



Long-term changes in endemic threshold populations for pertussis in England and Wales: A spatiotemporal analysis of Lancashire and South Wales, 1940-69

Alastair D. Munro^{*}, Matthew Smallman-Raynor, Adam C. Algar

School of Geography, University of Nottingham, University Park, Nottingham, NG7 2RD, UK

ARTICLE INFO

Keywords:

England and Wales
Endemic threshold
Disease persistence
Childhood infections
Geographical connectivity
Pertussis

ABSTRACT

Metapopulation dynamics play a critical role in driving endemic persistence and transmission of childhood infections. The endemic threshold concept, also referred to as critical community size (CCS), is a key example and is defined as the minimum population size required to sustain a continuous chain of infection transmission. The concept is fundamental to the implementation of effective vaccine-based disease control programmes. Vaccination serves to increase endemic threshold population size, promoting disease fadeout and eventual elimination of infection. To date, empirical investigations of the relationship between vaccination and endemic threshold population size have tended to focus on isolated populations in island communities. Very few studies have examined endemic threshold dynamics in 'mainland' regional populations with complex hierarchical spatial structures and varying levels of connectivity between subpopulations. The present paper provides the first spatially explicit analysis of the temporal changes in endemic threshold populations for one vaccine-preventable childhood infection (pertussis) in two dynamic regions of England and Wales: Lancashire and South Wales. Drawing upon weekly disease records of the Registrar-General of England and Wales over a 30-year period (January 1940–December 1969) regression techniques were used to estimate the endemic threshold size for pertussis in the two study regions. Survival analyses were performed to compare disease fadeout duration and probability for both regions in the pre-vaccine and vaccine eras, respectively. Our findings reveal the introduction of mass vaccination led to a considerable increase in threshold size for both Lancashire (~387,333) and South Wales (~1,460,667). Significant growth in fadeout duration was observed in the vaccine era for pertussis non-hotspots in both regions, consistent with geographical synchronisation of epidemic activity. Regional differences in endemic threshold populations reflect significant regional variations in spatial connectivity, population dispersion and level of geographical isolation.

1. Introduction

The endemicity of childhood infections continues to be an area of considerable interest for a broad spectrum of disciplines engaged with the study of disease persistence. Determining the endemic nature of infections can make a valuable contribution to global disease eradication programmes. For example, identifying urban centres and regions which facilitate disease persistence by acting as 'permanent disease reservoirs' (Cliff et al., 2000: 85) is key to enabling effective strategies to be developed which successfully eliminate infection. One concept which holds profound implications for the persistence of an infection is that of the endemicity threshold. This concept has previously received attention

in the work of geographers who have attempted to elucidate the spatial structure and geographical spread of childhood infections (Cliff et al., 1992, 1993; 2000).

The endemic threshold (sometimes referred to as 'Critical Community Size', CCS) is the minimum population required for an infection to persist endemically within a geographical area without reintroduction from external sources (Schenzle and Dietz, 1987; see glossary in Table 1). Bartlett (1956) was the first to observe this phenomenon, finding the time to extinction for measles was an increasing function of population size; settlements above the 'critical size' of 250,000–300,000 experienced no disease fadeout. The endemic threshold is a stochastic threshold for which fluctuations in population size, migration and birth

^{*} Corresponding author. School of Geography, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom.

E-mail address: alastair.munro@nottingham.ac.uk (A.D. Munro).

<https://doi.org/10.1016/j.socscimed.2020.113295>

Received in revised form 23 July 2020; Accepted 12 August 2020

Available online 22 August 2020

0277-9536/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Table 1
Glossary of key terms.

Key term	Definition
Endemic threshold	Minimum population size required for an infection to persist within a geographical area without fadeout.
Fadeout	A period of three or more weeks with no reported cases of infection within a population.
Metapopulation	A group of loosely coupled, spatially separated local populations.
Rescue effect	Transmission of infection from an external source which prevents elimination of disease in a local population.

rates have repeatedly been demonstrated to play a fundamental role (Bartlett, 1957; Earn et al., 1998; Metcalf et al., 2013).

As Bolker and Grenfell (1996) observe, endemic threshold size is inherently geographical and is ultimately dependent on the spatial structure and connectedness of a regional population. The persistence of infection based on the stochastic models formulated by Bartlett (1957, 1960), and supported by subsequent work (Black, 1966; Cliff et al., 1993; Grenfell et al., 2001), strongly implies a spatial transmission of infection between geographical units as the population size of cities and towns fall below the endemic threshold. Hierarchical travelling waves spread from large urban centres that act as ‘endemic reservoirs’, maintaining the circulation of disease and re-infecting regions where infection has either been locally eliminated or has faded out (Cliff and Haggett, 1989). In metapopulation terms (see glossary, Table 1), urban centres act as ‘core’ patches, where the large, dense population maintains circulation of infection, which spreads outwards to surrounding ‘satellite’ patches (Bolker and Grenfell, 1996; Grenfell and Harwood, 1997). For example, Grenfell et al. (2001) use metapopulation models to analyse the spatiotemporal persistence of measles across all districts in England and Wales as a function of population size. Their findings support the conclusions of Bartlett (1957, 1960), revealing that measles persisted in urban settlements with populations above 300,000 during the pre-vaccine era and with prolonged periods of fadeout in small communities. Grenfell et al.’s (2001) work builds upon the ‘cities and villages’ model, outlining a clear urban/rural hierarchy in disease persistence (Anderson and May 1992). It also expands upon the geographical work of Cliff et al. (1992, 1993), whose study of the spatial dynamics of measles in the United States indicated that infection diffuses from major cities to settlements in surrounding rural areas.

Islands have received considerable attention as settings in which to analyse the endemic threshold concept (Black, 1966; Cliff et al., 2000; Metcalf et al., 2013). For example, Cliff et al.’s (2000) research at the meso-geographical level in Iceland provides a methodological template for examining long-term temporal changes in endemic threshold populations. They calculated an average endemic threshold estimate for pertussis in Iceland of approximately 106,000 over a 100 year-long study period (1888–1987). The highest endemic threshold estimates were detected at the beginning and end of the study period, reflecting the relative internal isolation of the Icelandic population and the impact of mass vaccination. Previous empirical research has repeatedly demonstrated that mass vaccination serves to increase endemic threshold size (Anderson, 2016) by disrupting chains of disease transmission within a population; larger populations are required to prevent the local elimination of the disease (Lavine and Rohani, 2012). Past work has also demonstrated the positive impact of vaccination on reducing disease persistence by observing changes in the frequency and duration of fadeouts (Bolker and Grenfell, 1996; Lavine; Rohani, 2012). For instance, Rohani et al. (1999) demonstrate an increase in the number of fadeouts and duration of each individual fadeout associated with the introduction of the pertussis vaccine in England and Wales.

Beyond studies of endemic thresholds in island populations, little geographical work has been undertaken which analyses both temporal and spatial changes in the size of endemic threshold populations in regions that operate as independent epidemiological systems, featuring

complex patterns of spatial mobility and hierarchical spatial structures. We aim to address this research gap. It has generally been assumed that the population size at which disease becomes endemic in regional populations in mainland metapopulations such as England and Wales are *masked* (Black, 1966; Cliff et al., 2000; Brouin et al., 2005). The reintroduction of infection from small settlements to large towns and cities due to constant commuter-related travel, alongside the spatial transfer of infection via population flows in the other direction, blurs the point at which disease would be expected to fadeout, making the calculation of threshold estimates problematic. However, regions in England and Wales possess distinctive and unique characteristics in terms of a settlement hierarchy, demography, connectivity, dispersion of susceptibles and population densities, as well as the nature of spatial interaction between communities.

The Registrar-General’s *Weekly Return* provides highly accessible long-term, spatially resolved, historical incidence data at district-level in England and Wales. These qualities permit the identification and investigation of unique endemic threshold populations at finer spatial scales over time. This is important since the endemic threshold, in an applied context, has been forwarded as a guide for disease control. Regions form part of a larger mainland metapopulation. By identifying significant regional differences in endemic threshold populations, it is possible to locate epidemically important regions and local populations at the district-level where control efforts can be focused, thereby reducing transmission events that work to undermine the success of national vaccination campaigns. The *Weekly Return* also represents an extremely useful resource for investigating the impact of interventions on the spatial dynamics of disease, providing a consistent record of notifications at the same spatial and temporal scales before and after the onset of mass vaccination. This data enables the study of rescue effects (see glossary, Table 1), with potential to inform strategies of spatially-targeted immunisation programs.

Alongside quantifying long-term spatiotemporal changes in endemic threshold populations, we also seek to identify hotspots of disease incidence and reintroductions, districts critical to maintaining regional disease persistence and potential sources of rescue effects after the onset of vaccination. Such hotspots can play a profound role in shaping the size of endemic threshold populations. An argument has been made for ignoring populations below the threshold value if vaccines or associated resources are limited (Beyer et al., 2011). Yet this position only holds weight if hotspots are few in number and rescue effects are scarce (Metcalf et al., 2013), with disease persistence in regional populations largely dependent upon local persistence in large urban centres, operating as endemic reservoirs. Finally, a multivariate modelling framework for disease counts is implemented to investigate how spatial interaction between local populations may inhibit or promote disease persistence in regional populations, with the aim of generating greater understanding of potential drivers for spatiotemporal changes in endemic threshold populations. As this paper will demonstrate for one disease (pertussis), profound differences in the spatial structure and spatial hierarchy of settlements are associated with marked differences in the size of endemic threshold populations and persistence of infection.

