Echinococcosis surveillance: Bayesian time-space analysis of *Echinococcus multilocularis*-infections in foxes in Thuringia, Germany

C. Staubach[1], L. Hoffmann[2], V. Schmid[3], M. Ziller[1], K. Tackmann[1] and **F.J. Conraths**[1*]

Abstract

The prevalence of *Echinococcus multilocularis* in foxes in Thuringa, Germany, was monitored from 1990 through 2009, by sampling 26,220 foxes. Data were analysed in space and time using a hierarchical Bayesian model. The prevalence increased from 11.9% (9.9-14.0%) in 1990 to 42.0 (39.1-44.1%) in 2005. The infection had spread from North-western parts over the entire state by 2004. The model helped to overcome problems of missing data and varying sample sizes in different spatial and temporal units and to evaluate the monitoring strategy.

Keywords: *Echinococcus multilocularis*, alveolar echinococcosis, fox, monitoring, Bayesian statistics.

Introduction

Alveolar echinococcosis is a rare human disease that ends often lethal if left untreated. It is caused by the larval stage of the cestode parasite, *Echinococcus multilocularis*. In central Europe, the main definitive host of *E. multilocularis* is the red fox (*Vulpes vulpes*), but occasionally the parasite is also found in dogs and cats [1, 2]. Raccoon dogs (*Nyctereutes procyonoides*), which invaded large territories in Europe in recent decades, represent a new definitive host whose epidemiological role is not yet clear. To assess the risk for human infection, it is important to monitor the epidemiological situation of *E. multilocularis* in its definitive hosts in time and space.

The spatial and temporal analysis of wildlife diseases is particularly challenging, because the population size of the target species is usually unknown, random sampling of animals is often not possible, and the spatial boundaries of epidemiologic units may be artificially chosen and not relevant to disease spread [3]. Hunting foxes, the most important definitive hosts of *E. multilocularis* in Europe, is not a random selection. With respect to their spatial origin, samples of foxes obtained by hunting are usually heterogeneous and not randomly distributed.

Moreover, random variation in the estimated prevalence per geographic unit is possible, so that the significance of spatial and temporal changes is difficult to assess. Data may be missing for several spatial or temporal units leading to increased uncertainty and different sample sizes may be available from the spatial units. The sample size in some spatial units or strata may even be too low to obtain reliable prevalence estimates [3].

To overcome these constraints, a hierarchical Bayesian space-time model [3] was used and the distribution of the model parameters and their variability estimated by using a Markov chain Monte Carlo (MCMC) simulation technique, on the basis of the sample size, number of cases per spatial unit and time interval, and the adjacency matrix of the municipalities respectively.

Materials and methods

A total of 26,220 foxes that were hunted or found dead in the state of Thuringia, Germany, between 1990 and 2009 were examined using an OIE/WHO standard protocol, the intestinal scraping technique [4]. Data on the foxes including the date of hunting/death, location (local community; district) and the result of the examination for *E. multilocularis* were recorded [5] and exported into a data base (Microsoft Access) that could be linked to a geographical information system. Further data management and mapping was done in ArcGIS Arcview 9.3 (ESRI, Redlands, CA, USA).

A hierarchical Bayesian space-time model was set up to analyse the data from the state of Thuringia from 1990-2009, and to test for significance of temporal and spatial effects in the study area. The model is based on a Bernoulli observation model and is an extension of an existing hierarchical model proposed for disease mapping [Staubach *et al.*, 2002], with the addition of time effects. In this model, the number of cases x is binomial distributed in each spatial unit i at each time t

$$x_{it} \sim \text{Bin}\,(n_{it}\,,\,\pi_{it})$$

where n is the sample size and π the prevalence. The parameter π it is modelled with a logistic link

$$\eta_{it} = \log\,(\pi_{it} / 1\text{-}\,\pi_{it}) = \mu + \upsilon_i + \nu_i + \phi_t$$

where

μ is the intercept (average logit prevalence),

υ_i is a structured spatial effect in the area *i*,

ν_i is an unstructured random effect, taking into account the heterogeneity mainly based on the sample size per spatial unit

ϕ_t is a time effect, as deviation from the average logit prevalence at the time *t*.

The effect υ_i covers spatial dependencies, which can be described by a Gaussian Markov Random field (GMRF):

$$\upsilon_i \sim \text{N}\,(\upsilon_{\text{neigh}} / \text{k}\,,\,\sigma_\upsilon / k\,)$$

with $\upsilon_{\text{neigh}} = \upsilon_i$

and k = number of neighbouring units *j* of unit *i*.

Only the first neighbours of each spatial unit were taken into account. The unstructured random effect ν_i has a priori a Gaussian distribution:

$$\nu_i \sim \text{N}\,(0\,,\,\sigma_\nu).$$

In units with no observation, ν_i is set to 0.

[1] Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Wusterhausen, Germany - * franz.conraths@fli.bund.de
[2] Thuringian State Authority for Food Safety and Consumer Protection, Bad Langensalza, Germany.
[3] Ludwig-Maximilians-University, Department of Statistics, Munich, Germany

The time effect ϕ_t is a special case of the random walk (RW), that is, the time effect has a priori a Gaussian distribution. We assume a linear trend, which can be described as random walk of second order:

$\phi_t \sim N(2\,\phi_{t-1} - \phi_{t-2}, \sigma_\phi)$.

