

Complex survival modeling with INLA

Janet van Niekerk

Håvard Rue, Divan Burger, Denis Rustand, Emmanuel Lesaffre, Elias Krainski and others

VII Workshop em Análise de Sobrevivência e Aplicações
October 2024

King Abdullah University of
Science and Technology



جامعة الملك عبد الله
للعلوم والتكنولوجيا





QR code for slides



https://github.com/JanetVN1201/WASA_2024



Outline I

- 1 Survival models and Bayesian inference
- 2 Survival models as Latent Gaussian models
- 3 The INLA methodology
- 4 AML survival (using Matern field)
- 5 Adherence study (Quantile joint model)

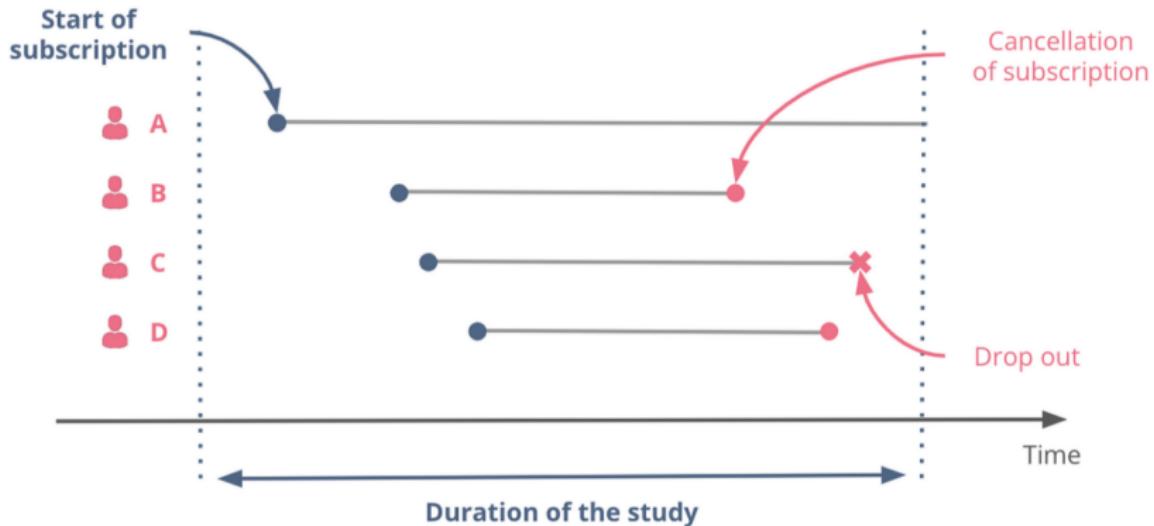


BayesComp group at KAUST





Time-to-event data



Hazard rate, Survival probabilities etc.



Bayesian inference

In GLM usually we model the mean with a linear model

$$E(y) = g(\beta_0 + \beta X)$$

But in survival analysis or reliability we usually model the hazard rate or survival function

$$h(t) = g(\beta_0 + \beta X)$$

or

$$S(t) = h(\beta_0 + \beta X)$$



Computational aspects

- Analytical methods - conjugacy (pre-computer era)
- Approximate methods - Laplace (can be inaccurate)
- Exact methods - MCMC (very slow for complex models or large data)

Now, due to computing resources approximate methods are gaining popularity - INLA, VB, EP etc.



What is INLA?

INLA - Integrated Nested Laplace Approximations

- Deterministic approximations instead of sampling
- LGM - Latent Gaussian models
- Three internal strategies - Gaussian, simplified Laplace, Laplace (pre 2021)
- R package "INLA"

Now there is a new default strategy combining Laplace approximations with Variational Bayes.¹ (2021+)

¹van Niekerk, J. and Rue, H., 2024. Low-rank variational Bayes correction to the Laplace method. *Journal of Machine Learning Research*, 25(62), pp.1-25.





INLA versus MCMC

For small models and data:

	INLA	MCMC/HMC
Method	Deterministic (Mathematical)	Sampling-based
Memory cost	Low	Low
Computational cost	Low	Low

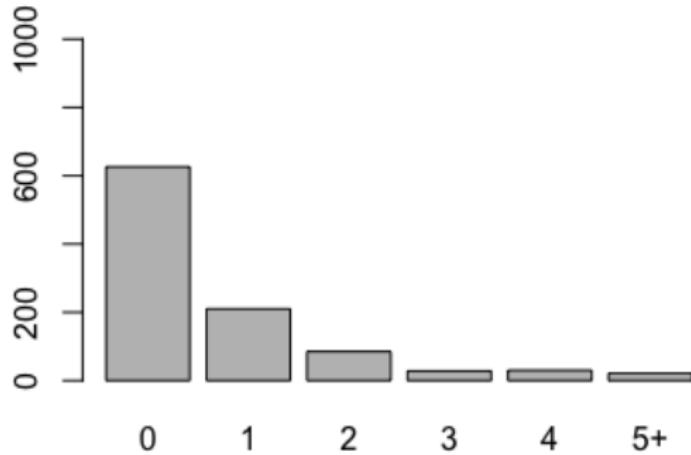
For large models and/or data:

	INLA	MCMC/HMC
Method	Deterministic (Mathematical)	Sampling-based
Memory cost	Low	High
Computational cost	Low	Very high



