Air pollution and health in Scotland: a multicity study

DUNCAN LEE*, CLAIRE FERGUSON
Department of Statistics, University of Glasgow, Glasgow, G12 8QQ UK
duncan@stats.gla.ac.uk

RICHARD MITCHELL
Public Health and Health Policy, University of Glasgow, Glasgow, G12 8QQ UK

SUMMARY
This paper presents an epidemiological study investigating the effects of long-term air pollution exposure on public health in Scotland, focusing on the 4 major urban areas, Aberdeen, Dundee, Edinburgh, and Glasgow. In particular, the associations between respiratory hospital admissions in 2005 and exposure to both PM$_{10}$ and NO$_2$ between 2002 and 2004 are estimated using a small-area ecological design. The implementation of such studies requires careful consideration of a number of statistical issues, including how to model spatial correlation, identifiability of the model parameters, and the possible effects of ecological bias. The results show that long-term exposures (over 3 years) to PM$_{10}$ and NO$_2$ are significantly associated with respiratory hospital admissions in Edinburgh and Glasgow, whereas the risks for Aberdeen and Dundee are generally positive but nonsignificant.

Keywords: Air pollution and health; Bayesian spatial modeling; Ecological bias.

1. INTRODUCTION
The adverse effects on health associated with air pollution exposure came to public prominence in the mid 1900s as a result of high air-pollution episodes that caused large numbers of excess deaths (Ministry of Public Health, 1954; Firket, 1936). Since then, numerous studies investigating both the short- and the long-term effects of various constituents of air pollution have been conducted, with the majority finding positive associations between exposure and both mortality and morbidity events. Short-term studies are typically based on an ecological (at the population level) time series design, in which counts of mortality or morbidity events on a given day are related to pollution exposures on the preceding few days. Examples include the National Morbidity, Mortality, and Air Pollution Study (Dominici and others, 2002) and Air Pollution and Health: a European Approach (Katsouyanni and others, 2001), which are multicity studies based on America and Europe, respectively.

Long-term studies estimate the cumulative effects on health of exposure over a number of years and are based on either individual or ecological designs. Individual-level studies relate to a large cohort of people whose health status is periodically assessed and subsequently related to ambient pollution concentrations.
In contrast, ecological studies have a small-area spatial design, with annual numbers of health events in each area being regressed against average pollution concentrations from preceding years. Examples include Maheswaran and others (2005) and Elliott and others (2007), with the latter investigating the relationship between black smoke and a range of mortality metrics at the electoral ward level in Great Britain.

This paper presents a Scottish study of the health effects associated with long-term exposure to air pollution and is based on an ecological small-area spatial design. Our study focuses on the 4 main urban centers in Scotland. We discuss the issues that arise in such studies, including the presence of spatial correlation, identifiability of the model parameters, and the problem of ecological bias. The remainder of this paper is organized as follows. Section 2 reviews the background of the study as well as its design. Section 3 describes our statistical modeling approach. Section 4 presents the results of our study. Section 5 is a concluding discussion.

2. Background and study design

A number of recent studies have investigated the association between air pollution and public health in Great Britain, looking at the possible effects of short-term (Lee and Shaddick, 2008) and long-term (Elliott and others, 2007) exposure. However, the majority of these studies relate to England, and especially to London, with relatively few assessing the effects in Wales or Scotland. In Scotland, Prescott and others (1998) and Carder and others (2008) have investigated the association between short-term exposure to black smoke and respiratory hospital admissions and mortalities, respectively, reporting increased risks of 3.1% (admissions) and 5.36% (mortalities) for a 10 µg m⁻³ increase in black smoke concentrations. In addition, Yap and others (2005) have investigated the effects of long-term black smoke exposure via a cohort study and reported a hazard ratio for lung cancer mortality of 1.227 for a 10 µg m⁻³ increase.

In this paper, we present a study that investigates the detrimental effects on health that are associated with long-term pollution exposure, using a small-area spatial design and making use of a recently created database of small-area statistics. In February 2007, the Scottish Executive introduced the Scottish Neighbourhood Statistics database and website (Scottish Government, 2007, http://www.sns.gov.uk/), which was an initiative to “improve the availability, consistency, and accessibility of small-area statistics in Scotland”. The database is free to access and includes information about the environment, demography, deprivation, and health at the small-area level throughout Scotland. The data take the form of yearly summaries at a number of spatial resolutions, ranging in size from 14 health boards to 1235 intermediate geographies (IGs) or 6505 data zones (DZs). The latter 2 are administrative regions designed specifically for the distribution of small-area statistics in Scotland, containing median populations of 3956 (IGs) and 768 (DZs) and covering median areas of 173 (IG) and 23 (DZ) ha, respectively. IGs are therefore similar in size to “electoral wards,” which have been used as the basic areal unit in a recent study in Great Britain (Elliott and others, 2007).

2.1 Spatial resolution

In this study, we estimate the pollution–health associations at both IG and DZ resolutions. This allows us to compare our results across spatial scales. DZs are nested exactly within IGs, with between 2 and 9 DZs comprising each IG.

