Journal Pre-proof

The spatial distribution of pertussis, but not measles or smallpox, in pre-industrial Finland matches dialects.

Aïda Nitsch, Virpi Lummaa, Tarmo Ketola, Terhi Honkola, Outi Vesakoski, Michael Briga

PII: S2589-0042(25)00791-6

DOI: https://doi.org/10.1016/j.isci.2025.112530

Reference: ISCI 112530

To appear in: *ISCIENCE*

Received Date: 2 December 2024

Revised Date: 28 February 2025

Accepted Date: 23 April 2025

Please cite this article as: Nitsch, A., Lummaa, V., Ketola, T., Honkola, T., Vesakoski, O., Briga, M., The spatial distribution of pertussis, but not measles or smallpox, in pre-industrial Finland matches dialects., *ISCIENCE* (2025), doi: https://doi.org/10.1016/j.isci.2025.112530.

Question

Are dialect groups a **relevant proxy** to **regionally** structure epidemics from infectious diseases?

Hypothesis

ADMINISTRATIVE REGIONS

- Commonly used
- Not ecologically relevant

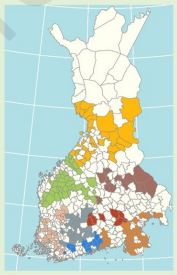
DIALECT GROUPS

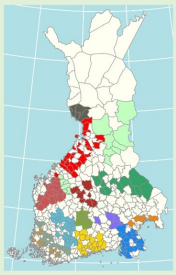
- Never tested
- Ecologically relevant

Methods

- Church records
- Finland
- 1800-1850
- 215 parishes
- 3 diseases
- Compare fit of models with different regional proxies (none, regions, dialects)

8 administrative regions





14 dialects

Results

Best regional proxy?

Smallpox

Pertussis

Measles

TITLE PAGE

Title: The spatial distribution of pertussis, but not measles or smallpox, in pre-industrial Finland matches dialects.

Author list:

Aïda Nitsch (ORCID: 0000-0002-1312-9677)[1]*, Virpi Lummaa[1], Tarmo Ketola[2], Terhi Honkola[3], Outi Vesakoski[4], Michael Briga (ORCID: 0000-0003-3160-0407)[1,5,6,7]

Author affiliations:

[1]Department of Biology, University of Turku, Vesilinnantie 5, 20014 Turku, Finland

[2]Department of Forest Sciences, University of Helsinki, P.O. Box 24, 00014 University of Helsinki, Finland

[3]Department of Finnish, Finno-Ugrian and Scandinavian Studies, University of Helsinki, P.O. Box 24, 00014 University of Helsinki, Finland

[4]School of Languages and Translation Studies, University of Turku, Hämeenkatu 1 A, 20014 Turku, Finland

[5]PandemiX Center of Excellence, Roskilde University, Universitetsvej 1, Postbox 260, 4000 Roskilde, Denmark

[6]Charité Centre for Global Health, Charité Universitaetsmedizin, Charitéplatz 1 10117, Berlin, Germany

[7]Turku Collegium for Science, Medicine and Technology, University of Turku, Turku, Finland

Email address of corresponding author:

Email: ainitsch@proton.me

* Lead Contact

SUMMARY

Infections spreading from host-to-host are a burden of social lifestyle mostly documented at the local scale (within groups). The influence of social structure at a broader scale (e.g., between groups or regions) on infectious disease dynamics is less understood partly due to the difficulty to identify the relevant social groups at this scale. Dialect groups encompass long-held human contacts and could indicate social groups relevant to infections. Using nationwide individual-level mortality records from pre-industrial Finland (1800–1850), we investigated which social grouping best predicted spatial variation in smallpox, pertussis and measles mortality by comparing models with no regional information, administrative regions and dialect groups. Dialect groups explained spatial variation of pertussis, adminsitrative regions for smallpox, while measles showed no broader scale spatial variation. These results highlight the complex spatial structuring of infectious diseases and stress the need for studies to identify the relevant social structure.

INTRODUCTION

Infections spreading from host to host are one of the main costs of sociality[1–5]. The size and structure of the social organisation (e.g., the existence of social hierarchy), as well as more detailed metrics of the social network (e.g., number of social partners, frequency of interactions), have been linked to disease incidence and dynamics both theoretically and empirically[6–11]. The importance of local social structure in disease distribution has also been documented in humans: the dynamics of directly transmitted pathogens were for instance found to be affected by the existence of socially connected groups in municipalities, such as schools[12–14], and households and villages[15].

However, the impact of social organisation on disease dynamics extends beyond this local scale and can influence spatial patterns of diseases at a broader regional scale[3,16,17]. For instance, social species show variation in anti-parasite strategies[18,19] and in immune defences[20] between populations. Similarly to non-human species, the epidemiological consequences of human social structure at the regional scale remain not well understood. One limitation stems from the difficulty of identifying the relevant regional population structure as it requires detailed information on regional variation between groups (e.g., on group characteristics, contacts between groups, people's movements)[8,21–25]. Due to its easy availability, infectious disease studies in humans often use administrative units such as counties, regions, states or countries to control for regional variation[e.g., 26,27] even if administrative borders are, to a large extent, the outcomes of political or historical events[28]. However, several studies show that biological processes do not align with regional borders, for instance in the case of genetic clustering of the British population[29], the distribution of plague patterns and its consequences on land use in middle ages[30] or regarding the ecological niche of the lumpy skin disease virus in the middle-east countries[31]. Therefore, although the consequences of using administrative regions to model epidemiological processes and its potential mismatch with population characteristics have not been assessed, it is likely that administrative regions may only

partially represent shared characteristics of the population, and thereby not well predict the geographic scale of outbreaks or the interventions required to contain them[32–34].

Dialects, which are partly formed as a consequence of long-held social contacts, could be an informative way to identify regional variation in human populations. Here, we define dialects as mutually intelligible, regional variants of a language that arose from spatial segregation of speaker populations followed by linguistic divergence[35]. The spatial segregation of dialects may be a result of various social processes and geographical distance[e.g., 35,36] and may reflect regional-level differences in social contacts[37], and in cultural or in environmental characteristics[38]. Therefore, regional differences in language (i.e., dialects) could be a good variable to characterize regional variation in infectious disease dynamics. However, studies investigating this hypothesis are currently lacking.

In this study, we tested whether regional variation in the Finnish language can explain regional variation in the mortality from three childhood infections (smallpox, pertussis, and measles) in pre-industrial Finland. These three childhood infections are directly transmitted from person to person (through direct contact or airborne droplets)[39] and were leading causes of childhood mortality at that time[15]. This population is suited for this study for several reasons: (1) Finland has excellent church records, which document individual-level deaths and their causes across Finland since 1749, thereby enabling the comparison of different types of geographical clustering of mortality risk across the whole population; (2) mortality from smallpox, pertussis and measles are well-documented in this population[40,40,41] and studies have shown that geographical distance and demographic factors explain spatial variation in childhood infections at a local scale[15]; (3) linguistic variation of Finnish has been extensively studied reviewed[42] and quantitative dialectometric studies have documented robust dialect groups[38,43], thereby allowing the use of these groups in this study.

We investigated whether dialect groups could explain regional variation in mortality by comparing the fit of models using different variables to cluster social groups. Specifically, we compared the fit of three models: (1) with dialect groups, (2) with administrative regions, (3) without

regional clustering. We hypothesized that dialect groups capture regional clustering of infectious diseases mortality better than administrative regions or when no information on regional clustering is provided for all three diseases[28].

RESULTS

Overall, our results show that for pertussis the models including dialect groups fitted the data better than the model containing only the municipality or the model containing the regional administrative borders. For smallpox, the best model was the model containing the administrative regions, whereas the control model, i.e., without regional clustering, was the best model for measles. Our results were not confounded by variation in municipalities' surface area, population size, number of villages, or number of households, which were corrected for in all of our models. Neither were they linked to spatial autocorrelation, which was also controlled for in all of our models.

