Long-Term Spatiotemporal Changes in Endemic Threshold Populations in England and Wales – A Multi-Disease Study

by Alastair Donald Munro, BA (Hons), MSc

(University of Nottingham)

Thesis submitted to the University of Nottingham,

for the Degree of Doctor of Philosophy

March 2021

Word count: 100,000

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

CB	County Borough
CCS	Critical Community Size
LGD	Local Government District
MB	Municipal Borough
RD	Rural District
UD	Urban District
CB	County Borough
TFR	Total Fertility Rate
MOH	Medical Officer of Health
PL	Power-Law
RI	Random Effects
RGWR	Registrar–General's *Weekly Return*

ACKNOWLEDGEMENTS

I would first like to express my profound gratitude to my supervisors, Professor Matthew Smallman-Raynor and Dr Adam Algar, for their constant support, guidance and patience over the last four-and-a-bit years. I am extremely grateful for the time they have invested in me, the valuable insights they have shared, and the encouragement they have provided, time and time again. Their supervision has gently pushed me forwards year-after-year and provided me with the confidence to continue challenging myself and make the most of my ability.

I am very grateful to Professor Sarah Metcalfe for her much appreciated comments and suggestions during past annual reviews, providing a different perspective on ways to approach my PhD research. I also appreciate the efforts of all those at the School of Geography who have contributed in various ways towards the completion of my research and this thesis.

I would like to sincerely thank all those involved at the ESRC DTC at the University of Nottingham for their support since I was awarded the 1+3 studentship in 2015. Without their generous financial support, it simply would not have been possible for me to pursue a postgraduate education at this stage of my life, let alone undertake the research documented in this thesis. This work was supported by the Economic and Social Research Council [grant number ES/J500100/1].

Finally, I would like to dedicate this work to my parents, Paul and Julie. Their nurturing, steadfast support and tremendous patience has been invaluable to me during this research, particularly in the more challenging and sometimes unforgiving moments that I have encountered along the way to completing my doctoral studies.

ABSTRACT

Metapopulation dynamics play a critical role in driving endemic persistence and transmission of childhood infections. The endemic threshold 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 geographical investigations of endemic threshold populations have tended to focus on isolated populations in island communities. Few studies have examined endemic threshold dynamics in 'mainland' regional populations with divergent spatial structures and varying levels of connectivity between subpopulations.

This thesis presents a geographical analysis of spatiotemporal changes in endemic threshold populations for three childhood infections (measles, pertussis and scarlet fever) in two regional metapopulations 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), empirical regression techniques were used to estimate the endemic threshold populations for childhood infections in the two study regions. Hotspot and survival analyses were performed to compare disease fadeout duration and probability for both regions in the pre-vaccine and vaccine eras, respectively. Endemic-epidemic modelling was undertaken to identify and analyse potential drivers of disease persistence.

The findings reveal strong regional differences in estimates of endemic threshold populations over time and space for all three childhood infections. Regional differences in endemic threshold populations reflect significant regional variations in spatial connectivity, population dispersion and level of geographical isolation. 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.

Chapter 1: Introduction & Review

1 Background

The endemicity of childhood infectious diseases, such as those which blighted the public health of industrialised societies in the early-twentieth century, remains an area of considerable interest for geographers engaged in the study of epidemiological theory from a spatial perspective. According to Dietz (1995), determining the endemicity of infections can make a valuable contribution to global disease eradication programmes. For instance, identifying the urban centres and regions which facilitate the persistence of an infection 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 and control of an infection is that of the endemicity threshold, which has previously received mention 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; Murray and Cliff, 1977; Trevelyan et al., 2005).

Most commonly referred to as 'critical community size' in epidemiological literature, an endemic threshold is the minimum population required for an infection to persist endemically within a geographical area (Schenzle and Dietz, 1987), without the reintroduction from external sources. According to the seminal work of Bartlett, stochastic processes play a fundamental role in determining the endemic threshold size (Anderson, 2016; Metcalf et al., 2013a). The persistence of infection, based on the stochastic models formulated by Bartlett (1957, 1960), strongly implies a spatial transmission of infection between geographical units as the population size of cities and towns fall below the endemic threshold, with hierarchical travelling waves spreading across the landscape from large urban centres. These settlements act as 'endemic reservoirs' which maintain the persistence of infection, re-infecting regions where disease has either been locally eliminated or faded out (Cliff and

Haggett, 1989). However, the value of the threshold population also depends on the spatial structure of communities and other heterogeneities, as well as on various epidemiological parameters such as infectious and latent periods (Keeling and Grenfell, 2007). The endemic threshold size concept has played a central role in a vast number of studies which seek to understand the spatiotemporal dynamics and endemic persistence of infection, most notably in the sub-field of measles studies. Numerous epidemiological models have been developed over recent decades attempting to accurately calculate the endemic threshold value for measles and other infections, accounting for mechanisms such as seasonality, age structure and spatial heterogeneity (Aron and Schwartz, 1984; Keeling and Grenfell, 2002; Bolker and Grenfell, 1996).

Over recent decades, the study of disease persistence and critical community size from a spatial perspective has largely involved the use of metapopulation models formulated by ecologists and population biologists. These studies have tended to focus on analysing the temporal patterns of fadeouts in greater detail, explaining the spatiotemporal persistence of measles in England and Wales as a function of urban population size (Earn et al., 1998; Grenfell and Harwood, 1997; Harrison, 1991). The work of Grenfell and Bolker (1998) supports the conclusions of Bartlett (1957, 1960), finding measles to persist during inter-epidemic periods in large urban centres with an endemic threshold population above 300,000 during the pre-vaccination era, with prolonged periods of endemic fadeouts mostly occurring in small communities. They interpreted their findings as evidence that a spatial hierarchy in the host population structure served as a vital prerequisite for the measles epidemic waves which travelled the length and breadth of England and Wales during the pre-vaccine era. Exploring the implications led by the endemic threshold concept for the metapopulation dynamics of measles, Grenfell and Harwood (1997) discuss the endemic persistence of measles in large settlements (‘core patches’), and regular fade-out at a local-level in smaller communities (‘satellite patches’) as representing a mainland-island metapopulation, with source-sink dynamics explaining the persistence and recurrent

outbreaks of infection. Bolker and Grenfell (1996) highlighted the importance of spatial decorrelation between the measles epidemics of 10 major urban centres in England after the onset of mass vaccination in sustaining the persistence of infection and the pre-vaccine era endemic threshold size for the infection. The spatiotemporal decorrelation of city epidemics after the introduction of vaccination coupled with the stability of the endemic threshold value suggested the occurrence of a 'rescue effect' between cities, maintaining the circulation of infection. This finding indicates that rescue effects alongside epidemiological coupling between geographical units may prevent positive changes in endemic threshold population size. Undoubtedly, metapopulation models that incorporate the spatial structure of regional populations provide a template for studies which aim to develop an explicitly geographical understanding of disease persistence.

Yet, according to Bolker and Grenfell (1996), endemic threshold size is inherently geographical, ultimately dependent on local spatial structure and the connectedness of regional populations. Geographical research on endemic infections has attempted to understand the processes that enable this spatial transfer of infection, as such knowledge may aid the development of control strategies that could be implemented to interrupt the disease diffusion process (Cliff et al., 1992), helping to facilitate the elimination of infectious disease for which the tools exist to do so. Developing an analytically sound platform for understanding the geographical nature of epidemiological data within a modelling framework is ultimately crucial to realising these ambitions.

This introductory chapter will present a detailed review of the empirical and theoretical literature concerning the concept of endemic thresholds. This will be followed by a statement of research; a statement of the problems the thesis seeks to address, a justification for the research, aims and objectives of the thesis as well as stating the research questions that seek to be answered, before finally presenting a chapter outline for the thesis.

1.1 Literature Review

The literature review is split into three parts. Part one introduces and unpacks the study of endemic thresholds, discussing seminal work on critical community size and disease persistence, and detailing previous empirical studies of endemic thresholds for childhood diseases. This is followed by a discussion of the relationship between endemic thresholds and vaccination. The second part details the quantification and modelling of endemic thresholds, with reference to empirical work, and discusses the utility of metapopulation theory to inform analysis of endemic thresholds. The final part of the chapter provides an account of the epidemiology of the three childhood infections under analysis.

Part One: The Study of Endemic Thresholds

1.2 Understanding Endemic Threshold Populations

According to Bolker and Grenfell (1996), the size of endemic threshold populations is inherently geographical, ultimately dependent on the spatial structure and connectedness of a regional population. It has been recognised that the persistence of infection, based on the stochastic models formulated by Bartlett (1957, 1960), implies the spatial transmission of infection between geographical units as the population size of cities and towns fall below the endemic threshold, with hierarchical travelling waves spreading across the landscape from large urban centres, the engines of infection. Geographers have in the past aimed to identify ‘endemic reservoirs’ which maintain the persistence of infection, re-infecting regions where disease has either been locally eliminated or faded out (Cliff and Haggett, 1989). Geographical research on endemic infections has attempted to understand the processes which enable this spatial transfer of infection, as such knowledge may aid the development of control strategies that could be implemented to interrupt the disease diffusion process (Cliff et al., 1992), facilitating the elimination of infectious disease. Key to achieving these

endeavours is the establishment of an analytically sound platform for understanding the geographical nature of epidemiological data within a modelling framework.

Several factors may affect the estimates for endemic threshold size. Firstly, subpopulations within the study area may artificially inflate the threshold value if most contact between susceptibles and infected individuals occurs in these subpopulations, rather than in a large homogenously mixed population. This effect generally increases in line with the number of subpopulations in existence. The likelihood of subpopulation mixing increases as the study area is extended geographically to include several towns and cities (Cliff et al., 1993). Related to this is the issue of geographical isolation, which reduces the level of mixing between susceptibles and infected persons compared to that expected in a single homogenously mixed population. Population density and turnover may also play influential roles in determining the value of endemic threshold size (Cliff et al., 2000). High birth rates produce a large pool of susceptibles, enhancing the likelihood that a disease will be able to persist endemically (Bartlett, 1957). High population density increases the transmission probability between susceptible and infected individuals, as well as the speed at which disease spreads and fades out, complicating the estimation of endemic threshold size (Black, 1966). Underreporting however is a key issue with regards to the calculation of endemic threshold size, since it may bias such estimates by suggesting apparent fade-outs where there were cases of infection which went unobserved (Metcalf et al., 2013a).

Endemic thresholds have been characterised empirically for relatively few childhood diseases (Hanski and Gaggiotti, 2004). By far the most well-documented of these is the endemicity threshold for measles, for which persistence of the infection is considered a function of urban population size (Grenfell and Harwood, 1997).

1.3 Study of Endemic Thresholds: Measles

In the early 20th century, it was recognised that in order to explain the recurrence of outbreaks of childhood infectious diseases, both demographic and epidemic forces had to be accounted for (Hamer, 1906). A deterministic theoretical model for recurrent measles epidemics, the Hamer–Soper model, was forwarded by Herbert Soper in 1929. This epidemic model predicted measles would approach an endemic stable equilibrium level above a threshold through the damping of epidemic oscillations, as a result of shorter periods of disease infectivity. The infectivity of measles, according to Soper (1929), begins instantly when an individual is infected and is spread over time. This leads to a decline in the incidence of the disease, thus indicating the damping of epidemic waves. Fundamentally, the Hamer–Soper model suggests disease extinctions do not occur at all (Nåsell, 2005). It is therefore insufficient for explaining recurrent epidemics of infection. Wilson and Worcester's (1945) state the incidence of recurrent outbreaks of measles provide no tangible evidence to support the theory of damping of epidemic oscillations. Bartlett (1956) believed that the Hamer–Soper model was problematic since it failed to account for demographic stochasticity, the inherent unpredictability in the timing and nature of births, mortality and migration (Conlan and Grenfell, 2007). This resulted in major inconsistencies between theoretical models of measles cases and recorded observational measles data.

Bartlett (1957) proposed a stochastic reformulation of the Hamer–Soper model (Nåsell, 2005), emphasising two key features: the theoretical tendency for successive epidemics in large communities to damp down could be offset by random variability, and the tendency for diseases to fade-out in small or isolated communities when the number of susceptibles had dropped below an infection's threshold value. It is the latter that led to Bartlett's (1957) introduction of the notion of 'critical community size', a threshold concept defined as the smallest host population size above which an infection can persist endemically.

1.3.1 Maurice Bartlett

Bartlett (1956) was first to observe that time to extinction for measles was an increasing function of the community size. In a seminal work, Bartlett (1957) explored the periodicity of measles in relation to community size using a sample of 19 English towns of varying population size (see Fig. 1.1). Measles notifications for each town were extracted from the Registrar General's Weekly Returns for England and Wales in the period 1940-1956.

