1 Introduction

The quantitative combination of information from a variety of studies or sources is an important problem in many areas of scientific research. The main difficulty in integrating the results from various studies stems from the sometimes diverse nature of the studies, both in terms of design and methods employed. A common issue in combining evidence is that one may need to use studies reporting different sets of variables to answer the same question.

In this paper we focus on the combination of studies reporting categorical variables when different sets of variables are recorded. There are a number of ways in which data relevant to the association between categorical variables might be arrayed in several contingency tables with different dimensions. For example, this scenario occurs when we want to combine $I \times J \times K \times L$ tables in which the joint distribution of the four categorical variables — X_1, X_2, X_3 and X_4 say — is not reported for all studies, while the marginal distributions of either X_1, X_2, X_3 or X_1, X_2, X_4 or every other subset of the four dimensions is always available.

We introduce a multinomial-logistic normal hierarchical model to combine contingency tables with different dimensions. We suppose that at least one of the studies to be combined records all the categorical variables. We model jointly the full set of categorical variables, and we treat the variables that are not reported as missing dimensions of a common underlying contingency table.

We apply our modeling strategy to a data base of air pollution, mortality and weather from 1987 to 1994 for six U.S. cities (Pittsburgh, Chicago, New-York, Miami, Philadelphia, Minneapolis). In Pittsburgh and Chicago, days are cross-classified according to low, medium, and high levels of non-accidental deaths (D), particulate matter (PM_{10}), ozone (O_3,) and carbon monoxide (CO). For the remaining four cities, not all variables were recorded. More specifically, because in New York, Philadelphia and Miami more than 80% of days are missing PM_{10} measurements, we consider days cross-classified according to D, O_3 and CO only. In Minneapolis, there are no monitors that record O_3 levels and therefore days are cross-classified according to D, PM_{10} and CO only.

Despite the large literature on the health effects of air pollution, the joint effects of the pollutants on health is currently inadequately understood. Epidemiological studies have shown that current levels of air pollution causes excesses in mortality (American Thoracic Society, 1996a; American Thoracic Society, 1996b; Kelsall et al., 1997; Samet et al., 1997). The associations have been found using Poisson regressions and other statistical methods for time-series data (Liang and Zeger, 1986). Pooled analyses for combining evidence about particulate matter-mortality relationships across the 20 largest US cities have recently been developed by Dominici et al. (1999). Modeling approaches are controversial in regard to model specification; which pollution variables include in the analysis; which lags of pollutant values to use; and to how handle missing air pollution variables. See Vedal (1997) for a critical review on ambient particles and health.

We aim at developing a novel statistical strategy to: 1) investigate the association between air pollution and mortality in a flexible way that take into account the correlation structure among the pollutants; 2) reconstruct the tables for the cities with a missing pollution variable; 3) make prediction on the association between air pollution and mortality for a city other than the six sampled; and 4) estimate for each city, which pollutant combination is the most harmful.

The outline of the paper is as follows. In section 2, we introduce a multinomial-logistic normal hierarchical model. We show that the conditional distribution of the missing counts, given the observed counts and the study-specific parameters, is available in closed form for any missing variables pattern. We also propose a Bayesian approach to reconstruct the "incomplete" contingency tables. In section 3, we give all the distributional theory necessary to implement a Gibbs sampler to approximate the posterior distributions of the parameters of interests. We illustrate how to obtain an accurate approximations of the Aitchison family that works well in large dimensional spaces. In section 4, we apply our method to the air pollution and mortality data set described earlier, and we address sensitivity to the choice of the prior distribution.

