

Practice of Epidemiology

Estimating Country-Specific Incidence Rates of Rare Cancers: Comparative Performance Analysis of Modeling Approaches Using European Cancer Registry Data

Diego Salmerón, Laura Botta, José Miguel Martínez, Annalisa Trama, Gemma Gatta, Josep M. Borràs, Riccardo Capocaccia, and Ramon Clèries* for the Information Network on Rare Cancers (RARECARENet) Working Group

* Correspondence to Dr. Ramon Clèries, Cancer Plan, Institut d'Investigació Biomèdica de Bellvitge (IDIBELL), Catalan Institute of Oncology, Avenida Gran Vía 199-203, 08908 Hospitalet de Llobregat, Spain (e-mail: r.cleries@iconcologia.net).

Initially submitted August 19, 2021; accepted for publication September 9, 2021.

Estimating incidence of rare cancers is challenging for exceptionally rare entities and in small populations. In a previous study, investigators in the Information Network on Rare Cancers (RARECARENet) provided Bayesian estimates of expected numbers of rare cancers and 95% credible intervals for 27 European countries, using data collected by population-based cancer registries. In that study, slightly different results were found by implementing a Poisson model in integrated nested Laplace approximation/WinBUGS platforms. In this study, we assessed the performance of a Poisson modeling approach for estimating rare cancer incidence rates, oscillating around an overall European average and using small-count data in different scenarios/computational platforms. First, we compared the performance of frequentist, empirical Bayes, and Bayesian approaches for providing 95% confidence/credible intervals for the expected rates in each country. Second, we carried out an empirical study using 190 rare cancers to assess different lower/upper bounds of a uniform prior distribution for the standard deviation of the random effects. For obtaining a reliable measure of variability for country-specific incidence rates, our results suggest the suitability of using 1 as the lower bound for that prior distribution and selecting the random-effects model through an averaged indicator derived from 2 Bayesian model selection criteria: the deviance information criterion and the Watanabe-Akaike information criterion.

credible interval; Poisson regression; random effects; rare events; uniform prior

Abbreviations: DIC, deviance information criterion; INLA, integrated nested Laplace approximation; IR, incidence rate; LB, lower bound; MCMC, Markov chain Monte Carlo; RARECARENet, Information Network on Rare Cancers; RC, rare cancers; RMSE, root mean squared error; UB, upper bound; WAIC, Watanabe-Akaike information criterion.

Editor's note: An invited commentary on this article appears on page 499, and the authors' response appears on page 503.

Estimation of epidemiologic indicators of incidence, survival, and prevalence for rare cancers (RCs) is challenging, particularly for exceptionally rare entities and in countries with small populations. Because of the small-case-count data, the pros and cons of directly providing estimates based on very unstable empirical data or derived from modeling

approaches are unclear, inherently less intuitive, more complex, and to some extent dependent on subjective choices.

Investigators in the Information Network on Rare Cancers (RARECARENet) (1–3) calculated incidence, prevalence, and survival estimates for an operative list of 190 RC entities, defined as cancers with incidence rates (IRs) less than 6 cases per 100,000 person-years, for the period 2000–2007 (3). Since many countries could only provide data for cancer cases in some regional areas, not at a national level, RARECARENet investigators estimated country-specific IRs (number of cases per 100,000 person-years) for RCs on the basis of cases recorded by 83 population-based

cancer registries across 27 European countries (3). More recently, in the context of the Joint Action on Rare Cancers, which is generating policy recommendations on RC that can be implemented by European Union member states (1–3), the burden of RC incidence counts in Europe was compared with another burden derived by a model-based approach that used a simple Poisson random-effects model under the Bayesian framework (4), through integrated nested Laplace approximations (INLAs) implemented in the INLA platform (5). Evidence suggests that INLA is appropriate for estimating the distribution of fixed-effect parameters, but it could fail to yield good estimates in a random-effects model (6). This last shortcoming could be related to the numerical method used in INLA to estimate the posterior distribution: Laplace approximation (4, 6). This approximation is good for models close to a Gaussian distribution, but this may underestimate the variance of the random effects when modeling Poisson or binary data (4, 6). In that situation, reasonable approximations to the posterior are achieved when modeling high counts (4) or by altering INLA's default settings (6).

A previous study showed very small differences in the precision of certain indicators when implementing a simple Poisson random-effects model with different platforms and approaches (4); the authors called for a simulation study to properly analyze the differences in the context of RCs. These 190 entities represent different specific cancers, and their distribution may vary across countries; however, it is not feasible to use a different statistical model for each. We therefore aim to propose an approach that: 1) provides an overall better model fit for all of the entities; 2) shows the implication of using the Bayesian approach for deriving the corresponding 95% credible intervals; and 3) proposes prior distributions for the precision of random effects in specific situations related to small case counts.

METHODS

We performed 2 studies. We designed study 1 as a simulation, comparing 4 RC scenarios to assess the performance of the frequentist, empirical Bayes, and Bayesian approaches in the presence of overdispersion. Study 2 assessed the choice of appropriate lower and upper bounds for a uniform prior distribution on the standard deviation, σ , of the random-effects model, depending on the IRs from 2000–2007. The accompanying Web material (available at <https://doi.org/10.1093/aje/kwab262>) provides technical details, R (version 3.5.1; R Foundation for Statistical Computing, Vienna, Austria) (7) and WinBUGS (MRC Biostatistics Unit, Cambridge, United Kingdom, and Imperial College School of Medicine, London, United Kingdom) (8) software code, and additional results.

The Poisson random-effects model

In the 1980s, a relative risk regression model was proposed, assuming that the observed cases C_i in the i th area (in our study country) were distributed according to a Poisson

distribution (9), such that

$$C_i \sim \text{Poisson}(R_i E_i), i = 1, \dots, I = 27, \quad (1)$$

where R_i is a relative risk or standardized incidence ratio and $R_i = E(C_i)/E_i$, where $E(C_i) = \lambda_i$ and E_i are, respectively, the expected rate for C_i and the expected number of cases calculated from internal standardization or from an external source of age- and sex-specific rates. For each RC considered, the observed number of cases for the i th area/country covered by the cancer registries was modeled as in equation 1, but accounting for extra Poisson variability through a Poisson random-effects model presenting a hierarchical structure (9–15):

$$\begin{aligned} \theta_i &= \log(R_i) = \mu + v_i, i = 1, \dots, I = 27, \\ v_i &\sim N(0, \tau), \end{aligned} \quad (2)$$

where the parameter μ is the unknown intercept and v_i are the random effects representing the unstructured residual for each country, assuming a Gaussian distribution with mean 0 and precision τ , whereas θ_i is the log relative risk, that is, $R_i = e^{\theta_i}$. In this approach, also known as the exchangeable random intercept approach, the distribution of the random effects allows for extra-Poisson variability in the marginal distribution of the C_i 's.

Once the model parameters were estimated, the expectation of the predicted number of incident cases for the k th country was calculated through $\hat{\lambda}_k = \hat{R}_k E_k$, where \hat{R}_k is an estimate of R_k .

Study 1: simulation study

To evaluate the performance of the different approaches, we simulated data to establish a realistic ground truth for disease risk variation based on the validated and published data used in the RARECARENet project (1–4). Four cancer entities were selected from the RARECARENet database, and for each cancer entity, we simulated $J = 1,000$ data sets, those derived from fitting the original data set to model 2 (equation 2) through WinBUGS. Each simulated data set included $I = 27$ predicted values (one for each country) C_i^j , where $j = 1, \dots, J$ and $i = 1, \dots, I$. We considered it realistic to include the 27 European Union member states in the study, since the aggregation of observed cases provided by the cancer registries within each country was previously validated to derive the most recent indicators: the country-specific number of RCs and their corresponding rates (1–4). In addition, in the context of RCs, an outcome of interest is to estimate the variability of the country-specific counts (4).

