International Journal of Health Geographics

BioMed Central

Methodology

Open Access

Geostatistical analysis of disease data: estimation of cancer mortality risk from empirical frequencies using Poisson kriging

Pierre Goovaerts*

Address: BioMedware, Inc., Ann Arbor, MI, USA

Email: Pierre Goovaerts* - goovaerts@biomedware.com

* Corresponding author

Published: 14 December 2005

International Journal of Health Geographics 2005, **4**:31 doi:10.1186/1476-072X-4-31

Received: 10 October 2005

Accepted: 14 December 2005

This article is available from: http://www.ij-healthgeographics.com/content/4/1/31

Abstract

Background: Cancer mortality maps are used by public health officials to identify areas of excess and to guide surveillance and control activities. Quality of decision-making thus relies on an accurate quantification of risks from observed rates which can be very unreliable when computed from sparsely populated geographical units or recorded for minority populations. This paper presents a geostatistical methodology that accounts for spatially varying population sizes and spatial patterns in the processing of cancer mortality data. Simulation studies are conducted to compare the performances of Poisson kriging to a few simple smoothers (i.e. population-weighted estimators and empirical Bayes smoothers) under different scenarios for the disease frequency, the population size, and the spatial pattern of risk. A public-domain executable with example datasets is provided.

Results: The analysis of age-adjusted mortality rates for breast and cervix cancers illustrated some key features of commonly used smoothing techniques. Because of the small weight assigned to the rate observed over the entity being smoothed (kernel weight), the population-weighted average leads to risk maps that show little variability. Other techniques assign larger and similar kernel weights but they use a different piece of auxiliary information in the prediction: global or local means for global or local empirical Bayes smoothers, and spatial combination of surrounding rates for the geostatistical estimator. Simulation studies indicated that Poisson kriging outperforms other approaches for most scenarios, with a clear benefit when the risk values are spatially correlated. Global empirical Bayes smoothers provide more accurate predictions under the least frequent scenario of spatially random risk.

Conclusion: The approach presented in this paper enables researchers to incorporate the pattern of spatial dependence of mortality rates into the mapping of risk values and the quantification of the associated uncertainty, while being easier to implement than a full Bayesian model. The availability of a public-domain executable makes the geostatistical analysis of health data, and its comparison to traditional smoothers, more accessible to common users. In future papers this methodology will be generalized to the simulation of the spatial distribution of risk values and the propagation of the uncertainty attached to predicted risks in local cluster analysis.