PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Dissecting recurrent waves of pertussis across the boroughs of London

Arash Saeidpour[1]*, **Shweta Bansal**[2], **Pejman Rohani**[1,3,4]

1 Odum School of Ecology, University of Georgia, Athens, Georgia, United States of America, **2** Department of Biology, Georgetown University, Washington, D.C., United States of America, **3** Department of Infectious Diseases, College of Veterinary Medicine, University of Georgia, Athens, Georgia, United States of America, **4** Center for Influenza Disease & Emergence Research (CIDER), Athens, Georgia, United States of America

* arashs@uga.edu

OPEN ACCESS

Citation: Saeidpour A, Bansal S, Rohani P (2022) Dissecting recurrent waves of pertussis across the boroughs of London. PLoS Comput Biol 18(4): e1009898. https://doi.org/10.1371/journal.pcbi.1009898

Editor: Alex Perkins, University of Notre Dame, UNITED STATES

Received: March 17, 2021

Accepted: February 4, 2022

Published: April 14, 2022

Data Availability Statement: We have made available all source code together with a synthetically generated dataset through OSF platform at https://osf.io/b5xuw/ DOI 10.17605/OSF.IO/B5XUW.

Funding: S.B. is supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM123007. P.R. is supported with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under

Abstract

Pertussis has resurfaced in the UK, with incidence levels not seen since the 1980s. While the fundamental causes of this resurgence remain the subject of much conjecture, the study of historical patterns of pathogen diffusion can be illuminating. Here, we examined time series of pertussis incidence in the boroughs of Greater London from 1982 to 2013 to document the spatial epidemiology of this bacterial infection and to identify the potential drivers of its percolation. The incidence of pertussis over this period is characterized by 3 distinct stages: a period exhibiting declining trends with 4-year inter-epidemic cycles from 1982 to 1994, followed by a deep trough until 2006 and the subsequent resurgence. We observed systematic temporal trends in the age distribution of cases and the fade-out profile of pertussis coincident with increasing national vaccine coverage from 1982 to 1990. To quantify the hierarchy of epidemic phases across the boroughs of London, we used the Hilbert transform. We report a consistent pattern of spatial organization from 1982 to the early 1990s, with some boroughs consistently leading epidemic waves and others routinely lagging. To determine the potential drivers of these geographic patterns, a comprehensive parallel database of borough-specific features was compiled, comprising of demographic, movement and socio-economic factors that were used in statistical analyses to predict epidemic phase relationships among boroughs. Specifically, we used a combination of a feed-forward neural network (FFNN), and SHapley Additive exPlanations (SHAP) values to quantify the contribution of each covariate to model predictions. Our analyses identified a number of predictors of a borough's historical epidemic phase, specifically the age composition of households, the number of agricultural and skilled manual workers, latitude, the population of public transport commuters and high-occupancy households. Univariate regression analysis of the 2012 epidemic identified the ratio of cumulative unvaccinated children to the total population and population of Pakistan-born population to have moderate positive and negative association, respectively, with the timing of epidemic. In addition to providing a comprehensive overview of contemporary pertussis transmission in a large metropolitan population, this study has identified the characteristics that determine the spatial spread of this bacterium across the boroughs of London.

Contract No. 75N93021C00018 (NIAID Centers of Excellence for Influenza Research and Response, CEIRR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Recent years have witnessed a resurgence of pertussis, a vaccine preventable bacterial disease, in countries with high estimated immunization rates. Here, to understand the timing of pertussis epidemics in different populations, we employed signal processing techniques to analyze a rich dataset of weekly incidence across the boroughs of London from 1982 to 2013. We observed 4-year epidemic cycles with a distinctly consistent spatial organization across the boroughs from 1982 to 1990, with some boroughs typically ahead of the wave and others consistently lagging. We identified putative demographic and socio-economic mechanisms that determined this spatial organization of these outbreaks using an interpretable machine learning approach. The reemergence of pertussis from 2006 onward, on the contrary, did not exhibit regular epidemic waves, instead there was a transition from spatially organized waves in the 1980s to a largely unstructured mosaic in the resurgence era.

Introduction

Understanding the spatial dynamics and diffusion of infectious diseases is increasingly recognized as an important component in shedding light on their underlying epidemiology and developing effective control policies [1–6]. This is true both for emerging infectious diseases whose spatial trajectory needs to be predicted [7–11] and those vaccine-preventable infectious diseases whose elimination is stymied by the stubborn mosaic of under-vaccination [12–16] or asynchronous metapopulation persistence [17, 18]. Key to this effort has been attempting to understand the drivers of the observed spatial ecology of these systems, including the impact of host movement [5, 11, 19, 20] and workflows [7], urban-rural gradients [1], regional variation in climate seasonality [4, 21–23] or population demography [3] and the impact of road networks [24]. These studies have identified insights that may, in principle, inform control policies including the implementation of spatially synchronised pulsed vaccination to increase the likelihood of elimination when epidemics are asynchronous [17, 25, 26], quantifying the impacts of border closures [9] and geographic barriers [27, 28] on transmission.

Here, we examined the spatial ecology of recurrent pertussis epidemics in the boroughs of London from 1982–2013. Pertussis is a bacterial pathogen once considered a disease of childhood [16, 29, 30] and whose recent resurgence in a number of countries with high estimated immunization coverage has generated a great deal of debate [31–37]. Prior studies of the spatial epidemiology of pertussis have demonstrated a range of distinct patterns [16]. Dynamics in the pre-vaccine era (prior to 1957) in the cities of England & Wales were shown to be characterized by spatial asynchrony, which gave way to largely synchronous 4-year epidemics in the early whole cell (wP) vaccination era [18]. Similarly, studying incidence reports from provinces of Thailand, Blackwood *et al.* [38] documented regular annual pertussis outbreaks that were highly in-phase with each other. Finally, a study of pertussis incidence in US states found evidence of traveling waves in the early years of wP vaccination, 1950s-1960s [5, 39]. These waves spread inwards from two distinct regional foci located at each coast, one in the Northeast and the other in the Northwest [5]. Comparable analyses of pertussis data for the US since its re-emergence over the past four decades [39, 40] did not identify any discernible spatial organization [5].

To dissect the spatial epidemiology of pertussis at a finer spatial scale than these previous analyses, we took advantage of a rich and well-documented disease incidence dataset collected

by Public Health England (PHE) [41–44]. These weekly *Notifications of Infectious Diseases (NOIDs)* data from 1982 to 2013 confirmed the evolving nature of pertussis epidemiology. Over this period, there were three distinct stages in the incidence of pertussis in the boroughs of London: (i) a period of declining trends with four-year inter-epidemic cycles from 1982 to 1994, (ii) a period characterized by a deep trough in incidence from 1994 until 2006, and (iii) a subsequent reemergence of the disease leading to a large outbreak in 2012. In the first phase, increasing infant wP vaccination coverage was associated with a drop in incidence, an increase in the age of reported infected and a higher frequency of weeks with no pertussis notifications [45]. These data also clearly identified regular 4-year epidemics in London boroughs, with surprisingly consistent phase relationships among locations through time. In contrast, the reemergence of pertussis from 2006 onward did not exhibit periodic epidemic waves.

We sought to identify putative mechanisms that determined the spatial organization of these outbreaks using an interpretable machine learning approach. We integrated the NOIDs data with a comprehensive database of borough-specific demographic, movement and socio-economic factors to carry out statistical analyses of epidemic phase relationships among the boroughs.

Materials and methods

Incidence data

Weekly pertussis incidence for the 32 boroughs of Greater London were obtained from *Notifications of Infectious Diseases* (NOIDs), spanning 1982–2013. Note that prior to 1994, case confirmations were based solely on the isolation and culture of the causative bacterium, *Bordetella pertussis* [46]. Since then, an enhanced surveillance system with epidemiological follow up for laboratory confirmation of cases has been established, with confirmation of clinically suspected cases achieved by culture, use of PCR to detect bacterial DNA from nasopharyngeal samples or antibody detection performed on oral fluid or serum [41, 46].

Geographic, demographic and socioeconomic variables

We compiled a comprehensive set of demographic and socio-economic factors from aggregated census data collected in 1981, 1991, 2001 and 2011 at the borough level (see Table 1). Since the census is carried out every 10 years, covariate values falling in-between census years were obtained by interpolation using B-splines, as shown in S1 Fig and mapped in S2–S5 Figs.

Household age composition. Past work has indicated a shift in the age distribution of cases in England and Wales from 1982 to 1994 [45]. While children and infants accounted for the majority of cases in the early years of this era, there was higher incidence in adults and adolescents in later years [41]. In order to examine the potential contribution of different age groups to epidemic phase lags, we complied data on household age composition from the census data, which provide *Number of households with no children*, *Number of households with children aged 0–4* and *Number of households with children age 5–16* for each borough. We used these data as predictor features in our statistical models.

Movement and geographic features. It has been shown that correlation between disease prevalence in connected populations is dependent on the flow of individuals between them [7, 20, 47]. This prompted us to examine the potential relationship between public and private commuter travel patterns and epidemic phase lags. We have incorporated the sum of all public (British rail, other rail, bus) and private (car, motorcycle, bicycle) modes of travel to work (*Travel public* and *Travel private*, respectively). We were also motivated to study the potential impact of gravitational coupling, whereby the magnitude of epidemiological exchange between two locations is determined by their distance and respective population sizes. Previous studies

Table 1. Description of geographic and census variables used in this study.

Variable	Definition	Source
Number of households with no children	-	[58]
Number of households with children aged 0–4	-	[58]
Number of households with children aged 5–16	-	[58]
Born in Africa	Number of people born in Africa	[58]
Born in Caribbean	Number of people born in Caribbean	[58]
Born in India	Number of people born in India	[58]
Born in Pakistan	Number of people born in Pakistan	[58]
Pres. & Res. comm. estbls.	Number of people present and all usual residents in communal establishments	[58]
Households with >1.5 PPR	Number of households with >1.5 people per room	[58]
Not self contained houses	Number of not self-contained housesholds (self-containment means that all rooms, including the kitchen, bathroom and toilet are behind a door (but not necessarily a single door) only that household can use.)	[58]
SEG 1–4	Number of people in SEG*1 to 4 (Employers, Managers and Professional Workers)	[58]
SEG 8,9,12	Number of people in SEG*8,9 and 12 (Skilled Manual Workers)	[58]
SEG 7–10	Number of people in SEG*7 to 10 (Skilled Manual Workers)	[58]
SEG 11	Number of people in SEG*11 (Unskilled Manual Workers)	[58]
SEG 13–15	Number of people in SEG*13 to 15 (Farmers and Agricultural Workers)	[58]
SEG 16–17	Number of people in SEG*16 and 17 (Armed Forces and Not Stated)	[58]
Travel public	Number of people who use means of public transportation to travel to work (British Rails, other rail, bus)	[58]
Travel other modes	Number of people who use other means of transportation to travel to work (car, motorcycle, bicycle, other)	[58]
Inland Area (Hectares)	Borough's inland area (excluding bodies of water)	[59]
Longitude	Longitude of borough's geometric centroid	[60]
Latitude	Latitude of borough's geometric centroid	[60]

* Socio-economic group, see [52].

https://doi.org/10.1371/journal.pcbi.1009898.t001

have identified a role of gravity models in the epidemics of measles in England and Wales [1, 19], waves of influenza in the US [7], the spread of ebola in West Africa [9] and hierarchies of dengue haemorrhagic fever in Thailand [48]. Thus, to allow for this mechanism, the *Latitude* and *Longitude* of the geographical centroid of each borough and *Inland Area* (Hectares) were added as geographical factors.