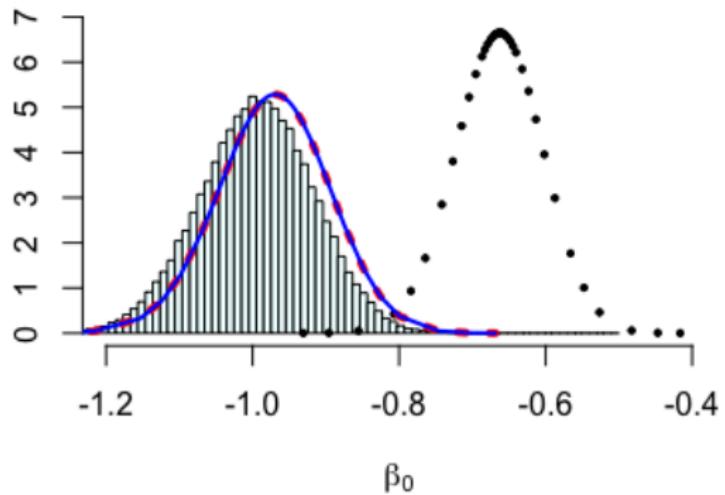
Example - INLA and MCMC I

$$y_i \sim \text{Poisson}(\exp(\eta_i)), \quad \eta_i = \beta_0 + \beta_1 x_i + u_i$$



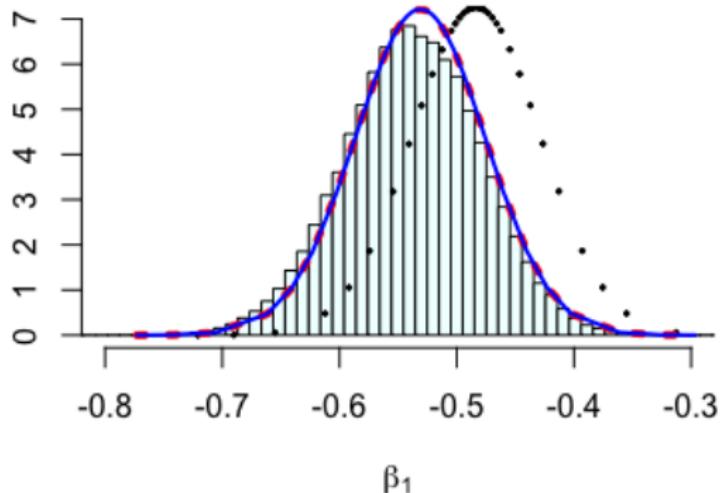


Example - INLA and MCMC II



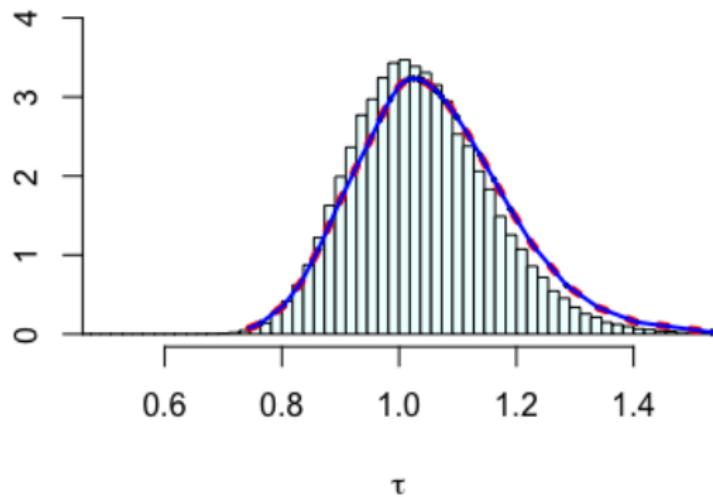


Example - INLA and MCMC III





Example - INLA and MCMC IV





Example - INLA and MCMC V

	INLA	MCMC
β_0	-0.972	-0.934
β_1	-0.531	-0.529
τ	1.056	1.037
Time(s)	5.718	207.445



Why is INLA so accurate and so fast?

- LGM structure
- Sparse precision matrix
- Specialized matrix algebra for sparse matrices
- NEW: VB (low-rank) correction [?]

Use precision matrix instead of covariance matrix → natural occurrence

But what about survival and reliability models?
Can we use INLA to perform Bayesian inference of such models?



Model definition - GAMM

Suppose we have response data $\mathbf{y}_{n \times 1}$ (conditionally independent) with density function $\pi(y|\mathbf{X}, \boldsymbol{\theta})$ and link function $h(\cdot)$, that is linked to some covariates \mathbf{Z} through linear predictors

$$\boldsymbol{\eta}_n = \beta_0 + \mathbf{Z}_\beta \boldsymbol{\beta} + \sum f^k(\mathbf{Z}_f) = \mathbf{A}\mathbf{X}$$

The inferential aim is to estimate the latent field $\mathbf{X}_m = \{\beta_0, \boldsymbol{\beta}, \mathbf{f}\}$, and $\boldsymbol{\theta}$.



GAMM → LGM

Assume

$$\boldsymbol{X}|\boldsymbol{\theta} \sim N(\boldsymbol{0}, \boldsymbol{Q}(\boldsymbol{\theta})^{-1})$$

where $\boldsymbol{Q}(\boldsymbol{\theta})$ is a sparse matrix (\boldsymbol{X} is a GMRF).

$p(\boldsymbol{X}, \boldsymbol{\theta}) = p(\boldsymbol{X}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ and $p(\boldsymbol{\theta})$ can be non-Gaussian.



LGM

INLA is designed to work for LGM's

- ① Data y_i with some likelihood $L(\mathbf{X}, \boldsymbol{\theta} | \mathbf{y}) = \prod_{i=1}^n f(y_i | \eta_i = h(\mathbf{X}, \boldsymbol{\theta}))$
- ② $\eta_i = h(\beta Z + \mathbf{u}(\boldsymbol{\theta}) \mathbf{A})$
- ③ $\mathbf{X} | \boldsymbol{\theta} \sim N(\mathbf{0}, \mathbf{Q}^{-1})$
- ④ $\boldsymbol{\theta} \sim \text{hyperprior}$



Survival models as LGM's

Censoring and Truncation affects the likelihood only. For censoring,

- ① Right : $L_i(t_i|d_i) = S_i(t_i)$
- ② Event : $L_i(t_i|d_i) = S_i(t_i)f_i(t_i)$
- ③ Left : $L_i(t_i|d_i) = 1 - S_i(t_i)$
- ④ Interval : $L_i(t_{1i}, t_{2i}|d_i) = S_i(t_{1i}) - S_i(t_{2i}),$

so then (still)