Smallpox

On average, 3.6% (±0.1SE) of the total deaths in each municipality were due to smallpox (figures 2a, 3a, table S1). As we found no spatial autocorrelation (Moran's I<0.05), the importance of spatial distance between municipalities in smallpox mortality was low and hence we did not include any PCNM variable in the model (see Methods for details). The results of the model selection indicated that administrative regions clustered smallpox mortality better than the other models: the model including administrative regions was a better fit than the control model containing only municipality (ΔAIC=14.95), and the model containing dialect groups (ΔAIC=2.46, tables 1, S2).

Pertussis

On average, 3.9% (±0.2SE) of deaths were due to pertussis (figures 2b, 3b, table S1). Controlling for the spatial autocorrelation required the fit of four PCNM vectors thereby indicating a strong importance of distance in mortality patterns. Our results show strong support for dialect groups to cluster pertussis mortality at the regional scale. Indeed, the model including dialect groups fitted the data better than the control model including only municipality (ΔAIC=37.32) and the model including the administrative regions (ΔAIC=28.74, tables 2, S3).

Measles

On average, 2.1% (±0.1SE) of total deaths occurring in a municipality were due to measles (figure 2c, 3c, table S1). Models required the fit of two PCNM variables to account for the autocorrelation due to the distance between municipalities. Results of the model selection indicate that among the set of models fitted, the control model containing only municipality fitted the data slightly better than the model including administrative regions (ΔAIC=1.89) and the model containing dialect groups (ΔAIC=2.18, tables 3, S4), thereby indicating that all three models fitted the data almost equally well.

DISCUSSION

Social interactions have a strong impact on epidemiological dynamics, especially for pathogens directly transmitted between hosts of the same species. However, the interplay between social dynamics and epidemiological patterns has often been documented only at a local scale (within groups) but less at the regional one (between groups). The reason for the lack of studies on the regional scale is, at least partly, due to the difficulty of identifying the regional social structure. Here we tested in a pre-industrial human population from Finland whether we could grasp the regional differences in the mortality from childhood infectious diseases using dialect groups. We selected three infectious diseases which are directly transmitted from person to person (through direct

contact or airborne droplets) and were leading causes of mortality in children under 15 years old[15,40]. We expected a regional structure to exist for these epidemics as previous studies have documented variation in the spatial distribution of mortality for each disease at the local[e.g. 12,15], regional[44–48] and even at the country scale[49].

Contrary to our predictions, our results do not support the hypothesis that regional social clustering may generally be grasped by linguistic variation but rather highlight disease-specific patterns. Our study adds to previous studies by documenting the regional spatial distribution of childhood infectious diseases and its variation between diseases. Acknowledging such variation emphasizes the need to identify the relevant spatial scale when studying epidemics and can be helpful when planning public health interventions[32].

Overall, our ability to compare our results to other studies on regional clustering is limited. Although spatial structuring of epidemics has been investigated previously, the scale of spatial clustering, the methodology, the population, and the factors included in the models varied between studies[26,50,51], thereby preventing the identification of the most relevant scale of regional structuring of epidemics. For instance, although Fridlizius and Ohlsson[46] compared the epidemics of smallpox, pertussis and measles in a similar period in historical Sweden (1750-1800) and reported variation between regions, they did not investigate which factors could mediate it. Similarly, Ketola et al.[15] found that several local spatial factors (e.g., the number of households or villages in the municipality) were linked to the spatial variation in childhood mortality from smallpox, pertussis, and measles, but this study did not study the regional structuring of mortality patterns. More importantly, studies other than ours investigating which variables best explain the the spatial variation of infectious diseases are lacking, thereby precluding the comparison and generalization of our results to other populations.

As we study mortality from different pathogens in the same population, variation in regional spatial structuring is likely to be linked to the characteristics of the pathogen. These differences

between pathogens can result from different processes, in particular variation in transmission or outcome of an infection[52]. The outcomes of infection encompass the risk of individuals dying from an infection and can be measured by the case fatality rate (the proportion of cases of a specified condition that are fatal within a specified time)[53]. Comparing variation in transmission and outcomes of infection across regions would therefore provide valuable insights to decipher the contributions of these processes and potential mediators to our results. However, this approach is beyond the scope of this study, and we will here only discuss putative mediators of our results for each disease.

Measles mortality was better explained by the model including neither of the regional clustering variables and the difference of fit between models was small, thereby indicating that measles mortality was not strongly regionally structured. As measles is highly contagious (R_0=12-14)[54], a measles epidemic outbreak may contaminate the entire country, which will subsequently lead to limited variation in the relative total death toll of measles between regions. However, spatial variation in the spread and impact of measles may still exist, but would require a study of the spatial spread, such as traveling waves, of the disease to be detected [see 55 for an example].

Surprisingly, the model with administrative regions was a better fit to smallpox data than those including dialect groups. However, the difference in fit between the model with administrative regions and dialect groups was quite small, which indicates more broadly a regional structuring of smallpox epidemics. One possible hypothesis is that the smallpox vaccination campaign in Finland may have been linked to the administrative structure of Finland, which could have let to smallpox mortality patterns following Finnish regional divisions. Smallpox vaccinations began in Finland in 1802, and by 1825, the country was divided into vaccination districts[56,57]. Although detailed comparison of vaccination coverage between regions during our study period is currently lacking[56,57], a previous study examining a subset of Finnish parishes (1837-1899) highlighted variation in vaccination coverage between parishes[41].

Our hypothesis that linguistic variation is a relevant proxy to structure variation in mortality was supported for pertussis. Studies focusing only on transmission (e.g. in measles[27], smallpox[45] or influenza[58]) successfully managed to explain the spatio-temporal dynamics of infectious diseases, thereby suggesting that variation in transmission patterns may mediate this result. This interpretation is also in line with studies on the same period and area: indeed Ketola et al.[15] found that the factors associated with disease spread (e.g., population density, and sub structuring of the population to villages and households) explained the local spatial distribution of mortality from infectious diseases. This could indicate that linguistic areas may represent a contact network through which infections could spread. Furthermore, a key difference between pertussis and the other two infections, is that pertussis is to a large extent transmitted by teens and adults because of waning immunity after natural infection[59,60]. Conversely, measles and smallpox mostly occur in children under 5 years and confer long-lasting immunity[39]. As older individuals are more likely to move across Finland, their movements may lead to regional scale spatial structuring of epidemics from pertussis whereas local spatial structure may be more important for the other diseases studied. In addition to variation in transmission, our results could be linked to factors specific to each dialect area and thereby being driven by differences in mortality rates from infections between dialect regions. Indeed, some group characteristics (e.g., variation in environmental conditions, food resources, population characteristics such as density, genetics or cultural dimensions), may shape the variation in mortality from pertussis[61,62]. For instance, a study focusing on Finnish dialect groups and other cultural (e.g., presence of slash-and-burn agriculture, chimneyless huts) and environmental features (e.g., percentage of land area covered by lakes or clay soil) identified a correlation between differences in dialects and differences in cultural traits[38]. However, studies investigating the correlation between dialect areas and ecological or cultural variables are currently lacking and remain to be explored.

Overall, the strong variation of regional structuring between diseases highlights the need for comparative studies to deepen our understanding of factors linked to spatial structuring. Therefore,

future studies could build on this work to closely investigate the regional structuring of similar diseases in other populations and investigate whether some specific characteristics of these areas may impact survival from infectious diseases.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Aïda Nitsch (ainitsch@proton.me).

Material availability

This study did not generate any new materials.

Data and Code Availability

- Data have been deposited at OSF and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- The code has been deposited at OSF and is publicly available as of the date of publication at: doi.org/10.17605/OSF.IO/C76DQ.
- Any other items reported in this manuscript are available from the lead contact upon request.