Socio-economic features. We sought to examine whether socio-economic heterogeneity among boroughs can explain epidemic phase relationships. Thus, we included the Socio-Economic Group *(SEG)*, which is a classification based on the employment status and occupation of individuals. According to the UK census, the population of each borough is stratified into 6 aggregate *SEG* classes, as shown in Table 1. Additionally, the number of people present and all usual residents in communal establishments outside the medical and care sector (*Pres. & Res. comm. estbls.*) and the number of not self-contained houses (*Not self-contained houses*) were also considered as potential indicators of economic status. To investigate possible statistical association between the composition of population ethnicity and epidemic timing of outbreaks, the number of individuals by place of birth for four major immigrant populations (Africa, Caribbean, India and Pakistan) were considered.

Household characteristics. The role of household characteristics has been identified in spatial spread of influenza across Europe [49] and pertussis epidemics in England and Wales [50] and the Netherlands [51]. To evaluate the combined effects of household occupancy, the Number of households with > 1.5 people per room (*Households with* > 1.5 *PPR*) was included among predictor variables. As shown in S5 Fig, we identified pairwise linear associations between a number of the described variables (see [52] for more details about the census).

Vaccine coverage estimates. Intuitively, a potentially important factor in shaping pertussis epidemics, especially the timing of outbreaks, may be susceptible births [53–55], which has been previously shown to determine the seasonal outbreaks of rotavirus in the US [3]. We have, however, been unable to obtain borough-specific routine immunization coverage estimates spanning the 1982–1996 period. It is plausible that some of our socio-economic features serve as a proxy for vaccination coverage [56].

We obtained immunization rates from the former Health Protection Agency (HPA) starting in 1997–98, which was the earliest available data at the borough level [57]. We carried out statistical analyses to examine potential association between these vaccine coverage data and epidemic phase lags among boroughs.

Epidemic time lag analysis

To analyze data from the first phase (1982–1994), weekly case count time series from each borough were first square rooted to stabilize the variance and then normalized to have zero mean. Normalized time series were then zero-padded to the next higher power of two to minimize edge effects [61]. We applied the continuous Morlet wavelet transform on the normalized time series to determine the range of dominant periods of the epidemic wave. We then filtered the time series with a band-pass filter corresponding to the dominant period and applied the Hilbert transform to filtered time series to extract the instantaneous phase angles. An important advantage of the Hilbert transform is that phase angles can be extracted for an arbitrary broad-band signal, where signal power is mainly distributed in a specified range of frequencies (see [62] for a comparison of different signal processing methods for quantification of phase synchrony). This contrasts with the continuous wavelet transform, which traditionally has been used for decomposition of ecological time series [1, 5]. We extracted the instantaneous phase angle of each borough at three time points (the outbreaks of 1982, 1986, 1990) and defined the phase lag of a given borough at each snapshot as the difference between the borough's instantaneous phase angle and the mean phase angle of all boroughs at the corresponding instant (the pre-processing and transformation steps are illustrated in S6 Fig).

Wavelet analysis of resurgence incidence data (2006–2013) did not exhibit periodic epidemic waves across the boroughs. To examine the spatial ecology of the 2012 epidemic, therefore, we used Kernel density estimation (KDE) [63] to quantify the epidemic time lag as the mean of a Gaussian distribution.

Fade-out frequency profile

We used a 4-year sliding window to calculate the average number of fade-outs per year. Three consecutive weeks with no reported cases was counted as one fade-out [64–66]. The total number of fade-outs in the counting window were then divided by four to obtain the average number of weekly fade-outs. The counting window is successively shifted 6-months ahead to count the number of fade-outs through time.

Feature importance

To examine the potential statistical association between the epidemic phase lag of boroughs and the predictor variables, we employed different regression methods with varying levels of complexity. The marginal contribution of each variable to the model predictions is deemed as feature importance. The variables are standardized to have mean zero and unit variance. All the regression models are fitted to the standardized data.

We first performed univariate regression to investigate the individual importance of each variable. This was followed by ordinary least squares (OLS) multivariate linear regression analysis to examine the combined effects of variables. Multivariate linear analysis was repeated with L1 (lasso) and L2 (ridge) regularizations [67]. Finally, to capture any potential nonlinear effects, we used a feed-forward neural network model (FFNN). The model was implemented in the 'Keras' [68] Python package with a 'TensorFlow' [69] backend. We constrained our architecture search to a maximum of 3 hidden layers and up to 64 nodes per layer with *tanh* activation. Dropout regularization was used to prevent overfitting [70]. We used Adam optimizer [71] and limited the number of epochs to 2000.

For model evaluation and hyperparameter tuning, we randomly split the data into 64 training samples and 32 test samples. The regularization parameter for lasso and ridge regression analysis was optimized via 10-fold cross validation over the training samples. For the FFNN model hyperparameter optimization, we randomly selected 16 samples out of 64 training samples as the validation set. The model's hyperparameters were tuned via a grid search over the validation set.

Upon training the FFNN model, we used the SHapley Additive exPlanations (SHAP) method [72, 73] to calculate the marginal contribution of each feature to the predicted value by the model. SHAP is a novel, model-agnostic method for measuring local interactions [74], rooted in Shapley values proposed in the theory of cooperative games [75]. Model-agnostic methods provide a *post hoc* interpretation for arbitrary machine learning models by treating them as black-boxes, where interpretation is obtained by fitting an explainable model to the predictions of the black-box model, or measuring the variation in the black-box model predictions with small perturbations in inputs or both [74]. These methods allow for local interpretation of a single input instance or a group of input data (See S1 Text for more information on SHAP).

Results

Epidemiological characteristics

We illustrate pertussis incidence in Greater London during the study period in Fig 1A. Data from the first phase (1982–1994) are characterized by regular 4-year epidemics with a decreasing trend in the amplitude. Routine vaccination coverage, which was less than 60% prior to 1988, rose to approximately 90% from 1990 onwards and was associated with a near-elimination of pertussis reports, coincident with frequent fade-outs. Three consecutive regular epidemics in the years 1982, 1986 and 1990 are evident. We depict the square-rooted weekly incidence across the 32 boroughs in Fig 1B, which shows a comparable pattern of declining incidence throughout the 1990s and frequent fade outs until 2012. Thus, the epidemiology of pertussis over this time span is well described by periods of declining incidence (1982–1994), near absence of reports (1994–2012) and resurgence, culminating in the 2012 epidemic. This re-emergence occurred after the switch in 2004 from the whole-cell (DTwP) to the acellular (DTaP) vaccine in the routine schedule. This switch has been cited as one of the main potential explanations for the recent resurgence of pertussis in the UK [36, 76, 77]. We note, however,

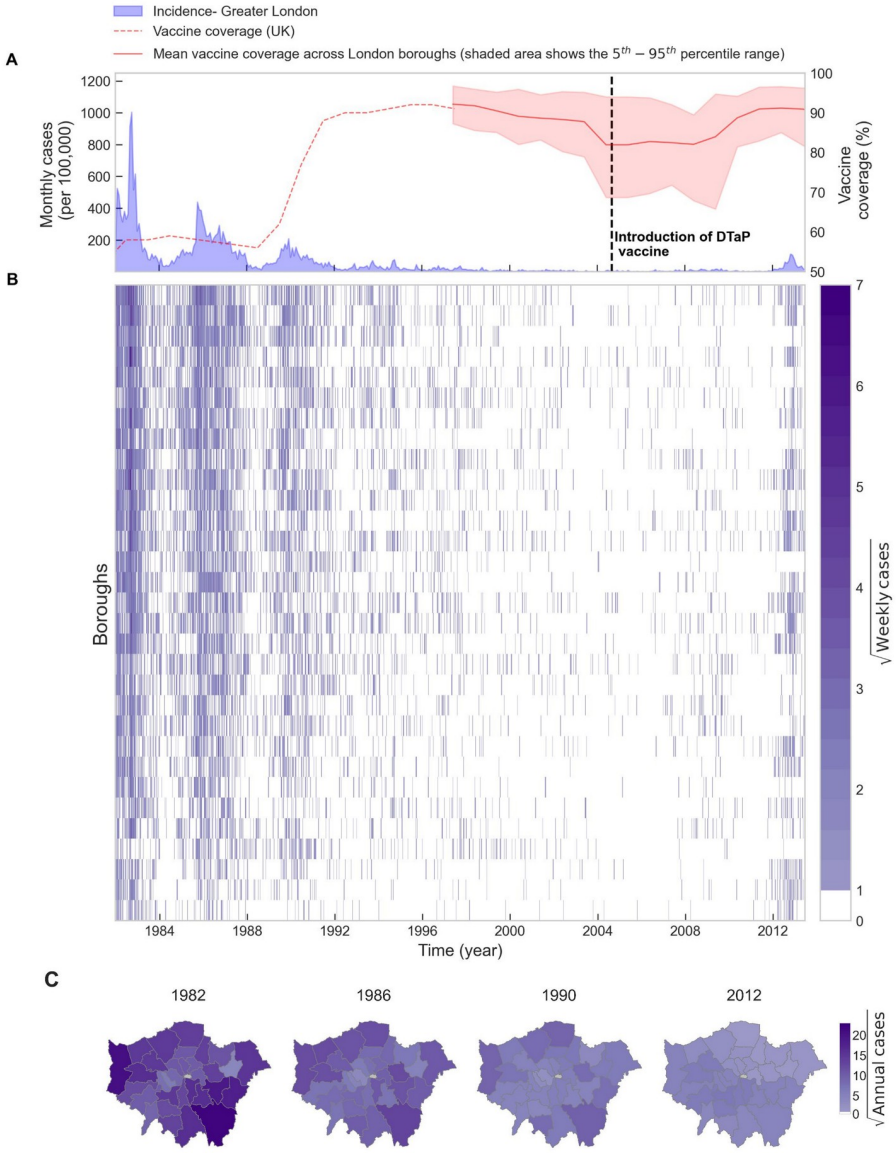

Fig 1. Evolution of pertussis incidence in Greater London. (A) Time series of monthly reported cases in Greater London and annual vaccine coverage rate in the UK (B) Square root of weekly cases of pertussis in London boroughs (C) Choropleth maps of square root of annual cases in the London boroughs at four time snapshots. Map base layers are obtained from London Datastore (https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london) and are available from https://data.london.gov.uk/download/statistical-gis-boundary-files-london/9ba8c833-6370-4b11-abdc-314aa020d5e0/statistical-gis-boundaries-london.zip (The digital boundary file contains Office for National statistics data Crown copyright and database (2012) and contains Ordnance Survey data Crown copyright and database (2012)).

https://doi.org/10.1371/journal.pcbi.1009898.g001

that most boroughs experienced a notable decline in vaccine coverage in the years leading to the vaccine switch (Fig 1A), with a drop in the mean vaccine coverage rates from 92% in 1997 to 82% in 2005. Finally, Fig 1C the annual incidence in boroughs is mapped (note square root scale).