Table 1 shows the distribution of the original numbers of cases for the selected cancers: 2 with relatively large case counts (tumors of the central nervous system and adenocarcinoma with variants of ovary) and 2 with smaller counts (adenocarcinoma with variants of middle ear and adenocarcinoma with variants of trachea).

In order to simulate data sets similar to the original data, the estimates of the model 2 (equation 2) parameters were used as the true values. Specifically, for each entity, if μ_0

Table 1. Numbers of Observed Cases of 4 Selected Cancers Obtained From the RARECARENet Database^a During 2000–2007

Cancer Entity	No. of Observed Cases		
	Range	Median	IR ^b
Tumors of the CNS ^c	174–28,732	2,758.0	73.800
AC with variants of ovary ^c	96–23,957	3,446.0	59.700
AC with variants of middle ear	0–17	1.8	0.003
AC with variants of trachea	0–39	6.0	0.011

Abbreviations: AC, adenocarcinoma; CNS, central nervous system; IR, incidence rate; RARECARENet, Information Network on Rare Cancers.

^a Numbers of cases were provided by 83 population-based cancer registries from 27 European countries (3).

^b Number of cases per 100,000 person-years, age-standardized to the European standard population (see Botta et al. (4)).

^c This cancer site was not considered a rare entity (large number of cases).

and τ_0 were the posterior estimates for μ and τ given the original data used in the RARECARENet project (obtained using the WinBUGS code of Web Appendix 1), where $\sigma \sim \text{Uniform}(\text{LB}, \text{UB})$ with lower bound (LB) = 0 and upper bound (UB) = 500 and $\tau_0 = \frac{1}{(\sigma^2)}$, then for $j = 1, \dots, J = 1,000$, and for $i = 1, \dots, I = 27$, we simulated $v_i^j \sim N(0, \tau_0)$, we computed $R_i^j = \exp(\mu_0 + v_i^j)$, and we generated the “new” observed cases $C_i^j \sim \text{Poisson}(R_i^j E_i)$. The simulated distribution of the “new” observed cases for each country is graphically depicted in Web Appendix 2 and Web Figures 1–4.

We compared the performance of 13 computational approaches (see Web Appendices 3–5) for estimating the mean number of cases and their confidence/credible intervals by using maximum likelihood approaches, empirical Bayes approaches, and modeling (12–22). The Bayesian models were fitted using WinBUGS and INLA (see Web Appendices 1, 4, and 5 for implementation and Web Appendices 3, 6, and 7 for details and computational issues (23–27)). There were 7 WinBUGS models fitted by assuming that σ follows a uniform prior with different LBs and UBs (see their code implementation in Web Appendix 1): In model 1, LB = 0 and UB = 500; in model 2, LB = 0.1 and UB = 500; in model 3, LB = 0.2 and UB = 500; in model 4, LB = 0.3 and UB = 500; in model 5, LB = 0.5 and UB = 500; in model 6, LB = 1 and UB = 500; and in model 7, there was sampling from LB and UB (see Web Appendix 5) such that $\text{LB} \sim \text{Uniform}(0, 5)$ and $\text{UB} \sim \text{Uniform}(\text{LB}, 500)$. INLA models were fitted by assuming 2 priors on the precision and standard deviation of the random effects: model 8, in which τ followed a γ prior with $\alpha = \beta = 0.000001$ (the INLA model used by Botta et

al. (4)) and model 9 (see Web Appendix 4), where σ had a uniform prior distribution between 0 and ∞ .

These 9 models were also compared with another 4 approaches: model 10—an empirical Bayes approach assuming a gamma distribution (14–16) for R_i^j ; model 11—a generalized linear mixed model estimated by maximum likelihood and 2 standard approaches not requiring modeling; model 12—exact Poisson distribution; and model 13—Byar’s Poisson approximation (11, 13). The last 2 approaches were used to compare the coverage and width of their 95% confidence intervals with the intervals obtained using modeling.

For each cancer entity and simulated data set, we calculated the 95% confidence interval or credible interval for the posterior number of cases in each country, $\hat{\lambda}_i^j = \hat{R}_i^j E_i$, measuring 3 indicators: 1) width, 2) coverage, and 3) the root mean squared error (RMSE) between λ_i and $\hat{\lambda}_i^j = \hat{R}_i^j E_i$ (assuming that R_i was the “real” value for the relative risk in the i th country, whereas \hat{R}_i^j was its estimate).

Bayesian and non-Bayesian approaches were assessed in terms of confidence and credible intervals by comparing the performance of these intervals through the percentage of their respective coverage of the true λ_i parameter across the 1,000 simulated data sets for each of the 4 cancer entities considered. However, one has to distinguish the interpretation of confidence and credible intervals, since they are conceptually different. The 95% confidence intervals refer to how often these intervals, computed from repetitions of the experiment under study, would contain the true parameter—considered as fixed—if model assumptions were valid (28). On the other hand, if one is interested in computing an interval with a 95% probability of containing a model parameter, considered as random, where each numerical value contained within this interval has its own probability mass, then the resulting interval is the Bayesian credible interval (18). For credible intervals, the achievement of the frequentist coverage when replicating a study is a desirable property for the prior’s assessment, in order to yield reliable posterior inference (29). It also guarantees a frequentist validity of these intervals when compared with confidence intervals (29). An adequate confidence/credible interval is expected to have coverage values of at least its nominal value, and if we calculate the average width of these intervals for all data sets, the one producing the narrowest width is preferred (29).

In addition to these 3 indicators, the effective number of model parameters (pD), the mean deviance, and the deviance information criterion (DIC) were also calculated (23) and compared between Bayesian models (see Web Appendices 6 and 7).

Study 2: applied study

In study 2, we compared the performance of models 1–7 across the 190 cancer entities by assessing the bounds of the uniform prior distribution in WinBUGS. For each cancer entity, we calculated 1) the average width of the 95% credible intervals of the 27 posterior estimates of λ_i , 2) the expected value of the posterior σ , and 3) common

measures used by data analysts for assessing model fit under the Bayesian framework: 3) the DIC and 2 versions of the Watanabe-Akaike information criterion (WAIC) (30–32), designated 4) $WAIC_1$ and 5) $WAIC_2$ (see Web Appendices 8–10).

We compared these indicators between models in each cancer entity as follows. First, we determined the minimum value of the average width, the posterior σ , $WAIC_1$, $WAIC_2$, and DIC across models. Second, for each model, we calculated the difference between these 5 indicators and their corresponding minimum. We stored these 5 indicators in a matrix of 190 rows, one for each cancer site selected from the RARECARENet database, and 40 columns (5 indicators \times 8 models considered). We summarized the results (median and 2.5th and 97.5th percentiles of each difference) for a specific selection of cancer entities in order to assess model performance.

The selection was based on the magnitude of the IRs (number of cases per 100,000 person-years), since these showed a minimum of 0.0004 and a maximum of 5.9692, with a median of 0.1100. We chose 4 scenarios: scenario A, where the IR was less than 0.03 (quartile 1 of the IRs (47/190 cancer entities)); scenario B, where the IR was less than 0.12 (up to quartile 2 of the IRs (97/190 entities)); scenario C, where the IR was less than 0.5 (up to quartile 3 of the IRs (143/190 entities)); and scenario D, where the IR was greater than or equal to 0.5 (quartile 4 of the IRs).