2. Background: the disease and the study setting

2.1. Disease

Pertussis (also known as ‘whooping cough’) is a contagious, acute respiratory infection caused by the bacterium *Bordetella pertussis* (Warfel et al., 2012). The bacterium is transmitted via airborne droplets. The average incubation period for pertussis typically lasts eight days, whilst the infection period lasts 14 to 21 days. The disease has two distinct phases: (1) a catarrhal stage lasting approximately seven to ten days, with a fever often the most notable symptom; and (2) a paroxysmal stage, characterised by periods of intense coughing producing a distinctive ‘whooping’ sound (Amirthalingam et al., 2013). Since

humans are the only known reservoir for the disease, pertussis is theoretically a candidate for eradication via vaccination. However, unlike other vaccine-preventable childhood diseases, natural infection does not confer life-long immunity and vaccination does not guarantee long-lasting protection (Amirthalingam et al., 2013). Mass vaccination was not implemented in England and Wales until 1957 (Milward, 2019), when the trivalent diphtheria-tetanus-whole-cell-pertussis (DTwP) vaccine became part of the routine childhood vaccination schedule nationwide.

2.2. Study setting

The two regions selected for the present analysis are the historic county of Lancashire, located in North West England (Fig. 1B), and South Wales, comprising of four historic counties of Wales (Fig. 1C). Both regions are geographically and topographically diverse with large overall populations. However, they contrast significantly in terms of their spatial connectivity and isolation, thereby providing an interesting case study of the persistence of infection in different regional contexts within one nation.

Historically polycentric, Lancashire comprises numerous settlements with substantial populations which operate as economic and connectivity hubs (Fig. 2A). By the mid-20th century, Manchester and Liverpool represented two of the most complex, polycentric functional urban systems in the United Kingdom outside of London (Freeman, 1966). The majority of Lancashire's urban population has long resided in metropolitan areas located within and surrounding the South Lancashire coalfield, with expansive, sparsely populated hinterlands lying in the northern portion of the region. In contrast to Lancashire, urbanisation and connectivity in South Wales have historically been tightly constrained by the region's topography. In the middle years of the 20th century, South Wales was home to several densely populated, peri-urbanised valleys where the narrow valley floors and countless ridges inhibited settlement formation, producing hybrid landscapes with fragmented urban and rural characteristics. Communities and transport routes were linearly organised, running parallel to rivers and tributaries

(Jones, 1969). Beyond the valleys, South Wales was characterised by a clearly defined urban-rural hierarchy, as it remains today. The largest urban centres, Cardiff and Swansea, are located in Glamorgan (Fig. 2B), with populations approximately one-third the size of Manchester in 1960. Yet their combined metropolitan areas accounted for the vast majority of the region's urban population. In the hinterlands, South Wales remained sparsely inhabited, with high levels of internal isolation (Jenkins, 1992).

2.3. Study period

Our study period extends over a 30-year interval from January 1940 to December 1969. January 1940 represents the beginning of the first complete year of statutory notification of pertussis cases. For this period, the routine disease surveillance reports of the Registrar-General for England and Wales include an unbroken sequence of pertussis notifications suitable for analysing temporal changes in endemic threshold size. Lancashire and South Wales reported a high number of pertussis cases throughout the 30-year period, allowing changes in endemic threshold size to be detected across time using a 'moving window' empirical regression approach. The length of the period allows us to analyse nine time windows with an equal number of reporting weeks, before the abolition of the reporting units in April 1974. The period also represents a unique point in the epidemiological history of 20th century England and Wales, with the demographic upheaval of the post-war baby-boom and introduction of mass vaccination transforming the epidemic activity of pertussis.

The overall population of Lancashire grew by approximately 11% over the study period, from 4,630,000 in 1940 to 5,130,000 in 1969. South Wales experienced a similar level of growth over the same period, increasing by roughly 8.5% (from 1,833,000 in 1940 to 1,988,000 in 1969). Time-series of pertussis activity in Lancashire and South Wales suggest that epidemic outbreaks occurred every two to four years in the pre-vaccine era, and three to five in the vaccine era (See Fig. 3). After the introduction of routine immunisation in 1957, there was a dramatic fall in pertussis incidence in both regions. Excluding wartime years, the



Fig. 1. Location map of the study regions. A: Administrative counties of England and Wales, 1961. Lancashire and South Wales regions are highlighted in grey. B: Lancashire, with LGD boundaries. C: South Wales, with county and LGD boundaries. The Counties that compose South Wales are Pembrokeshire (1); Carmarthenshire (2); Glamorgan (3); and Monmouthshire (4).

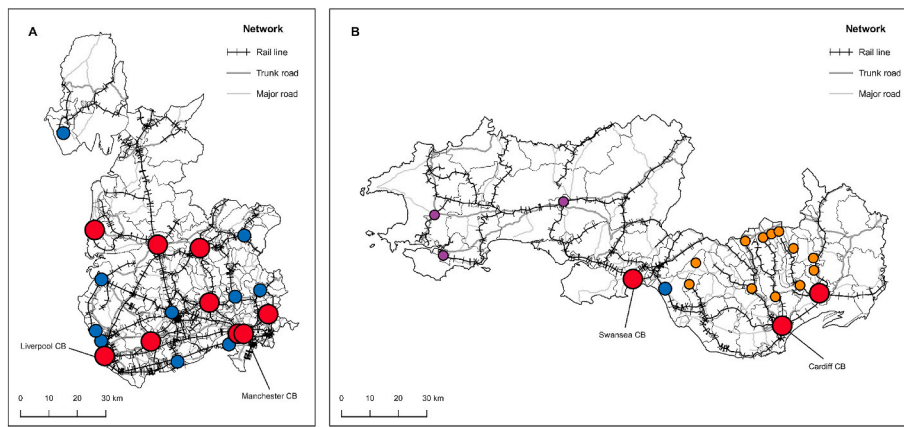


Fig. 2. Regional population centres and connectivity networks. A: Lancashire, B: South Wales. Red circles denote major urban centres with mean population size >100,000 over the study period 1940–69. Blue circles denote regional population centres with mean population sizes of 50,000–100,000. Orange circles denote urban settlements in the Valleys. Purple circles denote principal market towns. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

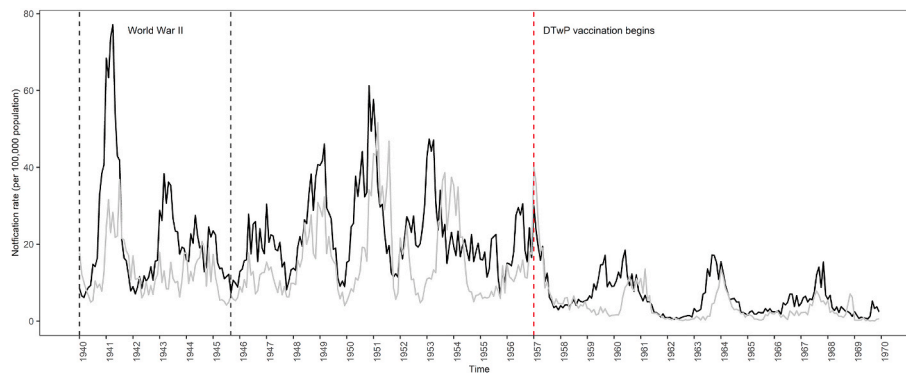


Fig. 3. Monthly time series of pertussis notification rates per 100,000 population, January 1940–December 1969. Black = Lancashire, grey = South Wales. Source: Registrar–General’s Weekly Returns (HMSO: London).

number of reported cases in Lancashire fell from 165,032 (1946–57) to 40,943 (1958–69); in South Wales, the number of reported cases fell from 42,974 (1946–57) to 8,835 (1958–69).

3. Materials and methods

3.1. Disease sources

3.1.1. Epidemiological data

Pertussis case notifications were abstracted from the Registrar–General for England and Wales’ *Weekly Returns* (HMSO: London). This disease record spans 1560 epidemiological weeks, from the week ended January 6, 1940 to the week ended January 2, 1970. Notification data were collected for all 125 local government districts (LGDs) in Lancashire, and 74 LGDs in South Wales. Here we note that, until April 1974, LGDs were the basic geographical reporting unit for disease notifications to the Registrar–General. Four principal sub-categories of LGD existed for local government: County Borough (CB); Municipal Borough (MB); Urban District (UD); and Rural District (RD). Overall, the *Weekly Return* provides an extraordinarily rich time-series of geographically aggregated infectious disease data at fine spatial and temporal scales.

Subtle changes were made to boundaries and land areas of LGDs in both regions throughout the study period, reflecting temporal changes in population size and density. In Lancashire, a new district was created mid-way through the study period, while another was abolished. These districts (Kirkby UD and Limehurst RD) were excluded to ensure a complete time-series of case reports across the study period for all LGDs under analysis. Separate regional datasets consisting of 125 Lancashire LGDs and 74 South Wales LGDs with associated demographic data were

constructed for each time window (see Section 3.2.1 below). Percentage endemicity for LGDs, along with their mean population sizes, forms the basis of the endemic threshold analysis.

3.1.2. Demographic data

Data on population size and population density for Lancashire and South Wales LGDs were abstracted from annual editions of the Registrar–General’s *Statistical Review for England and Wales* (HMSO: London). These data provide vital information on demographic stochasticity and susceptible dynamics, enabling a detailed analysis of observed epidemiological patterns within their ecological context. Since population density is measured in persons per acre in the *Statistical Review*, the measurement of population density was transformed to persons per square kilometre (per km²).

