PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Ketola T, Briga M, Honkola T, Lummaa V. 2021 Town population size and structuring into villages and households drive infectious disease risks in pre-healthcare Finland. *Proc. R. Soc. B* **288**: 20210356. https://doi.org/10.1098/rspb.2021.0356

Received: 11 February 2021
Accepted: 25 March 2021

Subject Category:
Ecology

Subject Areas:
health and disease and epidemiology, evolution, ecology

Keywords:
infection, meta-population, social lifestyle

Author for correspondence:
Tarmo Ketola
e-mail: tketola@jyu.fi

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5368975.

THE ROYAL SOCIETY PUBLISHING

Town population size and structuring into villages and households drive infectious disease risks in pre-healthcare Finland

Tarmo Ketola[1], Michael Briga[2], Terhi Honkola[2,3] and Virpi Lummaa[2]

[1]Department of Biological and Environmental Science, University of Jyväskylä, PO Box 35, 40014 Jyväskylä, Finland
[2]Department of Biology, University of Turku, Turku 20014, Finland
[3]Department of Anthropology and Archaeology, University of Bristol, Bristol BS8 1UU, UK

TK, 0000-0001-5151-7980; MB, 0000-0003-3160-0407; TH, 0000-0002-3330-0857; VL, 0000-0002-2128-7587

Social life is often considered to cost in terms of increased parasite or pathogen risk. However, evidence for this in the wild remains equivocal, possibly because populations and social groups are often structured, which affects the local transmission and extinction of diseases. We test how the structuring of towns into villages and households influenced the risk of dying from three easily diagnosable infectious diseases—smallpox, pertussis and measles—using a novel dataset covering almost all of Finland in the pre-healthcare era (1800–1850). Consistent with previous results, the risk of dying from all three diseases increased with the local population size. However, the division of towns into a larger number of villages decreased the risk of dying from smallpox and to some extent of pertussis but it slightly increased the risk for measles. Dividing towns into a larger number of households increased the length of the epidemic for all three diseases and led to the expected slower spread of the infection. However, this could be seen only when local population sizes were small. Our results indicate that the effect of population structure on epidemics, disease or parasite risk varies between pathogens and population sizes, hence lowering the ability to generalize the consequences of epidemics in spatially structured populations, and mapping the costs of social life, via parasites and diseases.

1. Introduction

Understanding the health consequences of living in large groups by some species versus the solitary life of others is a key conundrum in behavioural biology that also has strong implications for public health in human societies [1]. To explain the large variability in group size, research has concentrated on identifying the costs and benefits of group living, with the greatest costs of group living thought to be increased disease and parasite risk [2,3]: larger and denser groups or populations are more prone to infectious diseases and parasitism [4–6]. Surprisingly, however, strong positive correlations between group size and parasite prevalence are rarely found in the wild [3]. For example, in their meta-analysis of natural populations, Rifkin *et al*. [3] found only a small ($r = 0.187$) positive relationship between group size and parasite prevalence, with the highest values for vector-mediated diseases (0.396) followed by directly transmitted infectious diseases (0.231). However, these results were mainly driven by birds that have much larger group sizes than many other species, including for example mammals for which such associations were absent [3].

One reason why group size can be a rather poor predictor of parasite or disease prevalence in the wild is structuring, i.e. the existence of (sub)groups within groups or populations. Structuring can arise from kin associations, dominance hierarchies and behavioural specialization [2]. Disease transmission between these (sub)groups may be limited owing to reduced encounter rates

and imperfect mixing [7]. Thus, structuring can determine the spread of diseases more than the actual population or group size [8–10]. For example, according to the 'social bottle-neck hypothesis', infections and parasites could be more abundant when the population structure is more uniform [2,10], and in the most extreme case of structuring, infections and parasites will be restricted to spreading within subgroups rather independently from each other, lowering the total number infected in the population [7].

The effects of group size and the number of subgroups are however often entangled: when group or population size increases, there is often an increase in the number of subgroups [2]. For example, in primates, larger populations were found to be more modular and less connected (e.g. [11]). Such a relationship between population size and number of subgroups requires studies that can statistically separate both effects. However, this can be difficult to accomplish with natural populations as social groups or population structure could be hard to detect in routine population size censuses. The biology of the diseases in question can also have an effect on how population size or structure affects the transmission and prevalence of diseases (e.g. [12]; [13]). Therefore, to understand how population structure affects disease dynamics, a comparison of several diseases in the same landscape, or population structure, is necessary. Yet, very few natural populations have been explored simultaneously for several diseases, and in modern human societies where such information could be relatively easily accessible, early medical intervention prevents the natural transmission of diseases.

Here, we investigate the effects of population structuring on mortality owing to three childhood diseases that were major causes of death in populations of pre-healthcare societies: smallpox, pertussis and measles. We use nationwide mortality registers from pre-healthcare Finland between 1800 and 1850, combined with data on population structuring to towns, villages and separate households across the country. In the nineteenth century, healthcare in Finland was minimal [14]. Smallpox vaccinations started in 1802 and were slowly progressing during the study period. In addition, general healthcare was almost non-existent, with for example in 1820 only 373 hospital places towards 1.2 million inhabitants [15]. Owing to this, infections spread more naturally and were possibly more sensitive to social and environmental factors than in many contemporary populations with access to medical care. The extensive records on causes of death maintained by the church for the entire population of Finland enable us to describe the epidemics of three distinguishable directly transmitted childhood infections: smallpox, pertussis and measles.

Measles, pertussis and smallpox infections were important causes of death in pre-healthcare Finland (see Material and methods). Despite similar droplet mediated transmission, these diseases are different in many ways. For example, pertussis and measles have high R_0 in comparison to smallpox [16], and smallpox and measles are strongly immunizing while pertussis immunity is known to wane [17]. Such variation has been shown to have a strong role in epidemics when explored in the same population [18]. However, to our knowledge, no systematic study has been conducted in which these diseases are studied across a range of differently sized towns, with differing amounts of structuring to houses and villages.

