Journal of Exposure Analysis and Environmental Epidemiology (2001) 11, 459–469

www.nature.com/jea

Estimation of long-term average exposure to outdoor air pollution for a cohort study on mortality

GERARD HOEK,[a] PAUL FISCHER,[b] PIET VAN DEN BRANDT,[c] SANDRA GOLDBOHM[d] AND BERT BRUNEKREEF[a]

[a]*Environmental and Occupational Health Group, Utrecht University, Wageningen, The Netherlands*
[b]*Laboratory of Exposure Assessment and Environmental Epidemiology, National Institute of Public Health and the Environment, Bilthoven, The Netherlands*
[c]*Department of Epidemiology, University of Maastricht, Maastricht, The Netherlands*
[d]*Department of Nutritional Epidemiology, TNO Nutrition and Food Research, Zeist, The Netherlands*

Recent prospective cohort studies have suggested that long-term exposure to low levels of particulate matter (PM) air pollution is associated with increased mortality due to, especially, cardio-pulmonary disease. Exposure to ambient air pollution was estimated mostly as city average concentrations, assuming homogenous exposure within the city. We used an ongoing cohort study — The Netherlands Cohort Study (NLCS) on diet and cancer — to investigate the relationship between traffic-related air pollution and mortality. The baseline data collection took place in 1986. A study was conducted to develop methods for exposure assessment and evaluate the contrast in exposure to air pollution within the cohort. Assessment of long-term exposure to two traffic-related air pollutants, Black Smoke (BS) and Nitrogen Dioxide (NO_2), consisted of separate estimation of regional background, urban background, and local traffic contributions at the home address. Interpolation of concentration data from a routine monitoring network was used to estimate the regional background concentration. A regression model relating degree of urbanization to air pollution was used to allow for differences between different towns/neighborhoods of cities. Distance to major roads was calculated to characterize local traffic contributions, using a Geographic Information System (GIS). Interpolation resulted in reasonably precise regional background estimation when distant sites were not used and distance squared was used as the weight. Cross-validation showed that prediction errors were about 15% of the range in regional background concentration. Urban and local scales contributed significantly to the contrast within the cohort. Prediction errors for estimating the urban background were about 25% of the range in background concentrations. When the developed model was applied to the study cohort, there was substantial contrast in estimated exposure to BS and NO_2. About 90% of the study population lived 10 years or more at its 1986 home address — supporting the use of the estimated concentration at the 1986 address as a relevant exposure variable. *Journal of Exposure Analysis and Environmental Epidemiology* (2001) **11,** 459–469.

Keywords: *air pollution, Geographic Information System, mortality, NO_2, traffic.*

Introduction

Two recent cohort studies have suggested that long-term exposure to low levels of particulate matter (PM) air pollution is associated with increased mortality due to, especially, respiratory and cardiovascular disease (Dockery et al., 1993; Pope et al., 1995). A more recent study has partly, but not completely, supported these findings (Abbey et al., 1999). The reduced survival in the more polluted cities implies a reduction in life expectancy of 1–2 years, a substantial public health impact (Brunekreef, 1997). The three cohort studies have been conducted in the US, and it is not clear as to what extent the findings apply in European circumstances with different air pollution mixtures. The PEACE project documented that in many European regions, including the Netherlands, PM_{10} concentrations are currently at or above the highest concentrations observed in the two US cohort studies (Hoek et al., 1997). There is, therefore, a clear need to replicate/refute the findings of the US cohort studies in Europe. New cohort studies to further investigate the relationship between air pollution and mortality are expensive, and we would have to wait for many years before a conclusion can be reached. For these reasons, it seems useful to use data from ongoing cohort studies.

Accurate estimation of exposure to ambient air pollution has been a difficult task in epidemiological studies of long-term effects of air pollution. The mortality cohort studies and most previous epidemiological studies (e.g., Dockery et al., 1996; Ackermann-Liebrich et al., 1997) have compared several large study areas — such as cities or metropolitan areas— with different ambient air pollution concentrations. Exposure has been estimated as the average

1. Address all correspondence to: Gerard Hoek, Institute for Risk Assessment Sciences, Division of Environmental and Occupational Health, Utrecht University, PO Box 238, Wageningen 6700 AE, The Netherlands. Tel.: +31-317-482-080. Fax: +31-317-485-278. E-mail: g.hoek@iras.uu.nl
Received 13 August 2001.

of stations within the city/metropolitan area, assuming homogenous exposure within the study area. This assumption may result in a significant error in exposure, which, depending on the contrast in exposure between the cities, may lead to substantial bias in estimated exposure response relationships. A recently conducted study in four European countries found important variation of the concentration of NO_2 on a small scale within cities (Lebret et al., 2000). Several other studies have documented important within-city variation of concentration, especially related to vicinity to motorized traffic and location within the city, e.g., center versus suburb (Bernard et al., 1997; Raaschou-Nielsen et al., 2000). Several more recent epidemiological studies have attempted to relate the spatial variation in air pollution concentration within cities, mostly related to traffic, to health. Exposure to air pollution has been characterized with a variety of methods in these studies. Most studies have used *exposure indicators*, such as distance to a major road, objectively determined or self-reported traffic intensity (Wjst et al., 1993; Edwards et al., 1994; Weiland et al., 1994; Brunekreef et al., 1997; van Vliet et al., 1997). In some studies, *measurements* of traffic-related air pollutants were conducted (Brunekreef et al., 1997; van Vliet et al., 1997; Schindler et al., 1998). A few studies have used *dispersion modeling* of traffic-related air pollution as a measure of exposure. A recent Danish study suggested that dispersion modeling resulted in a more accurate estimate of exposure than traffic intensity data only, especially in rural areas (Raaschou-Nielsen et al., 2000). A problem with the application of dispersion models is the availability of input data for a large number of residential addresses.

A recent development is the use of Geographic Information Systems (GIS) to estimate individual exposure to air pollution in large-scale epidemiologic studies (Vine et al., 1997). In the SAVIAH study, individual exposure estimates for study subjects were generated based on concentration measurements at a limited number of sites and prediction of these concentrations using data on explanatory variables available in a GIS (Briggs et al., 1997). Regression models were developed using factors such as traffic intensity near the sites, distance to major roads, population density, and sampling height to explain the measured concentrations. This regression model was next applied to the addresses of the study subjects (Briggs et al., 1997). A similar approach has been applied to describe the variation of the NO_2 concentration in a dense network spread over the UK (Stedman et al., 1997).

