REVSTAT – Statistical Journal
Volume 9, Number 1, March 2011, 99–114

MIXTURES OF FACTOR MODELS FOR MULTIVARIATE DISEASE RATES

Authors: T.C. Bailey
– College of Engineering, Mathematics and Physical Sciences, University of Exeter, U.K.
t.c.bailey@exeter.ac.uk

P.J. Hewson
– School of Mathematics and Statistics, University of Plymouth, U.K.
paul.hewson@plymouth.ac.uk

Abstract:

- A range of different approaches have been suggested for the multivariate modelling of the geographical distribution of different but potentially related diseases. We suggest an addition to these methods which incorporates a discrete mixture of latent factors, as opposed to using CAR or MCAR random effect formulations. Our proposal provides for a potentially richer range of dependency structures than those encompassed in previously used models in that it is capable of representing an enhanced range of correlation structures between diseases at the same time as implicitly allowing for less restrictive spatial correlation structures between geographical units. We illustrate results of using the model on data taken from cancer registries on four carcinomas in some 300 UK geographical areas.

Key-Words:

- *mixture models; factor analysis; multivariate disease rates.*

AMS Subject Classification:

- 49A05, 78B26.

1. INTRODUCTION

The literature in spatial epidemiology contains a growing number of references to multivariate modelling of the geographical distribution of morbidity or mortality rates for potentially related diseases. Dabney and Wakefield [9] suggest that the two main motivations for this interest are firstly, to explore similarities or dissimilarities in the geographical risk distribution for the different diseases, and, secondly, to 'borrow strength' across disease rates to shrink the uncertainty in geographical risk assessment for any one of the individual diseases. Regardless of the relative balance of interest between these two motivations, it is clear that achievement of either objective will be limited if the structure of the multivariate model used is inadequate and/or its related assumptions are unrealistic. In particular, the possible dependency structure (either between diseases, or across geographical units) should not be overly constrained by the model structure. For example, it may not be the case that relationships between diseases are the same in, say, rural versus urban environments, nor that dependency between disease rates in neighbouring small areas will be homogeneous at a larger geographical scale where spatial discontinuities may well be present. The model structure should provide a sufficiently rich range of dependency structures to encompass such possibilities.

Various approaches to spatial modelling of multivariate disease rates have been proposed. Many of these may be characterised as generalised linear mixed models (GLMMs) of varying descriptions in which the dependence structure not explained by covariates is represented in terms of random effects which are correlated between diseases and across geographical units. The Multivariate Conditional Autoregressive (MCAR) model is one popular approach for dealing with multivariate disease rates in small areas (e.g. [13, 26, 10]) but some have commented that MCAR formulations remain difficult to fine tune because the correlations in random effects between diseases and/or across spatial units are not easy to disentangle. Models which incorporate more explicit latent structure also feature in multivariate modelling of disease rates. Held *et al.* [22] review a range of approaches to joint disease modelling including shared latent processes. Early examples include that used by Knorr-Held and Best [24] to identify a shared spatial component in the geographical distribution of bivariate disease rates and the simple latent variable formulation employed by Wang and Wall [36] to model Minnesota cancer rates. More recently, Liu *et al.* [27] have proposed a Structural Equation Model (SEM) for cancer rates where the three cancers have a single shared spatially structured latent variable. Latent structure models are not restricted to area applications. Christensen and Amemiya [6] have suggested an approach applicable to point data which has been illustrated by Minozzo and Fruttini [30] who examined bivariate point measures of types of diabetes morbidity.

MCAR versus explicit latent structure aside, there are two assumptions which dominate most of these previously described multivariate models. First, that spatial dependency across small areas is essentially 'smooth' and not subject to global spatial discontinuities. Second, that it is reasonable to assume that a single relationship between diseases counts applies to all types of areas. In some (perhaps many) applications the dependence structure may well be more complex than that implied by these rather broad assumptions. Normand *et al.* [31] highlight that in the absence of adequate covariate information, simple exchangeability assumptions across areas may not be valid in many of the GLMMs used to analyse healthcare provision. In a multivariate setting these exchangeability concerns across areas remain, but are compounded by additional concerns over whether the dependence structure between diseases varies geographically.

We therefore propose a model for use in such contexts which potentially provides a richer range of dependency structures than those encompassed by previous approaches. Rather than representing the dependence structure not explained by covariates in terms of correlated random effects, we suggest that it is preferable to formulate correlations in terms of an explicit latent structure similar to that arising in factor analysis. Our model is based on latent structure mixtures and we argue that incorporating a discrete mixture into the latent structure loadings in the model simultaneously provides potential to represent an enhanced range of correlation structures between diseases, at the same time as allowing for less restrictive spatial correlation structures between geographical units.

