Seroprevalence and incidence of Puumala orthohantavirus in its bank vole (Myodes glareolus) host population in northeastern France: Between-site and seasonal variability

Elodie Monchatre-Leroy, Franck Yves Sauvage, Franck Boué, Denis Augot, Philippe Marianneau, Viviane Hénaux, Laurent Crespin

Epidemics 40 (2022) 100600

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Seroprevalence and incidence of Puumala orthohantavirus in its bank vole (*Myodes glareolus*) host population in northeastern France: Between-site and seasonal variability

E. Monchatre-Leroy [a,*], F. Sauvage [b,c], F. Boué [a], D. Augot [a,d], P. Marianneau [e], V. Hénaux [f], L. Crespin [c,g,h]

[a] Anses, Laboratoire de la rage et de la faune sauvage, Nancy 54220, France
[b] SEENOVATE, 8 Rue de la Bourse, 69002 Lyon, France
[c] UMR–CNRS 5558 Biométrie et Biologie évolutive, Université C. Bernard Lyon-1, Villeurbanne, France
[d] Usc VECPAR-ANSES LSA, EA 7510, SFR Cap Santé, Université Champagne-Ardenne, Reims, France
[e] Université de Lyon, Anses, Laboratoire de Lyon, Unité de virologie, Lyon 69364, France
[f] Université de Lyon, Anses, Laboratoire de Lyon, Unité épidémiologie et appui à la surveillance (EAS), Lyon 69364, France
[g] Université Clermont Auvergne, INRAE, VetAgro Sup, UMR EPIA, 63122 Saint-Genès-Champanelle, France
[h] Université Lyon, INRAE, VetAgro Sup, UMR EPIA, 69280 Marcy-L'Etoile, France

ARTICLE INFO

Keywords:
Seroprevalence
Incidence
Bank voles
Puumala orthohantavirus
Suppress normalized difference vegetation index
Generalized estimating equations

ABSTRACT

Given the difficulty of measuring pathogen transmission in wildlife, epidemiological studies frequently rely on cross-sectional seroprevalence. However, seropositivity indicates only exposure to a pathogen at an unknown time. By allowing to obtain repeated test results from individuals sampled multiple times over an extended period, longitudinal data help reduce this uncertainty. We used capture-mark-recapture data on bank vole (*Myodes glareolus*) individuals collected at four sites over ten years in northeastern France to investigate the impact of environmental variables on seroprevalence and incidence of Puumala orthohantavirus (PUUV). PUUV causes a chronic infection without apparent symptoms, that may however impair survival of its rodent host in the wild. Viral transmission between rodents may occur through direct contact or via the environment. Principal component analysis was used to deal with multicollinearity among environmental variables. Incidence and seroprevalence were investigated with either generalized estimating equations or Poisson regression models depending on the number of observations for each season. In spring, only the factor site was found to be significant for seroprevalence, while a principal component including meteorological conditions of the previous winter and the normalized difference vegetation index (NDVI) of both the previous winter and spring had a significant effect on incidence. In autumn, only the factor site was significant for incidence, while two principal components, including either the meteorological conditions of the autumn and previous spring or NDVI of the autumn significantly affected seroprevalence. We discuss these results in light of the particular demography of small mammals. We encourage other researchers to investigate the relationships between demographic parameters of wild host populations and the environment, by using both incidence and seroprevalence.

1. Introduction

Emerging infectious zoonotic diseases are mainly caused by pathogens from wildlife (Jones et al., 2008; Karesh et al., 2012). To prevent human diseases, it is essential to monitor pathogens circulating in wildlife and to understand the risk factors for wildlife infection. However, epidemiological studies in wildlife are limited by the difficulty of observing and collecting data on animal interactions in wildlife and imperfect detection of cases (Craft and Caillaud, 2011). Facing these issues, epidemiological studies frequently rely on cross-sectional seroprevalence data. However, seropositivity is only the result of exposure of an animal to a pathogen at some time in the past (Cleaveland et al., 2007). This moment will be even more difficult to estimate if the persistence of antibodies is long. Using serological data from short-term

* Corresponding author.
E-mail address: elodie.monchatre-leroy@anses.fr (E. Monchatre-Leroy).

https://doi.org/10.1016/j.epidem.2022.100600
Received 22 June 2021; Received in revised form 2 May 2022; Accepted 14 June 2022
Available online 18 June 2022

and one-shot case studies, common for wildlife, is therefore not sufficient to estimate the time of infection and to evaluate risk factors for infection. In domestic animals, a recent study highlighted that imperfect detection of cases in logistic or zero-inflated Poisson models increased the probability of identifying false risk factors (Combelles et al., 2019). Imperfect case detection can lead to confusion between risk factors for disease occurrence and factors related to detection, especially if the spatial distribution of the disease is heterogeneous. (Vergne et al., 2014), and biases in estimates of the survival probability of individuals in different infection states and in the probabilities of infection and recovery (Benhaiem et al., 2018). To deal with this imperfect detection of cases, some authors (e.g. Hazel et al., 2000; MacCallum, 2000) have recommended repeated monitoring of susceptible animals in longitudinal serology studies of wildlife. Accordingly, a recent study has demonstrated that combining repeated test results from individuals sampled multiple times over an extended period improved detection of *Mycobacterium bovis* in badgers (Buzdugan et al., 2017). In this way, long-term monitoring of individuals, combined with systematic testing of captured individuals for the pathogen of concern, provides an important tool to estimate epidemiological parameters in wildlife (Craft and Caillaud, 2011).

Orthohantaviruses are zoonotic pathogens that have caused an increasing number of clinically apparent human infections in Europe over the past ten years (Heyman et al., 2011). The most prevalent orthohantavirus in Western Europe is Puumala virus (PUUV), whose reservoir is the bank vole (*Myodes glareolus*), a wide-spread rodent in this region. Viral transmission between rodents may occur through direct contact (biting, grooming, shared nesting, etc.) or via the environment (Kallio et al., 2006a). Rodents develop chronic infections, but clinical symptoms have never been reported (Bernshtein et al., 1999). However, previous studies have documented tissue pathologies in infected hosts (Lyubsky et al., 1996; Netski et al., 1999 but see Botten et al., 2000) and more recent papers have suggested that orthohantaviruses decrease the probability of survival of their hosts (Luis et al., 2012; Tersago et al., 2012). Humans are contaminated indirectly, through inhalation of aerosolized excreta of infected bank voles, and usually develop a mild form of hemorrhagic fever with renal syndrome (HFRS) called nephropathia epidemica (NE, Penalba et al., 2001). Approximately 100 human hospitalized cases are detected annually in France, all of them located in the northeastern part of the country (Reynes et al., 2019).

As the number of NE cases has been found to be related to seroprevalence of infection in the reservoir host (Drewes et al., 2017; Swart et al., 2017; Tersago et al., 2011; Voutilainen et al., 2016), there has been strong interest in modeling rodent seroprevalence to predict human risk. Several studies have investigated the role of environmental factors, including food availability, meteorological variables (temperature, precipitation, and humidity), and habitat conditions, on rodent population dynamics and virus survival in the environment (reviewed in Monchatre-Leroy et al., 2017). However, using PUUV seroprevalence as an indicator of rodent infection is not without flaws. First, young bank voles may be positive to serological testing because of maternal antibodies (that may last for about two and a half months; Kallio et al., 2006b). Second, life-long persistence of PUUV antibodies in infected bank voles (Voutilainen et al., 2015) contributes further to blur the exact moment when an individual was infected. As a consequence, the association between seroprevalence and infection calculated at a given time is confused by the uncertain time since rodent exposure and infection with PUUV. Third, there is actually great variability in the findings of all studies investigating seroprevalence (Drewes et al., 2017; Linard et al., 2007; Olsson et al., 2005; Piechotowski et al., 2008; Reil et al., 2015; Schwarz et al., 2009; Swart et al., 2017; Tersago et al., 2008; Thoma et al., 2014), certainly because of profound differences in methodological approaches and the short time span of most studies (see Monchatre-Leroy et al., 2017). To alleviate these issues, incidence, i.e. the number of new cases of infected bank voles, that allows us to define epizootic dynamics and to identify populations at risk, may be considered.

Using longitudinal capture-mark-recapture data on individual bank voles, we set out to investigate the impact of environmental risk factors on the epidemiology of PUUV infection in reservoir populations monitored seasonally over a ten-year period at four different sites in northeastern France. We used the normalized difference vegetation index (hereafter NDVI, see Pettorelli et al., 2005), temperature, and cumulative rainfall to assess the impact of the environment on both seroprevalence, as has frequently been done in previous studies, and incidence of PUUV infection.

2. Methods

2.1. Data collection

Rodents were trapped from 2000 to 2009 at four sites (referred to as sites A, B, C, and D hereafter) located in the Ardennes department, located on the border of France and Belgium. Sites A and B, 2 km apart in Elan forest, are located in the middle of the department (near Charleville-Mézières) and sites C and D (5 km apart in Croix-Scaille forest) are about 30 km to the north. Rodent trapping was conducted five times per year (typically sessions occurred in April, June, July, September, and October) (for further details, see Augot et al., 2008). The trapping grid was based on 49 live traps (7 × 7 Ugglan Live Trap). The distance between two traps was 15 m (Augot et al., 2008). For each session, traps were baited with pieces of carrots and sunflower seeds and set for three consecutive nights. Trapped rodents were individually marked by toe-clipping and released at their original site of capture after collecting a blood sample from the retro-orbital sinus and identifying the species. All rodents were weighted and sexed. All the procedures were carried out according to Council Directive 86/609/EEC and the French transposition of this directive, Décret 2001–486 of June 2001 that were in force during the experimentation. Other rodent species were caught in the field, e.g. *Apodemus spp.,* but only data from bank voles were included in the analysis given that it is the only reservoir species for PUUV.

2.2. PUUV seroprevalence and incidence data

Sera were tested by enzyme linked immunosorbent assay (ELISA) on PUUV and Hantan virus antigen (Augot et al., 2006). From these serological data, we wish to compute two epidemiological measures, incidence and seroprevalence. Both incidence and seroprevalence are usually defined as ratios with the population size of the population under concern at the denominator. Ideally, population size should have been estimated directly from the capture-mark-recapture data collected in the field and a particularly popular method to estimate population size from field data for small mammals is the closed population models, that is based on a minimal set of assumptions (Otis et al., 1978). In their seminal paper, Otis et al. (1978: 78) warned that a number of distinct animals per trapping session between 10 and 20 was too low to apply these models. Unfortunately, numbers of distinct animals were below this threshold of 20 for most trapping sessions (see Table S4). Thus we did not try to estimate population size and use instead the numbers of bank voles captured in the computations of incidence and seroprevalence, as detailed below.