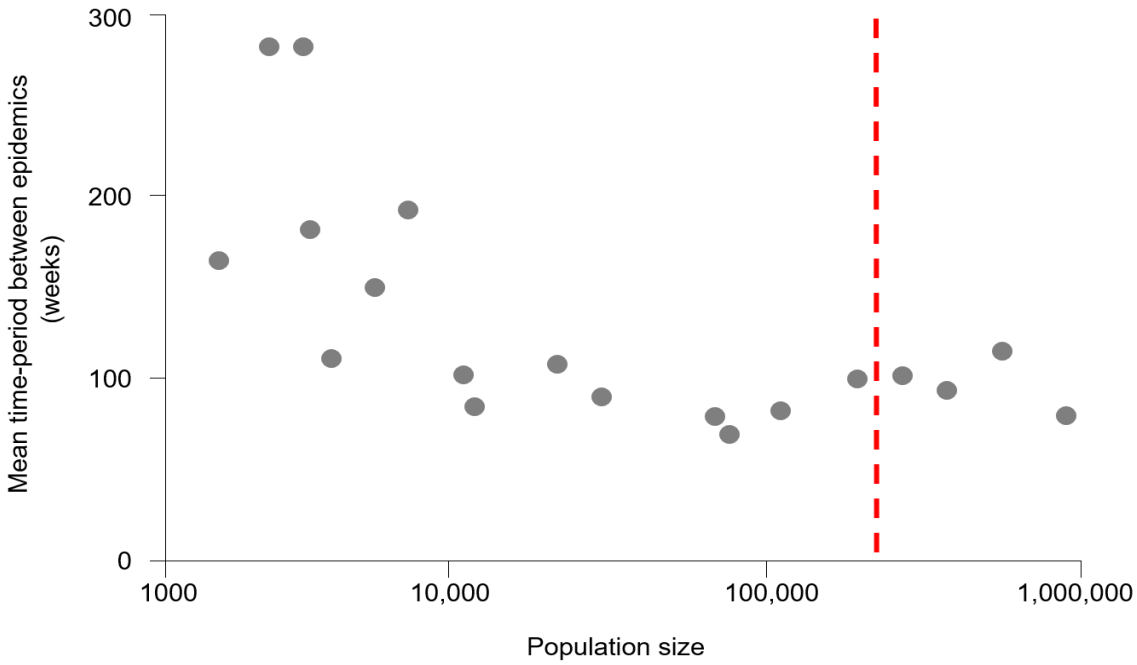

Figure 1.1 *Impact of population size on the periodicity of measles epidemics for 19 English and Welsh towns and cities. Red-sashed line marks the endemic threshold value (Adapted from Cliff et al., 2000: 89).*

Bartlett observed that large towns, such as Birmingham, Bristol and Hull, experienced recurring measles outbreaks with no fade-outs (a period of three or more consecutive weeks without a measles notification) after epidemics. In these towns, measles circulates endemically, and they were therefore considered to be of 'critical size'; established to be in the order of 250-300,000. Towns below the threshold population total, such as Carlisle and

Barrow-in-Furness, experienced a complete endemic fade-out of measles in troughs between epidemics. The smallest and most isolated towns with a population below 10,000, such as Cardigan and Llanrwst, were found to have extremely irregular, epidemic patterns. They would often go several years without experiencing a measles outbreak, indicating that the infection suffers stochastic extinction in troughs between epidemics in small communities (Earn et al., 1998), until its reintroduction from outside sources. Bartlett estimated 2,500 cases per annum was the minimum needed for measles to persist within urban areas during troughs in epidemic cycles.

Bartlett's (1957) identification of a threshold population size below which infection would fade-out is critical in understanding the recurrent epidemic patterns of measles observed in England and Wales during the pre-vaccine era (see Fig. 1.2). For settlements with a population size below the endemic threshold value where infection has faded out, outbreaks can only occur if the infection is reintroduced by index cases (infected individuals) via Type I epidemic waves emanating from the largest population centres above the threshold level (e.g., Birmingham, Hull), which act as reservoirs of disease (Lloyd and Sattenspiel, 2009). Type I epidemic waves result in fade-out free epidemics in populations above the endemic threshold. In intermediate-sized towns below the endemic threshold (e.g., Carlisle, Barrow-in-Furness), discrete but regular type II epidemic waves are observed, with recurrent outbreaks in sync with those experienced in the large urban centres. Irregularly spaced type III epidemic waves affect communities with the smallest populations (e.g., Cardigan, Llanrwst), resulting in sporadic outbreaks and extended periods of endemic fade-out (Cliff and Haggett, 1989). This pattern of recurrent epidemics, and the generalised persistence of measles is a clear illustration of how endemics infections are geographically transmitted between populations of varying size within regions.

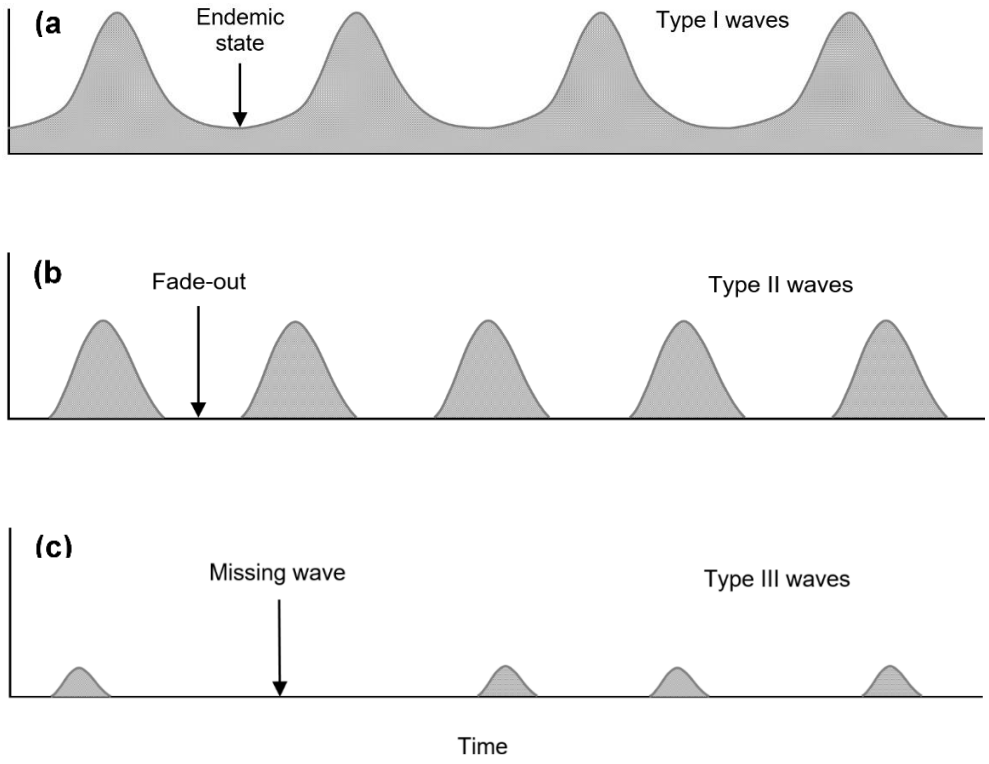

Figure 1.2 *Bartlett model of epidemic patterns of measles spread according to varying population size. (a) Type I waves affect large settlements above the endemic threshold, follow a regular cycle, and infection does not become extinct in the inter-epidemic period, (b) type II waves affect intermediate sized towns below the endemic threshold value and mirror type I waves in regularity, but infection fades out between epidemic outbreaks, (c) type III waves affect small settlements which tend to experience highly irregular epidemics with extensive periods of disease fade-out (Source: Cliff et al., 2000: 89).*

In a subsequent paper aiming to determine the critical community size for measles within the context of the urban United states, Bartlett (1960) calculated the endemic threshold size to be around the figure of 300,000 inhabitants using data on measles for 24 North American cities between 1921 and 1940. This was in broad agreement with the results of his previous paper on the endemicity threshold for measles in English and Welsh settlements. Bartlett (1960) also found that measles fade-outs would occur in cities where there were fewer than 4,000 to 5,000 cases per annum.

1.3.2 Francis Black

Building on the foundational work of Bartlett, Francis Black (1966) made a seminal contribution to the study of endemic disease, seeking to establish the threshold endemicity of measles in insular communities. Black believed Bartlett had failed to correct for the masking of fade-out in cities by the reintroduction of measles from external sources, and the damping effect of geographic dispersion, and therefore sought to confirm and refine Bartlett's estimates concerning the population thresholds below which measles would fade-out in towns and cities. Monthly measles reports from 19 island communities between 1949 and 1964 were analysed, and frequent extinctions of measles were found among all insular communities under study, apart from Hawaii (see Fig. 1.3).

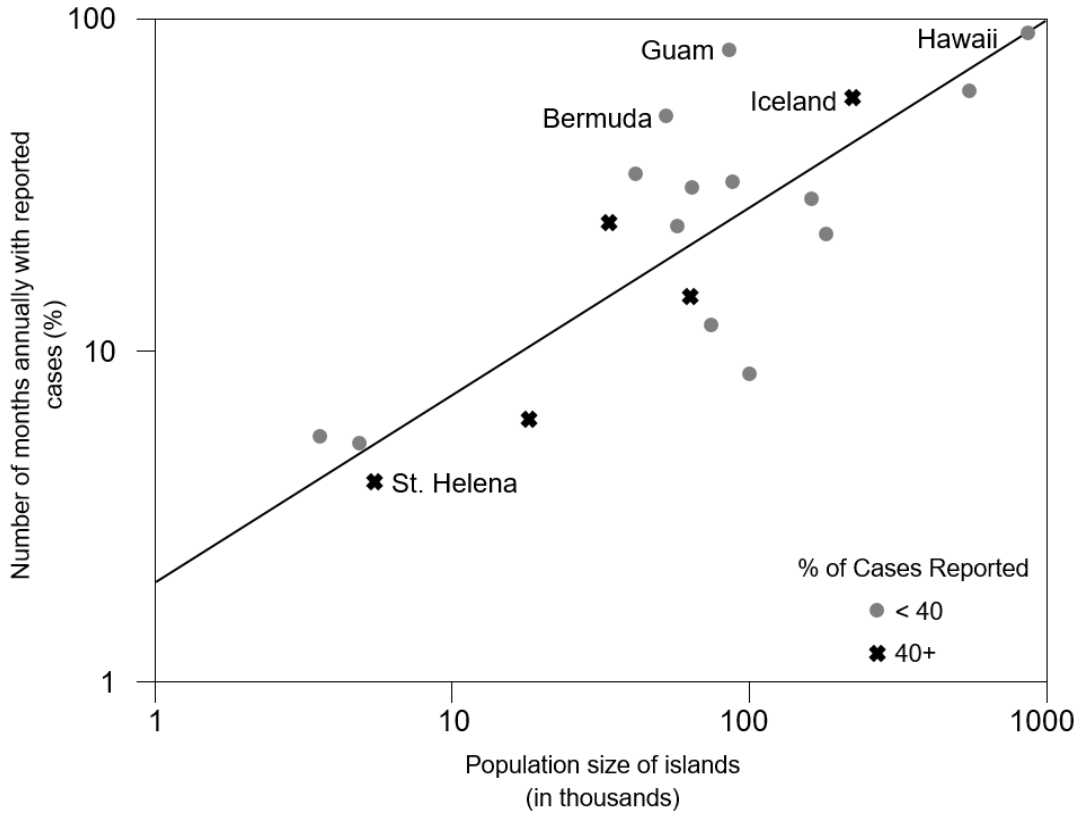

Figure 1.3 *Black's model of the relationship between the duration of fade-outs after epidemics and population size for 19 island communities between 1949–1964 (Adapted from Cliff and Haggett, 1989: 320).*

Frequent air travel between Hawaii and the rest of the United States, as well as the island's transient military population, is cited as a key factor in the reintroduction of measles, masking fade-outs that have occurred on the island. Due to the dubious quality of case reporting on other islands with populations comparable in size to Hawaii's, such as Fiji and Mauritius, Black was unable to test this hypothesis with confidence. Black states that measles may fade-out in island populations as large as 350,000, possibly over 500,000, if closely settled and without the re-introduction of the disease from outside. These endemic thresholds are similar to the critical population size for measles in UK and US cities forwarded by Bartlett (1957, 1960). However, depending on the spatial structure and connectedness of the island population, Black notes that measles may be able to persist in smaller populations, but not endemically in communities with less than 5,000 cases per annum.

Since the seminal works of Bartlett (1957, 1960) and Black (1966), the concept of critical community size has been much-cited in the epidemiological study of the measles, providing a theoretical basis for research exploring the geographical spread, patterns and persistence of the infection.

1.4 Study of Endemic Thresholds: Other Childhood Infections

As illustrated in the previous section, much of the extant literature on endemicity thresholds is mostly discussed and examined in relation to the spatial and population dynamics of measles. The commonly cited threshold size of 250,000-500,000 for measles however does not necessarily extend to other directly transmitted childhood infections, such as pertussis and poliomyelitis. Cliff and Haggett (1989) provide theoretical estimates for the endemic threshold size of five viral and bacterial diseases in relation to their serial interval, which is defined as the average time between the observation of symptoms in one case of infection

and symptoms in a second case directly infected by the first. These are summarised in Table 1.1.

Table 1.1 *Relationship between infectious periods of five directly transmitted viral and bacterial infections and estimated endemic threshold size (Source: Cliff and Haggett, 1989: 321).*

Disease	Estimated infectious period (Number of days)	Theoretical threshold population
Hepatitis A	35	22,000–152,000
Rubella	18	132,000
Pertussis	14	150,000–200,000
Measles	12	250,000–500,000
Influenza	4	1,000,000,000

1.4.1 Pertussis

Wearing and Rohani (2009) plotted the proportion of weeks with zero cases of pertussis against population size, the most common measure of disease fadeout, exploring whether endemic threshold size can act as a signature for waning immunity (whether that be naturally acquired or through vaccination). In the pre-vaccine era, analysis of fadeouts of pertussis in 50 towns and cities in England and Wales, varying in population size from 75,000 to 1,500,000, suggested an endemic threshold size between 150,000 and 250,000. After the onset of mass vaccination in 1957, Wearing and Rohani's (2009) data suggests an endemic threshold size in the region of 800,000 to 1,000,000 is required to maintain the

endemic persistence of pertussis, consistent with a substantially increased period of immunity rather than permanent immunity. This is in line with the findings of Rohani et al., (2000) who, in their paper on the impact of immunisation on pertussis transmission, demonstrated a significant increase in the observed endemic threshold size for pertussis in England and Wales after the introduction of mass vaccination for the infection (see Fig. 1.4).