- ① Data t_i with some likelihood $L(\mathbf{X}, \boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^n L(t_i|\eta_i = h(\mathbf{X}, \boldsymbol{\theta}))$
- ② $\eta_i = h(\beta\mathbf{Z} + \mathbf{u}(\boldsymbol{\theta})\mathbf{A})$
- ③ $\mathbf{X}|\boldsymbol{\theta} \sim N(\mathbf{0}, \mathbf{Q}^{-1})$
- ④ $\boldsymbol{\theta} \sim \text{hyperprior}$



Possibilities

- Frailty models - unit-specific or clustered
- Spatial survival models - areal or continuous in space
- Nonlinear effects of covariates using splines
- Joint models with continuous or discrete longitudinal biomarker(s)
- Competing risks or multi-state models
- and many more...



Website for R-INLA library

<https://www.r-inla.org/>



Bayesian inference

Data \mathbf{y} (with covariates \mathbf{Z}), depend on \mathbf{X} and $\boldsymbol{\theta}$ such that, $E[Y] = h(\mathbf{A}(\mathbf{Z})\mathbf{X})$.

Bayes' theorem:

$$\begin{aligned} q(\mathbf{X}, \boldsymbol{\theta} | \mathbf{y}) &\propto L(\mathbf{y} | \mathbf{X}, \boldsymbol{\theta}) p(\mathbf{X}, \boldsymbol{\theta}) \\ \text{Posterior} &\propto \text{Likelihood} \times \text{Prior} \end{aligned}$$



Posterior approximations by INLA

$$\begin{aligned}\pi(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{y}) &= \pi(\boldsymbol{\theta})\pi(\boldsymbol{X}|\boldsymbol{\theta}) \prod_{i=1}^n \pi(y_i | (\boldsymbol{AX})_i, \boldsymbol{\theta}) \\ \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) &\propto \frac{\pi(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi_G(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{y})} \Big|_{\boldsymbol{X}=\boldsymbol{\mu}(\boldsymbol{\theta})} \\ \tilde{\pi}(\theta_j|\boldsymbol{y}) &= \int \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta}_{-j} \\ \tilde{\pi}(\boldsymbol{X}_j|\boldsymbol{y}) &= \int \tilde{\pi}(\boldsymbol{X}_j|\boldsymbol{\theta}, \boldsymbol{y}) \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta},\end{aligned}$$

$\tilde{\pi}(\boldsymbol{X}_j|\boldsymbol{\theta}, \boldsymbol{y})$ depends on the approximation used, for Gaussian it is straightforward for the Laplace approximation we do another Gaussian approximation to $\tilde{\pi}(\boldsymbol{X}_{-j}|\boldsymbol{\theta}, \boldsymbol{y})$.



Modern INLA

The Gaussian approximation $\pi_G(\mathbf{X}|\boldsymbol{\theta}, \mathbf{y})$ to $\pi(\mathbf{X}|\boldsymbol{\theta}, \mathbf{y})$ is calculated from a second order expansion of the likelihood around the mode of $\pi(\mathbf{X}|\boldsymbol{\theta}, \mathbf{y})$, $\mu(\boldsymbol{\theta})$ as follows

$$\begin{aligned}\log(\pi(\mathbf{X}|\boldsymbol{\theta}, \mathbf{y})) &\propto -\frac{1}{2}\mathbf{X}^\top \mathbf{Q}(\boldsymbol{\theta})\mathbf{X} + \sum_{i=1}^n \left(b_i(\mathbf{AX})_i - \frac{1}{2}c_i(\mathbf{AX})_i^2 \right) \\ &= -\frac{1}{2}\mathbf{X}^\top (\mathbf{Q}(\boldsymbol{\theta}) + \mathbf{A}^\top \mathbf{D}\mathbf{A})\mathbf{X} - \mathbf{b}^\top \mathbf{AX}\end{aligned}$$

where \mathbf{b} is an n -dimensional vector with entries $\{b_i\}$ and \mathbf{D} is a diagonal matrix with n entries $\{c_i\}$. Note that both \mathbf{b} and \mathbf{D} depend on $\boldsymbol{\theta}$, so the Gaussian approximation is for a fixed $\boldsymbol{\theta}$.



Modern INLA

The process is iterated to find \boldsymbol{b} and \boldsymbol{D} that gives the Gaussian approximation at the mode, $\mu(\boldsymbol{\theta})$, so that

$$\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{y} \sim N\left(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{Q}_{\boldsymbol{X}}^{-1}(\boldsymbol{\theta})\right).$$

The graph of the Gaussian approximation consists of two components,

- ① \mathcal{G}_p : the graph obtained from the prior of the latent field through $\boldsymbol{Q}(\boldsymbol{\theta})$
- ② \mathcal{G}_d : the graph obtained from the data based on the non-zero entries of $\boldsymbol{A}^\top \boldsymbol{A}$



Modern INLA

Next, the marginal conditional posteriors of the elements of \boldsymbol{X} is calculated from the joint Gaussian approximation as

$$\boldsymbol{X}_j | \boldsymbol{\theta}, \boldsymbol{y} \sim N \left((\boldsymbol{\mu}(\boldsymbol{\theta}))_j, (\boldsymbol{Q}_{\boldsymbol{X}}^{-1}(\boldsymbol{\theta}))_{jj} \right).$$

and the marginals

$$\tilde{\pi}(\boldsymbol{X}_j | \boldsymbol{y}) = \int \pi_G(\boldsymbol{X}_j | \boldsymbol{\theta}, \boldsymbol{y}) \tilde{\pi}(\boldsymbol{\theta} | \boldsymbol{y}) d\boldsymbol{\theta} \approx \sum_{k=1}^K \pi_G(\boldsymbol{X}_j | \boldsymbol{\theta}_k, \boldsymbol{y}) \tilde{\pi}(\boldsymbol{\theta}_k | \boldsymbol{y}) \delta_k.$$