2.2 Study region

For this study, we focus on the 4 main urban centers in Scotland: Aberdeen, Dundee, Edinburgh, and Glasgow. These areas have been selected because they contain a substantial proportion of the population
and have a range of pollution concentrations. In order to obtain a sufficient number of IGs in each of these locations, we specify the boundary of each region as the health boards that contain these cities, namely Grampian, Tayside, Lothian, and Greater Glasgow and Clyde (now shortened to Greater Glasgow). These health boards contain both rural and urban areas, giving a greater contrast in pollution concentrations within each region. A map showing the locations of the 4 health boards within Scotland is provided in Figure 1, while a summary of the regions is given in Table 1. The population density is not evenly distributed throughout Scotland, with 23% of the population residing in Greater Glasgow, which covers only 1.5% of the land area. In population terms, Tayside (90 IGs and 496 DZs) is the smallest, while Greater Glasgow (271 IGs and 1473 DZs) is the largest.

2.3 Health data

In this paper, we investigate the association between air pollution and hospital admissions due to respiratory diseases, which are classified by the International Classification of Disease Tenth Revision as J00-J99 and R09.1. These data are available as yearly total of admissions for 2005 at both DZ and IG resolutions and are summarized in Table 2 and Figure 2. The average number of admissions in an IG is higher in Greater Glasgow and Tayside than in Grampian and Lothian, which is most likely due to the higher levels of socioeconomic deprivation in these cities. This hypothesis is reinforced by the relative levels of income deprivation observed in Table 2 and is due to the fact that poorer populations have a higher prevalence of smoking and a greater likelihood of comorbidity. Glasgow and Dundee contain marked concentrations of deprivation and poor health, with Glasgow in particular containing some of Europe’s poorest and least healthy people (Leyland and others, 2007).

2.4 Pollution data

Ambient air pollution concentrations are routinely measured at a number of monitoring sites across Scotland, details of which are available from http://www.scottishairquality.co.uk/index. These data are concentrations at single fixed locations at a point in time, rather than being aggregated averages over time and space. As a result, only modeled average pollution concentrations, based on these data and dispersion models, are available for this study and therefore are subject to uncertainty. However, there is no available information about the size of this uncertainty, so throughout this study we treat the pollution data as known average concentrations. These modeled estimates are of particulate matter (measured as PM$_{10}$) and nitrogen dioxide (NO$_2$, µg m$^{-3}$) and relate to average exposure over the 3-year period from 2002 to 2004. They are summarized in Table 2 and Figure 3 and relate to concentrations at urban background locations. Such background concentrations are more indicative of population average exposure than are roadside concentrations and are therefore more appropriate for our study. The table shows that the concentrations are higher in Lothian and Greater Glasgow than in Tayside and Grampian, which is probably due to the former containing larger populations and increased traffic volume. Further details of the exposure estimates and their calculation are available from http://www.scottishairquality.co.uk/maps.php. The use of such modeled pollution estimates is common in recent spatial studies, as individual exposure data are not available and pollution monitoring networks are not dense at the small-area scale; see, for example, Jerrett and others (2005) and Maheswaran and others (2005).

2.5 Covariate data

In addition to hospital admissions and pollution concentrations, the database contains measures of demography and deprivation, both of which are likely to be important covariates. The age and sex distributions
Fig. 1. The locations and sizes of the 4 health boards. The dots show the population density.
Table 1. Summary of the population, area, number of DZs and IGs for each of the 4 study regions. The population and area are also provided as percentage of the Scottish population and mainland area

<table>
<thead>
<tr>
<th>Health board</th>
<th>City</th>
<th>Population</th>
<th>Area (ha)</th>
<th>DZ (no.)</th>
<th>IG (no.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grampian</td>
<td>Aberdeen</td>
<td>525 930</td>
<td>874 493 (11%)</td>
<td>684</td>
<td>128</td>
</tr>
<tr>
<td>Tayside</td>
<td>Dundee</td>
<td>389 740</td>
<td>755 179 (10%)</td>
<td>496</td>
<td>90</td>
</tr>
<tr>
<td>Lothian</td>
<td>Edinburgh</td>
<td>792 600</td>
<td>172 937 (2.2%)</td>
<td>992</td>
<td>177</td>
</tr>
<tr>
<td>Greater Glasgow</td>
<td>Glasgow</td>
<td>1 187 062</td>
<td>115 808 (1.5%)</td>
<td>1473</td>
<td>271</td>
</tr>
</tbody>
</table>

Table 2. Summary of the mean (standard deviation) number of admissions, pollution concentrations, and socioeconomic deprivation (illustrated by the percentage of the population who are income deprived) for each of the 4 study regions across all IGs

