A Bayesian model of time activity data to investigate health effect of air pollution in epidemiological studies

Blangiardo, Marta
MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College,
Norfolk Place,
London W2 1PG, UK
E-mail: m.blangiardo@imperial.ac.uk

Richardson, Sylvia
MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College,
Norfolk Place,
London W2 1PG, UK
E-mail: sylvia.richardson@imperial.ac.uk

Introduction

In the epidemiological literature, the short term effect of air pollution on the health of individuals has been extensively studied using time series data, with particular regard to fine particles (Samet et al. 2000a, Samoli et al. 2008). A typical approach considers the ambient measure of particulates (e.g. PM_{10}, $PM_{2.5}$) at a set of locations as an approximation of the exposure of each individual in the population under study, typically a city or a region. Considerable progress has been made in interpolating pollution fields to get estimates at a higher spatial resolution (Gryparis et al., 2007) so that an ambient pollution estimate is often available close to where individuals live. Nevertheless in applying this estimate to quantify the exposure of individuals, the underlying assumption is that people remain in the same place throughout the day and are only exposed to the ambient concentration (equivalent to being outside all day). In reality, people move around indoor and outdoor environments, characterized by different concentrations of pollutants and engage in behaviors (e.g. smoking, cooking) that may also affect exposure.

Direct measures of personal exposure in individuals in a study can be obtained using a small monitor attached to each participant (Williams, 2008) and complex statistical models have been proposed to utilize such data.

Dominici et al., (2000) proposed a Bayesian meta-analytical framework using available data on personal exposure from several studies, and infers the personal exposure distribution for an area where such data are not available: (i) a general additive model is specified for the relationship between mortality and the latent (unobserved) personal exposure; (ii) a linear regression model is specified for the unobserved variable as a function of the ambient concentration, where the coefficients are given a distribution obtained from a meta-analysis of available studies on personal exposure.

Recently, time activity diaries have become a popular tool to obtain information about individual exposure: Jantunen et al., (2002) combined information from diaries and personal monitors to estimate the individual exposure in the EXPOLIS study, but several approaches have been proposed where only diaries are used to estimate the exposure at the individual level, when personal monitors are not available or impractical to be used. In this framework, the most popular approach is based on simulators: the environment is split in microenvironments (e.g. home, work, outdoor, etc.) that an individual could visit during a typical day and that are characterized by different levels of air pollution; the diaries are then used to simulate the activity patterns of a random individual in the population

combined with pollutant concentrations of the microenvironments they visit, to obtain an estimate of the personal exposure.

In this framework Burke et al., (2001) presented the Stochastic Human Exposure and Dose Simulator (SHEDS-PM), a statistical tool for estimating the population distribution of PM implemented by combining the activity pattern of individuals, obtained from diaries, and repeated sampling from the distribution of some input parameters. The SHEDS method specifies several microenvironments and different functions of the ambient PM are used to obtain their concentration. Distributions are assumed on all the function parameters and from these a distribution of exposure is estimated for each individual using a Monte Carlo simulation approach.

More recently Zidek et al., (2007) and Shaddick et al., (2008) presented a two-step Bayesian model that (i) estimates the individual exposure by combining diaries of activities and time spent in each microenvironment and (ii) links the probability distribution of the individual exposure to the values of the health outcome, which is available at an aggregated level, for strata of interest (e.g. defined with respect to age groups, sex), in an ecological regression analysis. Holloman et al., (2004) proposed a hierarchical exposure simulator, based on SHEDS, for estimating personal exposure, but adding the link with cardiovascular mortality in North Carolina. Reich et al., (2008) extended this framework, using a Bayesian model to link ultrafine PM_{10} distributions, obtained using SHEDS simulative approach, to daily mortality. Activity patterns are incorporated into SHEDS framework and approximated output distributions for exposure are then used in further models.

We follow the simulator approach but differently from the previous methods presented we model directly time activity patterns in differing microenvironments at the strata level to estimate *a distribution of exposure* for the population in each stratum. Our model is framed in a personal exposure perspective as the inference target is the exposure of groups of individuals, but it avoids intensive individual-based simulative processes and allows a simple and more transparent characterization of variability of personal exposure in different subgroups. Working at the group level it is easy to evaluate and compare different scenarios, for instance modifying the exposure for some groups as a result of spending more or less time in some microenvironments, and to assess how these changes affect the risk of disease (or death) for those groups, which would be cumbersome in an individual based simulator.