2 Methods

2.1 Multinomial-Logistic Normal hierarchical model

Consider S studies each reporting observations arranged in a contingency table with at most I cells. Let $\boldsymbol{n}^s = \{n_i^s;\ i = 1, \ldots, I\}$, and $\boldsymbol{\pi}^s = \{\pi_i^s;\ i = 1, \ldots, I\}$ denote the vectors of the counts and of the probabilities in the study s. We focus on studies where the sample size $N^s = \sum_i^I n_i^s$ is fixed by design, and assume that the $\boldsymbol{n}^s$ can be modeled by a multinomial distribution with cell probabilities $\boldsymbol{\pi}^s$, i.e. $\boldsymbol{n}^s \mid \boldsymbol{\pi}^s \sim \mathcal{M}_I(\boldsymbol{\pi}^s, N^s)$.

In Bayesian analyses of contingency tables, it is common to model the cell probabilities using a conjugate Dirichlet distribution. Although a Dirichlet prior for the cell probabilities is the most tractable choice for the multinomial data, it implies a negative correlation among any two cell probabilities, thereby making it inappropriate to fit data for which some correlations might be positive.

An alternative and richer class of distribution is the logistic normal family (LN), (Leonard, 1975; Aitchison and Shen, 1980; Nazaret, 1987; Knuiman and Speed, 1988; Goutis, 1993). The LN distribution can model dependence more flexibly than the Dirichlet distribution. This is because the LN distribution allows varying degrees of uncertainty in the prior cell means, and can incorporate prior knowledge about the correlation structure between the log-ratios of the cell probabilities. Its density function, denoted here by $LN_{I-1}(\boldsymbol{\mu}, \Sigma)$ is proportional to:

$$p(\boldsymbol{\pi} \mid \boldsymbol{\mu}, \Sigma) \propto \left(\prod_{i=1}^{I} \pi_i \right)^{-1} \exp\left[-\frac{1}{2} (\boldsymbol{\theta} - \boldsymbol{\mu})' \, \Sigma^{-1} \, (\boldsymbol{\theta} - \boldsymbol{\mu}) \right] \quad (1)$$

where $\boldsymbol{\theta}$ is the multivariate logit defined by $\log(\boldsymbol{\pi}_{-I}/\pi_I)$, $\boldsymbol{\pi}$ is defined on the $(I-1)$ - dimensional simplex, and $\boldsymbol{\pi}_{-I} = (\pi_1, \ldots, \pi_{I-1})$. The parameters $\boldsymbol{\mu}$ and Σ can be interpreted as the mean vector and the covariance matrix for the multivariate logit $\boldsymbol{\theta}$. For $\boldsymbol{\mu}$ we specify a multivariate normal distribution, and for Σ we specify an inverse Wishart distribution which is conjugate to the likelihood $LN_{I-1}(\boldsymbol{\mu}, \Sigma)$. Several strategies for non-conjugate Bayesian estimation of the covariance matrix in hierarchical models are addressed by Daniels (1997) .

In summary, our assumptions define the following hierarchical model:

$$\begin{aligned}
\boldsymbol{n}^s \mid \boldsymbol{\pi}^s &\sim \mathcal{M}_I(\boldsymbol{\pi}^s, N^s) \quad s = 1, \dots, S \\
\boldsymbol{\pi}^s \mid \boldsymbol{\mu}, \Sigma &\sim LN_{I-1}(\boldsymbol{\mu}, \Sigma) \quad s = 1, \dots, S \\
\boldsymbol{\mu} &\sim N_{I-1}(\boldsymbol{m}, \boldsymbol{M}) \\
\Sigma &\sim IW_{I-1}(b, \boldsymbol{B})
\end{aligned}$$

where $\boldsymbol{m}, \boldsymbol{M}, b, \boldsymbol{B}$ are fixed hyperparameters.

Under this model, we estimate study-specific parameters $\boldsymbol{\pi}^s$ for a particular table, as well as population parameters $\boldsymbol{\mu}$ and Σ which characterize the underlying table.

2.2 Missing dimensions

Consider now the situation in which some studies do not report or record the full set of categorical variables. We assume here that the data collection mechanism is ignorable, i.e. that 1) the missingness mechanism does not depend on the missing values and that 2) the parameters governing the distribution of the missing data mechanism and the parameters of the statistical model are a priori independent. Under these assumptions, the posterior distribution of the model parameters and of the missing data are determined by the specification of the statistical model and the observed data (Rubin, 1976).