Furthermore, to suggest a strategy for different scenarios, we established a ranking for each model and cancer entity by calculating the indicators and then averaging the rankings of indicators for each model in the corresponding scenario (see Web Appendix 11, where Web Figure 5 depicts this procedure).

Finally, we made a graphical comparison between the best model derived from study 2 and model 8.

RESULTS

Study 1

Table 2 presents modeling indicators for cancer sites with large case counts. For tumors of the central nervous system, we noted that 1) the mean coverage for the generalized linear mixed model and the empirical Bayes strategies barely reached 95% and 2) the 2 nonmodeling strategies, exact and Byar's Poisson approximation, yielded wide confidence intervals. This phenomenon was also detected for an entity with large variability and relatively large counts: adenocarcinoma with variants of ovary. The Bayesian models showed an average of almost 95% coverage, with similar widths and RMSEs.

The indicators for cancer entities with very small counts and a large number of zeros are presented in Table 3. The exact Poisson and Byar's Poisson strategies produced large confidence intervals. WinBUGS and INLA strategies using a uniform prior on σ showed good performance compared with the aforementioned strategies, with narrower intervals and RMSEs, and coverage above 96%. Among WinBUGS models, we noted that model 6 showed the largest

width (3.61 cases) and RMSE (RMSE = 0.62) and model 1 the smallest values. On the other hand, model 8 (INLA) presented low mean coverage (79.56%), well below the expected 95%, indicating that this model would not be appropriate for adenocarcinoma with variants of middle ear. These conclusions related to coverage can also be applied to the generalized linear mixed and empirical Bayes models.

Similarly, for adenocarcinoma with variants of trachea, a cancer entity with a median of 6 cases per country during the study period and just 3 countries with 0 counts, the coverage was below 95% in all modeling strategies except for models 4–7 (WinBUGS). Among these models, model 6 again showed the largest width (7.60 cases) and RMSE (RMSE = 1.55), model 7 the smallest width, and models 4 and 5 the smallest RMSEs. Therefore, changes in the lower bound on the uniform prior assumed for σ might improve model performance. Differences between WinBUGS and INLA models were only detected for pD values (see Web Figure 6).

Study 2

Since the results derived from the simulation study suggested that using a uniform prior distribution on σ might be a useful strategy, we assessed the performance of models with different LBs on the uniform prior distribution for models 1–7. Figures 1–4 depict the distribution of these indicators across all scenarios. Model 6 showed the smallest differences on average, but it presented the largest variability for this last indicator. Model 6 ranked first among RC sites that had an IR of less than 0.5 cases per 100,000 person-years during the study period (Figure 1F, 2F, and 3F), whereas model 6 was surpassed by models 3 and 4 when the IR exceeded 0.5 (Figure 4F). Notably, model 6 performed best when DIC and WAIC rankings were averaged into one “overall indicator” (see Web Figure 7).

Finally, we also assessed the impact of using model 6 versus model 8 when comparing the 95% credible intervals for λ_i . For adenocarcinoma with variants of middle ear, the coverage of λ_i using model 6 was clearly better than that with model 8, since the credible interval derived from model 8 did not cover the observed rate provided by RARECARENet in 5 countries (Web Figure 8). However, we noted that differences between model 6 and other models could also be minimal even for entities with very small case counts, such as adenocarcinoma with variants of trachea (Web Figure 9).

DISCUSSION

In the Bayesian modeling of the burden of RC between countries, our study showed that using a uniform prior for the standard deviation of the random effects, σ , with a lower bound of 1, improves the performance of a simple Poisson random-effects model. However, when dealing with very large counts, the influence of the choice of the prior distribution or the non-Bayesian modeling might have a very small influence on the final results, as shown in study 1. In addition, model selection could be improved by selecting

Table 2. Mean Values (and Standard Deviations) of Indicators Considered for Assessing the Performance of Models Used to Estimate Incidence of 2 Types of Cancer With Large Case Counts, RARECARENet Database, 2000–2007^a

Cancer Type and Model Used	CI/CrI Coverage, %	CI/CrI Width, no. of cases	RMSE
Tumors of the central nervous system			
WinBUGS models ^b			
Model 1	95.16 (21.88)	223.62 (128.03)	46.38 (48.69)
Model 2	94.16 (22.18)	223.56 (127.57)	46.36 (48.76)
Model 3	95.33 (21.10)	224.08 (127.33)	46.43 (48.90)
Model 4	95.26 (21.26)	224.67 (127.30)	46.51 (46.51)
Model 5	95.26 (21.26)	224.67 (127.30)	46.51 (49.04)
Model 6	95.11 (21.57)	224.99 (127.35)	46.64 (49.11)
Model 7	94.89 (22.03)	223.66 (127.85)	46.44 (48.92)
INLA models ^c			
Model 8	95.18 (22.18)	223.81 (127.93)	46.44 (48.83)
Model 9	95.18 (22.18)	223.79 (127.90)	46.44 (48.84)
Empirical Bayes approach	94.96 (21.88)	222.33 (128.90)	46.38 (48.68)
Generalized linear mixed model	94.67 (22.48)	223.51 (128.35)	46.39 (48.75)
Poisson approximation			
Exact	95.04 (21.73)	226.20 (127.26)	46.64 (49.08)
Byar's	95.11 (21.57)	225.22 (127.35)	46.64 (49.08)
Adenocarcinoma with variants of ovary			
WinBUGS models			
Model 1	95.04 (21.73)	204.59 (119.07)	40.98 (44.08)
Model 2	95.19 (21.42)	204.55 (118.76)	40.99 (43.85)
Model 3	95.04 (21.73)	204.56 (118.77)	40.92 (43.97)
Model 4	95.11 (21.57)	204.91 (118.57)	40.91 (43.94)
Model 5	95.11 (21.57)	204.91 (118.57)	40.91 (43.94)
Model 6	95.56 (20.62)	205.59 (118.47)	40.99 (43.88)
Model 7	95.11 (21.75)	204.55 (118.81)	40.98 (43.82)
INLA models			
Model 8	95.03 (22.03)	204.70 (118.71)	40.94 (43.88)
Model 9	95.03 (22.03)	204.77 (118.76)	40.93 (43.86)
Empirical Bayes approach	94.67 (22.33)	203.45 (118.84)	40.99 (43.87)
Generalized linear mixed model	94.83 (22.03)	204.32 (118.79)	40.93 (43.82)
Poisson approximation			
Exact	95.26 (21.26)	206.43 (118.08)	40.98 (43.88)
Byar's	95.48 (20.78)	205.50 (118.20)	40.98 (43.87)

Abbreviations: CI, confidence interval; CrI, credible interval; INLA, integrated nested Laplace approximation; RARECARENet, Information Network on Rare Cancers; RMSE, root mean squared error.

^a Coverage and width of the 95% CI or CrI and RMSE between the observed and predicted numbers of cases across 1,000 simulated data sets for 2 cancer sites with large case counts in all areas.

^b WinBUGS models: model 1, $\sigma \sim \text{Uniform}(0, 500)$; model 2, $\sigma \sim \text{Uniform}(0.1, 500)$; model 3, $\sigma \sim \text{Uniform}(0.2, 500)$; model 4, $\sigma \sim \text{Uniform}(0.3, 500)$; model 5, $\sigma \sim \text{Uniform}(0.5, 500)$; model 6, $\sigma \sim \text{Uniform}(1, 500)$; model 7, $\sigma \sim \text{Uniform}(a, b)$.

