XLV Jornadas de Economía de la Salud

Datos, evidencia, decisiones: generando valor para la gestión y las políticas sanitarias

Sevilla, 17 al 19 de junio de 2026

Cancer Mortality Atlas of Costa Rica, 2000–2023:

A Bayesian Spatial Estimation Approach

Kerry Loaiza Marín (UCR – BCCR – IU) (Corresponding Author)
Juan Rafael Vargas Brenes (UCR – CCP)
Jerlin Villalobos Ledezma (UCR – CCP)
Yanira Xirinachs Salazar (UCR – CCP)

XLV Health Economics Conference — Seville, June 2026

The results presented are solely the authors' own and do not represent the official position of any institution.

Motivation: the small-area estimation problem

Problem:
Crude SMR $= O_i/E_i \times 100$

- O_i: observed deaths in county i; E_i: expected deaths under national risk.
- In small-population counties, $E_i \approx 0$:
  - $O_i = 0 \Rightarrow$ SMR $= 0$ (false protection)
  - $O_i = 1 \Rightarrow$ SMR $\to \infty$ (false alarm)
- **Sampling noise $\neq$ real risk pattern.**

Solution: Methodological Contribution
Bayesian spatial smoothing (BYM, 1991):

- Neighbouring counties share information through a spatial prior.
- MCMC produces the full posterior distribution of risk.
- Smoothed SMRs with explicit uncertainty quantification.

Geographic scope:

- 82 counties of Costa Rica, 2000–2023
- 18 tumour sites (ICD-10 classification), year, sex, age, population
- Source: INEC Costa Rica

The Besag, York and Mollíe model (BYM, 1991)

Observation Model

Counts are Poisson: $O_i \mid R_i \sim \text{Poisson}(E_i R_i)$

Log-risk decomposition:

$$\log(R_i) = \underbrace{\mu}_{\text{global}} + \underbrace{\alpha_i}_{\text{spatial}} + \underbrace{\beta_i}_{\text{heterog.}}$$

- $\mu = \mathbf{X}_i\boldsymbol{\theta}$: global component (age, sex); diffuse prior $\boldsymbol{\theta} \sim \mathcal{N}(-\infty, +\infty)$.
- α_i: structured spatial effect (ICAR).
- β_i: unstructured (heterogeneous) effect.

Prior Structure

Structured spatial prior (ICAR):

$$\alpha_i \mid \boldsymbol{\alpha}_{-i}, \sigma^2 \sim \mathcal{N}\left(\bar{\alpha}_i, \frac{\sigma^2}{n_i}\right)$$

- $\bar{\alpha}_i$ = mean of α_j over neighbours of i.
- n_i = number of neighbours of county i.

Unstructured heterogeneity: $\beta_i \mid \tau^2 \sim \mathcal{N}(0, \tau^2)$

Hyperpriors: $\sigma^2, \tau^2 \sim \Gamma^{-1}(1, 0.01)$ (vague, non-informative)

Joint posterior distribution and MCMC rationale

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \tau^2 \mid \mathbf{O}) \propto \underbrace{\prod_{i=1}^{n} e^{-E_i e^{\log R_i}} \left(E_i e^{\log R_i}\right)^{O_i}}_{\text{Poisson likelihood}} \times \underbrace{(\sigma^2)^{-n/2} \exp\left\{-\tfrac{1}{2\sigma^2} \textstyle\sum_{i\sim j} (\alpha_i - \alpha_j)^2\right\}}_{\text{structured ICAR prior}}$$

$$\times \underbrace{(\tau^2)^{-n/2} \exp\left\{-\tfrac{1}{2\tau^2} \textstyle\sum_i \beta_i^2\right\}}_{\text{heterogeneous prior}} \times \underbrace{\pi(\sigma^2)\,\pi(\tau^2)}_{\Gamma^{-1}(1,\,0.01)}$$

- $i \sim j$: contiguous county pair (defined by Queen type adjacency matrix $\mathbf{W}$, i.e. $w_{ij} = 1$ if i, j share a border or vertex, 0 otherwise).
- $\sum_{i\sim j}(\alpha_i - \alpha_j)^2$: **penalises** abrupt spatial jumps.