The local wavelet power spectrum of aggregate data from Greater London is shown in Fig 2A and depicts a strong concentration of power within the 3.5–4.5 year band until the mid-1990s, after which no periodicity is apparent. A similar pattern is evident in local power

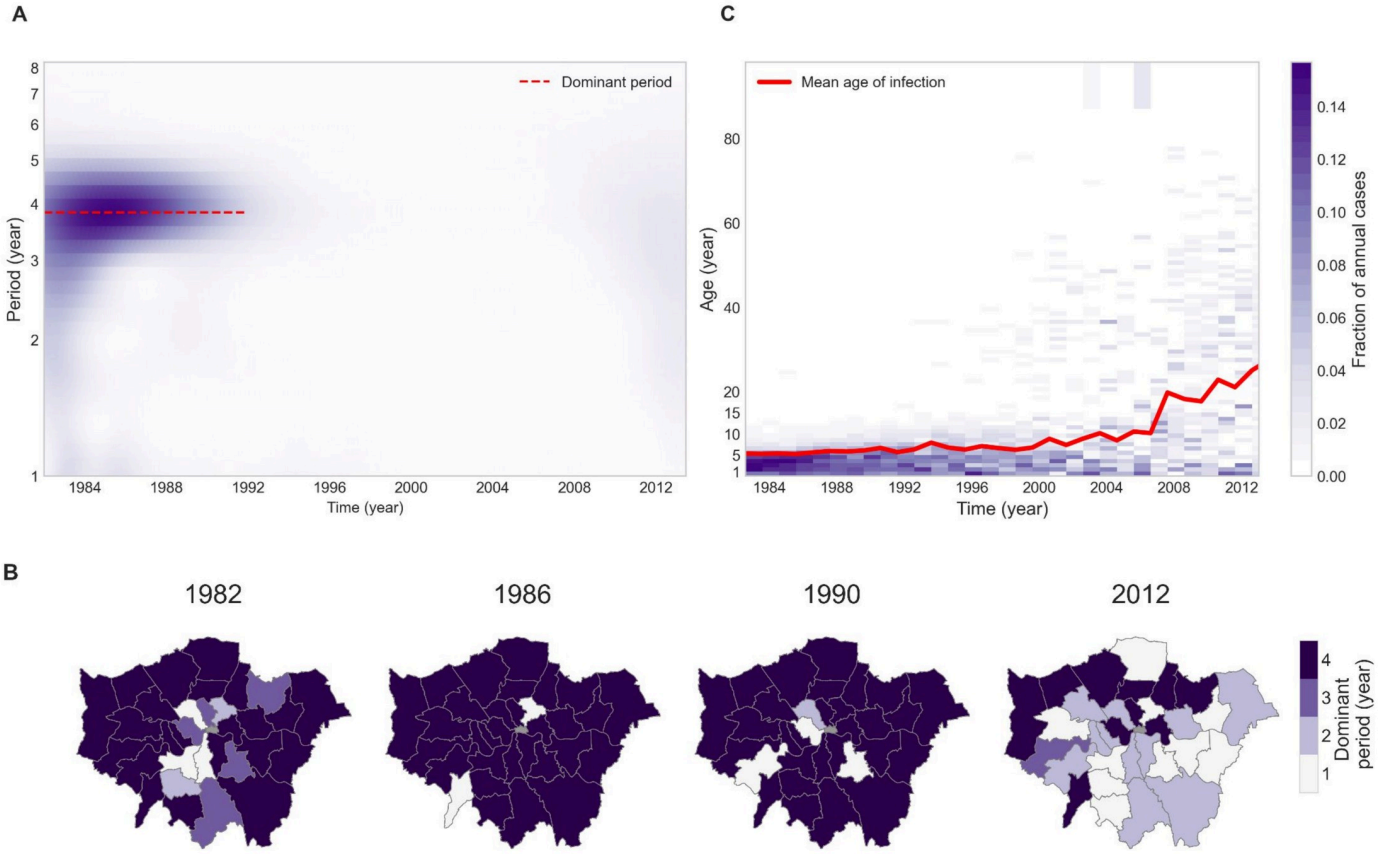

Fig 2. Changes in periodicity and age distribution of pertussis in Greater London and boroughs. (A) Local wavelet spectrum of cases time series of Greater London (B) Choropleth map of dominant period of the boroughs: In the first three epidemics (1982,1986,1990) most boroughs have dominant period of 3.5–4.5 years except a few with strong annual cycles. The epidemic cycles vanish by mid-1990s, and the 2012 resurgence does not exhibit as a periodic epidemic wave across the boroughs. Map base layers are obtained from London Datastore (https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london) and are available from https://data.london.gov.uk/download/statistical-gis-boundary-files-london/9ba8c833-6370-4b11-abdc-314aa020d5e0/statistical-gis-boundaries-london.zip (The digital boundary file contains Office for National statistics data Crown copyright and database (2012) and contains Ordnance Survey data Crown copyright and database (2012)). (C) Annual number of pertussis in Greater London, discretized by age. The red line depicts mean age of infection by year.

https://doi.org/10.1371/journal.pcbi.1009898.g002

spectrum of the boroughs (see S7 Fig). However, contrary to the trend in Greater London, fairly strong annual cycles are also observed in some of the boroughs (see regions shaded white in Fig 2A). We mapped the dominant period of the boroughs at the 1982, 1986 and 1990 epidemics in Fig 2B, which confirms that while most boroughs exhibit a strong quadriennial cycle, in some the annual component is stronger. Another notable observation in Fig 2B is the absence of any consistent periodicity across boroughs during the 2012 epidemic.

Increases in immunization coverage from late 1980s (Fig 1A) coincided with a notable shift in the age distribution of pertussis cases [45, 78, 79]. In Fig 2C, we depict the annual number of reported cases in Greater London discretized by age, illustrating that the burden of diseases is largely concentrated in 1–5 year-old children until 1988, after which it is restricted to very young infants and children aged 7 and older. Starting in the mid-1990s, a gradual increase in the proportion of adult cases is observed, followed by a sharp rise from 2006 onward. A similar trend has been previously reported for England & Wales [45], the US [80], and Canada [81]. This shift in the age distribution of cases to older age groups is reflected in the mean age of infection, which gradually increases from approximately 5 years in 1982 to $\sim$10 years by 2006; consistent with increasing national vaccine coverage rates over this period (Fig 1A). A striking

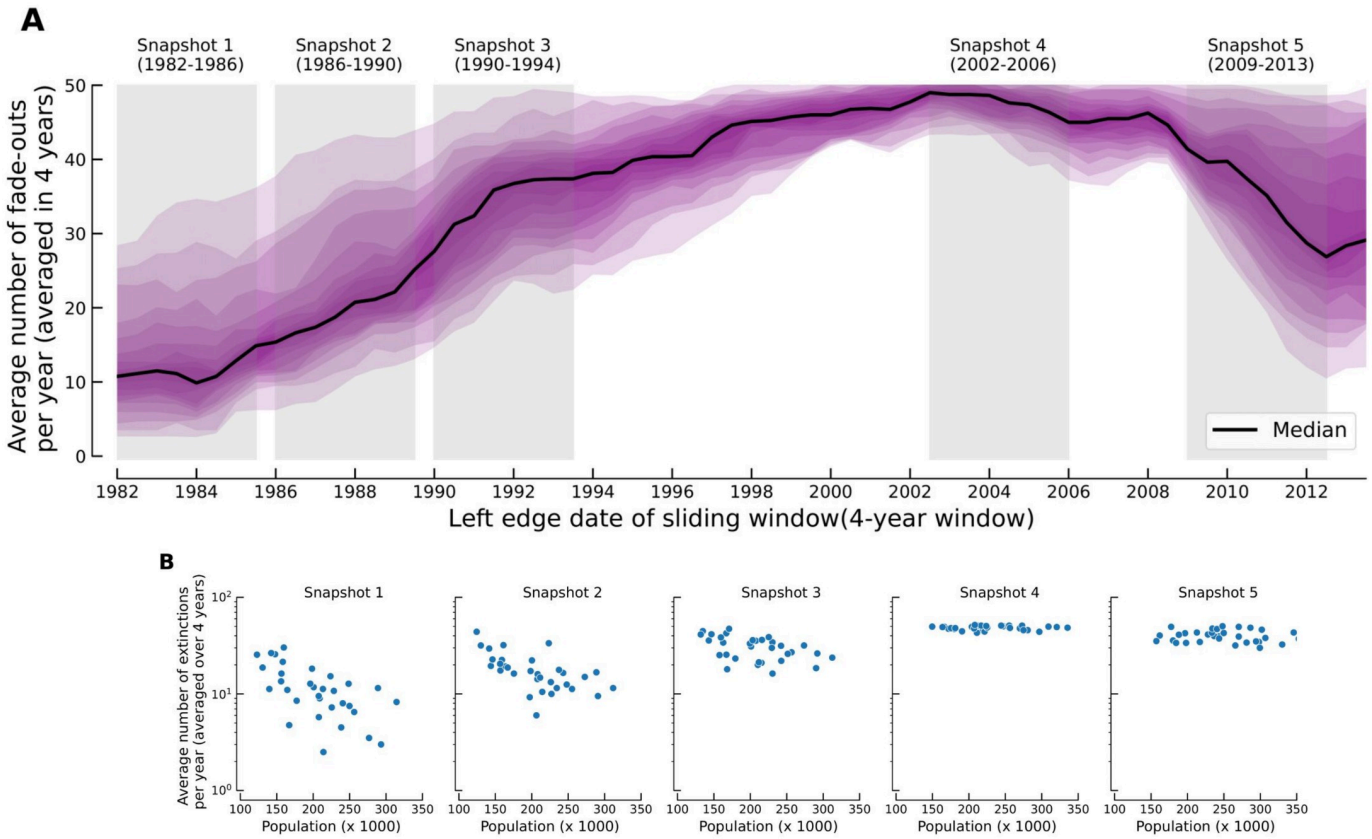

Fig 3. Fade-out profile of pertussis in Greater London and boroughs. (A) Distribution of number of fade-outs among boroughs of London. Shaded regions show the middle inter-percentile bands for different percentile values (inter-percentile range of 1% to 90%)(B) The number of fade-outs vs population size at four temporal snapshots (1982–1986, 1986–1990, 1990–1994, 2002–2006, and 2009–2013).

https://doi.org/10.1371/journal.pcbi.1009898.g003

observation from Fig 1C is the steep increase in the mean age of infection from 2006 onward, reaching 27-years old by 2013.