We use our model to evaluate through a realistic simulation study the impact on the relative risk of mortality occurring when the ambient concentration is used instead of group exposure estimated from time activity distribution.

Materials and Methods

Our model is illustrated using diaries from the Consolidated Human Activities Database (CHAD) and PM_{10} data from the National Morbidity, Mortality, and Air Pollution Study (NMMAPS).

In overview, it is based on the following four steps

1. Strata are defined, choosing characteristics that affect activity (and therefore exposure) patterns, e.g. sex, age, day of the week. Then, within a Bayesian hierarchical statistical model where group exposures are treated as unknown quantities:

2. Time-activity diary data are used to inform the distribution of the probability of spending time in each microenvironment for each stratum.

3. Sample values of the group exposure are obtained combining the concentration estimates for the microenvironments and the time spent in each of them.

4. Finally group exposure estimates are linked to the health outcome under study.

Steps 2 and 3 are iterated according to the hierarchical formulation producing a posterior distribution for the group exposures of each stratum, together with that of the regression coefficients that quantify the effect of group exposure on health.

The use of a Bayesian approach is particularly suitable for exposure studies, as it allows to combine several sources of data, namely the microenvironment concentrations and the time activity pattern, in a unified framework, taking into account the uncertainty in each of the sources and propagating it through the model. In addition, it enables the inclusion of any available information on one or more parameters through the specification of informed prior distributions.

Statistical formulation of personal exposure model

In ecological studies, the information about health and exposure data is available only at the group level, ($k = 1,\ldots, K$ throughout the paper), identified by a combination of covariates relevant for the study, like age, sex, time, location. Thus the outcome of interest is defined as the number of events occurring for the k^{th} group and the method we propose in this paper estimates the posterior distribution of exposure to air pollution for each group (group exposure). Throughout the paper we will use equivalently "stratum" or "group" when we refer to k.

To obtain an estimate for the group exposure X_k we take advantage of (i) information about the activity pattern of people through the day in each microenvironment, available from diaries collected in previous studies; (ii) known functions from the literature that link the microenvironment concentrations to the ambient value for the pollutant (Z).

Following Burke et al., (2001) we assume that the environment is divided in several microenvironments indexed by m, which differ either in the sources of pollution or in their concentration. Throughout the paper we refer to PM_{10} as the generic pollutant Z, but the same methodology can be applied to different types of particulates and different pollutants. The diaries for people who belong to each stratum are modeled as determinations of a multinomial distribution on the time spent in the microenvironments:

$$T_{ki} \sim Multinomial(\mathbf{p}_k, n) \tag{1}$$

where n represents the total amount of time in each unit (e.g. 1440 minutes in a day), $\mathbf{p}_k = (p_{1k}, \ldots, p_{Mk})$ is a vector representing the unknown probability of spending time during the day in microenvironment m for each stratum k and $i = 1, \ldots, I_k$ indexes the subjects in the k^{th} stratum. Similarly to McBride et al. (2007), a Dirichlet distribution is specified on the vector $\mathbf{p}_k$ with minimally informative prior:

$$\mathbf{p}_k \sim Dir(\alpha_1, \ldots, \alpha_m, \ldots, \alpha_M) \qquad \alpha_m = 1 \tag{2}$$

allowing the posterior distribution of $\mathbf{p}_k$ to be easily estimated.