The structure we propose could be considered similar to a 'mixture of factor analysers'. Such models (mostly for Gaussian responses) have been reported in other contexts in the statistical literature and elsewhere. For example, Mclachlan and Peel [29] discuss mixtures of factor analysers, Lee and Song [25] report on mixtures in relation to Structural Equation Modelling, and Viroli [35] describes 'independent factor analysis' based upon approaches developed in the signal processing literature. Many insights into the properties of GLMMs for multivariate disease rates can be gained from studying recent developments in factor analysis which has been enjoying somewhat of a methodological renaissance in a Bayesian setting [1] with a number of useful results emerging. The development of our latent structure mixture model for joint disease modelling in this paper is encouraged by these results and draws upon our belief in the value of viewing correlated random effects in a factor analysis framework.

In Section 2 we develop our model and describe fitting strategy. In Section 3 we introduce an illustrative data set on which to demonstrate results which concerns four cancers in some 300 geographical units in England, Scotland and Wales. We present model results for these data in Section 4 and then go on to discuss conclusions in Section 5.

2. MODEL FORMULATION

The basic structure of the problem we consider is that we have data, y_{ij}, representing the number of cases in area i for disease j $(i = 1, ..., n;\ j = 1, ..., p)$. The corresponding expected number of cases e_{ij} is also known, this being based on age/sex standardised rates for the whole of the study region, or for some appropriate alternative reference population (equivalently, we may know y_{ij} along with the standard morbidity ratio (SMR) y_{ij}/e_{ij}, for disease j in area i). Where appropriate we will refer to the vector of disease counts in each area as $\boldsymbol{y}_i = (y_{i1}, ..., y_{ip})$ and the corresponding vector of expected counts as $\boldsymbol{e}_i = (e_{i1}, ..., e_{ip})$. In a practical setting we may well also have a vector of covariates $\boldsymbol{x}_i$ $(i = 1, ..., n)$ measured in each area, but for simplicity of exposition we will assume throughout this paper that such covariates are not available. If required these can be included into the models we develop in an obvious and straightforward fashion.

It is usual to assume disease counts are Poisson distributed viz: $y_{ij}|\lambda_{ij} \sim \text{Pois}(e_{ij}\lambda_{ij})$, with the mean vector, $\boldsymbol{\lambda}_i = (\lambda_{i1}, ..., \lambda_{ip})$, in each area then being modelled through an appropriate link function by a suitable linear predictor. In developing our modelling framework we build upon proposals made by Wang and Wall [36] mentioned in Section 1 which used a log link and a simple linear predictor involving a single area specific latent variable with a disease specific loading, so that: $\log(\lambda_{ij}) = \phi_i \delta_j$, where δ_j is the disease specific loading and ϕ_i is the area specific latent (unmeasured) variable which was in turn assumed to follow a Conditional Auto-Regressive Gaussian (CAR) distribution over the areas. In this model correlation between diseases within an area is reflected through the shared latent variable and spatial correlation across areas is achieved via the CAR). However, the simple structure only provides for a limited range of correlation structures between diseases (same for all areas) and makes possibly unrealistic assumptions about the spatial dependence (it is 'smooth' — there is no possibility of global spatial discontinuity). We therefore consider ways to provide more complex possibilities for dependencies between diseases and across areas.

First, to allow potential for a more complex dependence structure between diseases, we include q latent variables. So that the model becomes

$$\log(\lambda_{ij}) \ = \ \sum_{h=1}^{q} \phi_{ih}\, \delta_{jh} \ , \tag{2.1}$$

where δ_{jh} is a disease specific loading for area specific latent variable ϕ_{ih} $(h = 1, ..., q)$. We can express this more succinctly as

$$\log(\boldsymbol{\lambda}_i) \ = \ \boldsymbol{\phi}_i \boldsymbol{\Delta} \ , \tag{2.2}$$

where it is understood that $\log(\boldsymbol{\lambda}_i) = (\log \lambda_{i1}, ..., \log \lambda_{ip})$ and where $\boldsymbol{\phi}_i = (\phi_{i1}, ..., \phi_{iq})$

is the vector of latent variables for area i and $\boldsymbol{\Delta}$ is the $q \times p$ matrix of loadings:

$$\boldsymbol{\Delta} = \begin{pmatrix} \delta_{11} & \cdots & \delta_{p1} \\ \vdots & \ddots & \vdots \\ \delta_{1q} & \cdots & \delta_{pq} \end{pmatrix}.$$