We also examined the fade-out frequency profile of pertussis over this period, as shown in Fig 3A. The black line shows the median number of fade-outs across Greater London and the purple shaded regions depict the middle inter-percentile bands for different percentile values. The figure illustrates that the number of fade-outs in all boroughs grows until $\sim$2002, which is consistent with the steady reduction in incidence shown in Fig 1A. However, the rate of growth in the number of fade-outs varies across boroughs. From 2002 onward, the number of fade-outs gradually declined until 2008, followed by a steep reduction leading up to the 2012 epidemic. Similar to 1980s, the rate of increase in the number of fade-outs during the resurgence varies significantly across boroughs. In Fig 3B, we depict the number of fade-outs at different time snapshots of Fig 3A as a function of the borough population size. In 1982, an exponential decline in the number of fade-outs with population size is evident. This relationship between fade-outs and population size becomes increasingly tenuous over time, ultimately disappearing in 2006.

Results of the epidemic time lag analysis are shown in Fig 4. The phase relationship among boroughs at the four epidemic snapshots is mapped in Fig 4A. Focusing on the early era (1982–1994), the most striking feature in this plot is the consistent spatial organization of boroughs across the three epidemics at 1982, 1986 and 1990. This is illustrated in Fig 4B, which

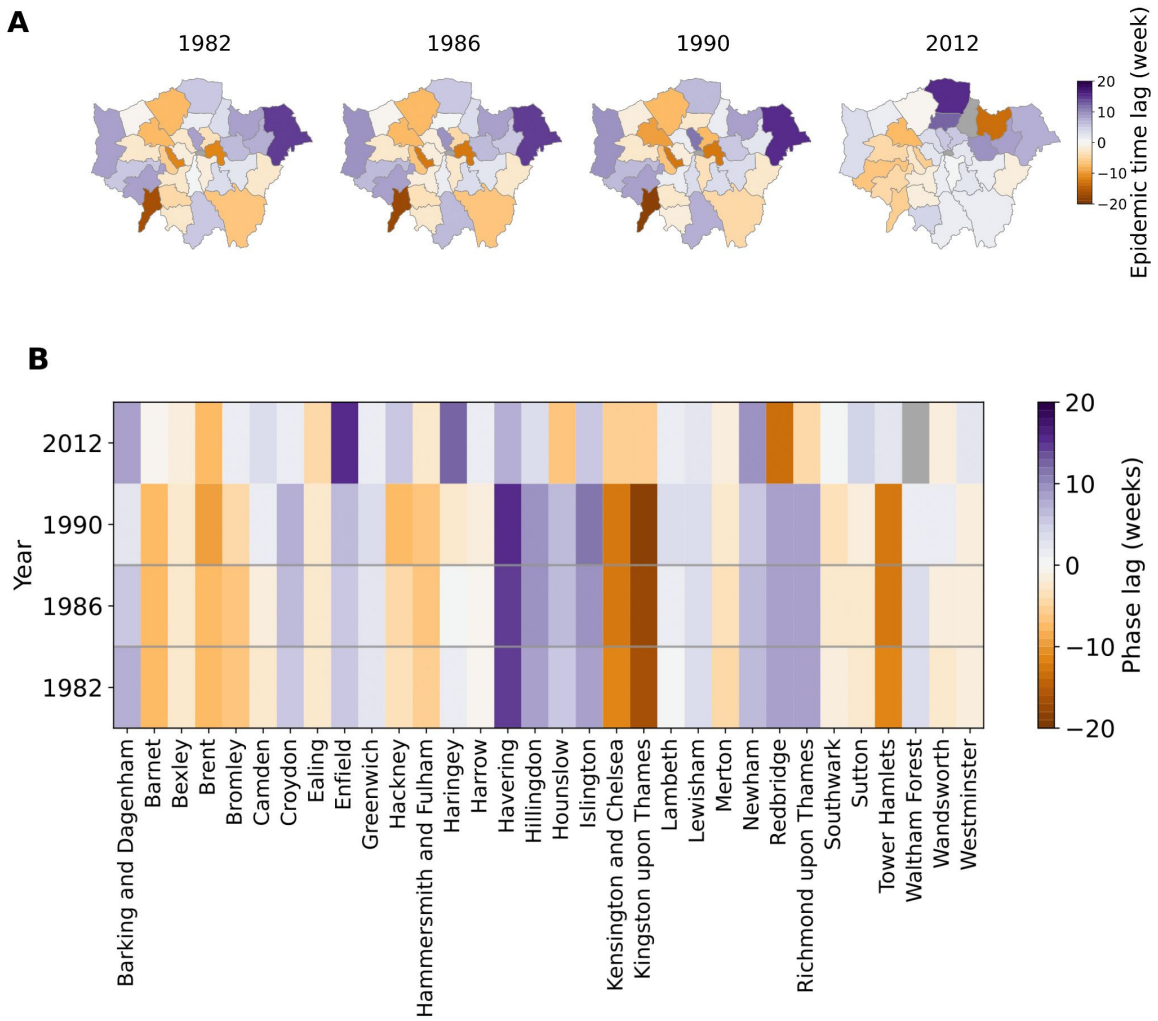

Fig 4. Pertussis epidemic phase lag among London boroughs in 1982, 1986, 1990 and 2012. Leading and lagging boroughs are shown in purple and orange respectively (A) Mapping of the phase lags at four different time snapshots. Map base layers are obtained from London Datastore (https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london) and are available from https://data.london.gov.uk/download/statistical-gis-boundary-files-london/9ba8c833-6370-4b11-abdc-314aa020d5e0/statistical-gis-boundaries-london.zip (The digital boundary file contains Office for National statistics data Crown copyright and database (2012) and contains Ordnance Survey data Crown copyright and database (2012)). (B) Phase lag across time, depicting a consistent spatial organization among the boroughs at the first three snapshots.

https://doi.org/10.1371/journal.pcbi.1009898.g004

shows the variation in borough phase lags across these three snapshots. Despite minor variations in relative phase lags in some boroughs, the spatial organization of the boroughs is consistent with each borough either consistently among the leaders of the epidemic wave, or lagging. The figure reveals four distinct foci (shaded purple) in the northeast, west, south, and the center of London that appear to lead the epidemic wave from 1982 to 1990. We found the spatial hierarchy of boroughs during the 2012 epidemic to be qualitatively different from the previous three snapshots. However, due to the sparsity of cases in some of the boroughs, the 2012 time lags must be interpreted with caution. The consistent spatial hierarchy of boroughs during the first three snapshots (1982, 1986 and 1990) prompted us to investigate the potential determinants of this spatial organization, by examining the statistical association between phase lag and demographic, socio-economic and geographical heterogeneity among boroughs.

Determinants of epidemic phase lag among boroughs

Here we describe the statistical analysis of phase lags against various demographic, socio-economic and geographical factors. Due to the omission of household composition variables from the 2011 census and the switch from Socio-economic group (SEG) classification to The National Statistics Socio-economic classification (NS-SEC) [82], we used the first three snapshots (1982,1986,1990) to perform multivariate analysis and conducted univariate analysis on 2012 snapshot.

1982–1990. The relative performance of the linear and FFNN regression models, as quantified by the coefficient of determination (R^2), is presented in S8 Fig. As shown in S8(A) Fig, OLS fails to generalize well to the test set and thus is overfitting. The lasso and ridge models perform slightly better in this respect, however, none of them outperforms the FFNN model, shown in S8(B) Fig. The FFNN model performed relatively well on all three data sets (train, validation and test). Comparing the coefficients of determination between the three datasets indicates that the model is slightly overfitted to the train dataset.

We used the trained FFNN model to make prediction for all the samples and estimated the local SHAP value for each sample to obtain a local feature importance for each feature at the corresponding sample. The local SHAP values quantify the linear association between each feature and the phase lag as the target variable at the corresponding sample point (see supplementary section 7 for more information on SHAP). After grouping the samples by year, we plotted the local SHAP values for each feature as shown in Fig 5. It is important to note that (i) a positive SHAP value indicates a positive association between that feature and a borough's phase lag, and (ii) boroughs with a positive phase lag lead the epidemic wave.

Overall, we find four distinct characteristics of a borough affect their relative phase position in London: household composition, socio-economic factors, transportation traits and latitude (Fig 5). Taking these in turn, we detect a notable difference in SHAP values among the three features representing household composition, specifically the *Number of households with children aged 0–4* has a positive relationship with phase in all three outbreak years, indicating an earlier epidemic take-off. In contrast, *Number of households with children age 5–16* is negatively associated with phase lag in the 1982 and 1986 outbreaks. We note that the *Number of households with no children* is positively associated with phase lags in 1986. *Households with* $>$ *1.5 PPR* also shows strong negative association, which counter-intuitively suggests that boroughs with less densely-populated household lead the epidemic wave. We also found socio-economic drivers of epidemic phase, namely *SEG 13,15* (quantifying the number of farmers and agriculture workers in a borough), which stands out as the feature with strongest positive association with phase in all three years. A weaker positive association is also observed between *SEG 7–10* (Skilled manual workers) and phase lag. The remaining features to highlight are *Travel public* (number of people who use public transportation to travel to work) and *latitude* which exhibit a positive association with phase, though the relationship with public transportation is only for the 1982 and 1990 outbreaks.