LIMITATIONS OF THE STUDY

The study dataset is limited to a specific period and location, which limits the generalisation of our findings to other populations. Furthermore, as we the study the total mortality over a period of 50 years, it is not possible to study the factors underpinnings our results.

ACKNOWLEDGMENTS

We thank the funding from NordForsk (104910), the NetResilience consortium funded by the Strategic Research Council at the Academy of Finland (345183), the Finnish Cultural Foundation, the Ella & Georg Ehrnrooth Foundation, the Turku Collegium for Science, Medicine and Technology, the Kone Foundation, the European Research Council (KinSocieties, ERC-2022-ADG, number 101098266), the Turku Institute for Advanced Studies and the Human Diversity consortium, Profi7 programme by the Research Council of Finland (grant 352727). We also thank the Genealogical Society of Finland for providing access to the data and to the volunteers who digitised the data from church records and two anonymous reviewers for their comments.

AUTHOR CONTRIBUTIONS

Conceptualization: M.B.; Resources: M.B., T.H., T.K., V.L., O.V.; Data curation: A.N., Statistical analyses: A.N.; Writing-original draft: A.N.; Writing-review and editing: M.B, T.H., T.K., V.L., A.N., O.V., Supervision: M.B., V.L. All authors were responsible for the final version of the manuscript.

DECLARATION OF INTEREST

The authors declare no competing interests.

FIGURES AND SCHEME LEGENDS

Figure 1. Maps of historical Finland divided into (a) eight administrative regions in 1831 and (b) fourteen dialect groups of the Finnish language, as identified by Honkola et al (2018) for municipalities included in the analyses (n=215). Municipalities in white were excluded from the analyses because they represent transitional dialect areas, Swedish-speaking or Saami-speaking areas or due to incomplete data coverage.

Figure 2. Proportion of total deaths owing to (a) smallpox; (b) pertussis and (c) measles in each municipality (n=215).

Figure 3. Averaged proportion of total deaths owing to (a) smallpox; (b) pertussis and (c) measles in each dialect group (n=14).

TABLES

Table 1. Summary of the best a priori models on the proportion of deaths in each municipality (n=215) due to smallpox including the total number of estimable parameters (K), the log-likelihood (LogLik), AIC differences relative to the minimum value in the model set (ΔAIC), the Akaike weight (w_i), and the coefficient of determination of the model (R^2). See Methods for details on each analysis and associate set of candidate models, and the Supplements for the model estimates of each variable (table S2).

Models	K	Log Lik	ΔAIC	w_i	R^2
Control + Administrative Region	7	-1185.2	0.00	0.77	0.0033
Control + Dialect Group	7	-1186.5	2.46	0.23	0.0028
Control	6	-1193.8	14.95	0.00	4.00E-04

Table 2. Summary of the best a priori models on the proportion of deaths in each municipality (n=215) due to pertussis including the total number of estimable parameters (K), the log-likelihood (LogLik), AIC differences relative to the minimum value in the model set (ΔAIC), the Akaike weight (w_i), and the coefficient of determination of the model (R^2). See Methods for details on each analysis and associate set of candidate models, and the Supplements for the model estimates of each variable (tables S3).

Models	K	Log Lik	ΔAIC	w_i	R^2
Control + Dialect Group	11	-1163.6	0	1	0.01
Control + Administrative Region	11	-1178	28.74	0	0.0058
Control	10	-1183.4	37.32	0	0.004

Table 3. Summary of the best a priori models on the proportion of deaths in each municipality (n=215) due to measles including the total number of estimable parameters (K), the log-likelihood (LogLik), AIC differences relative to the minimum value in the model set (ΔAIC), the Akaike weight (w_i), and the coefficient of determination of the model (R^2). See Methods for

280 details on each analysis and associate set of candidate models, and the Supplements for the model
281 estimates of each variable (tables S4).

Models	K	Log Lik	ΔAIC	w_i	R^2
Control	8	-1006.5	0	0.58	0.001
Control + Administrative Region	9	-1006.4	1.89	0.23	0.0011
Control + Dialect Group	9	-1006.5	2.18	0.20	1.00E-03

282

STAR METHODS

STUDY POPULATION

We combined two previously published datasets to investigate whether the spatial variation in mortality to infectious diseases at the regional scale could be explained by the dialect groups. The first dataset documents the mortality owing to smallpox, pertussis and measles across 19th century Finland[15] and the second presents the dialect groups of Finnish[38,43]. As these datasets have been described and analysed in detail elsewhere, we will only summarise them briefly here.

The mortality dataset was a subset from a large demographic dataset on historical Finnish populations used previously in Ketola et al.[15] and in Briga et al.[40,63]. Since 1749, each parish was obliged by law to document all births, deaths, causes of death, marriages and movements between parishes in the whole country[64,65]. The original parish records have been digitized by the Genealogical Society of Finland and are available at http://hiski.genealogia.fi/historia/indexe.htm. For each deceased, the church record included information on the date of death, the cause of death, the date of burial, the identity of the deceased and the parish of residence (hereafter referred to as *Municipality*).

In our study, we focused on three infectious diseases, smallpox, pertussis and measles, as their diagnosis is generally reliable due to distinctive symptoms[15,47]. However, the church records had a variety of ways how these diseases were registered in the records. Following Vuorinen[66], two authors (M.B. and T.K.) independently identified the causes of death (>50,000 different causes in the initial church records from the Genealogical Society of Finland) by combining typographical variants, abbreviations and synonyms of diseases in the different languages used (Finnish, German and Swedish) and obtained the same classification (data not shown). We limited the study period between 1800 and 1850 in order to maximise the number of municipalities that had records every year[15]. Only municipalities with at least one death documented in each year of the study were kept in the study dataset. Our study period is set before the spread of industrialism, urbanisation, the

transition to reduced birth and mortality rates. Vaccination against smallpox started in 1802 in Finland and access to modern health care started to take place in the late 19th century[67]. The study populations mostly depended on farming for their livelihood and were supplemented with fishing in the coastal areas. Regional variation across different dimensions has been documented, including in environmental and ecological conditions[38,68,69], genetic structure[70], household structure[71,72], cultural practices[73] or economic conditions[74,75]. Overall, the standard of living was low with both famines and diseases common[48]. Population reconstructions estimated an increase in the Finnish population from 1 million to 1.5 million over the study period (1800-1850)[76].

During this period, municipalities located above the Arctic Circle were mostly inhabited by Saami, who were mostly depending on reindeer herding, fishing and hunting for their livelihood. As these populations spoke Saami languages[77], these municipalities were excluded from the study[78]. When different parishes were part of a larger town, they were grouped to represent the same administrative unit (i.e., Turku, Viipuri, Helsinki, Heinola, Jyväskylä, Kuopio and Sortavala). Additionally, we used information on the demographic conditions of each municipality (population size, number of villages, number of households and municipality surface area) to control for their documented effect on disease dynamics[15].

The dialect group data used in this study is from Honkola et al.[38]. The initial linguistic data were extracted from the Dialect Atlas of Finnish, which represents linguistic variation of the Finnish language in each Finnish-speaking municipality in the beginning of the 20th century[79]. The digitised atlas is archived in the Fairdata-service[80] and a modified version is available in Santaharju et al.[81]. Using an analytical framework from population genetics, Honkola et al.[38] clustered the linguistic variation between 471 municipalities into 14 dialect groups. Each municipality obtained a set of membership coefficients (a value of Inferred Cluster (IC) ranging from 0 to 1) to each dialect group which can be interpreted as a percentage of membership to each dialect group[38,43]. Following Honkola et al.[38], we only included municipalities where the IC-value was above 0.75 (a municipality

dialect with an IC value of at least 0.75 to one specific dialect) in order to have reliably distinct dialect groups. Transitional areas between dialect groups were thus excluded from the dataset. Additionally, as the Dialect Atlas focuses on Finnish dialect data, the Swedish speaking municipalities were excluded from our study (western and southern coast of Finland). Honkola et al.[38] also contained data on historical administrative borders (national, bishopric and provincial borders) from ca. year 1250 to 1895. The provincial borders usually divided Finland into four or more administrative areas, whereas there were four or fewer bishopric areas. To capture the regional level of the administrative areas, we decided to use provincial borders from 1831 in our study. It divided Finland into eight provinces (referred to as *regions* in this study) and matched with the time period of the mortality dataset (the next change in provincial divisions took place in the early 20th century)[82].