In a typical time activity database, information from the diaries to be included in (1) is available for a restricted number of days, but the time series studies on short term effect of exposure to pollutants usually include daily measures of ambient concentration over a period of several years. As a first approximation, we assume that the probability of spending time in each microenvironment ($\mathbf{p}_k$) remains constant for the entire period under study and is informed by the diaries in (1); if detailed time activity data become available for different years, this assumption can be easily relaxed. For predictive purpose we allow the time spent in each microenvironment to vary, as for each day j we predict it from a Multinomial distribution parametrized by :

$$T^{pred}_{k1}(j), \ldots, T^{pred}_{kM}(j) \sim Multinomial(\mathbf{p}_k, n)$$

We assume that the concentration of each microenvironment for each time unit $j = 1, \ldots, J$ (e.g. day) is obtained through a linear function of the ambient value $Z(j)$

$$C_{km}(j) = a_{km} + b_{km} Z(j) \tag{3}$$

where a and b have a different specification for different types of microenvironment and they are allowed to depend on the stratum k. We consider six microenvironments and following Zidek et al., (2007) we distinguish between closed (house) and open (vehicle, office, other indoor like schools, stores, bar/restaurants and outdoor) ones.

An open microenvironment is assumed not to produce sources of PM_{10} (Zidek et al., 2007). In this case the intercept and the slope in (3) do not depend on the stratum, so we drop the index k. a_m and b_m are not observed and uncertainty on their values enters (3) by means of an informative normal prior distribution, obtained from the literature (Murray et al., 1997).

A closed microenvironment is characterized by its own sources of pollutant such as smoking and cooking emissions; in this case the concentration is obtained through a complex mass balance equation, where a_{km} and b_{km} are functions of several unknown quantities (volume of the house, air exchange rate, penetration factor, deposition rate).

Both in open and closed microenvironments, the concentration $C_{km}(j)$ for stratum k and microenvironment m, $(m = 1, \ldots, M)$, is functionally related to $Z(j)$ via unknown parameters that are given prior distributions and consequently $C_{km}(j)$ is a random variable.

Finally, for group k and day j, the time activity adjusted exposure is obtained as an average of the microenvironment concentrations weighted by the predicted time spent in each:

$$X_k(j) = \frac{\sum_m C_{km}(j) T_{km}^{pred}(j)}{\sum_m T_{km}^{pred}(j)} \tag{4}$$

Note that $X_k(j)$ is a random variable as both C_{km} and $T_{km}^{pred}(j)$ are random quantities.

It is important that $T_{ki}(j)$ has the same temporal resolution as $C_{km}(j)$ and consequently the same as the ambient concentration $Z(j)$, since they are combined to obtain $X_k(j)$. We return to this point and propose refinement of the temporal resolution in the Discussion.

A graphical representation of the model is presented in Figure 1 using a directed acyclic graph (DAG) where circles denote unobserved quantities and rectangles indicate observed quantities; solid arrows identify stochastic dependence, while dashed arrows indicate functional relationship.

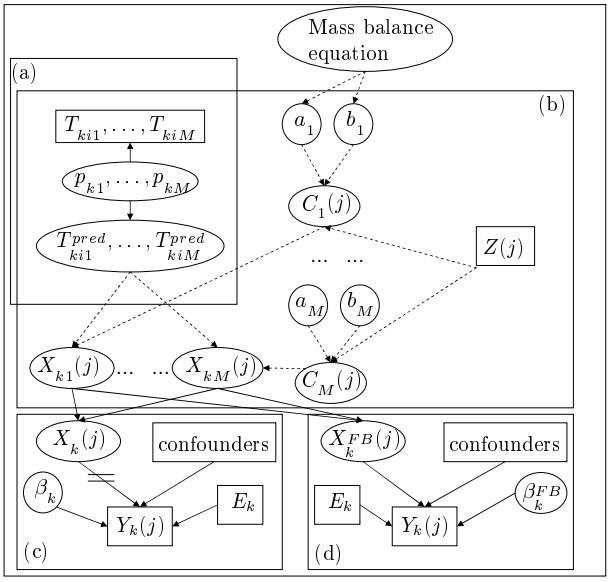


Figure 1: Directed Acyclic Graph of the model

Graphical representation of the model through a Directed Acyclic Graph (DAG). The figure shows (a) how the diaries are linked to the exposure to PM_{10} for the k^{th} stratum and a day j through the time spent in each microenvironment; (b) the model on the concentration, a function of ambient PM_{10} for day j; (c) the interrupted link between the exposure and the health outcome ($Y_k(j)$), i.e. the uncertainty on $Y_k(j)$ is not fed back to $X_k(j)$; (d) an alternative model where the link between exposure and health is not cut and the health outcome is allowed to influence the exposure $X_k^{FB}(j)$ in a fully Bayesian manner.