Universal tools in R-INLA - also for survival models

- Model selection metrics - WAIC, DIC
- Cross validation (1 and group) and model-based clustering
- Prediction of unobserved areas or new profiles
- Mean or quantile models
- Joint models
- Multiple imputation
- Coregionalization models
- etc..... ask at <https://groups.google.com/g/r-inla-discussion-group?pli=1> or
e-mail help@r-inla.org



Introduction

In this example we are studying the spatial distribution of leukemia mortality to inform public health policies, to gain insights for unmeasured covariates²



²van Niekerk, J. and Rue, H., 2024. Low-rank variational Bayes correction to the Laplace



Model

In this example we are studying the spatial distribution of leukemia mortality to inform public health policies, to gain insights for unmeasured covariates.

$$\begin{aligned} h(t|\beta, \boldsymbol{u}) &= h_0(t) \exp(\eta(s)) \\ \eta_i(s) &= \beta_0 + \beta_1 \text{Age}_i + \beta_2 \text{WBC}_i + \beta_3 \text{TPI}_i + u(s), \end{aligned}$$

to account for spatial variation we use a Gaussian effect \boldsymbol{u} with a Matérn covariance structure with hyperparameters, marginal variance σ_u^2 and nominal range $r = 2/\kappa$.



Posterior inference - fixed effects

Time used:

Pre = 4.1, Running = 2.58, Post = 0.156, Total = 6.84

Fixed effects:

	mean	sd	0.025quant	0.5quant	0.975quant	mode	kld
(Intercept)	-2.170	0.206	-2.569	-2.173	-1.753	-2.172	0
sex	0.072	0.069	-0.063	0.072	0.208	0.072	0
age	0.033	0.002	0.029	0.033	0.037	0.033	0
wbc	0.003	0.000	0.002	0.003	0.004	0.003	0
tpi	0.025	0.010	0.005	0.025	0.044	0.025	0

Random effects:

Name	Model
spatial	SPDE2 model

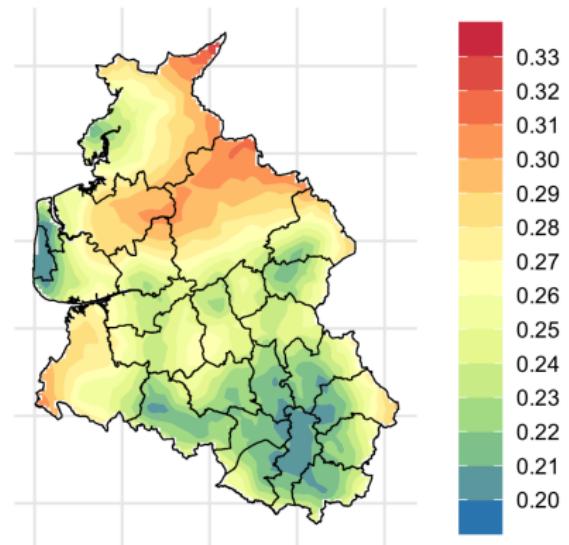
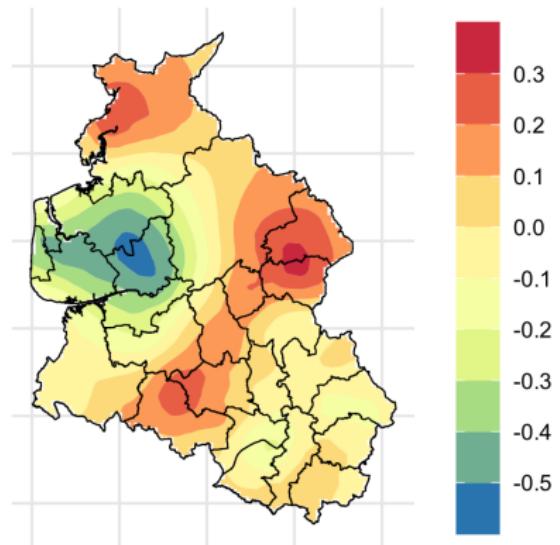
Model hyperparameters:

	mean	sd	0.025quant	0.5quant	0.975quant	mode
alpha parameter for weibullsurv	0.599	0.016	0.568	0.599	0.631	0.599
Range for spatial	0.310	0.156	0.114	0.276	0.709	0.220
Stdev for spatial	0.293	0.073	0.174	0.284	0.460	0.268

Marginal log-Likelihood: -839.92



Posterior inference - spatial field





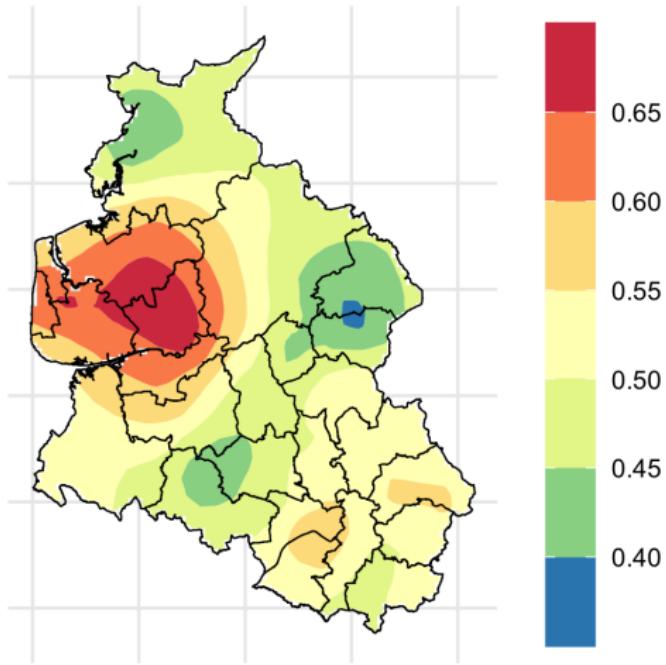
Posterior inference - with MCMC

	HMC	INLA
β_0	-2.189	-2.189
β_1	0.597	0.597
β_2	0.241	0.241
β_3	0.108	0.108
τ	0.340	0.340
σ_u	0.223	0.223
r	0.202	0.202
Time(s)	8214	26.3