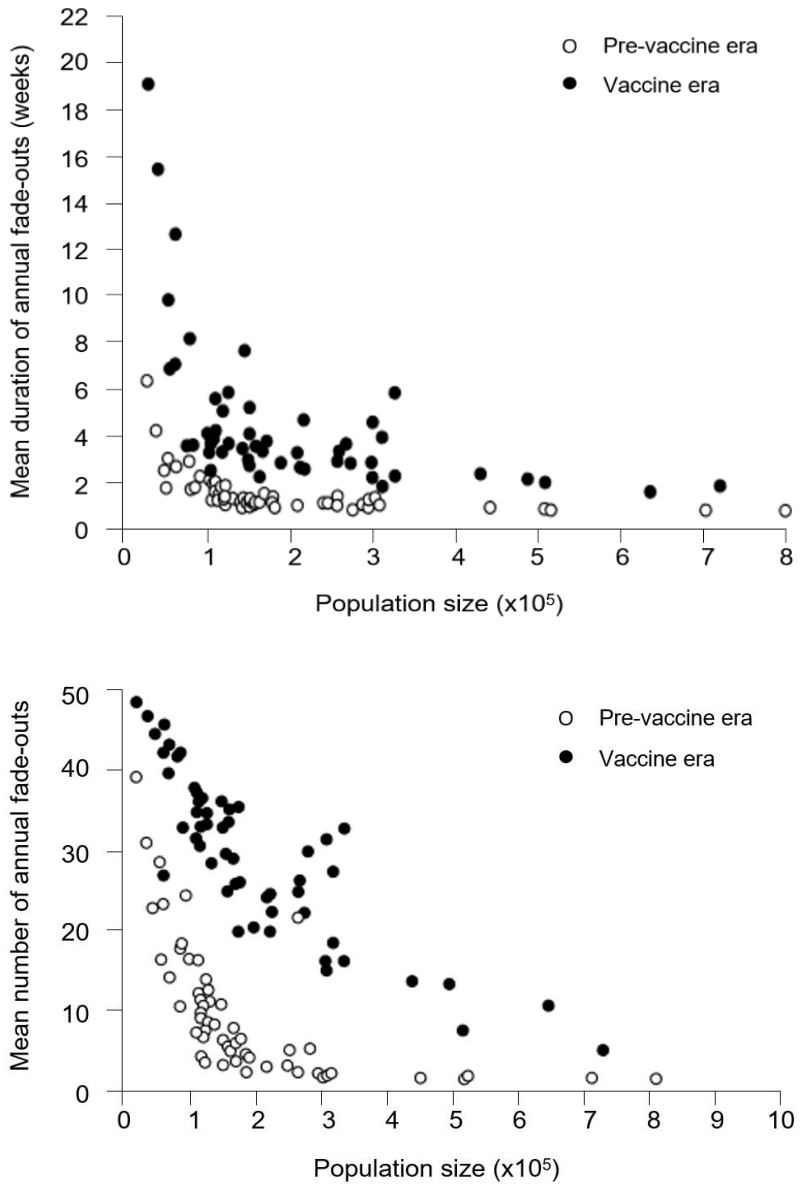

Figure 1.4 *Mean number and duration of annual fade-outs of pertussis against population of 60 towns and cities in England and Wales, 1944–1994 (Source: Rohani et al., 2000: 285).*

1.4.2 Poliomyelitis

Estimating the endemic threshold size for poliomyelitis is not a straightforward endeavour compared to other childhood infections (Cliff and Haggett, 1995), complicated by two major factors. Firstly, there are three distinct types of polioviruses and, secondly, a large number of cases do not present symptoms (Cliff et al., 2000). They therefore go unreported or misidentified. Despite these issues, Eichner et al. (1995) conducted stochastic simulations to estimate the minimum population size required to maintain the endemic persistence of poliomyelitis. They found that the endemic threshold size for the virus in regions with a relatively high standard of hygiene, where the opportunities for the faecal-oral transmission of poliomyelitis is limited, is of the order of 200,000–500,000. This contrasts with populations with poor standards of hygiene, where the endemic threshold size was found to be as low as 50,000, with the upper bound at 150,000. However, the simulations carried out by Eichner et al. (1995) also suggest that the low values estimated for the endemic threshold size of poliomyelitis are determined more by high population turnover (due to a rapid influx of susceptibles via a high birth rate) rather than poor standards of hygiene.

1.5 Temporal Changes in Endemic Thresholds: Island Populations

Cliff and Haggett (1995) and Cliff et al. (2000) discuss the utility of islands as laboratories to analyse the endemic threshold concept and to explore the relationship between endemicity and population size at the macro-geographical (through inter-island comparisons), meso-geographical (through time-series analysis of single islands), and micro-geographical levels (through within-island comparisons).

Building on the work of Black (1966) (Section 1.3.2), Cliff et al. (2000) explored changes in endemic threshold size over time at the meso-geographical level for just one of these

islands: Iceland. Monthly measles data were collected for 47 medical districts, for which an unbroken sequence of reporting exists from January 1888 to December 1987. These medical districts range significantly in population size from 100,000 in the district located immediately around the capital, Reykjavik, in 1990 to as low as 40 in an island medical district. The average endemic thresholds size over the course of 100 year-long study period was calculated to be around the figure of 259,000, a value consistent with the findings of both Bartlett (1957, 1960) and Black (1966). To monitor temporal changes over the period, the study period was broken down into a series of 8-year time–windows with a 2 year overlap between windows to produce a smoothing effect with the preceding and succeeding windows. This approach yielded 24 96-month time–windows, providing 24 thresholds estimates between the beginning of 1888 and end of 1987. The percentage of months in which cases were reported (percentage endemicity) along with their mean populations in the 47 districts in each time–windows form the basis of the temporal analysis of endemic threshold size. Cliff et al. (2000) found the highest estimates of endemic threshold size at the beginning and end of the study period. In the case of the former, they suggest the high threshold value is the result of Iceland's relative geographical isolation during this period, whilst high values calculated in the 1970s and 1980s indicate impact of mass vaccination.

Adopting the same methodological approach, Cliff et al. (2000) examine temporal changes in the endemic threshold size at the global level, by analysing the percentage endemicity of four childhood infections (measles, pertussis, diphtheria and scarlet fever) for 84 island populations using data extracted from archival records of the League of Nations and World Health Organisation. Between 1923 and 1990, the endemic threshold size was calculated for 12 time–windows of varying length, as well as three larger windows. Except for pertuusis after 1980, the three other infections display a similar pattern of sharply rising threshold values during the control periods. The pattern for pertussis is more complex, rising rapidly to 1980 before falling thereafter. Prior to vaccination and the widespread use of antibiotics, endemic threshold values for measles and scarlet fever had shown signs of increase.

The work of Cliff et al. (2000) on temporal changes in endemic threshold size in an island population provides a methodological template for examining long term temporal changes in endemic threshold populations. However, in the context of England and Wales, where the temporal changes in endemic threshold size for notifiable infections are yet to be studied in any great detail, unlike many island populations, settlements are strongly bound together by the movement of people and spatial coupling. This movement is hierarchically structured, and regular spatial flows from one geographical unit to another are known to reintroduce infection after endemic fade-outs in regions where infection is yet to be eliminated. It is therefore important to explore temporal changes in endemic threshold population size in relation to the spatial dynamics of infection, to understand the effect of spatial coupling, distance and movement between populations on the persistence of infection, since the rate and scaling of import of infected individuals has the ability to change the nature of the relationship between zero-incidence and population size (Conlan et al., 2009).

1.6 Vaccination & Endemic Thresholds

1.6.1 Theory

in the past, simple theoretical studies have suggested that vaccination should increase the endemic threshold population (Cliff et al., 2000), above which an infection can persist indefinitely. If vaccination acts to block the chain of transmission of an infection, the introduction of mass immunisation programs is predicted to increase endemic threshold size; a larger population will therefore be required to prevent the local elimination of the disease (Lavine and Rohani, 2012). For instance, in the case of measles, vaccination should multiply the endemic threshold size by a factor of $1/x^2$ (Griffiths, 1973). Therefore if 50% of the population is immunised against measles, the endemic threshold will rise from approximately 250,000 to 1,000,000; with 90% immunisation, the endemic threshold would be significantly increased to a population size of 25,000,000. Crucially, if vaccine uptake within a community is great enough to prevent R_0 from exceeding unity in value an infection will not be able to become endemic (Jansen et al., 2003).

1.6.2 Onset of Mass Vaccination

Mass vaccination programmes play a vital role in disease control strategies designed to eliminate contagious bacterial and viral childhood diseases. In simple terms, mass vaccination is aimed at reducing the incidence level of childhood diseases by slowing down the build-up of susceptibles in the general population, effectively reducing the recruitment rate of the infections (Earn et al., 1998). Vaccination in the community is the most effective mean for blocking the local chain of transmission.

The introduction of mass vaccination programmes for once-prominent childhood infections such as measles, diphtheria, pertussis and poliomyelitis from the mid-twentieth century in

developed nations, and the sustained immunisation efforts since, is widely considered to have had the most profound impact of all factors on reducing the transmission and endemicity of such diseases in industrialised societies (see Table 1.2) (Anderson, 2016; Roush et al., 2007). To ensure major epidemic outbreaks of childhood disease are prevented, the immunisation of large numbers of young children before they are exposed to natural infection in the community is vital, alongside sustaining a sufficiently high level of vaccination coverage year after year (Fine and Clarkson, 1982; Beyer et al., 2012).

Table 1.2 *Historical comparison of morbidity and mortality for selected vaccine-preventable childhood diseases in the United States, with dates of onset of mass vaccination for each infection (Source: Roush et al., 2007: 2156).*

Vaccine-preventable Diseases	**Pre-vaccine era Annual Average**		**Dates of Vaccine Licensure & Introduction**	**Most recent morbidity & mortality reports**		**Reduction in No. of Cases Since Vaccine Introduction (%)**	**Reduction in No. of Deaths Since Vaccine Introduction (%)**
	Cases	Deaths		Cases (2015)	Deaths (2015)		
Diphtheria	21,053 (1936-45)	1,822 (1936-45)	1928-1943	0	0	100	100
Measles	530,217 (1953-62)	440 (1953-62)	1963, 1967,1968	188	1	99.9	99.9
Pertussis	200,752 (1934-43)	4,034 (1934-43)	1914-1941	18,166	6	90.9	99.8
Acute Poliomyelitis	19,794 (1941-50)	1,393 (1941-50)	1955, 1961-1963, 1987	0	0	100	100
Paralytic Poliomyelitis	16,316 (1951-54)	1,879 (1951-54)	1955, 1961-1963, 1987	0	0	100	100
Smallpox	29,005 (1900-49)	337 (1900-49)	1798	0	0	100	100

Numerous empirical studies have demonstrated that the introduction of mass vaccination programmes in developed nations has successfully diminished the amplitude of epidemics

for several once-prominent childhood diseases, with massive reductions recorded in morbidity and mortality attributed to infections such as measles, poliomyelitis and pertussis, as well as changing their phase and periodicity (Anderson and May, 1982, 1992; Bolker and Grenfell, 1996; Griffiths, 1973; Magpantay and Rohani, 2015). Roush et al. (2007) conducted a simple and concise study comparing the historical morbidity preventable childhood infections in the US, which illustrates immense success of vaccination in the fight against infectious disease. For instance, they reveal in the pre-vaccine era, an average of 530,000 measles cases and 440 measles-related deaths were reported annually in the US; in 2006, the incidence rate had declined by 99.9% nationwide, and no deaths from measles were recorded (Table 1.2).

1.6.3 Towards a Geographical Understanding of Vaccination

Many studies in the epidemiological literature have revealed that the introduction of mass vaccination has resulted in dramatic shifts in the spatiotemporal dynamics of childhood infections from the pre-vaccine to the vaccine era. The decorrelation of epidemics between major urban centres may hold significant consequences for public health systems hoping to eliminate childhood infections via mass immunisation efforts. The likelihood of simultaneously eliminating an infection across all communities in a regional population is significantly reduced by the elimination of large epidemics, which served to synchronise population dynamics at the inter-city level in the pre-vaccine era (Allen et al., 1993). Continuous vaccination at an intermediate level may lead to occasional fadeouts of infection within large urban centres, but it also reduces the correlation of infective densities and peaks of infection at the inter-city level, as the total number of susceptibles within the population decreases, and the level of epidemiological coupling falls (Earn et al., 1998).

To understand the potential implications of spatiotemporal decorrelation on epidemic dynamics and the persistence of childhood diseases, the degree of epidemiological

coupling at various geographical scales, from within cities to between regions, is a key parameter to consider (Keeling and Rohani, 2002; Xia et al., 2004). Variation in the level of epidemiological contact between individuals in different subpopulations, with varying levels of spatial interaction and distance from regional urban centres, could amplify reduction in the level of correlation between epidemic outbreaks at the inter-city level. The cross-coupling of large urban centres suffering from major epidemic outbreaks of infection with spatially uncorrelated settlements that have experienced a fadeout of the disease may result in a 'rescue effect' (Bolker and Grenfell, 1996).

