*Epidemiol. Infect.* (2013), **141**, 1857–1866. © Cambridge University Press 2012
doi:10.1017/S095026881200249X

# Climatic, ecological and socioeconomic factors as predictors of Sindbis virus infections in Finland

K. JALAVA[1]*, J. SANE[2], J. OLLGREN[1], R. RUUHELA[3], O. RÄTTI[4], S. KURKELA[2,5], P. HELLE[6], S. HARTONEN[3], P. PIRINEN[3], O. VAPALAHTI[2,5,7] AND M. KUUSI[1]

[1] *Department of Infectious Disease Surveillance and Control, National Institute for Health and Welfare, Helsinki, Finland*
[2] *Department of Virology, Haartman Institute, Faculty of Medicine, University of Helsinki, Helsinki, Finland*
[3] *Finnish Meteorological Institute, Helsinki, Finland*
[4] *University of Lapland, Arctic Centre, Rovaniemi, Finland*
[5] *Department of Virology and Immunology, HUSLAB, Helsinki, Finland*
[6] *Finnish Game and Fisheries Research Institute, Helsinki, Finland*
[7] *Department of Veterinary Biosciences, Faculty of Veterinary Medicine, University of Helsinki, Finland*

*Received 9 May 2012; Final revision 10 September 2012; Accepted 17 October 2012; first published online 15 November 2012*

## SUMMARY

Mosquito-borne Sindbis virus (SINV) causes rash-arthritis syndrome in Finland. Major outbreaks with approximately 7-year cycles have caused substantial burden of illness. Forest dwelling grouse are suspected to be amplifying hosts, with the infection transmitted to humans by mosquito bites. SINV infection surveillance data for 1984–2010 were used to create a negative binomial hurdle model, with seasonality, long-term cycles, climatic, ecological and socioeconomic variables. Climatic factors during early summer and amount of snow in April described the occurrence and incidence of SINV infections. Regulated water shore and hatch-year black grouse density described the occurrence, while population working in agriculture, agricultural land (negative) and income (negative) described the incidence of the disease. The prediction for 2009 was 85 cases (95% prediction interval 2-1187), while the actual occurrence was 106. We identified novel and known risk factors. The prevention of SINV infections in regulated water areas by infected mosquito populations should be targeted.

**Key words**: Climate – impact of, emerging infections, epidemiology, modelling, virology (human).

## INTRODUCTION

Mosquito-borne Sindbis virus (SINV) of the genus *Alphavirus* is the causative agent of rash-arthritis syndrome in Finland [1, 2]. The emergence of a Chikungunya virus, a globally important mosquito-borne alphavirus, has increased the interest in climatic factors [3]. SINV antibodies have been detected globally in humans and birds. The clinical SINV infection is known as Pogosta disease in Finland, Ockelbo disease in Central Sweden and Karelian fever in Russian Karelia [4, 5], with similar microbiological and epidemiological characteristics [6]. SINV has a

* Author for correspondence: Dr K. Jalava, Department of Infectious Disease Surveillance and Control, National Institute for Health and Welfare, Mannerheimintie 166, 00300 Helsinki, Finland.
(Email: katri.jalava@thl.fi)

local circulation in Northern Europe and no significant changes in the genotype have been observed during the past decades [7].

The laboratory diagnosis of SINV infection is by serology [8]. Most cases occur between ages 40–60 years with female predominance [9, 10]. The incubation period of SINV infection is 4 days (range 2–18 days), and the disease is characterized by fever, myalgia, rash and joint symptoms [1, 2]. About 25–50% of cases have severe long-term articular symptoms [1, 11, 12]. The highest seroprevalence has been reported from Eastern Finland, while the disease is virtually non-existent in some parts of the country [9]. Major SINV infection outbreaks have occurred mostly in 7-year cycles: in 1981, 1988, 1995 and 2002 [9]. Cases occur exclusively within a seasonal cycle from the end of July until October [9]. The disease has probably only relatively recently (less than five decades ago) been introduced to Finland [9, 13], as no cases were identified and no SINV seroprevalence in animals or humans has been observed before 1965 [9, 13]. The major public health burden is due to the high number of asymptomatic infections and articular symptoms [9].

SINV infection is transmitted to humans exclusively by mosquito bites [2]. Wild birds have been suspected as viral reservoirs and as amplifying hosts, especially Passeriformes of the genera *Turdus* and *Fringilla* [14]. The introduction of SINV to Northern Europe probably occurred through migratory birds from South Africa [10]. Forest-dwelling grouse species (Tetraonidae), black grouse (*Tetrao tetrix*) and capercaillie (*Tetrao urogallus*) had high SINV antibody titres following outbreaks [9, 10]. Experimentally SINV-infected birds have had sufficient titres to infect mosquitoes [14]. The temperature between May and July and the depth of the snow cover in the preceding winter have coincided with increasing numbers of SINV infections in the following July–September [9]. Late summer *Culex* and *Culiseta* mosquito species are considered to be the primary vectors for SINV infection [15], although the more human-adapted *Ochlerotatus* species may also play a role [5, 7]. There is a paucity of up-to-date information about the distribution of the mosquito species in Finland, yet all the above-mentioned species are considered abundant in most of the country [7, 16].

