Extended Excess Hazard Model for Spatially Dependent Survival Data with Applications to Cancer Research

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

André Victor Ribeiro Amaral [1,*], Francisco Javier Rubio [2], and Paula Moraga [1]

[1] Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST).
[2] Department of Statistical Science, University College London (UCL).
[*] E-mail: andre.ribeiroamaral@kaust.edu.sa

GeoHealth

1 Introduction

In the relative survival framework, it is assumed an additive decomposition of the hazard function $h(\cdot)$ into two components, namely the hazard associated to other causes of death $h_O(\cdot)$, and the hazard associated to the study cancer $h_E(\cdot)$. That is, given covariates,

$$h(t; \mathbf{x}) = h_O(\text{age} + t) + h_E(t; \mathbf{x}), \ t \geq 0, \tag{1}$$

where "age" is the patient age when diagnosed, and $\mathbf{x}$ is a vector of risk factors. Also, as $h_O(\text{age} + t)$ is typically unknown in practice, it is common practice to approximate it by the population hazard $h_P(\text{age} + t; \mathbf{z})$, obtained from the life tables, such that $\mathbf{z} \subseteq \mathbf{x}$.

However, it is known that cancer registry data may present a spatially dependent structure, as patients from adjacent neighborhoods are likely to share environmental and social-economical factors.

2 Modeling Approach

Let t_{ij} be the observed survival times, such that $i = 1, \cdots, r$ indicates the region and $j = 1, \cdots, n_i$ denotes the individuals. Then,

$$h_E(t_{ij}; \mathbf{x}_{ij}, \tilde{u}_i, u_i) = h_0(t_{ij} \exp\{\tilde{\mathbf{x}}_{ij}^\top \boldsymbol{\alpha} + \tilde{u}_i\}) \exp\{\mathbf{x}_{ij}^\top \boldsymbol{\beta} + u_i\}, \tag{2}$$

where $h_0(\cdot)$ is the baseline function modeled in a parametric manner.

Model 2—Relative Survival Mixed-effects General Hazard (RS-MEGH) model—generalizes the following approaches

Table 1: Eight simpler models based on the Relative Survival Mixed-effects General Hazard (RS-MEGH) approach.

Name	Description	Name	Description
RS-MEGH-I	$\tilde{\mathbf{u}} = \mathbf{0}$	RS-GH	$\tilde{\mathbf{u}} = \mathbf{u} = \mathbf{0}$
RS-MEGH-II	$\tilde{\mathbf{u}} = \mathbf{u}$	RS-PH	$\tilde{\mathbf{u}} = \mathbf{u} = \mathbf{0}, \boldsymbol{\alpha} = \mathbf{0}$
RS-MEPH	$\tilde{\mathbf{u}} = \mathbf{0}, \boldsymbol{\alpha} = \mathbf{0}$	RS-AFT	$\tilde{\mathbf{u}} = \mathbf{u} = \mathbf{0}, \boldsymbol{\alpha} = \boldsymbol{\beta}$
RS-MEAFT	$\tilde{\mathbf{u}} = \mathbf{u}, \boldsymbol{\alpha} = \boldsymbol{\beta}$	RS-AH	$\tilde{\mathbf{u}} = \mathbf{u} = \mathbf{0}, \boldsymbol{\beta} = \mathbf{0}$

2.1 Spatial Model

We want to model $\tilde{\mathbf{u}}$ and $\mathbf{u}$ in such a way that these random effects smooth the data according to the observed neighborhood structure.

2.1.1 ICAR Model

For the Intrinsic Conditional Autoregressive (ICAR) model [1], the conditional distribution of u_i given u_j, such that $j \neq i$ (not. $\mathbf{u}_{-i}$), is

$$\pi(u_i | \mathbf{u}_{-i}) = \text{Normal}\left[\frac{\sum_{k \in \Lambda_i} u_k}{\lambda_i}, \frac{\sigma_i^2}{\lambda_i}\right],$$

where Λ_i and λ_i correspond to the neighbors and the number of neighbors of region i, respectively. The joint model can be written as

$$\pi(\mathbf{u}) \propto \tau_u^{\frac{r-1}{2}} \exp\left\{-\frac{\tau_u}{2} \sum_{i \sim j} (u_i - u_j)^2\right\}, \tag{3}$$

where τ_u is the precision parameter. From Equation (3), $\sum_{\text{all } i} u_i = 0$.

2.1.2 BYM2 Model

For the Besag-York-Mollié (BYM2) model [2], the unstructured and structured random effects can be written as

$$\mathbf{v} + \mathbf{u} = \sigma(\sqrt{1 - \rho}\mathbf{v}^\star + \sqrt{\rho}\mathbf{u}^\star),$$

where σ represents the overall standard deviation, $\rho \in [0, 1]$, $\mathbf{v}^\star \sim \text{Normal}(\mathbf{0}, I_r)$, and $\mathbf{u}^\star$ is the scaled ICAR model.