3.1.3. Digital shapefiles

ESRI polygon shapefiles, providing the administrative boundaries for counties and LGDs in Lancashire and South Wales for 1951, 1961 and 1971, were obtained from www.VisionofBritain.org.uk, part of the Great Britain Historical GIS Project run by the University of Portsmouth. Maps were produced using QGIS 3.10 ‘A Coruña’.

3.2. Endemic threshold population estimation

3.2.1. Time windows

Building on the empirical regression approach outlined by Cliff et al. (2000), the 30-year time series of weekly pertussis notifications was broken down into nine 312-week time windows to track temporal changes in endemic threshold size in Lancashire and South Wales across the study period. We selected time windows of 312 weeks duration to

ensure an adequate number of windows to detect temporal changes in threshold population size in a relatively short study period. This also allowed a sufficient number of weeks in each time window to produce estimates from model parameters without unreasonably large standard errors. A 312-week window ensured at least one epidemic was captured in all nine time windows; a 3-year overlap between time windows provides a smoothing effect between successive windows.

The time windows studied are as follows: 1940–45, 1943–48, 1946–51, 1949–54, 1952–57, 1955–60, 1958–63, 1961–66 and 1964–69. The first five time windows cover the pre-vaccine era of the study period, with the latter four time windows corresponding with the vaccine era. The population size of LGDs in each time window was defined as the mean population size for the 312-week period.

3.2.2. Model outliers

In the case of Lancashire, an important issue was the ‘swamping’ effect of two outliers which persisted in each of the nine-time windows: Liverpool CB and Manchester CB. Both districts possess populations approximately four times greater than the next largest district and yielded an excessive influence in the fitting of the models. As the only districts at the 100% endemicity threshold in each of the nine-time windows for pertussis, their presence in the analysis could act to constrain the form of the regression line, functioning as ‘a cap in a closed number system’ (Cliff et al., 2000; 93). Consequently, Liverpool CB and Manchester CB were omitted from all models before estimation. Although not reported, preliminary analysis indicated that the inclusion of Liverpool CB and Manchester CB in the modelling procedure resulted in only marginal increases in the regional threshold estimates across all time windows.

3.2.3. Estimating threshold size

To estimate the size of the endemic threshold population for each window, the approach of Cliff et al. (2000) was adopted. The overall threshold estimate was determined initially by using a simple linear regression model of the form:

$$(\text{percentage endemicity}) = \hat{b}_0 + \hat{b}_1(\text{mean population size}) \quad (1)$$

Since population size ranges over several orders of magnitude, it is unknown what form the regression relationship may take. To address this, linear-log and log-log regressions were also fitted. Using equation (1) and calculating the relevant parameter values, we can estimate the population size at which pertussis cases would be reported for all weeks in the time-series. This value is the endemic threshold.

To investigate the effect of geographical isolation on endemic threshold size, a binary distinction was made between the most connected and isolated LGDs according to distance from the nearest endemic centre. This acts as a coarse connectivity index in the absence of data on migration/population flows between LGDs for each region. Liverpool CB and Manchester CB in Lancashire and Cardiff CB in South Wales were defined as endemic centres since all three LGDs consistently reported at or very close to the 100% threshold across all nine-time windows. A Euclidean distance matrix was calculated using the centroids of each LGD for Lancashire and South Wales to establish the distance from the nearest endemic centre for each district. LGDs below the median distance from the nearest endemic centre were coded as zero and LGDs greater than the median distance from the nearest endemic centre as one. Similarly, for population density, LGDs were dichotomised according to their population density per square kilometre, with LGDs below the median population density coded as zero, and LGDs higher than the median population density as one.

3.3. Hotspot & survival analysis

3.3.1. Spatial coupling and correlation

OLS linear regression models were fitted between the total annual

number of fadeouts (TAFs) and mean population size for Lancashire and South Wales to estimate reintroduction events across the length of the pre-vaccine era (1946–1957). This analysis seeks to identify potential pertussis hotspots prior to the onset of vaccination and explore the effect of spatial coupling on endemic persistence. We therefore do not employ a moving window approach here, instead focusing on a pre-vaccine time-series of 624 weeks duration. The war years (1940–45) were excluded to ensure time-series for the pre-vaccine and vaccine eras consisted of an equal number of reporting weeks for comparative purposes. For the purposes of the present analysis, a fadeout was defined as a period of four weeks or more without reported cases of infection. Since the infectious period for pertussis may last up to three weeks, a fadeout period of four weeks ensured that chains of transmission had broken down locally. Here, we note that a negative correlation between the number of fadeouts and population size has been well-established in previous empirical research on measles and, to a lesser extent, pertussis (Bartlett, 1957; Earn et al., 1998; Rohani et al., 1999). Consequently, a negative relationship between local population size and TAFs is expected.

Following the method of Bharti et al. (2010), the analysis concentrates on residuals from OLS linear regression models to potentially reveal areas of critical importance for maintaining pertussis persistence in Lancashire and South Wales. LGDs with negative residuals indicate areas with fewer fadeouts relative to their population size and may begin to elucidate the role of spatial connectivity on influencing rates of disease reintroduction. Regression residuals were examined for spatial autocorrelation to initially assess the impact of spatial proximity and human mobility on pertussis persistence in each region by performing a Moran’s *I* test, using R package *spdep* (Bivand et al., 2013). Since data on population flows between LGDs in both regions during the study period is not available, a geographical proxy for mobility and interaction between districts was required. A contiguity-based spatial weighting was incorporated in Moran’s *I* tests for both regions to capture characteristics of contagious transmission. LGDs sharing administration boundaries with common borders were defined as neighbours.

3.3.2. Persistence in vaccine era

The onset of vaccination midway through the study period serves as a natural experiment to analyse mechanisms which fundamentally affect disease persistence. For instance, the detection of potential hotspots in the vaccine era can potentially reveal important information concerning the location of ‘corridors of infection’, where intra-regional movement between local populations is concentrated and spatial importation of infection is a regular occurrence. Pertussis hotspots in the vaccine era are defined as LGDs with more cases reported than the regional mean, adjusted for population size. We use pertussis incidence here rather than TAFs to discriminate between persistence and invasion dynamics. These LGDs were identified after calculating the total number of weekly cases notified in each district in the 12 years following the introduction of routine vaccination in 1957. To test for spatial autocorrelation amongst ‘hotspots’, a Moran’s *I* test was performed, with hotspots treated as a binary variable (i.e. 1 = hotspot; 0 = non-hotspot).

3.3.3. Rates of Re-Introductions

Comparing fadeout duration of hotspots and other districts in the pre-vaccination and vaccine era periods provides additional insights into spatiotemporal changes in pertussis persistence at the local level. The impact of vaccination on spatial coupling can be assessed and districts of vital importance to maintaining the regional circulation of pertussis over time can be identified. Long fadeout periods indicate low geographical connectivity, usually coupled with small populations with low susceptible density, resulting in infrequent to rare re-introduction of disease. Short fadeout periods suggest frequent reintroduction events due to rescue effects. Following the approach of Bharti et al. (2010), a survival analysis was performed by fitting a multivariate Cox proportional hazard regression model, with mean fadeout duration serving as

the waiting time, to determine the survival probability of disease fade-outs in hotspots and other districts in both regions. The waiting time represents the number of weeks without reported cases until a reintroduction event. The following independent variables were incorporated into each Cox regression model: population size; susceptible input (number of births); susceptible density; and distance from nearest endemic centre.

3.4. Endemic-epidemic modelling: 'hhh4' models

The moving window approach allows estimates of the endemic threshold to be calculated but sheds little light on the endemic-epidemic dynamics of pertussis in each time window and the concurrent factors influencing these dynamics, such as spatial interaction, population size, and random effects. To address this, the 'hhh4' modelling framework is utilised to construct a greater understanding of the nature of disease spread in each time window, identifying the drivers of persistence in each regional metapopulation and how these contribute to the emergence of hotspots, potentially influencing the temporal changes in endemic threshold populations. The hhh4 model is a multivariate time-series model for infectious disease counts that divides disease incidence into its endemic and epidemic components, modelling the expected baseline rate of notifications, while also capturing the influence of previous cases in the same and neighbouring districts (Held et al., 2005; Held and Paul, 2012). Basic hhh4 models were fitted for each regional time window, with successive models run with extensions included to assess the impact of geographical connectivity on disease persistence by capturing networks of spatial spread, and assorted random effects upon the endemic-epidemic dynamics of pertussis. The models were implemented using the R package surveillance (Meyer et al., 2017).