Our data consist of 317 differently sized towns, with varying numbers of villages and households, covering the less sparsely inhabited southern part of Finland. In this paper, we test whether population size and population structuring to villages and households affect the risk of death and the number of months of infection. More specifically, we test: (i) if the population size of a town is positively linked with the risk of death from infectious diseases [4–6], and (ii) if the division into subgroups, such as number of households and villages within a given town, decreases the number of casualties as is suggested by the 'social bottle-neck hypothesis' [2]. Moreover, we test if town characteristics also affect the number of months with infection, which summarizes how often and long epidemics were persistent in towns. For example, village-rich towns could slow down the spread of epidemics but yield larger numbers of months with infections present, in comparison to equally sized towns with smaller numbers of villages.

2. Material and methods

In Finland, the Lutheran Church has been obligated by law to maintain records of all births, deaths, marriages and migration events between parishes in the entire country since 1749. The original parish records have been digitized by the Genealogical Society of Finland and are available at http://hiski.genealogia.fi/historia/indexe.htm. Large towns (e.g. Turku, Viipuri Helsinki, Heinola, Jyväskylä, Kuopio and Sortavala) included several parishes, and we pooled these to reflect the same administrative unit. Hereafter, we refer to parishes or larger towns with several parishes pooled, as towns.

For the purpose of our research aims, we focused on information collected between 1800 and 1850, which is the time period best covered by the records, as the collection of death records had already been standardized over the first 50 years (from 1749). In our analyses, we concentrated on the most densely populated area of Finland south of the Arctic Circle with approximately 800 000 inhabitants in 200 000 km^2 (figure 1). The population was agrarian and subject to large fluctuations in mortality and fertility with 25% of children dying before the age of 1 and *ca* 50% by the age of 20, often from infectious diseases [19,20]. The excluded, northernmost part of Finland was sparsely populated mainly by nomadic Sami-people who depended mostly on fishing and hunting for their livelihood, leading to incomplete church records [21]. The study period largely predates the onset of industrialization, improved healthcare and demographic transition in Finland [19].

The data used here therefore contain a total of 1 223 308 registered deaths of which 52 834 are owing to pertussis (4.3%), 45 127 to smallpox (3.7%) and 26 123 to measles (2.1%). Fifteen per cent (185 399) of the deaths lacks a documented cause. Although overall disease diagnostics during the era have been criticized [22], smallpox, pertussis and measles are among the most recognizable of the diseases with clear and distinct symptoms [23]. We identified the causes of death from the database by combining typographical variants and abbreviations and synonyms of diseases in the different languages used in the church records (Finnish, Swedish and German) following Vuorinen [22]. This was done by two authors (T.K. and M.B.) independently and results were consistent.

The analysed dataset contains 317 towns, with data on diseases and explanatory variables (figure 1). In these towns, population size estimates based on data from the year 1882 ranged from 300 to 24 315 with a mean of 4030.24 inhabitants and s.d. of 2826.99 [24]. For analyses, we used town area [25], from which the area covered by lakes was subtracted (range = 15.9 to 7990.9, mean = 492.0 and s.d. = 771.1). We also extracted from the church records information on the number of villages for each town (range = 1–99, mean = 20.82, s.d. =