We used data from The Netherlands Cohort study on diet and cancer (NLCS) to evaluate the association between air pollution and mortality (van den Brandt et al., 1990). A study was conducted to develop methods for exposure assessment and assess the contrast in long-term exposure to traffic-related air pollutants in the cohort within a random subcohort of 5000 subjects. In this paper, we will illustrate how, using GIS methods, it is feasible to improve assessment of exposure to outdoor air pollution for cohort studies of the effect of long-term average air pollution on mortality by taking into account spatial variation of outdoor air pollution on finer spatial scale.

Methods

NLCS

The NLCS on diet and cancer (van den Brandt et al., 1990) is a cohort study that was started in 1986, with 120,852 subjects who were 55–69 at enrollment (48.2% male). The cohort is spread out over the Netherlands (Figure 1). The study sample was recruited from 204 municipalities that had computerized population registries in 1986, and were sufficiently covered by cancer registries.

In the initial questionnaire, information was collected about a large number of risk factors besides diet for the development of cancer. Many of these risk factors (smoking habits, passive smoking, nutritional factors, occupation) are also of potential importance for the development of respiratory as well as cardiovascular disease and mortality. The exact address of all study subjects in 1986 is known. In addition, a residential history, which allows identification of up to four previous cities of residence, is available. The NLCS study is being analyzed as a case–cohort study for which a random subsample (n~5000) is drawn from the total cohort. The subcohort, which is followed up at regular intervals for migration and vital status by contacting the participants and the municipalities, is used to estimate the person–time experience of the cohort. The current study used this subcohort for methods development.

Pollutants

At the onset of the study, it was unclear whether the contrast in outdoor air pollution in the Netherlands was sufficient to allow an epidemiological study. Measurements in the framework of the National Air Quality Monitoring Network (NAQMN) of PM_{10} have documented that the contrast in PM_{10} concentrations in the Netherlands is relatively small — 30–40% (van der Wal and Janssen, 2000). NAQMN measurements have also documented that there is virtually no contrast in long-term average O_3 concentrations within the Netherlands. The contrast in secondary aerosol components (sulfate and nitrate) is small as well (Hoek et al., 1996).

As indicators of traffic-related air pollution, Black Smoke (BS) and Nitrogen Dioxide (NO_2) are used. BS is measured by light absorbance of exposed filters and is mostly related to elemental carbon emissions from diesel engines, whereas NO_2 reflects emissions from all motorized vehicles. BS is not a good proxy for PM_{10} or $PM_{2.5}$, but is

Figure 1. Distribution of the municipalities (black) of the NLCS study population over the Netherlands. Area of the Netherlands is 41,526 km^2.

highly correlated with the elemental carbon content of particles (Edwards et al., 1983; Chow, 1995; Janssen et al., 2001). Therefore, it is a more direct indicator of exposure to (diesel) traffic-related particles than PM_{10} or $PM_{2.5}$. The major source contributing to BS concentrations is Dutch road traffic: 65% in urban environments (Bloemen et al., 1998). Approximately 50% of the national NO_2 emissions is due to motorized traffic. Due to the low emission height of traffic-related air pollutants and the concentration of traffic in urban environments, the contribution to ambient NO_2 concentrations to which human beings are actually exposed is higher than 50%.

Estimation Method

Exposure to air pollution has been estimated on the basis of the 1986 home address and residential history information to generate indicators, on an individual basis, of long-term exposure to traffic-related air pollutants. Arguments for concentrating on this period are in consistency with the two US cohort studies that focused on the first years of the cohort study as well and the changes in 1986 of the network configuration (methods, sites).

The long-term average exposure to outdoor air pollution is considered to be a function of the regional background, additional pollution from urban sources (resulting in an urban background), and additional pollution from local sources (nearby streets). In formula:

$$\text{Exposure} = \text{C(regional)} + \text{C(urban)} + \text{C(local)}$$

The methodology is similar to the methods used to estimate nationwide NO_2 concentrations (Stedman et al., 1997). The basic idea is that because NO_2 concentrations vary on a fine spatial scale, direct interpolation will result in systematic prediction errors related to emission sources in the area

(Stedman et al., 1997). Therefore, regional background, urban background, and local source contributions are estimated separately.

The six-position postal code of the home address has been transformed into standard Dutch geographical coordinates (Rijksdriehoeksmeting, RDM) using a GIS system available at the National Institute of Public Health and the Environment (RIVM). The earliest available database was used (for 1995). Due to privacy regulations, the exact address could not be used. The six-position postal code comprises approximately 25 addresses only, so it is a geographically close approximation of the actual home addresses. Especially within the (densely built) Dutch cities, the accuracy is high. In rural areas, the accuracy is lower because addresses are located further apart, but the spatial variability of air pollution concentrations is also lower in rural areas. RDM coordinates with a precision of 100×100 m of the monitoring stations of the NAQMN were taken from an RIVM report (Elskamp, 1989).

The *regional* component at the home address has been estimated using interpolation of BS and NO_2 measurements at regional background stations of the NAQMN. There are 13 regional stations for BS and 24 regional stations for NO_2. This small number has been quantitatively evaluated by RIVM (using kriging procedures) to be sufficient to describe regional variation in air pollution. The small size of the country (41,526 km^2) and the lack of mountain ranges contribute to this. The elevation is generally less than 50 m, with a few exceptions in the south with hills up to 321 m. Therefore, the Netherlands can be considered as one airshed.

Table 1. Dependence of the RMSE of prediction ($\mu g/m^3$) of concentrations at network sites, using interpolation from other network sites.

Distance criterion (km)[a]	Power[b] 1	2	3
NO_2			
75	4.07	4.06	4.04
100	4.54	4.33	4.18
150	5.26	4.76	4.41
200	5.67	4.99	4.53
None	5.84	5.10	4.58
BS			
100	1.49	1.48	1.49
150	1.85	1.72	1.64
200	2.04	1.84	1.70
None	2.11	1.87	1.72

[a]Radius of circle in which network sites were used for interpolation.
[b]Power of the weight used for interpolation (inverse of distance to a certain power).