Table: Posterior means from HMC and INLA

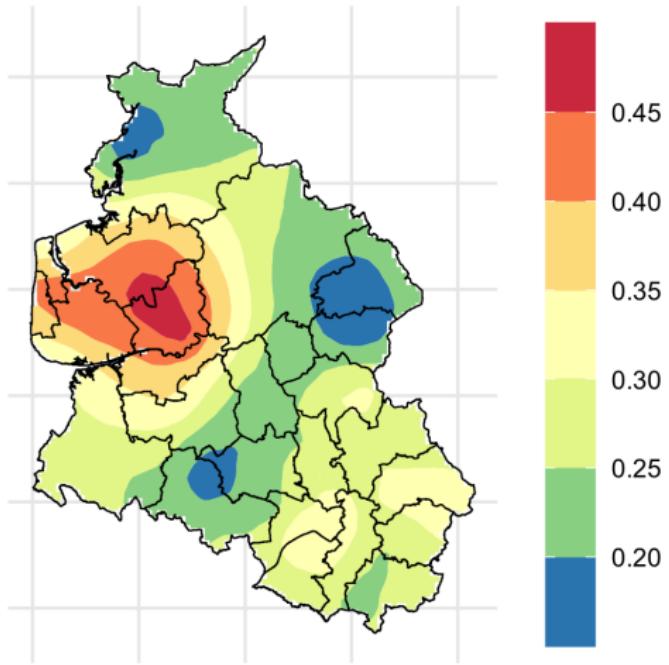
Survival probabilities

Survival function at year 1



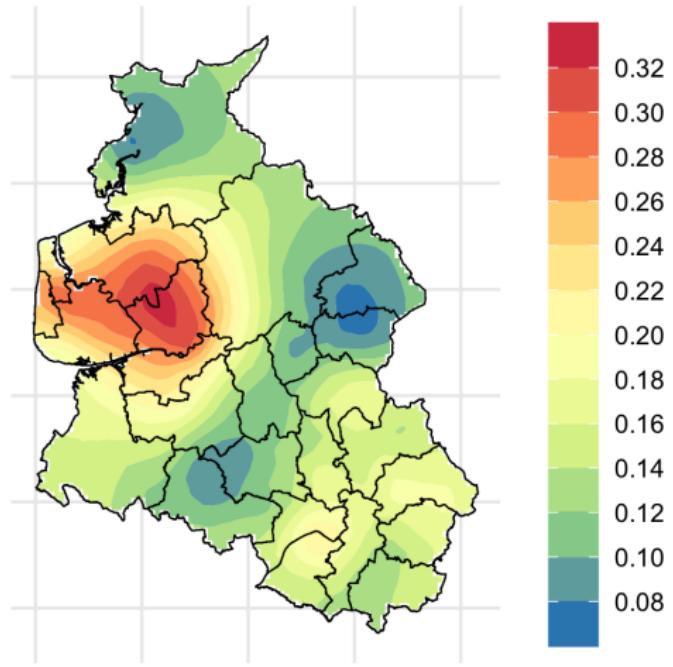
Survival probabilities

Survival function at year 2



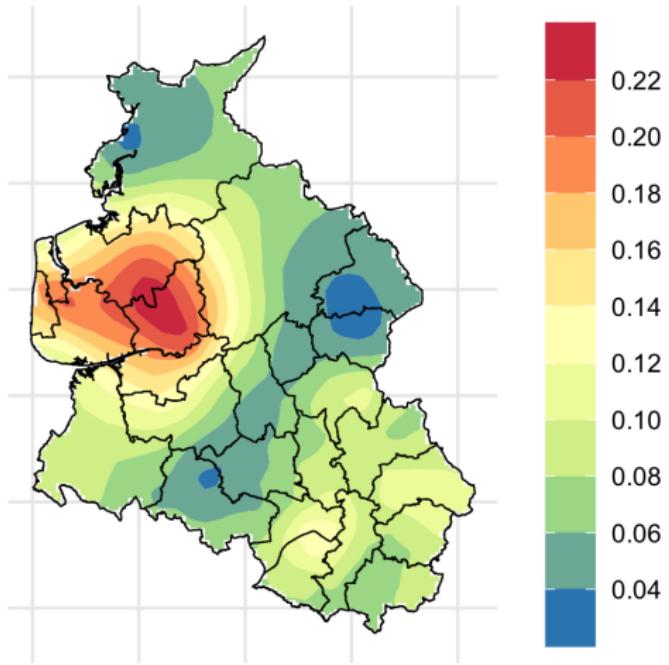
Survival probabilities

Survival function at year 3



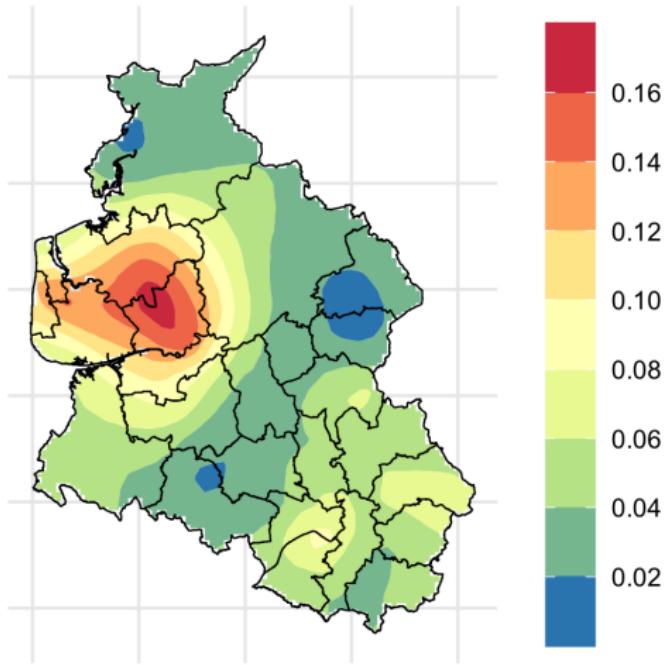
Survival probabilities

Survival function at year 4



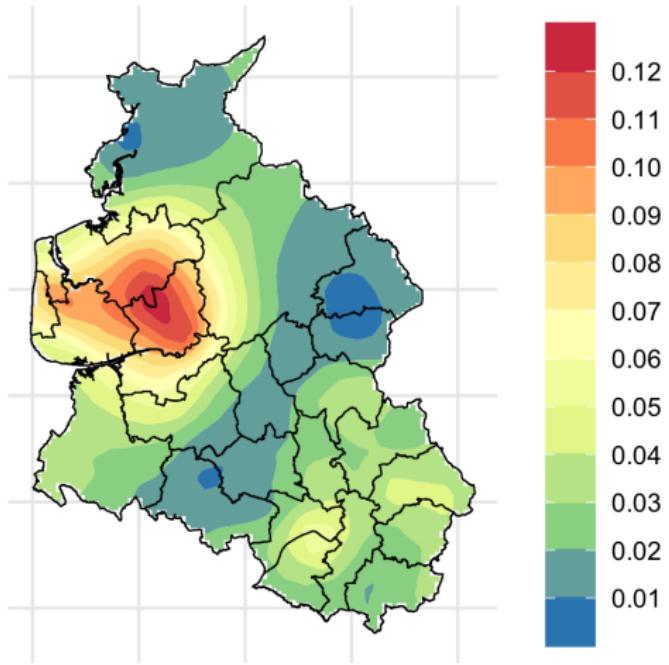
Survival probabilities

Survival function at year 5