For the unknown variance parameters, Inverse Gamma priors with the following hyper parameters were used:

$\sigma_u \sim$ Inv-gamma (1 , 0.001)

$\sigma_v \sim$ Inv-gamma (1 , 0.05), and

$\sigma_\phi \sim$ Inv-gamma (1 , 0.05).

The joint distribution of the full Bayesian model is analytically intractable. Therefore, a MCMC algorithm was used in form of a combination of Metropolis-Hastings and Gibbs steps in order to estimate the distribution of the model parameters and their variability on the basis of the sample size, the number of cases per spatial unit and time interval, and the adjacency matrix of the municipalities in a C++ program with Fortran code [3, 6, 7]. After 1,000 iterations, which were considered as burn-in phase, every 50th iteration of the next 50,000 iterations was saved as a sample from the posterior distribution.

Statistical analysis and evaluation of the model output was done in the R software package (R Development Core Team, 2010). Convergence of the MCMC chain was assessed using standard diagnostic plots and tests in the CODA library of R.

Result

In a total of 26,220 foxes sampled in Thuringia between 1990 and 2009, 6853 animals were found infected with *E. multilocularis*. An explorative analysis of sample sizes at district level for 1990-2009 showed that the sampling sizes per spatial and time unit were often too low to obtain reliable prevalence estimates (Figure 1), although efforts had been made to make the sample size as large as possible. Moreover, the sample sizes obtained from different spatial units varied considerably over time.

Spatial analysis of the data using the Bayesian model clearly showed a substantial expansion of the area where *E. multilocularis* infections occurred in foxes. In 1990, the highest prevalence was estimated in the Northwest of Thuringia, whereas most of the eastern districts were still free from the parasite or the estimated prevalence of infected foxes was lower than 10%. Until 1993, the prevalence had increased in the western districts. In 1998, the estimated prevalence in two south-eastern districts, where only 11 or 1 infected foxes had been detected in the years before, had risen to more than 30%. By 2004, the infection had spread over the entire state. This situation remained unchanged until 2009, although the estimated prevalence varied in some districts over time. The results show that the prevalence increased in particular in the western part of the state and that the infection spread in eastern direction.

Temporal analysis indicates a clear increase of the estimated prevalence from 11.9% (95% confidence interval 9.9-14.0%) to values between 30 and approximately 40 percent in recent years with a peak of 42.0 (95% confidence interval 39.1-44.1%) in 2005. The time effect as analysed by the Bayesian model shows a significant increasing trend during the whole observation period, even if years with lower sample sizes are taken into account (Figure 2).

Figure 1: Distribution of test results on foxes for *E. multilocularis* per spatial unit and year.

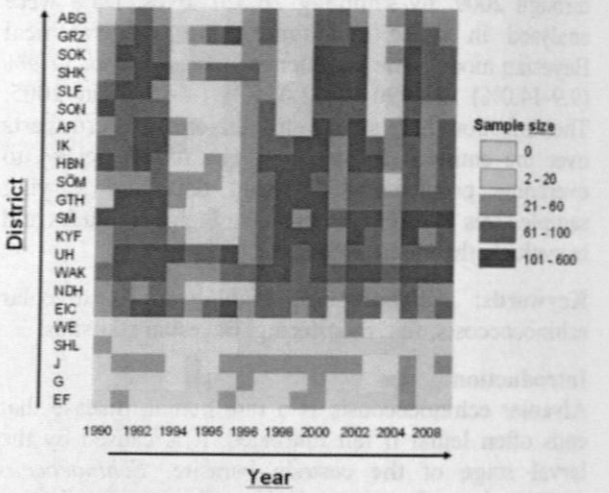
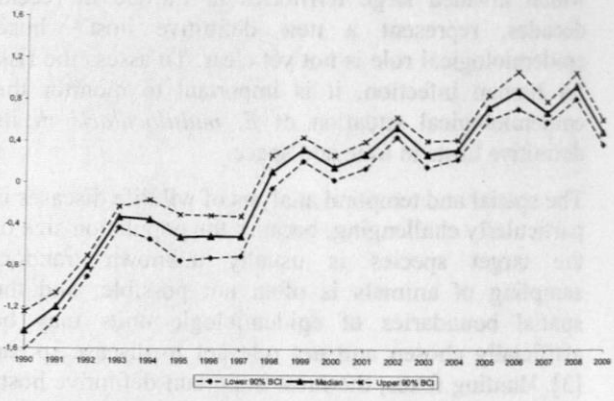


Figure 2: Median time effect ϕ_t and 90 Bayesian confidence intervals (BCI) of the Bayesian model for the period 1990-2009