Table 2. Dependence of the RMSE of prediction ($\mu g/m^3$) of concentrations at network sites, using interpolation from other network sites: impact of exclusion of slightly urban sites.

Pollutant	Distance criterion (km)	Power	Sites used for interpolation	RMSE[a]
NO_2	75	2	All (n=24)	3.42
	75	2	w/o urb 3 (n=20)	3.16
BS	100	2	All (n=13)	1.35
	100	2	w/o urb 3 (n=10)	0.97

[a]RMSE calculated only for the sites that are not slightly urban (urbanization degree 3), that is, 20 sites for NO_2 and 10 sites for BS.

First, the average concentration between 1987 and 1990, at all available background network sites, was calculated for NO_2 and BS. Before calculating the average concentration for a station, missing values were estimated to prevent bias in the comparisons across sites. Missing data were estimated by calculating the mean of the daily ratios of the concentration at a station to the daily average concentration at the other stations; this mean ratio was used to multiply the average concentration of all sites of a day with a missing value. The procedure resulted in small adaptations to the mean concentration per station. The correlation between the unadjusted and adjusted 1987–1990 average was 0.971 and 0.998 for BS and NO_2, respectively.

Next, we used inverse distance weighed interpolation to calculate concentrations at the home addresses. In this calculation, the weight for the contribution of a specific network site was the inverse of distance from home address to the site, the inverse of distance squared, or distance to the power three. We studied how the prediction error depended on the power used for weighing distance to a site and the distance criterion used for inclusion of network sites. Prediction errors were estimated by cross-validation using regional network sites. The concentration at all network sites was predicted using regression models developed excluding the data from that site. The difference of the predicted and measured concentration was calculated and squared (Nikiforov et al., 1998). The square root of the mean squared error (RMSE) was finally calculated.

Table 3. Distribution of distance of 1986 home address to major regional roads ("freeways") and major urban roads, N=4973.

	Distance to freeways (m)	Distance to major urban roads (m)
Minimum	0.1	3
1%	43	42
5%	159	104
10%	295	208
25th percentile	661	515
50th percentile	1,366	1,287
Maximum	26,152	23,226

The *urban* component has been estimated on the basis of a regression model with measured concentrations as the dependent variable and address density ("degree of urbanization") of the four-digit postal code area of the site as the independent variable. Address density is a standardized variable developed in 1992 by the Central Bureau of Statistics as a proxy for the intensity of human activities in an area. It measures the number of addresses in a defined area. Data on address density are available for four-digit postal code areas ("neighborhood scale"). For example, the city of Rotterdam (about 600,000 inhabitants) consists of 64 different four-digit postal code areas; small communities typically have only one or two codes. The earliest available GIS database was used (1993). The model has been developed on the basis of data from the NAQMN. From the NAQMN, only sites in the west of the country were used since all four urban background sites in the largest cities are located there. The long-term average (1987–1990) BS and NO_2 concentration data were the dependent variables.

A simple approach has been taken to estimate the *local* component. Indicators of living near major roads were developed. A quantitative estimate for living near a major road was accomplished based on previous measurement campaigns. In the Netherlands, the CAR model (a simple dispersion model for gaseous pollutants) has been developed to estimate street level concentrations (Eerens et al., 1993). However, input data (such as traffic intensity and

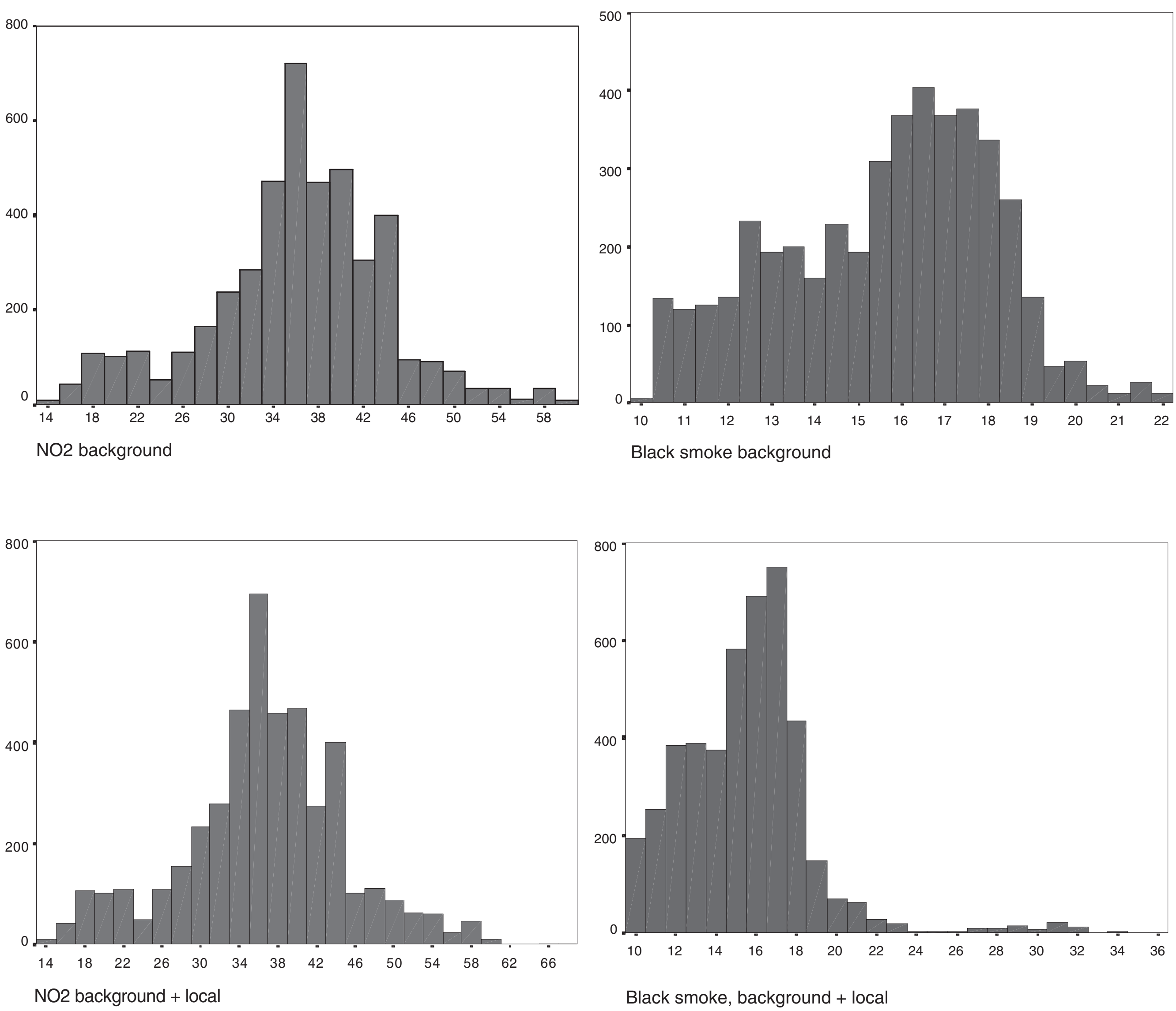

Figure 2. Distribution of long-term average BS and NO_2 exposures at the 1986 home address ($\mu g/m^3$) in the NLCS subcohort.