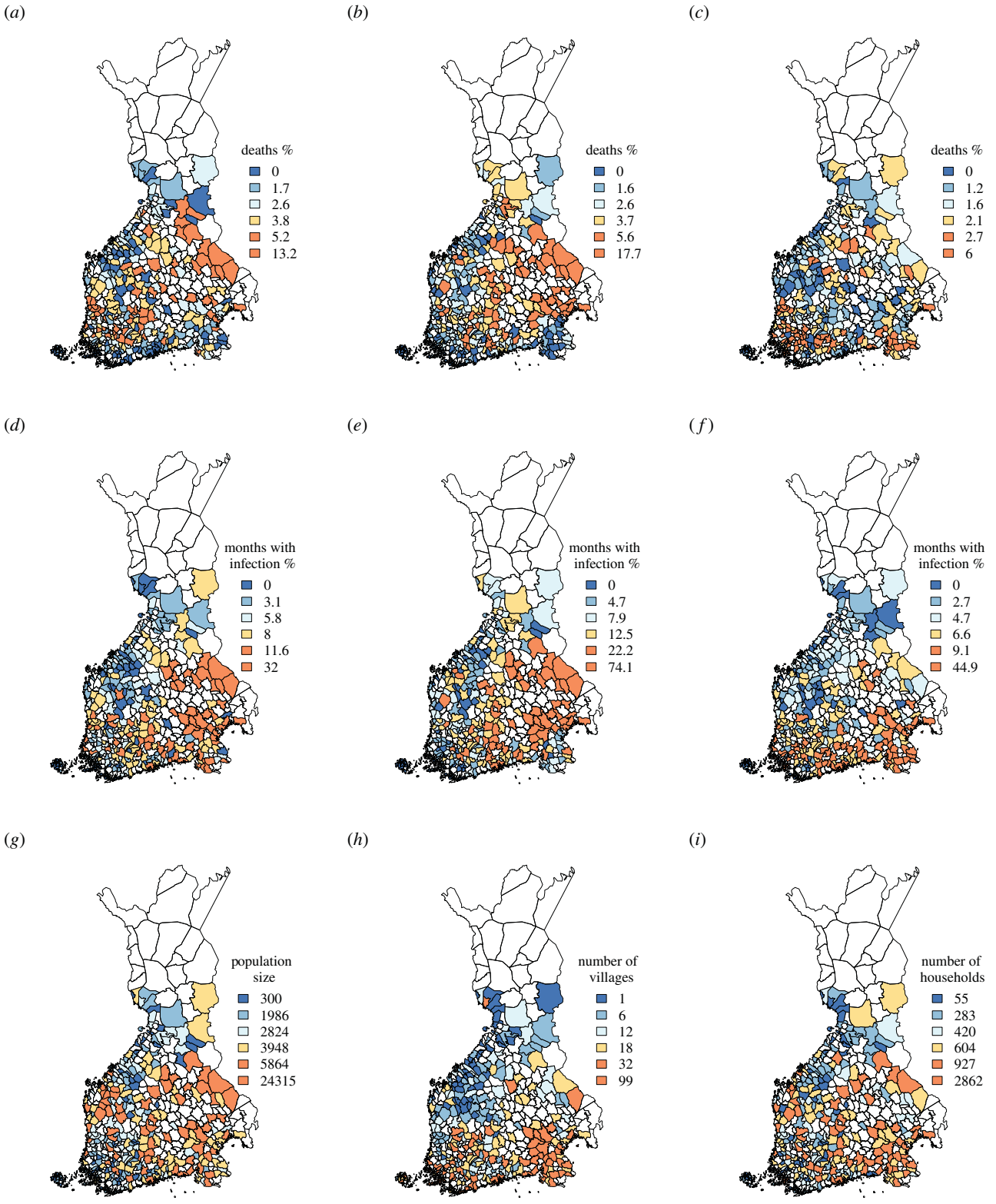

Figure 1. Maps depicting the raw data from 317 Finnish towns in Finland below the Arctic Circle (66° 33′) included in this study. White areas indicate either missing data or excluded areas. Panels show the proportion of deaths from all deaths owing to smallpox (*a*), pertussis (*b*) and measles (*c*). Proportion of months that infections were present (at least one casualty) in towns for smallpox (*d*), pertussis (*e*) and measles (*f*). In the analyses, we tested the effect of predictor variables: area, population size (*g*), number of villages (*h*) and number of households (*i*) on infectious disease mortality. The legends indicate lower limits of class (five quantiles), whereas for the largest class, the upper limit is also depicted. Parish borders depict those determined in 1930. (Online version in colour.)

18.58), which indicates how the town inhabitants were spread into smaller units. The number of households varied between towns from 55 to 2862 with a mean of 609.7 (s.d. = 404.0, [24]). Population size correlated positively with the number of villages: $r = 0.38$, $p < 0.001$, number of households: $r = 0.69$, $p < 0.001$, and town area $r = 0.26$, $p < 0.001$. In addition, the number of households was positively correlated with the number of villages: $r = 0.43$, $p < 0.001$ and town area: $r = 0.17$, $p < 0.001$. The number of villages was not associated with town area: $r = -0.03$, $p = 0.582$.

(a) Data analysis

To test how town structuring affects infectious disease risk in pre-healthcare Finland, we conducted two analyses. In the first, we explored the risk of dying of an infectious disease. Here,

the dependent variable was the number of deaths owing to a specific infection per town, which was included as a binomial factor (infection versus all other deaths). In the second, we tested the risk of having a month with at least one casualty per infection per town, which indicates how often and long a given town was affected by the epidemic. It is noteworthy that a larger threshold gave consistent results. Here, the dependent variable was the months between 1800 and 1850 coded as a binomial factor for when a month had at least one death owing to a specific infection or zero deaths for a particular infection. These were modelled using generalized linear models with a binomial logit link function using the glmer function in the lme4 package [26] in R (v. 3.6.3). We analysed each infectious disease separately, as building a more elaborate epidemic model with several diseases [27] was not feasible with this dataset including very small towns.

The explanatory variables describing town characteristics were town population size, number of villages, number of households and town area. These were standardized to a mean of zero and s.d. of one, to facilitate model convergence and to produce comparable coefficients. In addition, we fitted a random effect of town identity to control for the non-independency of observations arising from including several observations (persons, or months) from the same town. Owing to complex epidemics, and sparsely occurring infections in the smallest parishes, time dependency was not explicitly modelled and hence our estimates refer to average risks over the 50-year period.

The model building was done stepwise, starting from main effects, and gradually building up model complexity by testing various combinations of interactions (electronic supplementary material, table S1) and resolving model fit with log likelihood tests. In the case of significant interactions, we used the Johnson–Neyman procedure to find out where the interaction played a role in determining response variables. After the analysis, we checked the distribution of model residuals with the R package Dharma, all fulfilling the requirements of normally distributed and homoscedastic residuals.

However, all of the models indicated spatial autocorrelation of residuals (Moran's I $p < 0.05$, package ape) and hence we reran models with a method that can effectively correct the interpretational problems arising from spatially autocorrelated residuals. This straightforward method [28] is based on fitting fixed covariates describing the autocorrelation structures (principal coordinates of neighbourhood matrix (PCNM) variables) alongside descriptive variables and is flexible for the exact modelling method or the package used. This flexibility was also our main motivation for choosing this method. This method is frequently used in ecological studies and also in epidemiological modelling [29,30]. Altogether, 163 positively autocorrelated PCNM variables were obtained by R-package vegan, and they were fitted sequentially by adding one PCNM variable at the time to the base model (see above). The decision to include them in the model was based on their ability to correct the spatial autocorrelation of the residuals. The best model was found when Moran's I indicated no spatial autocorrelation of residuals. This model fitting method is considered the most accurate of the tested methods for modelling autocorrelation structures by PCNM, or related, methods [31]. In this study, we do not describe autocorrelation structures in detail and omit the PCNM variables in tables, but show the full tables containing PCNM variables in the electronic supplementary material, tables S2 and S3.

3. Results

The same models were found best for the risk of death and number of months with infection. We found out that best modes describing smallpox were models with effects of population size, village number, household number, land area, population size by household number interaction and village number by land area interaction. In pertussis and measles, the best models were otherwise the same but excluding village number by land area interaction (electronic supplementary material, table S1). We discuss these models in detail below.

(a) Risk of death

First, we investigated whether there was a positive association between the risk of death from childhood infections and town population size, and we found the expected positive association for all three infections (table 1 and figure 2*a–c*). For example, an average town size had 4000 inhabitants and an increase by 1000 inhabitants increased the odds of death from smallpox by 7.4% (rather than dying of something else), from measles by 4.0% and from pertussis by 1%.

The structuring of towns into a larger number of villages affected the risk of death from childhood infections. For pertussis, the risk of dying decreased when a town had more villages: when the number of villages increased by a factor of 10, the odds of dying from pertussis decreased by 3.8%, but this decrease was only tentatively significant ($p = 0.08$; table 1 and figure 3*b*). Interestingly, for measles, the effect was the opposite, but not statistically significant either ($p = 0.107$; table 1 and figure 3*c*). For smallpox, however, structuring into more villages clearly decreased the risk of dying ($p = 0.005$; table 1 and figure 3*a*). On average, a town had 20 villages and an increase by 10 villages (everything else remaining equal) decreased the odds of dying from smallpox by 6.2%. For smallpox, the risk-reducing effect of structuring into villages was more profound in towns with smaller areas and, thus, more densely inhabited towns (table 1 and figure 3*a*). To summarize, structuring a town into a larger number of villages tended to decrease the risk of dying from the childhood infections smallpox and pertussis, but not measles.

