item/Year | Data/Event |
|-----------|------------|
| Traffic counts, Stockholm city and suburbs | Linear increase from 210,000 to 490,000 vehicles/day passing city limits (+19,000/year) |
| 1955–1970 | Small changes |
| 1970–1980 | Linear increase from 470,000 to 620,000 vehicles/day passing inner-city limits (+19,000/year) |
| 1981–1989 | Estimated at a linear increase |
| Traffic counts, rest of county | Small changes |
| 1955–1970 | Linear increase from 530,000 to 720,000 vehicles/day passing outer city limits (+27,000/year) |
| 1971–1980 | 608, 901, 882, and 1,008 vehicle Gm/year* |
| 1981–1989 | 2018, 3,282, 2,876, and 3,370 vehicle Gm/year |
| Total traffic work, Stockholm city | 1,868, 3,305, 4,016, and 5,103 vehicle Gm/year |
| 1960, 1970, 1980, and 1993 | 2,678, 3,305, 4,016, and 5,103 vehicle Gm/year |
| Total traffic work, city and suburbs | Two parts of the western link open |
| 1960, 1970, 1980, and 1993 | Two parts of the western link open |
| Total traffic work, rest of county | Unified taxes in the whole region |
| 1960, 1970, 1980, and 1993 | 1971 |

*1 vehicle Gm/year = 10\textsuperscript{6} vehicle km/year.
NO₂ levels estimated from NOₓ and measured NO₂ levels was 0.91.

We performed the dispersion calculations in four different resolutions: 2,000 m × 2,000 m, 500 m × 500 m, 200 m × 200 m, and 100 m × 100 m grids. We used the highest resolution in central city areas and the lowest resolutions in the countryside. For each grid we estimated the average levels of the index pollutants. All illustrations are based on 500 m × 500 m grids. The isolines in the illustrations were not used for the calculations.

For main streets in the city center, we added a street canyon contribution to the assessed roof concentrations of NOₓ. We assessed this contribution by dispersion calculations with the AIRVIRO street canyon model, and summarized it in an 80% addition at the ground floor and a 40% addition for all other dwellings in the streets concerned. We gave 68 addresses (2% of 3,406) an addition to the 1960 dispersion estimate; 58 addresses an addition to the 1970 dispersion estimate, and 44 to the 1980 dispersion estimate. Additions to the assessed levels of NO₂ were made in the same way, but with a 50% addition at the ground floor and a 20% addition for other dwellings in these streets.

For 1980, we compared the dispersion model calculations including all sources of NOₓ—after transformation to NO₂ as indicated above—with measurements of NO₂ in five rooftop or background sites (one stationary and four mobile stations) taken in the early 1980s (Figure 1). In the first model run, all five annual values predicted by the model at the highest resolution (not shown in Figure 1) were within ±20% of the value measured at their respective locations. The model was successively calibrated to minimize this deviation. All available measurements for SO₂ had been used to calibrate the model, so no formal validation of the SO₂ modeling was possible.

To create annual levels of exposure to SO₂ and NO₂ from the modeled concentration fields, we computed extra- and intrapolated values for each address for each year between 1955 and 1990. We modeled the NO₂ concentration in a geographic point to increase linearly from 1955 to 1970, using the estimated values for 1960 and 1970 to obtain the slope. The rationale for linearity is that the traffic counts increased approximately linearly in this period. Between 1970 and 1980 the changes were small and were also modeled as linear. We modeled the increases for after 1980 as linear with the same slope as 1955–1970, motivated by a similar linear increase in traffic counts. In the vicinity of new major roads and the corresponding relieved roads, the slopes before and after the change were extracted from surrounding areas.

We modeled SO₂ concentrations as constant during the period 1955–1967 (1960

Table 2. Examples of input data for reconstruction of SO₂ emission data for 1960, 1970, and 1980, and local interpolation of SO₂ levels 1955–1959, 1961–1969, 1971–1979, and 1981–1990.