This formulation raises identifiability problems (e.g. rotational indeterminacy) so we follow Lopes and West [28] and constrain the loading matrix $\boldsymbol{\Delta}$ so that it is upper triangular with the diagonal strictly positive, i.e.:

$$\boldsymbol{\Delta} = \begin{pmatrix} \delta_{11} & \delta_{21} & \cdots & \delta_{p1} \\ 0 & \delta_{22} & \ddots & \delta_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_{pq} \end{pmatrix}.$$

To incorporate a richer range of dependency across areas we then further extend model (2.2) to a mixture model across s sets of q latent variables in each area, with $\boldsymbol{\phi}_i^{(k)} = \left(\phi_{i1}^{(k)}, ..., \phi_{iq}^{(k)}\right)$ denoting the k^{th} set of latent variables. So that the model becomes

$$y_{ij} \mid \lambda_{ij}^{(1)}, ..., \lambda_{ij}^{(s)} \sim \sum_{k=1}^{s} \pi_{ik} \operatorname{Pois}\left(e_{ij}\lambda_{ij}^{(k)}\right),$$

with π_{ik} denoting mixing probabilities ($\sum_k \pi_{ik} = 1$) and with

$$\log\left(\boldsymbol{\lambda}_i^{(k)}\right) = \boldsymbol{\phi}_i^{(k)} \boldsymbol{\Delta}^{(k)},$$

where $\boldsymbol{\Delta}^{(k)}$ is a $q \times p$ matrix of loadings for the the k^{th} set of latent variables ($k = 1, ..., s$) and with each such matrix is subject to the constraints described earlier. The latent variables $\phi_{ih}^{(k)}$ are assumed to follow independent Gaussian distributions with means $\mu_h^{(k)}$ for $k = 1, ..., s,\ h = 1, ..., q$ and $i = 1, ..., n$.

So each area is now a mixture of s types of areas, with each type of area being associated with a different set of q latent variables and corresponding loadings. Note there is no explicit spatial dependence (e.g. CAR or MCAR) in the above formulation. However, implicit spatial dependence arises through groups of areas being free to share a similar pattern of mixing probabilities over the sets of latent variables and loadings. This type of spatial dependence is potentially very flexible since it does not necessarily impose undue global spatial smoothness.

Finally, we incorporate additional unstructured area and disease specific random effects into the linear predictor of our formulation above in order to deal with possible overdispersion. These additional random effects are effectively equivalent to 'uniqueness' in the traditional factor analysis literature.

So the final model then becomes

$$\log\left(\boldsymbol{\lambda}_i^{(k)}\right) = \boldsymbol{\phi}_i^{(k)} \boldsymbol{\Delta}^{(k)} + \boldsymbol{\zeta}_i, \tag{2.3}$$

where the random effects, $\boldsymbol{\zeta}_i = (\zeta_{i1}, ..., \zeta_{ip})$, are independent zero mean Gaussian with variances drawn from inverse Gamma hyperpriors.

Using mixture models as in the above formulation, raises a number of fitting and identifiability issues Mclachlan and Peel [29]. We use an MCMC fitting approach with a flat hyperprior for the group means $\mu_h^{(k)}$ of the latent variables; and with loadings δ_{jh} (subject to the identifiability constraints described earlier) given zero mean Gaussian priors with inverse Gamma hyperpriors for their distinct associated variances. The model involves both unknown numbers of latent variables and mixture components, but there is considerable complexity in using dimension changing methods (e.g. RJMCMC) even with just unknown numbers of latent variables (e.g. [28]) let alone when this is compounded with an unknown number of mixture components. We therefore follow suggestions made by Green [19] and use a strategy whereby distinct models are fitted to distinct dimensionalities. In selecting the number of mixture components we draw on similarities between our latent structure mixture and the 'mixture of factor analysers' model [29, Chap 8]. We note that in the machine learning literature, 'Variational Bayes' approaches are used to fit 'mixtures of factor analysers' [17, 3] which are essentially equivalent to minimising the Kullback–Leibler distance between the factorised approximation and the joint posterior. We have therefore selected a strategy based on the Kullback–Leibler distance in order to select the number of mixture components. We use this measure to assess the distance between our fitted model and a model which assumes the two closest mixture components have been merged. An approximation to the Kullback–Liebler distance can be generated as a byproduct of Gibbs sampling [32] and we use this to guide model selection with respect to the number of substantive mixture components which may be supported by the data. We follow Celeux *et al.* [5] by not placing any constraints on the ordering of the mixture group means and deal with label-switching by post-processing the output of our MCMC sampler. Mixture group memberships for loadings and latent variables (where applicable) were assumed to be categorical variables with a Dirichlet prior for the probabilities, $\boldsymbol{\pi}_i = (\pi_{i1}, ..., \pi_{is})$, of belonging to the different groups, i.e.

$$p(\pi_{i1}, ..., \pi_{is}) \;=\; \frac{\Gamma(\alpha_1 \cdots \alpha_s)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_s)} \; \pi_{i1}^{\alpha_1 - 1} \cdots \pi_{is}^{\alpha_s - 1} \,,$$

where α_k represents the prior group weights for each of the $k = 1, ..., s$ mixing components.