Structuring a town into a larger number of households decreased the risk of dying (everything else remaining equal) from smallpox and measles, but this effect was only detectable when population sizes were high (i.e. above 7550 and 9761 inhabitants for smallpox and measles, respectively (table 1 and figure 2*a*,*c*)). For pertussis, the interaction was in the same direction as for the other infections, but it was not statistically significant ($p = 0.092$; table 1 and figure 2*b*). It is noteworthy that while for smallpox and measles the interaction is statistically significant, most towns in rural nineteenth-century Finland had population sizes smaller than the aforementioned thresholds for having a statistically significant household effect (figure 2). Thus, structuring towns into more households decreased the risk of death by smallpox or measles, but only in the biggest towns.

(b) Number of months with infection

We then investigated whether population structuring increased the number of months in infection, which we calculated as a risk of having a month with at least one infectious death (i.e. months with infection). Just as for the risk of death, the probability of having a month with an infection increased with increasing population size (table 2 and figure 2*d–f*). For example, increasing a town size by 1000 inhabitants increased

Table 1. The estimated risk of death owing to three contagious diseases, from total number of deaths, owing to town characteristics (population size, number of villages and households and land area) between 1800 and 1850 in Finland (below Arctic Circle, 66° 33′). (Estimates correspond to *z*-standardized values. Italics denote statistically significant result $p < 0.05$.)

		risk of death			
		estimate	s.e.	*z*-value	Pr(>|*z*|)
smallpox	intercept	−3.4441	0.0430	−80.0796	*<0.001*
	population size	0.1911	0.0597	3.2007	*0.0014*
	number of villages	−0.1227	0.0441	−2.7851	*0.0054*
	number of households	−0.0197	0.0540	−0.3656	0.7146
	area	0.0735	0.0538	1.3657	0.1720
	population size × number of households	−0.0793	0.0301	−2.6347	*0.0084*
	number of villages × area	0.1785	0.0765	2.3338	*0.0196*
pertussis	intercept	−3.4250	0.0390	−87.8175	*<0.001*
	population size	0.2559	0.0551	4.6447	*<0.001*
	number of villages	−0.0730	0.0417	−1.7493	0.0802
	number of households	−0.0227	0.0512	−0.4434	0.6575
	area	−0.0101	0.0379	−0.2659	0.7903
	population size × number of households	−0.0465	0.0275	−1.6878	0.0915
measles	intercept	−3.9594	0.0286	−138.3499	*<0.001*
	population size	0.1063	0.0399	2.6631	*0.0077*
	number of villages	0.0486	0.0302	1.6102	0.1074
	number of households	−0.0023	0.0361	−0.0634	0.9494
	area	0.0368	0.0264	1.3931	0.1636
	population size × number of households	−0.0459	0.0200	−2.2962	*0.0217*

the odds of having months with infection by 22% for smallpox, by 36.6% for pertussis and by 16.5% for measles.

For all three infections, the effect of structuring into households on having a month with infection interacted with population size (table 2 and figure 2*d–f*). At lower population sizes (below 3146, 5174 and 6011 inhabitants for smallpox, pertussis and measles, respectively), a higher number of households was associated with an increasing number of months with infection, consistent with the idea of a more persistent epidemic (figure 2*d–f*). By contrast, at high population sizes (i.e. above 6978, 9623 and 10 408 inhabitants for smallpox, pertussis and measles, respectively), a high number of households was associated with a lower risk of having months with infectious deaths suggesting faster epidemics (figure 2*d–f*). Thus, the effect of population structuring on the speed of the spread of diseases depended on population size and the expected slowing down of spread occurred only when population sizes were small.

For smallpox, towns with a small land area (less than 612.76 km^2), the high number of villages was associated with a smaller number of months with infection (table 2 and figure 4*a*). In towns with a very large area, the effect was estimated to be the opposite, but as only a few towns are within this range it is hard to conclude the biological significance of this result. The number of villages had no effect on the number of months with pertussis infection (table 2). By contrast, for measles, structuring towns into more villages greatly increased the number of months with infection, indicating possibly a longer spread of the epidemic (table 2 and figure 4*b*): everything else remaining equal, increasing a town with 10 more villages increases the odds of having at least one measles-casualty per month by 10%.

4. Discussion

In this study, using a novel dataset covering almost the entirety of pre-healthcare Finland, we found that people living in towns with large population sizes had an increased risk of dying from three infectious diseases—measles, smallpox and pertussis—and a higher risk of having to live more months with infection. Consistent with the expectation, the division of towns into villages and households decreased the risk of dying from smallpox and to some extent of pertussis, but having a larger number of villages slightly increased, if anything, the risk of dying from measles. Dividing towns into more households increased the number of months with an infection, showing the expected slower spread of epidemics, but only when population sizes were small. These results indicate that the effect of population structure is pathogen-specific and population size-dependent, hence complicating generalizations concerning links between population structure and disease or parasite risk. Here, we discuss the implications of our results for understanding the costs and benefits of group living and for the public health management of epidemics in human meta-populations.

We found that a larger population size increased both the risk of dying from infection and also the number of months that at least one person succumbed to a particular disease.

