

A Regional Approach for Modeling Dog Cancer Incidences with Regard to Different Reporting Practices

G. Boo^{1,2}, S. Leyk³, S. I. Fabrikant¹, A. Pospischi²

¹Department of Geography, University of Zurich, Zurich, Switzerland
Email: {gianluca.boo; sara.fabrikant}@geo.uzh.ch

²Collegium Helveticum, Zurich, Switzerland
Email: apos@vetpath.uzh.ch

³Geography Department, University of Colorado, Boulder, CO, USA
Email: stefan.leyk@colorado.edu

Abstract

Underreporting is a persistent limitation in research on environmental risk factors for dog cancer, impeding potential comparative investigations with human cancer. To address this challenge, we propose a regional modeling approach accounting for different reporting practices across the study area. In doing this, we demonstrate the need for new modeling strategies to improve statistical performance through more systematic assessments of spatial non-stationarity of statistical associations that can be linked to underreporting.

1. Introduction

Humans and dogs have been sharing their habitat for millennia by being exposed to similar environmental conditions over time. An interesting aspect of this co-evolutionary process is that the development of cancer in humans and dogs might be comparable. Environmental exposures associated with cancer in dogs could thus serve to timely identify risk for humans (Pinho *et al.* 2012). However, to date, only few studies have addressed possible linkages between environmental risk factors and dog cancer. This gap is due to uncertainty in most existing dog cancer databases, typically because of underreporting. For this reason, in this study, we investigate underreporting of dog cancer, and propose a regional modeling approach to account for different reporting practices across the study area.

We develop a case study based on the Swiss Canine Cancer Registry (SCCR), a unique dog cancer database that has been assembled for comparative investigations with humans. We model dog cancer incidences at the municipal level using demographic variables and indicators accounting for underreporting. Based on the model residuals, we decompose Switzerland into regions of similar model fit and build regional models of dog cancer incidences. We finally compare statistical performance and spatial distributions of model residuals to identify changes in the statistical associations across the study area. In doing this, we aim to demonstrate that underreporting challenges the use of statistical models of dog cancer incidences, and that a regional modeling approach improves statistical performance by mitigating spatial non-stationarity of statistical associations.

2. Data

The SCCR stores diagnostic records collected in Switzerland between 1955 and 2008, and it is in the process of being updated to more recent years. Comprising more than 120,000 records, this is the largest and most durable animal cancer database, to date. The present study is based on the 3,509 cancer examinations performed in 2008, which have been enumerated within the 2,351

Swiss municipal units to protect the privacy of the dog owners. We retrieved dog census data for the same year to assess the impact of demographic characteristics accountable for cancer development. This census has been established in 2006 following the nationwide obligation of dog registration. For our study purposes, we summarized information about age and sex of the dog population at the municipal level.

We used indicators of urban character and socio-economic status, as our prior research has shown that these variables successfully account for underreporting of dog cancer. The urban character is estimated based on human population densities in 2008, as provided by the Swiss Federal Statistical Office. The socio-economic status is derived from average national income tax information collected by the Swiss Federal Tax Administration in 2008. We also derived information on access to veterinary care from a hectometric distance-raster based on the addresses of the 938 veterinarians active in 2013. The distance-raster is constrained along roads to compute travel distances, and values are averaged to estimate access to veterinary care at the municipal level. We used more recent addresses of registered veterinarians because historical data are not available.

3. Methods

We model dog cancer incidences through a Poisson regression because this method is commonly used to identify risk factors for the spatial distribution of rare diseases (Frome 1983). We choose this model, in spite of possible over-dispersion, as the spatial distribution of model residuals can inform about potential misspecifications. We fit dog cancer incidences at the municipal level through the following variables: Dog Average Age, Female Dog Ratio, Average Income Tax, Human Population Density, and Distance to Veterinary Care. Dog Population is used as offset. Statistical performance is evaluated with the McFadden pseudo R-squared and the analysis of variance reduction, including a spatial decomposition based on model residuals (Cameron and Windmeijer 1997). We used Pearson residuals because absolute values exceeding 2 highlight lack of model fit.

The spatial distribution of model residuals is used to define regions of similar model fit, and inform about spatial non-stationarity of statistical associations (Brunsdon *et al.* 2008). We identify contiguous regions by means of a connectivity graph algorithm (minimum spanning tree), based on Queen contiguity to determine adjacent municipal units (Duque *et al.* 2007). The optimal number of regions is selected through the pseudo F-statistic, indicating the number of regions with maximum internal similarity and external dissimilarity (Ketchen and Shook 1996). We then fit models of dog cancer incidences for each region separately. We finally compare the statistical performance of the regional models with the global model and explore the spatial distributions of Pearson residuals to assess whether spatial stationarity in the statistical associations has improved.