| Item, year | Data/event |
|------------|------------|
| House heating fuel | Gradual shift from carbon to oil |
| House heating methods | Mainly independent for each building or property |
| October 1969 | Gradual shift to district heating |
| October 1969 | 50% of heating energy in Stockholm city from district heating |
| October 1969 | 70% of heating energy in Stockholm city from district heating, remaining areas converted to electric heating |
| Energy conservation | Oil crisis started a series of actions on energy conservation in house heating |
| Sulfur limit in oil, Stockholm city | Not regulated |
| Until 1968 | 1% in the city center |
| 1969–1990 | 1% in the rest of the city |
| Sulfur limit, Stockholm county except the city | Not regulated |
| Until 1968 | 3.5% |
| 1969–1970 | 1% gradually implemented |
| 1984–1990 | 1% |
| Estimated average sulfur content in fuel | 3% in all the county |
| 1955–1968 | 1% in the city; 3% in rest of the county |
| 1969–1970 | 1% in all the county |
| 1971–1990 | 1% in all the county |

Figure 1. Modeled average concentrations of NO₂ (µg/m³) from all sources, Greater Stockholm 1980. Grid resolution is 500 m. Higher-resolution model runs (for city center, not shown here) show better prediction of measured values.
Based on these exposure estimates are pre-
multiply weighted for average population.
interpreted as an annual area average, approxi-
mate to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
exposure indices for each of the three air pol-
lutations for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yielding
data to 10,800 individual address coordi-
ates for the relevant time intervals, yieldin

Results

The dispersion calculations for NO2 from road traffic show a substantial increase of the average level with time (Figures 2 and 3), mainly from 1960 to 1970. The increase affects most parts of greater Stockholm and is caused mainly by the increase in traffic volume (Table 1). The dispersion calculations for SO2 from heating show a substantial decrease of the levels with time, from above 130 µg/m3 in central Stockholm in the end of the 1960s to less than 50 µg/m3 in the whole area at the beginning of the 1980s (Figures 4 and 5). The main reasons for this decrease are restrictions in the sulfur contents of heating oil, massive investment in district heating, and investment in emission control at the energy plants (Table 2). The decrease is more pronounced in densely populated areas because of earlier implementation of more strict sulfur standards and a more rapid development of district heating.

Of the 10,800 addresses in the study, 56% were geocoded by the automatic geocoding function in MAPINFO. A further 27% were geocoded by the computer program with minor operator assistance. The remaining 17% of the addresses had to be geocoded manually; these addresses were either outside urban areas or too inexact to be geocoded interactively. Three percent of all the addresses could not be found and were given approximate coordinates, e.g., of a village center. We checked the result of the geocoding visually by examining maps where the geocoded address were plotted.

For a subset (n = 100) of the addresses, we performed a reliability test of the automatic coding of geographic coordinates using two external suppliers of geocoding services. In this subset, 68 addresses could be geocoded automatically with all three procedures. Both external codings agreed reasonably well with our automatic coding. The average distance from our coding was about 50 m in both; 80% of the addresses agreed within 100 m for all three codings, and our coding agreed within 100 m for almost 90% in comparison with any one of the other two codings. The maximal deviation was 168 m and 237 m, respectively, when our coordinates were compared with those from the external suppliers.

The estimated annual mean concentration levels for NO2 from road traffic and SO2 from heating, at all of the 10,800 addresses in the study (i.e., considering the whole period for every address), shows clearly the increase of NO2 and decline of SO2 in Stockholm County (Figures 6 and 7). The corresponding annual means for the study population individuals are similar to those of all the addresses. There is a slight tendency of temporal decrease in the individual means compared with the average address means, caused by a net migration from more polluted to less polluted areas (Figures 6 and 7).

All individual exposure indicators show approximate normal distribution (Table 3). Despite the increase in NO2 levels during the period, the average estimated individual traffic-related NO2 levels during the three decades before the end of follow-up are well below the current World Health Organization (WHO)
ambient concentration guideline of 40 µg/m³ for NO₂, which refers to contributions from all sources. This guideline defines an annual limit value for the protection of human health, and is numerically identical to that of the European Union recommendation proposed for 2010 (24). The maximum of the individual period average is about 11 times higher than the minimum. The total range extends just over the WHO limit.

Despite the decline of emissions since 1960, the average estimated individual heating-related SO₂ level during the 30-year study period exceeds the current WHO recommended guideline of 50 µg/m³ for the annual mean ambient SO₂ concentration from all sources (25). About 58% of the individuals have study-period mean exposures from heating-related sources that exceed this level. The total range of individual average levels is quite large, with a maximum that is about 20 times higher than the minimum. That the levels decrease with time is reflected in differences in individual average levels in the three decades preceding individual end of follow-up. The mean and most other parameters decrease by about two-thirds from the first to the last decade (Table 3). The interquartile ratio is about 1.7 for all exposure metrics, and the ratio of the means of the upper and lower quartiles is about 3 (Table 3).