street configuration) are not available for a large number of streets in the Netherlands. Also, BS cannot be modeled.

Distance to different road types was calculated using a GIS system. Major roads were defined according to a standard classification system of all Dutch roads, available in a GIS system at RIVM (Basisnetwerk). Distances were calculated separately for freeways and major urban roads. A study conducted near two Dutch freeways showed that the largest contrast in concentration of NO_2 and BS occurred within the first 100 m from the freeway (Roorda-Knape et al., 1998). Based on this observation, subjects living within 100 m from a freeway were considered to be exposed, whereas subjects living more than 100 m away from the freeway were considered as unexposed from the specific freeway emissions. Living within 100 m of a freeway has been shown to be associated with increased respiratory symptoms in children (van Vliet et al., 1997). Since traffic intensities on major urban roads are substantially smaller than on freeways and buildings may limit the impact of roads, we considered subjects living within 50 m from a major urban road as exposed to the emissions from this road.

A quantitative estimate for living near a freeway was made by calculating the average contribution of the two studied freeways to the NO_2 and BS concentration at 50 m using the concentration–distance data from Roorda-Knape et al. (1998). The contribution from major inner-city roads was based upon the difference measured between major roads and small side streets in Amsterdam (Fischer et al., 2000). In this study, multiple major roads and side streets were included so that the estimate is not influenced by one specific street configuration only.

Calculations were performed with ArcInfo at RIVM (GIS), Microsoft Excel 97, the Statistical Analysis System (SAS 6.12), and SPSS 7.5.2 for Windows (plots).

Results

With the GIS system for 4973 of 5000 (99.5%) home addresses, geographical coordinates were found. Seven of 27 missing data were due to the fact that subjects only reported a mail box and not their actual address. For the other 20 subjects, no explanation was found.

Regional Component

The cross-validation showed that especially exclusion of sites, which are more than 75 km (NO_2) or 100 km (BS) away, resulted in smaller prediction errors (Table 1). Powers of two and three performed better, especially if distant sites were included in the calculations. To be consistent with work from others (e.g., Kuenzli et al., 1997; Abbey et al., 1999), we used a power of two in subsequent analyses. Exclusion of the three sites with a slightly higher degree of urbanization (30,000–40,000 inhabitants towns) than the other regional background stations resulted in lower prediction errors for BS and NO_2 (Table 2).

The 1987–1990 average BS concentration at the regional network sites ranged from 9.2 to 16.0, with a mean of 12.7 $\mu g/m^3$. The 1987–1990 average NO_2 concentration at the network sites ranged from 14.4 to 35.8, with a mean of 27.7 $\mu g/m^3$. The prediction errors are therefore about 10% of the mean concentration of both pollutants and 15% of the range in concentration.

Urban Background Component

The range in address density in the cohort was from 9 to 7456 addresses/km^2. The average address density was 1785 addresses/km^2; the 10th and 90th percentile were 355 and 3514 addresses/km^2. In the classification system used by CBS, address densities below 500/km^2 are considered as nonurban, whereas address densities above 2500/km^2 are

Table 4. Distribution of estimated long term average (1987–1990) concentrations at 1986 home address in $\mu g/m^3$ on the basis of different spatial scales (N=4973 addresses).

	NO_2			BS		
	Regional	Regional+ urbanization	Regional+ urbanization+local	Regional	Regional+ urbanization	Regional+ urbanization+local
Minimum	14.5	14.7	14.7	9.4	9.6	9.6
1st percentile	15.8	16.9	16.9	9.8	10.1	10.1
10th percentile	19.8	25.4	25.6	10.2	11.5	11.5
25th percentile	28.8	32.4	32.6	11.9	13.1	13.2
50th percentile	31.5	36.4	36.6	13.4	15.5	15.6
75th percentile	32.3	40.8	41.4	14.3	16.9	17.0
90th percentile	33.0	44.8	45.9	15.5	18.0	18.4
95th percentile	33.4	48.2	50.7	15.9	18.6	19.9
99th percentile	33.8	56.9	57.7	16.0	20.5	29.0
Maximum	35.4	59.5	67.2	16.0	21.6	35.8
Mean	29.5	36.1	36.6	13.2	15.1	15.5
SD	4.7	7.9	8.3	1.8	2.5	3.2

considered as very urban. For BS, the regression slope from a model developed on the basis of NAQMN sites was 1.07 (SE 0.36) $\mu g/m^3$ per 1000 addresses/km^2; the R^2 of the model was 0.69. For NO_2, the regression slope was 3.71 (SE 1.02) $\mu g/m^3$ per 1000 addresses/km^2; the R^2 of the model was 0.59.

Prediction errors (RMSE), estimated by cross-validation, were 1.8 and 6.9 $\mu g/m^3$ for BS and NO_2, respectively. The RMSE was 24.7% and 25.9% of the range in background concentration at the network sites for BS and NO_2, respectively.

Local Component

A small percentage of the subcohort lived close to a major road (Table 3). The percentage of subjects living within 100 m of a freeway was 3.0; 1.6% lived within 50 m of a major urban road. The percentage of subjects living within 100 m of a freeway and/or 50 m from a major urban road was 4.5.

The quantitative estimates of the contribution of living 50 m away from a freeway to the BS and NO_2 concentration were 4.4 and 11 $\mu g/m^3$, respectively. The contribution from major inner-city roads was estimated to be 13 and 8 $\mu g/m^3$ for BS and NO_2, respectively. These estimates were assigned to each "exposed" address, independent of the actual distance to the road.