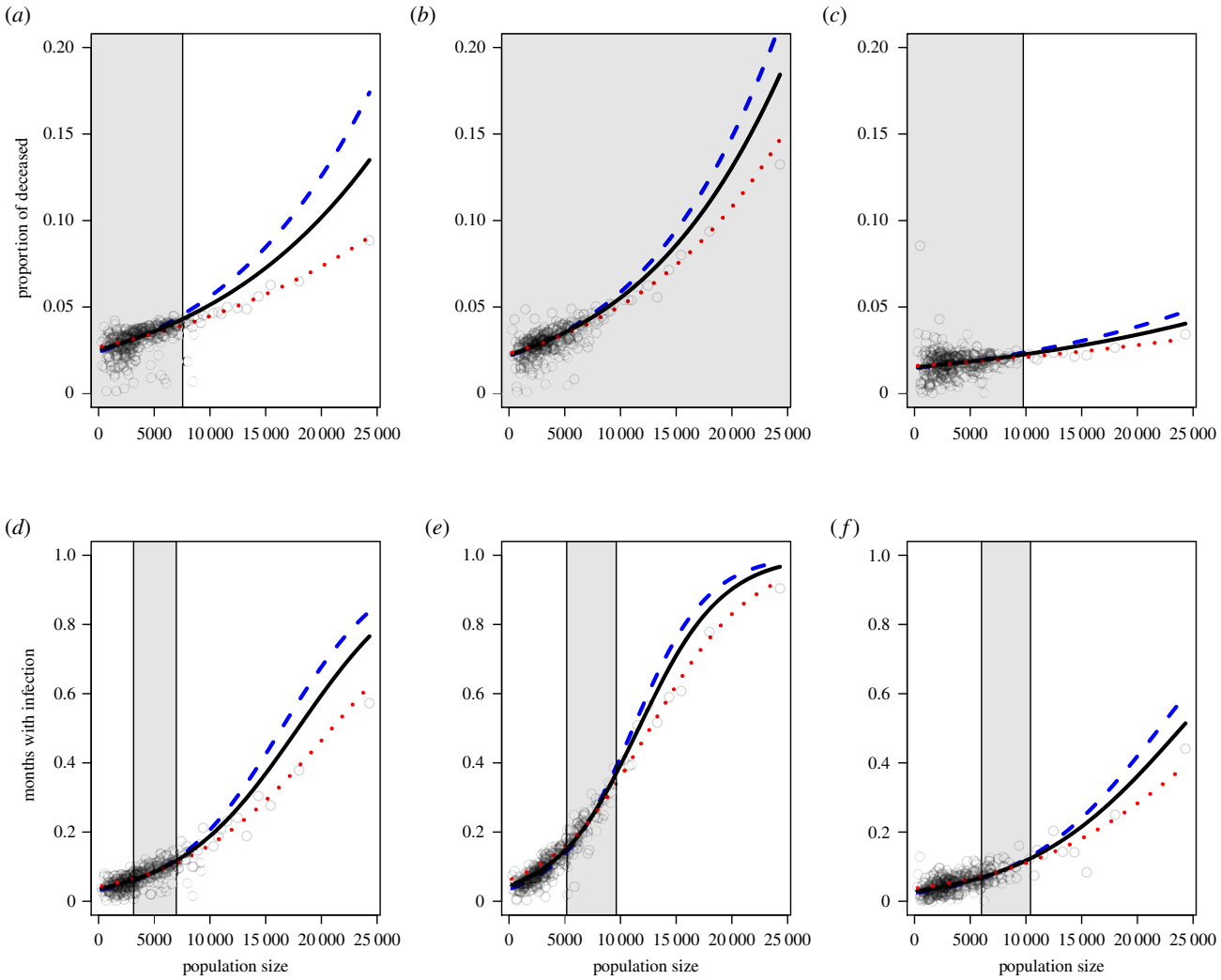

Figure 2. Larger town population size increased the risk of death by (*a*) smallpox, (*b*) pertussis, and (*c*) measles, and increased the risk of having months with infection (*d–f*) during the years 1800–1850 in Finland. Best fitting models included interaction with population size and number of households as depicted in the figure by lines, where black solid lines correspond to the effect of population size at the median (516) household number, and red dotted lines and blue dashed lines correspond to 75% (825) and 25% (310) quantiles of household numbers, respectively. Grey rectangles indicate the range of population sizes where the slope of household number on risk of the infectious death is non-significant ($p > 0.05$). This region was resolved following the Johnson–Neyman procedure. (Online version in colour.)

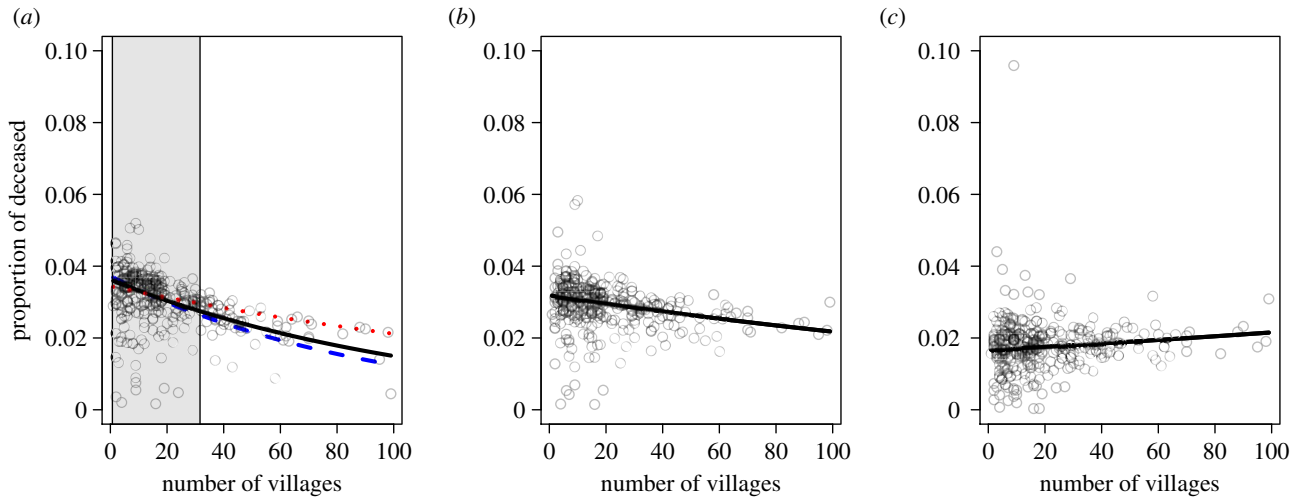

Figure 3. Town structuring into villages decreased the risk of dying of smallpox (*a*) but had no effect on the risk of death by pertussis (*b*) or measles (*c*) between 1800 and 1850 in Finland. The best model for the smallpox risk of death included the interaction with land area and the number of villages with three lines indicating different land areas: black solid line corresponds to effect of villages at the median (30 116.2 ha) land area, and red dotted line and blue dashed line correspond to 75% (56 127.2 ha) and 25% (16 890.4 ha) land area, respectively. Grey rectangles indicate the range of village numbers where the effect of area on the risk is non-significant ($p > 0.05$). This region was resolved following the Johnson–Neyman procedure. (Online version in colour.)