The individual 30-year average exposures to NO₂ and SO₂ are correlated (r = 0.65). Despite this correlation, there are obvious differences in individual exposure to the two agents. For example, at an individual average SO₂ exposure of 80 µg/m³, there is a 3-fold total variation (15–40 µg/m³) of the individual average exposure to NO₂.

**Figure 4.** Modeled average concentrations of SO₂ (µg/m³) from heating, Greater Stockholm 1960. Grid resolution in map is 500 m. Most central parts were modeled in a 100 m gridnet (not shown).

**Figure 5.** Modeled average concentrations of SO₂ (µg/m³) from heating, Greater Stockholm 1980. Grid resolution in map is 500 m. Most central parts were modeled in a 100 m gridnet (not shown).

**Figure 6.** Estimated annual mean ambient concentrations of NO₂ from road traffic in Stockholm County 1955–1990.

**Figure 7.** Estimated annual mean ambient concentrations of SO₂ from heating in Stockholm County 1955–1990.
We assessed both NO\textsubscript{X} and NO\textsubscript{2} as indices of pollution from road traffic. NO\textsubscript{X} represents the direct emissions and is, thus, in principle, a better index of traffic exhaust exposure. NO\textsubscript{2} is also important because of its toxicity and widespread use in exposure assessment. The relation between the two indices was derived from observations in Stockholm in the 1980s. The assessed relation is relevant for the period before catalytic converters were implemented, and showed good agreement with the measured relation in the beginning of the 1990s. However, it may not be used outside our study period and location, because it is dependent on composition of automobile fleet and climatologic conditions.

Geocoding addresses is often central to the use of a GIS in environmental epidemiology research. The success of the geocoding depends on the quality of the databases available. It became clear in the present study that the regional address database contained errors such as spelling mistakes and local differences in the completeness of data. This is likely to occur in most detailed databases of this kind. Also, the interpolation technique, used in the present study, yields a set of address coordinates close to the road, which should describe road traffic exposure more accurately than the method using the property centroid. A more formal validation of the address coordinates (e.g., by using the satellite Global Positioning System technique) was not performed, but we compared our method to geocoding using two other address databases, for a subset of the addresses. There was a difference of less than 100 m for 90% of the addresses. Since the highest resolution of the air dispersion model was 100 m × 100 m, the geocoding accuracy seems satisfactory.

### Table 3

Descriptive statistics for average estimated individual exposure to NO\textsubscript{X} from road traffic and SO\textsubscript{2} from house heating, used in a case-control study of lung cancer in Stockholm. Individual exposure is reported for three decades before the year in which follow-up ended. Lung cancer incidence was followed up in 1985–1990 and the individual decades fall within the calendar years 1955–1969, 1965–1979, and 1975–1989.

|                           | Decade 1 | Decade 2 | Decade 3 |
|---------------------------|----------|----------|----------|
| Average exposure to road traffic fumes (µg/m\textsuperscript{3} NO\textsubscript{X}) |          |          |          |
| Valid n                   | 3,406    | 3,406    | 3,406    |
| Mean                      | 20       | 18       | 21       |
| Median                    | 20       | 17       | 20       |
| Minimum                   | 4        | 4        | 4        |
| Maximum                   | 44       | 51       | 50       |
| Lower quartile cutpoint   | 15       | 13       | 16       |
| Upper quartile cutpoint   | 25       | 26       | 28       |
| Ratio of quartile cutpoints | 1.6     | 1.8      | 1.8      |
| Lower quartile average     | 9        | 11       | 11       |
| Upper quartile average     | 30       | 32       | 34       |
| Ratio of quartile averages | 2.8      | 3.1      | 3.1      |

| Average exposure to house heating fumes (µg/m\textsuperscript{3} SO\textsubscript{2}) |          |          |          |
| Decade 1-2                | 3,406    | 3,406    | 3,406    |
| Mean                      | 53       | 87       | 46       |
| Median                    | 64       | 88       | 66       |
| Minimum                   | 6        | 8        | 6        |
| Maximum                   | 110      | 185      | 108      |
| Lower quartile cutpoint   | 41       | 66       | 33       |
| Upper quartile cutpoint   | 67       | 110      | 59       |
| Ratio of quartile cutpoints | 1.6     | 1.7      | 1.8      |
| Lower quartile average     | 29       | 45       | 23       |
| Upper quartile average     | 78       | 129      | 71       |
| Ratio of quartile averages | 2.7      | 2.8      | 3.1      |