A time-series regression model with monthly indicators and sinusoidal terms for annual cycles has been used with time-series count data [17]. One important feature of the data was the excess of zeros and therefore, a two-part hurdle or zero-inflated model that models the zero counts and positive counts separately may be necessary [18, 19]. We have previously applied a hurdle model with Verotoxigenic *Escherichia coli* surveillance data with low case counts, which has distinguished between significant risk factors for the occurrence (binary part) and incidence of disease (count part) [20]. These types of models have also been applied to cholera prevalence [21] and bacterial counts [22].

The aim of this study was to model SINV infections by healthcare district (HCD) in Finland between 1984 and 2010, taking into account seasonal monthly fluctuation, seasonality, long-term cycles and time lags in the observations to identify risk factors and predict the number of SINV infections. Climatic, ecological and socioeconomic data were used as explanatory variables in a hurdle model to identify factors for the occurrence and incidence of SINV infection. Furthermore, we used the obtained model to predict the cases for the years 2009, 2010 and 2011, in order to further justify the model's applicability.

## MATERIALS AND METHODS

### Serology and surveillance data

The laboratory diagnosis of SINV infection is based on enzyme immunoassays and the haemagglutination inhibition test [8]. A case was defined as a person with serologically confirmed acute SINV infection by the Department of Virology, University of Helsinki (Haartman Institute) during the period 1984–1994. This included about 70–80% of the laboratory-confirmed SINV infections in Finland. All cases from Helsinki University Hospital Laboratory (HUSLAB, or its predecessors) and other clinical microbiology laboratories were reported to the National Infectious Disease Register (NIDR) after its establishment in 1995. These two datasets were combined, collating information on sex, year of birth, place of residence or a proxy by place of treatment and date of sampling. The date of sampling was established as the latter if paired samples were taken.

### Explanatory variables

We selected a range of explanatory variables based on a knowledge of known risk factors for SINV infections [2, 9]. Weather variables used in the models were precipitation, temperature and snow cover from the

database of the Finnish Meteorological Institute. Two types of weather variables were used; first, monthly precipitation and mean temperature for May, June, July and August as a mean for each HCD calculated from the gridded datasets [23, 24] and depth of snow cover on 15 March and 15 April, and second, weather variables with different monthly lags with respect to the reporting month of the case: precipitation and temperature ($t-1$ to $t-6$ months) and depth of snow cover ($t-5$ to $t-7$ months). The annual wildlife data for the density of grouse adults and juveniles (hatch-year grouse) in August were also included [25]. We also included more time-independent variables, such as habitation, agricultural, and land usage variables, as shown in Supplementary Table S1, with all explanatory variables adjusted by HCD ($n=21$), year and month.

### Hurdle model

We applied a two-part hurdle model with log of lag 1 of the outcome as an autoregression replacement term for SINV infection surveillance data and created a model with covariates explaining the occurrence of the cases in the binary part of the model and the incidence in the negative binomial regression part. With an excess of zeros in the data for SINV infections (mean number of cases 0·52 per HCD, month and year with variance of 18·1), we applied a negative binomial distribution-based hurdle model [19] as previously described in detail [20].

Serologically diagnosed acute SINV infections were included in the analysis as outcome variables for the 26-year follow-up period by HCD, month and year of diagnosis based on date of sampling. We included data from June to October, as there were essentially no cases outside this time-frame. We performed a univariable analysis with all explanatory variables in the hurdle model applying a clog log link function and using the data for 1984–2009. Those variables with $P$ values of $<0{\cdot}20$ in the univariate analysis were selected for the multivariable model. Of the correlated variables with correlation coefficients of $>0{\cdot}80$, only those with $P$ values of $<0{\cdot}20$, i.e. the most significant ones, were included in the final model within each group as shown in Supplementary Table S1. To identify explanatory variables in the final multivariable model, we used forward selection according to Akaike's Information Criteria (AIC) due to its ability to increase goodness of fit while simultaneously penalizing for increasing the number of estimated parameters to avoid overfitting. We also tested the linear trend with splines, estimated the sinusoidal terms, and performed collinearity diagnostics to identify possible multicollinearity for the final model. The autocorrelation and partial autocorrelation for scaled residuals by HCD with lags 1–10 (2 years) were checked by visual inspection for any remaining autocorrelation. We also tested for the spatial autocorrelation of the residuals for each year and month. The permutation test for Moran's $I$ statistics (Bonferroni) was used to assess the possible spatial autocorrelation in the scaled residuals by time. We evaluated further the spatial accuracy of the model by calculating the point estimates for the predictions of each HCD for 1984–2009 using other HCD as learning data. The comparison between Poisson *vs.* negative binomial hurdle model and hurdle *vs.* zero-inflated model was performed with Vuong's test [26]. The statistical packages included R version 2.14.1 (R Foundation, Austria), SPSS version 19 (SPSS Inc., USA) and Stata version 9.2 (StataCorp., USA).