Survival probabilities

Survival function at year 6



Introduction

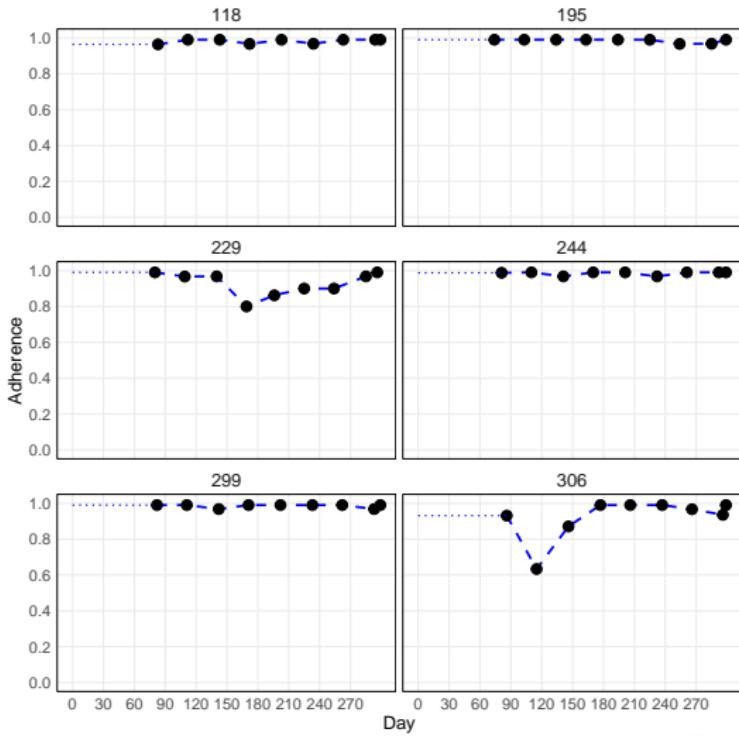
From the AARDEX Group - medication adherence and patient persistence is crucial for successful treatment regimes and regularized evaluation³.

Adherence is a proportion and persistence is defined based on a set of criteria unique to each drug.

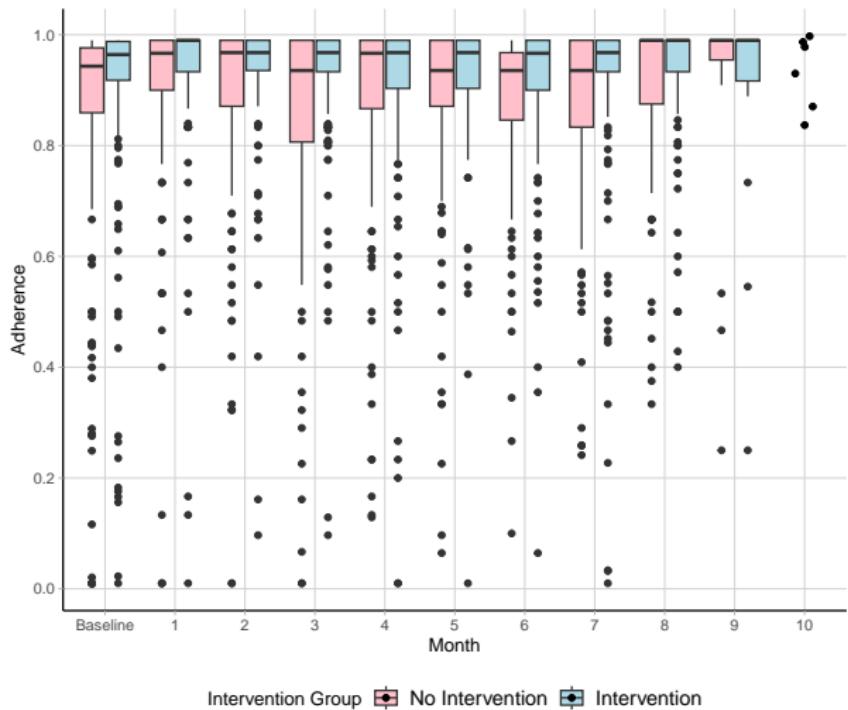
Research problem - outliers in the adherence bias the model, and can cause false positives for intervention.