In an epidemiological system, a rescue effect is defined as the transmission of an infection between-patches which facilitates the recolonisation of communities where the infection has faded out or been locally eliminated (Bolker and Grenfell, 1995). Rescue effects are the product of asynchronous epidemic outbreaks among spatially separate communities within a host population, which decorrelates dynamical fluctuations of disease in different patches. Spatial heterogeneity in vaccine coverage can also result in rescue effects due to spatiotemporal decorrelation between populations (Hagenaars et al., 2004). Vaccination reduces spatial coupling (Bolker and Grenfell, 1996), which at intermediate levels will enable the maximum persistence of an infection. If rescue effects are to effectively maintain the regional persistence of an infection, there must be a sufficient level of epidemiological coupling between patches to ensure subpopulations are connected and frequent contact between patches is established (Dalziel et al., 2016). However, rescue effects are absent if the level of coupling is too high, since they act to remove spatial heterogeneity between patches, and subpopulations act as a homogenous population (Grenfell and Harwood, 1997). In metapopulation terms, rescue effects are caused by migration from source patches (where infection persists due to a positive growth rate in the number of susceptibles) to sink patches (where the susceptible growth rate is negative), where infection would be expected to decrease to extinction (Harrison, 1991).

As the value of an endemic threshold grows, rescue effects become increasingly important for maintaining the circulation and regional persistence of an infection, especially as elimination thresholds are reached. It has in the past been suggested that rescue effects may explain the relative stability of the endemic threshold level in England and Wales in the decades after the introduction of mass vaccination for measles in the late 1960s, which remained at the pre-vaccine era level of approximately 250,000-300,000 (Fine and Clarkson, 1982; Bolker and Grenfell, 1996).

1.7 Study of Endemic Thresholds: Vaccination

1.7.1 Measles

The change in measles incidence and transmission dynamics following the onset of mass vaccination is among the best-documented in epidemiological literature. Before the introduction of the measles vaccine in England and Wales in 1968, epidemics of the disease at the city level exhibited a regular, and spatially coherent, biennial pattern (Anderson et al., 1984). Measles exhibited persistence fluctuations in different cities almost synchronously across regions, with fade-outs of infection observed only in smaller towns and rural settlements (Bjørnstad et al., 2002; Grenfell and Harwood, 1997). In the vaccine era of the 1970s and 1980s, this pattern altered radically, with a marked reduction in size and frequency of epidemic events and extended inter-epidemic intervals occurring less regularly being observed (Bolker and Grenfell, 1995). This has been attributed to a non-linear dynamic effect; the onset of mass vaccination resulted in the elimination of large epidemic outbreaks, which had previously acted to synchronise measles dynamics in different cities in the pre-vaccine era.

Bolker and Grenfell (1996) found that the level of epidemic decorrelation after the onset of mass vaccination varies according to spatial scale (see Fig. 1.5). For instance, in marked

contrast to the between-city decorrelation in epidemic outbreaks, the correlation between London boroughs remained high after the introduction of vaccination, on a par with the pre-vaccine era correlations. However, this is not a surprising observation when one considers the effect of spatial distance on disease transmission. The spatial transfer of an infection in contiguous boroughs where population mixing is high, will be at a far greater rate than what would be expected from the epidemiological coupling of cities, separated by many miles of countryside (Cliff and Haggett, 1980). Consequently, Bolker and Grenfell's (1996) findings suggest that the decorrelation experienced at the inter-city level in the post-vaccine era is lessened by intense coupling at the intra-city level.

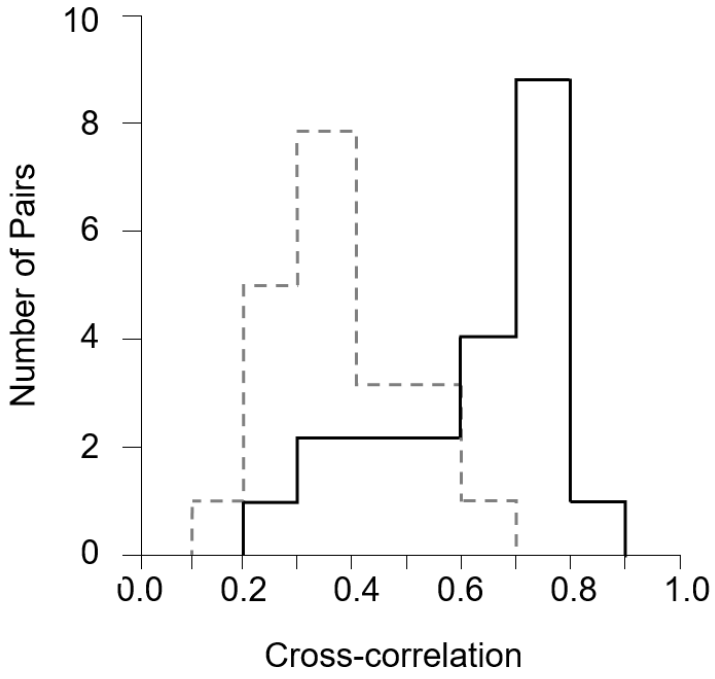

Figure 1.5 *Distributions of pairwise cross-correlations among the seven English cities, in the pre-vaccine era (black line, 1948-1968) and vaccine era (grey dash line,1968-1988). (Source: Bolker and Grenfell, 1996: 12650).*

An equivalent decline in the geographical coherence of measles epidemics after the end of the pre-vaccine era has also been observed using measles morbidity data available in the US (Cliff et al., 1992a, b). Cliff et al. (1992b) found a similarly substantial decorrelation

between 22 states of the north-eastern United States, including New York City and Washington D.C, using surveillance data available from 1962 to 1988.

Table 1.3 *Hypothetical numbers of people susceptible to measles before and after vaccination program based on predicted fade-outs (Source: Nathanson, 1982).*

Population	Before Immunisation	After Immunisation
9,000,000	900,000	400,000
900,000	90,000	40,000
500,000	50,000	22,000
300,000	30,000	13,000

Nathanson (1982) uses the hypothetical examples of New York City and Baltimore in the early vaccine era to explain the impact of immunisation on endemic persistence of measles. According to Nathanson (1982), before the introduction of the measles vaccine in the US in 1963, an estimated 10% of the US population was susceptible to measles. Nathanson calculated that a susceptible population within the community of around 50,000 individuals was required to propagate the infection. Towards the end of the pre-vaccine era, New York possessed a population estimated at 9 million; approximately 900,000 in the city were therefore susceptible to measles. Baltimore, a far smaller city yet with a considerable population of around 900,000, had an estimated 90,000 susceptibles before vaccination. However, after the introduction of measles immunisation, the number of susceptibles in Baltimore has fallen to only 40,000, below the value required to spread the virus. By contrast, in New York, even after immunisation, there would still be enough susceptibles, around 400,000, to maintain the circulation of measles during seasonal troughs.

Consequently, New York continued to report measles monthly, whilst Baltimore experienced a transition from endemic measles to regular fade-out of the disease.

1.7.2 Pertussis

Interestingly, the spatiotemporal dynamics of pertussis are in stark contrast to those of measles, which exhibits the opposite pattern of spatial synchrony in both the pre-vaccine era and after the onset of mass vaccination (Bolker and Grenfell, 1995; Cliff et al., 1992a). Rohani et al. (1999, 2000) found that the onset of mass vaccination for pertussis in 1957 in England and Wales coincided with a major change in the spatiotemporal patterns of pertussis incidence. The epidemic outbreaks of the pre-vaccine era showed little spatial correlation between urban centres and were replaced in the post vaccine era of the 1960s and 1970s by highly spatially synchronised outbreaks (Rohani et al., 1999). Moreover, fluctuations every to 2–2.5 years in pertussis outbreaks in towns and cities between 1944 and 1957 were superseded by highly synchronised, triennial epidemics across England and Wales after the introduction of mass vaccination. Unsurprisingly the inter-epidemic interval between outbreaks increased during the 1960s and 1970s due to the dramatic decline in pertussis incidence (Rohani et al., 1999). Reflecting the spatial synchronisation of epidemic troughs between pertussis outbreaks in the vaccine era, a substantial increase in both the frequency and duration of fade-outs in major cities was also observed, signalling a successful reduction in the transmission of pertussis since the onset of mass immunisation.

In their study of the epidemiological impact of vaccination on pertussis and measles dynamics in the Niakhar area of Western Senegal, Broutin et al. (2004) revealed a substantial increase in the mean number of fade-outs as a direct consequence of immunisation. The mean duration of fadeouts also increased significantly, with a considerable fall in the R_0 for both diseases signalling a clear decrease in the regional persistence of the infections.

Part Two: Modelling Endemic Thresholds

1.8 Quantifying Endemic Thresholds

Quantifying endemic threshold size, in principle, is a relatively straightforward process, and there is more than one method which may be utilised to achieve this task. Several studies have examined the relationship between population size and the mean number of fade-outs per year (see Fig. 1.6) (Bartlett, 1957; Black, 1966, Broutin et al., 2004a,b; Grenfell and Harwood, 1997).

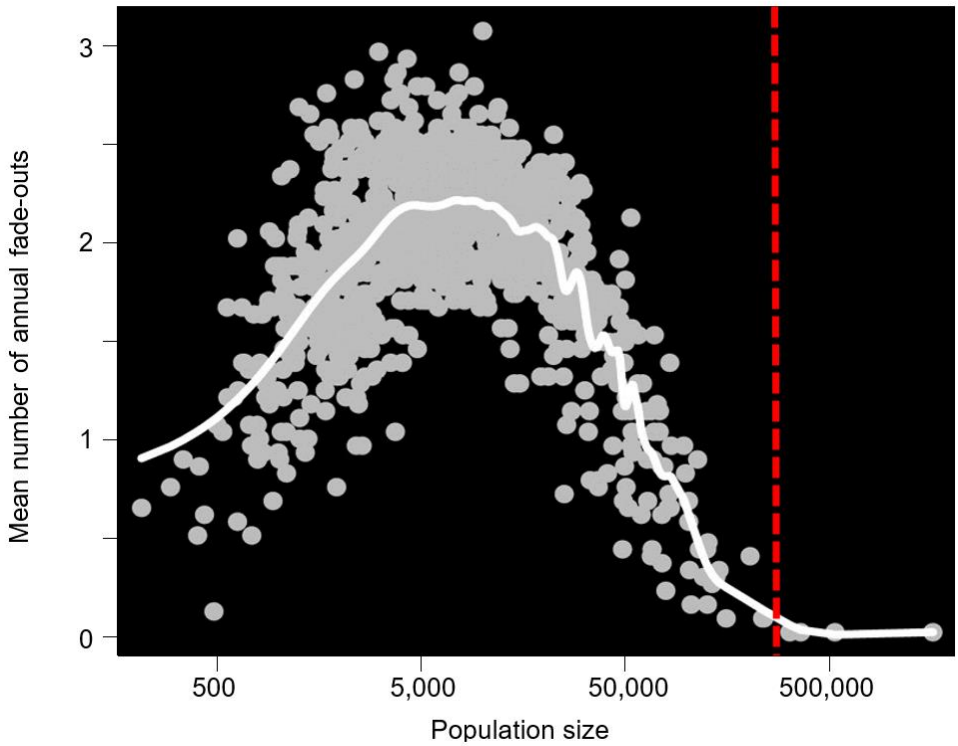

Figure 1.6 *Mean number of annual fade-outs of measles against population size in England and Wales, 1940–1964. The mean annual fade-out value for the 954 urban districts of England and Wales are plotted in grey, with a smooth trend line overlaid in white. The red-dashed line marks the endemic threshold value. Population size of each district represents the average population size over the period (Source: Conlan et al., 2009: 626).*

A fade-out is generally defined based on the infectious period of a disease; the average time required for an individual to recover from an infection after the initial moment of transmission of disease. If no new cases of either infection are reported after a period of three weeks or more, it is generally assumed that the chains of transmission have broken down, and the infection has become locally extinct (Broutin et al., 2007).