3.4.1. Model formulation

The proposed hhh4 multivariate time-series model is built upon an additive decomposition of disease incidence into three components: one endemic component and two observation-driven epidemic components. The epidemic component involves explicit autoregressions on the number of cases at the previous time-point (Meyer and Held, 2014), which can be viewed as a multivariate branching process. The first part of the epidemic component represents an autoregression of the disease incidence within a district, while the other part is an autoregression incorporating the number of cases from surrounding districts, capturing spatial spread. The autoregressive coefficients follow a threshold phenomenon, similar to the basic reproduction number R_0 in susceptible-infectious-recovered (SIR) models (Anderson and May 1992). The basic model assumes a mean structure for disease incidence across the time-series under analysis and assumes, conditional on past observations, that count data has a negative binomial distribution

$$Y_{it}|Y_{t-1} \sim \text{NegBin}(\mu_{it}, \psi), \quad (2)$$

where Y is the time series of weekly count data, i is the geographical district, t is time-period (weeks), ψ is the overdispersion parameter and μ_{it} is the additively composed mean. The mean structure decomposes disease risk additively into three components

$$\mu_{it} = e_{it}v_{it} + \lambda_{it}Z_{t-1} + \varphi_{it} \sum_{j \neq i} \omega_{ji}Y_{j,t-1}, \quad (3)$$

where e_{it} is the offset of known counts reflecting population at risk and ω_{ji} is the weight for the neighbourhood component reflecting strength of transmission from district j to district i . The first (endemic) component represents variation in disease incidence which cannot be attributed to the previous number of cases

$$\log(v_{it}) = \alpha_i^{(v)} + \beta^{(v)\top} z_{it}^{(v)}, \quad (4)$$

where v is the unknown endemic parameter. The second (autore-

gressive) component accounts for autoregressive effects; the reproduction of disease within district i

$$\log(\lambda_{it}) = \alpha_i^{(\lambda)} + \beta^{(\lambda)\top} Y_{it}^{(\lambda)}, \quad (5)$$

where λ is the unknown autoregressive parameter. The final (spatio-temporal) component accounts for neighbourhood effects; the transmission of infection from surrounding districts

$$\log(\varphi_{it}) = \alpha_i^{(\varphi)} + \beta^{(\varphi)\top} Y_{it}^{(\varphi)}, \quad (6)$$

where φ is the unknown neighbourhood parameter. Without the epidemic components, the model would represent a standard negative binomial regression model for independent observations (Meyer and Held, 2014). Here, $\alpha_i^{(v)}$, $\alpha_i^{(\lambda)}$ and $\alpha_i^{(\varphi)}$ are component-specific intercepts and $\beta^{(\lambda)\top}$, $\beta^{(\varphi)\top}$ and $\beta^{(v)\top}$ are the vectors of the fixed effects for each component. The parameters v , λ , and φ are allowed to vary across districts to enable the inclusion of district-specific covariates and heterogeneity.

A common intercept is assumed across districts in the endemic component to prevent districts with zero case reports from being forcibly excluded. We use mean district population size as the endemic offset in the hhh4 models fitted for each time window. The importance of the two epidemic components are assessed using dominant eigenvalues ($\max EV$), a combined measure for epidemic potential. If the dominant eigenvalue is below unity (i.e. below 1), this value represents the epidemic proportion of total disease incidence. Likelihood inference is performed using generic numerical optimisation routines (Paul and Held, 2011). For data with overdispersion, we use maximum likelihood to estimate parameters and standard errors by maximising the negative binomial log-likelihood of the model. The hhh4 model framework allows for covariate effects on either the endemic or epidemic components of disease incidence to be included using model extensions.

3.4.2. Model extensions

Spatial Interaction. In its basic formulation, a hhh4 model assumes that the spread of infection is restricted to first-order neighbours; all districts have the same potential for importing cases from adjacent units (Meyer et al., 2017). However, the assumption that infection spreads only via adjacent regions is too simplistic; individuals can travel longer distances, with movement often concentrated around large urban centres in regions with hierarchical population structures (Bartlett, 1957). A more appropriate model of spatial interaction is the gravity model, which enables analysis of hierarchical transmissions between cities, towns, and villages according to spatial coupling patterns (Xia et al., 2004). In its most common form, the gravity model postulates that the population flow between two geographical units is log-linearly dependent on population size and distance (Jandarov et al., 2014), suggesting a scaling process in spatial interaction. Crucially, a gravity model can be calculated without detailed network data on population movement (Geilhufe et al., 2014). The modelling framework can be extended to account for short-range, commuter-driven spread and long-range transmission of cases between districts to incorporate a gravity model.

To reflect commuter-driven spread, we account for the district-specific population in the spatio-temporal epidemic component. District susceptibility to infection is scaled according to population size, multiplying the neighbourhood parameter (ϕ) by district population size (e_i^{pop}). To reflect long-range transmission, a power-law is included, with weights for the neighbourhood component (w_{ji}) estimated as a function of adjacency order (o_{ji}) in the neighbourhood graph of geographical units between districts (Meyer and Held, 2014). Formally, two districts, districts j and i , are k th-order neighbours if the shortest distance between them has k steps across distinct regions. Here, k represents the number or neighbours. The network of districts thus features

a symmetric $I \times I$ matrix of neighbourhood orders. According to this measure, the first-order weight matrix is generalised to higher-order neighbours with the power-law model assuming the form $w_{ji} = o_{ji}^{-d}$, for $j \neq i$ and $w_{ji} = 0$, where the decay parameter d is estimated. Higher-order neighbours diminish in importance as d increases.

3.4.2.1. Random effects. The final model extension incorporates random effects to account for unobserved covariates in regions with a large number of districts that influence heterogeneity in disease incidence (Paul and Held, 2011). We account for two unobserved heterogeneities: underreporting and edge effects. To address underreporting, independent random effects were introduced by including district-specific intercepts in the autoregressive and spatiotemporal epidemic components $b_i^{(\lambda)}$ and $b_i^{(\phi)}$. Districts on the region's edge are missing potential sources of infection from across the border. To capture unobserved heterogeneity due to 'edge effects', district-specific random effects were included in the endemic component $b_i^{(v)}$. Accounting for unobserved heterogeneities is expected to improve model fit by reducing residual heterogeneity.

3.4.3. Model assessment

We analyse model performance using successive one-step-ahead predictions assessed by strict proper scoring rules (Czado et al., 2009). Scoring rules measure predictive quality by assigning a numerical score based on the probability distribution from a fitted model and the later observed true value. Lower scores correspond to better predictions (Meyer et al., 2017). Two strictly proper scoring rules are used to assess model performance: the logarithmic score (logS) and the ranked probability score (RPS).

4. Results

4.1. Endemic threshold estimation

The endemic threshold analysis utilises Cliff et al. (2000)'s empirical linear regression approach as described in Section 3.2. Since percentage endemicity and population size range over many orders of magnitude, the log-log form of equation (1) provides the most effective representation of the endemic threshold estimates and it is the results for models using the log-log regression equation that we report here.

Fig. 4 displays the estimated threshold values calculated by fitting the percentage of weeks with zero case reports against population size for all districts using log-log regression models for Lancashire and South Wales. The basic shape of the trend line for South Wales is quadratic, with the minimum threshold value found towards the middle of the study period, falling from 293,000 in 1940–45 to 162,000 in 1952–27. This is in stark contrast to threshold estimates for Lancashire, which reveal the regional endemic threshold population remains relatively constant over the same period ($\sim 125,000$, see Fig. 4). The introduction of routine vaccination in 1957 sees remarkable growth in the size of the endemic threshold population in South Wales, rising to 1,720,000 in the final time window (1964–69), a tenfold increase from the 1952–57 estimate. Lancashire also experiences considerable endemic threshold population growth, rising steadily to 461,000 in 1964–69, a fourfold increase.

The effect of population size upon endemic threshold size reflects only a small component of a more complex epidemiological system. We consider the impact of internal and external factors such as connectivity and population density by estimating the endemic threshold size for LGDs with low or high connectivity, and low or high population densities.

Fig. 5A shows that districts with low connectivity in both regions experienced a dramatic increase in endemic threshold population size in the vaccine era, with immunisation disrupting chains of transmission and removing susceptibles from subpopulations. It is notable that by the

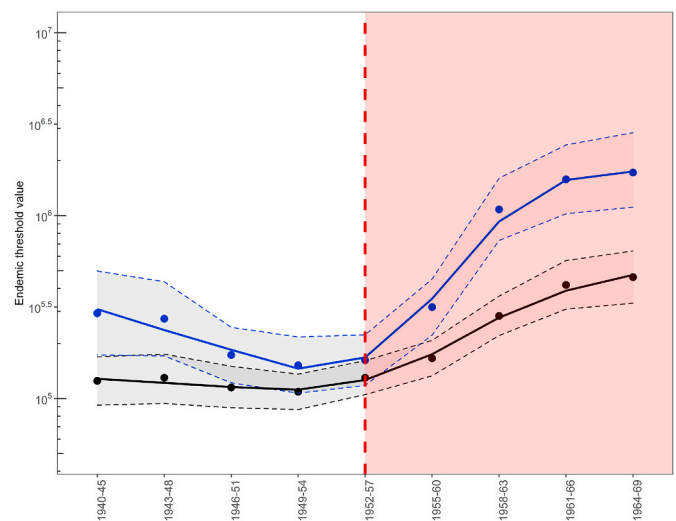


Fig. 4. Endemic threshold size estimates for nine time windows. Dots represent the endemic threshold estimate for each window. To show the time trend across time windows, solid LOESS lines have been fitted to the points for each region. The red shaded area denotes the vaccine era. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

final time window, a strong quadratic trend in endemic threshold size is present for communities with low connectivity in South Wales. Fig. 5A and B reveal very similar threshold estimates for LGDs in both connectivity categories in Lancashire before the onset of vaccination. There is little variation in the estimated threshold population size across the first five-time windows, despite the demographic shock of the baby-boom in the post-war years. Vaccination does not significantly increase the endemic threshold population in highly-connected districts (Fig. 5B), rising from 49,000 in 1940–45 to 145,000 in 1964–69. Fig. 5C shows a similar shape in the trend line for low-density LGDs in South Wales to that shown in Fig. 5A, with a constant decrease in threshold size during the post-war baby boom. Intriguingly, low-density districts in Lancashire see a significant increase in endemic threshold size in the two time windows before the introduction of vaccination. Fig. 5D also shows a quadratic-shaped trend line for high-density LGDs in South Wales, with increased density in less dispersed populations during the pre-vaccine era serving to reduce the spacing between susceptibles.