When merging the two sources, we only included municipalities which had information on both mortality and the dialect group. We obtained a dataset comprising 215 municipalities divided into 14 dialect groups and 8 administrative regions (Figure 1). It contained a total of 890,684 registered deaths of which 39,066 were due to pertussis (4.4% of all registered deaths), 33,697 to smallpox (3.8%) and 19,621 to measles (2.2%).

QUANTIFICATION AND STATISTICAL ANALYSIS

We tested which model would best explain variation in mortality from three different infectious diseases in Finland in 1800-1850. The dependent variable was the proportion of deaths due to a specific disease per municipality relative to the total number of deaths recorded in this municipality: it was treated as a binomial factor (deaths due to a specific disease vs all other causes of deaths). We fitted separate models on each disease (smallpox, pertussis and measles), as the relevance of dialect groups to cluster infectious disease mortality risk may vary between infections. We fitted Generalised Linear Mixed Models (GLMMs) with a binomial error structure and a logit link function.

As we did not focus on temporal dynamics here (see Briga et al.[63] for an example), time dependency was not explicitly modelled and our estimates refer to average risks over the 50-year period (following Ketola et al.[15]).

For each disease, we considered a set of 3 models: (1) a control model, *Control*, without information about regional clustering. This model tested if the municipality-wise variation in mortality could be explained only with information about demographic characteristics of each municipality (see details below), the spatial autocorrelation components when needed and the random term *Municipality*, (2) a model testing the relevance of administrative areas where *Municipality* was nested within the administrative region: *Control + Municipality/Region*, (3) a model testing the relevance of dialect groups where the *Municipality* was nested within the dialect group: *Control + Municipality/Dialect group*.

Each model included the following control variables. To avoid regional clusters reflecting municipality-level demographic characteristics, we included municipality population size, number of villages, number of households and municipality area as fixed factors in all models documented in Ketola et al.[15] and municipality as a random intercept to account for the potential dependency of deaths occurring in the same municipality. All continuous variables were standardised with a mean of zero and a standard deviation of one.

Following Ketola et al.[15], when model residuals indicated spatial autocorrelation (Moran's I $p<0.05$, package *ape*)[83], we corrected it using the PCNM approach[84] as implemented in the package *vegan*[85]. This method consists of sequentially adding to a statistical model a number of fixed covariates which grasp the spatial autocorrelation structures and reran this model until no spatial autocorrelation could be detected in the model residuals (Moran's I $p<0.05$). The PCNM approach is commonly used in ecology and epidemiology to correct for autocorrelation structures and its robustness has been confirmed by simulations[86]. For each disease, we initially ran a model including

only the control variables to determine the PCNM variables necessary. In our models, we initially obtained 115 PCNM variables and sequentially added a model-specific number of variables until there was no further spatial autocorrelation (see Results for details). The distribution of model residuals was checked with the R package *DHARMa*[87] and it fulfilled the requirements of homoscedasticity and was without influential datapoints.

After defining for each disease the PNCM variables required to correct for spatial auto-correlation, we used AIC model selection techniques (R package MuMIn)[88] and an a priori model set corresponding to the different hypotheses tested (see above). These models were ranked according to their goodness-of-fit to the data based on the Akaike Information Criterion (AICc)[89–91]. The difference in AIC (ΔAIC) between the model with the lowest AIC (considered as the best model) and the other models provides a measure of how much more likely the best model is than the other models. Following Symonds and Moussalli[91] and Burnham et al.[92], a difference in ΔAIC values above 2 indicates a difference in fit between models. For each fixed variable, we calculated the 95% confidence interval of the estimate (CI95%) to investigate the importance of its effect. When the CI95% overlaps zero, it indicates that the variable is not systematically associated with a higher or lower risk of dying from the disease investigated. For each model, we calculated the proportion of variance explained (R^2, the coefficient of determination) with the package *partR2*[93]. All statistical analyses were conducted on R software v.4.2.1[94] using generalized linear mixed effects models (GLMMs) function *glmer* in the package *lme4*[95].

SUPPLEMENTAL INFORMATION TITLES AND LEGENDS

Table S1. Descriptive statistics of dialect groups. Related to Figure 1.

Table S2. Proportion of total deaths due to smallpox for each municipality (n=215): averaged estimates and their 95% confidence intervals of the models including (a) municipality; (b) municipality and regions; (c) municipality and dialect groups as random factors. Estimates correspond to z-standardised values. See Methods and Table 1 for details of the model selection results. Variables with a 95% confidence interval not including 0 are highlighted in bold. Related to Table1.

Table S3. Proportion of total deaths due to pertussis for each municipality (n=215): averaged estimates and their 95% confidence intervals of the models including (a) municipality; (b) municipality and regions; (c) municipality and dialect groups as random factors. Estimates correspond to z-standardised values. See Methods and Table 1 for details of the model selection results. Variables with a 95% confidence interval not including 0 are highlighted in bold. Related to Table 2.

Table S4. Proportion of total deaths due to measles for each municipality (n=215): averaged estimates and their 95% confidence intervals of the models including (a) municipality; (b) municipality and regions; (c) municipality and dialect groups as random factors. Estimates correspond to z-standardised values. See Methods and Table 1 for details of the model selection results. Variables with a 95% confidence interval not including 0 are highlighted in bold. Related to Table 3.

420

421 ## REFERENCES

422 1. Alexander, R.D. (1974). The evolution of social behavior. Annu. Rev. Ecol. Syst. *5*, 325–383.

423 2. Krause, J., and Ruxton, G.D. (2002). Living in groups by Krause, Jens, Ruxton, Graeme D (Oxford
424 University Press).

425 3. Altizer, S., Nunn, C.L., Thrall, P.H., Gittleman, J.L., Antonovics, J., Cunningham, A.A., Dobson, A.P.,
426 Ezenwa, V.O., Jones, K.E., Pedersen, A.B., et al. (2003). Social organization and parasite risk in
427 mammals: Integrating theory and empirical studies. Annu. Rev. Ecol. Evol. Syst.
428 https://doi.org/10.1146/annurev.ecolsys.34.030102.151725.

429 4. Rifkin, J.L., Nunn, C.L., and Garamszegi, L.Z. (2012). Do animals living in larger groups experience
430 greater parasitism? A meta-analysis. Am. Nat. *180*, 70–82. https://doi.org/10.1086/666081.

431 5. Patterson, J.E.H., and Ruckstuhl, K.E. (2013). Parasite infection and host group size: a meta-analytical
432 review. Parasitology *140*, 803–813. https://doi.org/10.1017/S0031182012002259.

433 6. Kauffman, K., Werner, C.S., Titcomb, G., Pender, M., Rabezara, J.Y., Herrera, J.P., Shapiro, J.T., Solis, A.,
434 Soarimalala, V., Tortosa, P., et al. (2022). Comparing transmission potential networks based on social
435 network surveys, close contacts and environmental overlap in rural Madagascar. J. R. Soc. Interface *19*,
436 20210690. https://doi.org/10.1098/rsif.2021.0690.

437 7. Lucatelli, J., Mariano-Neto, E., and Japyassú, H.F. (2021). Social interaction, and not group size, predicts
438 parasite burden in mammals. Evol. Ecol. *35*, 115–130. https://doi.org/10.1007/s10682-020-10086-6.