Within each study, we rearrange the full set of the variables recorded in at least one of the data sources in the form of a contingency table, and we treat the categorical variables that are not reported as missing dimensions of the table.

Bayesian computation in a missing-data problem is based on the joint posterior distribution of parameters and missing data, given modeling assumptions and observed data. This data-augmentation strategy permits us to create a common scale to synthesize this incomplete information without understating uncertainty. Simulation-based methods for handling uncertainty about missing data in Bayesian analysis are discussed by Tanner (1991), and Gelman et al. (1995).

2.3 Conditional distribution of the missing counts

Suppose that study s reports the vector of "aggregated counts" $\boldsymbol{t}^s = \{t_1^s, \ldots, t_j^s, \ldots, t_{I^s}^s\}$ with $I^s \leq I$, and let $\{A_1^s, \ldots, A_j^s, \ldots, A_{I^s}^s\}$ be the partition of the set of cell indexes $\{1, \ldots, i, \ldots, I\}$ induced from $\boldsymbol{t}^s$. Under this notation we can rewrite the "aggregated counts" as functions of the "original counts", i.e., $t_j^s = \sum_{i \in A_j^s} n_i^s$ for all $j = 1, \ldots, I^s$. For any j such that i falls into A_j^s, let $\xi_i^s = \pi_i^s / \sum_{i \in A_j^s} \pi_i^s$ be the normalized probability of the cell i, and let $\boldsymbol{\xi}^s(j) = \{\xi_i^s : i \in A_j^s\}$, $\boldsymbol{n}^s(j) = \{n_i^s : i \in A_j^s\}$. Using this formulation, t_j^s and $\boldsymbol{n}^s(j)$ represent the observed and the missing part of the data for the study s. Because of the multinomial assumption at the first stage, the conditional distribution of the missing counts, given the observed counts and the study-specific parameters for study s, is $\boldsymbol{n}^s(j) \mid t_j^s, \boldsymbol{\xi}^s(j) \sim \mathcal{M}_{|A_j^s|}(\boldsymbol{\xi}^s(j), t_j^s), \quad \text{independently} \quad s = 1, \ldots, S.$ (Bishop et al., 1975), where $|A|$ denotes the number of elements in the set A. By this distribution, we take into account the uncertainty arising from the missing variables. We treat the missing counts $\boldsymbol{n}^s(j)$ as the unknown parameters, and we simulate them conditionally on the parameters and on the observed data.

3 Computational Methods

In this section, we describe how we sample from the posterior distribution of the parameters and unreported observations using a combination of the Gibbs sampler (Gelfand and Smith, 1990) and Metropolis-Hastings algorithm (Hastings, 1970). We partition the unknown parameters and missing data into groups, and sample each group in turn, given all the others. The associated full conditional distributions are given below:

$$
\begin{aligned}
\boldsymbol{n}^s(j) \mid t_j^s, \boldsymbol{\xi}^s(j) &\sim \mathcal{M}_{|A_j^s|}(\boldsymbol{\xi}^s(j), t_j^s), \text{ ind. } j = 1, \ldots, I^s,\ s = 1, \ldots, S \\
\boldsymbol{\pi}^s \mid \boldsymbol{n}^s, \boldsymbol{\mu}, \Sigma &\sim \mathcal{A}_I(\boldsymbol{n}^s, \Sigma)\ s = 1, \ldots, S \\
\boldsymbol{\mu} \mid \boldsymbol{\pi}^1, \ldots, \boldsymbol{\pi}^S, \Sigma &\sim N_{I-1}\left(\left(\Sigma^{-1} \textstyle\sum_{s=1}^S \boldsymbol{\theta}_{-I}^s + \boldsymbol{M}^{-1}\boldsymbol{m}\right) V, V\right) \\
\Sigma \mid \boldsymbol{\pi}^1, \ldots, \boldsymbol{\pi}^S, \boldsymbol{\mu} &\sim IW_{I-1}\left(b + S, \left[\boldsymbol{B} + \textstyle\sum_{s=1}^S \left(\boldsymbol{\mu} - \boldsymbol{\theta}_{-I}^s\right)\left(\boldsymbol{\mu} - \boldsymbol{\theta}_{-I}^s\right)'\right]^{-1}\right).
\end{aligned}
$$