^c INLA models: model 8, $\tau \sim \gamma(0.00001, 0.00001)$; model 9, $\sigma \sim \text{Uniform}(0, \infty)$.

the model with the smallest ranking indicator derived from an average of the models' rankings using DIC and WAIC. These 2 strategies are effective for modeling and obtaining a reliable measure of variability for country-specific RC IRs.

In the first study, we compared the coverage of the true λ_i between Bayesian and non-Bayesian approaches. A common means of evaluating an objective prior distribution is through the frequentist-matching approach: If posterior

Table 3. Mean Values (and Standard Deviations) of Indicators Considered for Assessing the Performance of Models Used to Estimate Incidence of 2 Types of Cancer With Small Case Counts, RARECARENet Database, 2000–2007^a

Cancer Type and Model Used	CI/CrI Coverage, %	CI/CrI Width, no. of cases	RMSE
Adenocarcinoma with variants of middle ear			
WinBUGS models ^b			
Model 1	96.04 (16.96)	2.54 (2.54)	0.43 (0.69)
Model 2	96.56 (15.45)	2.63 (2.57)	0.43 (0.69)
Model 3	96.22 (13.22)	2.74 (2.61)	0.43 (0.69)
Model 4	96.59 (11.78)	2.86 (2.67)	0.44 (0.70)
Model 5	96.59 (11.78)	2.86 (2.67)	0.44 (0.70)
Model 6	96.37 (12.67)	3.61 (2.94)	0.62 (0.84)
Model 7	96.37 (12.67)	2.84 (2.65)	0.44 (0.70)
INLA models ^c			
Model 8	79.56 (40.34)	1.78 (2.23)	0.46 (0.76)
Model 9	95.59 (18.15)	2.57 (2.56)	0.44 (0.70)
Empirical Bayes approach	51.33 (50.00)	1.23 (2.12)	0.45 (0.70)
Generalized linear mixed model	76.44 (42.45)	1.84 (2.33)	0.45 (0.71)
Poisson approximation			
Exact	96.74 (17.76)	6.11 (2.87)	0.91 (0.99)
Byar's	98.44 (12.38)	5.05 (2.94)	0.91 (0.99)
Adenocarcinoma with variants of trachea			
WinBUGS models			
Model 1	92.44 (26.44)	6.28 (5.40)	1.39 (1.85)
Model 2	92.81 (25.83)	6.34 (5.40)	1.38 (1.83)
Model 3	94.00 (23.76)	6.42 (5.39)	1.37 (1.82)
Model 4	95.48 (20.78)	6.57 (5.43)	1.36 (1.82)
Model 5	95.48 (20.78)	6.57 (5.43)	1.36 (1.82)
Model 6	96.22 (19.07)	7.60 (5.61)	1.55 (1.90)
Model 7	95.30 (23.20)	6.49 (5.42)	1.38 (1.83)
INLA models			
Model 8	87.04 (33.60)	5.92 (5.37)	1.44 (1.91)
Model 9	92.44 (26.44)	6.36 (5.41)	1.39 (1.83)
Empirical Bayes approach	84.15 (36.54)	5.58 (5.18)	1.39 (1.87)
Generalized linear mixed model	85.70 (35.02)	6.01 (5.38)	1.39 (1.84)
Poisson approximation			
Exact	95.19 (21.42)	9.68 (5.48)	1.78 (1.94)
Byar's	97.70 (14.98)	8.67 (5.51)	1.78 (1.94)

Abbreviations: CI, confidence interval; CrI, credible interval; INLA, integrated nested Laplace approximation; RARECARENet, Information Network on Rare Cancers; RMSE, root mean squared error.

^a Coverage and width of the 95% CI or CrI and RMSE between the observed and predicted numbers of cases across the 1,000 simulated data sets for 2 cancer sites with small case counts in all areas.

^b WinBUGS models: model 1, $\sigma \sim \text{Uniform}(0, 500)$; model 2, $\sigma \sim \text{Uniform}(0.1, 500)$; model 3, $\sigma \sim \text{Uniform}(0.2, 500)$; model 4, $\sigma \sim \text{Uniform}(0.3, 500)$; model 5, $\sigma \sim \text{Uniform}(0.5, 500)$; model 6, $\sigma \sim \text{Uniform}(1, 500)$; model 7, $\sigma \sim \text{Uniform}(a, b)$.

^c INLA models: model 8, $\tau \sim \gamma(0.00001, 0.00001)$; model 9, $\sigma \sim \text{Uniform}(0, \infty)$.

credible intervals have good coverage properties, posterior inference is reliable in the absence of past data or sources of information (29). On the other hand, the bounds of the confidence interval and those of the credible interval for a parameter θ might numerically coincide when using a flat

prior, $P(\theta) \propto 1$, since $P(\theta|X) \propto P(\theta)L(X|\theta)$ —the posterior probability for θ , $P(\theta|X)$ —coincides with the likelihood (33), $L(X|\theta)$. However, use of the Bayesian approach allows the researcher to calculate probabilities for assessing whether certain parameter values are more probable than others.

