Investigating spatial patterns of cancer in the administrative district of Münsterland (North Rhine-Wesphalia, Germany) linking GIS and R

Dorothea Lemke[1] and

Volkmar Mattauch[2], Edzard Pesemba[3], Hans-Werner Hense[2,4]

[1] Department of Geoscience, University of Münster, Robert-Koch-Str. 28, 48149 Münster, Germany
[2] Epidemiological Cancer Registry NRW, Robert-Koch-Str. 40, 48149 Münster, Germany
[3] Institute for Geoinformatics, University of Münster, Weselerstr. 253, 48151 Münster, Germany
[4] Institute of Epidemiology and Social Medicine, University of Münster, Domagkstr. 3, 48149 Münster, Germany

Background

Spatial cluster detection is an important tool in cancer surveillance to identify areas of elevated risk and to generate hypotheses about cancer etiology. There are many cluster detection methods used in spatial epidemiology to investigate suspicious groupings of cancer occurrences. However, the use of point pattern analysis (PPA) is less widely used within routine health services in Germany than might be expected given its potential usefulness demonstrated in research studies. This study is an effort to prepare the transition from research applications to widespread use within a health service, i.e. cancer registry, setting.
The present study has a clear methodological focus: It aims at integrating point pattern analysis into a geographical information system (GIS) framework in order to find spatial and temporal clusters of cancer incidence, based on non-aggregated data without any a priori hypothesis. It requires linking of a GIS with a powerful statistical software as well as developing a clear GIS- and spatial analysis strategy. The analysis includes three steps: data pre-processing (data import, development of a geodatabase), spatial data analysis and data post-processing (visualization/mapping of results).
Data sets to be used are incident cases of any cancer recorded in the Cancer Registry of the Regierungsbezirk Münster between 1990 and 2005 and Geodaten NRW (topographic maps, environmental data). We employ ArcGIS and R as software tools.

Methods:

The geocoding of the address data was done using the services of Geobasis.nrw provided by IT.NRW.
In this study, we will use methods of exploratory spatial data analysis (summaries of spatial distributions, measures of central tendency, dispersion, orientation, G-/F-function) to describe the location of events (that is, cases of incident cancer. Second, we will use methods to make inferences about the underlying process that generated the pattern. These inferential statistical analyses include tests of complete spatial randomness via Monte Carlo methods, and kernel density estimations (first order properties) to evaluate the probability of event occurrence, accounting for geographic variation in the density of the population at risk. Furthermore, measures of second order properties which reflect any tendency of the cases to appear clustered, independently or regular spaced (K-function for homogeneous or inhomogeneous Poisson processes) will be applied.
In order to assess the potential association between diverse cancer entities and the exposure to point sources (e.g. of pollution, like waste incinerators, radio masts, etc.) and other covariates (like social determinants), emerging clusters will be analysed exploratively with binary regression using GAM and with the raised incidence modelling approach. For the estimation of the spatial distribution of the resident population we consider a spatial case-control design where the controls will be selected at random from the resident population (Einwohnermeldeämter).

Results:

Because of the early state of the study, only the first step: the data pre-processing can be presented here. This refers to the data import of the geo-referenced address data into GIS and to the development of a geo-database. Moreover, the analysis environment, especially the data transfer between GIS and R, is presented as well as the preparation of the data for the spatial analysis in R (e.g. spatial data object handling).
Secondly, the further proceeding in the statistical point pattern analysis should be displayed and discussed. A major issue is how to account for the influence of the spatial variation of the underlying population at risk and the integration of direct and indirect standardizations in epidemiology to improve the feasibility and practicability of utilizing the available demographic information.