where $V = \left(S\Sigma^{-1} + \boldsymbol{M}^{-1}\right)^{-1}$. All the full conditionals density functions belong to known classes of probability distributions. The density $\mathcal{A}_I(\boldsymbol{n}, \Sigma)$, known as the Aitchison distribution is a broad class of distributions which includes the Dirichlet and logistic normal distributions as special cases. It was first introduced by Aitchison (1985a) in the analysis of compositional data, and then it was proposed as a class of prior distribution for multinomial models (Aitchison, 1985b). Unfortunately, this family doesn't have a closed-form expression for the normalizing constant. Using a result of Aitchison (Aitchison, 1986) on approximating the $\mathcal{A}$ family, it is possible to derive an efficient proposal distribution to be used in a Metropolis step. By definition, the Aitchison density $\mathcal{A}_I(\boldsymbol{n}, \Sigma)$ is equal to the Dirichlet density $\mathcal{D}(\boldsymbol{n})$ times the logistic normal density $LN(\boldsymbol{\mu}, \Sigma)$. Aitchison (1986) shows that the closest logistic normal approximation to $\mathcal{D}(\boldsymbol{n})$ in the sense of the Kullback-Liebler distance between the two densities is the one with parameters $\boldsymbol{\mu}_n$ and Σ_n, where $[\boldsymbol{\mu}_n]_i = \psi(n_i) - \psi(n_I)$, $[\Sigma_n]_{ii} = \psi'(n_i) + \psi'(n_I),\ i = 1, \ldots, I-1$ and $[\Sigma]_{ij} = \psi'(n_I),\ i \neq j = 1, \ldots, I-1$. Here ψ and ψ' are the digamma and trigamma functions. Using this approach, the Aitchison distribution can be approximated by a LN density with parameters $\Sigma^* = (\Sigma^{-1} + \Sigma_n^{-1})^{-1}$ and $\boldsymbol{\mu}^* = \Sigma^* \times (\Sigma^{-1}\boldsymbol{\mu} + \Sigma_n^{-1}\boldsymbol{\mu}_n)$.

In our experience, a Metropolis Algorithm with independent proposal equal to $LN(\boldsymbol{\mu}^*, \Sigma^*)$ provides an efficient strategy for sampling all cell probabilities in contingency tables of relatively high dimensions. The performance of this strategy in the context of the air pollution data analysis is discussed in section 4. Alternative MCMC strategies for generating samples from the Aitchison distribution are discussed by Forster and Skene (1994) .

Within each study, we can use sampled values from the MCMC to estimate the marginal densities of the missing counts, by averaging the final conditional densities from each sequence. We estimate $p(\boldsymbol{n}^s(j) \mid \text{ data})$ by $1/M \sum_{m=1}^{M} p(\boldsymbol{n}^s(j) \mid \boldsymbol{\xi}^{s^{(m)}}(j), t_j^s)$ where $\boldsymbol{\xi}^{s^{(1)}}(j), \ldots, \boldsymbol{\xi}^{s^{(M)}}(j)$ can be obtained by a variable transformation of the sequence of the sampled values $\boldsymbol{\pi}^{s^{(1)}}(j), \ldots, \boldsymbol{\pi}^{s^{(M)}}(j)$.