### Prediction of the occurrence of acute SINV infections

Data for the periods 1984–2008, 1984–2009 and 1984–2010 were used with the obtained model to predict cases for 2009, 2010 and 2011, respectively. As there were no packages available to calculate the prediction intervals (PI) for the hurdle model in R, we calculated predicted point estimate values and standard errors by simulation. Subsequently the partial likelihood estimators were calculated for the parameters. We assumed that the distribution of the prediction estimates could be approximated by normal distribution [27]. The predicted estimates and PIs were sampled 100 times using a normal distribution to simulate a fiducial distribution for parameters for the years 2009, 2010 and 2011. We used only marginal normal distributions for the parameters (R code in Supplementary Material S2, abbreviations in Table 1).

## RESULTS

### Descriptive epidemiology

There were 339 cases with acute SINV infection diagnosed by the Department of Virology, University of Helsinki (Haartman Institute) for the period 1984–1994. Altogether 3042 SINV infection cases

Table 1. *Significant explanatory variables in the multivariable hurdle model*

| Explanatory variable (Supplementary Material S2 abbreviation) | Time period (lags *t* in months if applicable) | Estimate (*P* value); hurdle model | | Relative change (%) in approximated odds for the occurrence (zero part)*: increase in the standardized incidence ratio (negative binomial part)† when changing the respective variable by 1 % | |
|---|---|---|---|---|---|
| | | Zero part | Negative binomial part | Zero part | Negative binomial part |
| Income per household (mean) (Income) | *t* | n.a. | −2·5 (<0·001) | n.a. | −2·5 |
| Agricultural area per total area (Agricultural_land) | *t* | n.a. | −6·1 (<0·001) | n.a. | −5·9 |
| Number of cases in previous month (ln transformed), (Cases_1month_ln) | *t* − 1 | 0·53 (<0·001) | 0·42 (<0·001) | 0·4 | 0·4 |
| Proportion of population working in agriculture (Agricultural_work_population) | *t* | n.a. | 20·6 (<0·001) | n.a. | 22·9 |
| Depth of snow cover on 15 April (Snowcover_april) | April | 0·019 (<0·001) | 0·018 (<0·001) | 0·01 | 0·02 |
| Monthly mean temperature in June (Temperature_June) | June | 0·19 (<0·001) | 0·24 (<0·001) | 0·1 | 0·2 |
| Monthly precipitation in June (Rainfall_June) | June | 0·0074 (<0·05) | 0·013 (<0·001) | 0·005 | 0·01 |
| Monthly mean temperature in May (Temperature_May) | May | 0·19 (<0·001) | 0·35 (<0·001) | 0·1 | 0·4 |
| Length of regulated water shore per total area (Reg_watershore) | *t* | 2·06 (<0·001) | n.a. | 1·5 | n.a. |
| Monthly precipitation in previous month (Rainfall_month_t_1) | *t* − 1 | 0·0048 (<0·05) | n.a. | 0·004 | n.a. |
| Hatch-year black grouse density (Hatch_year_grouse) | *t* | 0·14 (<0·001) | n.a. | 0·1 | n.a. |
| sin12 | | n.a. | −8·7 (<0·001) | n.a. | −8·3 |
| cos12 | | n.a. | −2·3 (<0·001) | n.a. | −2·3 |
| sin24 | | n.a. | 0·20 (0·22) | n.a. | 0·2 |
| cos24 | | n.a. | 0·61 (<0·001) | n.a. | 0·6 |
| sin36 | | −0·19 (<0·05) | −0·34 (<0·001) | −0·1 | −0·3 |
| cos36 | | −0·098 (0·16) | −0·10 (0·13) | −0·07 | −0·1 |
| sin60 | | −0·13 (0·077) | n.a. | −0·09 | n.a. |
| cos60 | | 0·098 (0·18) | n.a. | 0·07 | n.a. |
| sin72 | | n.a. | 0·33 (<0·05) | n.a. | 0·3 |
| cos72 | | n.a. | 0·10 (0·37) | n.a. | 0·1 |
| sin84 | | −0·43 (<0·001) | −0·35 (<0·001) | −0·3 | −0·3 |
| cos84 | | 0·17 (<0·05) | −0·11 (0·36) | 0·1 | −0·1 |
| July, month of notification (month2) | | 0·66 (0·059) | n.a. | 0·5 | n.a. |
| August, month of notification (month2) | | 3·25 (<0·001) | n.a. | 2·4 | n.a. |
| September, month of notification (month2) | | 3·52 (<0·001) | n.a. | 2·6 | n.a. |

Table 1 (*cont.*)

| Explanatory variable (Supplementary Material S2 abbreviation) | Time period (lags *t* in months if applicable) | Estimate (*P* value); hurdle model | | Relative change (%) in approximated odds for the occurrence (zero part)*; increase in the standardized incidence ratio (negative binomial part)† when changing the respective variable by 1 % | |
|---|---|---|---|---|---|
| | | Zero part | Negative binomial part | Zero part | Negative binomial part |
| October, month of notification (month2) | | 1·5 (<0·001) | n.a. | 1·1 | n.a. |
| Linear trend, 1984–1995 | | 2·72 (<0·001) | 2·32 (<0·001) | 2·0 | 2·3 |
| Linear trend, 1995–2002 | | 2·95 (<0·001) | 2·69 (<0·001) | 2·2 | 2·7 |
| Linear trend, 2002–2009 | | 2·07 (<0·001) | 2·5 (<0·001) | 1·5 | 2·5 |

n.a., Not applicable.
* $\exp(0{\cdot}01 \times \beta_x/1{\cdot}386) - 1$.
† $\exp(0{\cdot}01 \times \beta_x) - 1$.