<table>
<thead>
<tr>
<th>Health board</th>
<th>Admissions</th>
<th>PM$_{10}$</th>
<th>NO$_2$</th>
<th>% Income deprivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grampian</td>
<td>53.4 (24.2)</td>
<td>11.9 (1.7)</td>
<td>10.0 (6.8)</td>
<td>9.1 (6.1)</td>
</tr>
<tr>
<td>Tayside</td>
<td>61.2 (27.3)</td>
<td>12.1 (1.0)</td>
<td>11.0 (4.9)</td>
<td>13.0 (8.3)</td>
</tr>
<tr>
<td>Lothian</td>
<td>59.7 (27.2)</td>
<td>14.2 (1.1)</td>
<td>18.2 (5.2)</td>
<td>11.5 (8.7)</td>
</tr>
<tr>
<td>Greater Glasgow</td>
<td>68.8 (29.9)</td>
<td>13.9 (1.6)</td>
<td>21.4 (7.2)</td>
<td>19.2 (12.1)</td>
</tr>
</tbody>
</table>

in each IG or DZ are unlikely to be identical. These differences are accounted for by calculating the expected number of admissions in each area based on external rates. These expected numbers are calculated from UK admissions rates, which in turn are based on data from the European hospital morbidity database (World Health Organization, 2006). Strong relationships between socioeconomic status and health have been demonstrated repeatedly at international, national, regional, and local scales (Mackenbach and others, 1997). Socioeconomic and environmental adversity are also closely associated (Brulle and Pellow, 2006), and exposure to harmful environments may be one way in which socioeconomic inequalities in health are created and sustained. A large number of area-level measures of deprivation are available in this study, including metrics within the general domains of income, health, housing, employment, education, access to services, and crime. A categorical urban–rural variable (with 6 levels) is also available.

3. Methods

In this section, we provide a brief overview of the spatial regression methods used in our study, including the general model specification (3.1), how we allow for ecological bias (3.2), and the method of inference used (3.3). Full details are provided in the supplementary material available at Biostatistics online, http://www.biostatistics.oxfordjournals.org.

3.1 Spatial regression models

Denote the vector of hospital admission counts by $y = (y_1, \ldots, y_n)$, where $y_k$ is the number of admissions in area $k$ in 2005. These counts are regressed against modeled average air pollution concentrations (over 2002–2004) $x = (x_1, \ldots, x_n)$ and a matrix of $q$ covariates $Z = (z_{1T}, \ldots, z_{nT})$ which include measures of socioeconomic deprivation and geographical location. A general model for such spatial count data was proposed by Besag and others (1991) and has been adopted in later reviews by Banerjee and others (2004)
Fig. 2. The standardized morbidity ratios (observed/expected) for respiratory hospital admissions at IG resolution.

and Wakefield (2007). It is given by

$$y_k \sim \text{Poisson}(E_k \theta_k), \quad \text{for } k = 1, \ldots, n,$$

$$\theta_k = \exp(z_k^T \beta_z + x_k \beta_x + V_k + U_k),$$ (3.1)

where $E_k$ denote the expected number of admissions in area $k$. The specification of a linear relationship represented by $(\beta_z, \beta_x)$ between each covariate and the response is not required, and the use of flexible additive formulations is a straightforward extension. The random effects $(U_k, V_k)$ model spatial correlation and overdispersion, respectively, in the response data. This dual specification is common in both disease mapping and spatial regression contexts. The nonspatial random effects $V_k$ are modeled as independent Gaussian random variables, $V_k \sim \mathcal{N}(0, \sigma^2)$, while spatial dependence can be modeled either by directly specifying a joint distribution for $U = (U_1, \ldots, U_n)$ or equivalently by a set of $n$ conditional distributions $U_k|U_{-k}$, where $U_{-k} = (U_1, \ldots, U_{k-1}, U_{k+1}, \ldots, U_n)$. 
Under a joint specification, $U$ is assigned a multivariate Gaussian distribution, $U \sim N(0, \tau^2 \Sigma(\phi))$. We assume an isotropic exponential correlation structure because of its simplicity. In addition, a discrete uniform prior is specified for $\phi$ as adopted in Diggle and Ribeiro (2007), where the hyperparameters are chosen so that between 5% and 30% of pairs of areas have a substantial correlation above 0.5. A discrete prior is adopted here because it enables the covariance matrix to be inverted once prior to the simulation algorithm rather than at each iteration, easing the computational burden. For identifiability and ease of prior elicitation, we reparameterize from $(\sigma^2, \tau^2)$ to $(\lambda, \rho)$ as suggested by Wakefield (2007), where $\lambda = \sigma^2 + \tau^2$ and $\rho = \tau^2 / (\sigma^2 + \tau^2)$, so that $\lambda$ is the total variance associated with both random effects while $\rho$ is the proportion that is spatial.