4 Mortality-Air pollution association in six U.S. cities

In this section, we apply our modeling strategy to a data base of air pollution, mortality and weather for six US locations from 1987 to 1994. We investigate the association between pollution and mortality by looking at the cross-classifications of days according to L=low, M=medium, and H=high levels of number of deaths for respiratory diseases among elderly D, of particulate matter PM_{10}, of ozone O_3, and of carbon monoxide CO. The scientific focus is the dependence of mortality on the three pollutants, that is $[D \mid PM_{10}, O_3, CO]$.

Several epidemiological studies (Lipfert and Wyzga, 1993; Dockery et al., 1993; Lipfert, 1994; Li and Roth, 1995) have made the point that to investigate the mortality-pollutants association, it is appropriate to adjust for confounding factors such as weather, long term trends, and other variables. To do so, in our analysis we first calculate the predicted mortality counts by fitting a generalized additive model (Hastie and Tibshirani, 1990) with log link and Poisson error within each city, we adjust for smooth functions of temperature, dew point, time, and day of week. We then classify days according to low, medium, and high levels of the standardized residual number of deaths respect to such predictive values. Dominici et al. (1999) give details about the choices of functions used here for the adjustment.

For mortality, O_3, and CO, the cutoff points are chosen differently within each city, based on exploratory analyses. For PM_{10}, we locate the cutoff points at 20 and 50 $\mu g/m^3$. 50 $\mu g/m^3$ corresponds to the annual average of PM_{10} in the review of the national ambient air quality standard (Environmental Protection Agency (EPA), 1996). In this way, the "high' level of PM_{10} correspond to values exceeding the standard.

If little is known about the third-stage distribution, diffuse but proper priors can be used. We implemented this by selecting a multivariate normal prior for $\boldsymbol{\mu}$ and an Inverse Wishart prior for Σ, and by specifying hyperparameters so that the variation of the prior density in a plausible range of the parameters is small. We choose $\boldsymbol{m}$ having all its elements equal to zero, $\boldsymbol{M}$ having all the diagonal elements equal to 1, and all the off-diagonal elements equal to $.5$. This prior specification leads: 1) to the same marginal prior distributions for the cells probabilities π_i's; 2) same joint prior distributions of all the pairs π_i, π_j's; and 3) invariance with respect to the choice of the reference cell π_I. In addition, we choose b and $\boldsymbol{B}$ such that $E[\Sigma] = 2\boldsymbol{M}$, leading to 95 $\%$ HPD for the logits $\boldsymbol{\mu}$ wide enough to let the data drive the posterior results. This prior specification strategy can also be used in other problems as a sensible default choice for the multinomial-LN model introduced here, and it doesn't depend on the number of cells.

We run a chain of length 25,000. The first 1,000 draw were discarded for burn-in. A subchain made of every 5-th of the sampled values was used for making inferences. The autocorrelation of the posterior samples of the probabilities are roughly 0.1 at lag 1 and 0.06 at lag 2. The acceptance probabilities for the Metropolis Algorithm with an independent proposal equal to the $LN(\boldsymbol{\mu}^*, \Sigma^*)$ averaged between .4 and .7 depending on the city.

We start by looking at the mortality-PM_{10} association in New-York, conditionally on O_3=H and CO=H. The reconstructed four-way table in New-York is represented in Table 1 (at the top). Inside each cell, we plot the histogram of the marginal density for the corresponding count. The solid line is placed at the posterior median (number at the top). The sums by column of the posterior distributions are equal to the column margins 71,92,87. They correspond to the right column of table at the bottom; the row margins are found by taking the sums by row of the posterior medians of the three distributions. Table 1 shows a trend in the mortality-PM_{10}

association. At high mortality levels, the posterior medians of the missing counts increase from 9 to 42 as the PM_{10}-level increase.