4.2. Hotspot & survival analysis

4.2.1. Spatial correlation & coupling patterns

Linear regression analysis reveals a negative association between TAFs and mean population size in Lancashire ($R^2 = 0.26$) (Fig. 6A) and South Wales ($R^2 = 0.17$) (Fig. 6B). Negative residuals were detected for 55 Lancashire LGDs and 31 South Wales LGDs in the OLS regression model, indicating districts with less-than-expected fadeout events related to population size, suggesting higher levels of pertussis persistence than would otherwise be expected. Moran's I tests reveal that both Lancashire LGDs (Moran's I statistic = -0.03 , $p = 0.77$) and South Wales LGDs (Moran's I statistic = -0.14 , $p = 0.15$) with negative residuals are not significantly spatially auto-correlated. In Fig. 7A, a clearly defined spatial pattern of districts with inflated rates of pertussis reintroduction is not evident but does provide initial visual indication of infection fanning out from conurbations that serve as geographical reservoirs of infection for surrounding areas. In South Wales, districts with inflated pertussis reintroduction rates are not concentrated spatially and are distributed unevenly across the region in South East Carmarthenshire, South East Pembrokeshire, and patches of 'the Valleys' in Glamorgan and Monmouthshire. In general, these districts represent small and

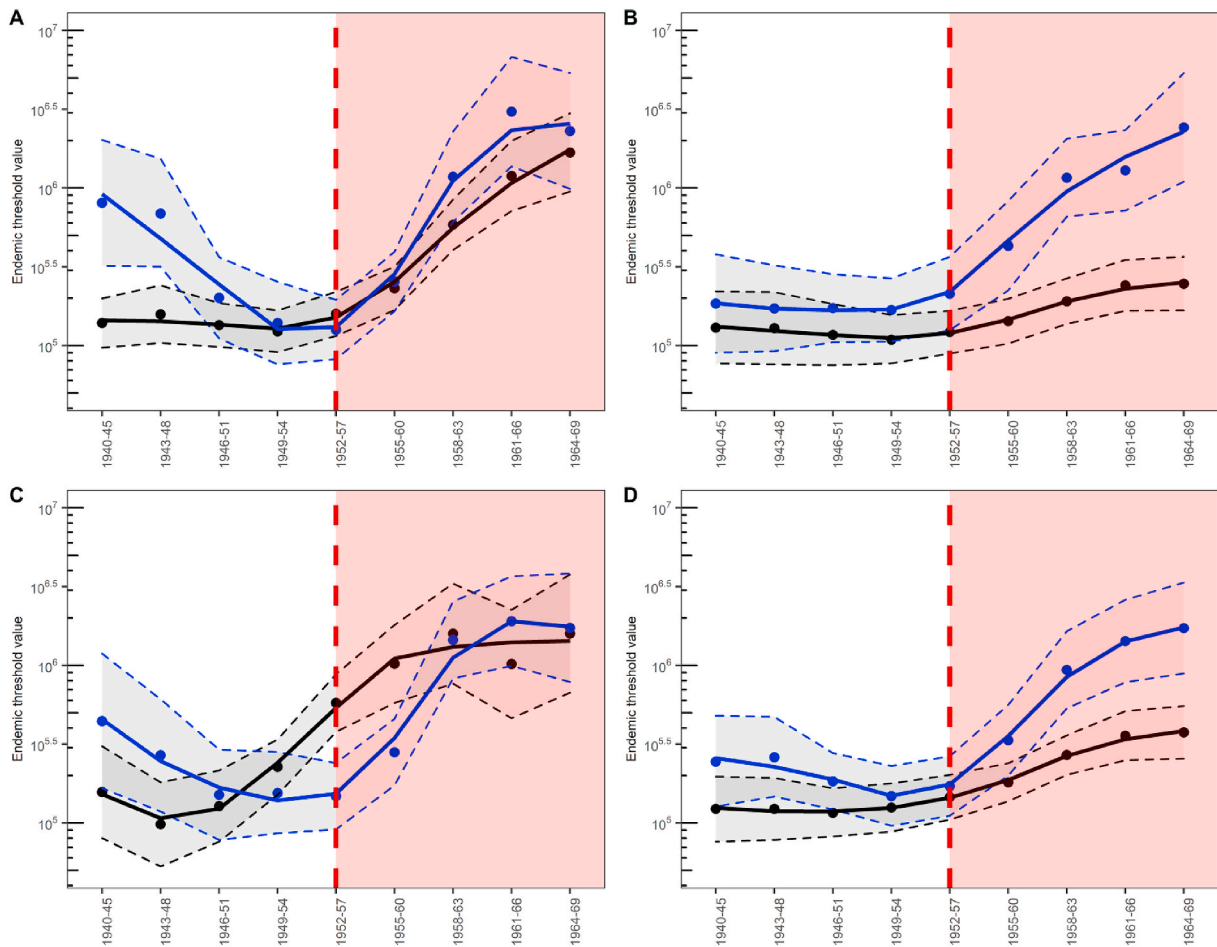


Fig. 5. Endemic threshold size estimates for nine time windows, accounting for geographical connectivity and population density. A: Low connectivity LGDs. B: High connectivity LGDs. C: Low density LGDs. D: High density LGDs. Black = Lancashire, blue = South Wales. The red shaded area denotes the vaccine era. Dotted lines represent the 95% confidence intervals. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

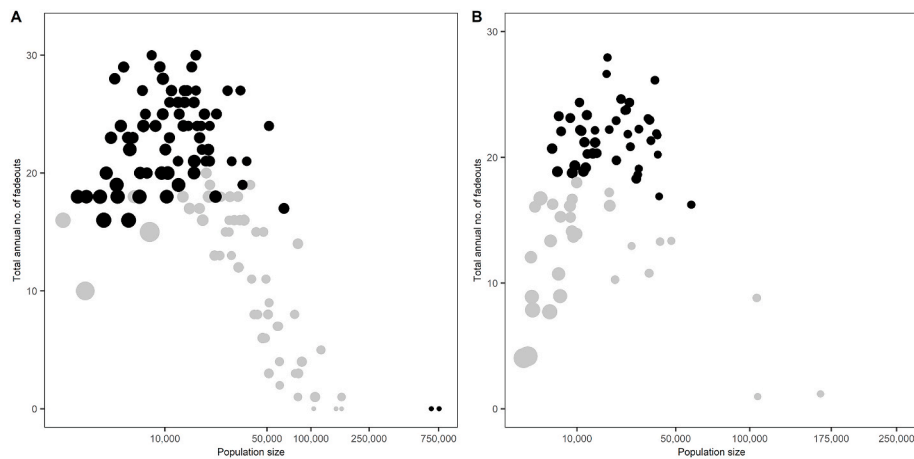


Fig. 6. Total annual number of fadeouts (TAFs) against mean population size for the pre-vaccine era. Circle size increases with mean fadeout duration. Black = positive residuals, grey = negative residuals. A: Lancashire, B: South Wales.

medium sized urban centres and market towns (See Fig. 7B).

4.2.2. Disease persistence

LGDs with cases above the mean after the onset of vaccination act as a flag for likely importation due to enhanced spatial connectivity, and hotspots are categorised according to this criteria. Thirty-eight

Lancashire LGDs reported more pertussis cases than the regional mean following the introduction of mass vaccination. Thirty-five of these hotspots had higher-than-expected rates of reintroduction in the pre-vaccine era. The hotspots vary considerably in population size (16,655–728,271) and display significant positive spatial autocorrelation (Moran's *I* statistic = 0.14, *p* < 0.01). Fig. 8A shows pertussis

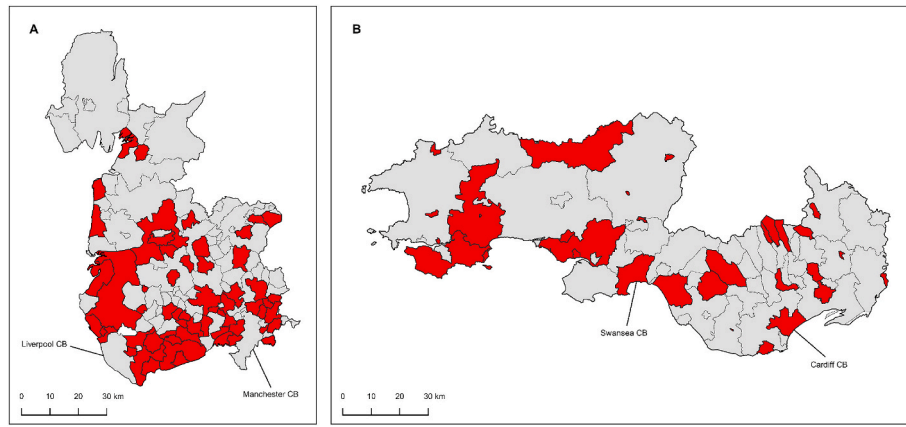


Fig. 7. Pre-vaccine era: LGDs with higher rates of pertussis reintroduction in relation to population size, identified by red shading ('hotspots'). A: Lancashire, B: South Wales. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

hotspots in the vaccine era are concentrated mainly in the Greater Manchester and Merseyside conurbations, incorporating Manchester CB, Liverpool CB, and their surrounding districts. Several CBs which represent key regional centres in Eastern and Northern Lancashire are also notable hotspots of pertussis persistence.