439 8. Nunn, C.L., Jordán, F., McCabe, C.M., Verdolin, J.L., and Fewell, J.H. (2015). Infectious disease and
440 group size: more than just a numbers game. Philos. Trans. R. Soc. B Biol. Sci. *370*, 20140111.
441 https://doi.org/10.1098/rstb.2014.0111.

442 9. Sah, P., Mann, J., and Bansal, S. (2017). Disease implications of animal social network structure: A
443 synthesis across social systems. J. Anim. Ecol. *87*, 546–558. https://doi.org/10.1111/1365-2656.12786.

444 10. Salazar, M.F.M., Waldner, C., Stookey, J., and Bollinger, T.K. (2016). Infectious disease and grouping
445 patterns in mule deer. PLOS ONE *11*, e0150830. https://doi.org/10.1371/journal.pone.0150830.

446 11. Stroeymeyt, N., Grasse, A.V., Crespi, A., Mersch, D.P., Cremer, S., and Keller, L. (2018). Social network
447 plasticity decreases disease transmission in a eusocial insect. Science *362*, 941–945.
448 https://doi.org/10.1126/science.aat4793.

449 12. Bjørnstad, O.N., Finkenstädt, B.F., and Grenfell, B.T. (2002). Dynamics of measles epidemics: Estimating
450 scaling of transmission rates using a time series SIR Model. Ecol. Monogr. *72*, 169–184.
451 https://doi.org/10.2307/3100023.

452 13. London, W.P., and Yorke, J.A. (1973). Recurrent outbreaks of measles, chickenpox and mumps: I.
453 Seasonal variation in contact rates. Am. J. Epidemiol. *98*, 453–468.
454 https://doi.org/10.1093/oxfordjournals.aje.a121575.

455 14. Martinez, M.E. (2018). The calendar of epidemics: Seasonal cycles of infectious diseases. PLOS Pathog.
456 *14*, e1007327. https://doi.org/10.1371/journal.ppat.1007327.

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

16. Johnson, P.T.J., de Roode, J.C., and Fenton, A. (2015). Why infectious disease research needs community ecology. Science *349*, 1259504. https://doi.org/10.1126/science.1259504.

17. Nunn, C.L., Craft, M.E., Gillespie, T.R., Schaller, M., and Kappeler, P.M. (2015). The sociality–health–fitness nexus: synthesis, conclusions and future directions. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20140115. https://doi.org/10.1098/rstb.2014.0115.

18. Cremer, S., Armitage, S.A.O., and Schmid-Hempel, P. (2007). Social immunity. Curr. Biol. *17*, R693–R702. https://doi.org/10.1016/j.cub.2007.06.008.

19. Stockmaier, S., Ulrich, Y., Albery, G.F., Cremer, S., and Lopes, P.C. (2023). Behavioural defences against parasites across host social structures. Funct. Ecol. *37*, 809–820. https://doi.org/10.1111/1365-2435.14310.

20. Becker, D.J., Albery, G.F., Kessler, M.K., Lunn, T.J., Falvo, C.A., Czirják, G.Á., Martin, L.B., and Plowright, R.K. (2020). Macroimmunology: The drivers and consequences of spatial patterns in wildlife immune defence. J. Anim. Ecol. *89*, 972–995. https://doi.org/10.1111/1365-2656.13166.

21. Albery, G.F., Kirkpatrick, L., Firth, J.A., and Bansal, S. (2021). Unifying spatial and social network analysis in disease ecology. J. Anim. Ecol. *90*, 45–61. https://doi.org/10.1111/1365-2656.13356.

22. Albery, G.F., Eskew, E.A., Ross, N., and Olival, K.J. (2020). Predicting the global mammalian viral sharing network using phylogeography. Nat. Commun. *11*, 2260. https://doi.org/10.1038/s41467-020-16153-4.

23. Cantor, M., Maldonado-Chaparro, A.A., Beck, K.B., Brandl, H.B., Carter, G.G., He, P., Hillemann, F., Klarevas-Irby, J.A., Ogino, M., Papageorgiou, D., et al. (2021). The importance of individual-to-society feedbacks in animal ecology and evolution. J. Anim. Ecol. *90*, 27–44. https://doi.org/10.1111/1365-2656.13336.

24. He, P., Montiglio, P.-O., Somveille, M., Cantor, M., and Farine, D.R. (2021). The role of habitat configuration in shaping animal population processes: a framework to generate quantitative predictions. Oecologia *196*, 649–665. https://doi.org/10.1007/s00442-021-04967-y.

25. Somveille, M., Firth, J.A., Aplin, L.M., Farine, D.R., Sheldon, B.C., and Thompson, R.N. (2018). Movement and conformity interact to establish local behavioural traditions in animal populations. PLOS Comput. Biol. *14*, e1006647. https://doi.org/10.1371/journal.pcbi.1006647.

26. Churakov, M., Villabona-Arenas, C.J., Kraemer, M.U.G., Salje, H., and Cauchemez, S. (2019). Spatio-temporal dynamics of dengue in Brazil: Seasonal travelling waves and determinants of regional synchrony. PLoS Negl. Trop. Dis. *13*, e0007012. https://doi.org/10.1371/journal.pntd.0007012.

27. Grenfell and Bolker (1998). Cities and villages: infection hierarchies in a measles metapopulation. Ecol. Lett. *1*, 63–70. https://doi.org/10.1046/j.1461-0248.1998.00016.x.

28. Diener, A., and Hagen, J. (2012). Borders: a very short introduction (Oxford University Press).

29. Leslie, S., Winney, B., Hellenthal, G., Davison, D., Boumertit, A., Day, T., Hutnik, K., Royrvik, E.C., Cunliffe, B., Lawson, D.J., et al. (2015). The fine-scale genetic structure of the British population. Nature *519*, 309–314. https://doi.org/10.1038/nature14230.

30. Izdebski, A., Guzowski, P., Poniat, R., Masci, L., Palli, J., Vignola, C., Bauch, M., Cocozza, C., Fernandes, R., Ljungqvist, F.C., et al. (2022). Palaeoecological data indicates land-use changes across Europe linked to spatial heterogeneity in mortality during the Black Death pandemic. Nat. Ecol. Evol., 1–10. https://doi.org/10.1038/s41559-021-01652-4.

31. Alkhamis, M.A., and VanderWaal, K. (2016). Spatial and temporal epidemiology of lumpy skin disease in the Middle East, 2012–2015. Front. Vet. Sci. *3*. https://doi.org/doi.org/10.3389/fvets.2016.00019.

32. Gastañaduy, P.A., Banerjee, E., DeBolt, C., Bravo-Alcántara, P., Samad, S.A., Pastor, D., Rota, P.A., Patel, M., Crowcroft, N.S., and Durrheim, D.N. (2018). Public health responses during measles outbreaks in elimination settings: Strategies and challenges. Hum. Vaccines Immunother. *14*, 2222. https://doi.org/10.1080/21645515.2018.1474310.

33. Skoff, T.H., Hadler, S., and Hariri, S. (2019). The epidemiology of nationally reported pertussis in the United States, 2000-2016. Clin. Infect. Dis. *68*, 1634–1640. https://doi.org/10.1093/cid/ciy757.

34. Tomljenovic, M., Lakic, M., Vilibic-Cavlek, T., Filipovic, S.K., Vucina, V.V., Babic-Erceg, A., Ljubic, M., Novosel, I.P., Ilic, M., Tabain, I., et al. (2020). Measles outbreak in Dubrovnik-Neretva County, Croatia, May to June 2018. Eurosurveillance *25*, 1900434. https://doi.org/10.2807/1560-7917.ES.2020.25.7.1900434.

35. Chambers, J.K., and Trudgill, P. (1998). Dialectology 2nd ed. (Cambridge University Press) https://doi.org/10.1017/CBO9780511805103.

36. Wieling, M., Nerbonne, J., and Baayen, R.H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. PLOS ONE *6*, e23613. https://doi.org/10.1371/journal.pone.0023613.