Why MCMC?

- Poisson $\times$ ICAR $\times$ Normal is **not** conjugate $\Rightarrow$ no closed-form posterior.
- We use **Metropolis-within-Gibbs** with 110,000 iterations (10,000 burn-in).

MCMC: Metropolis-within-Gibbs algorithm

1. Initialise $\boldsymbol{\Theta}^{(0)} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \tau^2)$ arbitrarily; fix random seed.
2. **For** $t = 1, \ldots, 110{,}000$:
   - **Block 1 – $\boldsymbol{\alpha}$ (ICAR) [MH]:** propose $\alpha_i^\star \sim \mathcal{N}(\alpha_i, \delta_\alpha^2)$; accept with $\rho_i = \min(1, p(\alpha_i^\star|\cdot)/p(\alpha_i|\cdot))$.
   - **Block 2 – $\boldsymbol{\beta}$ (heterog.) [MH]:** same random-walk proposal.
   - **Block 3 – $\boldsymbol{\theta}$ (covariates) [MH]:** random-walk proposal.
   - **Block 4 – σ^2, τ^2 [Gibbs]:** conjugate $\Gamma^{-1} \Rightarrow$ direct sampling.
3. Discard first 10,000 draws (burn-in); retain $M = 100{,}000$.

- Proposal variance δ^2 tuned adaptively to keep acceptance rates in the optimal **20%–65%** window.
- The ICAR MH ratio combines a Poisson likelihood factor and a spatial penalty factor.
- Convergence checked via **Geweke Z-test**.

From MCMC to SMRs: posterior inference

Posterior Quantities:
MCMC yields $M = 100{,}000$ draws of expected deaths: $\tilde{E}_i^{(m)} = E_i\, e^{\hat{R}_i^{(m)}}$

- **Point estimate (posterior mean):** $\text{SMRs}_i = 100 \cdot O_i / \bar{\tilde{E}}_i$
- **90% credible interval** (quantiles inverted since SMRs $\propto 1/\tilde{E}_i$):
  - Lower: $100 \cdot O_i \,/\, q_{95\%}(\tilde{E}_i)$
  - Upper: $100 \cdot O_i \,/\, q_{5\%}(\tilde{E}_i)$

Posterior Excess-Risk Probability:

$$\Pr(\text{SMRs}_i > 100) = \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\left(\frac{O_i}{\tilde{E}_i^{(m)}} > 1 \right)$$

- ≈ 1: strong evidence of excess risk.
- ≈ 0.5: ambiguous evidence.
- ≈ 0: evidence of below-average risk.

Joint interpretation is mandatory: same SMRs can imply very different statistical certainty.

County	SMRs	$\Pr > 100$
San Pablo	158.9	0.98
County X	158.9	0.50

Selected results: notable territorial patterns

County	Cancer site	Smoothed SMRs	$\Pr(\text{SMRs}>100)$
Montes de Oro	Melanoma/skin	160.4	0.99
Flores	Male genital	310.8	1.00
Turrubares	Breast	180.7	0.99
Dota	Prostate	125.8	0.95
San Mateo	Colon	169.9	0.99

Territorial heterogeneity:

- Substantial variation across 82 counties by tumour site.
- **Chorotega region:** cluster in ill-defined sites and soft tissues.
- **Central Pacific:** pattern in urinary tract and endocrine glands.

Value of Bayesian smoothing:

- Smoothing recovers genuine patterns where the crude SMR would be unreliable.
- San Mateo (15 colon cases): SMRs stabilised using neighbours San Ramón and Esparza.
- Older ages are associated with larger risks overall. Sex matters too.