3 Estimation procedure

Let $\mathcal{D} = \{(t_{ij}, \delta_{ij}, \mathbf{x}_{ij}, \mathbf{z}_{ij}); i = 1, \cdots, r, \text{ and } j = 1, \cdots, n_i\}$ be the observed data. In that case, the likelihood function for the vector of unknown parameters $\boldsymbol{\xi}$ can be written as proportional to

$$\Pi_{i=1}^r \Pi_{j=1}^{n_i} \left\{h_P(\text{age}_{ij} + t_{ij}; \mathbf{z}_{ij}) + h_E(t_{ij}; \mathbf{x}_{ij}, \tilde{u}_i, u_i)\right\}^{\delta_{ij}} \times$$
$$\times \exp\{-H_E(t_{ij}; \mathbf{x}_{ij}, \tilde{u}_i, u_i)\},$$

where $h_P(\text{age}_{ij} + t_{ij}; \mathbf{z}_{ij})$ is previously obtained from the life tables.

4 Case Study

We will analyze a data set that contains survival information about $n = 1,043$ patients diagnosed with acute myeloid leukemia between 1982 and 1998 in the 24 areas in the North West region of the UK.

After fitting possible model combinations, the highest ranked approach was the RS-MEGH (LN ICAR). The results are based on it.

To compare the areas, we will compute the "net survival" as follows

$$S_N(t_i) = (n_i)^{-1} \sum_{j=1}^{n_i} \exp\left\{-\int_0^{t_{ij}} h_E(\ell; \mathbf{x}_{ij}, \tilde{u}_i, u_i) d\ell\right\}, \ \forall i.$$

Figure 1 shows the estimated net survival function $S_N(t_i)$, $\forall i$, where $t_i = 1$ (years), such that the data was stratified by deprivation level.

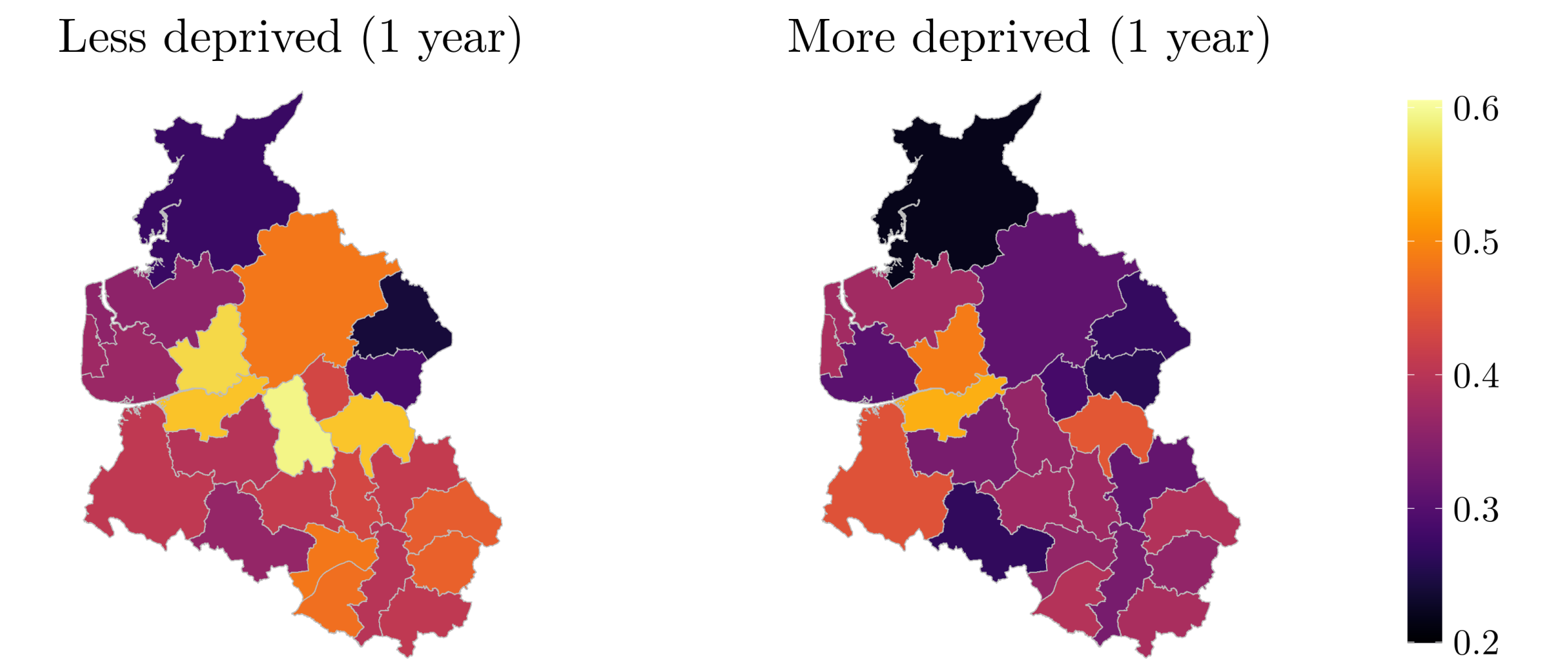Less deprived (1 year)   More deprived (1 year)

Figure 1: Estimated net survival function, after 1 year, for two groups of patients.

In conclusion, it seems that, based on the analyzed data, the patients' location contribute to their chances of surviving when diagnosed with leukemia. Therefore, modeling the spatially dependent random effects might be important to understand how these individuals are treated in different manners across the North West region of the UK.

References

[1] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, 1974.

[2] J. Besag, J. York, and A. Mollie. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20, 1991.