In summary, the estimation method consisted of interpolation of regional background stations from the NAQMN; estimation of urban background increase as BS=1.07×address density/1000 and NO_2=3.71×address density/1000 (BS, NO_2 in $\mu g/m^3$; address density in km^{-2}); characterization of the local contribution with indicator variables indicating whether a subject lived within 100 m from a freeway or 50 m from a major city road. For subjects living within 100 m from a freeway or 50 m from a major city road, quantitative estimates were added to the urban and regional background: for a freeway NO_2 11 $\mu g/m^3$ and BS 4.4 $\mu g/m^3$ and for a major inner city road NO_2 8 $\mu g/m^3$ and BS 13 $\mu g/m^3$.

Exposure Contrast in the Subcohort

The average of the smallest distance to a monitoring site for the subcohort members was 38 and 21 km for BS and NO_2, respectively. None of the subcohort members lived more than 100 km from a BS site or 75 from an NO_2 site. Thus, an estimate of the regional background was obtained for all cohort members for whom geographical coordinates were found.

When the above model was applied to the 1986 home addresses of the subcohort, substantial variability existed in the estimated long-term exposure of the components BS and NO_2 (Table 4, Figure 2). If only the regional scale is taken into account, the range of exposure is approximately a factor 2. The difference in concentration between the 10th and 90th percentile is about 52% for BS and 67% for NO_2. There is only a small a contrast between the 25th and 75th percentiles. If differences in urbanization degree are taken into account, the contrasts within the cohort increase, especially for NO_2. The difference in concentration between the 10th and 90th percentile is 57% for BS and 76% for NO_2. The effect of the local scale is mostly in the highest concentrations, consistent with the small number of subjects living near a major road. Figure 3 demonstrates that the local scale is more important for BS than for NO_2. This is consistent with the fact that BS is a primary pollutant, whereas NO_2 is largely a secondary pollutant.

Figure 3 illustrates the impact of ignoring spatial variation related to degree of urbanization and presence of a major road. The difference for individual addresses can be substantial. On average, the "exposure error" (assuming

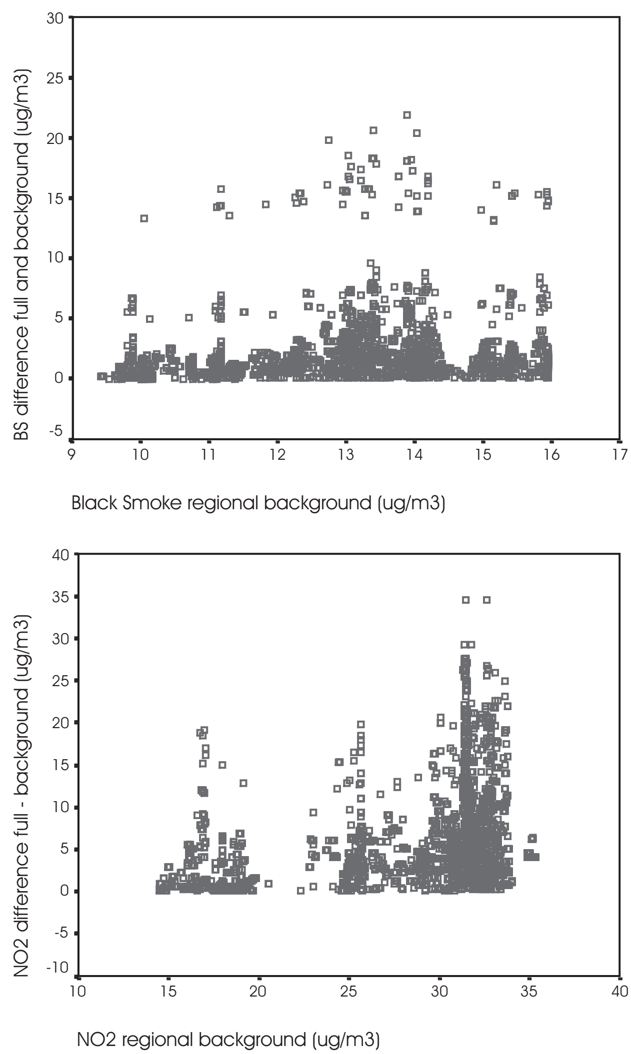

Figure 3. Difference between full air pollution exposure (sum of regional, urban, and local scale) and regional background versus regional background.

Table 5. Distribution of duration of residence (year) at 1986 address (N=4379).

Percentile	Residence duration (year)
0	0
10	9
25	19
50	33
Mean	35.3
75	56
90	63
100	70

that the full exposure incorporating the three spatial levels is the true exposure) was slightly higher for homes with higher regional background. The correlation between the regional background and the difference between full and regional exposure was 0.14 for BS and 0.31 for NO_2. Note that due to the separate estimation of three elements, regional exposures cannot be smaller than the full exposure.

Residential Address History

The question on past towns of residence was not completed by 409 subjects (8.2%). An additional 577 subjects did not fill out the question completely, that is, the final year was not equal to 1986. A substantial fraction of the latter subjects had not completed the last "until year" but did complete the "from year" and the town of residence. We assumed that if the town of residence was the same as the 1986 town, subjects implied that they still lived at their current address. This was the case for 365 subjects. Thus, duration of residence could be calculated for 4379 subjects (87.6%). A large percentage (89.8%) of the study participants lived 10 years or more at its present address (Table 5), suggesting that estimates of recent (1986) exposure are relevant for historic exposure as well.

Discussion

We have documented possibilities to take into account within-city variation of air pollution concentration in an ongoing cohort study of the effect of long-term exposure to air pollution on mortality. Previous mortality cohort studies assigned one concentration value to subjects living within the same city/metropolitan area (Dockery et al., 1993; Pope et al., 1995) or used interpolation of concentrations from the nearest network sites at the zip code centroid of the residential address (Abbey et al., 1995). In the latter study, differences within metropolitan areas were thus taken into account if the network was sufficiently dense. An explicit distinction between site types (e.g., regional versus urban background sites) was not made. None of the three studies took proximity to major roads into account.