Concerning seroprevalence, all individuals weighting 14 g or less were considered young individuals still protected by maternal antibodies (Kallio et al., 2006b) and were therefore excluded from the number of seropositive individuals. Seroprevalence was calculated as the proportion of seropositive rodents among all individuals captured during a given trapping session:

$$Pr_t = \frac{Number\ \ of\ \ seropositive\ \ individuals > 14g\ \ \ at\ \ \ t}{Number\ \ of\ \ individuals\ \ \ captured\ \ \ at\ \ \ t}.$$

An incident rodent is a new case of PUUV infection between two

captures. Seronegative rodents were considered as susceptible. Seropositive rodents were considered as incident or susceptible or excluded from the rodents' counts, depending on their weights and their serological histories (see Table 1). Incidence rates (IRs) between two successive trapping sessions were calculated as the ratio of incident individuals among susceptible individuals:

$$IR_{t \ to \ t+1} = \frac{Number \ of \ incident \ indiviuals \ between \ t \ and \ t+1}{(Number \ of \ susceptible \ individuals \ at \ t + Number \ of \ susceptible \ individuals \ at \ t+1)/2}.$$

The denominator represents then the average number of individuals at risk (Thrusfield, 2005: 55).

Given that the relationships between incidence rate or seroprevalence and environmental covariates depend on the season (Monchatre-Leroy et al., 2017), we split the data into two seasons. However, because a measure of incidence was calculated over a time interval separating two trapping sessions, while a measure of seroprevalence was calculated for each trapping session, seasons were slightly different between incidence and seroprevalence. Concerning seroprevalence, "spring" included the three sessions from April to July (ending August 10, 2000) and "autumn" the two sessions from September to October (starting August 28, 2001 and August 27, 2003). Concerning incidence, "spring" included the first three measures and "autumn" only the last one (between September and October, see Fig. 1).

2.3. Environmental data

We gathered data on environmental variables reported in the literature as risk factors for rodent infection (reviewed in Monchatre-Leroy et al., 2017). NDVI measures the photosynthetic activity of vegetation and is therefore considered an indicator of the quantity of green vegetation. Green vegetation is an extremely important component of the environment for bank voles because it provides both food and cover from predators. Data were acquired from the Global Agriculture Monitoring (GLAM) project (https://ipad.fas.usda.gov/glam.htm), with a spatial resolution of 250 m and a 16-day compositing period. The French national meteorological Institute (Météo France) provided daily temperature and precipitation data for sites A and B, respectively at 16.5 and 15.5 km from the meteorological station of Charleville-Mézières (indicative number 08105005). The French national forest Institute (ONF) provided meteorological data for sites C and D, respectively at 2.2 and 3.3 km from the private station of Croix-Scaille forest (EPC08) in Ardennes. Minimum and maximum daily temperatures and daily cumulative rainfall were available from each meteorological station.

Table 1
Decision rules to identify incident individuals, depending on the weight of rodents and their serological histories. '+' denotes seropositive individuals and '-' denotes seronegative individuals. A seropositive status for individuals with a weight less than or equal to 14 g was not considered as a result of an infection (incident) but due to maternal antibodies.

Weight> 14 g		
Capture at t	Recapture at t + 1	Counted as incident between t and t + 1
+	+	No (excluded from the data)
+	–	No (this transition does not exist in the data)
–	+	Yes
–	–	No
Weight≤ 14 g		
Capture at t	Recapture at t + 1	Counted as incident between t and t + 1
+	+	No (maternal antibodies)
+	–	No (this transition does not exist in the data)
–	+	Yes
–	–	No
Weight≤ 14 g	Weight> 14 g	
Capture at t	Recapture at t + 1	Counted as incident between t and t + 1
+	+	Yes (maternal antibodies)
+	–	No (maternal antibodies)
–	+	Yes
–	–	No

Because correlation was strong between minimum and maximum daily temperatures at each meteorological station (Pearson correlation: 0.94, p-value $< 2.10^{-16}$), only maximum daily temperatures were used. We calculated mean values for each variable for the previous winter (from November to January), spring (from April to July) and autumn (from September to October). We also calculated mean annual maximum temperature values two years before ("$Temp_{Y\text{-}2}$"), as this value may be considered an indicator of seed production during the year preceding the captures, and thus a surrogate variable for food availability of bank voles (see Clement et al., 2009). Variables for the spring analyses included NDVI ("$NDVI_w$", "$NDVI_s$"), temperatures ("$temp_w$", "$temp_s$"), and rainfall ("$rain_w$", "$rain_s$") of the previous winter and spring, respectively, and temperature two years before ("$Temp_{Y\text{-}2}$"). Variables for the autumn analyses included NDVI ("$NDVI_s$", "$NDVI_a$"), temperatures ("$temp_s$", "$temp_a$"), and rainfall ("$rain_s$", "$rain_a$") of the spring and autumn, respectively (Table 2).

2.4. Statistical analyses

2.4.1. Environmental variables

Evaluation of correlations among variables underlined strong multicollinearity; correlations and partial correlations were assessed with the ppcor Package in R (R Core Team, 2018). Among variables for the spring analysis, there was a correlation structure between $rain_w$, $temp_w$ and $temp_s$ (correlations and partial correlations significant with $p < 0.05$). Among variables for the autumn analysis, there was a correlation between $temp_a$, $temp_s$, $NDVI_a$ and $rain_s$ (correlations and partial correlations significant with $p < 0.05$). To alleviate this difficulty, new variables were defined using a principal component analysis (PCA) in R (FactoMineR Package). Based on the decrease of eigenvalues, we decided to keep the first two principal components for spring and the first three for autumn. Taken together, these principal components explained 65.5% of the variance in spring (Fig. S1, Table S1) and 85.2% of the variance in autumn (Fig. S1, Table S2).

To interpret each principal component, we considered the variables with the highest contribution (i.e. highest squared cosine (cos^2) values (Abdi and Williams, 2010)). For the PCA on spring variables, meteorological data (both temperatures and rainfall) in spring and the proxy for food availability contributed most to Principal Component 1, and meteorological data during the previous winter and the NDVI in spring and winter contributed most to Principal Component 2 (Table 3). Similarly, for the PCA on autumn variables, meteorological data in both spring and autumn contributed most to Principal Component 1, NDVI in autumn to Principal Component 2, and NDVI in spring to Principal Component 3 (Table 4).

2.4.2. Generalized estimating equation and Poisson regression models

We stress that site D was excluded from all statistical analyses because of a limited number of captures (see Results below). The data were counts of seropositive or incident rodents trapped at one or several trapping sessions at each site, within each season. These counts were likely to be correlated within a season for a given site (see below), i.e. to

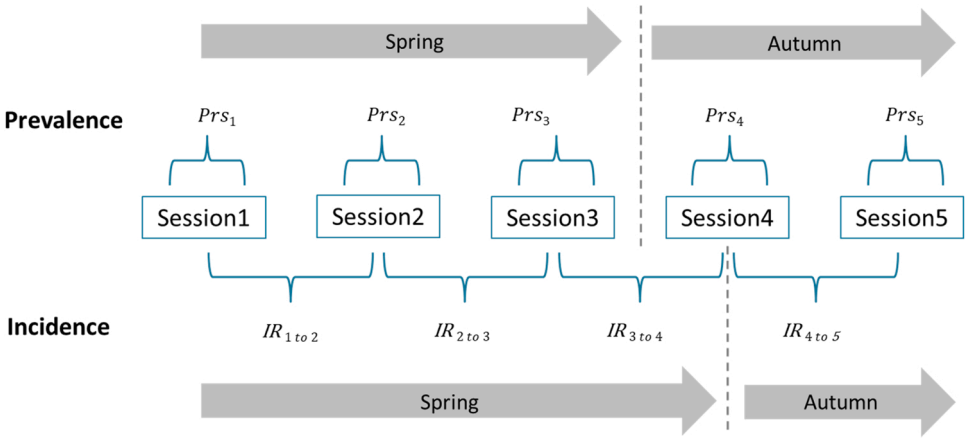

Fig. 1. Definition of season in seroprevalence and incidence. Prs_n: Seroprevalence in session n, $IR_{n\ to\ n+1}$: incidence rate from session n to n + 1.

Table 2
Description of variables used in the modeling of seasonal incidence or seroprevalence during 2000–2009.

Variables	Method of calculation
$temp_s$	Average of daily maximum temperatures from April to July
$rain_s$	Cumulative rainfall from April to July
$NDVI_s$	Average of NVDI values acquired every 16 days from April to July
$temp_a$	Average of daily maximum temperatures from August to October
$rain_a$	Cumulative rainfall from August to October
$NDVI_a$	Average of NVDI values acquired every 16 days from August to October
$temp_w$	Average of daily maximum temperatures during the previous November–January period
$rain_w$	Cumulative rainfall during the previous November–January period
$NDVI_w$	Average of NVDI values during the previous November–January period
$temp_{Y-2}$	Average daily maximum temperatures from June to August 2 years before

be longitudinal data with panels consisting of the combination of year by site (Hardin and Hilbe, 2003). Generalized estimating equations (GEE) are statistical models particularly suitable to deal with this kind of data as they allow the different measurements within a panel to be correlated (Hardin and Hilbe, 2003; Liang and Zeger, 1986). Besides GEE models make the assumption of no correlation between panels, which appears to be reasonable at first approximation, given the short dispersal and life expectancy of bank voles, relative to the experimental design we used (Bernshtein et al., 1999; Karlsson and Potapov, 1998; Yoccoz et al., 1998).

A limitation of GEE models is that their performances decrease greatly if the number of panels is too small (see below), and a standard rule of thumb is to use at least 30 panels (see p. 132 of Hardin and Hilbe, 2003; even 40 for randomized clustered trials according to Li and Redden, 2015). The number of panels in our data sets was either 28 (incidence at spring) or 29 (seroprevalence at spring and autumn). As a result, we decided to compute all standard errors from the best model using the Kauermann and Carroll (2001) correction for small samples. However, this small sample correction is not currently implemented in PROC GENMOD, the SAS procedure designed to run "classic" GEE models, i.e. using asymptotic estimators, estimated by the method of moments as proposed by Liang and Zeger (1986). Thus, we had to use another procedure, PROC GLIMMIX that makes use of a different estimation method, i.e. residual pseudo-likelihood, to be able to work out the Kauermann and Carroll (2001) correction. As already reported by other authors facing the very same problem (McNeish and Harring, 2017; McNeish and Stapleton, 2016), estimates from PROC GENMOD and PROC GLIMMIX before the correction were, however, nearly identical.