To dissect the relationship between the age composition of households and phase, we analyzed the age-specific seasonality in reported cases for the focal epidemic years of 1982, 1986 and 1990. In Fig 6, we present the within-year variation in reported cases for each age group and find that incidence is 1.7-fold higher among young children (0–5 years old), compared with 5–16 year-olds. While the broad patterns of seasonality across age are comparable, with troughs in May-July and a peak in September, the overall amplitude of seasonal variation in incidence—calculated as (max-min)/mean—is higher for the youngest age group (0.974 Vs 0.847). We note that the amplitude of seasonal variation is similarly higher in the adult age group (0.926) than 5–15 year-olds. Taken together, these findings suggest that boroughs with

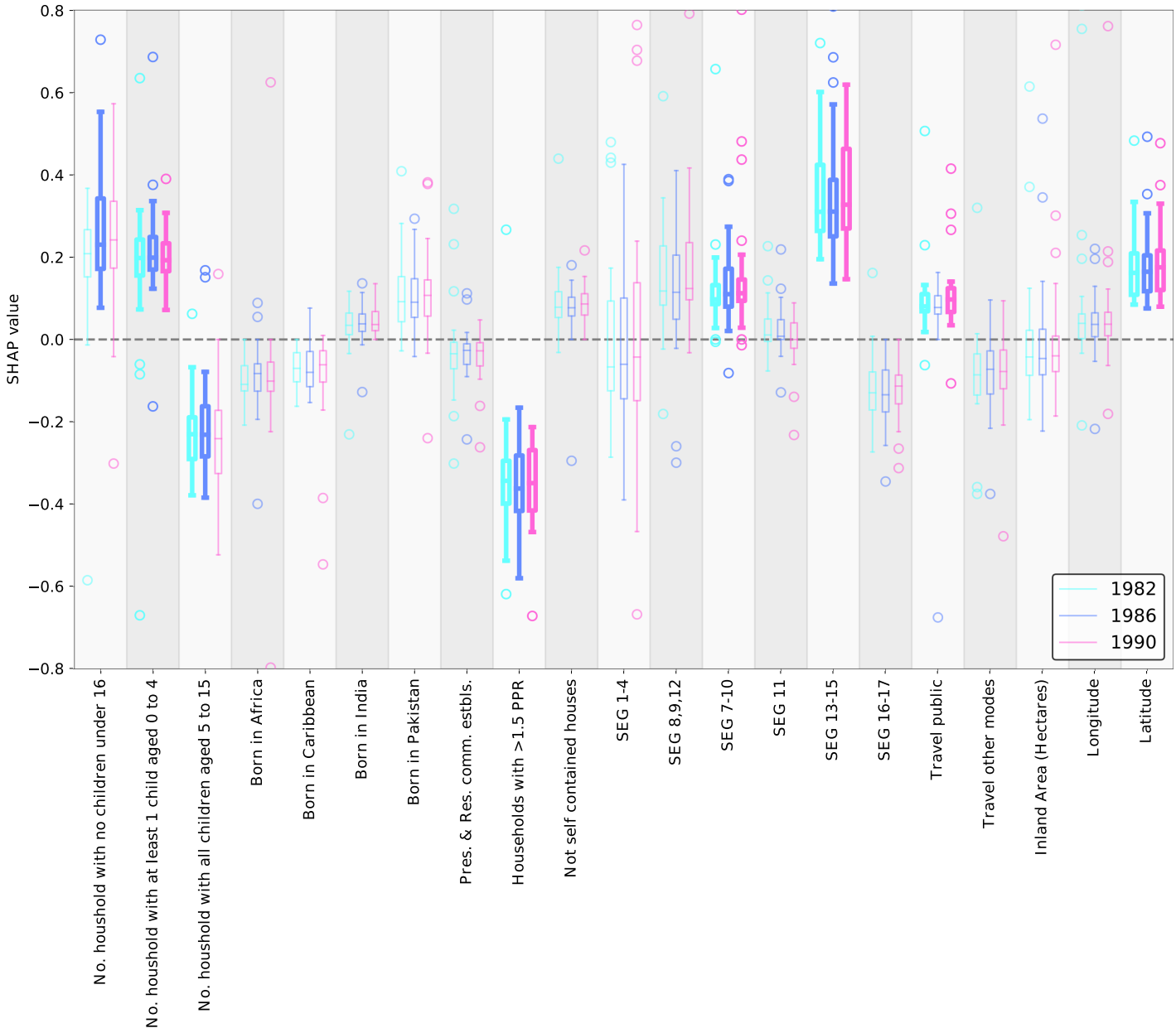

Fig 5. Aggregated SHAP values, representing the marginal contribution of each feature to predicted phase lags by the FFNN model. After fitting the FFNN model on the entire dataset, local SHAP values are obtained for each data point and aggregated over each time snapshot, distinguished with different colors. Values plotted in bold indicate statistical significance.

https://doi.org/10.1371/journal.pcbi.1009898.g005

many households composed of young children account for most intense transmission, experience the shallowest summer troughs and thus lead epidemics in London.

2012. The results of univariate regression analysis of 2012 epidemic is shown is S3 Table. Among the census variables, we found a moderate negative correlation between epidemic timing and *Born in Pakistan* ($r_{Pearson} = -0.4$, $P = 0.007$), i.e. boroughs with a greater Pakistan-born population lag the epidemics. We did not find any significant association between other census variables and epidemic phase.

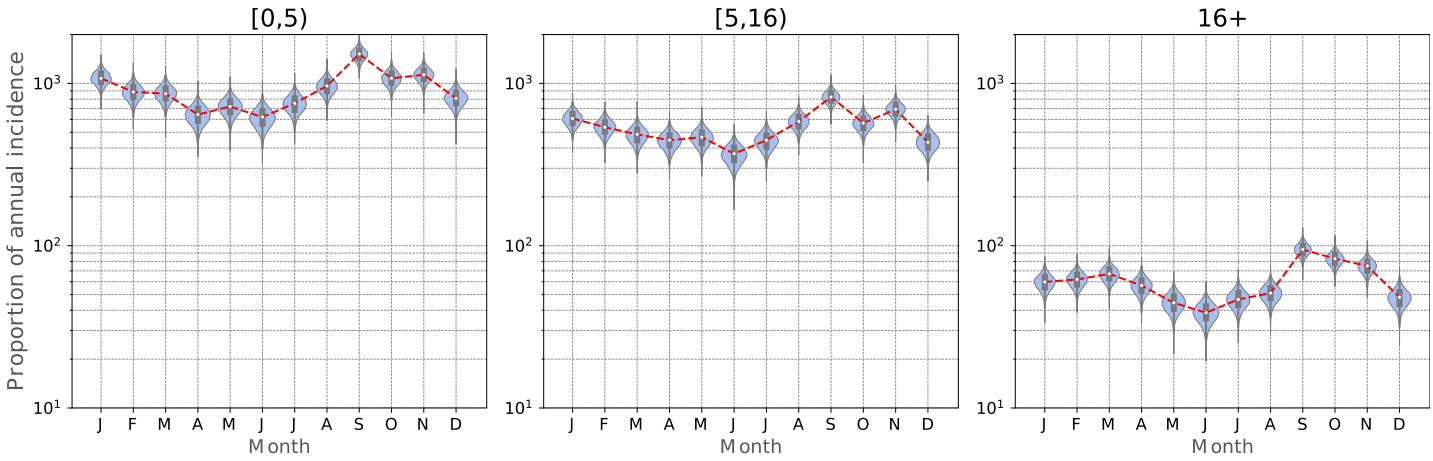

Fig 6. Age-specific seasonality in incidence (1982–1990). The figure demonstrates the year-to-year variation in the reported cases during each month for each age group. The distributions are generated from 10000 bootstrap samples of size 1000. The proportions for each age group is scaled by total cases in the corresponding age group. The relative incidence in each age group (scaled by the number of cases in the adult age group) is: 15.2, (11083 total cases), 8.88 (6474 total cases), and 1 (729 total cases), respectively. The amplitude of seasonality for each age group (calculated as (max-min)/average) is: 0.974, 0.847 and 0.926, respectively. Similarly, the standard deviation in intra-annual incidence for each age group is: 0.0228, 0.0196 and 0.0231.

https://doi.org/10.1371/journal.pcbi.1009898.g006

To examine the potential association between the decline in vaccine coverage rates and epidemic time lags, we used the annual birth and immunization data to calculate the cumulative number of unvaccinated children from 1997 to 2012 in each borough, and divided the total count by the population of each borough (as reported in census 2011) to find the *Ratio of cumulative unvaccinated children to total population*. As shown in S3 Table, we found a moderate positive correlation between this feature and phase lag ($r_{Pearson} = 0.4$, $P = 0.019$); that is, boroughs with greater ratio of unvaccinated children tend to lead the epidemic.

Discussion and conclusion

In this paper, we have examined the spatio-temporal dynamics of pertussis incidence in the boroughs of Greater London from 1982 to 2013. The dynamics of pertussis over this period can be characterized by three stages: a period of declining trends with four-year inter-epidemic cycles from 1982 to 1994, a deep trough until 2006, and the re-emergence of the disease leading to an outbreak in 2012. There was an increase in the estimated immunization coverage in Greater London from 55–60% in the early to mid-1980s to approximately 90% by the early 1990s (Fig 1A). Our study indicates that coincident with increasing vaccine coverage until mid-1990s there were predictable systematic trends in pertussis epidemiology, namely an overall decline in pertussis incidence (Fig 1A), smaller epidemic sizes (Fig 2A), a shift in the age distribution of cases towards older individuals (Fig 2B) and a rise in the frequency of fade-outs (Fig 3). The subsequent re-emergence stage is characterized by a steep decline in fade-out frequency (Fig 3) and a sharp increase in the mean age of infection (Fig 2B), however, we did not detect any periodicity in the wavelet spectrum of reported incidence (Fig 2A).

Various potential explanations have been offered in the literature for the recent re-emergence of pertussis in highly vaccinated populations such as England and Wales, including changes in the vaccine schedule and diagnostics [32], vaccine failure [36, 37, 83–87], waning of vaccine immunity [37, 85, 88–90], differences in the efficacy of acellular and whole cell vaccines [36, 37, 86], changes in the population contact structure [36, 45], the impact of asymptomatic transmission [89, 91–94] perhaps associated with the introduction of acellular

vaccines [37, 77], and failure in the degree of vaccine protection [32, 83, 95], possibly arising from vaccine-driven evolution [96, 97]. In addition, it is worth pointing out that there was a notable decreasing trend in the vaccine coverage rates across the boroughs of London in the years preceding the 2012 epidemic (Fig 1A), whose potential role in the re-emergence of pertussis requires further systematic investigation.

Despite the pronounced trends in pertussis incidence from 1982 to 1990, the relative ranking of the timing of outbreaks in boroughs, quantified through their relative phases (Fig 4A), did not exhibit substantial variation and instead revealed a reasonably consistent pattern of spatial organization (Fig 4B). To understand the mechanisms that may underpin the observed geographic motif, we assessed the association between borough-specific phases and a range of demographic, socio-economic and geographical factors using a combination of approaches, including linear regression and neural networks.