Under a conditional specification, $U_k|U_{-k}, W \sim N(\bar{U}_k, \tau^2/n_k)$ for all $k$, which is the intrinsic conditionally autoregressive model (ICAR) proposed by Besag and others (1991). Here, $W$ is a neighborhood matrix containing 1s if two areas are neighbors and 0s otherwise, and we adopt the specification that areas are neighbors if they share a common border. Thus, the full conditional of $U_k$ has expectation equal to the
mean of its neighbors and a variance that is inversely proportional to the number of neighbors. However, in common with the joint model we reparameterize from \((\sigma^2, \tau^2)\) to \((\lambda, \rho)\), which requires a proper joint distribution for \(U\). Therefore, we replace \(U\) with the \(n-1\) random variables \(T = (T_1, \ldots, T_{n-1})\), where \(T_k = U_k - U_n\) as suggested by Besag and Kooperberg (1995). This arbitrarily sets \(T_n\) equal to 0, although it is not clear which of the \(n\) areas this constraint should be applied to. Here, we choose the area with the least neighbors, although we present a sensitivity analysis to this choice in Section 4.4. Further details of these models are given in Section B of the supplementary material available at Biostatistics online, http://www.biostatistics.oxfordjournals.org.

3.2 Accounting for ecological bias

The data in our study are at the ecological rather than individual level, and the effects of ecological bias on our results should be investigated; for a full discussion, see Wakefield and Salway (2001). We focus on pure specification bias here, although we note that there are other types of ecological bias such as within-group and between-group confounding. Pure specification bias arises because a nonlinear individual-level risk model changes its form under aggregation. Therefore, the mean model in (3.1) needs to be adjusted. We adopt the parametric approach of Richardson and others (1987) and assume that the within-area exposure distribution is given by \(X_k \sim N(x_k, s_k^2)\), resulting in the mean function

\[
E[y_k|x_k, s_k^2] = E_k \exp(z_k^T \beta + V_k + U_k) \exp \left( x_k \beta_1 + \frac{1}{2} s_k^2 \beta_1^2 \right) .
\] (3.2)

Note that the ecological association from (3.1) \(\beta_x\) may not equal the corresponding individual-level association \(\beta_I\) from (3.2). For this study, only the mean pollution concentration is available at the DZ level, meaning that the spatial variation in exposure, \(s_k^2\), across a DZ cannot be estimated. In contrast, DZs are nested exactly within IGs, meaning that for a given IG its spatial variation in exposure, \(s_k^2\), can be estimated by the variance of the DZ concentrations that lie within it. Therefore, we assess the potential impact of ecological bias at the IG scale only, by using the pollution concentrations at the DZ level. Further details of the impact of ecological bias is presented in Section B.3 of the supplementary material available at Biostatistics online, http://www.biostatistics.oxfordjournals.org.

3.3 Inference

Inference for all models is implemented within a Bayesian framework via Markov chain Monte Carlo (MCMC) simulation, using a mixture of Gibbs sampling steps and block Metropolis–Hastings moves. The results from our study are based on 40 000 posterior samples generated from 2 Markov chains, both of which are initialized at dispersed locations in the sample space. The chains are burnt-in until convergence, as measured using the diagnostic methods of Gelman and others (2003), after which 20 000 samples are generated from each Markov chain. The MCMC simulation is implemented in the statistical package R, using functions written by the authors. Full details of the implementation as well as the data and code to run the models are presented in Sections C and D of the supplementary material available at Biostatistics online, http://www.biostatistics.oxfordjournals.org.

4. Multicity Air Pollution and Health Study

This section presents the results of the study introduced in Section 2, which is based on data from the 4 health boards Grampian, Tayside, Lothian, and Greater Glasgow. In each board, the association between
respiratory hospital admissions in 2005 and long-term exposure to PM$_{10}$ and NO$_2$ between 2002 and 2004 is estimated, at both IG and DZ resolutions. The robustness of our results is assessed by applying the general model (3.1) with 3 different specifications for the spatial random effects $U$. The first removes these effects and assumes independence, the second specifies a joint model based on intercentroidal distances, while the third is a conditional structure based on areas sharing a common border. Finally, the impact of pure specification bias is investigated by comparing the above results with those obtained by adopting a model with the extended mean function (3.2).

4.1 Model building

The model-building process first selected an appropriate set of covariates, then modeled any residual spatial dependence, and finally estimated the effects of air pollution. A Bayesian approach was adopted, based on the deviance information criterion (Spiegelhalter and others, 2002), posterior predictive residual distributions (Gelman and others, 2003), and personal judgment. The expected numbers of admissions $E_k$ were calculated first, based on external rates from the United Kingdom (World Health Organization, 2006). Then, hospital admissions data were examined for the presence of large-scale spatial trend, and a linear trend surface model was fitted in all 4 cities. At this point, the significance of a categorical urban–rural variable with 6 levels was assessed, where the first level represents areas within an urban city, while the sixth relates to remote settlements (further details are available from http://www.sns.gov.uk/).