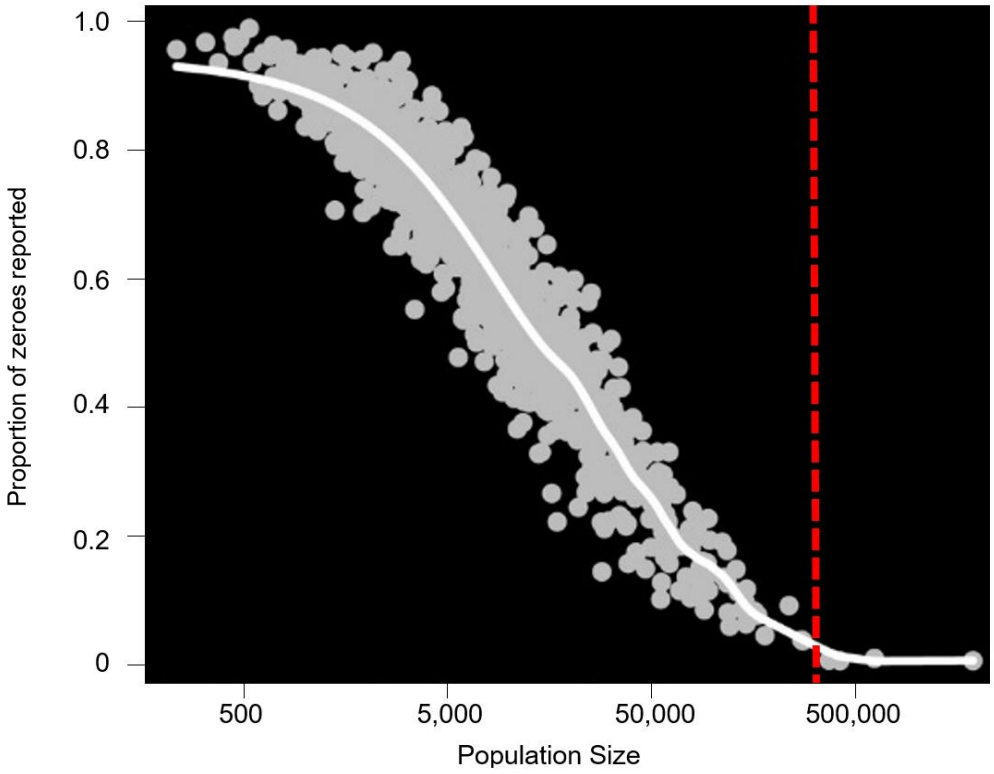

Figure 1.7 *Proportion of weeks with no reported measles cases against population size in England and Wales, 1940–1964. Proportion of zero reports for the 954 urban districts of England and Wales are plotted in grey, with a smooth trend line overlaid in white. The red-dashed line marks the endemic threshold value, the population size above which measles is persistent. Population size of each district represents the average population size over the period (Source: Conlan et al., 2009: 626).*

When the proportion of weeks with no reported cases is plotted against population size (a complementary measure of persistence), the length of fade-out experienced increases with decreasing population size (see Fig. 1.7). The mean number of fade-outs is consequently

limited by the rate of re-introduction of infection from external sources in populations experiencing fade-out (Conlan et al., 2009), which is much greater in larger populations. It is therefore typical for empirical studies to feature a measure of fade-out length (duration of fade-outs/proportion of weeks with no reports) alongside a plot of mean annual fade-outs against population size (Bjørnstad et al., 2002; Broutin et al., 2005). Although these measures provided account for the frequency and proportion of zero reports, they do not clearly discriminate between the relative roles of persistence and invasion dynamics, and this limitation has received attention in the extant literature (Conlan et al., 2009).

In the extant literature on childhood infections, a fade-out has traditionally been defined as a period of at least three weeks without reported cases of infection (Bartlett, 1957), due to the wealth of studies on measles and, to a much lesser extent, pertussis. However, it is not an arbitrary definition. For instance, influenza has an infectious period of approximately four to seven days (Cliff and Haggett, 1989), therefore fade-out for the disease would be defined as a period of two weeks or more without a reported case. The length of a fade-out period is measured by the number of consecutive weeks without cases corresponds to the length of a fade-out period (Broutin et al., 2004a). Wearing and Rohani (2009) provide an alternative definition; the number of times at least three consecutive weeks have zero cases per epidemic. However, they found that estimates for the endemic thresholds size for pertussis using both definitions were often very similar in range.

1.9 Modelling Approaches

A variety of mechanistic transmission models, either deterministic or stochastic in nature, have been developed over several decades to capture the spatiotemporal dynamics of childhood infections. Formulation of these models has been made possible by the extensive records of measles notifications which exist in developed nations. To a varying degree, these models incorporate mechanisms such as seasonality (Aron and Schwartz, 1984; Earn

et al., 2000), age structure (Schenzle, 1984; Bolker and Grenfell, 1993), spatial structure (Bolker and Grenfell, 1995; Hagenaars et al., 2004) and variable infectious and latent period distributions (Keeling and Grenfell, 1997). Regarding measles data in England and Wales, these models have been relatively successful in capturing the temporal pattern of recurrent epidemic outbreaks of the infection (Conlan et al., 2009). The use of these models for accurately estimating the value of endemic threshold size has, however, increasingly been questioned over recent decades (Bolker and Grenfell, 1995; Conlan et al., 2009).

1.9.1 The Stochastic SIR Model

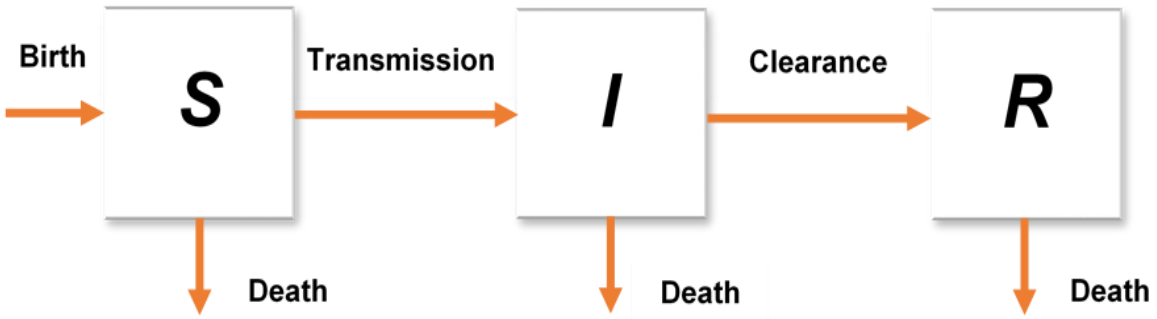

Figure 1.8 *Flow diagram of classic deterministic SIR model (Adapted from Bonds and Rohani, 2009: 542).*

The standard model used by Maurice Bartlett in the late 1950s to estimate endemic threshold size is a continuous time stochastic susceptible-infected-recovered (SIR) model (see Fig. 1.8). The SIR model was originally devised by Kermack and McKendrick (1927), as a simple compartmental deterministic model to analyse the mass-action transmission of a directly transmitted infection with an exponentially distributed infectious period in a closed population.

A stochastic SIR model enables estimates to be made for potential outcomes whilst allowing for the effect of demographic stochasticity and the random nature of population events on

inputs over time (Keeling and Ross, 2008). They are essentially individual based. In relation to the study of infectious disease, stochastic models consider the potentiality for chance events to cause the number of susceptibles to fall to zero, resulting in the local extinction of infection (Bartlett, 1956; Keeling, 2000). Stochastic models tend to assume that the statistical distribution of residence times in disease categories, such as the susceptible and infectious periods, follow an exponential distribution (Grenfell and Harwood, 1997). Random fluctuations in the timing of birth, recovery and transmission events have been demonstrated to play an essential role in the recurrence and extinction of infection (Bartlett, 1957), and fundamental in the persistence of infection in small populations (Trottier and Philippe, 2001). Since Bartlett's formulation of a continuous time stochastic SIR model, such models have provided the means to estimate endemic threshold size (Bartlett, 1957, 1960; Black, 1966). Yet, as Conlan et al. (2009) note, the resultant value for this threshold ultimately depends on both the assumptions of the stochastic model devised and the parameters of the infection itself.

1.9.2 Metapopulation models

When one looks at the extant literature in relation to the concept of endemic threshold size, much research conducted from a spatially explicit perspective has involved the use of metapopulation models formulated by ecologists, as evidenced most notably by the work of Benjamin Bolker, Bryan Grenfell and Matt Keeling on measles throughout the 1990s and early 2000s. For childhood infections, of which many are acutely immunising, it has been recognised that metapopulation dynamics may play an important role in enabling such infections to persist locally within a host population (Grenfell and Harwood, 1997; Keeling et al., 2004). Diseases such as mumps and measles are extremely efficient in using a pool of susceptibles after they have invaded a community, hence their tendency to rapidly fade-out after the supply of susceptibles has been exhausted, and local chains of transmission collapse (Keeling, 2000; Metcalf et al., 2013b).

1.10 Endemic Thresholds & Metapopulation Theory

Over recent decades, metapopulation theory in infectious disease modelling has been pivotal for developing a greater understanding of the processes which operate over time and space that play a fundamental role in the persistence of childhood infections. The regional persistence of measles has become seen as a classic metapopulation problem (Levins, 1969), the outcome of which depends on the balance between the frequency of local extinctions of infection and the rate at which infection is reintroduced from a community where infection remains persistent, to a community where the infection has long faded out.

Metapopulation models have been shown to be effective for exploring the effect of spatial heterogeneity on disease persistence. Spatial heterogeneity, in metapopulation terms, is defined as the 'patchiness' of a host population (Hagenaars et al., 2004). A host population is often divided into spatially separate subpopulations, commonly referred to in ecological and epidemiological literature on metapopulations as 'patches'. Spatial heterogeneity is determined by the level of contact and mixing between patches. Investigating the interplay between local extinctions of infection between-patch transmissions in spatially separate subpopulations is critical to an understanding of how heterogeneous mixing patterns in a large population maintains the endemic persistence of infectious diseases.

1.10.1 Concept

Basic deterministic and stochastic epidemiological models tend to assume all individuals are from a single, homogeneously mixed population, based on the principle of mass-action, yet real systems very rarely experience this assumed mixing (Anderson and May, 1984; Bolker and Grenfell, 1995). Consequently, these models only offer limited approximations of real-world epidemics since they overlook the spatial structure of host populations, which can strongly affect transmission and resulting disease dynamics (Bolker and Grenfell, 1995;

Lloyd and May, 1996). Metapopulation theory suggests that the persistence of infection increases or decreases according to the level of spatial heterogeneity in transmission and rate of contact between individuals.

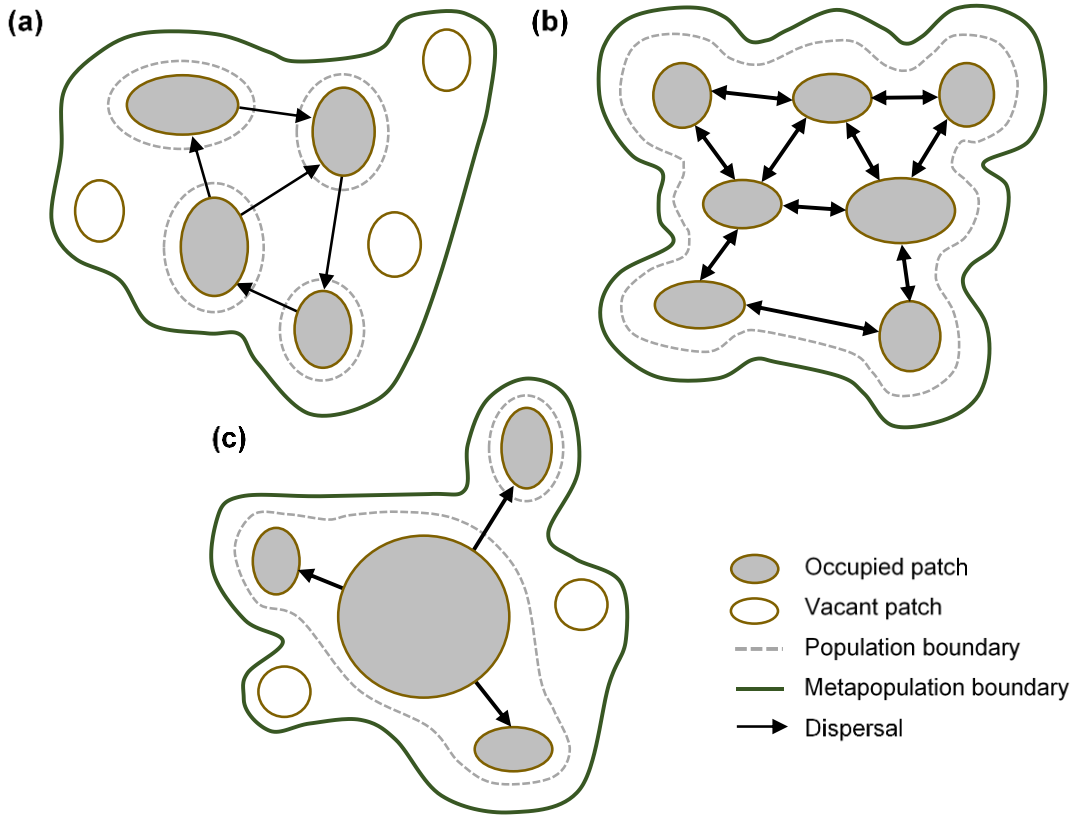

Figure 1.9 *Metapopulation structures.* ***(a)*** *Classical metapopulation comprising of relatively small patches that are sufficiently large and in near proximity so that recolonisation balances extinction,* ***(b)*** *'Patchy' metapopulation consist of several patches of varying size with high levels of dispersal between each patch, functioning together as one unit* ***(c)*** *Mainland-island metapopulation consists of small 'island' patches within dispersal distance of a much larger 'mainland' patch. The probability of local extinction is much greater in island patches, and lower in the mainland patch. This structure can explain source-sink dynamics observed in some metapopulations (Source: Harrison, 1991; Harrison and Taylor, 1997).*

A metapopulation is an ecological concept defined as a population composed of subpopulations (Hanski and Gaggiotti, 2004), in which discrete subpopulations occupy

spatially separate ‘patches’ (Glass et al., 2004). Various metapopulation structures exist, the three most common of which are presented in Fig 1.9. The dynamics of a metapopulation depends on the rate of extinction and recolonisation of its constituent patches (Broutin et al., 2007), and interaction between subpopulations in the form of population flux (migration). By inserting a disease transmission model or epidemiological time series into a metapopulation framework, a large population can be divided into a collection of loosely coupled patches, representing local communities.

Often homogenous mixing is assumed within subpopulations, rather than the regional population itself (Beyer et al., 2012); infected individuals have much greater contact with other individuals from within the same spatially defined subpopulation, rather than with those from other patches (Hagenaars et al., 2004). Transmission between patches is determined by a variety of relevant factors, such as the level of epidemiological coupling, distance and sub-population size (Keeling, 1999, 2000).