Routine checks for MCMC convergence were used involving Gelman and Rubin's R [15], Geweke's statistic [16] and Heidelberger and Welch's statistic [21]. These are slow models to fit, running multiple chains is therefore somewhat time consuming but it is essential given the level of cross-correlation possible due the model formulation. Whilst somewhat slower than using customised code, we used the widely accessible WinBUGs software package [34] to fit models.

Overall model fit was assessed by a number of measures. In addition to assessing mixture group membership in terms of Kullback–Leibler distance, we also considered Posterior Predictive Loss as proposed by Gelfand and Ghosh [14] as well as a variant on this proposed by Gneiting and Raftery [18] which enjoys the advantage of being a proper scoring rule and has been examined specifically in respect of count data by Czado *et al.* [8]. We refer to this proper score as the 'Dawid and Sebastiani score' following earlier work reported in Dawid and Sebastiani [11].

We also validated the performance of our model by means of out of sample predictions using the posterior predictive distribution of the relative risks for each disease for each area studied. We removed one tenth of the observations at random, and fitted the model to these data. The posterior predictive density for these deleted observations was collected during the model fitting process, and compared with the observed data.

3. DATA

As an illustrative application to demonstrate model performance, we consider data on reported numbers of cases of four types of cancer in some 300 geographical units covering England, Wales and Scotland. These data were obtained from the 9 cancer registries in England as well as the cancer registries in Scotland and Wales and comprise the number of cases reported between 1999 and 2001 of, 'Lung cancer' (ICD-10 classified sites C33-C34, i.e. Trachea, Bronchus and Lung cancer), 'Oral cancer' (C00-C14, i.e. Lip, Oral Cavity and Pharynx), Breast cancer (C50) and Cancer of the Cervix (C53). Data were collected on prevalence for males and females for the first two cancers, but only for females in respect of breast and cervical cancer. Direct standardisation [7] was used to estimate associated expected morbidity based on quinary age bands for the whole of the study region.

These data refer to the smallest administrative geographical unit available, i.e. the 303 Primary Care Trusts (PCT) in England, the 22 Local Health Boards in Wales and the 14 National Health Service Boards in Scotland. For convenience we will subsequently refer to all such units by the name given to the majority, namely 'PCTs'. The English and Welsh entities are comparable in size, for example the mean population within an English PCT is 163,000 with a minimum 63,700 and a maximum 372,600 whereas the Welsh Local Health Board mean population was 131,900 with a minimum of 56,500 and a maximum of 310,300. Scotland is dominated by a couple of very large NHS Boards, the mean population was 720,320 with a minimum of 38,400 and a maximum of 1,736,300. Some caution may therefore be needed when comparing results from England and Wales with

those of Scotland due to aggregation effects alone. We concentrate here on those 335 'PCTs' which are entirely based on the mainland, i.e we exclude islands. It should be noted for later reference that one of the mainland PCTs in Cornwall (the far South West of our maps) contains an aggregate of data from the Isles of Scilly.

Basic information about the number of cancers registered under each diagnosis in each PCT are contained in Table 1. As is usual with administratively collected data, there are some provisos over accuracy. Particular problems with UK cancer registry data are documented in Best and Wakefield [4], and we note that it may not be entirely reasonable to assume that each cancer registry collects the data in exactly the same way.

Table 1: Summary information on the mean, standard deviation, minimum and maximum number of cancer cases registered for each of the six diagnosis groups in each of the 335 non-island PCTs in England, Scotland and Wales.

	Oral (F)	Lung (F)	Oral (M)	Lung (M)	Breast	Cervix
Mean	16.25	132.81	28.65	205.06	359.99	26.68
Std. Dev.	11.62	107.06	23.04	136.67	184.47	17.62
Minimum	1	29	3	41	107	4
Maximum	133	1333	276	1653	1832	175

Table 2 gives the observed correlation coefficients between the various cancer rates. It can be seen for example that Lung cancer rates are highly correlated between males and females (0.88), the same is not so true of oral cancer rates (0.31). Figure 1 provides the same information in graphical form.

Table 2: Observed correlation between cancer rates for the four cancers, male and female data shown separately.