*Decades before the individual year of the end of follow-up. Example: For an individual with end of follow-up in 1988, Decades 1, 2, and 3 correspond to the calendar years 1958–1967, 1968–1977, and 1978–1987.*
NO₂. When the subjects are divided in quartiles, the ratio between the average exposure in the upper group and that in the lower group is almost 3 for both indices. Within the different decades, this contrast is somewhat higher. It thus seems possible to categorize this population in different levels of exposure, which permits the detection of exposure response relations. Furthermore, despite the overall correlation between the two agents, there were substantial differences in the relation between exposures to NO₂ and SO₂ for study individuals, which should make it possible to distinguish between these two pollutants in epidemiologic analyses.

However, the outdoor levels assessed in residential areas to a varying extent reflect true individual total exposure. Exposure contributions at work and in traveling are not included. In addition, perfectly measured individual levels of these pollutants have proxy properties, because they are single components in very complex mixtures of pollutants from road traffic and heating. The relation of NO₂ and SO₂ with the other components is likely to have changed over time, and may also vary across the territory. Any positive finding of exposure-response based on this exposure information is therefore likely to be an underestimate of the total effects of air pollution from these sources.

Conclusion

To our knowledge, this is the first attempt to use reconstructed emission data together with dispersion modeling and GIS for retrospective individual exposure assessment of multi-source air pollution in an epidemiologic study. The results show that the technique may be useful for exposure assessment in environmental epidemiologic studies, given that detailed emission data or environmental exposure data are available for study-relevant time periods and with good spatial resolution. Another prerequisite is that the biologically relevant true individual exposure contrast must be high enough for individual contrast to remain in a geographically based assessment of proxy exposures.

References and Notes

1. Croner CM, Sperling J, Broome FR. Geographic information systems (GIS): new perspectives in understanding human health and environmental relationships. Stat Med 15:1961–1977 (1996).
2. Vine MF, Degnan D, Hanchette C. Geographic information systems: their use in environmental epidemiologic research. Environ Health Perspect 105:598–605 (1997).
3. Briggs DJ. Mapping environmental exposure. In: Geographical and Environmental Epidemiology: Methods for Small-area Studies (Elliott P, Cuzick J, English D, Stern R, eds). Oxford:Oxford University Press, 1992:158–176.
4. Briggs DJ, Elliott P. The use of geographical information systems in studies on environment and health. World Health Stat Q 48:85–94 (1995).
5. Stockwell JR, Sorensen W, Eckett JW Jr, Carreras EM. The U.S. EPA Geographic Information System for mapping environmental releases of Toxic Chemical Release Inventory (TRI) chemicals. Risk Anal 13(2):155–164 (1993).
6. Vailus J, Hongisto M, Verkasalo P, Järvinen P, Heikkinä K, Koskenvuo M. Residential exposure to magnetic fields generated by 110–400 kV power lines in Finland. Bioelectromagnetics 16(6):365–376 (1995).
7. Moussisopoulos N, Berge E, Bøhler T, de Leeuw F, Grønskei K-E, Mylona S, Tombrou M. Ambient Air Quality, Pollutant Dispersion and Transport Models. Topic Report 19/96. Copenhagen, Denmark:European Environment Agency, 1996.
8. SMHI: Indic Airviro. Swedish Meteorological and Hydrological Institute. Technical Description of the Dispersion Models. Norrköping, Sweden: SMHI: Indic Swedish Meteorological and Hydrological Institute, 1993.
9. SMHD: Swedish Meteorological and Hydrological Institute. Airviro User Documentation. Norrköping, Sweden: Swedish Meteorological and Hydrological Institute, 1997.
10. SMHI: Indic Airviro. Swedish Meteorological and Hydrological Institute. Available: http://www.indic-airviro.smhi.se [cited 22 December 1999].
11. Higgins IT. Epidemiologic evidence on the carcinogenic risk of air pollution (Review). JARC Sci Publ 134(1):52 (1976).
12. Doll R. Atmospheric pollution and lung cancer. Environ Health Perspect 22:23–31 (1978).
13. Friberg L, Cederlöf R. Late effects of air pollution with special reference to lung cancer. Environ Health Perspect 22:45–66 (1978).
14. Hemminki K, Pershagen G. Cancer risk of air pollution: epidemiologic evidence. Environ Health Perspect 104(suppl 1):187–192 (1996).