**SUPPLEMENTARY MATERIAL**

Emerging Infectious Diseases

**Joint estimation of relative risk of dengue and Zika infections**

D.A. Martínez-Bello 1, \*, A. López-Quílez 1, A. Torres Prieto 2,

1 Department of Statistics and Operations Research, Faculty of Mathematics, University of Valencia, Spain

2 Office of Demography and Epidemiology, Health Office, Department of Santander, Colombia

\* Author for correspondence: D. Martínez-Bello, Department of Statistics and Operations Research, Faculty of Mathematics, University of Valencia, C/ Dr. Moliner, 50, 46100 Burjassot, Valencia, Spain, E-mail: danieladyro@gmail.com

**Introduction**

In this supplementary material we present to the interested reader the statistical formulation for the joint models in the paper “Joint estimation of relative risk of dengue and Zika infections” and the model diagnostic measures. Together with the supplementary material, the reader can use the .txt files located at the same repository of this file, containing the WinBUGS 1.4 code, data and initial values necessary to reproduce the results shown in the manuscript as a result of fitting the joint models of relative risk 1 to 8 for the department of Santander, and the city of Bucaramanga.

**Statistical formulation of the joint models of relative risk**

Let us assume the observed counts of dengue or Zika virus disease (ZVD) are Poisson distributed with mean parameter () where *i* is the aggregation area (*i* = 1,..,*n*, and *n* = 87 municipalities at departmental level; or *n* = 293 census section for the municipal level), and *j* is the disease (*j* = 1,…,*p*, and *p* = 1 for Dengue or *p* = 2 for ZVD), then,

Thus, the mean parameter is equal to the product of the expected values and the relative risk with linear predictor. The linear predictor includes an intercept for every disease, and the parameter which could be spatially clustered or uncorrelated random effects or covariates. Spatially clustered random effects are unobserved variables recovering spatial clustered patterns of risk, or the fact that the one area’s risk is highly associated with the neighboring areas. The lack of spatial association is accounted by the spatially uncorrelated random effects (1).

All the models shown in the following rely in different formulations for the parameter, because every formulation accounts for a different association structure between the disease-specific parameters. Models 1 and 2 in Table 1 contain spatially uncorrelated random effects linearly un-associated (Model 1) or associated (Model 2) between both diseases. Models 3 and 4 in Table 2 contain spatially clustered random effects linearly associated (Model 4) or not (Model 3). Models 5 and 6 in Table 3 contain disease-specific spatially uncorrelated (Model 5) and clustered (Model 6) random effects including a spatially clustered shared-component for both diseases. Models 7 and 8 in Table 4 fit spatially clustered random effects of dengue (or ZVD) conditioned on spatially clustered random effects of ZVD (or dengue).

In Figure 1, we can appreciate the different spatial patterns of risk that each joint model accounts.

Table 1. Statistical formulation of joint models of relative risk 1 and 2.

|  |  |
| --- | --- |
| ­­Model 1 | Model 2 |
| ­This model contains disease-specific independent and identically distributed (IID) Normal spatially uncorrelated random effects for dengue and ZVD. | This model contains disease-specific independent and identically distributed (IID) Normal spatially uncorrelated random effects linearly correlated for both diseases. |
|  |  |
| The are spatially uncorrelated random effects, the are intercepts, are variance parameters of the , and ***I*** is an identity matrix, and is the relative risk in area *i* and disease *j*. | is an variance-covariance matrix accounting for the linear association of the spatially uncorrelated random effects , corresponds to the Kronecker product of two matrices, **R** is the parameter matrix for the Wishart distribution, and the other parameters similar to Model 1. |

Table 2. Statistical formulation of joint models of relative risk 3 and 4

|  |  |
| --- | --- |
| Model 3 | Model 4 |
| This model accommodates conditionally autoregressive (CAR) Normal spatially clustered random effects for every disease, which are not linearly associated. | This model contains CAR Normal spatially clustered random effects linearly associated for both diseases. |
|  |  |
| In this model, the vectors are spatially structured random effects Normally distributed with mean zero vector, variance parameters and structure matrix , the are intercepts, and the ***I*** matrix is an identity matrix. The matrix is a structure matrix where **D** is a diagonal matrix of the total number of neighbors of every area, and **W** is a proximity matrix. In the W matrix, the symbol means “neighbor of”. | This time, the are spatially structured random effects Normally distributed with zero mean vector and variance covariance matrix , where is the variance-covariance matrix of the , the are intercepts for every disease, and is the correlation parameter in the variance covariance matrix **.** The matrix is a structure matrix where **D** is a diagonal matrix of the total number of neighbors of every area, and **W** is a proximity matrix. In the W matrix, the symbol means “neighbor of”. The inverse of the variance covariance matrix is assumed Wishart distributed with **R** matrix and 2 degrees of freedom. |