were reported to NIDR for the period 1995–2010. Major outbreaks of SINV infection occurred in a cyclic manner with the largest outbreak recorded in 1995 (1310 cases) (Fig. 1). Of all cases, 3320 (98 %) occurred between the end of July and October. The first cases were detected every year around week 30 (end of July), peaking in week 33 (mid August) and declining by weeks 37–47 (September–October). There is a delay of about 2 days between onset of symptoms and first medical contact [2]. Half of the cases, 1965 (58 %) were female, while 1135 (34 %) were identified in the 45–54 years age group. Cases were geographically clustered in central latitudes of approximately 61°–64° (Fig. 2).

## Hurdle model

The splines for the piecewise linear trend with knots at years 1995 and 2002 (major outbreak years), several sinusoidal sin/cos terms, and month of reporting were significant, as shown in Table 1. As some of the variables were most likely to be on the same pathway for infection, the ecological variables (grouse and water-related variables) were used mostly in the binary part and population variables (income, working in agriculture) in the negative binomial part, as judged by AIC. In the multivariable analysis, a high monthly mean temperature in May and June, high monthly precipitation in June, thick snow cover in April (melting waters) and a high number of cases in the previous month (natural logarithm) were positively significant both for occurrence of the disease (zero part) and incidence (negative binomial part) of SINV infection. The early summer weather conditions and depth of the previous winter's snow cover probably predict the number of mosquitoes in late summer. In addition for the occurrence of the disease, hatch-year black grouse density, regulated water shore area and previous month's precipitation were significant.

In addition to the those mentioned above, the most significant variables for the incidence of the disease were the level of income (negative), proportion of population working in agriculture (positive), and proportion of agricultural land (negative). The people most likely to become exposed are those working in agriculture or picking berries or mushrooms or hunting in the forest, possibly reflecting a lower level of income. The negative binomial-based hurdle model was chosen as best fitting the data based on Vuong's test, the AIC for the fully fitted model was 3866. The residuals were satisfactory for autocorrelations and

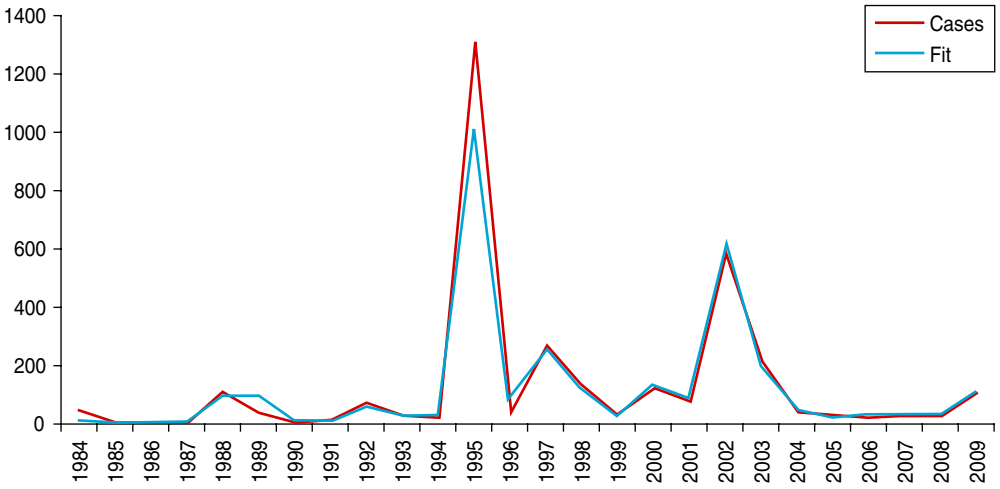


**Fig. 1** [*colour online*]. SINV infection cases and the fit of the final model in Finland for the period 1984–2009. The fit was calculated from the original model by healthcare district, month and year of diagnosis, subsequently these were summed up by year for the whole of Finland.

partial autocorrelations as analysed by visual inspection for any lags over the 95% confidence intervals. There was essentially no remaining spatial autocorrelation in the scaled residuals. The predictions by HCD were within the magnitude of the true total counts and therefore considered satisfactory (Table 2). The fit of the final model is presented in Figure 1.

### Prediction of cases for 2009–2011

Using the hurdle model for the period 1984–2009, a prediction for the number of cases for 2009 was 85 (95% PI 2-1187) cases, with 106 actual cases occurring. The prediction for 2010 was 37 cases (95% PI 5-241), with an actual occurrence of 54 cases and for 2011, 44 cases (95% PI 11-392), with an actual occurrence of 63. It should be noted that the predictions can be done only in the preceding month.