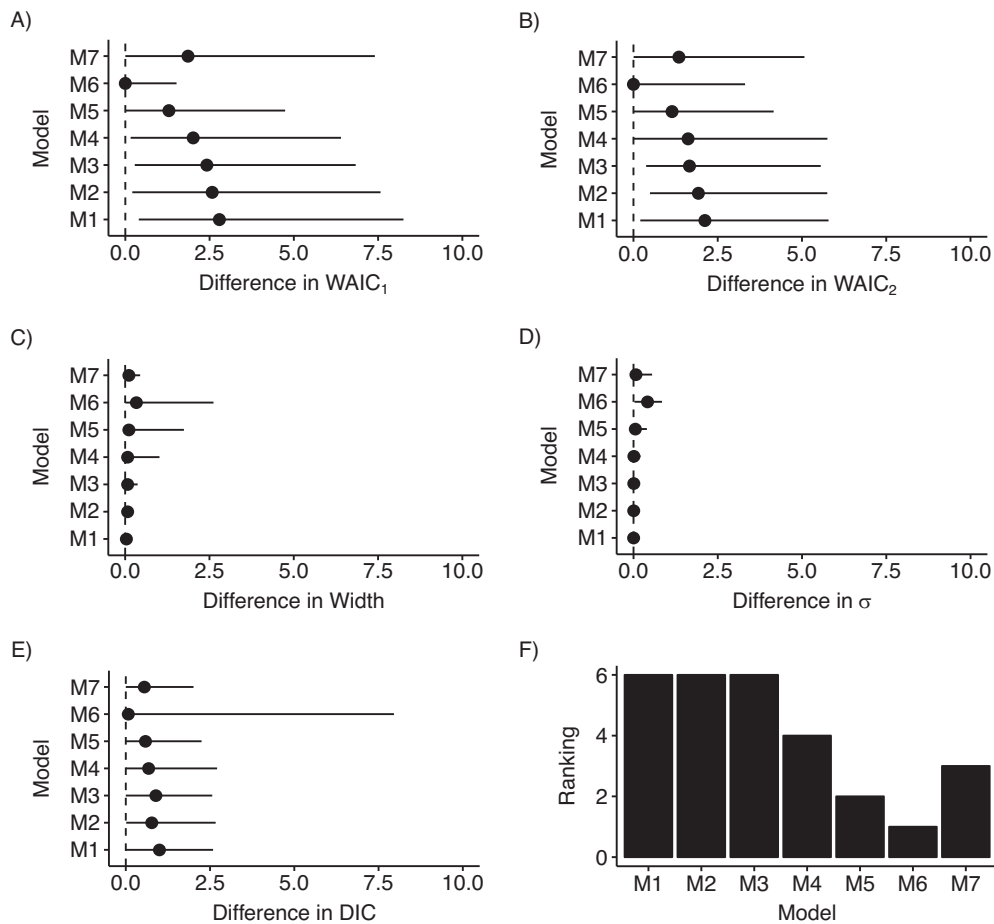


Figure 1. Assessment of models' indicators (differences) and model (M) ranking in scenario A, where the incidence rate (IR) (number of cases per 100,000 person-years) was less than 0.03 (quartile 1 of the IRs (47/190 cancer entities)), RARECARENet database, 2000–2007. The graphs show median values (dots) and 95% credible intervals (lines) for the difference between the estimate of the corresponding indicator according to the specific model and its minimum across all models. A) Difference between the first version of the Watanabe-Akaike information criterion (WAIC₁) for the corresponding model and the minimum WAIC₁ across all models; B) difference between the second version of the Watanabe-Akaike information criterion (WAIC₂) for the corresponding model and the minimum WAIC₂ across all models; C) difference between the width of the 95% credible interval for λ_j derived from the corresponding model and the minimum width of that interval across all models; D) difference between σ estimated using the corresponding model and the minimum σ estimated across all models; E) difference between the deviance information criterion (DIC) for the corresponding model and the minimum DIC across all models; F) ranking of the models according to the average of the rankings derived from the aforementioned indicators. See text for detailed descriptions of models 1–7. RARECARENet, Information Network on Rare Cancers.

Despite this fact, the Bayesian approach also has costs: 1) It adds dependency to the results due to the choice of prior distributions, and 2) it incurs a computational burden when using Markov chain Monte Carlo (MCMC) methods (6, 24, 34). The INLA platform performs approximate Bayesian inference based on the multiple use of Laplace approximations combined with numerical integration, providing faster computation than MCMC methods (6, 24, 25, 34–38), making its use appropriate for simple models. However, a limitation of INLA is related to the use of a bounded uniform prior distribution, since the end user cannot modify the bounds of a uniform prior (24, 25) as in WinBUGS (6, 24, 25, 34).

In the applied study, we used several indicators to assess the performance of the WinBUGS model for different LBs on the uniform prior distribution of σ . We gave the same weight to each of these indicators, and we ranked the models accordingly. However, we believe that the use of all of these indicators might be complex for applied data analysts. Since the most commonly used indicators for Bayesian model choice in epidemiologic studies are DIC and WAIC (30–32), our results suggest that averaging the models' rankings using these indicators could be a good strategy. The DIC can easily be obtained from WinBUGS and INLA, whereas one can use the R code in Web Appendix 10 to carry out this calculation for WAIC indicators. In our study, we reached the same

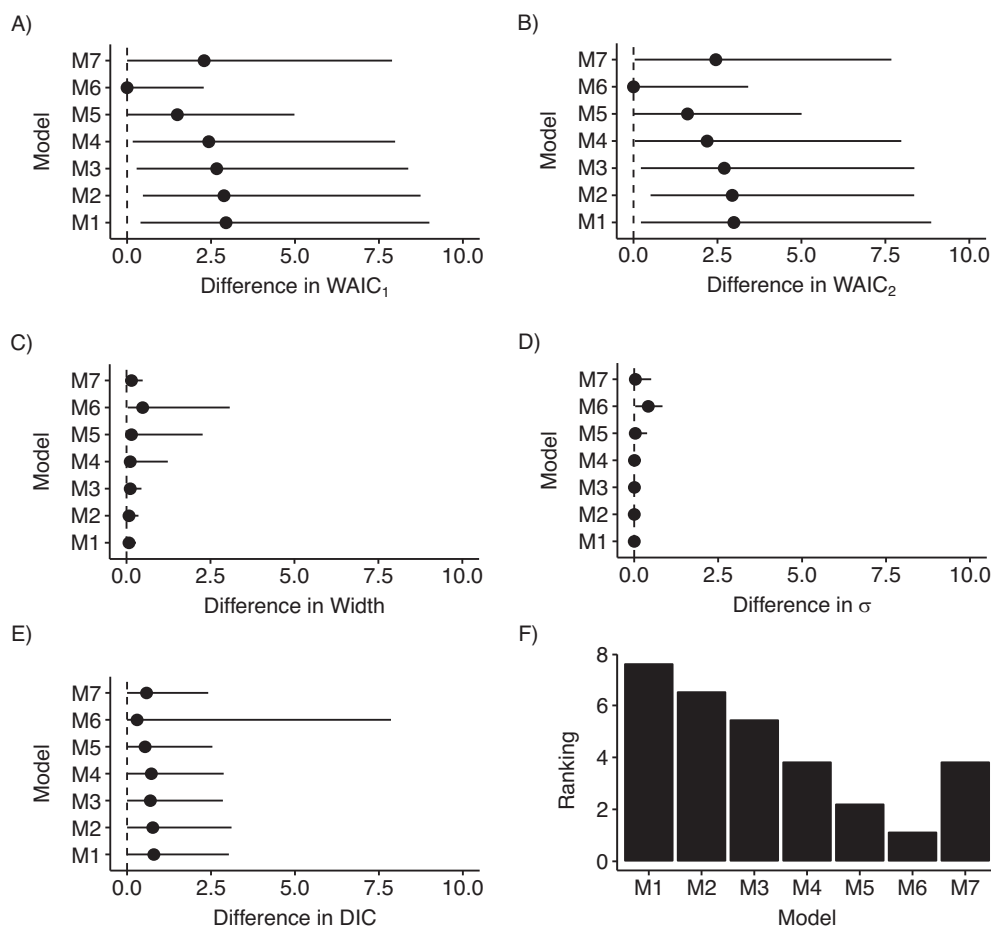


Figure 2. Assessment of models' indicators (differences) and model (M) ranking in scenario B, where the incidence rate (IR) (number of cases per 100,000 person-years) was less than 0.12 (up to quartile 2 of the IRs (97/190 cancer entities)), RARECARENet database, 2000–2007. The graphs show median values (dots) and 95% credible intervals (lines) for the difference between the estimate of the corresponding indicator according to the specific model and its minimum across all models; A) Difference between the first version of the Watanabe-Akaike information criterion (WAIC₁) for the corresponding model and the minimum WAIC₁ across all models; B) difference between the second version of the Watanabe-Akaike information criterion (WAIC₂) for the corresponding model and the minimum WAIC₂ across all models; C) difference between the width of the 95% credible interval for λ_j derived from the corresponding model and the minimum width of that interval across all models; D) difference between σ estimated using the corresponding model and the minimum σ estimated across all models; E) difference between the deviance information criterion (DIC) for the corresponding model and the minimum DIC across all models; F) ranking of the models according to the average of the rankings derived from the aforementioned indicators. See text for detailed descriptions of models 1–7. RARECARENet, Information Network on Rare Cancers.

conclusions using these indicators as we did when using all indicators, except in the situation of relatively “large” IRs ($IR > 0.5$). We found that the best-performing model was that using $\sigma \sim \text{Uniform}(1, 500)$, so we suggest model 6 as a model to start with. However, there might be models that could perform better than model 6 for particular cancer entities (see Web Figure 10).