The effect of group exposure estimates from (1)-(4) on health outcomes can be investigated linking these two quantities in an epidemiological perspective. We have included this step in the DAG. In many cases, the exposure model and the health model are fitted separately and the uncertainty in the exposure is only "fed forward" to the health outcome - we have depicted this case in panel (c) of Figure 1 using an interrupted arrow to indicate that the two models are not estimated jointly. We present a simulation study following (c) in the "Simulation set up" section. In a fully Bayesian

perspective, the joint distribution of exposure and health parameters is obtained and feedback from health to exposure is allowed, corresponding to panel (d) that we follow in the "Sensitivity Analysis" section.

Air pollution and diary data

Data on ambient $PM_{10}(j)$ concentration were obtained from the NMMAPS study, the largest study on pollutants in US, which covers 108 cities (Samet et al., 2000) and analysis was limited to the last five years available (1996-2000). For the sake of simplicity, the US was divided into five regions (North East, South, MidWest, Mountain, Pacific) and in each region the city characterized by the smallest proportion of missing values for $PM_{10}(j)$ was chosen for illustration (5% for Pittsburgh, 9% for Houston, 0.08% for Chicago, 0.05% for Denver, 9% for Sacramento). For each city the ambient concentration was estimated as the daily mean value of $PM_{10}(j)$ ($\mu g/m^3$), obtained as the average over all the monitoring sites in that city. Missing values were readily imputed by specifying a distribution on $PM_{10}(j)$ (e.g. $PM_{10}(j) \sim LogN(\mu, \sigma)$) with hyper parameters that ensure a proper but vague prior).

The diaries of activities were derived from the CHAD database, using information available for 7100 people older than 16. Individuals were stratified into 8 groups, according to age (< 65, ≥ 65), sex and day of the week (weekday, weekend), characteristics which influence the pattern of activities and that are generally available for a typical health outcome. Six microenvironments were considered (House, Vehicle, Office, Other indoor, Bar/Restaurant and Outdoor) and the time spent in each microenvironment was calculated for each group. The average number of smoked cigarettes and the average time spent cooking were also included in the model as observed quantities obtained from the diaries; the volume of the house was assigned a prior distribution based on the type of dwelling (detached house, semi-detached/terraced house, flat, other) and for each city the average volume was considered, weighted by the proportion of different dwellings in that city.

The model for the five cities built in equations (1)-(4) and by the prior distributions has been implemented in the free software WinBUGS.

Comparison of exposure-response functions (ERF)

In an epidemiological study, when the health outcome is available only at the group level, a Poisson log-linear model is commonly used to link the exposure to it. A Poisson distribution for the number of outcomes $Y_k(j) \sim Pois(E_k \mu_k(j))$ is assumed for each stratum k, where E_k is the expected number of outcomes for the k^{th} stratum, and the logarithm of the parameter $\mu_k(j)$ is a function of the exposure variable. Then typically the ERF is assumed log-linear with respect to the group exposure variable:

$$\log \mu_k(j) = \mu_0 + X_k(j)\beta_k^X + confounders \tag{5}$$

where μ_0 is the mean risk of experiencing the outcome in the population under study and β_k^X is the log-relative risk representing the change in the risk of the outcome when the average group exposure $X_k(j)$ changes by one unit. Linear or non linear functions of potential confounders can be included in the model.

The group exposure $X_k(j)$ can be interpreted as the average exposure for all the individuals belonging to the k^{th} stratum during the day j. When $X_k(j)$ is replaced by the ambient concentration $Z(j)$, it leads to the substitution of $X_k(j)\beta_k^X$ by $Z(j)\beta^Z$ in (5). In this context we want to show how the estimates of relative risk of death differ when the effect of the group exposure $X_k(j)$ is the prime interest or when the focus of the inference is the crude effect of the ambient concentration $Z(j)$, the typical exposure measure adopted in time series studies.