	Oral (F)	Lung (F)	Oral (M)	Lung (M)	Breast	Cervix
Oral (F)	1.00	0.23	0.31	0.24	0.04	0.22
Lung (F)	0.23	1.00	0.52	0.88	−0.32	0.47
Oral (M)	0.31	0.52	1.00	0.52	−0.18	0.35
Lung (M)	0.24	0.88	0.52	1.00	−0.39	0.49
Breast	0.04	−0.32	−0.18	−0.39	1.00	−0.18
Cervix	0.22	0.47	0.35	0.49	−0.18	1.00

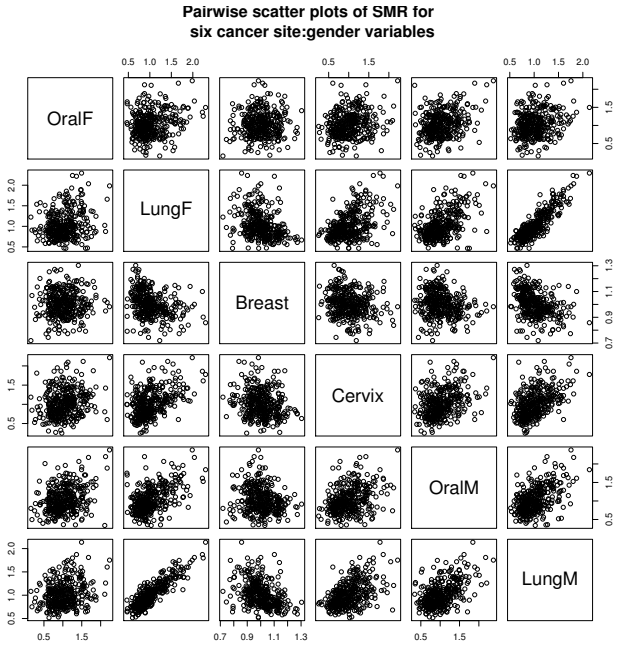

Figure 1: Pairwise scatter plots for the six age standardised cancer rates.

4. RESULTS

Models were fitted as described in Section 2, using the priors and convergence criteria indicated there. A standard burn in period of 50,000 iterations was used, a further 100,000 samples thinned by a factor of 20 were used for posterior inference. As mentioned previously, all results reported here were obtained using the WinBUGs software package [34].

We fitted a range of models with differing numbers of latent variables and differing numbers of mixture components. Given $p = 6$ disease counts we considered all the classically identifiable possibilities, i.e. $k = 1, ..., 3$ latent variables. It was found feasible to fit a three latent variable model and we prefer that both because it has the greatest potential to model a complex dependence structure and because it has the lowest posterior predictive score. In general, the posterior predictive Dawid and Sebastiani score tends to favour models with a larger number of mixture components. However, Kullback–Leibler tends to favour a two component solution. Figure 2 contains a density plot of the sampled values for the approximate Kullback–Leibler distance between a two component mixture and a one component model and indicates a considerable distance between the two and one component means of the second latent variable. Given this support from the Kullback–Leibler distance we accept a two component solution despite

the fact that this fits slightly less well than higher numbers of components on the Dawid and Sebastiani score.

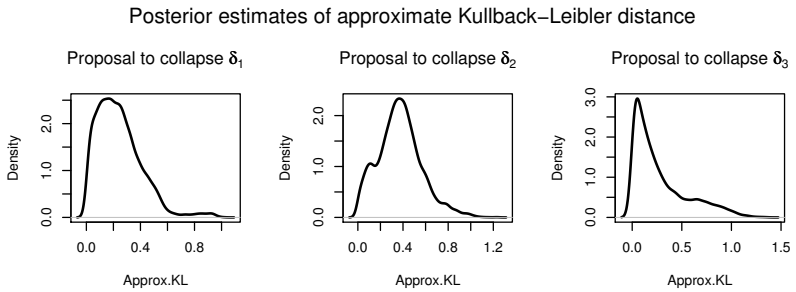

Figure 2: Approximate Kullback–Leibler distance between two and one component latent structure mixture model with three latent variables when considering each of the latent variables.

Figure 3 presents maps of the geographical distribution of raw and model estimated posterior mean relative risk for Breast cancer. The model achieves a degree of shrinkage in terms of the posterior mean of the relative risks when compared with the raw data. Maps for the other cancer counts reveal a similar story.