Linear regression models have been extensively used in the literature to investigate the drivers of epidemic asynchrony for various infectious disease systems [1, 3, 5]. These models are easy to implement and are inherently interpretable—that is, the relative importance of each variable for prediction is transparently quantified via the estimated coefficients. However, this convenience in modeling and interpretation comes at the expense of accuracy as these models, by definition, are incapable of capturing potentially nonlinear relationships between predictor variables and the target variable (S8 Fig). On the other hand, more complex models may provide improved prediction accuracy, but their results might not be readily interpretable. To overcome this hurdle, we adopted a novel methodology that has allowed consideration of possible nonlinearities in the model, while yielding straightforward interpretability between underlying demographic, socio-economic and geographical factors and the spatial organization of London boroughs. We have taken advantage of SHAP-based feature importance to quantify the relative contribution of each factor to model predictions. Our initial univariate regression analysis identified an association between 11 of the predictor variables and relative phase lags ($P < 0.05$) (See S1 Table and S9–S15 Figs). However, the subsequent SHAP-based analysis indicated that only a small subset of features contributed to predictive performance. This difference was not surprising given the high degree of correlation among the variables (as shown in S5 Fig).

The important features during the 1982–1990 epidemics, identified via SHAP analysis, were the age composition of households, agricultural workers, skilled manual workers, latitude, population of public transport commuters and high-occupancy households as key predictors of borough phase. We found the differences in age-specific seasonality of incidence as a key determinant of epidemic phase lag among the boroughs during this period. An earlier epidemic peak in September in children 0–5 years of age and adults causes the boroughs with higher population of these age groups to lead the epidemic wave. A second peak in November in adolescents can explain the epidemic lag in boroughs with more households with children 5–15 years old of age. A similar pattern in age-specific seasonality has been reported in a past study in Massachusetts [98, 99] and in the US overall [100]. During the 2012 epidemic, our univariate regression analysis indicated the size of Pakistan-born population to be negatively correlated with timing of epidemic. This may be indicative of some underlying correlation between infant immunization coverage and these socio-economic characteristics.

As mentioned previously, prior work leads us to suspect the outbreak phase of a borough to be impacted in part by susceptible recruitment, which in turn is affected by vaccine uptake [3]. However, we have been unable to obtain vaccine coverage at the borough level for the 1982–1990 period, with 1997 the first year for which these data can be obtained. Therefore, we examined the association between the vaccine coverage and epidemic phase in 1994 (borough-level incidence data post-1994 were too sparse to quantify phase lags). As shown in S17 Fig, no

significant association was observed between the two. On the other hand, we found a moderate positive association between *Ratio of cumulative unvaccinated children to total population*, a proxy for susceptible pool size, and the timing of 2012 epidemic across the boroughs.

Our study paints a different picture of the spatial ecology of an infectious disease at fine spatial scale. Past work on traveling waves of measles in England and Wales [1] uncovered a *source-sink* relationship between large cities serving as the source, leading biennial outbreaks and their satellite towns and villages acting as the sink, with lagging epidemics. In contrast, our study shows that at the borough level, epidemic foci located in the northeast, west, south, and center of London lead the 4-year epidemic wave of 1982–1990. In another study of the timing of seasonal rotavirus outbreaks across the US, it was shown that state-specific birth rates can explain epidemic timing [3]. We found incidence in infants and children 0–4 years of age as a determinant in timing of the epidemics spanning 1982–1990. Finally, a number of previous studies have argued for a *pacemaker mechanism* in driving epidemic cycles [5, 101–103], where a group of dynamically synchronous foci act as local rhythm setters. The identification of the aforementioned foci in the northeast, west, south, and center of London would be consistent with this explanation. Understanding the mechanisms that determine which populations assume the role of pacemaker remains an important area of future research.

More broadly, we report a transition from spatially organized waves in the decline phase of this period to a largely unstructured mosaic in the resurgence era. This observation is consistent with a previous study of spatial synchrony in pertussis epidemics in US states, where organized waves in the wP era gave way to spatially disorganized contemporary epidemics [5]. It is tempting to speculate regarding whether the switch to acellular vaccines for the routine schedule in 2004 [41] played a part in these transitions, especially given the concern over increased likelihood of asymptomatic infections in aP vaccinees [37, 77, 91]. This is an interesting possibility that warrants careful consideration as we strive to understand the factors that may have played a role in the transformation of the spatial epidemiology of pertussis in these populations in recent decades.

Supporting information

S1 Fig. Interpolation of demographic and socioeconomic variables from decennial censuses.
(PDF)

S2 Fig. Variation of demographic and socioeconomic variables across boroughs of London—1982.
(PDF)

S3 Fig. Variation of demographic and socioeconomic variables across boroughs of London—1986.
(PDF)

S4 Fig. Variation of demographic and socioeconomic variables across boroughs of London—1990.
(PDF)

S5 Fig. Correlation matrix of geographic and census variables.
(PDF)

S6 Fig. Preprocessing and transformation steps of case count time series to obtain phase lags.
(PDF)

S7 Fig. Evolution of pertussis incidence in the boroughs of Greater London exhibited in the frequency domain.
(PDF)

S8 Fig. Comparison of true phase lag vs predicted phase lag by different models.
(PDF)

S9 Fig. Univariate linear regression coefficients.
(PDF)

S10 Fig. Univariate regression, bivariate Gaussian kernel density contours 1.
(PDF)

S11 Fig. Univariate regression, bivariate Gaussian kernel density contours 2.
(PDF)

S12 Fig. Univariate regression, bivariate Gaussian kernel density contours 3.
(PDF)

S13 Fig. Univariate regression, bivariate Gaussian kernel density contours 4.
(PDF)

S14 Fig. Univariate regression, bivariate Gaussian kernel density contours 5.
(PDF)

S15 Fig. Univariate regression, bivariate Gaussian kernel density contours 6.
(PDF)

S16 Fig. Multivariate linear regression coefficients.
(PDF)

S17 Fig. Ranking of the London boroughs by phase lag at 1994 vs their ranking by the vaccine coverage in 1997 (Immunization rate at the first birthday).
(PDF)

S1 Table. Results of univariate regression between epidemic phase lag of London boroughs and census features, 1982–1990.
(PDF)

S2 Table. Results of multivariate regression between epidemic phase lag of London boroughs and census features, 1982–1990.
(PDF)

S3 Table. Results of univariate regression between epidemic phase lag of London boroughs and census features, 2012.
(PDF)

S1 Text. SHAP values.
(PDF)

Author Contributions

Conceptualization: Arash Saeidpour, Pejman Rohani.

Investigation: Arash Saeidpour.

Validation: Shweta Bansal, Pejman Rohani.

Writing – original draft: Arash Saeidpour, Pejman Rohani.

Writing – review & editing: Arash Saeidpour, Shweta Bansal, Pejman Rohani.

References

1. Grenfell BT, Bjørnstad ON, Kappey J. Travelling waves and spatial hierarchies in measles epidemics. Nature. 2001; 414(6865):716–723. https://doi.org/10.1038/414716a PMID: 11742391
2. Cummings DAT, Irizarry RA, Huang NE, Endy TP, Nisalak A, Ungchusak K, et al. Travelling waves in the occurrence of dengue haemorrhagic fever in Thailand. Nature. 2004; 427(6972):344–347. https://doi.org/10.1038/nature02225 PMID: 14737166
3. Pitzer VE, Viboud C, Simonsen L, Steiner C, Panozzo CA, Alonso WJ, et al. Demographic Variability, Vaccination, and the Spatiotemporal Dynamics of Rotavirus Epidemics. Science. 2009; 325 (5938):290–294. https://doi.org/10.1126/science.1172330 PMID: 19608910
4. Alonso WJ, Viboud C, Simonsen L, Hirano EW, Daufenbach LZ, Miller MA. Seasonality of influenza in Brazil: a traveling wave from the Amazon to the subtropics. American Journal of Epidemiology. 2007; 165:1434–1442. https://doi.org/10.1093/aje/kwm012 PMID: 17369609
5. Choisy M, Rohani P. Changing spatial epidemiology of pertussis in continental USA. Proceedings of the Royal Society B. 2012; 279:4574–4581. https://doi.org/10.1098/rspb.2012.1761 PMID: 23015623
6. Wesolowski A, Qureshi T, Boni MF, Sundsøy PR, Johansson MA, Rasheed SB, et al. Impact of human mobility on the emergence of dengue epidemics in Pakistan. Proceedings of the National Academy of Sciences of the United States of America. 2015; 112(38):11887–11892. https://doi.org/10.1073/pnas.1504964112 PMID: 26351662
7. Viboud C, Bjørnstad ON, Smith DL, Simonsen L, Miller MA, Grenfell BT. Synchrony, waves, and spatial hierarchies in the spread of influenza. Science. 2006; 312(5772):447–451. https://doi.org/10.1126/science.1125237 PMID: 16574822
8. Gog JR, Ballesteros S, Viboud C, Simonsen L, Bjørnstad ON, Shaman J, et al. Spatial Transmission of 2009 Pandemic Influenza in the US. PLoS Computational Biology. 2014; 10(6):e1003635. https://doi.org/10.1371/journal.pcbi.1003635 PMID: 24921923
9. Kramer AM, Pulliam JT, Alexander LW, Park AW, Rohani P, Drake JM. Spatial spread of the West Africa Ebola epidemic. Royal Society Open Science. 2016; 3(8):160294. https://doi.org/10.1098/rsos.160294 PMID: 27853607
10. Yang W, Zhang W, Kargbo D, Yang R, Chen Y, Chen Z, et al. Transmission network of the 2014–2015 Ebola epidemic in Sierra Leone. Journal of The Royal Society Interface. 2015; 106(112):6872–6877. https://doi.org/10.1098/rsif.2015.0536 PMID: 26559683
11. Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science. 2020; 368(6490):493–497. https://doi.org/10.1126/science.abb4218 PMID: 32213647
12. Omer SB, Enger KS, Moulton LH, Halsey NA, Stokley S, Salmon DA. Geographic clustering of nonmedical exemptions to school immunization requirements and associations with geographic clustering of pertussis. American Journal of Epidemiology. 2008; 168(12):1389–1396. https://doi.org/10.1093/aje/kwn263 PMID: 18922998
13. Phadke VK, Bednarczyk RA, Salmon DA, Omer SB. Association Between Vaccine Refusal and Vaccine-Preventable Diseases in the United States. Journal of the American Medical Association. 2016; 315(11):1149. https://doi.org/10.1001/jama.2016.1353 PMID: 26978210
14. Graham M, Winter AK, Ferrari M, Grenfell B, Moss WJ, Azman AS, et al. Measles and the canonical path to elimination. Science. 2019; 364(6440):584–587. https://doi.org/10.1126/science.aau6299 PMID: 31073065
15. Lau MSY, Becker AD, Korevaar HM, Caudron Q, Shaw DJ, Metcalf CJE, et al. A competing-risks model explains hierarchical spatial coupling of measles epidemics en route to national elimination. Nature Ecology & Evolution. 2020; 4(7):934–939. https://doi.org/10.1038/s41559-020-1186-6 PMID: 32341514
16. Rohani P, Scarpino SV. Pertussis: Epidemiology, Immunology & Evolution. Oxford: Oxford Univ Press; 2019.
17. Earn D, Rohani P, Grenfell B. Persistence, chaos and synchrony in ecology and epidemiology. Proceedings of the Royal Society B: Biological Sciences. 1998; 265(1390):1–4. https://doi.org/10.1098/rspb.1998.0256 PMID: 9470213