Attractive feature of our model is that it easily allows for making predictions for a city other than the six sampled. To draw from the posterior predictive distribution of the new counts, say $\boldsymbol{n}^*$, we first draw the parameters from their joint posterior distribution, including the vector of the cells probabilities of the "future" city, say $\boldsymbol{\pi}^*$, and then we simulate $\boldsymbol{n}^*$ from the multinomial distribution $\mathcal{M}_{81}(\boldsymbol{\pi}^*, N^*)$, where N^* is fixed by design, and chosen here to be equal to 2800 days.

We now focus on investigating which combination of pollutant levels is more likely to be associated with a high number of deaths for a "future" city. We rank pollutant level combination depending on the probability of leading to the highest mortality. We estimate these probabilities with the empirical frequencies in the MCMC. Table 2 shows the predictive posterior probabilities that each pollutant mix is the most harmful within each city. Generally, the most harmful combinations corresponds to high levels of PM_{10}. In particular, the two most harmful combinations for PM_{10}, O_3 and CO are HHL and HLM with probabilities 0.18 and 0.17 in Chicago, HLM and HLL with probabilities 0.17 and 0,13 in Pittsburgh, and HLM and HHH with probabilities 0.12 and 0.10 in New-York. In the other cities the probabilities smoothly decrease from 0.09 to 0.01 providing no clear indication on which pollutant mix is the most harmful.

We now turn to assessing how robust our posterior inferences are with respect to prior specifications. We can address this question using a sensitivity analysis. Our strategy is based on selecting three alternative scenarios and one outcome of primary interest, and evaluating the outcome for each scenario; here the outcome of interest is the predictive probability that, for $O_3 = H$ and $CO = H$, the number of days with high mortality and high PM_{10} is greater than the number of days with high mortality and low PM_{10}. Below is a detailed description of the three scenarios: 1) as in the default, except that the logit means are smaller, with all the elements of $\boldsymbol{m}$ equal to

-0.69; 2) as in the default, except that the logit variances, are larger, with the diagonal elements of $\mathbf{M}$ equal to 3; 3) as in the default, except that the logit correlations are set to zero, by setting all the off-diagonal elements of $\boldsymbol{M}$ equal to zero. The posterior predictive probabilities defined below under the default prior, scenario 1, scenario 2, and scenario 3 are respectively $0.8, 0.8, 0.9$, and 0.9 revealing that they are quite stable under the different prior specifications.

5 Discussion

In this paper we describe a hierarchical multinomial-LN model for combining studies reporting categorical variables, when not all of the variables are observed in every study. The strategy considers flexible models for the cell probabilities and allows us to incorporate prior information on the correlations among the logit of the cells probabilities; it provides for practical implementation of MCMC algorithm for sampling from the posterior distributions; it maintains a consistent interpretation of study-specific parameters across studies; and it addresses the uncertainty arising from the missing dimensions. Our model also provides a novel approach for Bayesian reconstruction of contingency tables. In addition, as with all MCMC approaches, the posterior inferences of any measure of association or interaction — within studies, and across studies — are obtained easily by a transformation on the sampled values.

A strength of our approach is that we are considering a saturated model that includes all variables and interactions without making any assumption of independence among the categorical variables. To assure identifiability of high order interactions, and to properly impute the missing counts for the incomplete tables, the data must provide at least one complete contingency table.

We have proposed an original modeling strategy for investigating the association between air pollution and mortality for six Eastern U.S. cities. Our approach can accommodate non-linearities in the exposure-response relationship, can identify change points in pollutant concentrations that cause a significant adverse health impact, and can lend itself easily to the inclusion of interactions of every order between the pollutants. A drawback is the loss of information resulting from the discretization. In regards to what conclusions can be drawn from this analysis: 1) Tables 1 indicate that – for high levels of O_3 and CO – there is a trend in the effects of PM_{10}, e.g. number of cases with of high number of deaths increases as the levels of PM_{10} increases; 3) Table 2, indicates that the most harmful combinations corresponds to high levels of PM_{10}, confirming prior findings that levels of PM_{10} higher than the NAAQS standard are harmful.