Except for the analysis of incidence in autumn, full GEE models included the logarithm of the number of susceptible rodents or the number of rodents captured (calculated as described in Fig. 1) as an offset, factors site (*site*), time interval (*ti*) for incidence or trapping session (*session*) for seroprevalence, the two or three first principal components (*PC1*, *PC2* and *PC3*) as continuous covariates, and all two-way interactions between site and principal components (*PC1 *site*, *PC2 *site* and *PC3 *site*). Additionally, to deal with the fact that the numbers of incident (or seropositive) rodents may not be independent between time intervals (or sessions) within a given year, we used an exchangeable process or an autoregressive process of first order as a working correlation matrix. To make a choice between an exchangeable and an autoregressive process of first order structure as the working correlation

Table 3
Contribution of variables to principal components (spring analysis).

Variables	PC1			PC2		
	cos^2	correlation	*p*-value	cos^2	correlation	*p*-value
NDVIs	0.05	/	ns	0.50	0.71	< 0.001
$NDVI_w$	0.09	/	ns	0.48	0.70	< 0.001
$temp_s$	0.69	0.83	< 0.001	0.01	/	ns
$temp_w$	0.27	0.52	0.001	0.60	0.77	< 0.001
$rain_s$	0.56	-0.74	< 0.001	0.05	/	ns
$rain_w$	0.18	-0.42	0.020	0.63	0.80	< 0.001
$temp_{Y-2}$	0.47	0.68	< 0.001	0.01	/	ns
Interpretation	Meteorological conditions in spring and the proxy of food availability			Meteorological conditions and NDVI of previous winter and NDVI in spring		

cos^2: quality of variables projection on each principal component, ns: not significant
$NDVI_s$: average of NVDI values from April to July / $NDVI_w$: average of NVDI values during the previous November–January period / $temp_s$: average of daily maximum temperatures from April to July / $temp_w$: average of daily maximum temperatures during the previous November–January period / $rain_s$: cumulative rainfall from April to July / $rain_w$: cumulative rainfall during the previous November–January period / $temp_{Y-2}$: average daily maximum temperatures from June to August two years before.

Table 4
Contribution of variables to principal components (autumn analysis).

	PC1			PC2			PC3		
	cos^2	correlation	*p*-value	cos^2	correlation	*p*-value	cos^2	correlation	*p*-value
$NDVI_s$	0.10	/	ns	0.14	0.37	0.050	0.75	0.87	< 0.001
$NDVI_a$	0.30	0.54	0.002	0.55	0.74	< 0.001	0.04	/	ns
$temp_s$	0.56	0.75	< 0.001	0.25	-0.49	0.010	0.07	/	ns
$temp_a$	0.83	0.91	< 0.001	0.00	/	ns	0.02	/	ns
$rain_s$	0.70	-0.83	< 0.001	0.08	/	ns	0.06	/	ns
$rain_a$	0.43	-0.65	< 0.001	0.26	0.51	0.050	0.00	/	ns
Interpretation	Meteorological conditions of spring and autumn of the same year			NDVI in autumn			NDVI in spring		

cos^2: quality of variables projection on each principal component, ns: not significant
$NDVI_s$: average of NVDI values from April to July / $NDVI_a$: average of NVDI values from August to October / $temp_s$: average of daily maximum temperatures from April to July / $temp_a$: average of daily maximum temperatures from August to October / $rain_s$: cumulative rainfall from April to July / $rain_a$: cumulative rainfall from August to October.

matrix, we worked out the correlation between Pearson residuals in the same season, from the full model with an independence structure correlation. In case of an autoregressive process of first order, we would expect a decrease in the correlation from lag 1 (consecutive observations) to lag 2 (observations separated by two time intervals).

For the analysis on incidence in autumn, we used a Poisson regression model because there was only one observation by year and site. The data were sparse, however, and it was not possible to estimate site-specific slopes for the environmental covariates *PC1* and *PC2* for sites A and B. Therefore, we included a common slope for sites A and B (both sites were 2 km away in the same Elan forest) and a specific slope for site C plus site-specific slopes for *PC3* in the full model.

We carried out a backward model selection. At each step, the term with the highest *p*-value was removed on the next step until all effects remaining in the model were significant. Regarding incidence in autumn, likelihood ratio tests were used in model selection. For the three other data sets, generalized robust score tests known to be conservative (Guo et al., 2005) were used. Model selection was carried out using the "mmmgee" package (Ristl et al., 2019) in R. All estimates are displayed with a 95% confidence interval constructed from t-quantiles, as recommended by several authors for GEE estimates (Li and Redden, 2015; Pan and Wall, 2002; Teerenstra et al., 2010).

3. Results

3.1. Seroprevalent and incident rodents

Overall, we trapped 2056 individual bank voles. Among them, 672 were recaptured once at least (952 recaptures in total). The percentage of recaptured rodents by trapping session and site was on average 0.33 (SD=0.29). The number of captures was very different between years, seasons, and sites (Fig. 2 and Table S3). However, site D displayed markedly fewer captures than other sites. For instance, the maximum annual number of captures at site D was 64 (in 2005) versus more than 100 at site C for two years (maximum: 310), at site A for six years (maximum: 278), and at site B for three years (maximum: 242). Thus, because of this limited number of captures, site D was excluded from all statistical analyses.

Very few animals were trapped in years 2004 and 2006, while years 2001, 2003 and 2005 were years with many captures at all sites. Captures were more regular at site A (seven years with captures in all sessions) than at sites B and C (five years with captures in all seasons). Most captures were made at site A (n = 1174) in comparison with site B (n = 827), and site C (n = 748).

There were more seroprevalent rodents than incident ones (Fig. 3 and Table S4). There was no or only one incident rodent at site C for six years (2000, 2001, 2003, 2006, 2007 and 2009) and at sites A and B for seven years (2000, 2002, 2004, 2006, 2007, 2008 and 2009). Some seroprevalent rodents were trapped every year at each site, except in 2006 for all sites, in 2000 for site A and in 2009 for site B.

3.2. Model selection and estimates from best model

3.2.1. Spring models

3.2.1.1. Seroprevalence. There were 74 observations pooled into 28 panels (combinations of year by site). The correlation between Pearson residuals from the full model did not decrease from lag 1 to lag 2 but instead seemed relatively constant (lag 1: 0.464, lag 2: 0.488). This suggested an exchangeable working correlation matrix. The correlation estimate from the best model was high (0.70). The best model included a single variable *site* (Table S5). The seroprevalence estimate at site C (0.31 [0.25; 0.40]) was the highest and was significantly different from that at site A (0.10 [0.04; 0.23], $p = 0.01$), but not from the seroprevalence estimate at site B (0.23 [0.14; 0.39], $p = 0.29$). The seroprevalence estimate at site A was marginally significantly different from the one at site B ($p = 0.08$).

3.2.1.2. Incidence. There were 82 observations pooled into 28 panels. The correlation between Pearson residuals from the full model decreased markedly from lag 1 to lag 2 (lag 1: 0.223, lag 2: 0.008). This pattern suggested an autoregressive process of first order as the working correlation matrix. The corresponding correlation estimate from the best model was rather small (0.31). The best model included principal component *PC2* and factor *session*. Principal component *PC2* was actually marginally significant ($p = 0.06$), but because it is known that the generalized robust score test on which the model selection was based is conservative (Guo et al., 2005), we decided to keep it in the best model. Spring incidence increased (0.27 [0.01; 0.53]) with increasing principal component *PC2,* i.e. with increasing spring and winter NDVI, rainfall, and temperature of previous winter. Incidence varied between sessions during the spring season: it was at its lowest level in April–June (0.01 [0.01; 0.03]), increased significantly to a maximum in June–July (0.06 [0.04; 0.08] $p < 0.001$) and then decreased slightly in July–September (0.04 [0.02; 0.07] $p < 0.05$, see Fig. 4). The incidence estimate in June–July was not significantly different from incidence in July–September ($p = 0.26$).

3.3. Autumn models

3.3.1. Seroprevalence

There were 53 observations pooled into 29 panels. The correlation estimate from the best model of seroprevalence was small (0.36). We kept the marginally significant interaction *site*PC1* in the best model ($p = 0.08$) because the generalized robust score test is known to be conservative (Guo et al., 2005). The best model thus included principal components *PC1* and *PC2* as well as factor *site* and interaction *site*PC1* (Table S6). Autumn seroprevalence increased with increasing principal component *PC1* (i.e. with increasing spring and autumn temperatures and decreasing spring and autumn rainfalls) at sites A and B (with slopes of 1.15 [0.49; 1.80] and 0.78 [0.20; 1.36], respectively) while there was