## DISCUSSION

This study identified ecological cycles and variables explaining the occurrence and incidence of SINV infections and provided accurate predictions. SINV infection dynamics between 1984 and 2011 were characterized by regular annual cycles between late July and October with larger outbreaks in 1988, 1995 and 2002. The disease most likely spreads between black grouse, mosquitoes and humans when suitable climatic conditions for the reproduction of mosquitoes occur, such as warm temperature and high precipitation with thick snow cover during the previous winter. These conditions are likely to be met in the area of Central-Eastern Finland. The development of infected mosquitoes is probably further expedited by the amount of regulated waters. Hatch-year black grouse may be one of the main amplification hosts for SINV. People are likely become infected during outdoor activities when exposed to infected mosquitoes. As SINV surveillance data were characterized by an overdispersion of zeros, we were able to model SINV infections by applying a hurdle model.

The higher occurrence of SINV infections in middle-aged women was as expected [2]. Cases are known to cluster in Eastern Finland [9, 28], but we found them also concentrated within Central Finland, indicating the importance of local geographical and climatic factors for the incidence of the disease. Ockelbo disease also clusters in certain latitudes in Sweden [6] and Karelian fever in Russia [5]. However, host genetic factors may also contribute to the geographical distribution of clinical SINV infections, producing potentially more clinical disease in Eastern Finland [29]. The study combined two datasets, one prior to initiation of surveillance of infectious diseases in Finland, and the other the NIDR data from 1995 to 2010. We tested the distribution of cases by HCD during these two periods by $\chi^2$ test and found that they were not comparable (data not shown), yet the number of cases during the early period was only about 10% of the total number of cases. This was mostly due to the history of the disease as it was initially discovered in one HCD only in the beginning of

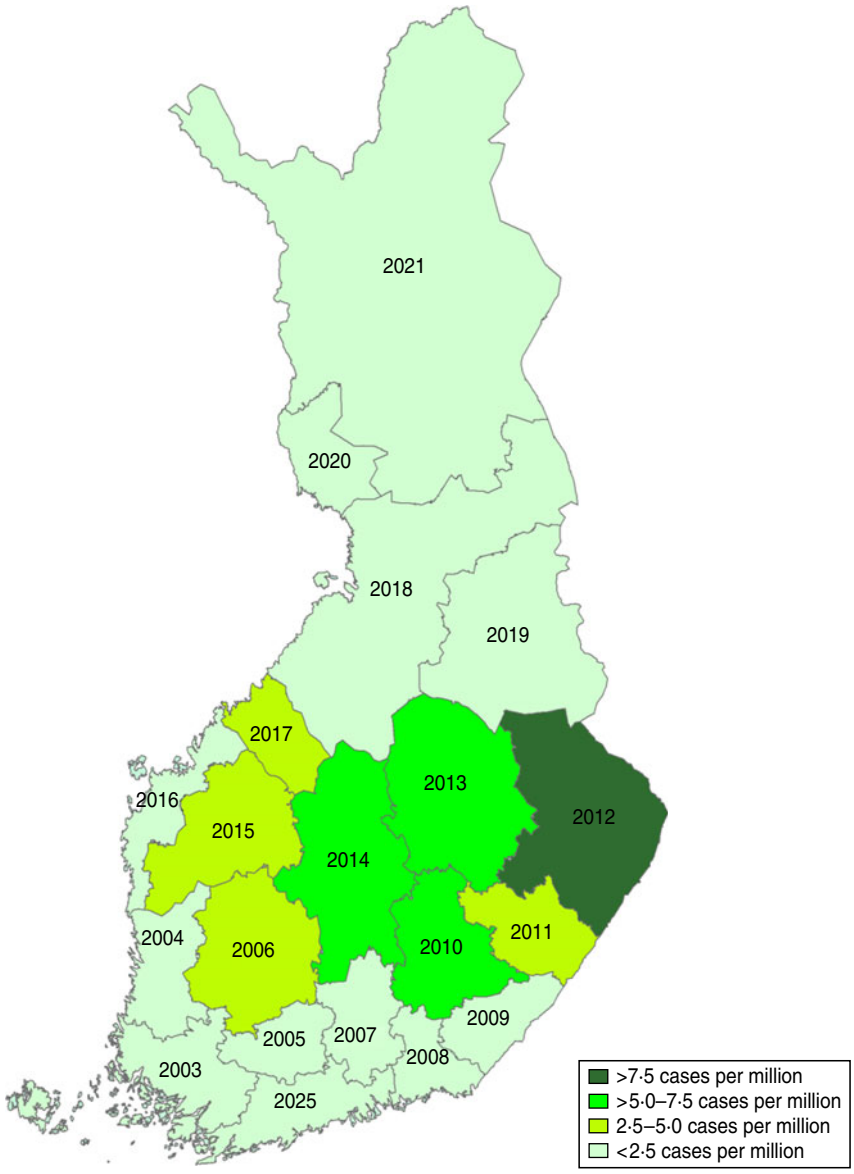


**Fig. 2** [*colour online*]. Incidence of serologically confirmed SINV infections in Finland in the period 1984–2009 by healthcare district.