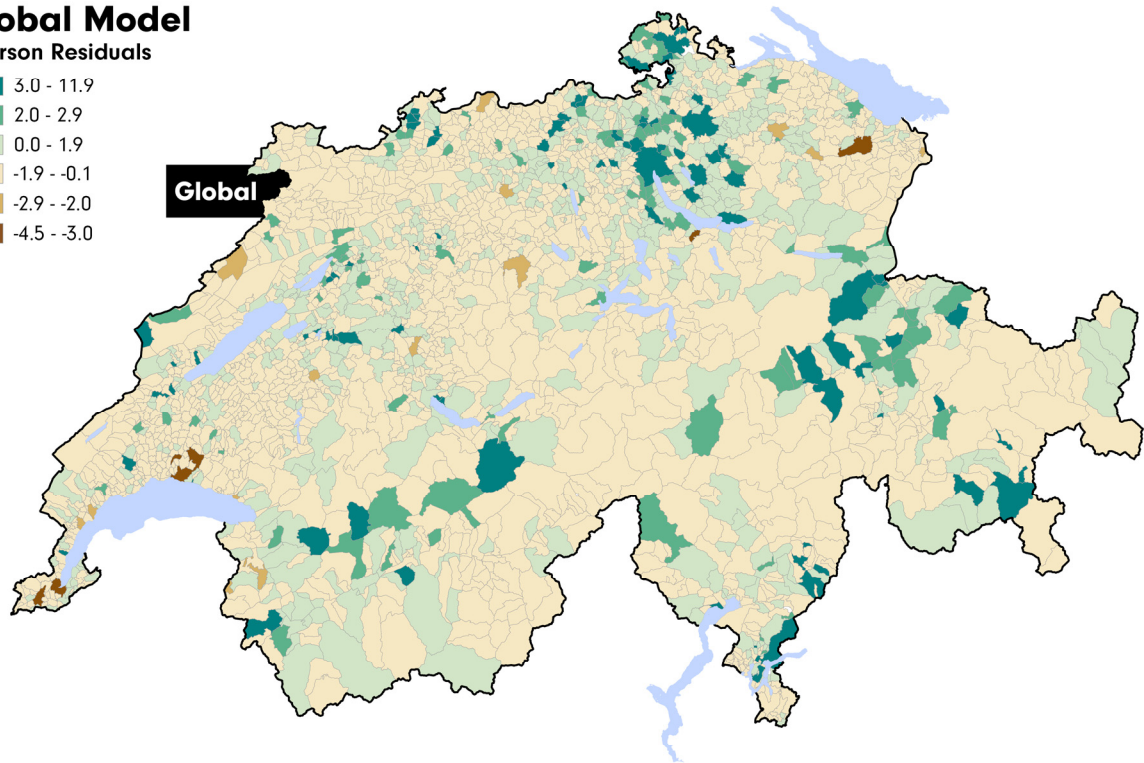
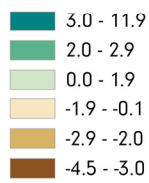
4. Results and Discussion

For the global model, the McFadden statistic shows a value of .50. This means that 50% of the total variability in dog cancer incidences is explained by the model. The variables accounting for underreporting of dog cancer are statistically significant ($p < .01$) and contribute to 87% of the overall variance reduction. The remaining 13% of variance reduction is explained by the demographic variables, which are also statistically significant ($p < .01$). Our results confirm the expected effects of indicators accounting for well-known sources of underreporting of dog cancer. Nevertheless, the spatial distribution of the model residuals presented in Figure 1a shows

several regions of poor model fit, suggesting that the global model might be affected by spatial non-stationarity of statistical associations.