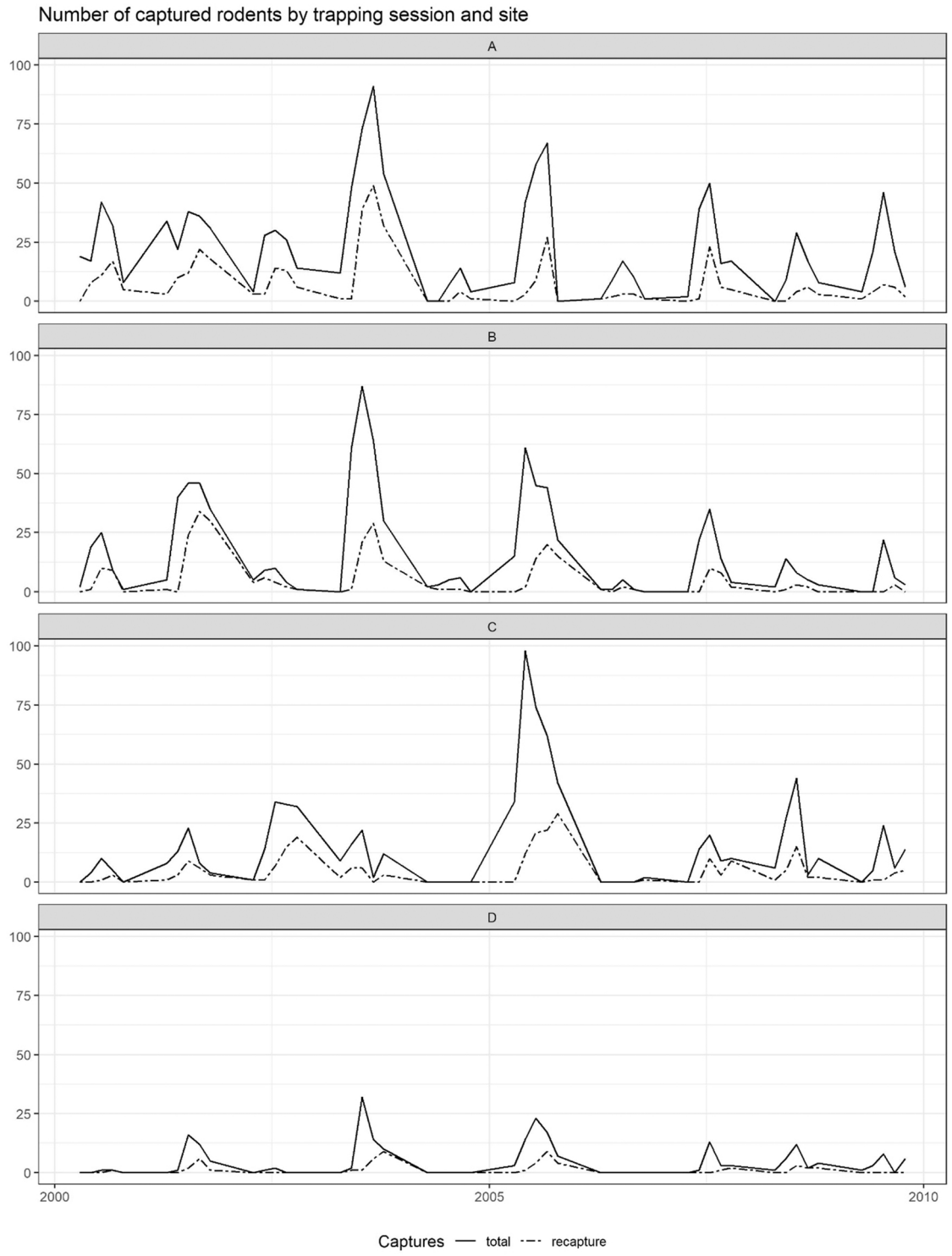

Fig. 2. Number of captures of bank voles per year, trapping session, and site.

no significant relationship at site C (slope: 0.13 [−0.44; 0.70]). Similarly, seroprevalence increased with increasing principal component *PC2* (i.e. with increasing NDVI in autumn) at all sites (slope: 0.32 [0.05; 0.59]). The seroprevalence estimate was the highest at site C (0.28 [0.08; 1.04]) and the lowest at site A (0.01 [0.00; 0.07], $p < 0.001$). The estimate at site B (0.09 [0.03; 0.27]) was marginally different from the estimate at site A ($p = 0.06$) and did not differ significantly from the estimate at site C ($p = 0.17$).

3.3.2. *Incidence*

Concerning incidence, there were 29 observations. The best model included only factor *site* (Table S6). The incidence estimate at site A (0.01 [0.00; 0.04]) was the lowest and was significantly different from that at site B (0.05 [0.02; 0.13], $p = 0.04$) and site C (0.05 [0.02; 0.13], $p = 0.04$). The incidence estimate at site B was not significantly different from the incidence estimate at site C ($p = 0.97$).

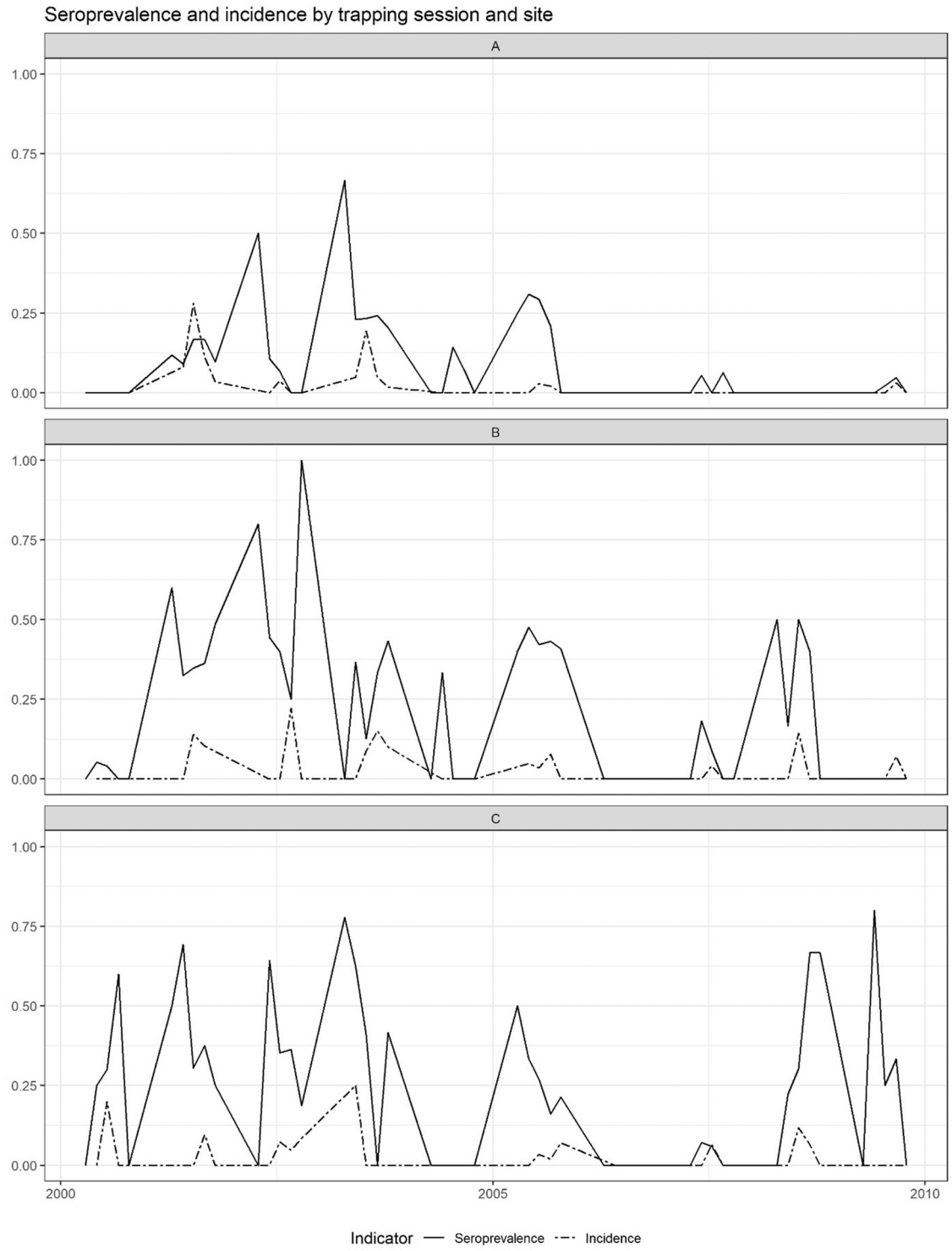

Fig. 3. Incidence and seroprevalence per trapping session and site.

4. Discussion

Our study assessed the impact of environmental variables on PUUV incidence and seroprevalence in bank vole populations, monitored over ten years at four different sites in northeastern France. Depending on the season, we found statistically significant relationships between principal components, including environmental variables like NDVI, rainfall or temperature, and incidence or seroprevalence epidemiological measures of PUUV infection.

4.1. Environmental variables

Incidence and seroprevalence are two different epidemiological measures of infection. The former has been much less frequently used to investigate links between environmental variables and PUUV epidemiology (Monchatre-Leroy et al., 2017). This is surprising given that antibodies remain life-long in infected bank voles (Bernshtein et al., 1999; Kallio et al., 2006b; Voutilainen et al., 2015), and thus seroprevalence may rather be considered a measure of cumulative infection over an

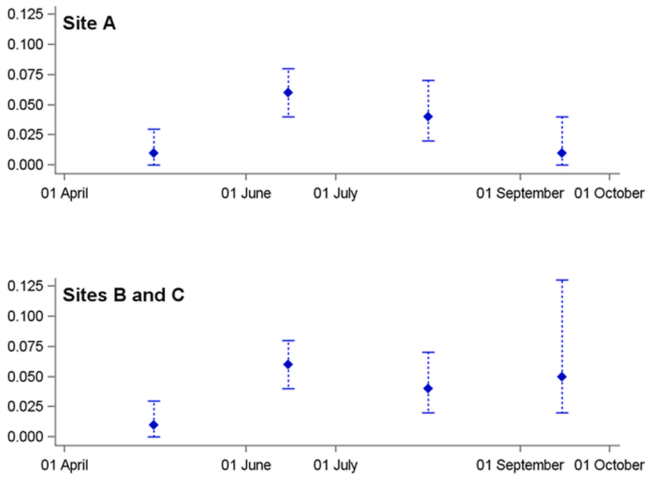

Fig. 4. Incidence estimates and their 95% confidence interval from best model.

extended period of time, while incidence is more representative of exposure and immediate infection by PUUV. Accordingly, the pattern of correlation among the three counts of rodent in spring was different for the seroprevalence and the incidence. The correlation was high for seroprevalence (0.70 from the GEE model) and relatively constant over the two lags while it was much lower for incidence (0.31 from the GEE model) and decreased much from first to second lag. Therefore, to investigate links between environment and PUUV epidemiology, incidence appears more suitable than seroprevalence because it will track down more closely the variability in environmental variables than will the seroprevalence.

Accordingly, we did not find any influence of environmental conditions on seroprevalence in spring, while we found that incidence increased with increasing rainfall and temperatures, i.e. relatively warm and wet meteorological conditions in the previous winter. As proposed by Flowerdew et al. (2017) and Vanwambeke et al. (2019), these meteorological conditions would promote better winter survival of rodents and an earlier and more intense breeding season in spring. The increase in the number of rodents would then facilitate PUUV transmission in the population (Clement et al., 2009; Escutenaire et al., 2000; Reil et al., 2017). Our results also highlighted that incidence increased with increasing winter and spring NDVI. NDVI is an indicator of the quantity of green vegetation that may represent, for bank voles, a food resource, protective cover against predators, or a source of materials to insulate themselves from the cold (Flower et al., 2019; Pettorelli et al., 2005). It is difficult to deduce which biological mechanisms are behind the impact of NDVI on incidence in our study. As far as we know, no other study has investigated the relationship between NDVI and incidence in any hantavirus/rodent pathosystem. However, other studies have shown a positive association between NDVI or other related indexes and incidence of human cases (NE), although at a different temporal scale, typically annual rather than seasonal (Barrios et al., 2013; Linard et al., 2007; Viel et al., 2011). Given the importance of seasonality in the demography of rodents, more work is needed to elucidate the link between NDVI, incident bank voles, and human cases.