Simulation set up

We use a realistic simulation study inspired by air pollution scenarios. Following Schwartz et al., (2000) we defined two possible true ERF:

- linear ($\beta_k^X X_k$)
- piecewise linear, with a slope of 0 for observations below 20 $\mu g/m^3$, which represents a hypothesized effect threshold (0 if $X_k(j) < 20\mu g/m^3$ and $\beta_k^X X_k$ if $X_k(j) \geq 20\mu g/m^3$).

For the sake of simplicity we did not run the simulation on all the 5 cities, but we considered only Pittsburgh, and we defined only four strata based on the age (< 65, ≥ 65) and day of the week (weekday/weekend) for each of these functions.

Using ambient data from the NMMAPS study for Pittsburgh and diaries of activities from the CHAD database

1. we ran the Bayesian personal exposure distribution model for one year ($j = 1, \ldots, 365$) and obtained posterior distributions for each day for the group exposure; ((a) and (b) in Figure 1)

2. we selected a reasonable spectrum of values for the true exposure-response coefficient β_k^X, the log-relative risk for group exposure to the pollutant, here $PM_{10}(j)$. The values were chosen to range between 0.005 and 0.1, where lower values are based on the WHO ERF (Ostro et al., 2004) which identifies an increment in relative risk for all cause mortality of 0.8% for an increase of ambient $PM_{10}(j)$ concentration of 10 $\mu g/m^3$ and higher values are seen in studies estimating air pollution exposure at finer as opposed to coarser spatial resolution (Willis et al., 2003);

3. we applied each ERF to the time activity adjusted group exposure distribution;

4. we simulated values for a hypothetical mortality outcome

$$Y_k^{sim}(j) \sim Pois(E_k f(X_k(j))) \tag{6}$$

where E_k is the average number of deaths for the two age groups (7.3 for < 65, 19 for ≥ 65) obtained again from NMMAPS and used as an offset in the simulation ((c) in Figure 1).

The aim of the simulation was to compare the results to those that would be obtained if the ambient concentration were used for all strata instead of the true group exposure. So we also analyzed the simulated data with a log-linear model where the outcomes $Y_k^{sim}(j)$ simulated in (6) are linked to the ambient concentration $Z(j)$:

$$Y_k^{sim}(j) \sim Pois(E_k f(Z(j))) \tag{7}$$

where $f(Z(j))$ can again be a linear ERF $f(Z(j)) = \mu_0 + Z(j)\beta_k^Z$ or a linear threshold ERF, with $f(Z(j)) = \mu_0 + Z(j)\beta_k^{Z1}$ if $X_k(j) < 20\mu g/m^3$ and $f(Z(j)) = \mu_0 + Z(j)\beta_k^{Z2}$ if $X_k(j) \geq 20\mu g/m^3$. The coefficients β_k^Z, β_k^{Z1} and β_k^{Z2} represent the misspecified log-relative risks for the stratum k for the two ERFs. Note that we assumed the same threshold ($20\mu g/m^3$) both for ambient concentration and group exposure. This would correspond to the analysis model presented in Figure 1 (c) or (d), but where $Z(j)$ replaces $X_k(j)$.

The entire simulation process was repeated 500 times using different values of the posterior distribution of the exposure obtained from the Bayesian model described in (1)-(4). Finally, we compared results for exposure-response coefficients obtained using the ambient concentration with the values set in the simulation scenarios.

Results

Time-activity patterns

Time spent in the different microenvironments was remarkably consistent among the five cities under study: for instance < 65 years old males spent the majority of time at home ($\sim 60\%$ of time during the weekday and 75% at the weekend), and in the office ($\sim 25\%$ overall in the week and 6-9% at the weekend).

As expected, there were differences in the time spent in each microenvironment by age and sex, with the office microenvironment mainly visited by those < 65 years, more time spent at home by men and women ≥ 65 during the week and more time spent by males than females outdoors, especially males < 65 years old.

Difference in relative risks: group exposure vs ambient concentrations

Average exposures using ambient concentrations were found to be higher than estimated group exposures in all the cities and this was consistent by sex and age (Figure 2). As expected, the time activity adjusted exposure, $X_k(j)$, shows heterogeneity between cities - again consistent by sex and age group. This is mostly driven by the difference in the values of the ambient PM_{10} concentration (the diamond shape in Figure 2), which enters (3) to estimate the concentration of the different microenvironments.