a. Global Model

Pearson Residuals



b. Regional Models

Pearson Residuals

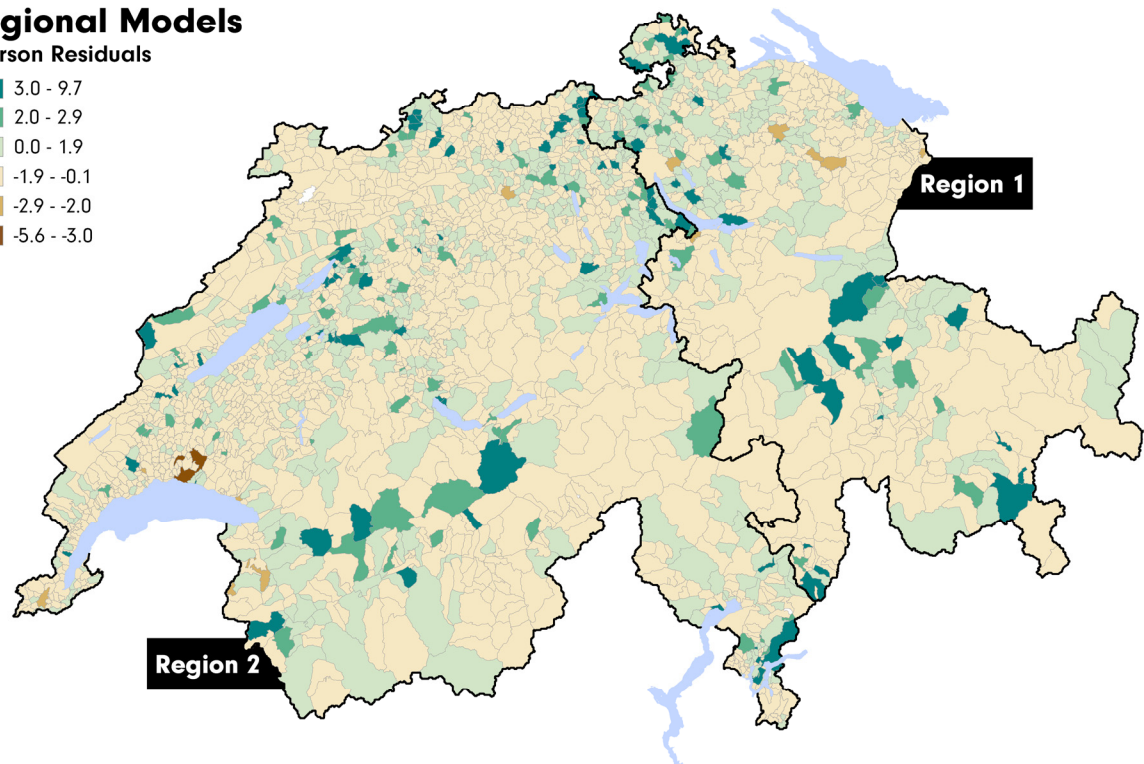
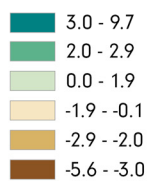


Figure 1. Spatial distribution of Pearson residuals for the global model (a) and the two regional models (b).

Our regional approach produced two macro-regions of distinct model fit: Region 1 and Region 2, consisting of 548 and 1,803 municipalities, respectively. When regional models of dog cancer incidences are fit to the two regions, most variables are statistically significant ($p < .01$) with a McFadden statistics of .58 for Region 1 and .46 for Region 2. Interestingly, the overall variance reduction explained by indicators of underreporting drops to 72% for Region 1, while it increases to 95% for Region 2. This confirms a decreased impact of underreporting of dog cancer for Region 1, but an increase for Region 2. Figure 1b presents the spatial distribution of Pearson residuals across both regions and shows a general improvement of model fit when compared with Figure 1a. This suggests that effects of spatial stationarity in statistical associations could be mitigated to some extent using the proposed regional modeling approach.

5. Conclusions and Outlook

The results of this case study demonstrate that the presence of different reporting practices across the study area challenges the use of statistical models for fitting dog cancer incidences. To address this issue, we propose a regional modeling approach to specifically delineate underreporting regimes across the study area. In our study, we identify two distinct regions with different characteristics. While model fit for Region 1 shows improved statistical performance, likely due to a reduced impact of underreporting, model fit for Region 2 deteriorates, suggesting increased underreporting and thus less reliable statistical associations. Still, employing two regional models reduces non-stationarity of statistical associations observed in the global model.

In future research, we aim to further refine our regional modeling approach by deepening the understanding of spatial non-stationarity of statistical associations linked to underreporting (see for example Leyk *et al.* 2012). In doing so, we intend to propose a new framework for systematically investigating regional changes in statistical associations across the study area, as a strategy to mitigate effects of spatial non-stationarity of statistical associations.

References

- Brunsdon C, Fotheringham AS and Charlton ME, 1996, Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity. *Geographical Analysis*, 28(4):281–298.
- Cameron AC and Windmeijer FAG, 1997, An R-squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models. *Journal of Econometrics*, 77(2):329–342.
- Duque JC, Ramos R and Suriñach J, 2007, Supervised Regionalization Methods: A Survey. *International Regional Science Review*, 30(3):195–220.
- Frome EL, 1983, The Analysis of Rates Using Poisson Regression Models. *Biometrics*, 39(3):665–674.
- Ketchen DJ and Shook CL, 1996, The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. *Strategic Management Journal*, 17(6):441–458.
- Leyk S, Norlund PU and Nuckols JR, 2012, Robust Assessment of Spatial Non-Stationarity in Model Associations Related to Pediatric Mortality due to Diarrheal Disease in Brazil. *Spatial and Spatio-temporal Epidemiology*, 3:95–105.
- Pinho SS, Carvalho S, Cabral J, Reis CA and Gärtner F, 2012, Canine Tumors: A Spontaneous Animal Model of Human Carcinogenesis. *Translational Research*, 159(3):165–172.