1.10.2 Application

The most common application of the metapopulation model in epidemiological literature has been to study the spatial transmission and persistence of childhood diseases nationally at the city-level (Bolker and Grenfell, 1996; Dalziel et al., 2016; Grenfell et al., 2001). At this geographical scale, mixing within cities plays an extremely important role in determining the persistence of an infection (Glass et al., 2004). In metapopulation terms, urban centres act as ‘core’ patches, where the large, dense population maintains the circulation of infection, which spreads outwards to surrounding ‘satellite’ patches (Grenfell and Harwood, 1997). Spatial patchiness has been cited as a key area of study if the circumstances in which childhood infections persist endemically are to be effectively elucidated (Mollison et al., 1994).

Grenfell and Bolker (1998) use metapopulation models to study regional spatial heterogeneity in measles transmission in England and Wales, in relation to urban-rural hierarchies in infection rates. Their findings support the conclusions of Bartlett (1957, 1960), finding measles to persist during inter-epidemic periods in large urban centres with an endemic threshold population above 300,000 during the pre-vaccination era, with prolonged periods of endemic fadeouts mostly occurring in small communities. Epidemic correlations revealed a complex urban–rural pattern in the pre-vaccination era; higher rates of infection were observed in large cities, with the proportion of urban-based cases rising significantly before major epidemic outbreaks in contrast to rural areas, yet small towns were found to have epidemic characteristics in-between those of large town and rural settlements. These results suggest a spatial hierarchy of infection from large, high density populations to smaller, low density communities. However, they also found that if urban and rural settlements are of equal population size, they will have the same propensity for local fade-outs of infection. This is considered a surprising finding since fade-out patterns during inter-epidemic periods are taken to be particularly sensitive to the degree of coupling to large centres (Finkenstädt and Grenfell, 1998), which would suggest fade-outs are less likely in urban areas due to the increased level of coupling with urban centres.

Grenfell and Bolker's (1998) work expand upon the 'cities and villages' model of Anderson and May, originally devised to explore the potential implications of spatial heterogeneity on vaccination efforts against childhood diseases in rural areas, where the rates of transmission are lower than in urban centres (May and Anderson, 1984; Anderson and May, 1992). However, Grenfell and Bolker's (1998) study can also be seen as building upon geographical work of Cliff et al. (1992, 1993), whose study of the spatial dynamics of measles in the United States indicated that the infection diffuses from major cities, 'endemic reservoirs' to settlements in the surrounding countryside, and the work of Bartlett (1957, 1960) on critical community size. Populations above the endemic threshold value, such as those of major urban centres, act as 'core patches' within a regional population.

Part Three: Childhood Infections Under Study

1.11 Pertussis

Pertussis, more commonly known as whooping cough, is an acute, microparasitic, childhood infection that mainly affects the respiratory system. Today, pertussis continues to be a major public health issue. Developing nations suffer high burdens of infant mortality relating to pertussis infection (Muloiwa et al., 2018), while several developed nations with long-established vaccination programmes have witnessed a resurgence of the disease amongst infants, adolescents and adults alike in recent decades (Cherry, 2012; Rohani and Drake, 2011; Martinón-Torres et al., 2018). Pertussis has remained a leading vaccine-preventable cause of hospitalisation and mortality in infants in England and Wales (Campbell et al., 2012).

1.11.1 Pathogenesis

Pertussis is caused by the Gram-negative bacterium *Bordetella (B.) pertussis,* first described by the French and Belgian immunologists Jules Bordet and Octave Gengou in 1906 (Bordet, 1906). Humans are the only recognised host for *B. pertussis* (Amirthalingam et al., 2013). It is a highly contagious, airborne disease, transmitted via small respiratory droplets which become aerosolised when infected persons sneeze or cough (Gopal et al., 2019). Pertussis cases are commonly characterised by a prolonged coughing illness that can last for several weeks (Edwards, 2005). The average incubation period for pertussis typically lasts around one week, whilst the infection period lasts around 14 days (Amirthalingam et al., 2013). A classic case of pertussis can be divided into three stages (Long, 2004). The first stage is the catarrhal stage when a pertussis case is most infectious. During this stage, infected individuals exhibit respiratory symptom such as a non-productive cough, a runny nose, and a mild fever (Amirthalingam et al., 2013). After seven to ten days,

these symptoms are succeeded by the paroxysmal stage, characterised by intermittent periods of intense coughing that end with a distinctive, high pitch 'whooping' sound as the individual inhales, which usually persist for two weeks (Wendelboe and Van Rie, 2006). It is during this stage that individuals are often clinically suspected of showing signs of pertussis infection. The final stage is recovery; this is gradual and can last between two weeks and several months.

In the vaccine era, the majority of pertussis infections remain undiagnosed as they tend to present with mild symptoms, while approximately two-thirds of cases are subclinical, presenting no symptoms (Long et al., 1990). Even when children present with a cough of two weeks or more, pertussis often goes undiagnosed; previous research has shown physicians only consider pertussis in 25% of clinical cases (Deeks et al., 1999). Older children, adolescents and adults may become infected with pertussis, due to a weakened immune system and close proximity to children (Gopal et al., 2019), or due to waning vaccine-induced immunity (Wendelboe et al., 2005; Plotkin, 2014). These age-groups often present subclinical symptoms because of past vaccination or different host response by age (Eidlitz-Markus, et al., 2007), resulting in frequent misdiagnosis. Some infants, despite generally exhibiting more severe illness, may also have atypical presentations of the disease, in particular lacking the protracted, spasmodic cough with the characteristic whoop during the paroxysmal stage (Tanaka et al., 2003).

1.11.2 Pertussis Elimination

Since humans are the only known disease reservoir for *B. pertussis*, pertussis is theoretically a prime candidate for disease eradication through the implementation of sustained mass vaccination campaigns. However, unlike other childhood infections, such as measles and poliomyelitis, natural infection does not confer life-long immunity against the disease (Amirthalingam et al., 2013), whilst immunisation does not guarantee protection

from infection, since the efficacy of the pertussis vaccine is known to decline over time (Wendelboe et al., 2005; Plotkin, 2014). It is therefore possible to contract pertussis later on in life and more than once, with the elderly in particular vulnerable to this particular outcome.

Throughout the interwar period, researchers in the United Kingdom and United States conducted numerous studies and field-trials in the quest to develop an effective inactivated vaccine, yielding positive results by the 1940s (Bell, 1948; Felton and Willard, 1944). Pertussis immunisation began in England and Wales on a localised basis in 1942 (Grenfell and Anderson, 1989). In 1957, the whole-cell pertussis (*wP*) vaccine was introduced across England and Wales for infants three months old and above (Amirthalingam et al., 2013), finally marking the introduction of a routine, nationwide pertussis immunisation programme (Griffith, 1978). It was combined with the diphtheria and tetanus vaccines to form the DTwP vaccine. Initially, there was a substantial fall in the magnitude of notifications following the introduction of routine pertussis immunisation (Rohani et al., 2000). However, by the mid-1960s, the number of pertussis notifications reported nationwide stabilised. The lack of a continued downward trend in pertussis notifications was attributed to either biological changes pertussis bacterium or the use of vaccines with low effectiveness during the period 1957 to 1968 (JCVI, 1977). A 1969 Public Health Laboratory Service survey indicated *wP* vaccines used before 1968 were far from effective; reporting efficacy rates were as low as 20% (PMC and API, 1969). Clarkson and Fine (1985) noted that the reported efficacy of pertussis vaccines varied greatly between the 1950s and 1980s, ranging from 20% to 95%. Their modelling suggested that a vaccine coverage rate of 88% for each birth cohort before one year of age would be necessary to will eliminate bacterial transmission, with repeated cohort immunisation necessary to eliminate transmission. Nevertheless, there was a general consensus within the UK Joint Committee on Vaccination and Immunisation in a report published in 1977 that onset of routine pertussis immunisation had been the vital factor behind the substantial decline in pertussis notifications (JCVI, 1977).

1.11.3 Dynamics in England & Wales

Epidemiologists have focused on the interplay between vaccination, nonlinearity in transmission and demographic stochasticity, to explain the noisy, irregular pertussis epidemics observed in the pre-vaccine era, and the dominant cyclic signature which arises after the introduction of routine pertussis immunisation (Rohani et al., 2002). In the context of England and Wales, much of this work has been undertaken analysing weekly or monthly pertussis incidence data at the national level and between major cities, overlooking the spatial structure and distribution of subpopulations at finer geographical scales, at the regional and local level. For instance, Rohani et al. (2000) examine the impact of the introduction of mass vaccination campaigns in the late 1950s on the spatio-temporal patterns of pertussis incidence, utilising notification data for major cities at the national level. Much of the exiting literature on pertussis in England and Wales has centred on the effectiveness of mass vaccination on the incidence of pertussis. In the pre-vaccination era, pertussis outbreaks accounted for an estimated 150,000 cases and contributing to approximately 300 deaths annually (Amirthalingam et al., 2013). Rohani et al. (2002) note that the *wP* vaccine may not some prevent subclinical cases of the disease, which go on to be unreported while potentially maintaining chains of transmission and thwarting elimination efforts. This echoes the sentiment of Cherry (1998), who remarked that the persistence of pertussis could not be controlled by contemporary immunisation programs. Nevertheless, based on the notifications reported, Rohani et al. (2000) conclude that mass vaccination has led to a large decrease in reported cases of whooping cough in England and Wales.

The effectiveness of pertussis vaccination can also be assessed by analysing the relative changes in the length of the inter-epidemic period after mass immunisation. Analysis of simple models indicates that a significant drop in transmission should be paralleled by an increase in the interval length between epidemic peaks (Anderson and May, 1991). The onset of pertussis vaccination corresponded with a considerable increase in the

interepidemic period, from around two years to nearly four years in the ten most populous cities of England and Wales. This repudiated previous findings in the studies undertaken by Fine and Clarkson (1982), who suggested that vaccination had a negligible effect on the inter-epidemic interval of pertussis in England and Wales, implying mass vaccination had failed to significantly curb pertussis transmission (Rohani, 2000). This was based on model projections which mirrored the apparent failure of mass vaccination to increase the inter-epidemic period of the infection, suggesting the adverse impact of partial vaccine efficacy (Fine and Clarkson, 1982). However, with access to a considerably more extensive, spatio-temporal dataset of pertussis notifications in England and Wale than Fine and Clarkson, Rohani et al. (2000) found that the onset of vaccination coincides with a significant transition in the spatial dynamics of pertussis, initiated by the decline in disease transmission, with spatially decorrelated epidemics in the 1950s succeeded by geographically synchronised outbreaks in the decades following the introduction of the *wP* vaccine. This finding is consistent with a considerable increase in both the number and duration of fade-outs in the vaccine era, consistent with spatial synchronisation of epidemic troughs (Rohani et al., 1999), and increase in critical community size required to maintain chains of infection transmission.

1.12 Measles

The measles virus has long posed a significant public health risk to human populations. Prior to the introduction of vaccination, measles blighted the populations of industrialised nations with regularity and impunity, and was a leading cause of infant mortality, with the annual number of measles-related deaths estimated in the range of five to eight million (Moss and Griffin, 2006). Since the 1960s, measles vaccines have facilitated the dramatic decrease in the incidence of the disease and its associated complications in many regions across the world (Strebel et al., 2012). The decline in mortality from measles in industrialised nations was associated with economic development, improved nutritional status and

supportive care, and emergence of antibiotic treatments for secondary infections such as bacterial pneumonia. Nevertheless, measles continues to be a major public health issue for the developed world, where low vaccination coverage leads to frequent local outbreaks at the community level, and for developing nations in regions such as sub-Saharan Africa, where elimination campaigns face significant geographical, socio-political and economic obstacles and the disease remains a leading vaccine-preventable cause of infant mortality (Moss and Griffin, 2006).

1.12.1 Pathogenesis

The measles virus is transmitted by aerosols or the respiratory droplets from infected individuals, primarily via sneezing or coughing. The incubation period for measles lasts between ten to 14 days, during which the virus replicates and spreads in the infected host (Moss and Griffin, 2006). The first clinical symptom of the disease is a prodromal illness characterised by a mild fever, cough, and conjunctivitis (Strebel et al., 2012). This is the most contagious stage of the disease. The prodromal symptoms intensify before the appearance of a distinctive rash, resembling red bumps on red patchy skin (Strebel et al., 2012). This rash tends to be generalised across the body. In uncomplicated measles, clinical recovery begins shortly after the appearance of the rash, normally lasting three to four days before fading.

Complications arising from measles infection are not uncommon, since the virus is known to cause profound immunosuppression (Moss and Griffin, 2006), which can last for several weeks and even months after recovering from the acute illness. Recent studies have found that measles infection can effectively erase 20 to 50 percent of antibodies against an array of viruses and bacteria, such as influenza, depleting a child's previous immunity (Mina et al., 2019; Petrova et al., 2019). In such instances, a measles-ravaged immune system has to relearn how to protect the body against infections. Unsurprisingly, measles patients are

susceptible to secondary bacterial and viral infections that cause pneumonia and diarrhoea and is responsible for much of the measles-related morbidity and mortality (Beckford et al., 1985; Greenberg et al., 1991).