Table 3. Statistical formulation of joint models of relative risk 5 and 6

|  |  |
| --- | --- |
| Model 5 | Model 6 |
| This model accommodates disease-specific IID Normal spatially uncorrelated random effects, and shared-components of spatial clustered patterns of risk for both diseases | This model accommodates disease-specific CAR Normal spatially clustered random effects, and shared-components of spatial clustered patterns of risk for both diseases |
|  |  |
| In this model, the vectors are Normally distributed unstructured spatial random effects for every disease with mean zero vector, variance parameters, and identity matrix ***I***. The vector is the spatially structured shared-parameter Normally distributed with zero mean vector, variance parameter, structure matrix, and is a scaling parameter. ***D*** and ***W*** were already defined above. | In this model, the vectors are Normally distributed structured spatial random effects for every disease with mean zero vector, variance parameters, and identity matrix ***I***. The vector is the spatially structured shared-parameter vector Normally distributed with zero mean vector, variance parameter, structure matrix, and is a scaling parameter. ***D*** and ***W*** were already defined above. |

Table 4. Statistical formulation of joint models of relative risk 7 and 8.

|  |
| --- |
| Models 7 and 8 |
| Model 7 accommodates the generalized multivariate conditionally autoregressive (CAR) model (1), where the disease- and area-specific CAR Normal spatially clustered random effects of ZVD are conditioned by the CAR Normal spatially clustered random effects of dengue. Model 8 presents the generalized multivariate CAR model [3], where the disease- and area- CAR Normal spatially clustered random effects of dengue per area are conditioned by the CAR Normal spatially clustered random effects of ZVD. |
|  |
| The parameters in Models 7 and 8 are: parameters are spatially structured random effects, where the random effects vector for disease 1 is conditioned on the random effects vector for disease 2. The parameters are Normally distributed with mean vector and variance covariance matrix. The mean vector depends of the parameters , , , and , the vector of one’s **1**, the identity matrix ***I***, and the proximity matrix ***W***, while the covariance matrix depends on the variance , the parameter, and the diagonal matrix **D** and the proximity matrix **W**. In contrast, the mean vector only depends of the parameter, while the variance covariance matrix depends of the variance , the , and the diagonal matrix **D** and the proximity matrix **W**. |