Table 2. The estimated risk of months with infection (at least one case of death) increased with town population size for all three infectious diseases and for measles also increased with a larger number of villages and households between 1800 and 1850 in Finland (below Arctic Circle, 66° 33′). (Estimates correspond to *z*-standardized values. Italics denote statistically significant result $p < 0.05$.)

		risk of months with infection			
		estimate	s.e.	*z*-value	Pr(>|*z*|)
smallpox	intercept	−2.5825	0.0338	76.4119	*<0.001*
	population size	0.4921	0.0471	10.4429	*<0.001*
	number of villages	0.0130	0.0368	0.3537	0.7236
	number of households	0.0481	0.0432	1.1150	0.2648
	area	0.0544	0.0421	1.2920	0.1964
	population size × number of households	0.1173	0.0594	1.9737	*0.0484*
	number of villages × area	−0.1313	0.0239	−5.5026	*<0.001*
pertussis	intercept	−2.0097	0.0426	47.1730	*<0.001*
	population size	0.7108	0.0650	10.9407	*<0.001*
	number of villages	−0.0200	0.0493	−0.4058	0.6849
	number of households	0.1823	0.0602	3.0299	*0.0024*
	area	−0.0737	0.0483	−1.5261	0.1270
	population size × number of households	−0.1863	0.0321	−5.8117	*<0.001*
measles	intercept	−2.8065	0.0287	97.9389	*<0.001*
	population size	0.3839	0.0397	9.6684	*<0.001*
	number of villages	0.1764	0.0302	5.8442	*<0.001*
	number of households	0.1480	0.0363	4.0807	*<0.001*
	area	−0.0244	0.0272	−0.8961	0.3702
	population size × number of households	−0.1103	0.0199	−5.5477	*<0.001*

These results directly demonstrate that living in large groups is costly in terms of increased parasite risks and are in concordance with many epidemiological studies in humans [4–6]. However, this effect has not been so clear in non-human species, at least in those species with generally small population sizes [2,3]. The effect of large population size on epidemics can be mediated via a higher density and contact network of individuals. Indeed, in our study, for smallpox, the effects of population structuring depended on town area or population size (table 1), which indicates that the epidemiological consequences of structuring are population size- and/or density-dependent.

Finnish towns varied in the number of villages they encompassed, which strongly affected the risk of dying of infectious diseases. In contrast with the rather straightforward prediction that group or population structuring decreases the disease or parasite risk [2], we found that different diseases differed in their relationship to the population structure. While the risk of dying of smallpox was lower in village-rich towns, these effects were not statistically significant for measles and pertussis. For pertussis, the effect was in the same direction as for smallpox, but tentative ($p = 0.08$; table 1), and for measles, the direction was opposite, if anything ($p > 0.1$; table 1) These results show that general predictions between population structure and disease or parasite load might be too simplistic, as different diseases can have varying relationships with population structure. These differences can arise if the incidence of some diseases is more driven by transmission and for other diseases by local extinctions and introduction, both of which are affected by population structure (e.g. [32,33]). For example, when a disease, such as smallpox has a broad range of susceptible age groups or a long infectious period [34,35], the disease could be maintained rather easily. In such a case, epidemics would be more controlled by the transmission between villages, which could lead to smaller epidemics in village-rich towns.

An influential characteristic for measles in structured populations could be its propensity for local extinctions i.e. fade-outs, an effect which is exacerbated in smaller towns [36]. Measles has a high R_0, infects a narrow subset of the population (i.e. children), is contagious over a relatively short period of time and causes strong immunity [37,38]. These biological characteristics can increase the chances of measles' local fadeouts and complicate generalizations. For example, Sah *et al.* [39] found that susceptible-infected-recovered models incorporating realistic animal network structures predicted that more slowly transmissible diseases are more prone to lead to smaller disease outbreaks, as some parts of the network are left completely uninfected. However, the spread of highly transmissible diseases is slowed down in structured networks. Therefore, populations structured with lots of villages or households should slow down the epidemic fadeout and allow measles incidences and casualties to become more persistent [40]. This was confirmed to be the case as we found out that measles had longer persistence in village-rich towns (larger number of months with infection, table 2). By contrast smallpox persistence decreased in village-rich towns, hence, being in concordance with the idea

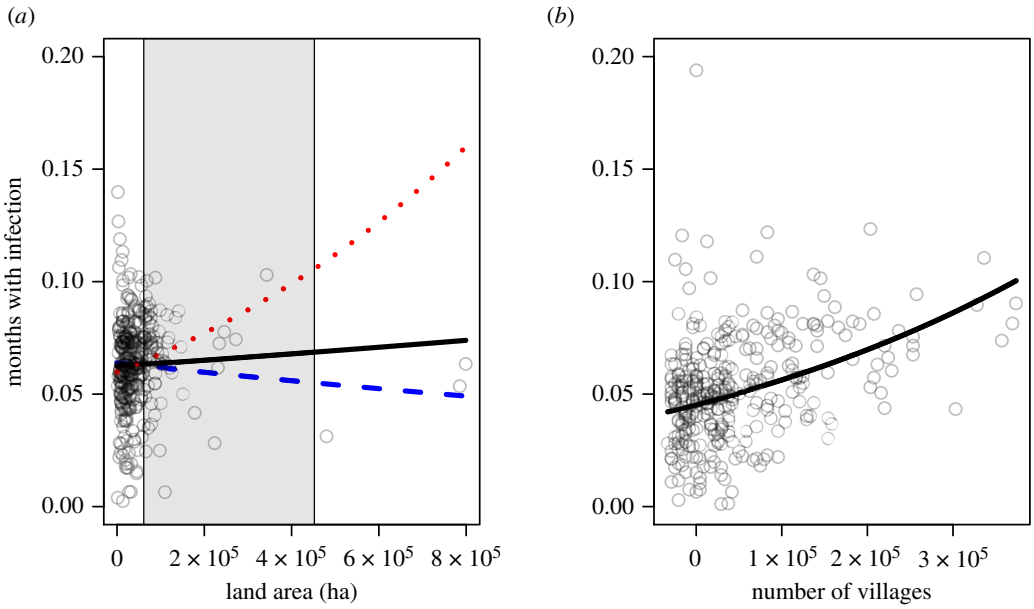

Figure 4. For smallpox, the risk of having months with infection decreased with number of villages if land area of the town was small (*a*), whereas in measles, a high number of villages was found to increase the number of months in infection (*b*) between 1800 and 1850 in Finland below the Arctic Circle (66° 33′). In (*a*), the red dotted line and blue dashed line corresponds to 75% (29) and 25% (8) quantiles for number of villages, respectively. The grey area indicates range of land area where village number has no effect on risk of having a month with infection. (Online version in colour.)

that structuring prevents transmission between the villages. However, this result held only in parishes with rather a small land area (figure 4*b*).

