Discrete versus continuous domain models for disease mapping and applications on childhood cancers

Garyfallos Konstantinoudis

MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College, St. Marys Campus, Norfolk Place, London W2 1PG, UK

The main goals of disease mapping is to calculate disease risk and identify high-risk areas. Such analyses are hampered by the limited geographical resolution of the available data. Typically data are counts of cases per spatial unit and the most common approach is the Besag-York-Mollié model (BYM). Less frequently, exact geocodes are available, allowing modelling a disease as a point process through Log-Gaussian Cox processes (LGCPs). The objective of this study is to examine in a simulation the performance of BYM and LGCPs for disease mapping. We simulated data in the Canton of Zurich in Switzerland sampling cases from the true population mimicking the childhood leukaemia incidence (n=334 during 1985-2015). We considered 39 different scenarios varying in the risk generating function (step-wise, smooth, flat risk), the size of the high-risk areas (1, 5 and 10km radii), the risk increase within the high-risk areas (2 and 5-fold) and the number of cases (n, 5n and 10n). We used the root mean integrated square error (RMISE) to examine the ability of the models to recover the true risk surface and their sensitivity/specificity in identifying high-risk areas. We found that, for larger radii, LGCPs recover the true risk surface with lower error across almost all scenarios (median RMISE: 9.17-27.0) compared to the BYM (median RMISE: 9.12-35.6). For radii = 1km and flat risk surfaces BYM performs better. In terms of sensitivity and specificity across almost all scenarios the median area under the curve (AUC) for LGCPs was higher (median AUC: 0.81-1) compared to the BYM (median AUC: 0.65-0.93). We applied these methods to childhood leukaemia incidence in the canton of

Zurich during 1985-2015 and identified two high-risk spatially coherent areas. Our findings suggest that there are important gains to be made from the use of LGCP models in spatial epidemiology.