1.12.2 Elimination

Measles is thought by many experts to meet the criteria forwarded as necessary for an infectious disease to be considered a candidate for eradication: first, it must be biologically feasible, with humans playing the crucial role in disease transmission; second, sensitive and specific diagnostic tools must exist; and finally, an effective intervention must be available (Orenstein et al., 2000; Quadros, 2004). The success of effective measles immunisation programmes in communities with high vaccination coverage has demonstrated that the interruption of chains of transmission in a large geographic area for significant periods of time supports the feasibility of measles elimination.

In terms of biological feasibility, measles is a good candidate for elimination. Humans are the only natural reservoir that can sustain the persistence of the virus. Although primates can be infected with measles and develop human-like illness, wild primate populations do not reach the critical size required to maintain chains of transmission for the infection to persist and pose no risk to measles elimination efforts. Based on the findings of Bartlett (1957, 1960) and Black (1966), a population size of 300,000 and above, with an input of 5,000 to 10,000 births annually is required to provide a sufficient number of new susceptible individuals to maintain chains of transmission.

The characteristic clinical features of measles, in particular the rash, makes the disease much more straightforward to diagnose compared to other childhood infections (Moss and Griffin, 2006). However, other infectious diseases which affect young children, such as

rubella and human herpes virus type 6 can all cause symptoms that can mimic, or to some extent replicate, those of measles which can lead to misdiagnoses (Davidkin et al., 1998).

A key difficulty with measles elimination relates to the high infectivity of the disease. The measles virus is an extremely contagious infection, able to infect individuals several days prior to exhibiting clinical symptoms, most notably the characteristic rash which distinguishes the disease from others. A key epidemiological metric of the infectivity of an infectious disease is the basic reproduction number (R_o). R_o is the mean number of secondary cases that would arise if an infectious agent were introduced into a completely susceptible population (Anderson and May, 1991). R_o *i*s a function not only of the infectious agent but also of the host population. The R_o for measles is often cited as 12–18 (Guerra et al., 2017). This is in contrast to five to seven or smallpox virus and two to three for coronaviruses (Moss and Griffin, 2006). In the 1951 measles epidemic of Southern Greenland, an infected individual, identified as the case of origin, attended a dance at a community gathering during the infectious period of the disease. This ignited a virgin soil epidemic which resulted in a R_o of 200 (Christensen et al., 1953). Given its high infectivity, measles thus requires a consistently high level of herd immunity of approximately 90-95% to effectively interrupt chains of transmission and significantly increase the critical size required to sustain the persistence of disease to enables elimination (Cutts and Markowitz, 1994). Local outbreaks can occur in populations in which only less than 10% of individuals are susceptible.

Although the use of the single-antigen live measles vaccine was introduced nationally in 1968, the vaccine had been available since 1966, and later studies generally accepted the vaccine have an efficacy rate of over 90% (Shelton et al., 1978; Marks et al., 1978). The recommended age for vaccination with the primary dose in England and Wales between 1968 and 1988 was 12 to 23 months of age (Strebel et al., 2012). Initial vaccine coverage

was poor, but it gradually increased to a level of approximately 80% by 1988 (Vyse et al., 2002).

1.12.3 Dynamics in England & Wales

Measles became a statutory notifiable infectious disease in England and Wales in January 1940, with clinicians legally required to report any cases diagnosed to the district Chief Medical Officer. There is considerable serological and historical evidence that prior to measles vaccination, at least 95% of individuals had experienced measles infection by the time they reached adolescence (Langmuir, 1962; Earn et al., 1998). An important exception is island populations, which can remain infection-free for variable periods of time and experience epidemic disease that involves all age groups not previously affected by the last wave of infection, once the virus has been reintroduced (Black, 1966; Cliff et al., 2000). Thus, whereas peak transmission usually occurs among young children, outbreaks in isolated communities can involve older generations. This was exemplified by 1846 measles epidemic on the Faroe Islands, where Panum noted the disease affected persons of all ages who were not alive during the last epidemic that had occurred 65 years earlier (Panum and Petersen, 1940). Before the onset of national vaccination campaigns for measles in 1968, England and Wales experienced regular measles epidemics, with the total number of notifications varying between 160,000 and 800,000 cases per year. Epidemics tended to exhibit seasonal cycles and longer-term, generally biennial major epidemics (Anderson and May, 1991; Grenfell et al., 2001). Young children of school age had the highest risk of infection and accounted for the largest proportion of cases. However, in densely populated urban areas, transmission among infants took on greater importance (Strebel et al., 2012).

The introduction of mass vaccination had an almost immediate and significant effect on the magnitude of measles cases, with annual notifications falling to between just 50,000 and 100,000 by the late 1980s. In total, 11,337,267 measles notifications were reported over a

28-year period in the pre-vaccine era between 1940 and 1968, with 7,863 cases resulting in death (PHE, 2017). After the introduction of vaccination, over the equivalent period of time to the year ending 1996, 2,424,836 cases and 413 measles-related deaths were reported. This represents a decline of 79% in measles morbidity and 95% in measles-related mortality, respectively. In their study of the efficiency of measles notification in England and Wales, Clarkson and Fine (1985) found a strong correlation between births and four-year interval to notification, coinciding with the age at school entry and maximum incidence for the disease.

1.13 Scarlet Fever

Scarlet fever is an acute bacterial childhood disease and once a leading cause of childhood mortality, in the nineteenth and early twentieth centuries (Duncan, 2019). It has been estimated that the mortality rate for scarlet fever in multiple locations around the world reached 25% of cases by 1900 (Guerrant et al., 2011). By the mid-twentieth century, scarlet fever experienced a significant decline in morbidity accelerated by loss of virulence, the introduction of antibiotics and improvements in hygiene. Sporadic outbreaks of scarlet fever were reported in England and Wales and other western nations throughout the mid and late-twentieth century (Walker and Brouwer, 2018), but were no longer associated with the high mortality rates of times past. In recent years, scarlet fever has made a dramatic return as a public health issue, albeit remains an issue that evades wider public recognition. In 2016, scarlet fever incidence in England and Wales was at its highest for over 50 years (Lamagni et., 2018).

1.13.1 Pathogenesis

Scarlet fever is caused by group A streptococcus (GAS) bacterium, which is responsible for a range of invasive and non-invasive infections (Duncan, 2019), and humans are the only known natural hosts of the GAS bacterium which causes scarlet fever (Ferretti et al., 2016). The association between streptococci and scarlet fever was confirmed by George and Gladys Dick in Chicago in 1923, who located the causative agent of scarlet fever in the toxins produced by GAS bacterium (Dick and Dick, 1924a). The following year, they went on to invent 'The Dick Test' in 1924, which was used to identify those susceptible to the disease (Dick and Dick, 1924b).

Scarlet fever is usually spread through aerosol transmissions, by people either coughing or sneezing, but it some instances can spread when a person comes into a contact with surfaces with GAS bacterium (CDC, 2016). The latent period for scarlet fever can vary from anywhere between two to seven days after initial exposure, although it has been known to be as short as 12 hours, with the infectious period lasting approximately seven days (Wong and Yuen, 2012). Scarlet fever cases are commonly characterised by several notable symptoms, such as strep throat and a 'scarlatiniform' rash which covers the body, leaving a sunburned appearance (Duncan, 2019). The rash begins within 48 hours of symptom onset. Other symptoms include fever and fatigue, which tend to pass within ten days after their initial presentation (Wong and Yuen, 2012). The disease most commonly affects children of school age, between five to fifteen years old (Bisno, 1995). Known complications arising from scarlet fever cases include acute rheumatic fever and inflammation of the kidneys (Duncan, 2007). Since their introduction in the early post-war period, antibiotics such as penicillin V and amoxicillin form the core treatment for scarlet fever and are often used to prevent children from developing potential complications arising from infection (Langlois, 2016).

1.13.2 Elimination

Unlike pertussis and measles, scarlet fever is not a vaccine-preventable disease. Historically, attempts to develop and effective vaccine for the disease have proved to be fruitless. George and Gladys Dick developed the first scarlet fever non-toxic vaccine in 1924 (Dick and Dick, 1924b). However, by the end of WWII, its use was discontinued due to poor efficacy and the arrival of antibiotics as an effective treatment for the disease (Ellis and Brodeur, 2012). Difficulties in developing an effective scarlet fever vaccine include accounting for the substantial number of GAS strains circulating in the environment and securing the necessary resources to ensure adequate safety and efficacy trials of potential future vaccines (Ellis and Brodeur, 2012).

1.13.3 Dynamics in England & Wales

Previous research on the dynamics of scarlet fever in England and Wales has mainly utilised mortality data contained in the Registrar General's Annual Reports from the latter half of the 19th century, which provide a consistent time series at the national level post-1847. Using this data adopted from the work of physician Charles Creighton in the 1890s, Duncan et al. (1996) revealed a significant increase in scarlet fever mortality in the nineteenth century which typified the second epidemiological phase of the disease (Katz and Morens, 1992). Prior to the early mid-nineteenth century, scarlet fever outbreaks had been lethal but sporadic in nature (Katz and Morens, 1992). A spectral analysis highlighted a regular inter-epidemic period nationally, lasting approximately five years, with large fatal epidemics ceasing by the beginning of the third epidemiological phase in the 1880s (Katz and Morens, 1992), when the endemic level also decreased markedly. These findings confirm those from work undertaken by physicians during the period by Johannessen (1884) on scarlet fever incidence and mortality rates in Oslo between 1863-1878, and by Whitelegge (1893)

documenting incidence and mortality rates in Boston between 1840-1904. Both found epidemics in urban centres corresponded to regular three-to-five year cycles. A longitudinal analysis by Lamagni et al. (2018), using morbidity data held by Public Health England from 1911 to 2016, which incorporates notification data reported by the Registrar General's Weekly Returns, also detected a periodicity of four to six years between epidemics across much of the twentieth century until the 1960s, when the widespread use of antibiotics and significant improvements in disease prevention pushed scarlet fever back into an endemic phase, characterised by low incidence and long periods between minor outbreaks.

Part Four: Statement of Research

1.14 Statement of The Problem

In an applied context, the endemic threshold has been proposed as a guide for control strategies, and an argument has been made for ignoring populations below the threshold value if vaccines are constrained or resources limited (Beyer et al., 2011; Haydon et al., 2006). However, this only hold value if the persistence of infections in a regional population is dependent on local persistence in large core urban communities which serve as endemic reservoirs, and rescue effects are rare. The implementation of mass vaccination should theoretically drive disease persistence away from the local scale towards the metapopulation scale. All local districts are embedded in a metapopulation in which rescue effects will take place, at least, to some extent, and the rate and scaling of import of infected individuals between districts influences the nature of the relationship between zero-incidence and population size. Despite the simple theoretical prediction that endemic threshold size should increase significantly with vaccination, there have been suggestions in the past that the threshold size for measles in England and Wales remained steady due to rescue effects (Bolker and Grenfell, 1996), born from spatial coupling between districts

and settlements, and highlighted by phase differences among disease activity in geographically separate districts caused by intermediate levels of vaccine uptake. Identifying rescue effects require hierarchical spatial data from districts above and below the endemic threshold, to establish on what spatial scale it occurs

The work of Cliff et al. (2000) on temporal changes in endemic threshold population size in an island population provides a methodological template for examining long term temporal changes in endemic threshold populations. However, in the context of England and Wales, where the temporal changes in endemic threshold size for notifiable infections are yet to be studied in any great detail, unlike island populations, settlements are strongly bound together by the movement of people and spatial coupling. This movement is hierarchically structured, and regular spatial flows from one geographical unit to another are known to reintroduce infection after endemic fade-outs in regions where infection is yet to be eliminated. It is therefore key to explore temporal changes in endemic threshold population size in relation to the spatial dynamics of infection, to understand the effect of spatial coupling, distance and movement between populations on the persistence of infection, particularly since the rate and scaling of import of infected individuals has been described as having the ability to change the nature of the relationship between zero-incidence and population size (Metcalf, 2013; Conlan et al., 2009).

As reflected in the work of Bolker and Grenfell (1996), there is limited empirical evidence that suggest rescue effects may prevent increases in endemic threshold size, and this may be fruitful area of research for geographers aiming to understand the role of geography in shaping the endemicity of disease. The only attempt to quantity rescue effects for childhood infections has been led by ecologists who use WHO incidence data to explore the rescue effects globally, comparing persistence of infection between island and mainland countries in a global metapopulation framework (Metcalf et al., 2013a).

Despite the vast literature that exists on the spatiotemporal dynamics and persistence of measles, very few these studies, with the possible exception of Bolker and Grenfell (1996) explicitly focus on or account for spatiotemporal changes in endemic threshold population size over time, beyond analyses of fade-out structure and behaviour of measles and pertussis metapopulations (Broutin et al., 2004a, Xia et al., 2004). Lesser still, endemic threshold populations have been poorly characterised for uniquely structured regional populations which exist within wider mainland metapopulations. These regional metapopulations represent fertile testbeds to explore disease persistence from a geographical perspective. Attempts to quantify endemic threshold populations for childhood infections have been mostly limited to studies conducted in island populations (Black, 1966; Cliff et al, 2000) or those conducted in England and Wales on a national scale (Keeling and Grenfell, 1997; Grenfell et al., 2001; Conlan and Grenfell, 2007). 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 with complex patterns of spatial mobility and hierarchical spatial structures which operate as independent epidemiological systems.