In autumn, we did not find any influence of environmental conditions on incidence, but this data set has markedly fewer observations than the three others (see results), and the analysis had likely much lower statistical power as a result. However, our results showed that seroprevalence in autumn increased with increasing autumn NDVI at all sites and that, at two sites, seroprevalence increased with increasing temperatures and decreasing rainfall in spring and autumn, i.e. relatively warm and dry meteorological conditions. Others studies about the influence of rain, temperature, or NDVI on rodent seroprevalence found mixed results (Linard et al., 2007; Tersago et al., 2008) and the mechanisms of action of environmental conditions on rodent seroprevalence are still poorly understood (Monchatre-Leroy et al., 2017). Long-term monitoring of rodent populations allowing a more accurate description of the variability in seroprevalence over seasons would greatly improve our understanding of the mechanisms acting in natural host populations.

4.2. Seasonnal patterns of seroprevalence and incidence

As generally reported in other studies (Escutenaire et al., 2000; Kallio et al., 2010; Khalil et al., 2017; Voutilainen et al., 2016), seroprevalence was higher in spring than in autumn although the difference between the two estimates was small at site C. This general pattern is thought to reflect the accumulation of infections through the course of a year in the cohorts of bank voles (Voutilainen et al., 2016). Interpreting the variation of incidence seems less straightforward. We found that incidence was low at the beginning of spring, then increased to a maximum between June and July, slightly decreased in late summer and then, depending on the site, decreased further (site A) or tended to level off in autumn (sites B and C, see Fig. 4). In the same vein, Bernshtein et al. (1999) had studied a population of bank voles for five years in Russia and reported a peak in incidence in June, and then a steady decrease until August for two high-density years. Douglass et al. (2007) monitored a population of deer mouse, *Peromyscus maniculatus*, the host of Sin Nombre Virus for ten years in Montana, USA and reported a steady increase in incidence from spring to early summer, and then a plateau until October.

In fact, the incidence pattern we have described may be explained mainly by the peculiar demography of small mammals. First, many authors have reported, for several hantavirus/rodent pathosystems, that individuals become infected at the onset of sexual maturity (Glass et al., 1988) or that breeding individuals are more infected than non-breeding individuals, which suggests that engaging in reproductive activities entails a higher risk of becoming infected by the virus for an individual (Bernshtein et al., 1999; Douglass et al., 2007; Escutenaire et al., 2002; Olsson et al., 2002; Tersago et al., 2012). Second, it is known that, in contrast to individuals born early in the breeding season, the individuals born later will not engage in reproductive activities, but will overwinter and breed the year after (Gliwicz, 1989; Prevot-Julliard et al., 1999; Wiger, 1979). In this study, the first trapping session, April, corresponds roughly to the start of the breeding season, as shown by the small numbers of juveniles trapped at this session over the ten years at the three sites (2001: 1 juvenile 2007: 7 juveniles on two sites), so the bulk of the population is made up of overwintered individuals. If, as suggested by Voutilainen et al. (2015), seroconversion is generally high during winter, most overwintered individuals are infected during winter. Given that the peak of shedding occurs within one month after inoculation in laboratory experiments (Hardestam et al., 2008 but see Voutilainen et al., 2015), these overwintered individuals will no longer shed much virus in spring. This, combined with the typically small population sizes observed in spring, may explain why virus transmission is low at the beginning of the reproductive season. Alternatively, if, as suggested by Bershtein et al. (1999) and Olsson et al. (2002), most adults are infected when breeding activities resume, the low estimate of incidence observed between April and June in this study may rather be due to the low population size and the delay between the infection and the appearance of antibodies in the blood. This delay is not well known but was estimated to be 18 days in a laboratory experiment by Yanagihara et al. (1985). This is a rather low estimate as Hardestam et al. (2008) further showed a large variance between individuals. The following strong increase in incidence observed between April to June and June to July is likely the consequence of reproduction by the first cohorts of bank voles, born in April and May. About 80% of all juveniles captured in this study were trapped about equally in June and July (132 and 146 juveniles pooled over the three sites over the whole period, respectively). This massive influx of young will result in a decrease of incidence by two mechanisms: (1) this massive influx of young may result in

a temporary decrease of incidence, observed between time interval June to July and time interval July to September in this study, by inflating the population of susceptible individuals. This is the juvenile dilution effect described previously by Mills et al. (1999); see also Adler et al. (2008) on seroprevalence. (2) since the circulation of the virus was high in the previous time interval (June to July), many mothers of these young were actually infected by the virus and then would have passed them maternal antibodies. Voutilainen et al. (2016) reported that young with maternal antibodies may represent a high proportion of seropositive animals in July (up to 25–35% in peak years) in a Finnish population of bank voles. Now, given that these maternal antibodies will protect the young against the infection for about 80 days (Bernshtein et al., 1999; Kallio et al., 2006b), there will be a pool of young immune to the infection for a relatively long period of time which may further hinder the spread of the virus in the population (Kallio et al., 2015, 2010; Reijniers et al., 2020) and contribute to the stronger decrease in incidence observed between July to September and September to October at site A. Nonetheless, this decrease may also partly stem from the fact that most individuals born in the late reproductive season do not engage in reproductive activities until spring of next year. The pattern observed at sites B and C may be different, with a plateau rather than a decrease in incidence in autumn. Although the wide confidence intervals of these two incidence estimates call for caution, the difference would be best explained in terms of emigration (see below).

4.3. Site effect

Finally, best models displayed a significant site effect for seroprevalence and incidence estimates in autumn and for seroprevalence estimates in spring. More specifically, site A displayed lower incidence and seroprevalence than both sites B and C. This is particularly puzzling as site A and site B are located 2 km apart in the same broad-leaved forest, Elan, while site C is located in another forest with spruce (Croix-Scaille), 30 km away. Therefore, the differences in seroprevalence and incidence did not match the distances between sites, and did not match the different tree composition of the two forests either. In the absence of a specific study about the physical and ecological characteristics of the sites, it is difficult to pinpoint the factor responsible for these differences. However, on demographic grounds, we can speculate that site A was a habitat of good quality (i.e. supporting a large and stable population of bank voles; Gliwicz, 1989 and van Apeldoorn et al., 1992), while both sites B and C were habitats of poorer quality. Out of ten years of data, captures were more regular at site A (out of 50 trapping sessions, site A had more captures than the two others site for 26 trapping sessions, site B for 10 and site C for 13) and overall, site A also had more captures than any other site (respectively +42% and +57% compared to sites B and C, see Results). Keeping this in mind, most small mammal ecologists would agree that demographic exchanges among sites may be described according to either a source-sink (Pulliam, 1988) or a metapopulation (Levins, 1970) demographic model. In an extensive study encompassing 51 woodlots in The Netherlands, van Apeldoorn et al. (1992) presented compelling arguments in favor of the source-sink model for bank voles. The additional observation reported by Khalil et al. (2017) that none of the 58 trapping 1-ha plots they monitored in northern Sweden "*harbored infected bank voles throughout the 10-year study period*" confirmed that movements between sites and stochasticity play an important role in the PUUV-bank vole pathosystem. Thus, according to the source-sink theoretical model, we would expect an influx of individuals from sites with a reproductive surplus (sites of high quality, i.e. site A) toward sites of poorer quality (sites B and C). As a matter of fact, the number of juveniles trapped was larger and more regular at site A than at sites B and C. Now, given that the virus is known to survive outside its host in the environment (Kallio et al., 2006a) and that the distribution of PUUV in the landscape must show considerable spatial variation (as does the distribution of PUUV-infected bank voles, see Khalil et al., 2017), the individuals moving between sites likely have a higher probability of exposure than those staying at their site of birth. Consequently, sites B and C may receive more (infected) immigrants than site A. Two lines of empirical evidence strengthen this scenario. For voles of the genus *Microtus*, more amenable to experimentation than bank voles (Gliwicz and Ims, 2000), Gundersen et al. (2001) by manipulating density in enclosures, have shown that individuals immigrated to habitat patches with lower density than the patch they came from and Gundersen et al. (2002) have further reported that the probability of settlement of immigrants decreased with the density of the immigration patch. For bank voles, evidence is less clear-cut; we are aware of a single experiment that is unfortunately inconclusive (Glorvigen et al., 2012) but Gliwicz (1989), in a comparison of three habitats based on small trapping grids and therefore small numbers of capture has described the successful immigration of young (settlement and breeding) from optimal to suboptimal habitats, likely because of better prospects of maturation in low density patches (maturation of young females is known to be suppressed at high densities, see Prevot-Julliard et al., 1999). Thus, in this perspective, site A displayed lower incidence and seroprevalence than site B and site C because its larger and more stable population made it more difficult for potential (infected) immigrants to settle in (the so-called "social resistance hypothesis", Armansin et al., 2020) and to bring the virus back into the population's site.

4.4. Study limitations

In our study, the numbers of incident and seroprevalent rodents were estimated several times per season at each site, and these counts were thus correlated. GEE models are designed to deal with correlated data and to yield a robust estimate of the variance/covariance matrix, using the modified sandwich variance estimator that has the advantage of being consistent even if the working correlation matrix has been misspecified (Liang and Zeger, 1986; Hardin and Hilbe, 2003). Unfortunately, this estimator has been shown to underestimate the variance in analyses based on small data sets and thus to lead to liberal Wald-type test results (Feng et al., 1996; Lu et al., 2007; Paik, 1988). As a consequence, corrections of the modified sandwich variance estimator or other estimators (e.g. jackknife or clustered bootstrap) have been proposed to improve the performances of GEE with small samples (e.g. Cheng et al., 2013; Gosho, 2014; Kauermann and Carroll, 2001; Lipsitz et al., 1990; Mancl and DeRouen, 2001; McNeish and Stapleton, 2016; Morel et al., 2003; Paik, 1988; Pan, 2001; Paul and Zhang, 2014; Wang and Long, 2011). Unfortunately, the performance of all these estimators has not been investigated concurrently and compared (McNeish and Stapleton, 2016). Importantly, most studies presented numeric simulations to compare a handful of corrections of the modified sandwich variance estimator (Li and Redden, 2015; Lu et al., 2007; McNeish and Harring, 2017; Pan and Wall, 2002; Teerenstra et al., 2010; but see Wang et al., 2016 who compared eight of these corrections). These comparisons have also shown that many factors could impact the performances of GEE in small samples which hampers the choice of a correction for a specific data set: number of panels, number of parameters in the model, variation in panel sample sizes, and nature of response and predictor variables.