The next step was to account for differences in socioeconomic deprivation between the areas, which has been an important confounder in similar studies (see, e.g. Elliott and others, 2007). In this study, a large number of variables in the domains of access to services, crime, education, health, housing, income, and employment were available, and the important variables in each domain were identified first, before combining variables from all the domains into a single model. From this model-building process, the percentage of adults who are income deprived, the percentage of children in ethnic minorities, the percentages of dwellings that are flats, terraced, and semidetached, educational attainment at school, and incidences of various categories of crimes (e.g., assault, burglary) were important covariates for all 4 cities. Modeled estimates of the proportion of the population who smoke were also available but are highly correlated with the percentage of the population who are income deprived (correlation of 0.82) and are therefore not required. The final covariates used in this study are listed in the R function “Overall.R” within the supplementary material available at Biostatistics online, http://www.biostatistics.oxfordjournals.org. The residuals from these models displayed low levels of spatial correlation and overdispersion, so spatial and nonspatial random effects were added to the model together with air pollution concentrations.

The neighborhood matrix for the ICAR spatial model used in this paper is defined by areas sharing a common border. Although this has been criticized by some authors when the spatial units are irregularly shaped (see, e.g. Wakefield, 2007), it does have a sensible spatial interpretation of proximity. In this paper, we adopt a distance-based measure of spatial similarity for the joint spatial model. The choice of a “border” specification is not ideal but enables the results to be compared across different spatial correlation structures.

4.2 Main results

The relationships between respiratory hospital admissions and exposure to both PM$_{10}$ and NO$_2$ are presented in Table 3, together with associated credible intervals. The results are presented on the relative risk scale for an increase in pollution concentrations of one spatial standard deviation, which is 1.7 $\mu g$ m$^{-3}$ for PM$_{10}$ and 8 $\mu g$ m$^{-3}$ for NO$_2$, to correspond to realistic increases in long-term exposure. Table 3 shows
Table 3. Posterior medians and 95% credible intervals for the effects of PM$_{10}$ and NO$_2$ on respiratory hospital admissions. The results are shown on the relative risk scale for an increase of one standard deviation in pollution concentrations (1.7 µg m$^{-3}$ for PM$_{10}$ and 8 µg m$^{-3}$ for NO$_2$)

<table>
<thead>
<tr>
<th>Spatial model</th>
<th>PM$_{10}$</th>
<th>Spatial resolution</th>
<th>NO$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DZ</td>
<td>IG</td>
<td>DZ</td>
</tr>
<tr>
<td>Grampian</td>
<td>1.03 (0.95, 1.10)</td>
<td>1.05 (0.96, 1.13)</td>
<td>1.01 (0.89, 1.14)</td>
</tr>
<tr>
<td>Joint</td>
<td>1.00 (0.92, 1.09)</td>
<td>1.05 (0.97, 1.14)</td>
<td>1.04 (0.94, 1.15)</td>
</tr>
<tr>
<td>Conditional</td>
<td>1.03 (0.95, 1.12)</td>
<td>1.04 (0.97, 1.13)</td>
<td>1.04 (0.93, 1.17)</td>
</tr>
<tr>
<td>Tayside</td>
<td>1.03 (0.92, 1.16)</td>
<td>1.06 (0.94, 1.20)</td>
<td>0.96 (0.84, 1.09)</td>
</tr>
<tr>
<td>Independence</td>
<td>1.04 (0.91, 1.21)</td>
<td>1.04 (0.90, 1.21)</td>
<td>0.97 (0.81, 1.16)</td>
</tr>
<tr>
<td>Joint</td>
<td>1.04 (0.93, 1.18)</td>
<td>1.05 (0.92, 1.21)</td>
<td>1.01 (0.88, 1.14)</td>
</tr>
<tr>
<td>Conditional</td>
<td>1.06 (1.01, 1.11)</td>
<td>1.07 (1.01, 1.14)</td>
<td>1.04 (0.97, 1.12)</td>
</tr>
<tr>
<td>Lothian</td>
<td>1.09 (1.01, 1.16)</td>
<td>1.09 (1.02, 1.16)</td>
<td>1.12 (1.02, 1.22)</td>
</tr>
<tr>
<td>Independence</td>
<td>1.08 (1.02, 1.14)</td>
<td>1.07 (1.01, 1.15)</td>
<td>1.10 (1.01, 1.21)</td>
</tr>
<tr>
<td>Joint</td>
<td>1.10 (1.06, 1.14)</td>
<td>1.08 (1.03, 1.14)</td>
<td>1.11 (1.07, 1.15)</td>
</tr>
<tr>
<td>Conditional</td>
<td>1.09 (1.04, 1.15)</td>
<td>1.07 (1.02, 1.13)</td>
<td>1.10 (1.04, 1.15)</td>
</tr>
</tbody>
</table>

convincing evidence for an effect of PM$_{10}$ in Lothian and Greater Glasgow, with relative risks ranging from 1.06 to 1.10. In contrast, there is no convincing evidence of such an effect in Tayside and Grampian, with risks between 1.00 and 1.06 and the credible intervals containing the null risk of 1. For NO$_2$ exposure, the results are similar, with risks ranging from 1.04 to 1.12 for Greater Glasgow and Lothian, whereas for Tayside and Grampian they are between 0.96 and 1.04. These results show that pollution exposure has a substantial public health impact and that reducing ambient concentrations would likely reduce the number of admissions per year. For example, our results suggest that a decrease in PM$_{10}$ concentrations of 1.7 µg m$^{-3}$ in Glasgow would have led to a reduction in respiratory admissions of about 7% in 2005, equating to approximately 1300 fewer admissions.