Seventeen LGDs in South Wales were found with more pertussis cases than the regional mean in the vaccine era. Of the 31 districts identified as hotspots in the pre-vaccine era, only nine districts continued to exhibit signs of pertussis persistence in the vaccine era. The 17 potential hotspots vary considerably in population size (21,039–264,663) and were not significantly spatially autocorrelated (Moran's *I* statistic = 0.09, *p* = 0.22). Vaccine era hotspots tend to be located within or surrounding the Valleys sub-region and include the largest population centres in South Wales (e.g. Cardiff CB, Swansea CB, Newport CB) and their neighbouring districts (Fig. 8B). Many of these districts contain key trunk roads and rail lines, important components of regional network connectivity.

4.2.3. Rates of Re-Introductions

In the pre-vaccine era, fadeout duration in pertussis hotspots was relatively consistent regionally, with fadeout periods on average lasting approximately two weeks before pertussis was reintroduced (see

Table 2). Mean fadeout duration for other districts in Lancashire during this period was half the fadeout duration for other districts in South Wales, indicating an overall higher number of transmission events

Table 2

Mean fadeout duration (in weeks) for hotspots and other LGDs in Lancashire and South Wales in the pre-vaccine (1946–57) and vaccine (1957–69) eras.

	Mean fadeout duration (in weeks)	
	Pre-vaccine	Vaccine
<i>Lancashire</i>		
Hotspots	1.96	2.73
N	55	38
SD	1.40	1.40
Other districts	3.27	10.62
N	70	87
SD	2.21	7.91
<i>South Wales</i>		
Hotspots	2.03	4.21
N	31	8
SD	0.71	1.81
Other districts	7.43	19.93
N	43	66
SD	6.62	26.55

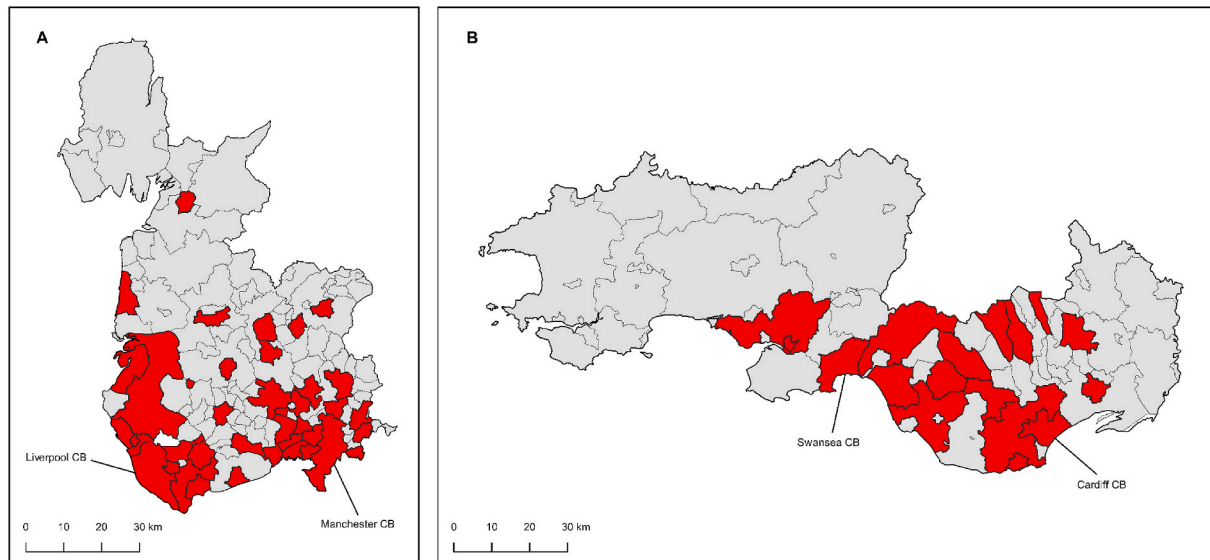


Fig. 8. Vaccine era: 'Hotspot' LGDs reporting a higher number of pertussis cases than the regional mean are identified by red shading. A: Lancashire, B: South Wales. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

between and within subpopulations in Lancashire. The introduction of vaccination sees mean fadeout duration almost double in length for hotspots and triple for other districts in South Wales (Table 2), indicating a significant decline in transmission events by decreasing the susceptible pool through immunisation. In contrast, vaccine era hotspots in Lancashire experienced a marginal increase in mean fadeout duration, from 2 to 2.7 weeks (SD = 1.40). There is a significant increase in fadeout duration in other districts in the region, but this is still far lower compared to South Wales, standing at 10.6 weeks (SD = 26.55).

The results of the multivariate Cox regression analyses reveal that population size makes a very marginal contribution to increasing the rate of disease reintroductions in hotspot districts in pre-vaccine Lancashire, a hazard ratio (HR) of 1.048 (95% CI, 1.003–1.096; $p < 0.05$). This would suggest that larger population size is positively associated with the rate of reintroductions. However, since the confidence interval for HR includes 1, these results indicate that population size makes a very small contribution. Neither susceptible input, density or distance from endemic centre are significant factors in relation to fadeout duration. This notably changes in the vaccine era: a HR of 0.96 (95% CI, 0.93–0.99; $p < 0.05$) suggests increasing distance from endemic centres leads to a 4% decrease in the rate of reintroductions among hotspots. Susceptible density is associated with an increased risk of pertussis in the vaccine era (HR = 1.209, 95% CI, 1.003–1.096; $p < 0.1$), although this association is weak in terms of statistical significance. In South Wales, there is no statistically significant difference between hotspots and other districts in terms of the rate of disease reintroduction. In vaccine era South Wales, population size was found to have a strong positive relationship with the rate of reintroductions in hotspots after holding the other covariates constant (HR = 1.198, 95% CI, 1.069–1.342; $p < 0.01$). A full breakdown of the parameters and hazard ratios for the multivariate Cox regression models are presented in Table 3.

4.3. Endemic-epidemic modelling

Three models were utilised to analyse the endemic and epidemic dynamics of pertussis in each time window across the study period in Lancashire and South Wales, accounting for spatial interaction and unobserved heterogeneities. These models are as follows: the basic hhh4 model, a gravity model with power law extension, and a random effects model (incorporating the gravity model and power law extensions). The full output of the hhh4 models can be found in the supplementary material in Appendix A. There is significantly less overdispersion in models extended to further include random effects in both regions (see in Table 4), indicating reduced residual heterogeneity. Incorporating

population into the spatiotemporal component of the endemic-epidemic models reveals that more populous districts attract a greater number of infections from neighbouring areas, with a significant reduction in the strength of the distance decay parameter when random effects are included (see Tables S7–S10, Appendix A). Commuter-driven spread produces an agglomeration effect, confining pertussis incidence to endemic centres and their surrounding areas. The hotspot analysis of vaccine era reported incidence captures the impact of this dynamic in the Lancashire region (See Fig. 8). The introduction of random effects describes the quadratic trend in estimated endemic threshold values in South Wales more effectively than the inclusion of a gravity model alone. The dominant eigenvalues are considerably higher in the 1943–48, 1946–51, 1949–54 and 1952–57 time windows (Table 4), indicating increased levels of epidemic activity during this period corresponding with declining time window threshold size estimates for South Wales. For both regions, the dominant eigenvalues are largest in the random effects model for each time window. This is unsurprising since the models explicitly account for the impact of underreporting and edge effects on epidemic incidence. Assessment of mean predictive scores for each model using strictly proper scoring rules reveal the random effects model gives the best one-week-ahead predictions across time windows in each region (see Table 5), registering the lowest mean scores (LogS = 1.119 and RPS = 1.002, respectively).

5. Discussion

Research on pertussis endemicity in England and Wales has been relatively limited, focused primarily on examining the impact of vaccination and analysing weekly or monthly pertussis incidence data at the national level and between major cities (Earn et al., 1998). The present paper has provided a comprehensive account of spatiotemporal changes in endemic threshold populations for pertussis in contrasting, dynamic regions in England and Wales by applying a methodology previously confined to the study of island populations (Cliff et al., 2000), elucidating the influence of spatial structure, connectivity, and dispersion on shaping the endemic persistence of pertussis through time and space.

Hagenaars et al. (2004) called for more efforts to be made to develop greater insights into how spatial heterogeneity affects persistence across a variety of population-dynamical regimes. The analysis of two populous regions in England and Wales with contrasting spatial structures, settlements of varying population size, density and dispersion, and networks of connectivity presented in this paper reveals significant regional differences do exist in terms of endemic threshold populations for pertussis, despite sharing metapopulation dynamics. These findings are notwithstanding the high level of connectivity nationwide and the

Table 3

Parameter estimates and measures of dispersion for multivariate Cox regression analyses of fadeout duration in Lancashire and South Wales in the pre-vaccine and vaccine eras.

	Era	Parameter	Estimate	SE	HR	95% CI
Lancashire	Pre-vaccine	Population size	0.047**	0.023	1.048	1.003–1.096
		Susceptible input	−0.521	1.473	0.594	0.033–1.065
		Susceptible density	0.186	0.116	1.205	0.960–1.512
		Distance from endemic centre	−0.008	0.011	0.992	0.970–1.014
	Vaccine	Population size	0.040	1.041	1.041	0.983–1.102
		Susceptible input	0.257	1.293	1.293	0.048–3.510
		Susceptible density	0.190*	1.209	1.209	0.976–1.498
		Distance from endemic centre	−0.042**	0.960	0.960	0.928–0.992
South Wales	Pre-vaccine	Population size	0.028	0.021	1.028	0.986–1.072
		Susceptible input	0.871	1.290	2.390	0.191–2.900
		Susceptible density	0.110	0.157	1.116	0.820–1.519
		Distance from endemic centre	−0.013	0.009	0.987	0.970–1.005
	Vaccine	Population size	0.181***	0.058	1.198	1.069–1.342
		Susceptible input	−0.774	1.604	0.461	0.020–1.070
		Susceptible density	−0.070	0.196	0.932	0.635–1.370
		Distance from endemic centre	0.005	0.014	1.005	0.977–1.034

Note: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$.