In this context, we decided to make use of the correction proposed by Kauermann and Carroll (2001), for two reasons. First, this correction has been extensively used in simulations and, for clustered randomized trials, it has been recommended in small samples for covariates varying at the cluster level with a moderate cluster sample size variation (Li and Redden, 2015; Lu et al., 2007; Teerenstra et al., 2010). These are characteristics shared by our three data sets. Second, the results of an admittedly limited simulation with a Poisson response variable and the second example presented in the study by Wang et al. (2016) about longitudinal data also showed that the Kauermann and Carroll (2001) correction performed satisfactorily with t-tests (see their Figs. 3 and 4). On the contrary, the corrections from Gosho (2014) or Morel et al. (2003) turned out to be conservative in Wang et al. (2016), which is in

agreement, for the latter at least, with the results of McNeish and Harring (2017) for binary predictor variables. The correction from Mancl and DeRouen (2001) also appeared to overcorrect in a number of simulations (Fay and Graubard, 2001; Li and Redden, 2015; Wang and Long, 2011). Finally, we discarded other corrections, e.g. the one from Pan (2001) or (Wang and Long, 2011) that require further assumptions difficult to test with the sparse data sets at hand.

Three other modeling choices are worth briefly commenting on. First, factor year was not included as such in the models because of the sparseness of data. However, the between-year variability was included somehow in the models through the offset variable and the principal components whose coordinates were computed from environmental variables with annual variability. Second, factor site was included as a fixed factor in the models because we do not believe that it is sensible to try to estimate a variance component from only three levels of a factor (sites A, B and C). In a different context, that is, capture-mark-recapture modeling, Burnham and White (2002: 259) run simulations and concluded that a minimum number of ten levels was to be used to estimate correctly a variance component. Third, it is intuitively appealing to use these data of seroprevalence and incidence in some epidemiological SIR models to gain more insight into the viral transmission. Yet the high heterogeneity observed in natural populations of small mammals makes the parameterization of such models from field data difficult. Further work is needed to bridge the gap between the requirements of such epidemiological SIR models and the amount of data that is reasonably possible to collect in the field.

In conclusion, our study highlighted the effects of bank vole demography and environmental characteristics on seasonal PUUV infection, using two epidemiological measures: incidence and seroprevalence. Biological interpretation still remained complicated as several mechanisms can be envisioned to explain the observed relationships between environment and epidemiological measures. Incidence is more precise than seroprevalence to evaluate the moment of rodent infection, but an accurate assessment of incidence requires longitudinal monitoring of individuals with repeated trapping sessions, which implies a heavy investment in both human and financial resources. It still remains unclear whether models including rodent incidence instead of rodent seroprevalence, should be used in addition to environmental risk factors to make decisions to help prevent human disease burden. Thus, we encourage other researchers working with wild host populations of pathogens to use both seroprevalence and incidence measures to enhance our understanding of the impact of environment and reservoir demography on the epidemiology of pathogens.

Funding

This work was supported by the AFSSET (now ANSES), program EST-75 and INSERM/MATE (No. EN99115).

Conflict of interest statement

None.

Acknowledgments

We have appreciated the valuable comments of the editor and two anonymous reviewers on original manuscript. We also thank Camille Delavenne (INRAE) for reading and commenting on an earlier version of the paper, Jean-Michel Demerson, Léo Legras, Claire Combeau, Delphine Muller, Christophe Caillot, Jean-Marc Boucher, Stéphanie Favier, Benoît Combes and Jean-Baptiste Pons for technical help in the field and/or laboratory work, Robin Ristl for his kind help with the mmmgee R-package, the 'Laboratoire d'excellence Dynamiques éco-évolutives des maladies infectieuses' (ANR-11-LABX-0048) and ONF for meteorological data for Croix-Scaille.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.epidem.2022.100600.

References

Abdi, H., Williams, L.J., 2010. Principal component analysis: principal component analysis. Wiley Interdiscip. Rev. Comput. Stat. 2, 433–459. https://doi.org/10.1002/wics.101.

Adler, F.R., Pearce-Duvet, J.M.C., Dearing, M.D., 2008. How host population dynamics translate into time-lagged prevalence: an investigation of Sin Nombre virus in deer mice. Bull. Math. Biol. 70, 236–252. https://doi.org/10.1007/s11538-007-9251-8.

Armansin, N.C., Stow, A.J., Cantor, M., Leu, S.T., Klarevas-Irby, J.A., Chariton, A.A., Farine, D.R., 2020. Social Barriers in Ecological Landscapes: The Social Resistance Hypothesis. Trends Ecol. Evol. 35, 137–148. https://doi.org/10.1016/j.tree.2019.10.001.

Augot, D., Muller, D., Demerson, J.M., Boué, F., Caillot, C., Cliquet, F., 2006. Dynamics of Puumala virus infection in bank voles in Ardennes department (France). Pathol. Biol. 54, 572–577. https://doi.org/10.1016/j.patbio.2006.07.039.

Augot, D., Sauvage, F., Boue, F., Bouloy, M., Artois, M., Demerson, J.M., Combes, B., Coudrier, D., Zeller, H., Cliquet, F., Pontier, D., 2008. Spatial and temporal patterning of bank vole demography and the epidemiology of the Puumala hantavirus in northeastern France. Epidemiol. Infect. 136, 1638–1643. https://doi.org/10.1017/S0950268808000423.

Barrios, J.M., Verstraeten, W.W., Maes, P., Aerts, J.M., Farifteh, J., Coppin, P., 2013. Relating land cover and spatial distribution of nephropathia epidemica and Lyme borreliosis in Belgium. Int. J. Environ. Health Res. 23, 132–154. https://doi.org/10.1080/09603123.2012.708918.

Benhaiem, S., Marescot, L., Hofer, H., East, M.L., Lebreton, J.-D., Kramer-Schadt, S., Gimenez, O., 2018. Robustness of eco-epidemiological capture-recapture parameter estimates to variation in infection state uncertainty. Front. Vet. Sci. 5, 197. https://doi.org/10.3389/fvets.2018.00197.

Bernshtein, A.D., Apekina, N.S., Mikhailova, T.V., Myasnikov, Y.A., Khlyap, L.A., Korotkov, Y.S., Gavrilovskaya, I.N., 1999. Dynamics of Puumala hantavirus infection in naturally infected bank voles (Clethrinomys glareolus). Arch. Virol. 144, 2415–2428.

Botten, J., Mirowsky, K., Kusewitt, D., Bharadwaj, M., Yee, J., Ricci, R., Feddersen, R.M., Hjelle, B., 2000. Experimental infection model for Sin Nombre hantavirus in the deer mouse (Peromyscus maniculatus). Proc. Natl. Acad. Sci. 97, 10578–10583. https://doi.org/10.1073/pnas.180197197.

Burnham, K.P., White, G.C., 2002. Evaluation of some random effects methodology applicable to bird ringing data. J. Appl. Stat. 29, 245–264.

Buzdugan, S.N., Vergne, T., Grosbois, V., Delahay, R.J., Drewe, J.A., 2017. Inference of the infection status of individuals using longitudinal testing data from cryptic populations: towards a probabilistic approach to diagnosis. Sci. Rep. 7, 1111. https://doi.org/10.1038/s41598-017-00806-4.

Cheng, G., Yu, Z., Huang, J.Z., 2013. The cluster bootstrap consistency in generalized estimating equations. J. Multivar. Anal. 115, 33–47. https://doi.org/10.1016/j.jmva.2012.09.003.

Cleaveland, S., Mlengeya, T., Kaare, M., Haydon, D., Lembo, T., Laurenson, M.K., Packer, C., 2007. The conservation relevance of epidemiological research into carnivore viral diseases in the Serengeti. Conserv. Biol. 21, 612–622. https://doi.org/10.1111/j.1523-1739.2007.00701.x.

Clement, J., Vercauteren, J., Verstraeten, W.W., Ducoffre, G., Barrios, J.M., Vandamme, A.-M., Maes, P., Van Ranst, M., 2009. Relating increasing hantavirus incidences to the changing climate: the mast connection. Int. J. Health Geogr. 8, 1. https://doi.org/10.1186/1476-072X-8-1.

Combelles, L., Corbiere, F., Calavas, D., Bronner, A., Hénaux, V., Vergne, T., 2019. Impact of imperfect disease detection on the identification of risk factors in veterinary epidemiology. Front. Vet. Sci. 6, 66. https://doi.org/10.3389/fvets.2019.00066.

Craft, M.E., Caillaud, D., 2011. Network models: an underutilized tool in wildlife epidemiology. Interdiscip. Perspect. Infect. Dis. 2011, 676949 https://doi.org/10.1155/2011/676949.

Douglass, R.J., Calisher, C.H., Wagoner, K.D., Mills, J.N., 2007. Sin Nombre virus infection of deer mice in Montana: characteristics of newly infected mice, incidence, and temporal pattern of infection. J. Wildl. Dis. 43, 12–22. https://doi.org/10.7589/0090-3558-43.1.12.

Drewes, S., Turni, H., Rosenfeld, U.M., Obiegala, A., Straková, P., Imholt, C., Glatthaar, E., Dressel, K., Pfeffer, M., Jacob, J., Wagner-Wiening, C., Ulrich, R.G., 2017. Reservoir-driven heterogeneous distribution of recorded human puumala virus cases in south-west Germany. Zoonoses Public Health 64, 381–390. https://doi.org/10.1111/zph.12319.

Escutenaire, S., Chalon, P., Verhagen, R., Heyman, P., Thomas, I., Karelle-Bui, L., Avsic-Zupanc, T., Lundkvist, A., Plyusnin, A., Pastoret, P., 2000. Spatial and temporal dynamics of Puumala hantavirus infection in red bank vole (Clethrionomys glareolus) populations in Belgium. Virus Res. 67, 91–107.

Escutenaire, S., Chalon, P., De Jaegere, F., Karelle-Bui, L., Mees, G., Brochier, B., Rozenfeld, F., Pastoret, P.-P., 2002. Behavioral, physiologic, and habitat influences on the dynamics of Puumala virus infection in bank voles (Clethrionomys glareolus). Emerg. Infect. Dis. 8, 930–936. https://doi.org/10.3201/eid0809.010537.