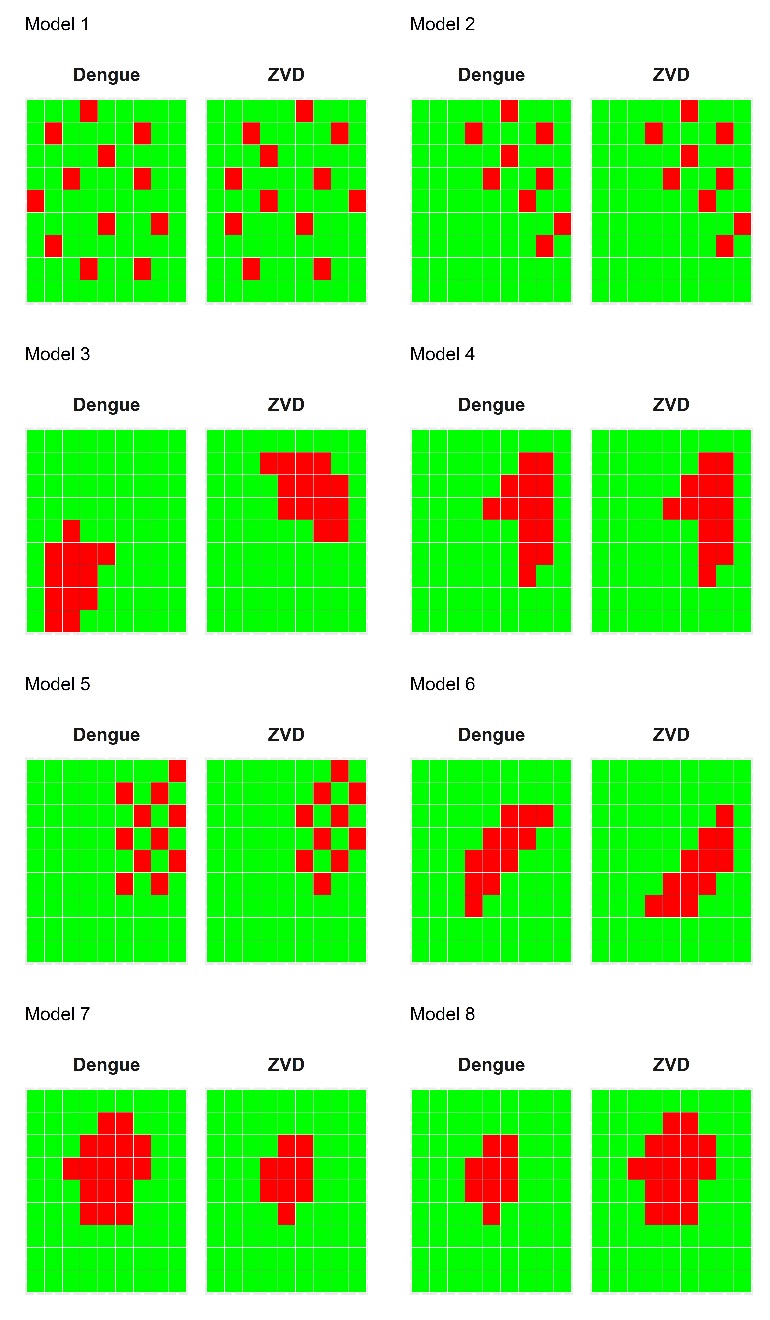


Figure 1. Spatial patterns represented by the joint models of relative 1 to 8 for dengue and ZVD.

**Calculating expected values of dengue and ZVD**

The joint models of relative risk require the disease- and area-specific observed counts and the expected counts. Expected counts are obtained by external or internal standardization. External standardization requires incidence rates (cases per 100,000 people) by age groups and sex obtained from a standard or reference population, while internal standardization uses incidence rates obtained from the same data. In the manuscript, we used internal standardization to compute the expected values. The incidence rate is computed by

where ***IRpq*** is the incidence rate in age groups p = 1,…,P, (1 = [0; 5); 2 = [5; 10),…, 14 = [65; 100)) and sex q = 1, 2, (1=female, 2= male); ***Casespq*** are the total number of ZVD or dengue cases; and ***Populationpq*** is the total population in age group *p* and sex *q* for the departmental or city level over the complete study period, as obtained from the 2016 projected population of the Colombian census 2005. Using the ***IRpq***, the expected values of dengue or ZVD per municipality at departmental level, or census section at city level were calculated using:

where ***Ei*** is the expected value in small area *i*, and ***Populationipq*** is the census population in small area *i*, age group *p* and sex *q*. At the end, we remain with four sets of expected values of dengue and ZVD: Santander's ZVD and dengue expected values, and Bucaramanga's ZVD and dengue expected values.

Using the observed and expected values of dengue and ZVD, a raw estimator of risk is the ratio between the observed and the expected values of dengue and ZVD per area *i* which is the standardized incidence ratio (SIR),

The standardized incidence ratio are plotted in maps in the manuscript, as a descriptive step of modeling relative risk at departmental and city level.

**Diagnostic measures for the joint models of relative risk**

For model diagnosis we use posterior predictive checks, residual histograms, and scatter plots and Spearman correlation coefficient of the observed versus the fitted counts. To assess whether the model recovers the over dispersion observed in the data we compare the ratio based on sampled new data with the corresponding ratio for the observed counts (2). This check was done at each iteration and a satisfactory model will have exceeding about 50% of the time.