1980s, so most of the early cases stem from that area. However, when the hurdle model was applied to the NIDR data only (1995–2009), the model remained essentially the same, temperature in May and sin/cos72 terms were not significant (data not shown).

The seasonal timing of the cases was identical each year, indicating a biological explanation for the amplification of the virus. We suspect that this may be due to hatching of black grouse chicks. Our data indicate that suitable climatic conditions enable the development of mosquito populations, allowing effective transmission between hosts, including amplification in hatch-year black grouse, thereby increasing the number of infective mosquito vectors that subsequently infect humans. This is also supported by a recent study on the prediction of tularaemia outbreaks by mosquito surveillance data where similar weather variables explained the development of mosquito populations [30]. More virological and epidemiological data is needed to validate this finding.

We expected weather variables to be important for the epidemiology of SINV infection as high temperature and increased precipitation have been shown to be important variables for SINV infections in South Africa [31, 32]. The early summer weather variables and melting water from the previous winter's snow cover were significant for both parts of the model, indicating their role in the occurrence and incidence of

Table 2. *Point estimates of the predictions of SINV infection cases by healthcare district for 1984–2009 between July and October*

| Healthcare district | Point estimate of prediction of cases (true number of total count) |
|---|---|
| 2003 | 118 (64) |
| 2004 | 111 (110) |
| 2005 | 41 (39) |
| 2006 | 244 (348) |
| 2007 | 76 (83) |
| 2008 | 81 (50) |
| 2009 | 74 (70) |
| 2010 | 163 (190) |
| 2011 | 64 (101) |
| 2012 | 344 (518) |
| 2013 | 621 (421) |
| 2014 | 322 (441) |
| 2015 | 102 (200) |
| 2016 | 37 (109) |
| 2017 | 36 (62) |
| 2018 | 253 (212) |
| 2019 | 48 (53) |
| 2020 | 16 (4) |
| 2021 | 15 (8) |
| 2025 | 368 (353) |

the infection. Mosquito larvae development depends on temperature and precipitation and melting water from snow [33]. The temperature in June is also an important factor for the survival of the newly hatched grouse chicks [34, 35]. The model supported the role of hatch-year black grouse as an amplifying host for SINV infections in Finland [2, 9]. The hatch-year birds have been found to be an important amplifying host for West Nile virus [36]. Despite the overall decline in black grouse density in the late 1980s and early 1990s in Finland, it is still comparatively high in areas with high SINV incidence. Grouse chicks feed on insects for the first few weeks of their life and hatching occurs around mid-June in Northern Finland and about 1 week earlier in Southern Finland [34, 35]. Hatch-year chicks are also more accessible to mosquitoes due to as yet undeveloped plumage and may be more susceptible to infection due to a lack of immunity. These factors and the density of the mosquito population during late summer may together explain the exact timing of the annual SINV infection outbreaks.

The negative effect of agricultural land in the model possibly indicates the importance of other land areas where people spend time and may become exposed to mosquitoes. The significance of the negative effect of income perhaps indicates that people with lower income are more likely to pick berries and mushrooms rather than purchase them, thereby becoming exposed to mosquito bites. The effect of income level may also be a coincidental finding, as the level of income is generally lower in high SINV infection rate areas. There was a negative correlation between income level compared to the proportion of population working in agriculture (data not shown). Spending time in the forest or swamps is an established risk factor for acquiring SINV infection [2]. It may be that overall outdoor activities are more common in the countryside where the level of income is also lower.

The significance of the proportion of regulated lake water shore-length for the occurrence of the disease is a novel finding. This finding should be interpreted with caution as this study is of an ecological nature and ecological fallacy may account for this finding. However, there is also strong supportive evidence that this association is valid. Water regulation probably creates ecological changes in the lake banks favouring the development of suitable environments for the development of mosquitoes or its bloodmeal supplier, i.e. birds. Regulated water areas have been found to be a risk factor for West Nile virus, which has a similar biological cycle to SINV infection [37]. Notably, the heavily regulated Koitere Lake is located in the hotspot of SINV infections, in the Ilomantsi municipality (the disease was named Pogosta disease by the Ilomantsi Centre) with one of the highest prevalence and incidence rates for SINV infection. A power plant was built in 1955, creating changes in the level of water surfaces, which was further enforced by initiation of water regulation in 1980 [38]. The lakeshore was trimmed of trees prior to regulation, but the regulation was conducted at a lower level than initially planned, creating bushy vegetation on the lakeshore [38]. The first major SINV infection outbreak was diagnosed by serology in Finland in 1981, although a probable SINV infection outbreak was described clinically in 1974 [9, 39]. With the geographical spread of the cases and the importance of local climatic factors and regulated waters, it is tempting to speculate that the prevalence of SINV infections in Sweden and Russian Karelia may be due to similar phenomena.