Fay, M.P., Graubard, B.I., 2001. Small-sample adjustments for Wald-type tests using sandwich estimators. Biometrics 57, 1198–1206.

Feng, Z., McLerran, D., Grizzle, J., 1996. A comparison of statistical methods for clustered data analysis with Gaussian error. Stat. Med. 15, 1793–1806. https://doi.org/10.1002/(SICI)1097-0258(19960830)15:16<1793::AID-SIM332>3.0.CO;2-2.
Flower, C.E., Dalton, J.E., Whelan, C.J., Brown, J.S., Gonzalez-Meler, M.A., 2019. Patch use in the arctic ground squirrel: effects of micro-topography and shrub encroachment in the Arctic Circle. Oecologia 190, 243–254. https://doi.org/10.1007/s00442-019-04400-5.
Flowerdew, J.R., Amano, T., Sutherland, W.J., 2017. Strong "bottom-up" influences on small mammal populations: State-space model analyses from long-term studies. Ecol. Evol. 7, 1699–1711. https://doi.org/10.1002/ece3.2725.
Glass, G.E., Childs, J.E., Korch, G.W., LeDuc, J.W., 1988. Association of intraspecific wounding with hantaviral infection in wild rats (Rattus norvegicus). Epidemiol. Infect. 101, 459–472. https://doi.org/10.1017/s0950268800054418.
Gliwicz, J., 1989. Individuals and populations of the bank vole in optimal, suboptimal and insular habitats. J. Anim. Ecol. 58, 237. https://doi.org/10.2307/4997.
Gliwicz, J., Ims, R.A., 2000. Dispersal in the bank vole. Pol. J. Ecol. 48, 51–61.
Glorvigen, P., Bjørnstad, O.N., Andreassen, H.P., Ims, R.A., 2012. Settlement in empty versus occupied habitats: an experimental study on bank voles. Popul. Ecol. 54, 55–63. https://doi.org/10.1007/s10144-011-0295-0.
Gosho, M., 2014. Robust covariance estimator for small-sample adjustment in the generalized estimating equations: a simulation study. Sci. J. Appl. Math. Stat. 2, 20. https://doi.org/10.11648/j.sjams.20140201.13.
Gundersen, G., Johannesen, E., Andreassen, H.P., Ims, R.A., 2001. Source-sink dynamics: how sinks affect demography of sources. Ecol. Lett. 4, 14–21. https://doi.org/10.1046/j.1461-0248.2001.00182.x.
Gundersen, G., Andreassen, H.P., Ims, R.A., 2002. Individual and population level determinants of immigration success on local habitat patches: an experimental approach. Ecol. Lett. 5, 294–301. https://doi.org/10.1046/j.1461-0248.2002.00320.x.
Guo, X., Pan, W., Connett, J.E., Hannan, P.J., French, S.A., 2005. Small-sample performance of the robust score test and its modifications in generalized estimating equations. Stat. Med. 24, 3479–3495. https://doi.org/10.1002/sim.2161.
Hardestam, J., Karlsson, M., Falk, K.I., Olsson, G., Klingström, J., Lundkvist, A., 2008. Puumala hantavirus excretion kinetics in bank voles (Myodes glareolus). Emerg. Infect. Dis. 14, 1209–1215. https://doi.org/10.3201/eid1408.080221.
Hardin, J.W., Hilbe, J.M., 2003. Generalized Estimating Equations. Chapman & Hall/CRC Company, Boca Raton, p. 222.
Hazel, S.M., Bennett, M., Chantrey, J., Bown, K., Cavanagh, R., Jones, T.R., Baxby, D., Begon, M., 2000. A longitudinal study of an endemic disease in its wildlife reservoir: cowpox and wild rodents. Epidemiol. Infect. 124, 551–562. https://doi.org/10.1017/s0950268899003799.
Heyman, P., Ceianu, C.S., Christova, I., Tordo, N., Beersma, M., João Alves, M., Lundkvist, A., Hukic, M., Papa, A., Tenorio, A., Zelená, H., Essbauer, S., Visontai, I., Golovljova, I., Connell, J., Nicoletti, L., Van Esbroeck, M., Gjeruldsen Dudman, S., Aberle, S.W., Avšič-Županc, T., Korukluoglu, G., Nowakowska, A., Klempa, B., Ulrich, R.G., Bino, S., Engler, O., Opp, M., Vaheri, A., 2011. A five-year perspective on the situation of haemorrhagic fever with renal syndrome and status of the hantavirus reservoirs in Europe, 2005-2010. Eur. Surveill. 16. https://doi.org/10.2807/ese.16.36.19961-en.
Jones, K.E., Patel, N.G., Levy, M.A., Storeygard, A., Balk, D., Gittleman, J.L., Daszak, P., 2008. Global trends in emerging infectious diseases. Nature 451, 990–993. https://doi.org/10.1038/nature06536.
Kallio, E.R., Klingström, J., Gustafsson, E., Manni, T., Vaheri, A., Henttonen, H., Vapalahti, O., Lundkvist, A., 2006a. Prolonged survival of Puumala hantavirus outside the host: evidence for indirect transmission via the environment. J. Gen. Virol. 87, 2127–2134. https://doi.org/10.1099/vir.0.81643-0.
Kallio, E.R., Poikonen, A., Vaheri, A., Vapalahti, O., Henttonen, H., Koskela, E., Mappes, T., 2006b. Maternal antibodies postpone hantavirus infection and enhance individual breeding success. Proc. Biol. Sci. 273, 2771–2776. https://doi.org/10.1098/rspb.2006.3645.
Kallio, E.R., Begon, M., Henttonen, H., Koskela, E., Mappes, T., Vaheri, A., Vapalahti, O., 2010. Hantavirus infections in fluctuating host populations: the role of maternal antibodies. Proc. Biol. Sci. 277, 3783–3791. https://doi.org/10.1098/rspb.2010.1022.
Kallio, E.R., Helle, H., Koskela, E., Mappes, T., Vapalahti, O., 2015. Age-related effects of chronic hantavirus infection on female host fecundity. J. Anim. Ecol. 84, 1264–1272. https://doi.org/10.1111/1365-2656.12387.
Karesh, W.B., Dobson, A., Lloyd-Smith, J.O., Lubroth, J., Dixon, M.A., Bennett, M., Aldrich, S., Harrington, T., Formenty, P., Loh, E.H., Machalaba, C.C., Thomas, M.J., Heymann, D.L., 2012. Ecology of zoonoses: natural and unnatural histories. Lancet 380, 1936–1945. https://doi.org/10.1016/S0140-6736(12)61678-X.
Karlsson, A.F., Potapov, E.R., 1998. Consistency and change in bank vole (*Clethrionomys glareolus*) home ranges across a decade. Can. J. Zool. 76, 1329–1334. https://doi.org/10.1139/cjz-76-7-1329.
Kauermann, G., Carroll, R.J., 2001. A note on the efficiency of sandwich covariance matrix estimation. J. Am. Stat. Assoc. 96, 1387–1396. https://doi.org/10.1198/016214501753382309.
Khalil, H., Olsson, G., Magnusson, M., Evander, M., Hörnfeldt, B., Ecke, F., 2017. Spatial prediction and validation of zoonotic hazard through micro-habitat properties: where does Puumala hantavirus hole – up? BMC Infect. Dis. 17. https://doi.org/10.1186/s12879-017-2618-z.
Levins, R., 1970. Extinction. In: Gerstenhaber, M. (Ed.). American Mathematical Society.
Li, P., Redden, D.T., 2015. Small sample performance of bias-corrected sandwich estimators for cluster-randomized trials with binary outcomes: P. LI AND D. T. REDDEN. Stat. Med. 34, 281–296. https://doi.org/10.1002/sim.6344.
Liang, K.-Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22. https://doi.org/10.1093/biomet/73.1.13.
Linard, C., Tersago, K., Leirs, H., Lambin, E.F., 2007. Environmental conditions and Puumala virus transmission in Belgium. Int. J. Health Geogr. 6, 55. https://doi.org/10.1186/1476-072X-6-55.
Lipsitz, S.R., Laird, N.M., Harrington, D.P., 1990. Using the jackknife to estimate the variance of regression estimators from repeated measures studies. Commun. Stat. Theory Methods 19, 821–845. https://doi.org/10.1080/03610929008830234.
Lu, B., Preisser, J.S., Qaqish, B.F., Suchindran, C., Bangdiwala, S.I., Wolfson, M., 2007. A comparison of two bias-corrected covariance estimators for generalized estimating equations. Biometrics 63, 935–941. https://doi.org/10.1111/j.1541-0420.2007.00764.x.
Luis, A.D., Douglass, R.J., Hudson, P.J., Mills, J.N., Bjørnstad, O.N., 2012. Sin Nombre hantavirus decreases survival of male deer mice. Oecologia 169, 431–439. https://doi.org/10.1007/s00442-011-2219-2.
Lyubsky, S., Gavrilovskaya, I., Luft, B., Mackow, E., 1996. Histopathology of Peromyscus leucopus naturally infected with pathogenic NY-1 hantaviruses: pathologic markers of HPS viral infection in mice. Lab. Investig. 74, 627–633.
MacCallum, H., 2000. Population parameters: estimation of ecological models. Methods in Ecology. Blackwell Science, Oxford.
Mancl, L.A., DeRouen, T.A., 2001. A covariance estimator for GEE with improved small-sample properties. Biometrics 57, 126–134.
McNeish, D., Harring, J.R., 2017. Correcting model fit criteria for small sample latent growth models with incomplete data. Educ. Psychol. Meas. 77, 990–1018. https://doi.org/10.1177/0013164416661824.
McNeish, D., Stapleton, L.M., 2016. Modeling clustered data with very few clusters. Multivar. Behav. Res. 51, 495–518. https://doi.org/10.1080/00273171.2016.1167008.
Mills, J.N., Yates, T.L., Ksiazek, T.G., Peters, C.J., Childs, J.E., 1999. Long-term studies of hantavirus reservoir populations in the southwestern United States: rationale, potential, and methods. Emerg. Infect. Dis. 5, 95–101. https://doi.org/10.3201/eid0501.990111.
Monchatre-Leroy, E., Crespin, L., Boué, F., Marianneau, P., Calavas, D., Hénaux, V., 2017. Spatial and temporal epidemiology of nephropathia epidemica incidence and hantavirus seroprevalence in rodent hosts: identification of the main environmental factors in Europe. Transbound. Emerg. Dis. 64, 1210–1228. https://doi.org/10.1111/tbed.12494.
Morel, J.G., Bokossa, M.C., Neerchal, N.K., 2003. Small Sample Correction for the Variance of GEE Estimators. Biom. J. 45, 395–409. https://doi.org/10.1002/bimj.200390021.
Netski, D., Thran, B.H., St Jeor, S.C., 1999. Sin Nombre virus pathogenesis in Peromyscus maniculatus. J. Virol. 73, 585–591.
Olsson, G.E., White, N., Ahlm, C., Elgh, F., Verlemyr, A.-C., Juto, P., Palo, R.T., 2002. Demographic factors associated with hantavirus infection in bank voles (Clethrionomys glareolus). Emerg. Infect. Dis. 8, 924–929. https://doi.org/10.3201/eid0809.020037.
Olsson, G.E., White, N., Hjältén, J., Ahlm, C., 2005. Habitat factors associated with bank voles (Clethrionomys glareolus) and concomitant hantavirus in northern Sweden. Vector Borne Zoonotic Dis. 5, 315–323. https://doi.org/10.1089/vbz.2005.5.315.
Otis, D.L., Burnham, K.P., White, G.C., Anderson, D.R., 1978. Statistical inference from capture data on closed animal populations. Wildl. Monogr. 62, 1–135.
Paik, M.C., 1988. Repeated measurement analysis for nonnormal data in small samples. Commun. Stat. Simul. Comput. 17, 1155–1171. https://doi.org/10.1080/03610918808812718.
Pan, W., 2001. On the robust variance estimator in generalised estimating equations. Biometrika 88, 901–906. https://doi.org/10.1093/biomet/88.3.901.
Pan, W., Wall, M.M., 2002. Small-sample adjustments in using the sandwich variance estimator in generalized estimating equations. Stat. Med. 21, 1429–1441. https://doi.org/10.1002/sim.1142.
Paul, S., Zhang, X., 2014. Small sample GEE estimation of regression parameters for longitudinal data. Stat. Med. 33, 3869–3881. https://doi.org/10.1002/sim.6198.
Penalba, C., Galempoix, J.M., Lanoux, P., 2001. Épidémiologie des infections à hantavirus en France. Méd. Mal. Infect. 31, 272–284. https://doi.org/10.1016/S0399-077X(01)80067-6.
Pettorelli, N., Vik, J.O., Mysterud, A., Gaillard, J.-M., Tucker, C.J., Stenseth, N.C., 2005. Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends Ecol. Evol. 20, 503–510. https://doi.org/10.1016/j.tree.2005.05.011.
Piechotowski, I., Brockmann, S.O., Schwarz, C., Winter, C.H., Ranft, U., Pfaff, G., 2008. Emergence of hantavirus in South Germany: rodents, climate and human infections. Parasitol. Res. 103 (Suppl. 1), S131–S137. https://doi.org/10.1007/s00436-008-1055-8.
Prevot-Julliard, A.-C., Henttonen, H., Yoccoz, N.G., Stenseth, N.Ch.R., 1999. Delayed maturation in female bank voles: optimal decision or social constraint. J. Anim. Ecol. 68, 684–697. https://doi.org/10.1046/j.1365-2656.1999.00307.x.
Pulliam, H.R., 1988. Sources, sinks and population regulation. Am. Nat. 652–661.
R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. ⟨https://www.R-project.org⟩.
Reijniers, J., Tersago, K., Borremans, B., Hartemink, N., Voutilainen, L., Henttonen, H., Leirs, H., 2020. Why hantavirus prevalence does not always increase with host density: modeling the role of host spatial behavior and maternal antibodies. Front. Cell. Infect. Microbiol. 10. https://doi.org/10.3389/fcimb.2020.536660.
Reil, D., Imholt, C., Eccard, J.A., Jacob, J., 2015. Beech fructification and bank vole population dynamics - combined analyses of promoters of human Puumala virus infections in Germany. PLOS ONE 10, e0134124. https://doi.org/10.1371/journal.pone.0134124.