Table 4

Overdispersion parameter (with standard errors) and dominant eigenvalues (*maxEV*) for power law + gravity hhh4 models and random effects hhh4 models for nine time-windows in Lancashire and South Wales. Dashes indicate models which failed to converge after reaching the maximum number of iterations ($n = 100$).

		Models								
		40–45	43–48	46–51	49–54	52–57	55–60	58–63	61–66	64–69
<i>Lancashire</i>										
Gravity + Power Law	ψ (se)	1.38 (0.02)	1.30 (0.02)	0.90 (0.02)	0.82 (0.01)	1.09 (0.02)	1.39 (0.03)	2.37 (0.06)	2.66 (0.08)	3.07 (0.09)
	<i>maxEV</i>	0.82	0.62	0.74	0.79	0.79	0.86	0.8	0.76	0.67
Random effects	ψ (se)	1.05 (0.02)	1.08 (0.02)	0.70 (0.01)	–	0.90 (0.02)	0.70 (0.02)	–	2.12 (0.07)	2.43 (0.07)
	<i>maxEV</i>	0.99	1.00	0.92	–	1.00	0.99	–	0.98	0.96
<i>South Wales</i>										
Gravity + Power Law	ψ (se)	2.80 (0.08)	2.81 (0.08)	2.26 (0.06)	2.36 (0.06)	2.81 (0.07)	3.66 (0.12)	4.57 (0.21)	4.59 (0.24)	5.76 (0.29)
	<i>maxEV</i>	0.69	0.70	0.84	0.82	0.71	0.71	0.78	0.80	0.76
Random effects	ψ (se)	2.23 (0.07)	2.39 (0.07)	1.87 (0.05)	–	2.45 (0.07)	2.89 (0.10)	3.47 (0.17)	–	4.27 (0.21)
	<i>maxEV</i>	0.77	1.04	0.80	–	0.99	0.76	0.76	–	0.71

Table 5

Mean score of predictive performance (measured by LogS and RPS) for three competing models in the temporal and spatial dimensions for pertussis incidence across all nine time-windows in Lancashire and South Wales.

Model	LogS	RPS
Basic	1.309	1.159
Gravity + Power law	1.307	1.161
+ Random effects	1.190	1.002

successful uptake of the pertussis vaccine following its introduction in 1957.

Endemic threshold estimates for the pre-vaccine era reveal stark regional differences in threshold population size, demonstrating the influence of geographical variability in population density and spatial connectivity on shaping the size of endemic threshold populations. Within the context of South Wales, the high endemic threshold values for time windows at the beginning of the study period can be attributed to a combination of low population density and high levels of internal isolation. Much of the region is rural, sparsely populated with high levels of dispersion. Immediately following WWII, the post-war baby boom resulted in substantial growth in the number and density of susceptibles within LGDs of all sizes across South Wales, leading to a dramatic fall in the threshold estimate. Intriguingly, low-density districts in Lancashire experienced a significant increase in the endemic threshold population size in the two time windows before the introduction of vaccination. One hypothesis is that migration from rural areas to urban centres and newly emerging suburbs throughout the 1950s further reduced the density of small rural communities and the number of transmission events. Overall, there was only modest growth in the endemic threshold in Lancashire during the pre-vaccine period, despite the post-war baby boom. High levels of spatial coupling between the Manchester and Liverpool conurbations and surrounding urban districts ensured a consistent transmission of infection to neighbouring areas and satellite towns further afield, similar to the spatio-temporal travelling waves of measles observed across England and Wales in the mid-to-late 20th century (Grenfell et al., 2001).

The introduction of mass vaccination served to increase regional endemic threshold populations by depleting the pool of susceptibles, thereby increasing the population size requirements for maintaining chains of pertussis transmission. In both regions, mean fadeout duration doubled in pertussis hotspots and increased threefold for other districts, echoing Rohani et al. (1999)'s finding that the introduction of the pertussis vaccine in England and Wales served to significantly increase the duration of each individual fadeout. The Cox regression analyses also indicated vaccination reduced the effect of geographical coupling on disease persistence in Lancashire, with increasing distance from endemic centres significantly associated with a fall in the rate of

reintroductions among hotspots during the vaccine era. Of the two regions, the impact of vaccination was considerably more dramatic in South Wales, with the endemic threshold population growing from 229,000 before routine pertussis vaccination to 1,461,000 in the vaccine era. This finding is unsurprising since the effective implementation of vaccination in more sparsely populated and widely dispersed communities effectively confined epidemic activity to Glamorgan, where a significant majority of the regional population resides in Cardiff CB, Swansea CB, and the Valleys urban areas. Limited connectivity across the wider region amplified the impact of vaccination, with districts such as Haverfordwest MB and Llandeilo MB, key local population centres in Pembrokeshire and Carmarthenshire, becoming isolated from the broader metapopulation in the vaccine era and no longer suggesting pertussis hotspots. In Lancashire, significant spatial interaction between endemic reservoirs, large towns, and surrounding communities' fuels rescue effects. Rescue effects restrict the growth of regional endemic threshold populations in the vaccine era with constant reinfection of districts. This is despite an effective vaccine uptake rate of approximately 80% in England and Wales after the onset of routine vaccination (Amirthalingam et al., 2013), which theoretically should have significantly reduced the rate of disease reintroduction in Lancashire by eliminating chains of transmissions.

Pertussis persistence dynamics in vaccine era South Wales and Lancashire generally conform to Bartlett (1960)'s model of disease persistence. The most populous settlements were more strongly coupled to the metapopulation at large, outweighing the adverse impact of distance decay. There is strong evidence that the Lancashire and South Wales regions represent classic examples of a mainland-island metapopulation, with source-sink dynamics explaining pertussis persistence and recurrent outbreaks of infection. These dynamics are clearly visible with the introduction of routine pertussis vaccination in 1957. The population flow between 'mainland' districts such as Manchester CB, Liverpool CB and Cardiff CB and 'island' districts below a median distance from endemic centres ensured pertussis persistence in these sub-populations via rescue effects.

The hhh4 models reveal commuter-driven spread as a key mechanism for disease persistence in both regional populations, with a gravity and power law model significantly improving the model fit to the observed data in successive time windows. In Lancashire and South Wales, disease persistence in communities with smaller populations was driven primarily by commuter traffic to and from endemic reservoirs, with incidence concentrated in districts with high levels of population mobility and spatial proximity. In South Wales, there was significant commuter spread between urban centres in Glamorgan and communities in the Valleys area. Districts immediately surrounding Liverpool and Manchester were of disproportionate importance in the spatial transmission of pertussis, suggestive of an 'agglomeration' effect highlighted in the vaccine era hotspot analysis. Waves of pertussis infection

radiated from the endemic reservoirs of Liverpool CB and Manchester CB to neighbouring districts which formed the urban overspill or suburbs of the conurbations. The consistent spatial interaction between Manchester CB, Liverpool CB, and surrounding settlements driven by commuter flows resulted in a positive feedback loop with both conurbations importing a large number of cases from neighbouring districts (Bartlett, 1957), amplifying already significant disease activity.

The hhh4 modelling framework has previously been applied as a standalone approach to describe, predict, and forecast influenza, measles, and meningococcal dynamics (Held and Paul, 2012; Meyer et al., 2014). For instance, the hhh4 model has been used to perform a spatio-temporal analysis of Influenza A spread in north Norway, with local air, road, and sea traffic data incorporated to explore the spatial dynamics of the disease (Geilhufe et al., 2014). An area for future work would be to explicitly account for the spatial impact of vaccination on pertussis dynamics in the vaccine era time windows using the hhh4 modelling framework. A standard hhh4 model could be extended to include regional vaccination coverage. The susceptible proportion of the population for each district across vaccine era time windows could be quantified by utilising district-level immunisation data abstracted from annual local Medical Officer of Health reports.

One limitation of the research presented concerns the uncertainties associated with the data used. Previous analyses have suggested that only 25% of the actual number of pertussis cases were notified and collated by the *Weekly Returns* (Clarkson and Fine, 1985), with sub-clinical cases of the disease being a common occurrence. Widespread under-reporting may result in an over-estimate of weeks experiencing fade-out. Additionally, the endemic threshold estimation also leaves out factors previously cited as potentially influencing threshold size, including seasonal term-time forcing, realistic age structure, and non-exponential waiting times. However, despite the relative crudeness of the 'moving window' approach, threshold estimates compare well with pre-vaccine and vaccine era estimates calculated for England and Wales in past work (Wearing and Rohani, 2009). Another limitation regards the fitting of hhh4 models. Ideally, the models would be stratified by age group since pertussis is predominantly spread by the mixing of school children, for whom social contacts are highly age-dependent. The nature of such contact acts to extend the pure distance decay of interaction.

The approach presented in this paper provides a straightforward method for analysing changes in regional disease persistence over time by studying long-term spatiotemporal changes in endemic threshold populations. This could be applied to other regional populations and the study of other vaccine-preventable, directly-transmitted childhood infections to uncover spatial heterogeneities in disease persistence. Recognising spatial heterogeneities such as rescue effects radiating from hotspots of epidemic activity is necessary to devise successful disease intervention strategies. Identifying and describing significant variations in threshold estimates for complex regional populations can better inform vaccination efforts in resource-constrained settings by highlighting the sometimes stark differences in persistence and invasion dynamics of a target disease in a mainland metapopulation. In such a context, district-based or regionally-targeted mass vaccination programmes might represent more effective strategies for disease elimination than nationwide mass vaccination.