We found the hurdle model suitable for characterizing the cyclic SINV infections. Distinguishing explanatory variables for the occurrence and incidence

of the infections was beneficial. Based on this and our previous study [40], we believe that the hurdle model provides more information on the nature of the explanatory variables for the pathways of infections. A model with an estimated mosquito variable based on weather variables would be useful [33]. We speculate that the variable for cases in the previous month (ln) is a useful proxy for the magnitude of the mosquito population. Several sinusoidal terms remained significant in the final model. We do not have biological explanations for all these terms, except the 12-month term, which reflects the seasonal variation of the infections and the 84-month term, which slightly reflects the cycles in the forest grouse populations detected prior to the mid-1980s [41]. Overall, our knowledge of SINV cycles is still rudimentary, while unknown and important factors may remain to be identified, including entomological variables. There may be cyclic fluctuations in the density of other potential amplifying hosts or similarly other factors influencing SINV epidemiology. We also fitted the model without the cyclic and trend terms to assess the adequate fit of the model (data not shown). All other variables remained significant, but temperature in May and hatch-year grouse density in the zero part and income in the negative binomial part became non-significant. However, leaving any of these variables out of the original model increases AIC by a factor of 25 units. Therefore we conclude that the model with the trend terms and external variables as listed above was adequately fitted.

We identified climatic, ecological and socio-economic determinants for both the occurrence and incidence of human SINV infections in Finland, applying a negative binomial distribution based hurdle model. For public health interventions, water regulation guidelines should be adapted to prevent an increase in infected mosquito populations. The study was hampered by the lack of mosquito surveillance data, although climatic factors are known estimators for mosquito quantities [30, 33, 42].

A shortcoming of our hurdle model is that data were unadjusted for age and sex distribution for the HCDs, and we ignored the possible change in effects by time and area. The spatial correlations of the residuals were tested with the limitations of the study material and found non-significant, this is in accordance with the intuitive observation that the spread of the disease is localized.

This study clarifies interactions between ecological and biological phenomena and the occurrence of SINV infections in Finland. Furthermore, the use of a hurdle model is justified for identifying risk factors for the occurrence and incidence of an infectious disease with an excess of zero counts in the data.

## SUPPLEMENTARY MATERIAL

For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S095026881200249X.

## DECLARATION OF INTEREST

None.

## REFERENCES


1. **Kurkela S, *et al.*** Clinical and laboratory manifestations of Sindbis virus infection: prospective study, Finland, 2002–2003. *Journal of Infectious Diseases* 2005; **191**: 1820–1829.
2. **Sane J, *et al.*** Epidemic sindbis virus infection in Finland: a population-based case-control study of risk factors. *Journal of Infectious Diseases* 2011; **204**: 459–466.
3. **Burt FJ, *et al.*** Chikungunya: a re-emerging virus. *Lancet* 2011.
4. **Lvov DK, *et al.*** Identity of Karelian fever and Ockelbo viruses determined by serum dilution-plaque reduction neutralization tests and oligonucleotide mapping. *American Journal of Tropical Medicine and Hygiene* 1988; **39**: 607–610.
5. **Lvov DK, *et al.*** Isolation of Karelian fever agent from *Aedes communis* mosquitoes. *Lancet* 1984; **2**: 399–400.
6. **Lundstrom JO, *et al.*** Geographical and temporal distribution of Ockelbo disease in Sweden. *Epidemiology and Infection* 1991; **106**: 567–574.
7. **Sane J, *et al.*** Complete coding sequence and molecular epidemiological analysis of Sindbis virus isolates from mosquitoes and humans, Finland. *Journal of General Virology* 2012; **93**: 1984–1990.
8. **Manni T, *et al.*** Diagnostics of Pogosta disease: antigenic properties and evaluation of Sindbis virus IgM and IgG enzyme immunoassays. *Vector Borne and Zoonotic Diseases* 2008; **8**: 303–311.
9. **Brummer-Korvenkontio M, *et al.*** Epidemiology of Sindbis virus infections in Finland 1981–96: possible factors explaining a peculiar disease pattern. *Epidemiology and Infection* 2002; **129**: 335–345.
10. **Kurkela S, *et al.*** Sindbis virus infection in resident birds, migratory birds, and humans, Finland. *Emerging Infectious Diseases* 2008; **14**: 41–47.
11. **Laine M, *et al.*** Prolonged arthritis associated with Sindbis-related (Pogosta) virus infection. *Rheumatology* 2002; **41**: 829–830.
12. **Kurkela S, *et al.*** Arthritis and arthralgia three years after Sindbis virus infection: clinical follow-up of a