Application of GIS

We have applied a GIS to convert address information into geographical coordinates efficiently, obtain data on address density in the neighborhood of the address, and calculate distances to objectively defined major roads. Geographical coordinates enabled us to interpolate concentrations from network sites at the residential address. The calculation of distances to major roads allowed us to efficiently obtain objective estimates of proximity to major roads. Several epidemiological studies on traffic emissions have relied on self-reported traffic density (e.g., Weiland et al., 1994), which may be biased depending on disease status of the subjects.

Address geocoding based upon the six-position postal code was successful for 99.5% of the study population. For studies in the US, it has been reported that automated address matching was poor for rural areas (match rate of 20–30%) and much higher (98%) for the largest urbanized county in North Carolina (Vine et al., 1997). A study in San Diego County reported a 85% address match, with missing data due in part to listing of PO boxes and incomplete addresses (English et al., 1999).

Several of the reported limitations of using a GIS (Vine et al., 1997) apply to our study as well. The major problem was the availability of historic data. The GIS system did not contain data for the period 1985–1990. The earliest address density data that were available were for 1993.

Regional Background

Interpolation using inverse distance weighing was used to obtain an estimate of the regional background. Cross-validation showed that the interpolation prediction error was about 10% of the mean and 15% of the regional contrast in concentration, which documents that interpolation was reasonably good. The prediction error is similar to the 13% prediction error (expressed as percent of the mean) obtained for estimated ozone concentrations across the US (Nikiforov et al., 1998). In that study, the interpolation error was 5% when expressed as a percentage of the range in concentration. The larger interpolation errors in our study are consistent with the difference in pollutants: NO_2 and BS have more local source impacts than ozone, which is largely a regional pollutant.

The smallest prediction error was obtained if stations far away were not used. In the study of Nikiforov et al. (1998), the interpolation methods resulted in the lowest prediction errors when three sites were included. Inclusion of more (distant) sites resulted in an increase in prediction errors. The study also showed that for the optimal number of sites, there was little difference between a regression-based approach and inverse distance squared weighted averages.

Other investigators have used kriging to interpolate concentrations (Wartenberg et al., 1991; Liu and Rossini, 1996). Kriging assumes that data from geographically close sites are more similar than more distant sites (see Ito et al., 2001). The procedure uses a function describing the correlation versus separation distance. We have not used this approach because RIVM has used kriging procedures to decrease the density of the network in 1986. Therefore, it is not possible to estimate the spatial covariance function anymore. Nikiforov et al. (1998) concluded that kriging and their regression-based approach were nearly identical.

Urban Background

Urban background concentrations were estimated using a regression model with address density in the four-position postal code area as the independent variable for three reasons. *First*, the NAQMN does not have monitoring sites in all cities. *Second*, we did not want to assign concentration values from the nearest measurement site without taking into account the nature of the site (regional, urban background). Because NO_2 concentrations vary on a fine spatial scale, direct interpolation will result in systematic prediction errors related to emission sources in the area (Stedman et al., 1997). On a smaller spatial scale, the SAVIAH study has found that regression-based methods using external information, such as distance to major roads, were superior to interpolation methods in explaining within-city variation of NO_2 concentrations (Briggs et al., 1997). In a UK study (Stedman et al., 1997), the percentage of urban and suburban land cover at two spatial scales (100 and 25 km^2 around the sites) and estimated emission from major road vehicle sources in an area of 4 km^2 around the site were used as predictors in a regression model for NO_2. The developed model correlated moderately high with independently measured concentrations (r=0.6). *Third*, differences within cities were accounted for as well. Several studies have documented differences in NO_2 concentrations between the center and suburb (Bernard et al., 1997; Raaschou-Nielsen et al., 2000). In a study in Copenhagen, NO_2 background concentrations declined exponentially from the city center: concentration ($\mu g/m^3$)=25.4×exp(−0.39×distance(km)). In a study in Montpellier, NO_2 background concentrations declined linearly with approximately 5 $\mu g/m^3$ for each kilometer distance from the city center (Bernard et al., 1997). One problem in several Dutch cities is how to define the city center: in some cities, the commercial center is located at the edge of the city. This difference between cities is taken into account by calculating address density in the four-digit postal code area.

The regression slope for NO_2 (3.71 $\mu g/m^3$ per 1000 addresses/km^2) compared well with the regression slope for outdoor measurements (2.97 $\mu g/m^3$ per 1000 addresses/km^2) from another Dutch study (Janssen et al., 1999). In that study, independent indoor (schools) and outdoor measurements were conducted in 1997 at 20 sites in the north and west of the country. The regression slope for BS (1.07 $\mu g/m^3$ per 1000 addresses/km^2) was much higher than for the outdoor measurements in the Janssen et al. study (0.22 $\mu g/m^3$ per 1000 addresses/km^2), but similar to the indoor measurements in two regions (0.78 and 0.74 $\mu g/m^3$ per 1000 addresses/km^2). The reasons for the discrepancy with the outdoor BS data are not clear. One of the least urban sites had higher concentrations than most of the other sites and may have affected the regression analysis.

Local Scale

Estimation of concentration contrasts at the local scale due to proximity of major roads is difficult because many factors determine the long-term average concentration. Important factors are light and heavy duty traffic intensity, traffic speed, street geometry, presence of trees, and distance of home front to the street (Eerens et al., 1993; Raaschou-Nielsen et al., 2000). In the Netherlands, input data for dispersion models, such as traffic intensity, are not available for a large number of streets. Since the number of subjects living near major roads was small, a simple approach based on measurement data was used that assigned the same estimate to subjects living near major roads. Because of the uncertainty of the quantitative estimate of the local contribution for these subjects, exposure on the local scale was also characterized by proximity indicator variables. An additional argument for not adding background and local scale concentrations is that they may be differently related to the personal exposure to outdoor air pollution. In a study in Switzerland, the effect of NO_2 on lung function estimated for within- and between-city air pollution exposure contrasts was different (Schindler et al., 1998). The magnitude of the NO_2 effect estimated from a between city comparison was about three times higher than that estimated from within city comparisons.

If actual traffic intensity data are more readily available than in the Netherlands, calculations of the traffic flow near the home address could be performed. In a study in San Diego, traffic flow in buffers of 550 ft was used to characterized local scale exposures (English et al., 1999).