37. Lynch, R., Loehr, J., Lummaa, V., Honkola, T., Pettay, J., and Vesakoski, O. (2022). Socio-cultural similarity with host population rather than ecological similarity predicts success and failure of human migrations. Proc. R. Soc. B Biol. Sci. *289*, 20212298. https://doi.org/10.1098/rspb.2021.2298.

38. Honkola, T., Ruokolainen, K., Syrjänen, K.J.J., Leino, U.-P., Tammi, I., Wahlberg, N., and Vesakoski, O. (2018). Evolution within a language: environmental differences contribute to divergence of dialect groups. BMC Evol. Biol. *18*, 132. https://doi.org/10.1186/s12862-018-1238-6.

39. Anderson, R.M., and May, R.M. (1991). Infectious diseases of humans: dynamics and control (OUP Oxford).

40. Briga, M., Ketola, T., and Lummaa, V. (2022). The epidemic dynamics of three childhood infections and the impact of first vaccination in 18th and 19th century Finland. Preprint at medRxiv, https://doi.org/10.1101/2022.10.30.22281707 https://doi.org/10.1101/2022.10.30.22281707.

41. Ukonaho, S., Lummaa, V., and Briga, M. (2022). The long-term success of mandatory vaccination laws after implementing the first vaccination campaign in 19th century rural Finland. Am. J. Epidemiol. *191*, 1180–1189. https://doi.org/10.1093/aje/kwac048.

42. Aarikka, L. (2023). Murre ja sen tutkimus: Näkökulmia fennistisen murteentutkimuksen historiaan ja kieli-ideologioihin 1871–2017.

43. Syrjänen, K., Honkola, T., Lehtinen, J., Leino, A., Vesakoski, O., and Leino, U. (2016). Applying population genetic approaches within languages. Lang. Dyn. Change *6*, 235–283. https://doi.org/10.1163/22105832-00602002.

44. Alimohamadi, Y., Zahraei, S.M., Karami, M., Yaseri, M., Lotfizad, M., and Holakouie-Naieni, K. (2020). Spatio-temporal analysis of Pertussis using geographic information system among Iranian population during 2012-2018. Med. J. Islam. Repub. Iran *34*, 22. https://doi.org/10.34171/mjiri.34.22.

45. Davenport, R.J., Satchell, M., and Shaw-Taylor, L.M.W. (2018). The geography of smallpox in England before vaccination: A conundrum resolved. Soc. Sci. Med. *206*, 75–85. https://doi.org/10.1016/j.socscimed.2018.04.019.

46. Fridlizius, G., and Ohlsson, R. (1984). Mortality patterns in Sweden 1751-1802'. In Pre-Industrial Population Change. Almqvist and Wiksell International, Lund.

47. Pitkänen, K.J., Mielke, J.H., and Jorde, L.B. (1989). Smallpox and its eradication in Finland: Implications for disease control. Popul. Stud. *43*, 95–111. https://doi.org/10.1080/0032472031000143866.

48. Turpeinen, O. (1978). Infectious diseases and regional differences in Finnish death rates, 1749-1773. Popul. Stud. *32*, 523–533. https://doi.org/10.2307/2173725.

49. Broutin, H., Guégan, J.-F., Elguero, E., Simondon, F., and Cazelles, B. (2005). Large-scale comparative analysis of pertussis population dynamics: periodicity, synchrony, and impact of vaccination. Am. J. Epidemiol. *161*, 1159–1167. https://doi.org/10.1093/aje/kwi141.

50. Panhuis, W.G. van, Choisy, M., Xiong, X., Chok, N.S., Akarasewi, P., Iamsirithaworn, S., Lam, S.K., Chong, C.K., Lam, F.C., Phommasak, B., et al. (2015). Region-wide synchrony and traveling waves of dengue across eight countries in Southeast Asia. Proc. Natl. Acad. Sci. *112*, 13069–13074. https://doi.org/10.1073/pnas.1501375112.

51. Tennant, W.S.D., Tildesley, M.J., Spencer, S.E.F., and Keeling, M.J. (2020). Climate drivers of plague epidemiology in British India, 1898–1949. Proc. R. Soc. B Biol. Sci. *287*, 20200538. https://doi.org/10.1098/rspb.2020.0538.

52. Ostfeld, R.S., Glass, G.E., and Keesing, F. (2005). Spatial epidemiology: an emerging (or re-emerging) discipline. Trends Ecol. Evol. *20*, 328–336. https://doi.org/10.1016/j.tree.2005.03.009.

53. Porta, M. (2014). A Dictionary of epidemiology (Oxford University Press).

54. Guerra, F.M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S.L., Li, Y., and Crowcroft, N.S. (2017). The basic reproduction number (R0) of measles: a systematic review. Lancet Infect. Dis. *17*, e420–e428. https://doi.org/10.1016/S1473-3099(17)30307-9.

55. Grenfell, B.T., Bjørnstad, O.N., and Kappey, J. (2001). Travelling waves and spatial hierarchies in measles epidemics. Nature *414*, 716–723. https://doi.org/10.1038/414716a.

56. Björkstén, J. (1902). Vaccinationens historia i Finland I (Helsingfors Centraltryckeri).

57. Björkstén, J. (1908). Vaccinationens historia i Finland II (Helsingfors Centraltryckeri).

58. Viboud, C., Bjørnstad, O.N., Smith, D.L., Simonsen, L., Miller, M.A., and Grenfell, B.T. (2006). Synchrony, waves, and spatial hierarchies in the spread of influenza. Science. https://doi.org/10.1126/science.1125237.

59. Domenech de Cellès, M., and Rohani, P. (2024). Pertussis vaccines, epidemiology and evolution. Nat. Rev. Microbiol. *22*, 722–735. https://doi.org/10.1038/s41579-024-01064-8.

60. Kilgore, P.E., Salim, A.M., Zervos, M.J., and Schmitt, H.-J. (2016). Pertussis: Microbiology, disease, treatment, and prevention. Clin. Microbiol. Rev. *29*, 449–486. https://doi.org/10.1128/CMR.00083-15.

61. Benton, M.L., Abraham, A., LaBella, A.L., Abbot, P., Rokas, A., and Capra, J.A. (2021). The influence of evolutionary history on human health and disease. Nat. Rev. Genet. *22*, 269–283. https://doi.org/10.1038/s41576-020-00305-9.

62. Jaadla, H., Potter, E., Keibek, S., and Davenport, R. (2020). Infant and child mortality by socio-economic status in early nineteenth-century England. Econ. Hist. Rev. *73*, 991–1022. https://doi.org/10.1111/ehr.12971.

63. Briga, M., Ukonaho, S., Pettay, J.E., Taylor, R.J., Ketola, T., and Lummaa, V. (2021). The seasonality of three childhood infections in a pre-industrial society without schools. medRxiv.

64. Pitkänen, K. (1977). The reliability of the registration of births and deaths in Finland in the eighteenth and nineteenth centuries: Some examples. Scand. Econ. Hist. Rev. *25*, 138–159. https://doi.org/10.1080/03585522.1977.10407878.

65. Luther, G. (1993). Suomen tilastotoimen historia vuoteen (WSOY).

66. Vuorinen, H.S. (1999). Suomalainen tautinimistö ennen bakteriologista vallankumousta. Hippokrates Suom. Lääketieteen Hist. Seuran Vuosik. *16*, 33–61.

67. Saarivirta, T., Consoli, D., and Dhondt, P. (2012). The evolution of the Finnish health-care system early 19th Century and onwards. Int. J. Bus. Soc. Sci. Stud. *3*, 243–257.

68. Mukula, J., and Rantanen, O. (1989). Climatic risks to the yield and quality of the field crops in Finland III. Winter rye 1969-1986. Ann. Agric. Fenn. *28*, 3–11.