The advantages of borrowing information from other studies is in this application illustrated how within the incomplete cities the model has provided estimates of all cell probabilities, without any observations being classified with respect to all the variables.

The approach described here relies crucially on the assumption that the missing dimensions are missing at random. For example for the Minneapolis data, this assumption implies that the probability of having a missing ozone level might depends on the other pollutants levels, but not on the missing ozone measurement itself. Generally this is a plausible assumption. The assumption of data missing at random would be violated if the measurements were censored, i.e. they were recorded using monitors that do not report pollutants levels below or above a fixed limit. Inference for non ignorable designs in the context of our model has not been addressed. One solution approach is based on the inclusion of additional stages in the hierarchical model, describing the missingness models.

Finally, in the analysis of the contingency tables, dimensionality increases results in a tremendous increase in the number of possible associations patterns and higher-order interactions. We provide a general strategy for combining high dimensional contingency tables with missing dimensions, that does not require any assumptions about high order interactions, and can be suitably implemented independently of the dimensionality of the contingency tables.

References

Aitchison, J. (1985a). A general class of distributions on the simplex. *Journal of the Royal Statistical Society, B* 136–146.

Aitchison, J. (1985b). Practical Bayesian problems in simplex sample spaces. In: *Bayesian Statistics 2*, 15– 31.

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman and Hall, London.

Aitchison, J. and Shen, S. M. (1980). Logistic normal distributions: Some properties and use. *Biometrika* **67**, 261–272.

American Thoracic Society, Bascom, R. (1996a). Health effects of outdoor air pollution, Part 1. *American Journal of Respiratory and Critical Care Medecine* **153**, 3–50.

American Thoracic Society, Bascom, R. (1996b). Health effects of outdoor air pollution, Part 2. *American Journal of Respiratory and Critical Care Medecine* **153**, 477–498.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analyses: Theory and Practice*. MIT Press.

Daniels, M. and Kass, R. (1997). *Non conjugate Bayesian estimation of covariance matrices and its use in hierarchical models*. Report 659, Department of Statistics, Carnegie Mellon University.

Dockery, D., Pope, C. I., and Xu X., e. a. (1993). An association between air pollution and mortality in six U.S. cities. *New England Medical Journal* **329(24)**, 1753—1759.

Dominici, F., Samet, J., and Zeger, S. (1999). Combining evidence on air pollution and daily mortality from the twenty largest us cities: A hierarchical modeling strategy. *Royal Statistical Society, Series C, read paper (to appear)* .

Environmental Protection Agency (EPA) (1996). Review of the National Ambient Air Quality Standards for Particulate Matter: Policy Assessment of Scientific and Technical Information. OAQPS Staff Paper. Research Triangle Park, North Carolina, U.S. Government Printing Office. *Environmental Protection Agency (EPA)* .

Forster, J. and Skene, A. (1994). Calculation of marginal densities for parameters of multinomial distributions. *Statistics ans Computing* **4**, 279–286.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**(410), 398–409.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (1995). *Bayesian Data Analysis*. London: Chapman and Hall.

Goutis, C. (1993). Bayesian estimation methods for contingency tables. *Journal of the Italian Statistical Society* **2**, 35–54.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*. Chapman and Hall, New York.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109.

Kelsall, J., Samet, J., and Zeger, S. (1997). Air pollution, and mortality in Philadelphia, 1974-1988. *American Journal of Epidemiology* **146**, 750—-762.

Knuiman, M. W. and Speed, T. P. (1988). Incorporating prior information into the analysis of contingency tables. *Biometrics* **44**, 1061–1071.

Leonard, T. (1975). Bayesian estimation methods for two-way contingency tables. *Journal of the Royal Statistical Society, Series B, Methodological* **37**, 23–37.

Li, Y. and Roth, H. (1995). Daily mortality analysis by using different regression models in Philadelphia county, 1973-1990. *Inhalation Toxicology* **7**, 45–58.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

Lipfert, F. (1994). *Air Pollution and Community Health: A Critical Review and Data Sourcebook*. New York: Van Nostrand Reinold.