Table 3 also shows that the majority of the estimated risks were robust to the choice of spatial correlation model. For example, the independence, joint, and conditional models give risks ranging from 1.06 to 1.09 for PM$_{10}$ exposure in Lothian. There are a small number of exceptions to this robustness, one of which is the risk associated with NO$_2$ exposure in Lothian at the DZ resolution. In this case, the independence model estimate is 1.04 compared with 1.12 and 1.10 for the joint and conditional models, respectively. In general, the results are also similar across the 2 spatial resolutions, with the majority of risks differing by less than 3%. The exception is for PM$_{10}$ exposure in Tayside and Grampian, which generally shows attenuation when changing from IG to DZ resolutions.

Table 4 summarizes the spatial and nonspatial random effects, presenting the total variation ($\lambda$) as well as the proportion that is spatial ($\rho$). The table relates to PM$_{10}$ exposure models, although the models with NO$_2$ give similar results. The total variation $\lambda$ is typically well estimated by the models, with the joint and conditional spatial models giving consistent estimates with small 95% credible intervals. In contrast, the estimates for the independence model are generally smaller, which reflects the fact that spatial variation is being ignored. The joint and conditional spatial random-effects models do not typically produce similar estimates for the proportion parameter $\rho$, which may be because different spatial dependence structures are being adopted in each case. In addition, the credible intervals for this parameter are generally wide, suggesting that there is relatively little information in the data about this quantity.
Table 4. Posterior medians and 95% credible intervals for the transformed random-effects parameters. The results refer to PM$_{10}$ exposures.

<table>
<thead>
<tr>
<th>Spatial model</th>
<th>Spatial resolution</th>
<th>( \lambda )</th>
<th>( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DZ</td>
<td>IG</td>
<td>DZ</td>
</tr>
<tr>
<td>Grampian</td>
<td>Independence</td>
<td>0.058 (0.042, 0.078)</td>
<td>0.007 (0.001, 0.017)</td>
</tr>
<tr>
<td></td>
<td>Joint</td>
<td>0.062 (0.045, 0.082)</td>
<td>0.009 (0.003, 0.023)</td>
</tr>
<tr>
<td></td>
<td>Conditional</td>
<td>0.060 (0.042, 0.081)</td>
<td>0.013 (0.003, 0.028)</td>
</tr>
<tr>
<td>Tayside</td>
<td>Independence</td>
<td>0.059 (0.040, 0.087)</td>
<td>0.004 (0.000, 0.018)</td>
</tr>
<tr>
<td></td>
<td>Joint</td>
<td>0.069 (0.049, 0.093)</td>
<td>0.009 (0.002, 0.029)</td>
</tr>
<tr>
<td></td>
<td>Conditional</td>
<td>0.073 (0.050, 1.102)</td>
<td>0.015 (0.005, 0.040)</td>
</tr>
<tr>
<td>Lothian</td>
<td>Independence</td>
<td>0.074 (0.060, 0.091)</td>
<td>0.019 (0.012, 0.029)</td>
</tr>
<tr>
<td></td>
<td>Joint</td>
<td>0.079 (0.062, 0.103)</td>
<td>0.023 (0.014, 0.039)</td>
</tr>
<tr>
<td></td>
<td>Conditional</td>
<td>0.074 (0.060, 0.089)</td>
<td>0.032 (0.018, 0.058)</td>
</tr>
<tr>
<td>Greater</td>
<td>Independence</td>
<td>0.077 (0.066, 0.089)</td>
<td>0.019 (0.014, 0.025)</td>
</tr>
<tr>
<td>Glasgow</td>
<td>Joint</td>
<td>0.080 (0.068, 0.093)</td>
<td>0.023 (0.016, 0.036)</td>
</tr>
<tr>
<td></td>
<td>Conditional</td>
<td>0.075 (0.064, 0.087)</td>
<td>0.022 (0.016, 0.031)</td>
</tr>
</tbody>
</table>

4.3 The effects of pure specification bias

To investigate the effects of pure specification bias on our results, we reapply the conditional spatial model to each data set with the modified mean function (3.2), which incorporates within-area variation in pollution exposure. We adopt a Gaussian within-area exposure distribution for simplicity but note that if a skewed distribution, such as the gamma, is required, then further details can be found in Wakefield and Salway (2001). We only estimate the effects of ecological bias at the IG scale because only one estimate of pollution concentrations is available for each DZ, meaning that the within-area variances cannot be estimated. For an IG, the within-area variances are estimated from the DZ concentrations. The relative risks from using the extended mean function (3.2) are presented in Table 5, together with the corresponding results from Table 3 (i.e. relative risks from the conditional model at the IG resolution). The risks are almost identical in most cases, which suggests that the effects of pure specification bias appear to be minimal for this study.