Figure 2: Posterior distribution of group exposure vs ambient concentration

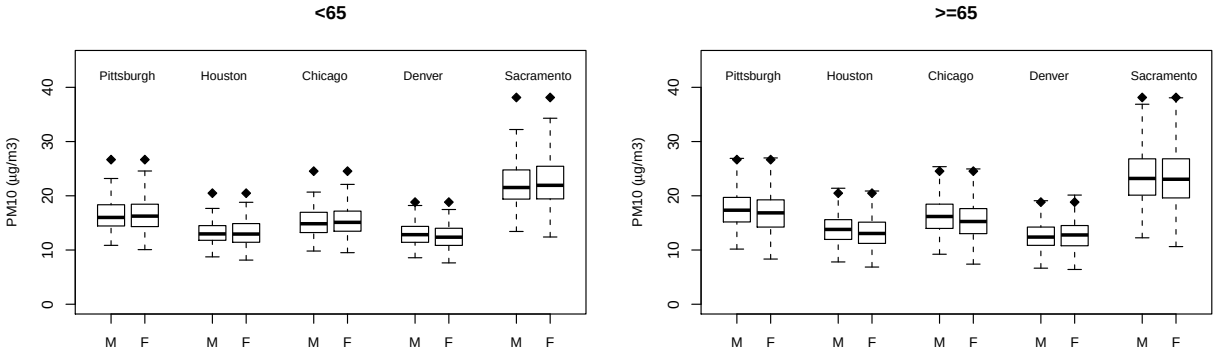
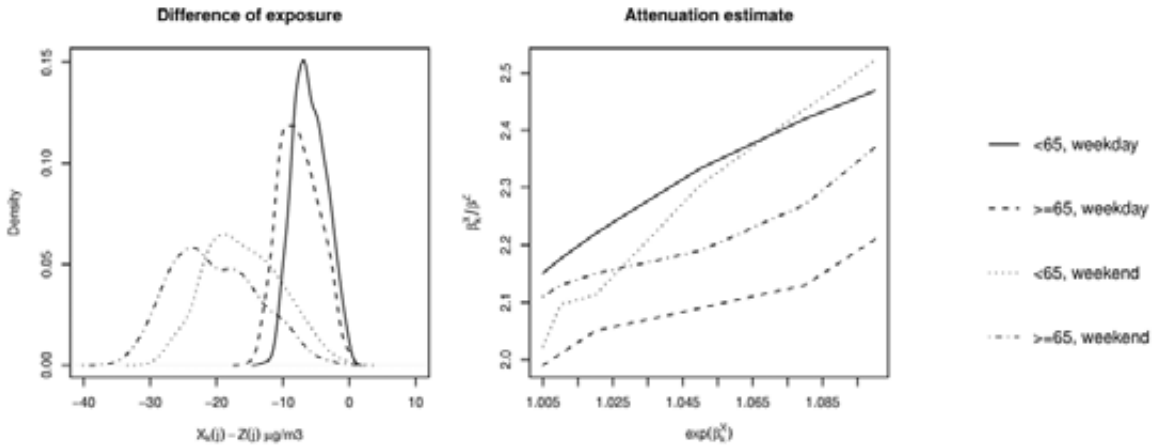


Figure 3 (left) shows the distribution of the difference between estimated group exposure and ambient concentration $Z(j)$ for the simulation study. Similarly to the real data presented in Figure 2, the difference is negative, i.e. the exposure is overestimated when the ambient concentration is used in place of the group exposure, where the assumption is effectively that people spend all their time outdoor.

A similar behavior is observed for the relative risks (RRs), with values obtained using the ambient concentration $Z(j)$ always smaller, ranging for instance between 1.002 and 1.003 for the different strata when the relative risk for the group exposure is set to 1.005. The difference increases as the RR gets larger. The ratio between log-relative risk using group exposure and log-relative risk using ambient concentration ($\frac{\beta_k^X}{\beta_k^Z}$) ranges between 1.99 for older people during weekday and 2.11 for older people during weekend. As the value of the relative risk set in the simulation increases, the ratio becomes larger, reaching 2.52 for younger people during weekend when the simulated RR=1.1. This trend in the ratios is illustrated in Figure 3 (right) for the four strata (age greater or less than 65 and weekday or weekend) using the linear ERF. The results show that even though the discrepancy of exposure does seem modest, the impact on the relative risk is noticeable, with always smaller values for the model that considers the ambient concentration as a surrogate measure of exposure for the group.