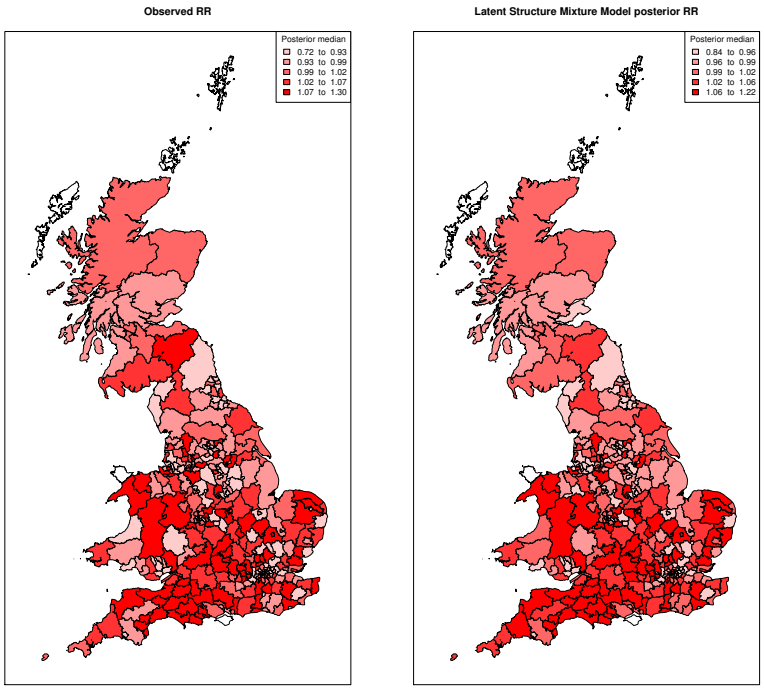

Figure 3: Breast cancer raw rates and posterior mean relative risk from latent structure mixture model.

Perhaps more interesting, is that with this model it is possible to examine posterior mean mixing probabilities for each PCT. Figure 4 gives the posterior groupings of PCTs associated with this measure. It is clear that although no explicit spatial structure is imposed in this model, the mixture groups appear to be highlighting a spatial pattern that has a substantive interpretation (Scotland and industrial areas in England and Wales). There does therefore appear appear to be some interesting possibilities in using this type of model formulation.

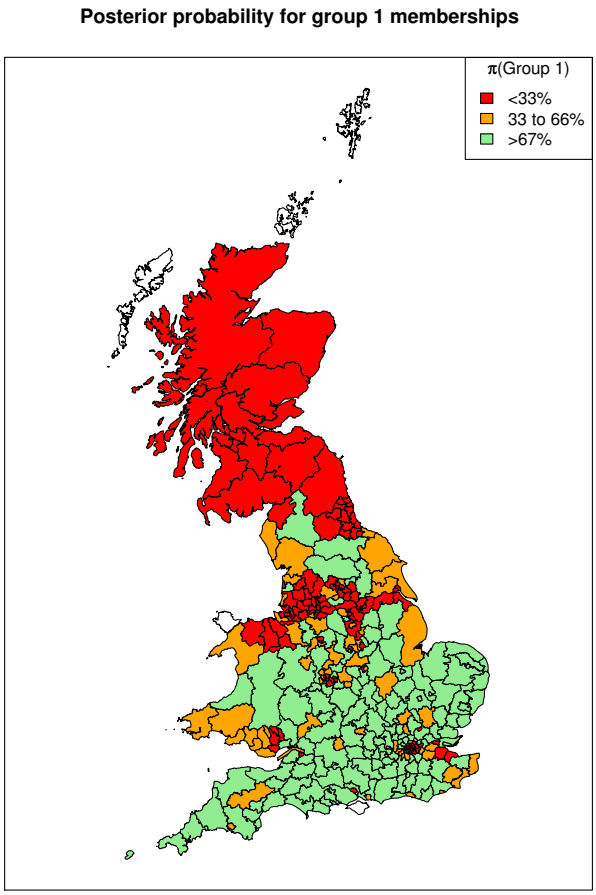

Figure 4: Posterior probability of PCT group membership for latent structure mixture model with two components and three latent variables.

Finally, we present illustrative results demonstrating the out of sample performance of the model. A random 35 PCTs had data removed for a a randomly selected cancer site (Female Lung Cancer). Results are depicted in Figure 5 which contrasts the posterior predictive density for the omitted data with the actual data that had been excluded from the model.