In addition to villages, households produce a second level of structuring in the towns and villages. Households in our study population often consisted of small-scale farms inhabited by the grandparents, their eldest son with his family and permanent or seasonal labourers including other unmarried siblings. However, owing to differences in farming practices, inheritance and culture across Finland, how the populations and villages were distributed into separate households varied widely across the country [41]. For all three infections, we found that structuring into more households increased the risk of having months with infection in a town. However, for all infections, this association only held in towns with small population sizes (tables 1 and 2 and figure 2*d–f*). It thus seems that structuring into many separate households can act by slowing down the spread of infections and hence maintaining infections in the towns for longer time periods. However, in towns with large population sizes, the high household numbers decreased the disease risk of having a month with an infection. Although this finding is in concordance with the idea that structuring should decrease the burden of infections, it is noteworthy that this result is significantly present only in the 10–20 largest towns (figure 2). This result could also indicate an effect of some other unfitted correlates that are present only in the largest towns and should be interpreted with caution.

We have here concentrated on population structure-related variables that could play a role in the spread of infections. However, several other factors such as urbanization, weather or socio-economic status can affect the risk of infections and partially explain some of the results and interactions (see above). Expansion of the modelling to more accurately account the village positioning, size differences of villages within the town [42] and simultaneous occurrence of several infections competing for the same susceptible hosts [27] are all effects that could improve the understanding of epidemics, and their determinants, in pre-healthcare Finland. More generally, the interaction between infectious diseases 'competing' for susceptible hosts [27] could provide an alternative explanation as to why the effect of population structuring might be the opposite between infections.

In summary, we found that the epidemics of infectious diseases are controlled both by population size and structure. Increasing population size leads to more devastating epidemics by increasing the risk of succumbing to disease and the months with infection. In addition, large numbers of households slowed down the epidemics in less populated towns. However, the effect of the town structuring into villages was found to have either a tempering or non-existing role on epidemics depending on the infection and on the population size. This shows that there might not be generalizations to be made concerning how overall disease or parasite loads explain the costs of living for species with either a social or solitary lifestyle, and pinpoint the need for taking into account the biology of the disease when predicting disease spread in natural, structured, populations.

Data accessibility. Data for church book records can be accessed via the Genealogical Society of Finland: https://hiski.genealogia.fi/hiski/oc13w?en. Other data used in this study are available from sources cited in the text.

Authors' contributions. T.K., V.L. and M.B. designed the work. T.K. compiled and analysed the dataset and wrote the first draft of the manuscript; T.H. provided key data and participated in writing the manuscript. All authors participated in finalizing the manuscript.

Competing interests. We declare we have no competing interests.

Funding. This research was funded by Academy of Finland (T.K.: 278751; V.L.: 292368), ERC (V.L.: ERC-2014-CoG 648766), Finnish Cultural Foundation (T.K.), Ella and Georg Ehrnrooth Foundation (M.B.) and Kone Foundation (projects SumuraSyyni and AikaSyyni for Outi Vesakoski) (T.H.).

Acknowledgements. We are grateful to the many volunteers of the Genealogical Society of Finland who collected the dataset from church records and to Miikka Voutilainen, Outi Vesakoski, Merja Elo and Andrew Park for discussions and comments on the manuscript.