Values of near 0 or 1 (above 0.9 or below 0.1) indicate discrepancy between the observations and the model, while values close to 0.5 mean that the observed data and the fitted data sampled from the model are closely comparable in terms of the over dispersion function. Then, Table 5 shows the predictive checks on over dispersion from the joint models 1 to 8 in Santander and Bucaramanga. Following the decision criteria, all the joint models for the department of Santander recovers the over dispersed data, because the predictive checks show an average probability of = 0.50, while the predictive checks for the city of Bucaramanga show some models with acceptable posterior predictive checks for over dispersion recovery (Models 1 to 7 for dengue, and Models 1, 3 and 8 for ZVD), and with acceptable but close to the boundary probabilities for Model 8 for dengue and Models 2, 4, 5, 6, and 7 for ZVD. The conclusion is that in the city of Bucaramanga all models clearly recovered the over dispersion, however ZVD data were more difficult to fit than dengue data, and in the department of Santander, all models recovered over dispersion very well. Figure 2 displays the residual histograms of the posterior mean of the fitted values in Santander (A) and Bucaramanga (B). At departmental level, model residuals shrinks to zero, and the residual boundaries are -1.5 to 1. In the city of Bucaramanga, model residuals are more dispersed than residuals in Santander, and the residual boundaries are from -2.5 to 5.0. From the residual examination, similar to the predictive checks in over dispersion, we observe that the joint models fitted the data worse for the city level than for the departmental level. The association between the posterior mean of the fitted values and the observed counts of dengue or ZVD is accounted using scatter plots of fitted versus observed and the assessment of the Spearman correlation coefficient. For the department of Santander, the association of fitted and observed values in Figure 3 (A) in Bucaramanga and (B) in Santander reveal that all models generate fitted values very close to the observed counts, which is ratified by the near to one Spearman correlation coefficients from Table 6. For the city of Bucaramanga, we observe that there are some variability in the association fitted- observed, so the models are missing something in following the observed counts, which is confirmed by the Spearman correlation coefficients near to 0.950 for most of the models. As an overall conclusion, joint models work very well by predicting the observed counts of dengue and ZVD, recovering the data over dispersion, and reducing the dispersion revealed in the residuals for the departmental level, although the performance could be improved at the city level. Reasons for the lower performance of the joint models at city level could be explained by the need to include other parameters in the linear predictor, for instance the inclusion of correlated as well uncorrelated heterogeneity parameters at the same time, or perhaps to difficulties of the model to fit high number of areas (293 at city level compared with 87 areas at departmental level). Although the lower performance of the joint models at city level is not an ideal predictive feature, we have observed that the linear association fitted-observed counts is greater than 0.92, which is not so bad, so we can proceed to make model selection based on the DIC.

|  |  |
| --- | --- |
| C:\Users\danieladyro\Dropbox\2015Latex\2017Reporte008_jointModelEID_version04\figresidSan.jpg | C:\Users\danieladyro\Dropbox\2015Latex\2017Reporte008_jointModelEID_version04\figresidBuc.jpg |
| (A) Santander | (B) Bucaramanga |

Figure 2. Histogram of the residuals for the joint Models 1 to 8 of (A) the department of Santander, and (B) the city of Bucaramanga.

|  |  |
| --- | --- |
| C:\Users\danieladyro\Dropbox\2015Latex\2017Reporte008_jointModelEID_version04\figFitObsSan.jpg | C:\Users\danieladyro\Dropbox\2015Latex\2017Reporte008_jointModelEID_version04\figFitObsBuc.jpg |
| 1. Santander | (B) Bucaramanga |

Figure 3. Scatter plot for the posterior mean of the fitted values versus the observed values of dengue and ZVD from the joint Models 1 to 8, (A) department of Santander, and (B) city of Bucaramanga.

Table 5. Predictive check on over dispersion by the joint Models 1 to 8, for the department of Santander and the city of Bucaramanga.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Bucaramanga | |  | Santander | |
| Model | Dengue | ZVD | Model | Dengue | ZVD |
| 1 | 0.71 | 0.78 | 1 | 0.50 | 0.50 |
| 2 | 0.51 | 0.88 | 2 | 0.51 | 0.50 |
| 3 | 0.62 | 0.68 | 3 | 0.51 | 0.49 |
| 4 | 0.43 | 0.81 | 4 | 0.50 | 0.49 |
| 5 | 0.70 | 0.83 | 5 | 0.51 | 0.50 |
| 6 | 0.41 | 0.82 | 6 | 0.51 | 0.50 |
| 7 | 0.39 | 0.86 | 7 | 0.51 | 0.50 |
| 8 | 0.48 | 0.81 | 8 | 0.50 | 0.49 |

Table 6. Spearman correlation coefficients for the posterior mean of the fitted values and observed values for the joint Models 1 to 8 of relative risk.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Santander | | | Bucaramanga | | |
| Model | Dengue | Zika | Model | Dengue | Zika |
| 1 | 0.995 | 0.994 | 1 | 0.962 | 0.972 |
| 2 | 0.995 | 0.987 | 2 | 0.938 | 0.958 |
| 3 | 0.995 | 0.992 | 3 | 0.944 | 0.957 |
| 4 | 0.992 | 0.976 | 4 | 0.925 | 0.945 |
| 5 | 0.995 | 0.984 | 5 | 0.932 | 0.949 |
| 6 | 0.994 | 0.980 | 6 | 0.922 | 0.945 |
| 7 | 0.994 | 0.988 | 7 | 0.928 | 0.952 |
| 8 | 0.995 | 0.984 | 8 | 0.929 | 0.947 |

**References**

[1] Banerjee S, Carlin BP, Gelfand AE. Hierarchical Modeling and Analysis for Spatial Data, Second Edition Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Boca Raton, Fl.; 2014. 584 Pages.

[2] Congdon P. Bayesian Models for Categorical Data. Wiley Series in Probability and Statistics. John Wiley & Sons, NY. 2005. 466 pages.