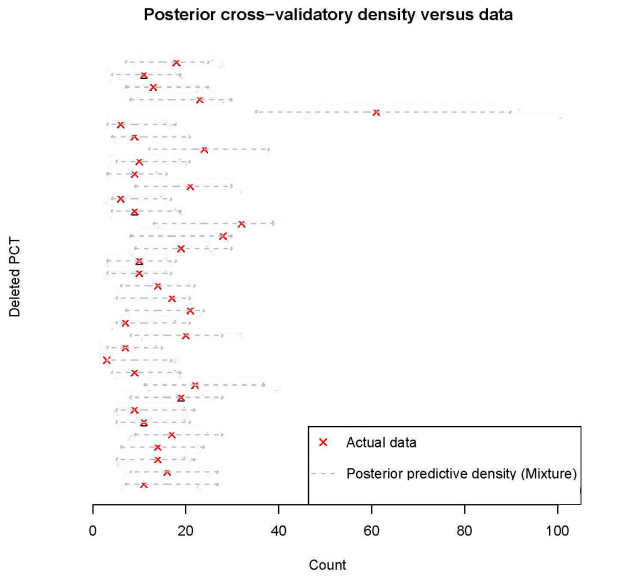

Figure 5: Out of sample posterior predictive density for 35 'PCTs' randomly removed from the Female Lung Cancer site. The actual removed data points have been superimposed.

5. DISCUSSION

We believe that the latent structure mixture model developed in this paper provide a tractable approach to handling situations in joint disease modelling where it may be anticipated that a single dependence structure, either between diseases or across geographical units, is overly restrictive. We also believe that such situations are not uncommon and that aspects of the illustrative cancer morbidity data we have examined substantiate the argument for employing mixture models as a way of avoiding unreasonable exchangeability assumptions.

Our primary focus has been on statistical methodology, rather than identifying any substantive epidemiological issues arising from the particular cancer morbidity data we have examined. That said, there could well be interesting epidemiological distinctions between the areas discriminated using our approach as reported in Section 4. It is quite striking that the areas with lowest probability of group 1 membership tending to correspond to former industrial areas of Scotland, North England and Wales, those PCTs with the highest probability of group 1 membership tending to correspond to more affluent and rural areas in Southern England. This fits well with the epidemiology of these diseases, lifestyle factors such as alcohol and tobacco consumption being more dominant in the

non-group 1 areas (hence greater lung and oral cancer) and other factors being responsible for greater breast cancer risk in the group 1 areas. However, we are cautious of over-interpretation in this regard. Further work is needed to deal with these models in such a way that the mixtures on the loadings can be disentangled from the mixtures on the latent variables, but it does appear from our results that the two structures do act differently.

We have concentrated on modelling dependency structure and not explicitly addressed use of additional covariate information on geographical units other than routine standardisation for age/sex population structure. We appreciate that in practice it is very likely that relevant additional covariate information will be available on the geographical units concerned. If so, then this can easily be handled by simply including relevant fixed effects into the linear predictor of the model we have proposed and does not present any additional methodological challenges.

In summary, we believe that incorporating a mixture distribution into a latent structure model has considerable potential in modelling multivariate disease rates. The advantages of using a latent structure model relate to the transparent way in which correlation structure is represented in the model allowing the modeller to tune this accordingly. It is less obvious how to do this within, say, the MCAR formulation where the latent structure is not explicit. We appreciate that in this paper we have not carried out any formal comparison of the fit of our proposed model to other formulations such as the MCAR. This topic is taken up and reported elsewhere in an expanded version of this paper (see [23]).

REFERENCES

[1] Aitkin, M. and Aitkin, I. (2005). *Bayesian inference for factor scores.* In "Contemporary Advances in Psychometrics" (A. Maydeu-Olivares and J.J. McArdle, Eds.), New York, Erlbaum.

[2] Banerjee, S.; Carlin, B.P. and Gelfand, A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, Chapman and Hall / CRC.

[3] Beal, M.J. and Ghahramani, Z. (2003). *The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures.* In "Bayesian Statistics 7" (J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith and M. West, Eds.), Oxford University Press.

[4] Best, N. and Wakefield, J. (1999). Accounting for inaccuracies in population counts and case registration in cancer mapping studies, *J.R. Statist. Soc. A*, **162**, 363–382.

[5] Celeux, G.; Hurn, M. and Robert, C.P. (2000). Computational and inferential difficulties with mixture posterior distributions, *J. Am. Statist. Ass.*, **95**, 957–970.

[6] Christensen, W.F. and Amemiya, Y. (2002). Latent variable analysis of multivariate spatial data, *J. Am. Statist. Ass.*, **97**, 302–317.

[7] Clayton, D.G. and Kaldor, J. (1987). Empirical Bayes estimates of age-standardised relative risks for use in disease mapping, *Biometrics*, **43**, 671–681.

[8] Czado, C.; Gneiting, T. and Held, L. (2007). Predictive model assessment for count data, *University of Washington Department of Statistics Technical Report*, Number 518.

[9] Dabney, A. and Wakefield, J. (2005). Issues in the mapping of two diseases, *Statistical Methods in Medical Research*, **14**, 83–112.

[10] Daniels, M.J. and Kass, R.E. (2001). Shrinkage estimators for covariance matrices, *Biometrics*, **57**, 1173–1184.