Limitations of the Method

The currently applied methodology attempts to estimate ambient exposure to air pollution at the residential address. Ideally, we would like to estimate the long-term average personal exposure to ambient air pollution. The data for such an estimate will in practice never be available. Most studies are therefore limited to estimating ambient exposure. An exception is the work by Kuenzli et al. (1997) who incorporated time–activity patterns of subjects affecting personal exposure into estimates of individual ozone exposures. The methods of the Kuenzli study were not feasible in the current study since we added air pollution exposure assessment to an already ongoing study (designed for another purpose) that had not collected time–activity data. An important observation in the SAVIAH study conducted in Amsterdam was that concentration differences between major and small side streets were found both in outdoor and indoor air (Fischer et al., 2000). For BS, the ratio between major and small streets was 1.84 for outdoor and 1.83 for indoor air (NO_2 was not measured indoors). Because of the large amount of time that people spent in their own home, a large fraction of the exposure to traffic-related air pollution may thus occur in the home.

A validation study conducted in the Netherlands in 1997–1998 showed that personal NO_2 exposures of children living in three communities with different degrees of urbanization differed significantly (Rijnders et al., 2001). The differences in outdoor and personal NO_2 concentration between the nonurban and the very urban location were 11 and 10 $\mu g/m^3$, respectively. This strongly supports the use of address density as a variable to estimate air pollution exposures.

Recently, the use of residence as an indicator of exposure in air pollution epidemiology has been criticized, mostly because the used indicators of exposure based on residence such as distance to a source do not necessarily reflect the true spatial pattern of exposure (Huang and Batterman, 2000). Most of the criticisms do not apply to the current study. For the regional and urban scale, we have used residence only to define the geographic position and assign quantitative concentrations based upon actual measurements. For the local scale, measurements have documented that proximity to a major road is related to increased concentrations of NO_2 and BS (Roorda-Knape et al., 1998; Fischer et al., 2000). One clear limitation that does apply is that we have assumed that everyone who lives close to any major road is exposed to the same concentrations. This is clearly a simplification.

In the current study, only home address was available. We were therefore not able to construct an estimate of air pollution exposure based upon both home and work addresses as in the AHSMOG study (Abbey et al., 1999).

The average concentration for 1987–1990 was used as the exposure estimate. These are the first years of the follow-up period of the cohort study. This is comparable to the Dockery et al. and Pope et al. studies. We did not estimate exposures before 1987.

Exposure Contrast

Within the subcohort, there was substantial contrast in estimated exposure to BS and NO_2. The (regional plus urban) background exposure contrast in the subcohort was somewhat smaller than in the two American cohort studies that formed the basis for initiating the current study. In the Six Cities study, $PM_{2.5}$ concentrations ranged from 11.0 to 29.6 $\mu g/m^3$ (Dockery et al., 1993). The range of exposure in the present study is larger than in the Six City study, but in that study, each individual within a city was assigned the same value, whereas in the present study, individual estimates have been generated. In the ACS study (Pope et al., 1995), the range of fine particles was from 9.0 to 33.5 $\mu g/m^3$, with 10th and 90th percentiles of about 10 and 25 $\mu g/m^3$.

Conclusions

We have illustrated possibilities to take into account within-city variation of air pollution concentration in an ongoing cohort study of the effect of long-term exposure to air pollution on mortality. The exposure model involved separate estimation of regional background, urban background, and local scale concentrations. The methodology included the use of interpolation and regression modeling supported by GIS methods. Interpolation resulted in reasonably precise regional background estimation, with prediction errors about 15% of the range in regional background concentration. Prediction errors for estimating the urban background were about 25% of the range in background concentrations. Ignoring the urban and local scale resulted in substantial errors that were moreover correlated with the regional background.

References

Abbey D.E., Lebowitz M.D., Mills P.K., et al. Long-term ambient concentrations of particulates and oxidants and development of chronic disease in a cohort of non-smoking, California residents. *Inhalation Toxicol* 1995: 7: 19–34.

Abbey D.E., Nishino N., McDonnell W.F., Burchette R.J., Knutsen S.F., Beeson W.L., and Yang J.X. Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *Am J Respir Crit Care Med* 1999: 159: 373–382.

Ackermann-Liebrich U., Leuenberger P., Schwartz J., Schindler C., Monn C., Bolognini G., Bongard J.P., et al. Lung function and long-term exposure to air pollutants in Switzerland. *Am Crit Care Med* 1997: 155: 122–129.

Bernard N.L., Astre C.M., Vuillot B., Saintot M.J., and Gerber M.J. Measurement of background urban nitrogen dioxide pollution levels with passive samplers in Montpellier, France. *J Expos Anal Environ Epidemiol* 1997: 7: 165–178.

Bloemen H.J.Th., van Bree L., Buringh E., Fischer P.H., Loos S. de, Marra M., and Rombout P.J.A. Fine particles in the Netherlands. Report 650010006. National Institute of Public Health and Environment, 1998 (in Dutch with English summary).

Briggs D.J., Collins S., Pryl K., Smallbone K., and van der Veen A. Mapping urban air pollution using GIS: a regression-based approach. *Int J Geogr Inf Sci* 1997: 11: 699–718.

Brunekreef B. Air pollution and life expectancy: is there a relation? *Occup Environ Med* 1997: 54: 781–784.

Brunekreef B., Janssen N.A.H., de Hartog J., Harssema H., Knape M., and van Vliet P. Air pollution from truck traffic and lung function in children living near motorways. *Epidemiology* 1997: 8: 298–303.

Chow J. Measurement methods to determine compliance with ambient air quality standards for suspended particles. *J Air Waste Manage Assoc* 1995: 45: 320–382.

Dockery D.W., Pope C.A. III, Xu X., Spengler J.D., Ware J.H., Fay M.E., Ferris B.G., and Speizer F.E. An association between air pollution and mortality in six US cities. *N Engl J Med* 1993: 329: 1753–1759.

Dockery D.W., Cunningham J., Damokosh A.I., Neas L.M., Spengler J.D., Koutrakis P., Ware J.H., and Speizer F.E. Health effects of acid aerosols on North American children: respiratory symptoms. *Environ Health Perspect* 1996: 104: 500–505.

Edwards J.D., Ogren J.A., Weiss R.E., and Charlson R.J. Particulate air pollutants: a comparison of British "smoke" with optical absorption coefficient and elemental carbon concentration. *Atmos Environ* 1983: 17: 2337–2341.