The relevance of choosing the prior distribution from an empirical Bayes perspective has been also reported (35) and mostly noted as a key factor from the full Bayesian perspective, where the comparative use of different priors is usually considered (19, 21, 22, 25, 34, 36). When estimating the rate of the RC counts through modeling the ratio between

the observed and expected numbers of cases in each country, our results are a compelling argument for using a uniform prior distribution for σ on the random effects. The end users can also assess the impact in the estimation and the variability of the model parameters by changing the LB of the uniform prior when using their own data.

Computational burden and differences in estimates could be an issue when comparing Bayesian platforms for modeling. INLA is much faster than platforms using MCMC methods, and this could be a key determinant of its use, since shorter computational time is a major advantage when analyzing large data sets. If computational time is not an issue, the flexibility in modeling priors through WinBUGS/

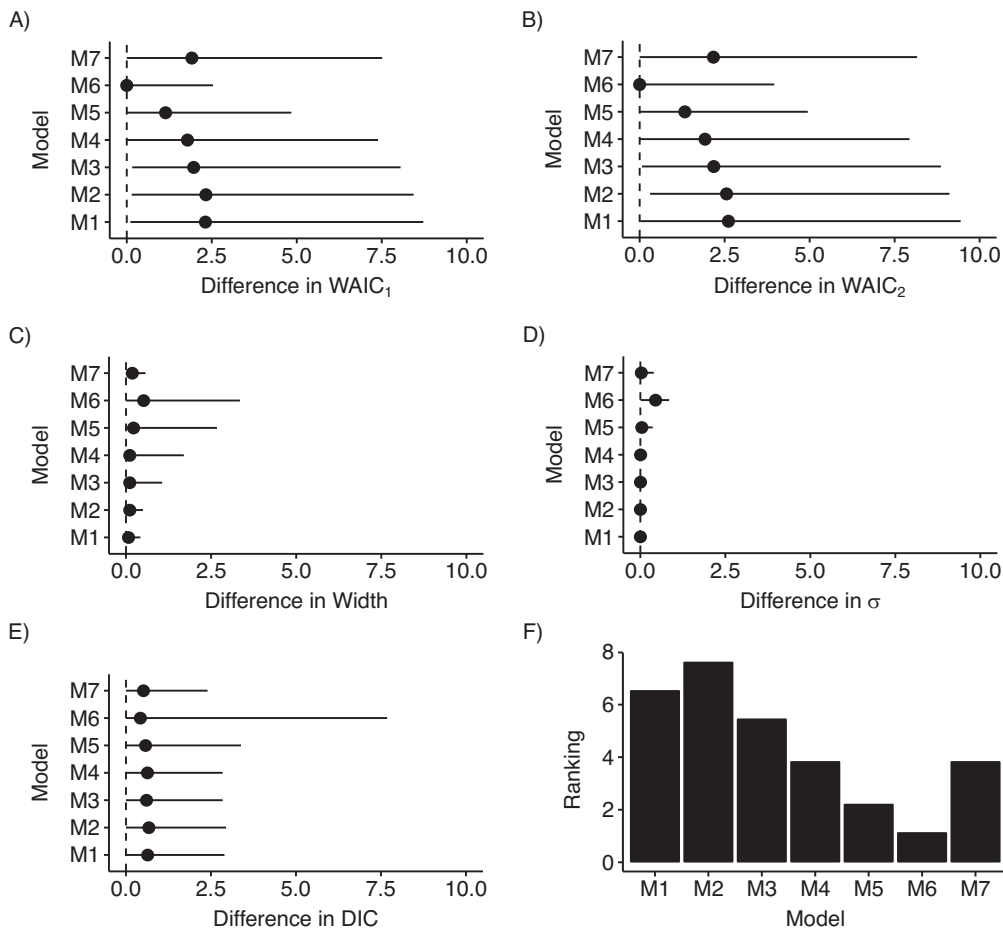


Figure 3. Assessment of models' indicators (differences) and model (M) ranking in scenario C, where the incidence rate (IR) (number of cases per 100,000 person-years) was less than 0.5 (up to quartile 3 of the IRs (143/190 cancer entities)), RARECARENet database, 2000–2007. The graphs show median values (dots) and 95% credible intervals (lines) for the difference between the estimate of the corresponding indicator according to the specific model and its minimum across all models. A) Difference between the first version of the Watanabe-Akaike information criterion (WAIC₁) for the corresponding model and the minimum WAIC₁ across all models; B) difference between the second version of the Watanabe-Akaike information criterion (WAIC₂) for the corresponding model and the minimum WAIC₂ across all models; C) difference between the width of the 95% credible interval for λ_j derived from the corresponding model and the minimum width of that interval across all models; D) difference between σ estimated using the corresponding model and the minimum σ estimated across all models; E) difference between the deviance information criterion (DIC) for the corresponding model and the minimum DIC across all models; F) ranking of the models according to the average of the rankings derived from the aforementioned indicators. See text for detailed descriptions of models 1–7. RARECARENet, Information Network on Rare Cancers.

MCMC is worth considering and can yield more accurate estimates of predictive probabilities than INLA. If computational time is an issue, we suggest using INLA and assessing a sensitivity analysis on the prior of τ by using γ -based priors (6, 14, 24, 25, 39–42). The use of INLA and combinations of MCMC and INLA have been suggested in situations dealing with hard-to-estimate conditional model parameters (6, 42, 43), such as in the case of geostatistical or spatiotemporal models (42). Here, using a simple random-effects model and $n = 27$ observations, all computations were carried out on a 4.8-GHz Intel Core i7 desktop personal computer (Intel Corporation, Santa Clara, California) with 16 GB of random access memory. The median computational time

using model 6, fitted across the 190 cancer sites, was 227 minutes, whereas model 8 took 98 minutes.

Further work

An overall European average for IRs was considered here, but spatial correlation between countries was not. In a scenario where yearly counts are available assuming between-area correlation, a model with identifiability constraints can help capture true spatial effects, and a Bayesian space-time model could also be used for modeling space \times time interaction terms (42–44).