4.4 Sensitivity analysis

This section presents a sensitivity analysis that assesses the robustness of our results to the choice of prior distribution for the random-effect parameters \((\lambda, \phi)\), as well as the neighborhood structure for the conditional spatial model.

Prior for \(\lambda\). The default variance prior for \(\lambda\) adopted in this paper is uniform on the standard deviation scale as suggested by Gelman (2006), rather than the more commonly used inverse gamma \((\epsilon, \epsilon)\) distribution for \(\epsilon = 0.001, 0.01, 0.1\). This is because the latter has minimal prior weight very close to 0, meaning that if the true variance is small, then its posterior distribution will be inflated by the prior. We assess the robustness of our results by comparing the flat prior on the standard deviation scale with an inverse gamma \((\epsilon, \epsilon)\) distribution, focusing on Dundee at the IG resolution. We chose this region as it has the smallest estimated variance from the main analysis (see Table 4) and is therefore likely to be the most sensitive to the prior specification. The estimated relative risk for PM$_{10}$ is invariant to the choice of prior...
Table 5. Posterior median relative risks and 95% credible intervals allowing for ecological bias at the IG resolution. The fixed exposures are taken from Table 3, while the Gaussian exposures result from adopting the mean function (3.2)

<table>
<thead>
<tr>
<th>Exposure</th>
<th>PM$_{10}$</th>
<th>NO$_{2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grampian</td>
<td>Fixed 1.04 (0.97, 1.13) 1.01 (0.91, 1.14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gaussian 1.04 (0.96, 1.13) 1.03 (0.92, 1.15)</td>
<td></td>
</tr>
<tr>
<td>Tayside</td>
<td>Fixed 1.05 (0.92, 1.21) 0.98 (0.85, 1.14)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gaussian 1.05 (0.92, 1.22) 0.99 (0.87, 1.16)</td>
<td></td>
</tr>
<tr>
<td>Lothian</td>
<td>Fixed 1.07 (1.01, 1.15) 1.08 (0.97, 1.20)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gaussian 1.07 (1.01, 1.15) 1.08 (0.98, 1.19)</td>
<td></td>
</tr>
<tr>
<td>Greater</td>
<td>Fixed 1.07 (1.01, 1.13) 1.09 (1.03, 1.15)</td>
<td></td>
</tr>
<tr>
<td>Glasgow</td>
<td>Gaussian 1.08 (1.02, 1.13) 1.09 (1.03, 1.15)</td>
<td></td>
</tr>
</tbody>
</table>

for $\lambda$, with the median value remaining unchanged. In addition, the posterior median of $\lambda$ is also fairly robust, showing no change for $\epsilon = 0.001$ but a slight inflation for $\epsilon = 0.01$ and 0.1 which is in keeping with intuition.

Prior for $\phi$. In this paper, we specify a discrete uniform prior for $\phi$ where the possible values $\psi_1, \ldots, \psi_6$ are initially chosen so that 5%, 10%, \ldots, 30% of the correlations between pairs of areas are above 0.5. This approach produces an informative posterior distribution when the number of areas is small; for example, the posterior distribution for Tayside at the IG level ($n = 90$ areas) is (35%, 34%, 17%, 7%, 4%, 3%). To assess the sensitivity of the results to this prior, we compare the above specification with (i) 5%, 6%, \ldots, 29%, 30% and (ii) 2%, 7%, \ldots, 92%, 97% pairs of areas having correlations above 0.5. In both cases, the relative risks, total variation, and spatial proportion remain unchanged, and the shape of the posterior for $\phi$ is consistent with decreasing posterior mass as the percentage of correlations greater than 0.5 increases. However, as the number of areal units increases, the amount of information in the data about $\phi$ also rises, which can result in the Markov chain remaining in one state as the posterior is very highly peaked. For example, at the DZ level in Edinburgh ($n = 992$ areas) the posterior for $\phi$ is (100%, 0%, 0%, 0%, 0%, 0%), when the default specification of $\psi_1, \ldots, \psi_6$ is adopted. Therefore, in such cases it may be preferable to adopt a finer grid of values, for example, allowing the percentage of pairs of areas with correlations greater than 0.5 to be 5%, 6%, \ldots, 29%, 30%. Adopting this prior results in a posterior with decreasing probability as the percentage of pairs of areas with correlations greater than 0.5 increases from 5% to 30%.

Neighborhood structure. The conditional spatial model is specified as a joint distribution for the $n - 1$ random effects $T$, where $T_k = U_k - U_n$ and $T_n$ is arbitrarily set to 0. Initially, we applied this constraint to the area with the least neighbors, and we compared this choice with the area that has the most and median number of neighbors as reassurance that this choice of area is arbitrary. We assessed this for both Dundee and Glasgow, and the relative risk, total variation, and proportion parameter $\rho$ are insensitive in both cases.