69. Rantanen, T., Tolvanen, H., Honkola, T., and Vesakoski, O. (2021). A comprehensive spatial model for historical travel effort - a case study in Finland. Fenn. - Int. J. Geogr. *199*, 61–88. https://doi.org/10.11143/fennia.98357.

70. Kerminen, S., Havulinna, A.S., Hellenthal, G., Martin, A.R., Sarin, A.-P., Perola, M., Palotie, A., Salomaa, V., Daly, M.J., Ripatti, S., et al. (2017). Fine-scale genetic structure in Finland. G3 Genes Genomes Genet. *7*, 3459–3468. https://doi.org/10.1534/g3.117.300217.

71. Voutilainen, M. (2017). Marriage and household structure in rural pre-famine Finland, 1845-65.

72. Waris, E. (1995). The extended family in the Finnish Karelia. The family system in Ruokolahti 1750–1850. Scand. J. Hist. *20*, 109–128. https://doi.org/10.1080/03468759508579298.

73. Raento, P., and Husso, K. (2002). Cultural diversity in Finland. Fenn. Int. J. Geogr.

74. Voutilainen, M. (2017). Poverty and tax exemptions in mid nineteenth century Finland. J. Finn. Stud. *20*.

75. Voutilainen, M., Turunen, R., and Ojala, J. (2020). Multi-currency regime and markets in early nineteenth-century Finland. Financ. Hist. Rev. *27*, 115–138. https://doi.org/10.1017/S0968565019000210.

76. Voutilainen, M., Helske, J., and Högmander, H. (2020). A Bayesian reconstruction of a historical population in Finland, 1647–1850. Demography *57*, 1171–1192. https://doi.org/10.1007/s13524-020-00889-1.

77. Valijärvi, D.A., Riitta-Liisa ed. (2023). The Uralic languages 2nd ed. (Routledge) https://doi.org/10.4324/9781315625096.

78. Itkonen, T.I. (1948). Suomen lappalaiset vuoteen 1945 2. painos. (Werner Söderström Oy).

79. Kettunen, L. (1940). Suomen murteet. 3 A, murrekartasto (Suomalaisen Kirjallisuuden Seura).

80. Kettunen, L. (2021). Murrekartasto, Lauri Kettunen (versio 1). Kotimaisten kielten keskus. http://urn.fi/urn:nbn:fi:csc-kata20151130145346403821. http://urn.fi/urn:nbn:fi:csc-kata20151130145346403821 http://urn.fi/urn:nbn:fi:csc-kata20151130145346403821.

81. Santaharju, J., Syrjänen, K., Honkola, T., Seppä, P., Vesakoski, O., and Unni, L. (2023). New version of the digitized Dialect Atlas of Finnish by Lauri Kettunen. Version 0.1 (Zenodo). https://doi.org/10.5281/zenodo.10078078 https://doi.org/10.5281/zenodo.10078078.

82. Kangasniemi, M. (1986). Lääninhallinto 350 vuotta = Länsförvaltningen 350 år (Sisäasiainministeriö).

83. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in {R}. Bioinformatics *35*, 526–528.

84. Voutilainen, A., Tolppanen, A.-M., Vehviläinen-Julkunen, K., and Sherwood, P.R. (2014). From spatial ecology to spatial epidemiology: modeling spatial distributions of different cancer types with principal coordinates of neighbor matrices. Emerg. Themes Epidemiol. *11*, 1–10.

85. Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., et al. (2022). vegan: Community Ecology Package. Version R package version 2.6-2.

86. Bauman, D., Drouet, T., Dray, S., and Vleminckx, J. (2018). Disentangling good from bad practices in the selection of spatial or phylogenetic eigenvectors. Ecography *41*, 1638–1649. https://doi.org/10.1111/ecog.03380.

87. Hartig, F. (2022). DHARMa: Residual diagnostics for hierarchical (multi-level / mixed) regression models. Version R package version 0.4.6.

88. Bartoń, K. (2016). MuMIn: Multi-model inference. R package version 1.15.6. Version R package version 1.15.6.

89. Burnham, K.P., and Anderson, D.R. (2002). Model selection and multi-model inference : A practical information-theoretic approach 2nd ed. (Springer-Verlag).

90. Grueber, C.E., Nakagawa, S., Laws, R.J., and Jamieson, I.G. (2011). Multimodel inference in ecology and evolution: challenges and solutions. J. Evol. Biol. *24*, 699–711. https://doi.org/10.1111/j.1420-9101.2010.02210.x.

91. Symonds, M.R.E., and Moussalli, A. (2011). A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behav. Ecol. Sociobiol. *65*, 13–21.

92. Burnham, K.P., Anderson, D.R., and Huyvaert, K.P. (2011). AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behav. Ecol. Sociobiol. *65*, 23–35. https://doi.org/10.1007/s00265-010-1029-6.

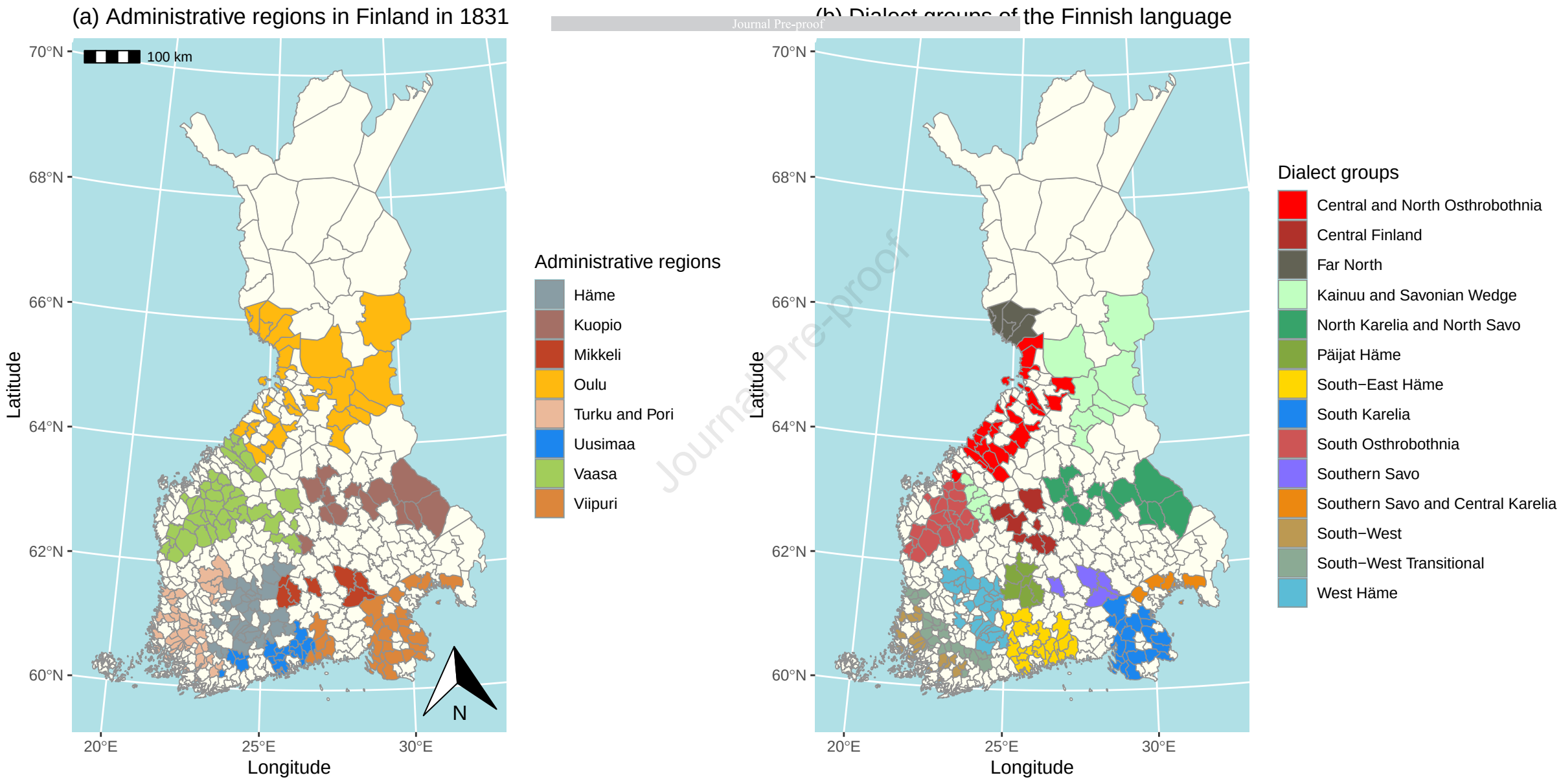
647 93. Stoffel, M.A., Nakagawa, S., and Schielzeth, H. (2021). partR2: partitioning R2 in generalized linear
648 mixed models. PeerJ *9*, e11414. https://doi.org/10.7717/peerj.11414.