Figure 3: Difference in the exposure and relative risk for different strata assuming a linear ERF: simulation study on Pittsburgh



Sensitivity analysis: Fully Bayesian model

The simulation study followed Figure 1 (c) and consisted of two separate steps: (i) the group exposure is estimated and (ii) is linked to the health outcome in the epidemiological study. This means that uncertainty from the health outcome cannot influence the exposure values, which is called in a Bayesian perspective "cut of feedback" (Lunn et al., 2009). To investigate the effect of allowing feedback on $X_k(j)$ we performed a sensitivity analysis building a unified global model following Figure 1 (d) in a fully Bayesian manner. We performed this investigation for a simplified version of the data, considering each of the five cities separately. We found that the average group exposure $X_k(j)$ is slightly smaller when the feedback is allowed. The results are consistent for the different cities; they go the same way for all the strata.
Consequently, allowing feedback leads to slightly increased RR of mortality, ranging from -0.003 to 0.009 where the model with feedback cut shows values between -0.003 and 0.005.
One drawback of implementing the fully Bayesian version described in Figure 1(d) is that the MCMC sampling used to estimate the refined posterior distribution for the full model becomes more computationally intensive and might require custom made programs if the number of strata is large.

Discussion

In this paper we propose a Bayesian model that uses time activity data to estimate the exposure to air pollution at the group level. It allows (i) integration of uncertainty in a principled way, specifying distributions both on parameters and on latent exposure quantities and (ii) use of individual level data (i.e. activity diaries) to estimate the probability of spending time in each microenvironment for each group.

Differently from the time series approach where the inference is done at the population level, we consider the individuals as the target, but as we obtain directly a distribution for the aggregated time that each group spent in each microenvironment, it does not require a complex simulation process for each individual exposure that then needs to be averaged when performing the link with the health outcome. Thus our approach can easily deal with thousands of diaries. Moreover, as inference is directly made at the group level of interest, it is easier to investigate and to interpret the main patterns of differences across the population strata.

In our example using US data, we found that diaries from different cities did not greatly influence the group exposure and that there was stability in the activity patterns. This might not be generalizable to other countries, especially if climate differs substantially between different locations. It is therefore good practice to include location specific diaries where possible, in order to avoid biases, even if small, due to the different behavior of people. As an additional benefit of our model of time activity diaries, when several geographic regions are available and reasonably similar, it would be possible to estimate time activity patterns in a hierarchical manner and borrow information across the different regions on some parameters if needed (e.g. if numbers by area are small), strengthening the inference.

In a realistic simulation study we showed that the relative risks obtained when ambient concentration is used are always closer to 1 than when the group exposure is considered and that the ratio between the two log-relative risks ($\frac{\beta_k^X}{\beta_k^Z}$) is always larger than 1.5. This suggests that when the interest is on the individuals, an attenuation of the exposure-response coefficient occurs using the ambient concentration. This attenuation can be framed in a measurement error perspective, as $Z(j)$ acts as a surrogate for the individual exposure. The range of attenuation found for the relative risks was comparable to that obtained by Shaddick et al., (2008) applying a detailed time activity simulator to time activity diaries in London during 1997.

The sensitivity analysis on the effects of allowing feedback shows smaller exposures for the 5 cities, with slightly larger regression coefficients β_k^X, suggesting that when information flows from the outcome to the exposure the attenuation, quantified by $\frac{\beta_k^X}{\beta_k^Z}$, is generally slightly higher and could have a more substantial impact on the epidemiological results. However, it would be simplistic to suggest that the same level of attenuation is always expected as it depends on different factors (e.g. time activity patterns, indoor emissions, distribution of the parameters in (3)).