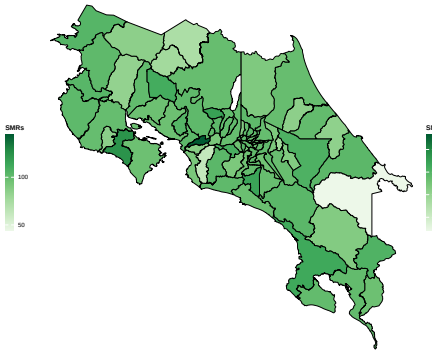
Jornadas de Economía de la Salud
Datos, evidencia, decisiones: generando valor para la gestión y las políticas sanitarias
Sevilla, 17 al 19 de junio de 2026

Conclusions

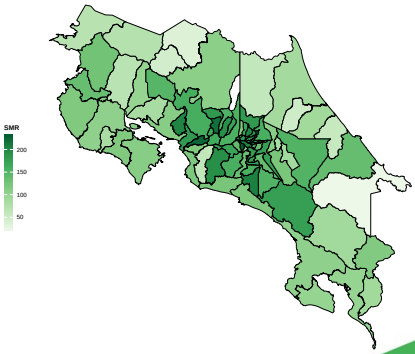
Contributions:

- The model separates **genuine spatial dependence** and resolves the small-area estimation problem.
- $\Pr(\text{SMRs}_i > 100)$ provides a **directly interpretable** uncertainty measure for decision-makers.

Smooth SMR



Non-smooth SMR



Thank you! Comments and questions are welcome.

AES Asociación de Economía de la Salud

MCMC convergence diagnostics and goodness of fit

- **Geweke test:** compares mean of first 10% vs last 50%; $Z \in [-1.96, 1.96]$ indicates convergence.
- **DIC:** DIC (Deviance Information Criterion) $= \overline{D(\boldsymbol{\Theta})} + p_D$ (p_D: effective complexity). Lower = better.
- **LMPL (Logarithmic Pseudo-Marginal Likelihood):** sum of conditional predictive log-densities. Higher = better.
- **WAIC (Watanabe-Akaike Information Criterion):** Bayesian counterpart of AIC. Lower = better.
- **Moran Index Test:** evaluates whether there is spacial autocorrelation in the residuals. -1 is perfect dispersion, $+1$ is perfect clustering, 0 indicates a completely random spatial distribution.

Robustness Checks

- **Non-smooth rates:** comparison with non-smooth Standard Mortality Rates. non-smooth rates contain more extreme values and moslty double the values of the smooth rates.
- **Estimation of non-spacial model:** comparison with estimates from a Poisson model without spacial component. Non-spacial estimates suffer from poor fit and non-convergence issues. Their residuals contain spacial autocorrelation.
- **Changes in number of iterations:** models with post burn-in 10,000 or 1,000,000 iterations perform similarly. Best estimation in terms of fit and computational efficiency contains 100,000 iterations.
- **MH acceptance rate:** optimal range **20%–65%**. Large $\delta \rightarrow$ low acceptance; small $\delta \rightarrow$ slow exploration. Not applicable to variances σ^2 and τ^2.
- **Traceplot:** visual stationarity check post burn-in.

Working Paper

References

Besag, J., York, J., & Mollíe, A. (1991). *Bayesian Image Restoration, with Two Applications in Spatial Statistics. Annals of the Institute of Statistical Mathematics*, 43, 1–20.

Martínez-Beneito, M. A., López-Quílez, A., & Botella-Rocamora, P. (2008). *An autoregressive approach to spatio-temporal disease mapping. Statistics in Medicine*, 27, 2874–2889.

Eurostat (2013). *Revision of the European Standard Population*. Publications Office of the European Union, Luxembourg.

Corpas Burgos, F., et al. (n.d.). *National Cancer Mortality Atlas of Spain (ANDEES)*.

Lee, D. (2013). *CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors. Journal of Statistical Software*, 55(13), 1–24.