cohort of 49 patients. *Scandinavian Journal of Infectious Diseases* 2008; **40**: 167–173.
13. **Brummer-Korvenkontio M, Saikku P.** Mosquito-borne viruses in Finland. *Medical Biology* 1975; **53**: 279–281.
14. **Lundstrom JO, Turell MJ, Niklasson B.** Viremia in three orders of birds (Anseriformes, Galliformes and Passeriformes) inoculated with Ockelbo virus. *Journal of Wildlife Diseases* 1993; **29**: 189–195.
15. **Francy DB, *et al.*** Ecologic studies of mosquitoes and birds as hosts of Ockelbo virus in Sweden and isolation of Inkoo and Batai viruses from mosquitoes. *American Journal of Tropical Medicine and Hygiene* 1989; **41**: 355–363.
16. **Utrio P.** Distribution of mosquitoes (Diptera, Culicidae) attracted by man in Finland in early July. *Notulae Entomologicae* 1978; **58**: 107–114.
17. **Schwartz J, *et al.*** Methodological issues in studies of air pollution and daily counts of deaths or hospital admissions. *Journal of Epidemiology and Community Health* 1996; **50** Suppl 1: S3–11.
18. **Hilbe JM.** *Negative Binomial Regression*, 1st edn. Cambridge: Cambridge University Press, 2007.
19. **Ridout M, Demétrio CGB, Hinde J.** Models for count data with many zeros. In: International Biometric Conference, Cape Town, 1998.
20. **Jalava K, *et al.*** Beef cattle density and proportion of unsalted water: the most significant explanatory variables for risk of Verotoxigenic *Escherichia coli* (VTEC) infection in Finland 1997–2006. In: European Scientific Conference on Applied Infectious Disease Epidemiology (ESCAIDE). Stockholm, Sweden, 2009: 77.
21. **Carrel M, *et al.*** Protection from annual flooding is correlated with increased cholera prevalence in Bangladesh: a zero-inflated regression analysis. *Environmental Health* 2010; **9**: 13.
22. **Gonzales-Barron U, *et al.*** Count data distributions and their zero-modified equivalents as a framework for modelling microbial data with a relatively high occurrence of zero counts. *International Journal of Food Microbiology* 2010; **136**: 268–277.
23. **Ylhaisi JS, *et al.*** Growing season precipitation in Finland under recent and projected climate. *Natural Hazards and Earth System Sciences* 2010; **10**: 1563–1574.
24. **Tietavainen H, Tuomenvirta H, Venalainen A.** Annual and seasonal mean temperatures in Finland during the last 160 years based on gridded temperature data. *International Journal of Climatology* 2010; **30**: 2247–2256.
25. **Lindén H, *et al.*** Wildlife triangle scheme in Finland: methods and aims for monitoring wildlife populations. *Finnish Game Research* 1996; **49**: 4–11.
26. **R Foundation.** The R project for statistical computing, 2011.
27. **Casella G, Berger RL.** *Statistical Inference*, 2nd edn. California: Duxbury Press, 2001.
28. **Sane J, *et al.*** Epidemiological analysis of mosquito-borne Pogosta disease in Finland, 2009. *Eurosurveillance* 2010; **15**.
29. **Sane J, *et al.*** Clinical Sindbis alphavirus infection is associated with HLA-DRB1*01 allele and production of antibodies. *Clinical Infectious Diseases* 2012; **55**: 358–363.
30. **Ryden P, *et al.*** Outbreaks of tularemia in a boreal forest region depends on mosquito prevalence. *Journal of Infectious Diseases* 2012; **205**: 297–304.
31. **Jupp PG, *et al.*** Sindbis and West Nile virus infections in the Witwatersrand-Pretoria region. *South African Medical Journal* 1986; **70**: 218–220.
32. **McIntosh BM, *et al.*** Epidemics of West Nile and Sindbis viruses in South Africa with Culex (Culex) univittatus Theobald as vector. *South African Journal of Science* 1976; **72**: 295–300.
33. **Trawinski PR, Mackay DS.** Identification of environmental covariates of West Nile virus vector mosquito population abundance. *Vector Borne Zoonotic Diseases* 2010; **10**: 515–526.
34. **Linden H.** Variations in clutch size and egg size of capercaillie and black grouse. *Suomen Riista* 1983; **30**: 44–50.
35. **Ludwig GX, *et al.*** Short- and long-term population dynamical consequences of asymmetric climate change in black grouse. *Proceedings of the Royal Society of London, Series B* 2006; **273**: 2009–2016.
36. **Hamer GL, *et al.*** Rapid amplification of West Nile virus: the role of hatch-year birds. *Vector Borne Zoonotic Diseases* 2008; **8**: 57–67.
37. **Liu H, Weng Q, Gaines D.** Spatio-temporal analysis of the relationship between WNV dissemination and environmental variables in Indianapolis, USA. *International Journal of Health Geography* 2008; **7**: 66.
38. **Tarvainen A, *et al.*** The influence and developmental challenges of regulating Lake Koitere [in Finnish]. In: *Suomen Ympäristö*. Helsinki: Suomen ympäristökeskus, 2006.
39. **Brummer-Korvenkontio M, Kuusisto P.** Onko Suomen länsiosa säästynyt 'Pogostalta'. *Suomen Laakarilehti* 1981; **32**: 2606–2607.
40. **Jalava K, *et al.*** Agricultural, socioeconomic and environmental variables as risks for human verotoxigenic Escherichia coli (VTEC) infection in Finland. *BMC Infectious Diseases* 2011; **11**: 275.
41. **Ranta E, Helle P, Lindén H.** Forty years of grouse monitoring in Finland [in Finnish]. *Suomen Riista* 2004; **50**: 128–136.
42. **Roiz D, *et al.*** Climatic factors driving invasion of the tiger mosquito (*Aedes albopictus*) into new areas of Trentino, northern Italy. *PLoS ONE* 2011; **6**: e14800.