References

1. Markel H, Lipman HB, Navarro JA, Sloan A, Michalsen JR, Stern AM, Cetron MS. 2007 Nonpharmaceutical interventions implemented by US cities during the 1918–1919 influenza pandemic. *JAMA* **298**, 644–654. (doi:10.1001/jama.298.6.644)
2. Nunn CL, Jordan F, McCabe CM, Verdolin JL, Fewell JH. 2015 Infectious disease and group size: more than just a numbers game. *Phil. Trans. R. Soc. B* **370**, 20140111. (doi:10.1098/rstb.2014.0111)
3. Rifkin JL, Nunn CL, Garamszegi LZ. 2012 Do animals living in larger groups experience greater parasitism? A meta-analysis. *Am. Nat.* **180**, 70–82. (doi:10.1086/666081)
4. Grenfell BT, Dobson AP. 1995 *Ecology of infectious diseases in natural populations*. Cambridge, UK: Cambridge University Press.
5. McCallum H, Barlow N, Hone J. 2001 How should pathogen transmission be modelled? *Trends Ecol. Evol.* **16**, 295–300. (doi:10.1016/S0169-5347(01)02144-9)
6. Bjornstad ON, Finkenstadt BF, Grenfell BT. 2002 Dynamics of measles epidemics: estimating scaling of transmission rates using a time series SIR model. *Ecol. Monogr.* **72**, 169. (doi:10.2307/3100023)
7. Ball F, Britton T, House T, Isham V, Mollison D, Pellis L, Tomba GS. 2015 Seven challenges for metapopulation models of epidemics, including households models. *Epidemics* **10**, 63–67. (doi:10.1016/j.epidem.2014.08.001)
8. Mossong J *et al.* 2008 Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* **5**, e74. (doi:10.1371/journal.pmed.0050074)
9. Salathé M, Jones JH. 2010 Dynamics and control of diseases in networks with community structure. *PLoS Comput. Biol.* **6**, e1000736. (doi:10.1371/journal.pcbi.1000736)
10. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. 2015 Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925–979. (doi:10.1103/RevModPhys.87.925)
11. Griffin RH, Nunn CL. 2012 Community structure and the spread of infectious disease in primate social networks. *Evol. Ecol.* **26**, 779–800. (doi:10.1007/s10682-011-9526-2)
12. Watwe MG, Jog MM. 1997 Epidemic diseases and host clustering: an optimum cluster size ensures maximum survival. *J. Theor. Biol.* **184**, 165–169. (doi:10.1006/jtbi.1996.0267)
13. North AR, Goodfray HCJ. 2017 The dynamics of disease in a metapopulation. The role of dispersal range. *J. Theor. Biol.* **418**, 57–65. (doi:10.1016/j.jtbi.2017.01.037)
14. Hayward AD, Rigby FL, Lummaa V. 2016 Early-life disease exposure and associations with adult survival, cause of death, and reproductive success in preindustrial humans. *Proc. Natl Acad. Sci. USA* **113**, 8951–8956. (doi:10.1073/pnas.1519820113)
15. Saarivirta T, Consoli D, Dhondt P. 2012 The evolution of the Finnish health-care system early 19th century and onwards. *Int. J. Bus. Soc. Sci.* **3**, 243–257. (doi:10.13039/501100002341)
16. Anderson RM, May RM. 1991 *Infectious diseases of humans*. Oxford, UK: Oxford University Press.
17. Rohani P, Keeling MJ, Grenfell BT. 2002 The interplay between determinism and stochasticity in childhood diseases. *Am. Nat.* **159**, 469–481. (doi:10.1086/339467)
18. Metcalf CJE, Bjornstad ON, Grenfell BT, Andreasen V. 2009 Seasonality and comparative dynamics of six childhood infections in pre-vaccination Copenhagen. *Proc. R. Soc. B* **276**, 4111–4118. (doi:10.1098/rspb.2009.1058)
19. Bolund E, Hayward A, Pettay JE, Lummaa V. 2015 Effects of the demographic transition on the genetic variances and covariances of human life-history traits. *Evolution* **69**, 747–755. (doi:10.1111/evo.12598)
20. Scranton K, Lummaa V, Stearns SC. 2016 The importance of the timescale of the fitness metric for estimates of selection on phenotypic traits during a period of demographic change. *Ecol. Lett.* **19**, 854–861. (doi:10.1111/ele.12619)
21. Ikonen T. 1948 *Suomen Lappalaiset vuoteen 1945*. Porvoo, Finland: Werner Söderström Oy.
22. Vuorinen HS. 1999 Suomalainen tautinimistö ennen bakteriologista vallankumousta. *Hippokrates* **16**, 33–61.
23. Pitkänen KJ, Mielke JH, Jorde LB. 1989 Smallpox and its eradication in Finland: implications for disease control. *Popul. Stud. (Camb.)* **43**, 95–111. (doi:10.1080/0032472031000143866)
24. 1882 *Suomenmaan tilastollinen vuosikirja*. Helsinki, Finland: Tilastollinen Toimisto.
25. 1916 *Maataloustiedustelu Suomessa vuonna 1910*. Helsinki, Finland: Maataloushallitus.
26. Bates D, Mächler M, Bolker B, Walker S. 2015 Fitting linear mixed-effects models using lme4. *J. Stat. Soft.* **67**, 1–48. (doi:10.18637/jss.v067.i01)
27. Rohani P, Green CJ, Mantilla-Beniers NB, Grenfell BT. 2003 Ecological interference between fatal diseases. *Nature* **422**, 885–888. (doi:10.1038/nature01542)
28. Borcard D, Legendre P. 2002 All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol. Model.* **153**, 51–68. (doi:10.1016/s0304-3800(01)00501-4)
29. Peres-Neto PR, Legendre P. 2009 Estimating and controlling for spatial structure in the study of ecological communities. *Glob. Ecol. Biogeogr.* **19**, 174–184. (doi:10.1111/j.1466-8238.2009.00506.x)
30. Voutilainen A, Tolppanen A-M, Vehviläinen-Julkunen K, Sherwood PR. 2014 From spatial ecology to spatial epidemiology: modeling spatial distributions of different cancer types with principal coordinates of neighbor matrices. *Emerg. Themes Epidemiol.* **11**, 11. (doi:10.1186/1742-7622-11-11)
31. Bauman D, Drouet T, Dray S, Vleminckx J. 2018 Disentangling good from bad practices in the selection of spatial or phylogenetic eigenvectors. *Ecography* **41**, 1638–1649. (doi:10.1111/ecog.03380)
32. Watve MG, Jog MM. 1997 Epidemic diseases and host clustering: an optimum cluster size ensures maximum survival. *J. Theor. Biol.* **184**, 165–169. (doi:10.1006/jtbi.1996.0267)
33. Caillaud D, Craft ME, Meyers LA. 2013 Epidemiological effects of group size variation in social species. *J. R. Soc. Interface* **10**, 20130206. (doi:10.1098/rsif.2013.0206)
34. Eichner M, Dietz K. 2003 Transmission potential of smallpox: estimates based on detailed data from an outbreak. *Am. J. Epidemiol.* **158**, 110–117. (doi:10.1093/aje/kwg103)
35. Milton DK. 2012 What was the primary mode of smallpox transmission? Implications for biodefense. *Front. Cell Infect. Microbiol.* **2**, 150. (doi:10.3389/fcimb.2012.00150)
36. Bartlett MS. 1957 Measles periodicity and community size. *J. R. Stat. Soc. Ser. A (General)* **120**, 48–70. (doi:10.2307/2342553?refreqid=search-gateway:778c544761643f5aab037f45d11848a0)
37. Perry RT, Halsey NA. 2004 The clinical significance of measles: a review. *J. Infect. Dis.* **189**(Suppl. 1), S4–16. (doi:10.1086/377712)
38. Coughlin M, Beck A, Bankamp B, Rota P. 2017 Perspective on global measles epidemiology and control and the role of novel vaccination strategies. *Viruses* **9**, 11–17. (doi:10.3390/v9010011)
39. Sah P, Leu ST, Cross PC, Hudson PJ, Bansal S. 2017 Unraveling the disease consequences and mechanisms of modular structure in animal social networks. *Proc. Natl Acad. Sci. USA* **114**, 4165–4170. (doi:10.2307/26480424?refreqid=search-gateway:426e7c08c77e3c04562cf36e76e97f30)
40. Keeling MJ, Grenfell BT. 1997 Disease extinction and community size: modeling the persistence of measles. *Science* **275**, 65–67. (doi:10.1126/science.275.5296.65)
41. Voutilainen M. 2017 Marriage and household structure in rural pre-famine Finland, 1845–65. In *The enormous failure of nature famine in nineteenth century Europe* (ed. A Newby), pp. 67–82. Helsinki Collegium for Advanced Studies.
42. Xia Y, Bjornstad ON, Grenfell BT. 2004 Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *Am. Nat.* **164**, 267–281. (doi:10.1086/422341)