It has been argued that in epidemiological studies only functions based on the ambient concentration (concentration-response functions) should be used, as they can be regulated, whereas human behavior, reflected in the ERF are not subjected to regulation. Moreover, as the indoor sources remain more or less constant across time, the fluctuations in the ambient sources determine the RR in the epidemiological study when ERF is included. On the other hand it can be argued back that disseminating the results of studies which use ERF could sensitize individuals about how different activities lead to differences in exposure to air pollution. In addition, the size of the RR using ERF better reflects the health impact of air pollution and the indoor sources do contribute to the health outcome even though ambient fluctuation determine the RR.

In this context we aimed at contributing to this debate presenting the range of differences that could occur using ambient concentration instead of group exposure and, as differences varied by strata (age group and weekday/weekends), stress that this could lead to different epidemiological conclusions with consequent policy implications.

REFERENCES (RÉFERENCES)

1. Burke J., Zufall M., Okaynak H., 2001. A population exposure model for particulate matter: case study results for PM2.5 in Philadelphia, PA. Journal of Exposure Analysis and Environmental Epidemiology 11, pp.470-489.

2. Dominici F., Zeger S., Samet J., 2000. A measurement error model for time-series studies of air pollution and mortality. Biostatistics 1(2), pp.157-175.

3. Gryparis A., Coull B., Schwartz J., Suh H., 2007. Semiparametric latent variable regression models for spatiotemporal modeling of mobile source particles in the greater Boston area. Journal of the Royal

Statistical Society Series C 56(2), pp.183-209.

4. Holloman C., Bortnick S., Morara M., Strauss W., Calder C., 2004. A Bayesian hierarchical approach for relating PM2.5 exposure to cardiovascular mortality in North Carolina. Environmental Health Perspectives 112(13), pp.1282-1288.

5. Jantunen M., Katsouyanni K., Knppel H., Knzli N., Lebret E., Maroni M. et al, 2002. Air pollution exposure in European cities: the EXPOLIS study, tech. rep, European Union.

6. Lunn D.J., Thomas A., Best N., Spiegelhalter D., 2000. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10, pp.325-337.

7. Murray D., 1997. Residential house and zone volumes in the United States: empirical and estimated parametric distributions. Risk Analysis 17, pp.439-445.

8. Ostro B., Preuss-Eusteun A., Campbell-Lendrum D., 2004. Outdoor air pollution - assessing the environmental burden of disease at national and local level. Environmental Burden of Disease Series 5, WHO, Geneva, Switzerland.

9. Reich B., Fuentes M., Burke J., 2008. Analysis of the effects of ultrafine particulate matter while accounting for human exposure. Environmetrics 20, pp.131-146.

10. Samet J.M., Dominici F., Curriero F., Coursac I., Zeger S.L., 2000. Fine particulate air pollution and mortality in 20 US cities: 1987-1994. New England Journal of Medicine (with discussion) 343, pp.1742-1757.

11. Samoli E., Peng R., Ramsay T., Pipikou M., Touloumi G., Dominici F., et al., 2008. Acute effects of ambient particulate matter on mortality in Europe and North America: results from the APHENA study. Environmental Health Perspectives 116, pp.1480-1486.

12. Schwartz J., and Zanobetti A., 2000. Using meta-smoothing to estimate dose-response trends across multiple studies, with application to air pollution and daily death. Epidemiology 11(6), pp.666-672.

13. Shaddick G., Lee D., Zidek J., Salway R., 2008. Estimating exposure response functions using ambient pollution concentrations. Annals of Applied Statistics 2(4),1249-1270.

14. Williams R., Rea A., Vette A., Croghan C., Whitaker D., Stevens C., et al, 2008. The design and field implementation of the Detroit Exposure and Aerosol Research Study. Journal of Exposure Science and Environmental Epidemiology 19, pp.643-659.

15. Willis A., Jerrett M., Burnett R.T., Krewski D., 2003. The Association between Sulfate Air Pollution and Mortality at the County Scale: An Exploration of the Impact of Scale on a Long-Term Exposure Study. Journal of Toxicology and Environmental Health Part A 66, pp.1605-1624.

16. Zidek J., Shaddick G., Meloche J., Chatfield C., White R., 2007. A framework for predicting personal exposures to environmental hazards. Environmental and Ecological Statistics 14, pp.411-431.