[11] Dawid, A.P. and Sebastiani, P. (1999). Coherent dispersion criteria for optimal experimental design, *Annals of Statistics*, **27**, 66–81

[12] Dreassi, E. (2007). Polytomous disease mapping to detect uncommon risk factors for related diseases, *Biometrical Jnl.*, **49**, 520–529.

[13] Gelfand, A. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis, *Biostatistics*, **4**, 11–15.

[14] Gelfand, A.E. and Ghosh, S.K. (1998). Model choice: a minimum posterior predictive loss approach, *Biometrika*, **85**(1), 1–11.

[15] Gelman, A. and Rubin, D. (1992). Inference from iterative simulation using multiple sequences, *Statistical Science*, **7**, 457–511.

[16] Geweke, J. (1992). *Evaluating the accuracy of sampling-based approaches to calculating posterior moments.* In "Bayesian Statistics 4" (J. Bernado, J. Berger, A. Dawid and A. Smith, Eds.), Oxford, Clarendon Press.

[17] Ghahramani, Z. and Beal, M. (2000). *Variational Inference for Bayesian Mixtures of Factor Analysers*, MIT Press.

[18] Gneiting, T. and Raftery, A.E. (2007). Strictly proper scoring rules, prediction and estimation, *J. Am. Statist. Ass.*, **102**, 359–378.

[19] Green, P.J. (2003). *Trans-dimensional Markov chain Monte Carlo.* In "Highly Structured Stochastic Systems" (P. Green, N.L. Hjort and S. Richardson, Eds.), Number 27 in Oxford Statistical Science Series, pp. 179–198, Oxford, Oxford University Press.

[20] Hardy, G.; Littlewood, J. and Pólya, G. (1988). *Some Theorems Concerning Monotonic Functions* (2nd ed.), pp. 83–84, Cambridge, UK, Cambridge University Press.

[21] Heidelberger, P. and Welch, P. (1983). Simulation run length control in the presence of an initial transient, *Opns. Res.*, **31**, 1109–44.

[22] Held, L.; Natario, I.; Fenton, S.; Rue, H. and Becker, N. (2005). Towards joint disease mapping, *Statistical Methods in Medical Research*, **14**, 61–82.

[23] Hewson, P.J. and Bailey, T.C. (2010). Modelling multivariate disease rates with a latent structure mixture model, *Statistical Modelling*, **10**, 241–264.

[24] Knorr-Held, L. and Best, N. (2001). A shared component model for joint and selective clustering of two diseases, *J. R. Statist. Soc. A*, **164**, 73–86.

[25] Lee, S.-Y. and Song, X.-Y. (2003). Bayesian model selection for mixtures of structural equation models with an unknown number of components, *British Journal of Mathematical and Statistical Psychology*, **56**, 145–165.

[26] Leonard, T. and Hsu, J.S. (1992). Bayesian inference for a covariance matrix, *The Annals of Statistics*, **20**(4), 1669–1696.

[27] Liu, X.; Wall, M.M. and Hodges, J.S. (2005). Generalized spatial structural equation models, *Biostatistics*, **6**(4), 539–557.

[28] Lopes, H. and West, M. (2004). Bayesian model assessment in factor analysis, *Statistica Sinica*, **4**, 41–67.

[29] McLachlan, G.J. and Peel, D. (2000). *Finite Mixture Models*, New York, J. Wiley and Sons.

[30] Minozzo, M. and Fruttini, D. (2004). Loglinear spatial factor analysis: an application to diabetes mellitus complications, *Environmetrics*, **15**, 423–434.

[31] Normand, S.-L.T.; Glickman, M.E. and Gatsonis, C.A. (1997). Statistical methods for profiling providers of medical care: issues and applications, *J. Am. Statist. Ass.*, **92**, 803–814.

[32] Sahu, S. and Cheng, R. (2003). A fast distance based approach for determining the number of components in mixtures, *The Canadian Journal of Statistics*, **31**, 3–22.

[33] Sargent, D.J.; Hodges, J.S. and Carlin, B.P. (2006). Structured Markov chain Monte Carlo, *Journal of Computational and Graphical Statistics*, **9**, 217–234.

[34] Spiegelhalter, D.; Thomas, A. and Best, N. (1998). WinBUGS version 1.1.1 User Manual, *MRC Biostatistics Unit Technical Report*, Cambridge.

[35] Viroli, C. (2004). Choosing the number of factors in independence factor analysis model, *Metodoloski zvezki*, **1**, 407–418.

[36] Wang, F. and Wall, M.M. (2003). Generalized common spatial factor model, *Biostatistics*, **4**, 569–582.