Reil, D., Rosenfeld, U.M., Imholt, C., Schmidt, S., Ulrich, R.G., Eccard, J.A., Jacob, J., 2017. Puumala hantavirus infections in bank vole populations: host and virus dynamics in Central Europe. BMC Ecol. 17, 9. https://doi.org/10.1186/s12898-017-0118-z.

Reynes, J.-M., Carli, D., Thomas, D., Castel, G., 2019. Puumala hantavirus genotypes in humans, France, 2012-2016. Emerg. Infect. Dis. 25, 140–143. https://doi.org/10.3201/eid2501.180270.

Ristl, R., Hothorn, L., Ritz, C., Posch, M., 2019. Simultaneous inference for multiple marginal generalized estimating equation models, 096228021987300 Stat. Methods Med. Res.. https://doi.org/10.1177/0962280219873005.

Schwarz, A.C., Ranft, U., Piechotowski, I., Childs, J.E., Brockmann, S.O., 2009. Risk factors for human infection with Puumala virus, southwestern Germany. Emerg. Infect. Dis. 15, 1032–1039. https://doi.org/10.3201/eid1507.081413.

Swart, A., Bekker, D.L., Maas, M., de Vries, A., Pijnacker, R., Reusken, C.B.E.M., van der Giessen, J.W.B., 2017. Modelling human Puumala hantavirus infection in relation to bank vole abundance and masting intensity in the Netherlands. Infect. Ecol. Epidemiol. 7, 1287986. https://doi.org/10.1080/20008686.2017.1287986.

Teerenstra, S., Lu, B., Preisser, J.S., van Achterberg, T., Borm, G.F., 2010. Sample size considerations for GEE analyses of three-level cluster randomized trials. Biometrics 66, 1230–1237. https://doi.org/10.1111/j.1541-0420.2009.01374.x.

Tersago, K., Schreurs, A., Linard, C., Verhagen, R., Van Dongen, S., Leirs, H., 2008. Population, environmental, and community effects on local bank vole (Myodes glareolus) Puumala virus infection in an area with low human incidence. Vector Borne Zoonotic Dis. 8, 235–244. https://doi.org/10.1089/vbz.2007.0160.

Tersago, K., Verhagen, R., Vapalahti, O., Heyman, P., Ducoffre, G., Leirs, H., 2011. Hantavirus outbreak in Western Europe: reservoir host infection dynamics related to human disease patterns. Epidemiol. Infect. 139, 381–390. https://doi.org/10.1017/S0950268810000956.

Tersago, K., Crespin, L., Verhagen, R., Leirs, H., 2012. Impact of Puumala virus infection on maturation and survival in bank voles: a capture-mark-recapture analysis. J. Wildl. Dis. 48, 148–156. https://doi.org/10.7589/0090-3558-48.1.148.

Thoma, B.R., Müller, J., Bässler, C., Georgi, E., Osterberg, A., Schex, S., Bottomley, C., Essbauer, S.S., 2014. Identification of factors influencing the Puumala virus seroprevalence within its reservoir in a Montane forest environment. Viruses 6, 3944–3967. https://doi.org/10.3390/v6103944.

Thrusfield, M., 2005. Veterinary Epidemiology, third ed. Wiley-Blackwell Edition, p. 624.

van Apeldoorn, R.C., Oostenbrink, W.T., van Winden, A., van der Zee, F.F., 1992. Effects of habitat fragmentation on the bank vole, Clethrionomys glareolus, in an agricultural landscape. Oikos 65, 265. https://doi.org/10.2307/3545018.

Vanwambeke, S.O., Zeimes, C.B., Drewes, S., Ulrich, R.G., Reil, D., Jacob, J., 2019. Spatial dynamics of a zoonotic orthohantavirus disease through heterogenous data on rodents, rodent infections, and human disease. Sci. Rep. 9, 2329. https://doi.org/10.1038/s41598-019-38802-5.

Vergne, T., Paul, M.C., Chaengprachak, W., Durand, B., Gilbert, M., Dufour, B., Roger, F., Kasemsuwan, S., Grosbois, V., 2014. Zero-inflated models for identifying disease risk factors when case detection is imperfect: application to highly pathogenic avian influenza H5N1 in Thailand. Prev. Vet. Med. 114, 28–36. https://doi.org/10.1016/j.prevetmed.2014.01.011.

Viel, J.-F., Lefebvre, A., Marianneau, P., Joly, D., Giraudoux, P., Upegui, E., Tordo, N., Hoen, B., 2011. Environmental risk factors for haemorrhagic fever with renal syndrome in a French new epidemic area. Epidemiol. Infect. 139, 867–874. https://doi.org/10.1017/S0950268810002062.

Voutilainen, L., Sironen, T., Tonteri, E., Bäck, A.T., Razzauti, M., Karlsson, M., Wahlström, M., Niemimaa, J., Henttonen, H., Lundkvist, Å., 2015. Life-long shedding of Puumala hantavirus in wild bank voles (Myodes glareolus). J. Gen. Virol. 96, 1238–1247. https://doi.org/10.1099/vir.0.000076.

Voutilainen, L., Kallio, E.R., Niemimaa, J., Vapalahti, O., Henttonen, H., 2016. Temporal dynamics of Puumala hantavirus infection in cyclic populations of bank voles. Sci. Rep. 6, 21323. https://doi.org/10.1038/srep21323.

Wang, M., Long, Q., 2011. Modified robust variance estimator for generalized estimating equations with improved small-sample performance. Stat. Med. 30, 1278–1291. https://doi.org/10.1002/sim.4150.

Wang, M., Kong, L., Li, Z., Zhang, L., 2016. Covariance estimators for generalized estimating equations (GEE) in longitudinal analysis with small samples. Stat. Med. 35, 1706–1721. https://doi.org/10.1002/sim.6817.

Wiger, R., 1979. Demography of a cyclic population of the bank vole Clethrionomys glareolus. Oikos 33, 373. https://doi.org/10.2307/3544325.

Yanagihara, R., Amyx, H.L., Gajdusek, D.C., 1985. Experimental infection with Puumala virus, the etiologic agent of nephropathia epidemica, in bank voles (Clethrionomys glareolus). J. Virol. 55, 34–38. https://doi.org/10.1128/JVI.55.1.34-38.1985.

Yoccoz, N.G., Mesnager, Séverine, Mesnager, Severine, 1998. Are alpine bank voles larger and more sexually dimorphic because adults survive better? Oikos 82, 85. https://doi.org/10.2307/3546919.