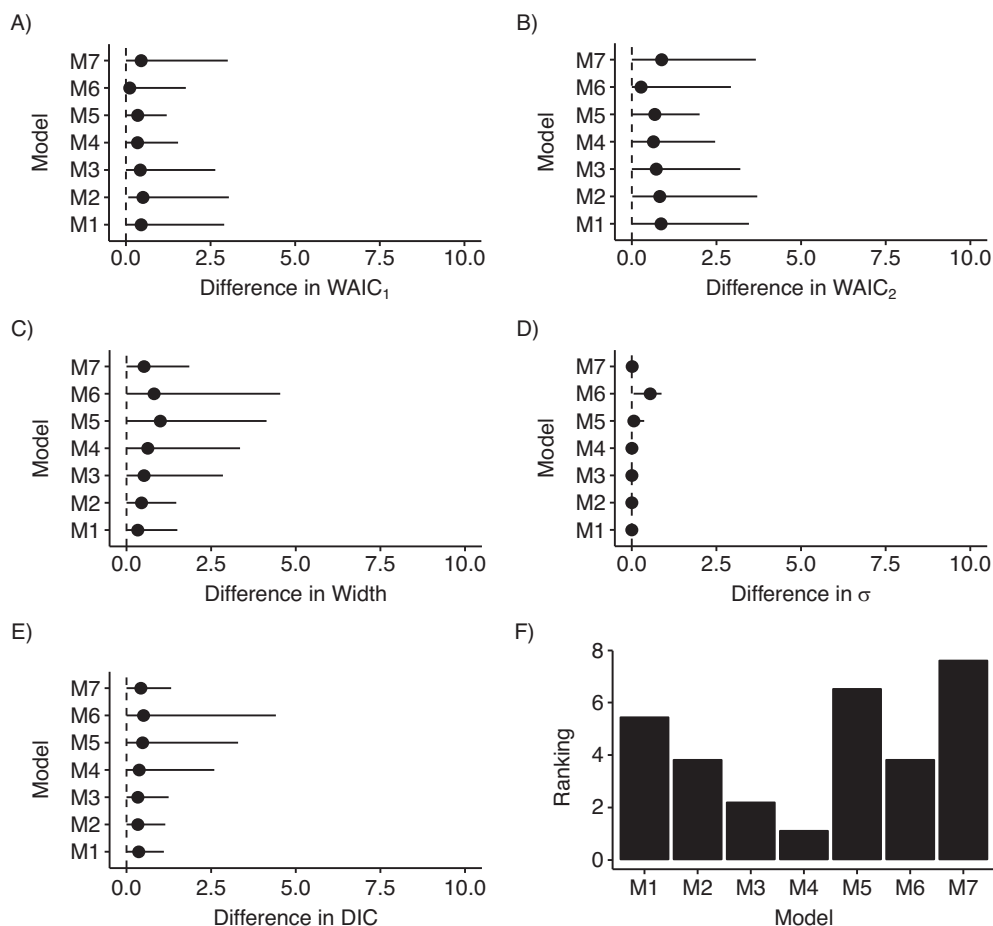


Figure 4. Assessment of models' indicators (differences) and model (M) ranking in scenario D, where the incidence rate (IR) (number of cases per 100,000 person-years) was greater than or equal to 0.5 (quartile 4 of the IRs), RARECARENet database, 2000–2007. The graphs show median values (dots) and 95% credible intervals (lines) for the difference between the estimate of the corresponding indicator according to the specific model and its minimum across all models. A) Difference between the first version of the Watanabe-Akaike information criterion ($WAIC_1$) for the corresponding model and the minimum $WAIC_1$ across all models; B) difference between the second version of the Watanabe-Akaike information criterion ($WAIC_2$) for the corresponding model and the minimum $WAIC_2$ across all models; C) difference between the width of the 95% credible interval for λ_i derived from the corresponding model and the minimum width of that interval across all models; D) difference between σ estimated using the corresponding model and the minimum σ estimated across all models; E) difference between the DIC of the corresponding model and the minimum DIC across all models; F) ranking of the models according to the average of the rankings derived from the aforementioned indicators. See text for detailed descriptions of models 1–7. RARECARENet, Information Network on Rare Cancers.

In this line, within-region variability is an important issue, especially for population-based cancer registry data. However, when dealing with RCs, the small counts often do not allow an in-depth study of the variability within a region from a statistical point of view, unless data are collected over a very long period of time (4). In addition, it is quite plausible that the within-region variability, if any, cannot be estimated with precision from a statistical viewpoint because of small counts. If our data were available disaggregated by local cancer registries within each country, model 2 (equation 2) could be easily modified to account for within-country variability by $\theta_{ir} = \log R_{ir} = \mu + v_i + \alpha_{ir}$. Here v_i and α_{ir} are random effects representing the between-country and within-country variability, respectively, where $v_i \sim N(0, \tau)$,

$\alpha_{ir} \sim N(0, \tau_\alpha)$, i refers to country, and r refers to a within-country unit (local cancer registries). Although the disaggregation of data was not available at this level in the RARECARENet search tool (1–4) and that model cannot be tested with our data set, we suggest that this additional level of variability requires studying assumptions about the prior distributions for τ and τ_α and how these can be combined and validated.

These assumptions must also take into account hypotheses—for example, that incidence is driven by an environmental risk factor present in one area but not another within the same country. In addition, population size has a significant influence on the measure of risk used in spatial modeling (44). These are challenges when applying a “one

size fits all” method, and their future study in the context of RC is warranted.

Limitations

The Joint Action on Rare Cancers strongly supported national/international coordination of clinical management of rare tumors, in terms of networking and physical centralization of treatments (1, 2). Therefore, provision of indicators at a subnational (local/regional cancer registry) level is of minor importance for health-care planning under the project’s aims. However, it remains an issue for etiological research on rare tumors with putative risk factors.

We studied the impact on the prior distributions used for the random effects, representing the unstructured residual for each country. The modeling and priors we used are only as reliable and valid as the data themselves (4). However, the modeling presented here might require additional assumptions in areas with poor cancer surveillance, where low occurrence might reflect issues with disease registration.

The database used here (March 10, 2020) was, unfortunately, the most updated one available at the time of this writing. This is largely due to the application of the European General Data Protection Regulation (45), which required a separate negotiation for data transfer with each one of the more than 100 participating cancer registries, so the speed of the entire data collection was determined by the slowest-reacting registry. Finally, the coronavirus disease 2019 pandemic emergency has entailed further delays over the entire process. We hope that the new database derived from the most recent “call for data” will become available during 2021, since it requires time for centralized processing and data quality checks.

Conclusion

In summary, our study shows that a simple Bayesian Poisson regression model using a uniform prior distribution on σ of the random effects with a lower bound of 1 yields reliable variability for the country-specific RC IRs when these vary around an overall IR. Despite this recommendation, in the context of RCs and small case counts, it is of the utmost importance to perform a sensitivity analysis combining precision with goodness of fit when the end user analyzes his/her own data. Along this line, we suggest selecting the random-effects model for each cancer site according to an averaged ranking indicator which uses DIC and WAIC.

ACKNOWLEDGMENTS

Author affiliations: Health and Social Sciences Department, Instituto Murciano de Investigación Biosanitaria Virgen de la Arrixaca, University of Murcia, Murcia, Spain (Diego Salmerón); Centro de Investigación Biomédica en Red Epidemiología y Salud Pública (CIBERESP), Murcia, Spain (Diego Salmerón); Evaluative

Epidemiology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy (Laura Botta, Annalisa Trama, Gemma Gatta, Riccardo Capocaccia); Department of Statistics and Operations Research, Technical University of Catalonia, Barcelona, Spain (José Miguel Martínez); Public Health Research Group, University of Alicante, Alicante, Spain (José Miguel Martínez); Cancer Plan, Institut d’Investigació Biomèdica de Bellvitge, Catalan Institute of Oncology, Hospitalet de Llobregat, Spain (Josep M. Borràs, Ramon Clèries); and Clinical Sciences Department, Faculty of Medicine and Health Sciences, University of Barcelona, Barcelona, Spain (Josep M. Borràs, Ramon Clèries).

This work was supported by the European Commission through the Consumers, Health, Agriculture and Food Executive Agency (grant 2000111201). We acknowledge the support of the Agència d’Avaluació d’Universitats i Recerca (grant 2017SGR00735) and the Centres de Recerca de Catalunya (CERCA) Program of the Generalitat de Catalunya. We also acknowledge support from the Fundación Séneca—Agencia de Ciencia y Tecnología de la Región de Murcia Program for Excellence in Scientific Research (project 20862/PI/18).