18. Rohani P, Earn DJ, Grenfell BT. Opposite patterns of synchrony in sympatric disease metapopulations. Science. 1999; 286(5441):968–971. https://doi.org/10.1126/science.286.5441.968 PMID: 10542154

19. Xia Y, Bjørnstad ON, Grenfell BT. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. The American naturalist. 2004; 164(2):267–81. https://doi.org/10.1086/422341 PMID: 15278849

20. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, et al. Quantifying the Impact of Human Mobility on Malaria. Science. 2012; 338(6104):267–270. https://doi.org/10.1126/science.1223467 PMID: 23066082

21. van Panhuis WG, Choisy M, Xiong X, Chok NS, Akarasewi P, Iamsirithaworn S, et al. Region-wide synchrony and traveling waves of dengue across eight countries in Southeast Asia. Proceedings of the National Academy of Sciences of the United States of America. 2015; p. 201501375. https://doi.org/10.1073/pnas.1501375112 PMID: 26438851

22. Martinez-Bakker M, King AA, Rohani P. Unraveling the Transmission Ecology of Polio. PLoS Biol. 2015; 13(6):e1002172. https://doi.org/10.1371/journal.pbio.1002172 PMID: 26090784

23. Pons-Salort M, Oberste MS, Pallansch MA, Abedi GR, Takahashi S, Grenfell BT, et al. The seasonality of nonpolio enteroviruses in the United States: Patterns and drivers. Proceedings of the National Academy of Sciences of the United States of America. 2018; Vol 2:201721159. https://doi.org/10.1073/pnas.1721159115 PMID: 29507246

24. Teurlai M, Huy R, Cazelles B, Duboz R, Baehr C, Vong S. Can Human Movements Explain Heterogeneous Propagation of Dengue Fever in Cambodia? PLoS Neglected Tropical Diseases. 2012; 6 (12):1–8. https://doi.org/10.1371/journal.pntd.0001957 PMID: 23236536

25. Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton University Press; 2008.

26. Browne CJ, Smith RJ, Bourouiba L. From regional pulse vaccination to global disease eradication: insights from a mathematical model of poliomyelitis. Journal of Mathematical Biology. 2015; 71 (1):215–253. https://doi.org/10.1007/s00285-014-0810-y PMID: 25074277

27. Russell CA, Real LA, Smith DL. Spatial Control of Rabies on Heterogeneous Landscapes. PLoS ONE. 2006; 1(1):e27. https://doi.org/10.1371/journal.pone.0000027 PMID: 17183654

28. Eisinger D, Thulke HH. Spatial pattern formation facilitates eradication of infectious diseases. The Journal of Applied Ecology. 2008; 45(2):415. https://doi.org/10.1111/j.1365-2664.2007.01439.x PMID: 18784795

29. Stewart GT. Pertussis Vaccine: the United Kingdom's Experience. International Symposium on Pertussis. U.S. Department of Health, Education, and Welfare Publication; 1978. p. 262–282.

30. Preston N. Pertussis today. In: Wardlaw AC, Parton R, editors. Pathogenesis and Immunity in Pertussis. John Wiley & Sons. John Wiley & Sons; 1988.

31. Clark TA, Messonnier NE, Hadler SC. Pertussis control: time for something new? Trends in Microbiology. 2012; 20(5):211—213. https://doi.org/10.1016/j.tim.2012.03.003 PMID: 22494804

32. Cherry JD. Pertussis: Challenges Today and for the Future. PLoS Pathogens. 2013; 9(7):e1003418. https://doi.org/10.1371/journal.ppat.1003418 PMID: 23935481

33. Jackson DW, Rohani P. Perplexities of pertussis: recent global epidemiological trends and their potential causes. Epidemiology and Infection. 2014; 142:672—684. https://doi.org/10.1017/S0950268812003093 PMID: 23324361

34. Riolo MA, King AA, Rohani P. Can vaccine legacy explain the British pertussis resurgence? Vaccine. 2013; 31(49):5903–5908. https://doi.org/10.1016/j.vaccine.2013.09.020 PMID: 24139837

35. Domenech de Cellès M, Magpantay FMG, King AA, Rohani P. The pertussis enigma: reconciling epidemiology, immunology and evolution. Proceedings of the Royal Society of London—Biological sciences. 2016; 283(1822):20152309. https://doi.org/10.1098/rspb.2015.2309 PMID: 26763701

36. Choi YH, Campbell H, Amirthalingam G, Van Hoek AJ, Miller E. Investigating the pertussis resurgence in England and Wales, and options for future control. BMC Medicine. 2016; 14:1–11. https://doi.org/10.1186/s12916-016-0665-8 PMID: 27580649

37. Domenech de Cellès M, Magpantay FMG, King AA, Rohani P. The impact of past vaccination coverage and immunity on pertussis resurgence. Science Translational Medicine. 2018; 10(434).

38. Blackwood JC, Cummings DAT, Broutin H, Iamsirithaworn S, Rohani P. The population ecology of infectious diseases: pertussis in Thailand as a case study. Parasitology. 2012; 139(14):1888–1898. https://doi.org/10.1017/S0031182012000431 PMID: 22717183

39. Magpantay FMG, Rohani P. Dynamics of Pertussis Transmission in the United States. American Journal of Epidemiology. 2015; 181(12):921–931. https://doi.org/10.1093/aje/kwv024 PMID: 26022662

40. Rohani P, Drake JM. The decline and resurgence of pertussis in the US. Epidemics. 2011; 3(3-4):183–188. https://doi.org/10.1016/j.epidem.2011.10.001 PMID: 22094341

41. Campbell H, Amirthalingam G, Andrews N, Fry NK, George RC, Harrison TG, et al. Accelerating control of pertussis in England and Wales. Emerging Infectious Diseases. 2012; 18(1):38—47. https://doi.org/10.3201/eid1801.110784 PMID: 22260989

42. Amirthalingam G, Gupta S, Campbell H. Pertussis immunisation and control in England and Wales, 1957 to 2012: a historical review. Eurosurveillance. 2013; 18(38):20587. https://doi.org/10.2807/1560-7917.ES2013.18.38.20587 PMID: 24084340

43. Campbell PT, McCaw JM, McIntyre P, McVernon J. Defining long-term drivers of pertussis resurgence, and optimal vaccine control strategies. Vaccine. 2015; 33(43):5794–5800. https://doi.org/10.1016/j.vaccine.2015.09.025 PMID: 26392008

44. Choi YH, Campbell H, Amirthalingam G, van Hoek AJ, Miller E. Investigating the pertussis resurgence in England and Wales, and options for future control. BMC Medicine. 2016; 14(1):121. https://doi.org/10.1186/s12916-016-0665-8 PMID: 27580649

45. Bento AI, Riolo MA, Choi YH, King AA, Rohani P. Core pertussis transmission groups in England and Wales: A tale of two eras. Vaccine. 2018; 36(9):1160–1166. https://doi.org/10.1016/j.vaccine.2018.01.046 PMID: 29395520

46. Pertussis Guidelines Group. Guidelines for the Public Health Management of Pertussis in England; 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/762766/Guidelines_for_the_Public_Health_management_of_Pertussis_in_England.pdf.

47. Bolker B, Grenfell B. Space, Persistence and Dynamics of Measles Epidemics. Philosophical Transactions of the Royal Society—Biological Sciences. 1995; 348(1325):309–320. https://doi.org/10.1098/rstb.1995.0070 PMID: 8577828

48. Cazelles B, Chavez M, McMichael AJ, Hales S. Nonstationary Influence of El Niño on the Synchronous Dengue Epidemics in Thailand. PLoS Medicine. 2005; 2(4):e106. https://doi.org/10.1371/journal.pmed.0020106 PMID: 15839751

49. Merler S, Ajelli M. The role of population heterogeneity and human mobility in the spread of pandemic influenza. Proceedings Biological sciences. 2010; 277(1681):557–65. https://doi.org/10.1098/rspb.2009.1605 PMID: 19864279

50. Van Buynder PG, Owen D, Vurdien JE, Andrews NJ, Matthews RC, Miller E. Bordetella pertussis surveillance in England and Wales: 1995-7. Epidemiology and Infection. 1999; 123(3):403–411. https://doi.org/10.1017/s0950268899003052 PMID: 10694150

51. De Greeff SC, Mooi FR, Westerhof A, Verbakel JMM, Peeters MF, Heuvelman CJ, et al. Pertussis Disease Burden in the Household: How to Protect Young Infants. Clinical infectious diseases. 2010; 50(10):1339–1345. https://doi.org/10.1086/652281 PMID: 20370464

52. Census definitions;. Available from: https://census.ukdataservice.ac.uk/use-data/censuses/definitions.aspx.

53. Earn D, Rohani P, Bolker B, Grenfell B. A simple model for complex dynamical transitions in epidemics. Science. 2000; 287(5453):667. https://doi.org/10.1126/science.287.5453.667 PMID: 10650003

54. Nguyen HTH, Rohani P. Noise, nonlinearity and seasonality: the epidemics of whooping cough revisited. Journal of the Royal Society, Interface. 2008; 5(21):403–413. https://doi.org/10.1098/rsif.2007.1168 PMID: 17878136

55. Broutin H, Viboud C, Grenfell BT, Miller MA, Rohani P. Impact of vaccination and birth rate on the epidemiology of pertussis: a comparative study in 64 countries. Proceedings of the Royal Society B: Biological Sciences. 2010; p. 1–7. https://doi.org/10.1098/rspb.2010.0994 PMID: 20534609

56. de Figueiredo A, Johnston IG, Smith DM, Agarwal S, Larson HJ, Jones NS. Forecasted trends in vaccination coverage and correlations with socioeconomic factors: a global time-series analysis over 30 years. The Lancet Global Health. 2016; 4(10):e726–e735. https://doi.org/10.1016/S2214-109X(16)30167-X PMID: 27569362

57. Annual Vaccine Coverage Statistics: England;. Available from: https://webarchive.nationalarchives.gov.uk/20140714092250/http://www.hpa.org.uk/webw/HPAweb&HPAwebStandard/HPAweb_C/1195733783627.