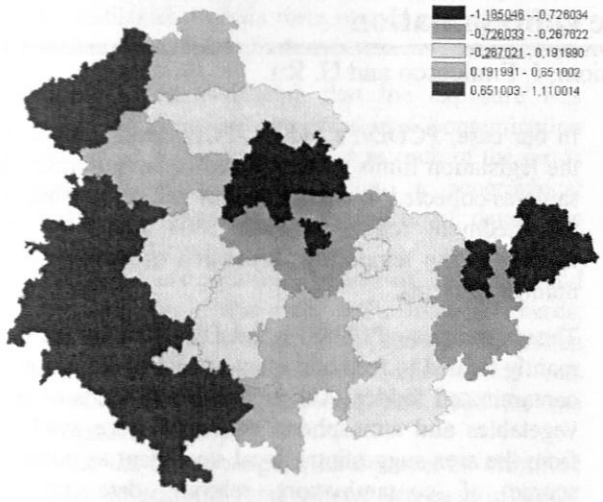


An increased risk of infection adjusted for the temporal trend in the Bayesian model can be observed particularly in the western districts of Thuringia, while a negative spatial effect is evident in the eastern part of the study region (Figure 3).

Discussion

Human alveolar echinococcosis is an emerging disease in Europe [8]. Surveys conducted in central Europe in the past 20 years revealed that *E. multilocularis* is more widely distributed than previously anticipated [2, 8]. The prevalence of *E. multilocularis* in its definitive hosts can vary widely between countries and regions from about 1 to over 60% [9]. Small endemic foci with a diameter of approximately 25 km can exist [5].

Figure 3: Median structured spatial effect v_i per district of the Bayesian model for the period 1990-2009



The spatial and temporal dynamics of the infection in a number of affected countries have been controversially discussed but the quality of the availably monitoring data was often not sufficient to assess whether the prevalence was increasing in particular regions or if endemic areas were extending. We addressed this issue by analysing monitoring data from Thuringia, Germany, as an example, since a comprehensive set of monitoring data for the period from 1990 through to 2009 was available and because circumstantial evidence suggested an increase of the prevalence, at least in some areas, and a widening of the endemic area.

The analysis of variation of the risk for a given disease in space and time is a key issue in descriptive epidemiology. An infectious disease typically spreads via contact between infected and susceptible animals. Often administrative structures are the only feasible way to map samples in wildlife population surveys. Therefore, available data are frequently aggregated on the level administrative units and regular time intervals. However, when data are stratified to reflect a potential space-time variation, maximum likelihood estimates of the area-specific prevalence and of its time-trend can be seriously affected by random variation due to the low number of cases and the corresponding sample size in a given unit of space and time [3]. Furthermore, the detection of spatial patterns is complicated by missing data as it is nearly impossible to cover of all areas and time intervals with samples in wildlife surveys. Therefore, the mapping of raw surveillance data may lead to false interpretations of disease clusters, disease-free areas and time-trends [3]. We therefore applied a model which is an extension of an existing hierarchical spatial model proposed for disease mapping that also accounts for time effects and potentially allows the inclusion of covariates (*e.g.* age, landscape proportions) and space-time interaction terms. The model uses spatial and temporal dependencies to make estimates in space-time units with low sample size more robust, and even allows estimating the prevalence in units with no samples at all if samples from neighbouring units exist.

By using an auxiliary variable approach [7], the MCMC simulation becomes more efficient [6] than in the original approach [3]. As most full conditional distributions in this model are multivariate Gaussian, we used a block update algorithm [10] to decrease simulation error and to save time. It has been shown, however, that approximate Bayesian inference using integrated nested laplace approximation (INLA) might reduce the computing time dramatically without increasing the error of the estimates significantly [11]. We are currently investigating if such algorithms can be used to estimate the parameters of our model.

We believe that the model we used to evaluate the data is widely applicable and can be applied to analyse data sets with gaps and variable sample sizes per spatial and temporal unit.

References

1. Thiess A. *et al.* - Berl. Münch. *Tierärztl. Wochenschr.*, 2001, **114**, 273-276.
2. Romig T. *et al.* - *Parasitol. Int.*, 2006, **55**, S187-191.
3. Staubach C. *et al.* - *Prev. Vet. Med.*, 2006, **56**, 75-87.
4. Deplazes P., Eckert J. - *Appl. Parasitol.*, 1996, **37**, 245-252.
5. Tackmann K. *et al.* - *Epidem. Infect.*, 1998, **120**, 101-109.
6. Schmid V. - http://edoc.ub.uni-muenchen.de/3000/1/Schmid_Volker.pdf, 2004.
7. Holmes C.C., Held L. - *Bayesian Analysis 1*, 2006, 145-168.
8. Torgerson R. *et al.* - PLoS Negl. *Trop. Dis.*, 2010, **4**, e722.
9. Eckert J. *et al.* - *Int. J. Parasitol.*, 2000, **30**, 1283-1294.
10. Rue H. - *J. Roy. Stat. Soc.*, 2001, **B 63**, 325-338.
11. Rue H. *et al.* - *J. Roy. Stat. Soc.*, 2009, **B 71**, 319-392.

Acknowledgements

We gratefully acknowledge our colleagues in the district veterinary authorities of Thuringia and the hunters who submitted foxes for this study. We are also indebted to the government of the State of Thuringia for funding the sampling and testing of the foxes.