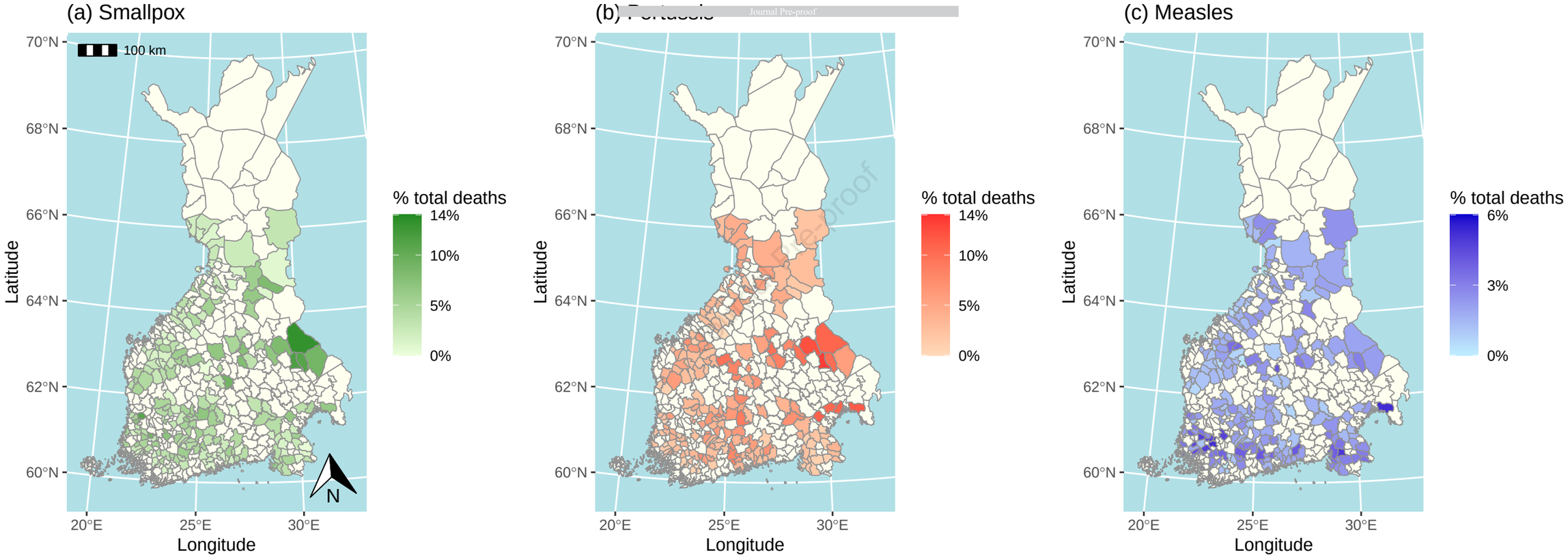
649 94. R Core Team (2022). R: A Language and Environment for Statistical Computing. Version 4.2.1 (R
650 Foundation for Statistical Computing).

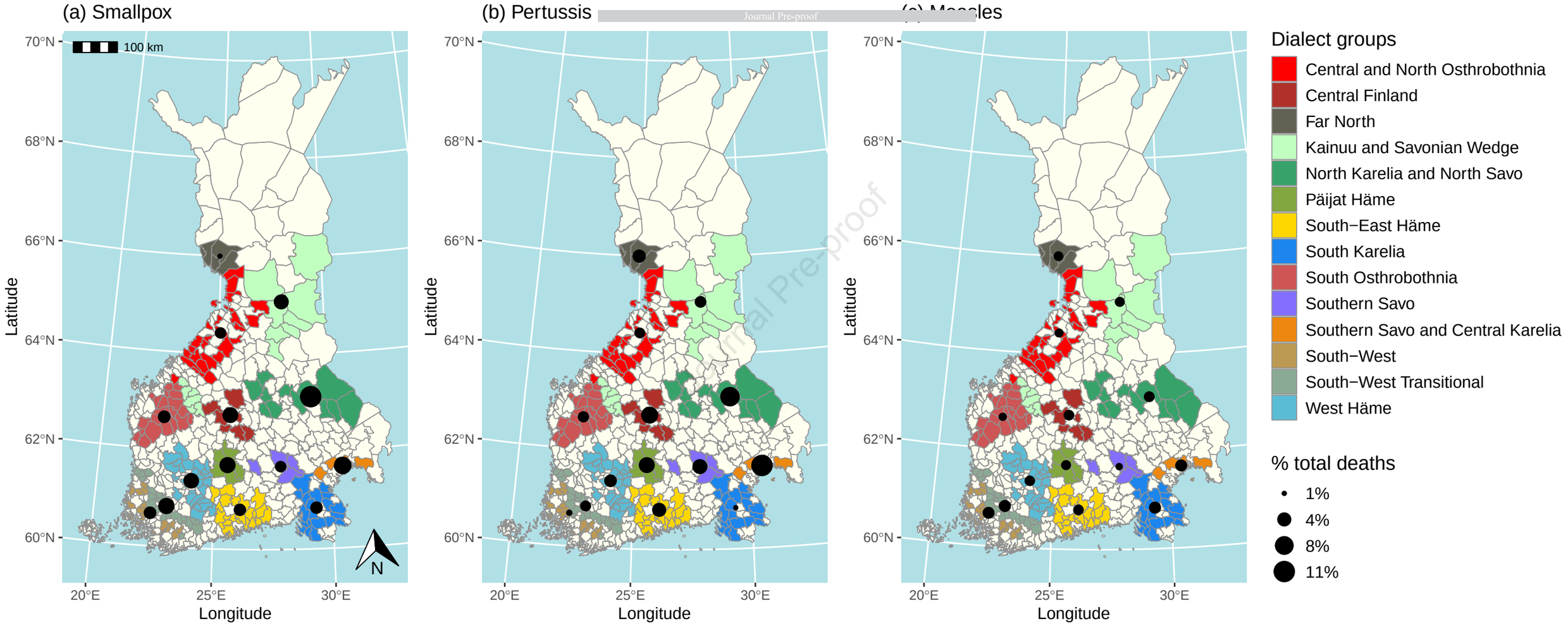
651 95. Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). lme4: Linear mixed-effects models using S4
652 classes. R package version 1.1-9. Version R package version 1.1-9.

653

654







RESEARCH HIGHLIGHTS

- The regional structure of diseases is critical to understand their dynamics
- Using dialect groups instead of administrative regions may be a better proxy
- We analysed mortality records from Finland (1800-1850) for 3 infectious diseases
- The best proxy was dialect for pertussis, regions for smallpox, none for measles

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Data	OSF	doi.org/10.17605/OSF.IO/C76DQ
Code	OSF	doi.org/10.17605/OSF.IO/C76DQ
Software and algorithms		
R software v4.2.1	R Core Team	https://cran.r-project.org/