58. Historical Census Tables;. Available from: https://data.london.gov.uk/dataset/historical-census-tables.

59. Land Area and Population Density, Ward and Borough;. Available from: https://data.london.gov.uk/dataset/land-area-and-population-density-ward-and-borough.

60. Population weighted centroids;. Available from: https://geoportal.statistics.gov.uk/datasets/output-areas-december-2011-population-weighted-centroids-1/explore.

61. Torrence C, Compo GP. A Practical Guide to Wavelet Analysis. Bulletin of the American Meteorological Society. 1998; 79:61–78. https://doi.org/10.1175/1520-0477(1998)079%3C0061:APGTWA%3E2.0.CO;2

62. Le Van Quyen M, Foucher J, Lachaux JP, Rodriguez E, Lutz A, Martinerie J, et al. Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony; 2001.

63. Parzen E. On estimation of a probability density function and mode. The annals of mathematical statistics. 1962; 33(3):1065–1076. https://doi.org/10.1214/aoms/1177704472

64. Rohani P, Earn D, Grenfell B. Impact of immunisation on pertussis transmission in England and Wales. The Lancet. 2000; 355(9200):285–286. https://doi.org/10.1016/S0140-6736(99)04482-7 PMID: 10675078

65. Broutin H, Mantilla-Beniers N, Simondon F, Aaby P, Grenfell B, Guégan J, et al. Epidemiological impact of vaccination on the dynamics of two childhood diseases in rural Senegal. American Journal of Epidemiology. 2005; 7(4):593–599. PMID: 15820150

66. Conlan AJK, Rohani P, Lloyd AL, Keeling M, Grenfell BT. Resolving the impact of waiting time distributions on the persistence of measles. Journal of the Royal Society, Interface. 2010; 7(45):623–640. https://doi.org/10.1098/rsif.2009.0284 PMID: 19793743

67. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. vol. 1. Springer series in statistics Springer, Berlin; 2001.

68. Chollet F, et al. Keras; 2015. Available from: https://github.com/fchollet/keras.

69. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems; 2015. Available from: http://download.tensorflow.org/paper/whitepaper2015.pdf.

70. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research. 2014; 15(56):1929–1958.

71. Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.

72. Lundberg SM, Allen PG, Lee SI. A Unified Approach to Interpreting Model Predictions. In: 31st Conference on Neural Information Processing Systems (NIPS 2017); 2017.

73. Lundberg SM, Nair B, Vavilala MS, Horibe M, Eisses MJ, Adams T, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. Nature Biomedical Engineering. 2018; 2(10):749–760. https://doi.org/10.1038/s41551-018-0304-0 PMID: 31001455

74. Ribeiro MT, Singh S, Guestrin C. Model-Agnostic Interpretability of Machine Learning. In: ICML Workshop on Human Interpretability in Machine Learning; 2016.

75. Shapley LS. In: Roth AEE, editor. A value for n-person games. Cambridge University Press; 1988. p. 31–40.

76. Shapiro ED. Acellular vaccines and resurgence of pertussis. JAMA. 2012; 308(20):2149–2150. https://doi.org/10.1001/jama.2012.65031 PMID: 23188034

77. Althouse BM, Scarpino SV. Asymptomatic transmission and the resurgence of Bordetella pertussis. BMC Medicine. 2015; 13(1):146. https://doi.org/10.1186/s12916-015-0382-8 PMID: 26103968

78. Rohani P, Scarpino SV. Introduction to pertussis transmission and epidemiological dynamics. In: Rohani P, Scarpino SV, editors. Pertussis: Epidemiology, Immunology, and Evolution. Oxford: Oxford University Press; 2019.

79. Rohani P, Zhong X, King AA. Contact network structure explains the changing epidemiology of pertussis. Science. 2010; 330(6006):982–985. https://doi.org/10.1126/science.1194134 PMID: 21071671

80. Zanardi L, Pascual MB, Bisgard K, Murphy T, Wharton M. Pertussis—United States, 1997–2000. MMWR Morb Mortal Wkly Rep. 2002; 51:73–76.

81. Skowronski DM, De Serres G, MacDonald D, Wu W, Shaw C, Macnabb J, et al. The changing age and seasonal profile of pertussis in Canada. The Journal of Infectious Diseases. 2002; 185(10):1448–1453. https://doi.org/10.1086/340280 PMID: 11992280

82. The National Statistics Socio-economic classification (NS-SEC);. Available from: https://www.ons.gov.uk/methodology/classificationsandstandards/otherclassifications/thenationalstatisticssocioeconomicclassificationnssecrebasedonsoc2010.

83. Halloran ME, Longini IM, Struchiner CJ. Design and analysis of vaccine studies. Springer Verlag. Springer Verlag; 2010.

84. Fine PEM, Clarkson J. The Recurrence of Whooping Cough: Possible Implications for Assesment of Vaccine Efficacy. The Lancet. 1982; 319(8273):666–669. https://doi.org/10.1016/S0140-6736(82)92214-0

85. Blackwood JC, Cummings DAT, Broutin H, Iamsirithaworn S, Rohani P. Deciphering the impacts of vaccination and immunity on pertussis epidemiology in Thailand. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110(23):9595—9600. https://doi.org/10.1073/pnas.1220908110 PMID: 23690587

86. Gambhir M, Clark TA, Cauchemez S, Tartof SY, Swerdlow DL, Ferguson NM. A Change in Vaccine Efficacy and Duration of Protection Explains Recent Rises in Pertussis Incidence in the United States. PLoS Computational Biology. 2015; 11(4):e1004138. https://doi.org/10.1371/journal.pcbi.1004138 PMID: 25906150

87. Magpantay FMG, Domenech de Cellès M, Rohani P, King AA. Pertussis immunity and epidemiology: mode and duration of vaccine-induced immunity. Parasitology. 2016; 143(07):835—849. https://doi.org/10.1017/S0031182015000979 PMID: 26337864

88. Heininger U, Cherry JD, Stehr K. Serologic Response and Antibody-Titer Decay in Adults with Pertussis. Clinical Infectious Diseases. 2004; 38(4):591–594. https://doi.org/10.1086/381439 PMID: 14765356

89. Wearing HJ, Rohani P. Estimating the Duration of Pertussis Immunity Using Epidemiological Signatures. PLoS Pathogens. 2009; 5(10):e1000647. https://doi.org/10.1371/journal.ppat.1000647 PMID: 19876392

90. Lavine JS, Rohani P. Resolving pertussis immunity and vaccine effectiveness using incidence time series. Expert Review of Vaccines. 2012; 11(11):1319—1329. https://doi.org/10.1586/erv.12.109 PMID: 23249232

91. Warfel JM, Merkel TJ. Bordetella pertussis infection induces a mucosal IL-17 response and long-lived Th17 and Th1 immune memory cells in nonhuman primates. Mucosal Immunity. 2012; p. 1—10. PMID: 23187316

92. de Graaf H, Ibrahim M, Hill AR, Gbesemete D, Vaughan AT, Gorringe A, et al. Controlled human infection with Bordetella pertussis induces asymptomatic, immunising colonisation. Clinical infectious diseases. 2019.

93. Craig R, Kunkel E, Crowcroft NS, Fitzpatrick MC, Melker Hd, Althouse BM, et al. Asymptomatic Infection and Transmission of Pertussis in Households: A Systematic Review. Clinical infectious diseases. 2020; 70(1):152—161. https://doi.org/10.1093/cid/ciz531 PMID: 31257450

94. Gill CJ, Gunning CE, MacLeod WB, Mwananyanda L, Thea DM, Pieciak RC, et al. Asymptomatic Bordetella pertussis infections in a longitudinal cohort of young African infants and their mothers. eLife. 2021; 10:e65663. https://doi.org/10.7554/eLife.65663 PMID: 34097599

95. Magpantay FMG, Riolo MA, Domenech de Cellès M, King AA, Rohani P. Epidemiological Consequences of Imperfect Vaccines for Immunizing Infections. SIAM Journal on Applied Mathematics. 2014; 74(6):1810—1830. https://doi.org/10.1137/140956695 PMID: 25878365

96. Kallonen T, He Q. Bordetella pertussis strain variation and evolution postvaccination. Expert Review of Vaccines. 2009; 8(7):863—875. https://doi.org/10.1586/erv.09.46 PMID: 19538113

97. Jayasundara D, Lee E, Octavia S, Lan R, Tanaka MM, Wood JG. Emergence of pertactin-deficient pertussis strains in Australia can be explained by models of vaccine escape. Epidemics. 2020; 31:100388. https://doi.org/10.1016/j.epidem.2020.100388 PMID: 32387895

98. Lavine J, Broutin H, Harvill ET, Bjørnstad ON. Imperfect vaccine-induced immunity and whooping cough transmission to infants. Vaccine. 2010; 29(1):11–16. https://doi.org/10.1016/j.vaccine.2010.10.029 PMID: 21034823

99. King AA, Domenech de Cellès M, Magpantay FMG, Rohani P. Pertussis immunity and the epidemiological impact of adult transmission: Statistical evidence from Sweden and Massachusetts. In: Rohani P, Scarpino SV, editors. Pertussis: Epidemiology, Immunology, and Evolution. Oxford: Oxford University Press; 2019.

100. Tanaka M, Vitek CR, Pascual FB, Bisgard KM, Tate JE, Murphy TV. Trends in pertussis among infants in the United States, 1980-1999. Journal of the American Medical Association. 2003; 290(22):2968–2975. https://doi.org/10.1001/jama.290.22.2968 PMID: 14665658

101. Sasaki A, Hamilton W, Ubeda F. Clone mixtures and a pacemaker: new facets of Red-Queen theory and ecology. Proceedings of the Royal Society of London—Biological Sciences. 2002; 269 (1493):761–772. https://doi.org/10.1098/rspb.2001.1837

102. Litvak-Hinenzon A, Stone L. Spatio-temporal waves and targeted vaccination in recurrent epidemic network models. Journal of the Royal Society Interface. 2009; 6(38):749–760. https://doi.org/10.1098/rsif.2008.0343 PMID: 18957362

103. Rohani P, Lewis TJ, Grünbaum D, Ruxton GD. Spatial self-organisation in ecology: pretty patterns or robust reality? Trends in Ecology & Evolution. 1997; 12(2):70–74. https://doi.org/10.1016/S0169-5347(96)20103-X PMID: 21237982