Credit author statement

Alastair Munro: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualisation. Matthew Smallman-Raynor: Conceptualisation, Writing - review & editing, Supervision. Adam C. Algar: Conceptualisation, Writing - review & editing, Supervision.

Acknowledgements

This work was supported by the Economic and Social Research Council [grant number ES/J500100/1]. The authors are grateful to the two anonymous referees for their insightful comments and suggestions on an earlier version of this manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.socscimed.2020.113295>.

References

- Amirthalingam, G., Gupta, S., Campbell, H., 2013. Pertussis immunisation and control in England and Wales, 1957 to 2012: a historical review. *Euro Surveill.* 18, 20587. <https://doi.org/10.2807/1560-7917.ES2013.18.38.20587>.
- Anderson, R.M., 2016. The impact of vaccination on the epidemiology of infectious diseases. In: Bloom, B.R., Lambert, P.H. (Eds.), *The Vaccine Book, second ed.* Academic Press, London, pp. 3–31.
- Anderson, R.M., May, R.M., 1992. *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press, Oxford.
- Bartlett, M.S., 1956. Deterministic and stochastic models for recurrent epidemics. In: Neyman, J. (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability.* University of California Press, Berkeley, pp. 81–109.
- Bartlett, M.S., 1957. Measles periodicity and community size. *J. Royal Stat. Soc. A. Gen.* 120, 48–70. <https://doi.org/10.2307/2342553>.
- Bartlett, M.S., 1960. The critical community size for measles in the United States. *J. R. Stat. Soc. A. Gen.* 123, 37–44. <https://doi.org/10.2307/2343186>.
- Beyer, H.L., Hampson, K., Lembo, T., Cleaveland, S., Kaare, M., Haydon, D.T., 2011. Metapopulation dynamics of rabies and the efficacy of vaccination. *Proc. R. Soc. B.* 278, 2182–2190. <https://doi.org/10.1098/rspb.2010.2312>.
- Bivand, R.S., Pebesma, E., Gomez-Rubio, V., 2013. *Applied Spatial Data Analysis with R, second ed.* Springer, New York.
- Bharti, N., Djibo, A., Ferrari, M.J., Grais, R.F., Tatem, A.J., McCabe, C.A., Bjornstad, O. N., Grenfell, B.T., 2010. Measles hotspots and epidemiological connectivity. *Epidemiol. Infect.* 138, 1308–1316. <https://doi.org/10.1017/S0950268809991385>.
- Black, F.L., 1966. Measles endemicity in insular populations: critical community size and its evolutionary implication. *J. Theor. Biol.* 11, 207–211. [https://doi.org/10.1016/0022-5193\(66\)90161-5](https://doi.org/10.1016/0022-5193(66)90161-5).
- Bolker, B.M., Grenfell, B.T., 1996. Impact of vaccination on the spatial correlation and persistence of measles dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 93, 12648–12653. <https://doi.org/10.1073/pnas.93.22.12648>.
- Brouhin, H., Simondon, F., Guégan, J., 2004. Whooping cough metapopulation dynamics in tropical conditions: disease persistence and impact of vaccination. *Proc. R. Soc. B.* 271, S302–S305. <https://doi.org/10.1098/rspb.2004.0173>.
- Clarkson, J.A., Fine, P.E., 1985. The efficiency of measles and pertussis notification in England and Wales. *Int. J. Epidemiol.* 14, 153–168. <https://doi.org/10.1093/ije/14.1.153>.
- Cliff, A.D., Haggett, P., 1989. Spatial aspects of epidemic control. *Prog. Hum. Geogr.* 13, 315–347. <https://doi.org/10.1177/030913258901300301>.
- Cliff, A.D., Haggett, P., Smallman-Raynor, M., 1993. *Measles: an Historical Geography of a Major Human Viral Disease: from Global Expansion to Local Retreat, 1840–1990.* Blackwell, Oxford.
- Cliff, A.D., Haggett, P., Smallman-Raynor, M., 2000. *Island Epidemics.* Oxford University Press, Oxford.
- Cliff, A.D., Haggett, P., Stroup, D.F., Cheney, E., 1992. The changing geographical coherence of measles morbidity in the United States, 1962–88. *Stat. Med.* 11, 1409–1424. <https://doi.org/10.1002/sim.4780111102>.
- Czado, C., Gneiting, T., Held, L., 2009. Predictive model assessment for count data. *Biometrics* 65, 1254–1261. <https://doi.org/10.1111/j.1541-0420.2009.01191.x>.
- Earn, D.J., Rohani, P., Grenfell, B.T., 1998. Persistence, chaos and synchrony in ecology and epidemiology. *Proc. R. Soc. B.* 265, 7–10. <https://doi.org/10.1098/rspb.1998.0256>.
- Freeman, T.W., 1966. *The Conurbations of Great Britain.* Manchester University Press, Manchester.
- Grenfell, B.T., Harwood, J., 1997. (Meta) population dynamics of infectious diseases. *Trends Ecol. Evol.* 12, 395–399. [https://doi.org/10.1016/S0169-5347\(97\)01174-9](https://doi.org/10.1016/S0169-5347(97)01174-9).
- Grenfell, B.T., Bjornstad, O.N., Kappey, J., 2001. Travelling waves and spatial hierarchies in measles epidemics. *Nature* 414, 716–723. <https://doi.org/10.1038/414716a>.
- Geilhufe, M., Held, L., Skrvøseth, S.O., Simonsen, G.S., Godtliebsen, F., 2014. Power law approximations of movement network data for modelling infectious disease spread. *Biom. J.* 56, 363–382. <https://doi.org/10.1002/bimj.201200262>.
- Hagenaars, T.J., Donnelly, C.A., Ferguson, N.M., 2004. Spatial heterogeneity and the persistence of infectious diseases. *J. Theor. Biol.* 229, 349–359. <https://doi.org/10.1016/j.jtbi.2004.04.002>.
- Held, L., Paul, M., 2012. Modelling seasonality in space-time infectious disease surveillance data. *Biom. J.* 54, 824–843. <https://doi.org/10.1002/bimj.201200037>.
- Held, L., Höhle, M., Hofmann, M., 2005. A statistical framework for the analysis of multivariate infectious disease surveillance counts. *Stat. Model. Int. J.* 5, 187–199. <https://doi.org/10.1191/1471082X05st098oa>.

- Jandarov, R., Haran, M., Bjørnstad, O., Grenfell, B., 2014. Emulating a gravity model to infer the spatiotemporal dynamics of an infectious disease. *J R Stat Soc Ser C Appl Stat* 63, 423–444. <https://doi.org/10.1111/rssc.12042>.
- Jenkins, P., 1992. *A History of Modern Wales, 1536–1990*. Longman, London.
- Jones, P.N., 1969. *Colliery Settlement in the South Wales Coalfield, 1850–1926*. University of Hull Publication, Hull.
- Lavine, J.S., Rohani, P., 2012. Resolving pertussis immunity and vaccine effectiveness using incidence time series. *Expert Rev. Vaccines* 11, 1319–1329. <https://doi.org/10.1586/erv.12.109>.
- Metcalf, C.J.E., Hampson, K., Tatem, A.J., Grenfell, B.T., Bjørnstad, O.N., 2013. Persistence in epidemic metapopulations: quantifying the rescue effects for measles, mumps, rubella and whooping cough. *PLoS One* 8, E74696. <https://doi.org/10.1371/journal.pone.0074696>.
- Meyer, S., Held, L., 2014. Power-law models for infectious disease spread. *Ann. Appl. Stat.* 8, 1612–1639. <https://doi.org/10.1214/14-AOAS743>.
- Meyer, S., Held, L., Höhle, M., 2017. Spatio-temporal analysis of epidemic phenomena using the R package surveillance. *J. Stat. Software* 77, 1–55. <https://doi.org/10.18637/jss.v077.i11>.
- Milward, G., 2019. *Vaccinating Britain: Mass Vaccination and The Public Since the Second World War*. Manchester University Press, Manchester.
- Paul, M., Held, L., 2011. Predictive assessment of a non-linear random effects model for multivariate time series of infectious disease counts. *Stat. Med.* 30, 1118–1136. <https://doi.org/10.1002/sim.4177>.
- Rohani, P., Earn, D.J., Grenfell, B.T., 1999. Opposite patterns of synchrony in sympatric disease metapopulations. *Science* 286, 968–971. <https://doi.org/10.1126/science.286.5441.968>.
- Schenzle, D., Dietz, K., 1987. Critical Population Sizes for Endemic Virus Transmission. In: Fricke, W., Hinz, E. (Eds.), *Räumliche Persistenz und Diffusion von Krankheiten*, 83. Heidelberg Geographical Studies, Heidelberg, pp. 31–42.
- Warfel, J.M., Beren, J., Merkel, T.J., 2012. Airborne transmission of Bordetella pertussis. *J. Infect. Dis.* 206, 902–906. <https://doi.org/10.1093/infdis/jis443>.
- Wearing, H.J., Rohani, P., 2009. Estimating the duration of pertussis immunity using epidemiological signatures. *PLoS Pathog.* 5, e1000647 <https://doi.org/10.1371/journal.ppat.1000647>.
- Xia, Y., Bjørnstad, O.N., Grenfell, B.T., 2004. Measles metapopulation dynamics: a gravity model for epidemiological coupling and dynamics. *Am. Nat.* 164, 267–281. <https://doi.org/10.1086/422341>.