Lipfert, F. and Wyzga, R. (1993). Air pollution and mortality: Issues and uncertainty. *Journal Air Waste Manage. Assoc* **45**, 949—966.

Nazaret, W. A. (1987). Bayesian log linear estimates for three-way contingency tables. *Biometrika* **74**, 401–410.

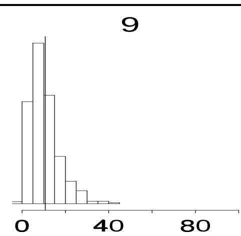
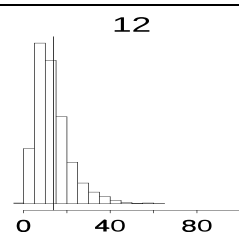
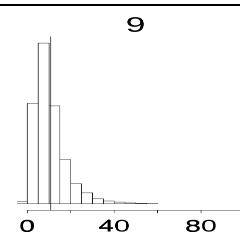
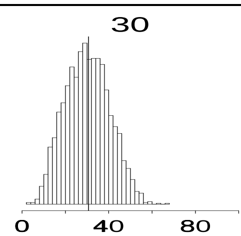
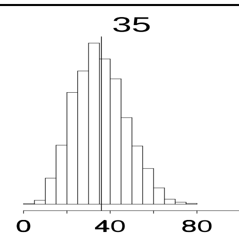
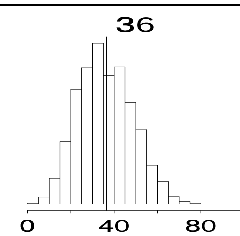
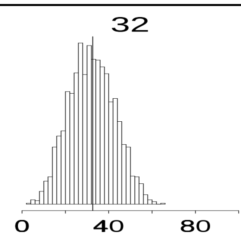
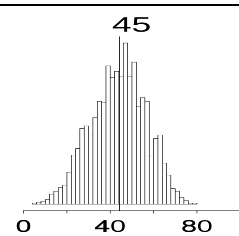
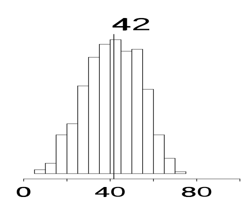
Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–590.

Samet, J., Zeger, S., Kelsall, J., Xu, J., and Kalkstein, L. (1997). *Air pollution, weather and mortality in Philadelphia, In Particulate Air Pollution and Daily Mortality: Analyses of the Effects of Weather and Multiple Air Pollutants. The Phase IB report of the Particle Epidemiology Evaluation Project*. Health Effect Institute, Cambridge, MA.

Tanner, M. A. (1991). *Tools for Statistical Inference – Observed Data and Data Augmentation Methods*, volume 67 of *Lecture Notes in Statistics*. New York: Springer-Verlag.

Vedal, S. (1996). Ambient particles and health: Lines that divide. *J. Air and Waste Manage. Assoc.* **47**, 551—-581.

Table 1: *At the top, marginal posterior distributions for the missing counts in New-York conditionally at O_3=H and CO=H. The solid line is placed at the rounded posterior median (number at the top). The sums by column of the posterior distributions are equal to the column margins 71,92,87. They correspond to the right column of table at the bottom; the row margins are found by taking the sums by row of the posterior means of the three distributions. For clarity of exposition the ranges of the histograms are truncated at 100.*

	D			
PM_{10}	L	M	H	tot
L				30
M				101
H				119
tot	**71**	**92**	**87**	250

CO	L			M			H		
	O_3								
D	L	M	H	L	M	H	L	M	H
L	54	128	82	110	143	106	113	63	**71**
M	69	177	115	130	182	142	140	116	**92**
H	37	130	81	121	140	92	100	85	**87**

Table 2: Posterior probability that each pollutant combination is the most harmful