³Burger, D.A., Van der Merwe, S., Van Niekerk, J., Lesaffre, E. and Pironet, A. Joint quantile regression of longitudinal continuous proportions and time-to-event data: application in medication adherence and persistence, *Statistical Methods in Medical Research*, Accepted

Dataset



Dataset



Model

Longitudinal model:

$$f(y_{ij}; \kappa_{q,i}(t_{ij}), \psi) = \alpha_{ij_1} \alpha_2 y_{ij}^{\alpha_{ij_1}-1} \left(1 - y_{ij}^{\alpha_{ij_1}}\right)^{\alpha_2-1},$$

$$\kappa_{q,i}(t_{ij}) = g^{-1}(\eta_{q,i}(t_{ij}))$$

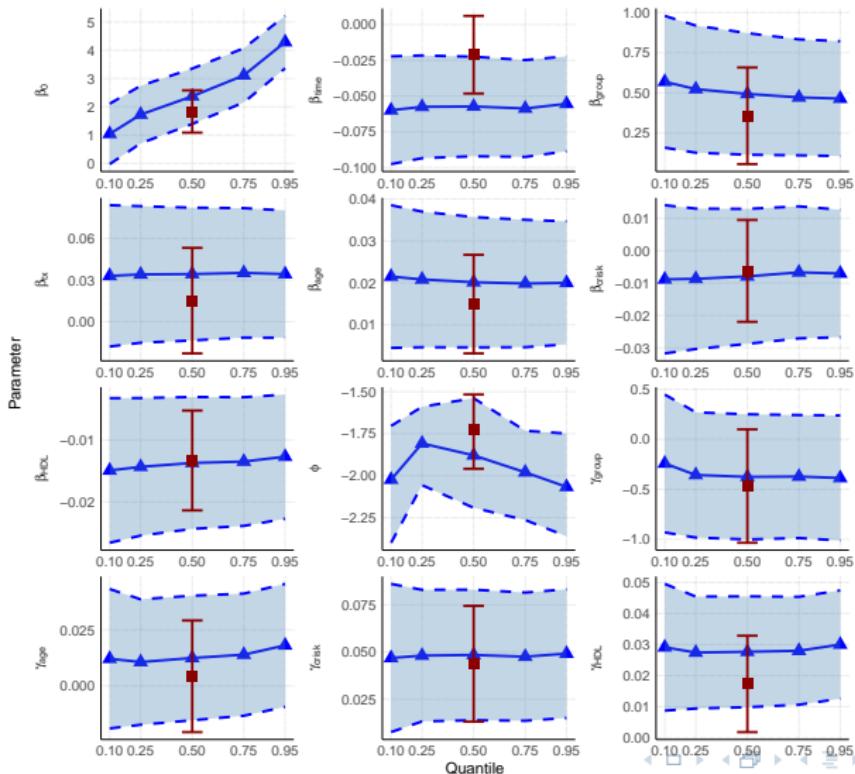
$$\eta_{q,i}(t_{ij}) = (\beta_0 + b_{0_i}) + (\beta_{\text{time}} + b_{\text{time}_i} + \mathbf{z}'_i \boldsymbol{\beta}_{\text{tx}}) t_{ij} + \mathbf{x}'_i \boldsymbol{\beta}_{\text{cov}}$$

with $\alpha_{ij_1} = \frac{\log\left(1-(1-q)^{\frac{1}{\alpha_2}}\right)}{\log(\kappa_{q,i}(t_{ij}))}$ and $\alpha_2 = \frac{\log(1-q)}{\log(1-e^{-\psi})}$.

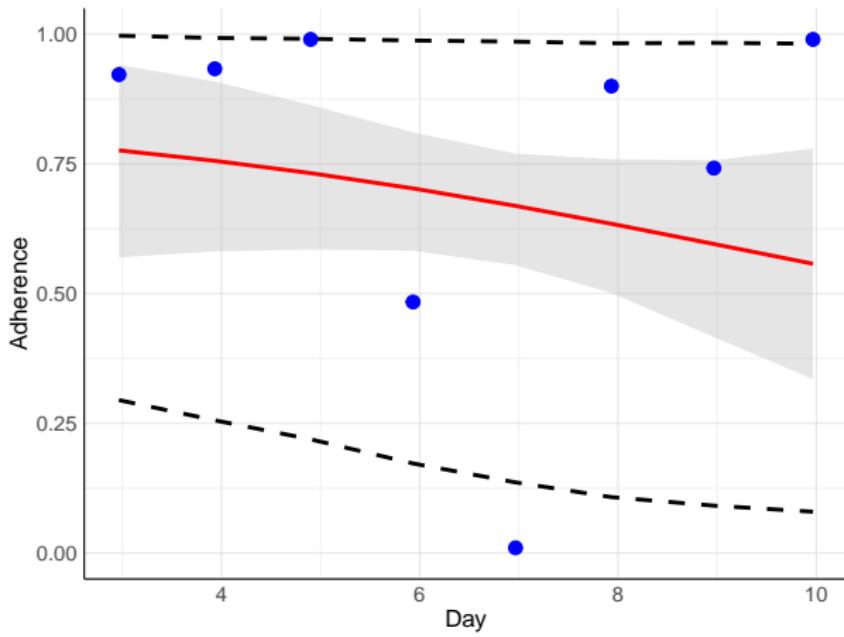
Survival model:

$$h_i(t) = h_0(t) \exp(\phi \eta_i(t) + \mathbf{w}'_i \boldsymbol{\gamma}).$$

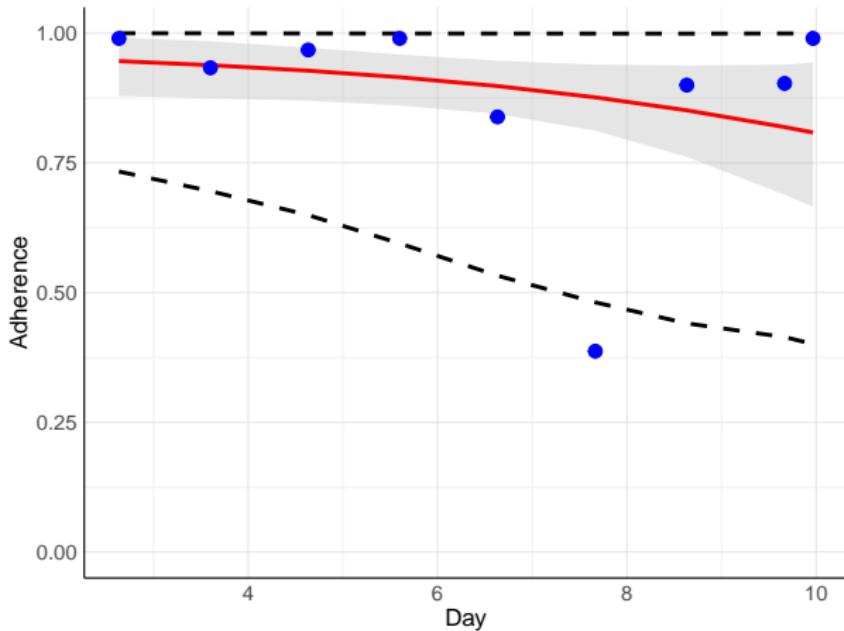
Posterior inference - fixed effects



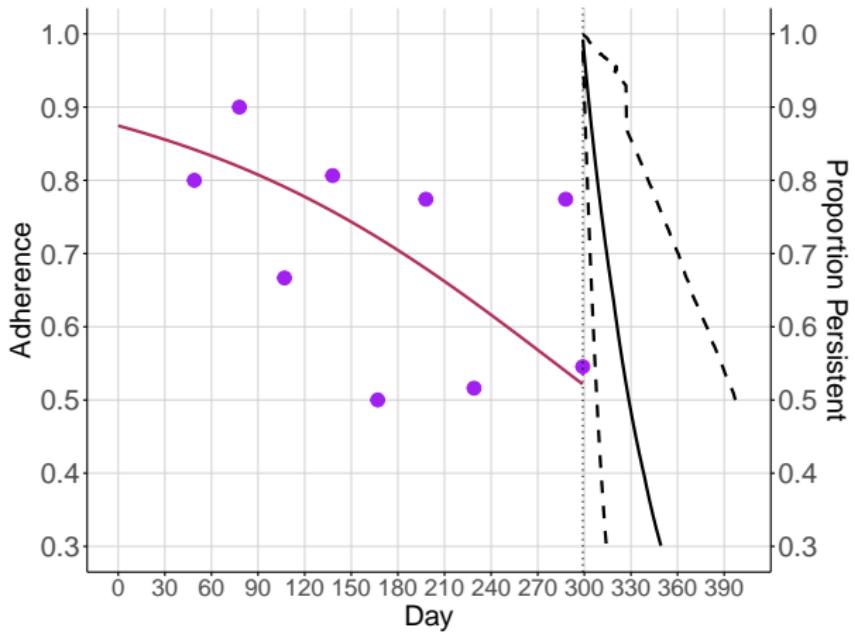
Posterior inference - Patient 46



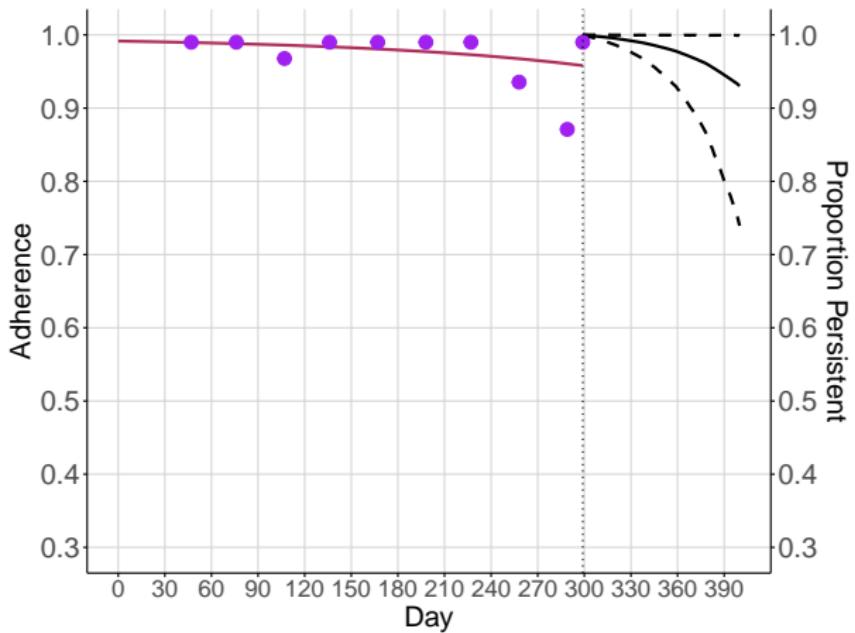
Posterior inference - Patient 56



Dynamic predictions - Patient 27



Dynamic predictions - Patient 10



Discussion

- Many survival analysis problems can be formulated as Latent Gaussian Models
- → INLA
- Analysis and dynamic predictions in near-real time
- → Personalized medicine

Many more exciting opportunities await...

شكراً • Thank you



جامعة الملك عبد الله
للعلوم والتكنولوجيا
King Abdullah University of
Science and Technology