The data underlying this article were provided by RARECARENet by permission. Data will be shared upon request to the corresponding author with the permission of RARECARENet (<http://rarecarenet.istitutotumori.mi.it/rarecarenet/>).

We thank Meggan Harris for her editorial support.

The funders played no role in the study design; data collection, analysis, and interpretation; or the writing of the report.

Conflict of interest: none declared.

REFERENCES

- Gatta G, van der Zwan JM, Casali PG, et al. Rare cancers are not so rare: the rare cancer burden in Europe. *Eur J Cancer*. 2011;47(17):2493–2511.
- Information Network on Rare Cancers (RARECARENet). Cancer list. <http://rarecarenet.istitutotumori.mi.it/rarecarenet/index.php/cancerlist>. Accessed June 10, 2018.
- Gatta G, Capocaccia R, Botta L, et al. Burden and centralised treatment in Europe of rare tumours: results of RARECARENet—a population-based study. *Lancet Oncol*. 2017;18(8):1022–1039.
- Botta L, Capocaccia R, Trama A, et al. Bayesian estimates of the incidence of rare cancers in Europe. *Cancer Epidemiol*. 2018;54:95–100.
- Rue H, Martino S, Chopin N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B Stat Methodol*. 2009;71(2):319–392.
- Carroll R, Lawson AB, Faes C, et al. Comparing INLA and OpenBUGS for hierarchical Poisson modeling in disease mapping. *Spat Spatiotemporal Epidemiol*. 2015;14:45–54.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020.

8. Lunn DJ, Thomas A, Best N, et al. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput.* 2000;10(4):325–337.
9. Breslow NE, Day NE. Statistical methods in cancer research. Volume II—the design and analysis of cohort studies. *IARC Sci Publ.* 1987;(82):1–406.
10. Leyland AH, Davies CA. Empirical Bayes methods for disease mapping. *Stat Methods Med Res.* 2005;14(1):17–34.
11. Swift B. Comparison of confidence intervals for a Poisson mean—further considerations. *Comm Stat Theory Methods.* 2009;38(5):748–759.
12. Waller LA, Carlin BP. Disease mapping. In: Gelfand AE, Diggle PJ, Fuentes M, et al., eds. *Handbook of Spatial Statistics*. Boca Raton, FL: Chapman & Hall/CRC Press; 2010:217–243.
13. Patil VV, Kulkarni HV. Comparison of confidence intervals for the Poisson mean: some new aspects. *REVSTAT.* 2012; 10(2):211–227.
14. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics.* 1987;43(3):671–681.
15. Marshall RJ. Mapping disease and mortality rates using empirical Bayes estimators. *J Roy Stat Soc C.* 1991;40(2): 283–294.
16. McCullagh P, Nelder JA. *Generalized Linear Models*. 2nd ed. (Monographs on Statistics and Applied Probability, vol. 37). Boca Raton, FL: Chapman & Hall/CRC Press; 1989.
17. Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia, PA: Lippincott-Raven Publishers; 1998.
18. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health Care Evaluation*. New York, NY: John Wiley & Sons, Inc.; 2004.
19. Lambert PC, Sutton AJ, Burton PR, et al. How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Stat Med.* 2005;24(15):2401–2428.
20. Waller LA. Bayesian thinking in spatial statistics. In: Dey DK, Rao CR, eds. *Bayesian Thinking: Modeling and Computation*. 1st ed. (Handbook of Statistics, vol. 25). Amsterdam, the Netherlands: Elsevier BV; 2005:589–622.
21. Browne WJ, Draper D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Anal.* 2006;1(3):473–514.
22. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* 2006;1(3):515–534.
23. Lunn D, Jackson C, Best N, et al. *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Boca Raton, FL: Chapman & Hall/CRC Press; 2003.
24. Krainski ET, Gómez-Rubio V, Bakka H, et al. *Advanced Spatial Modeling With Stochastic Partial Differential Equations Using R and INLA*. Boca Raton, FL: Chapman & Hall/CRC Press; 2019.
25. Gómez-Rubio V, Palmí-Perales F, López-Abente G, et al. Bayesian joint spatio-temporal analysis of multiple diseases. *SORT.* 2019;43(1):51–74.
26. Sturtz S, Ligges U, Gelman A. R2WinBUGS: a package for running WinBUGS from R. *J Stat Softw.* 2005;12(3):1–16.
27. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci.* 1992;7(4):457–511.
28. Greenland S, Senn SJ, Rothman KJ, et al. Statistical tests, *P* values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol.* 2016;31(4):337–350.
29. Bayarri MJ, Berger JO. The interplay of Bayesian and frequentist analysis. *Stat Sci.* 2004;19(1):58–80.
30. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res.* 2010;11: 3571–3594.
31. Watanabe S. A widely applicable Bayesian information criterion. *J Mach Learn Res.* 2013;14:867–897.
32. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. *Stat Comput.* 2016; 24(6):997–1016.
33. Jaynes ET. Confidence intervals vs Bayesian intervals. In: Harper WL, Hooker CA, eds. *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*. Dordrecht, the Netherlands: D. Reidel Publishing Company; 1976:175–211.
34. Grilli L, Metelli S, Rampichini C. Bayesian estimation with integrated nested Laplace approximation for binary logit mixed models. *J Stat Comp Sim.* 2015;85(13):2718–2726.
35. Yasui Y, Liu H, Benach J, et al. An empirical evaluation of various priors in the empirical Bayes estimation of small area disease risks. *Stat Med.* 2000;19(17-18):2409–2420.
36. Taylor BM, Diggle PJ. INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *J Stat Comp Sim.* 2014;84(10):2266–2284.
37. Fong Y, Rue H, Wakefield J. Bayesian inference for generalized linear mixed models. *Biostatistics.* 2010;11(3): 397–412.
38. Bivand RS, Gómez-Rubio V, Rue H. Spatial data analysis with R-INLA with some extensions. *J Stat Soft.* 2015;63(20): 1–31.
39. Kelsall JE, Wakefield JC. Discussion of “Bayesian models for spatially correlated disease and exposure data” by N. Best, I. Waller, A. Thomas, E. Conlon and R. Arnold. In: Bernardo JM, Berger JO, Dawid AP, et al., eds. *Bayesian Statistics 6: Sixth Valencia International Meeting on Bayesian Statistics, June 6–10, 1998*. London, United Kingdom: Oxford University Press; 1999:131–156.
40. Crainiceanu CM, Diggle PJ, Rowlingson B. Bayesian analysis for penalized spline regression using WinBUGS. *J Am Stat Assoc.* 2008;102:21–37.
41. Roos M, Martins TG, Held L, et al. Sensitivity analysis for Bayesian hierarchical models. *Bayesian Anal.* 2015;10(2): 321–349.
42. Ugarte MD, Adin A, Goicoa T, et al. On fitting spatio-temporal disease mapping models using approximate Bayesian inference. *Stat Methods Med Res.* 2014;23(6): 507–530.
43. Gómez-Rubio V, Rue H. Markov chain Monte Carlo with the integrated nested Laplace approximation. *Stat Comput.* 2018; 28(5):1033–1051.
44. Fernández-Navarro P, González-Palacios J, González-Sánchez M, et al. Ranking spatial areas by risk of cancer: modelling in epidemiological surveillance. *Ann Cancer Epidemiol.* 2020;4:10.
45. European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (text with EEA relevance). (OJ L 119, 4.5.2016, pp. 1–88). <https://eur-lex.europa.eu/eli/reg/2016/679/oj>. Accessed April 24, 2021.