Edwards J., Walters S., and Griffith R.K. Hospital admissions for asthma in preschool children: relationship to major roads in Birmingham, United Kingdom. *Arch Environ Health* 1994: 49: 223–227.

Eerens H.C., Sliggers C.J., and van Hout K.D. The CAR model: the Dutch method to determine city street air quality. *Atmos Environ* 1993: 27B: 389–399.

Elskamp H.J. National air quality monitoring network. Technical description. Report 228702017. National Institute of Public Health and Environmental Protection, Bilthoven, the Netherlands, 1989.

English P., Neutra R., Scalf R., Sullivan M., Waller L., and Zhu L. Examining associations between childhood asthma and traffic flow using a geographical information system. *Environ Health Perspect* 1999: 107: 761–767.

Fischer P.H., Hoek G., Reeuwijk H., Briggs D.J., Lebret E., van Wijnen J.H., Kingham S., and Elliott P.E. Traffic-related differences in outdoor and indoor concentrations of particles and organic compounds in Amsterdam. *Atmos Environ* 2000: 34: 3713–3722.

Hoek G., Mennen M.G., Allen G.A., Hofschreuder P., and vd Meulen T. Concentrations of acidic air pollutants in the Netherlands. *Atmos Environ* 1996: 30: 3141–3150.

Hoek G., Forsberg B., Borowska M., et al. Wintertime PM_{10} and Black Smoke concentrations across Europe: results from the PEACE study. *Atmos Environ* 1997: 31: 3609–3622.

Huang Y.L., and Batterman S. Residence location as a measure of environmental exposure: a review of air pollution epidemiology studies. *J Expos Anal Environ Epidemiol* 2000: 10: 66–85.

Ito K., Thurston G.D., Nadas A., and Lippmann M. Monitor-to-monitor temporal correlation of air pollution and weather variables in the North Central US. *J Expos Anal Environ Epidemiol* 2001: 11: 21–32.

Janssen N.A.H., van Vliet P.H.N., Harssema H., Brunekreef B., and Fischer P.H. Ontwikkeling methodiek voor het schatten van langdurige blootstelling aan verkeersgerelateerde luchtverontreiniging, voor toepassing in onderzoek naar chronische effecten van luchtverontreiniging. Rapport 1999-485. Wageningen Universiteit, afdeling Gezondheidsleer (in Dutch).

Janssen N.A.H., van Vliet P.H.N., Aarts F., Harssema H., and Brunekreef B. Assessment of exposure to traffic related air pollutants of children attending schools near motorways. *Atmos Environ* 2001: 35: 3875–3884.

Kuenzli N., Lurmann F., Segal M., Ngo L., Balmes J., and Tager I.B. Association between lifetime ambient ozone exposure and pulmonary function in college freshmen — results from a pilot study. *Environ Res* 1997: 72: 8–23.

Lebret E., Briggs D., van Reeuwijk H., Fischer P., Smallbone K., Harssema H., Kriz B., Gorynski P., and Elliott P. Small area variations in ambient NO_2 concentrations in four European areas. *Atmos Environ* 2000: 34: 177–185.

Liu L.J.S., and Rossini A.J. Use of kriging models to predict 12-hour mean ozone concentrations in metropolitan Toronto — a pilot study. *Environ Int* 1996: 22: 677–692.

Nikiforov S.V., Aggarwal M., Nadas A., and Kinney P.L. Methods for spatial interpolation of long-term ozone concentrations. *J Expos Anal Environ Epidemiol* 1998: 8: 465–481.

Pope C.A. III, Thun M.J., Namboodiri M.M., et al. Particulate air pollution as a predictor of mortality in a prospective study of US adults. *Am J Respir Crit Care Med* 1995: 151: 669–674.

Raaschou-Nielsen O., Hertel O., Vignati E., Berkowicz R., Jensen S.S., Larsen V.B., Lohse C., and Olsen J.H. An air pollution model for use in epidemiological studies: evaluation with measured levels of nitrogen dioxide and benzene. *J Expos Anal Environ Epidemiol* 2000: 10: 4–14.

Rijnders E., Janssen N.A.H, Vliet P.H.N, and Brunekreef B. Personal and outdoor nitrogen dioxide concentrations in relation to degree of urbanization and traffic density. *Environ Health Perspect* 2001: 109 (Suppl): 411–417.

Roorda-Knape M.C., Janssen N.A.H., de P van Hartog J.J., Harssema H., and Brunekreef B. Air pollution from traffic near major motorways. *Atmos Environ* 1998: 32: 1921–1930.

Schindler C., Ackermann-Liebrich U., and Leuenberger P., et al. Associations between lung function and estimated average exposure to NO_2 in eight areas of Switzerland. *Epidemiology* 1998: 9: 405–411.

Stedman J.R., Vincent K.J., Campbell G.W., Goodwin J.W.L., and Downing C.E.H. New high resolution maps of estimated background ambient NO_x and NO_2 concentrations in the UK. *Atmos Environ* 1997: 31: 3591–3602.

van den Brandt P.A., Goldbohm R.A., van't Veer P., Volovics A., Hermus R.J.J., and Sturmans F. A large-scale prospective cohort study on diet and cancer in the Netherlands. *J Clin Epidemiol* 1990: 43: 285–295.

van der Wal J.T., and Janssen L.H.J.M. Analysis of spatial and temporal variations of PM_{10} concentrations in the Netherlands using Kalman filtering. *Atmos Environ* 2000: 34: 3675–3687.

van Vliet P., Knape M., de Hartog J., Janssen N., Harssema H., and Brunekreef B. Motor vehicle exhaust and chronic respiratory symptoms in children living near motorways. *Environ Res* 1997: 74: 122–132.

Vine M.F., Degnan D., and Hanchette C. Geographic Information Systems: their use in environmental epidemiologic research. *Environ Health Perspect* 1997: 105: 598–605.

Wartenberg D., Uchrin C., and Coogan P. Estimating exposure using kriging: a simulation study. *Environ Health Perspect* 1991: 94: 75–82.

Weiland S.K., Mundt K.A., Rueckmann A., and Keil U. Self-reported wheezing and allergic rhinitis in children and traffic density on street of residence. *Ann Epidemiol* 1994: 4: 243–247.

Wjst M., Reitmeir P., Dold S., et al. Road traffic and adverse effects on respiratory health in children. *Br Med J* 1993: 307: 596–600.