5. Discussion

We have investigated the effects of long-term air pollution exposure on public health in Scotland, focusing on the health boards containing the 4 major urban centers, Aberdeen, Dundee, Edinburgh, and Glasgow.
In particular, we have estimated the associations between respiratory hospital admissions in 2005 and exposure to average PM$_{10}$ and NO$_2$ concentrations over the preceding 3 years, using a small-area ecological design. We adopted a general Bayesian hierarchical modeling framework, which should be more widely applicable.

Our study presents convincing evidence for an association between long-term exposures to PM$_{10}$ and NO$_2$ and respiratory hospital admissions in Lothian and Greater Glasgow, with relative risks of 1.06–1.10 for a 1.7 µg m$^{-3}$ increase in PM$_{10}$ concentrations compared with risks of 1.04–1.12 for an 8 µg m$^{-3}$ increase in NO$_2$. In contrast, the risks for Grampian and Tayside are generally lower and do not provide strong evidence of an effect in these health boards. These differences between the 4 health boards are consistent across both pollutants and all modeling approaches. A possible explanation is that the pollution concentrations observed in Grampian and Tayside are lower than those in Lothian and Greater Glasgow (see Table 2) and may be too low to have a health impact. The differences could also be due to confounding factors that have not been accounted for in the covariate model.

A further issue relates to the modeled pollution estimates and the fact that no associated measures of uncertainty are available. One approach to incorporating this uncertainty would be to estimate the pollution exposures and relative risks with admissions simultaneously within a combined Bayesian framework, although this is likely to be computationally prohibitive.

The estimated relative risks between air pollution and health appear to be robust to the statistical modeling approach, showing little change when the spatial random effects or the priors are altered. This is reassuring in scientific terms but does raise the question of whether the spatial correlation models are required at all. The major differences between the spatial and independence models are that the latter underestimates the random-effects variance, while the former is computationally more demanding. Thus, if the single goal of the study is estimating the relative risks and associated credible intervals, then the independence model seems to be adequate, whereas if the residual structure in the data is also of interest, then additionally incorporating a spatial model would be more appropriate.

The majority of estimated relative risks are also similar at DZ and IG resolutions. This spatial invariance is initially satisfying and suggests added confidence in our findings. However, as the median population sizes differ by a factor of 5 (3956 for IG and 768 for DZ) between the 2 resolutions, the true risks at each spatial scale are different quantities, and there is no reason why they should be the same. Indeed, taking this shrinking of the spatial areas to its limit results in an individual-level study, which under certain conditions (see the supplementary material at Biostatistics online, http://www.biostatistics.oxfordjournals.org) may have a different true risk from the ecological analysis due to the possible presence of ecological bias. This may explain why the risks for both PM$_{10}$ and NO$_2$ exposure at Grampian and Tayside are different at DZ and IG resolutions, which is generally not the case for Lothian and Greater Glasgow. Another possible explanation is that Grampian and Tayside contain the least amount of data (fewest areal units), and combined with small numbers of observed events in each DZ, the instability of small numbers may be amplified, possibly leading to less stable estimates. To overcome this problem, the admissions over a longer time period could be modeled, either within a spatial model or using a spatiotemporal extension.

The effects of pure specification bias appear to be negligible in this study, although we did not consider other types of ecological bias such as within-group and between-group confounding as well as errors in variables. The lack of an effect of pure specification bias is not surprising for our study because both the pollution parameter estimates $\beta_x$ and the within-area concentration variances are small.

While the health outcome in this study was respiratory admission, respiratory disease is not the only health consequence of long-term exposure to air pollution. Clear links have been established between air pollution exposure and cardiovascular morbidity and mortality. In addition, because poor respiratory health can contribute to other serious diseases, the true health burden to which air pollution may contribute is likely to be far larger than that estimated here. The concentrations observed in this study are generally
below the recommended annual averages for Scotland, which for PM$_{10}$ is especially stringent at 18 µg m$^{-3}$ compared with 40 µg m$^{-3}$ for the rest of the United Kingdom. The results of this study thus add urgency to policies designed further to reduce pollutant emissions. Air quality has, in the past, been a key public health priority. As we learn more about its health impacts at even relatively low concentrations, it must become so again.

**SOFTWARE**

All analyses in this paper have been carried out in R. Full details of the model fitting along with R code and data are provided in the supplementary material available at *Biostatistics* online, http://www.biostatistics.oxfordjournals.org.

**ACKNOWLEDGEMENTS**

The authors gratefully acknowledge the helpful and valuable comments and suggestions made by the editor and 2 referees, all of which have contributed significantly to the improvement of our paper. The data and shapefiles used in this study were provided by the Scottish Government (2007). *Conflict of Interest:* None declared.

**SUPPLEMENTARY MATERIAL**


**REFERENCES**


[Received June 20, 2008; first revision November 26, 2008; second revision January 7, 2009; third revision January 23, 2009; accepted for publication March 3, 2009]