Research on endemic thresholds in specific regions in England and Wales has been absent in the context of childhood infections. It has generally been assumed that the population size at which disease fade-out may occur in regional populations in mainland metapopulations such as England and Wales are masked. Regions are assumed to have a complex spatial hierarchy of communities of varying sizes, ranging from large cities and towns to rural hamlets, often with a high degree of spatial interaction driven by a dense network of localised travel movements. In this geographical context, the spatial transfer of infection via population flows from cities to neighbouring towns and villages may occur, but disease may also be reintroduced from surrounding, smaller settlements to large towns and cities due to the consistent movement of individuals between closely connected communities, regardless of population size. However, this assumes regions in mainland

populations such as England and Wales do not possess distinctive, unique characteristics in terms of settlement hierarchy, their demography, level of internal geographical isolation, the geographical dispersion of susceptibles and population densities, and the nature of spatial interaction between communities.

1.15 Research Justification

Identifying and studying endemic threshold populations at finer spatial scales enables one to explore how spatial interactions between local districts contribute to regional disease persistence, identify hotspots of infection and assess the impact of intervention on spatial dynamics of disease in light of potential rescue effects that may potentially inhibit control efforts. Identifying persistence hotspots with high endemic activity and export of infection after the onset of vaccination emphases the importance of geographically targeted immunisation programs, particularly in those regions of the world where vaccine-preventable diseases continue to re-emerge. To this author's knowledge, there has been very little work which quantifies temporal changes in endemic threshold populations in complex, hierarchically structured regional populations.

Historical data, in the form of the Registrar-General's *Weekly Return*, provides highly accessible long-term, spatially resolved, and disaggregated incidence data for pertussis in England and Wales. These qualities permit the identification and investigation of unique endemic threshold populations at finer spatial scales. The *Weekly Return* also represents an extremely useful resource for investigating the impact of vaccination on the spatial dynamics of disease, since it provides a consistent record of notifications at the same spatial and temporal scales before and after the onset of mass vaccination. This has the potential to inform strategies of spatially-targeted immunisation programmes. The study of historical data has played an essential role in developing current understandings of the effects of seasonality and stochasticity on disease patterns, as well as shedding light on the spatial

synchrony of epidemics and traveling waves in disease systems for childhood infections, most notably concerning measles (Grenfell et al., 2001). With the quality of historical infectious disease data accessible from the *Weekly Returns*, there is significant scope to explore from a geographical perspective long term spatiotemporal changes in endemic threshold populations in England and Wales, where the spread of infection is hierarchically structured, and data exists on a fine spatial scale.

It is clear from the research discussed that the epidemic dynamics of scarlet fever have been comprehensively established by a relatively concise body of literature. This is certainly the case when one looks at the research undertaken on the disease at the metropolitan and national levels, particularly for the nineteenth century, and in terms of mortality. However, it is also clear that the endemic and regional, spatial dynamics of scarlet fever, especially in the 20th century, have received scant attention. Katz and Morens (1992) note the urban-rural epidemic dynamics observed in the 19th century, with regular cyclical scarlet fever outbreaks found in metropolitan areas in contrast to sporadic, less severe epidemics in rural areas, which were sometimes several years apart. However, the regional spatial dynamics of scarlet fever have not been examined beyond this observation, and this could prove a fruitful area of investigation. The work of Cliff et al. (1992, 1993) on the geographical structure of measles epidemics in the North-Eastern United States, as well as research on measles persistence by population biologists in England and Wales (Grenfell and Bolker, 1998; Grenfell et al., 2001) has already demonstrated that infection may diffuses progressively from urban centres down to the surrounding rural areas, following the ‘cities and villages’ model forwarded by (Anderson and May, 1991). Within this model, the endemic threshold size is a key concept with regards to disease persistence. After an exhaustive search and review of past literature, there is a notable absence of work which estimates endemic threshold size for scarlet fever over time within the context of highly connected regions with hierarchical population structures. The endemic threshold size for

scarlet fever has only been estimated temporally for a historically closed, isolated island population, namely Iceland (Cliff et al., 2000).

1.16 Aims & Objectives

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. This study aims 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; Broutin 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 two regions selected for the present analysis are the historic county of Lancashire, located in Northwest England and South Wales, comprising of four historic counties of Wales. Lancashire and South Wales represent suitable candidates for studying the endemic persistence of childhood infections for several reasons. Both regions are geographically and topographically diverse, with significant overall populations and importantly they are regions of great contrast in terms of spatial structure, connectivity and isolation. Moreover, both experienced significant demographic transformations during the postwar period, as a result of the ‘Baby Boom’, population decentralisation driven by urban slum clearance and

significant economic upheaval as a consequence of deindustrialisation. These events transformed the ways in which local subpopulations interacted with each other on an everyday basis and the distribution and replenishment of susceptible individuals.

The approach presented in this thesis aims to provide a straightforward method for detecting and tracking changes in disease persistence over time in the form of the endemic threshold estimates, which can be applied to a range of diseases for which an adequately detailed data record exists. Regional persistence of infection is rooted in local transmission patterns. Identifying geographical corridors of infection that contribute to recurring epidemics helps define and predict outbreak patterns. Detecting epidemiologically important districts which play a significant role in facilitating the persistence of disease within a regional metapopulation via rescue effects is vital to the success of national vaccination campaigns. The approach followed in this thesis could be applied to other regional populations to detect spatial heterogeneities in disease persistence to help achieve successful outcomes for disease intervention.

1.17 Research Questions

1) Are there significant regional differences in the size of the endemic threshold populations for the same childhood infections?
2) What is the spatial impact of vaccination on endemic threshold populations for pertussis in geographically divergent regional metapopulations?
3) How does the historical spatial structure and spatial distribution of a regional metapopulation influence endemic threshold population size?
4) What are the differences in drivers of diverging levels of disease endemic among regional metapopulations?
5) To what extent does geographical mobility and connectivity affect the estimation of endemic threshold populations?

1.18 Chapter Outline

Chapter 2 is concerned with the research design of the study. The chapter consists of two parts. The first describes the data sources utilised, process of data collection and provides an assessment of data quality. The second part of the chapter details the multiple components and quantitative techniques used to perform the geographically-centred analysis in the study, describing the methodological procedures and providing rationales for selection of methods used in the analysis.

Chapter 3 is composed of three parts. Part one provides a description of the study period selected to facilitate the analysis. Part two provides a breakdown of the geographies of the regional metapopulation of Lancashire and South Wales. Part three provides a description of the demographic profile of the two regions and a summary of the evolving regional demography throughout the study period.

Chapter 4 presents an exploratory and descriptive spatial analysis of time-series of monthly notification data for measles, pertussis and scarlet fever in the Lancashire and South Wales regions, alongside a select subset of districts in each region (January 1940–December 1969).

Chapters 5, 6 and 7 present and discuss the main findings. Chapter 5 details and analyses spatiotemporal changes in endemic threshold populations for measles, pertussis and scarlet fever in the regional metapopulations of Lancashire and South Wales, using the 'moving window' empirical regression approach to estimate endemic thresholds. Chapter 6 details the findings of the hotspot and survival analyses, examining geographical patterns of pertussis persistence in the pre-vaccine and vaccine-eras, to explore the spatial impact of mass vaccination on pertussis endemicity in the Lancashire and South Wales regions. Chapter 7 presents the results of endemic-epidemic modelling of pertussis, measles and

scarlet fever incidence across nine time–windows in Lancashire and South Wales, using various sub-model formulations of the HHH model, a multivariate regression time series model for infectious disease count data. This empirical-based regression is used to analyse geographically aggregated count data, decomposing disease risk additively into endemic and epidemic components to account for spatial and other heterogeneities in disease spread within a regional metapopulation.

Chapter 8 is the final chapter of the thesis, presenting a summary of the research findings, a discussion of potential limitations and areas for future research, followed by concluding marks.

Chapter 2: Research Methodology

2 Introduction

This chapter consists of two parts that together outline the research design. The first part focuses on data sources, collection and data quality. This section documents the secondary data sources utilised to enable the construction of spatially-aggregated datasets of weekly measles, pertussis and scarlet fever notification data for Lancashire and South Wales. Demographic and geospatial data sources used are also detailed. A description of the data collection, digitisation and entry process is provided. The quality of the notification data for the three childhood infections are assessed and data limitations are discussed. The second part of the chapter details the various strands of the quantitative analyses performed throughout undertaking the research focused on detailing the methodological procedures and outputs produced during analysis. Rationales are provided for the selection of each method utilised with reference to supporting literature.

Part One: Data Sources, Collection & Quality

2.1 Data Sources

2.1.1 Disease Data

The data fundamental to conducting the research presented in this thesis has been abstracted from the Registrar-General's *Weekly Return*. The origins of the *Weekly Return* can be traced back to the mid-nineteenth century, with the publication of the *Weekly Return of Births and Deaths for London* by the General Register Office (GRO), By the 1890s, returns for other prominent cities and large towns in England and Wales had been incorporated. It was around this time that a regular series of weekly returns of certain

infectious diseases in London began to be published. The desire for more statistical information on the nation's health in the late nineteenth and early twentieth centuries encouraged the collection of surveillance data on the most severe and common communicable diseases in major cities and towns across the country (Mooney, 2015). From the beginning of 1922, notification data for selected directly transmissible diseases, primarily childhood infections, for all districts in England and Wales were combined with the returns of births, deaths in the Registrar-General's *Weekly Return* (Earn et al., 1998). Local medical officers of health across the country would collect notification data for selected infectious diseases within their district from general practitioners, who would make individual records of cases as the first point of contact with infected individuals (Cliff et al., 1981). These records were collected weekly, with the reporting week running from Friday-to-Friday and collated by the GRO in London before publication. Copies of the *Weekly Return* provide detailed tables which offer a rudimentary quantification of disease morbidity from notifiable infectious diseases for all reporting LGDs in each administrative county, as well as cases recorded by port authorities.

Local government districts (LGDs) are the basic geographical reporting unit for disease notifications and can be broken down into four sub-categories for local government: County Borough (CB), Municipal Borough (MB), Urban District (UD) and Rural District (RD). Local government districts were used as the reporting unit for communicable diseases until the 1972 Local Government Act came into effect on 1st April 1974, which saw the abolition of numerous county and municipal boroughs, urban districts and rural districts. The statutory notification of pertussis and measles cases in administrative districts in England and Wales first appeared in the *Weekly Return* beginning in early November 1939. For scarlet fever, statutory notifications go back further, to the week ended 7th January 1922, the first *Weekly Return* to include data on communicable disease in LGDs nationally to be published (Earn et al, 1998). During the study period (January 1940–December 1969), the Lancashire region consists of 125 local government districts (17 CBs, 26 MBs, 68 UDs, 14 RDs). The four

counties that compose the South Wales region (Carmarthenshire, Glamorgan, Monmouthshire, and Pembrokeshire) consist of 75 districts (four CBs, 14 MBs, 37 UDs, 20 RDs) over the same period. Together, these 200 districts form the bedrock of the geographical framework for the analyses laid out in this chapter.

During the study period, the names and boundaries of the districts were intermittently reviewed by Local Government Commissions, the role of which would be to examine the areas, status and functions of local authorities. Although there were subtle changes in the boundaries and land areas districts in South Wales during the study period, reflecting changes in population size and density over time, no new districts were introduced. Regarding Lancashire, a new district was created mid-way through the study period, whilst another district was abolished. In response to the exponential growth of Kirkby, an overspill estate for Liverpool where the population had swelled from around 3,000 inhabitants in 1951 to over 40,000 by the late 1950s, Kirkby UD was created in 1958. Limehurst RD was abolished in 1954 due to the increasingly urbanised nature of the area. The land area formerly covered by the district was divided between Ashton-under-Lyne MB, Oldham CB, Failsworth UD, Droylsden UD and Mossley MB. Ulverston RD was renamed North Lonsdale RD in 1960. Kirkby UD and Limehurst RD were excluded from the data collection process to ensure all districts included in the analysis had a complete time series of case notification data for all three diseases across the study period, along with accompanying demographic data.

2.1.2 Demographic Data

Until the implementation of the 1972 Local Government Act, annual estimates for the absolute number of births, population size and population density per acre, among other measures, at the local government district level were collated and published in the

Registrar–General's *Statistical Review of England and Wales*. These reports collected a wide range of data; vital statistics (such as causes of mortality), demography, labour migration and related economic measures, and civil statistics, concerning marriages and divorce. The *Statistical Review* is a valuable source for regular annual data on key demographic measures required to effectively analyse the endemic and epidemic dynamics of childhood infections at the district level when using historical infectious disease data.

2.1.3 Geospatial Data

To facilitate the construction of geospatial datasets to facilitate spatial modelling of infectious disease dynamics, digital shapefiles of administrative county and local government district boundaries for Lancashire, Monmouthshire, Glamorgan, Pembrokeshire and Carmarthenshire were downloaded from the Vision of Britain website (https://www.visionofbritain.org.uk/data/). The Vision of Britain website serves as the home for the '*A Vision of Britain through Time'*, which brings together historical surveys of Britain to create a historical, geographical and quantitative record of how the nation and its localities have changed since the mid-nineteenth century to the early 1970s. It was created by Humphrey Southall and the Great Britain Historical GIS Project, based at the University of Portsmouth (Gregory et al., 2002).

The shapefiles contain GIS polygons for each administrative unit at the district and county levels. For practical purposes and to maintain consistency among the map-based visualisations across the study period, all maps created utilise the boundary shapefiles for Counties of England and Wales and Districts of England and Wales for 1961. Each boundary dataset uses the OSGB National Grid. Relationships to container units have been omitted as there are frequently relationships to more than one higher-level unit. More information on the digital boundary data accessed is provided by online documentation located in the publicly accessible Great Britain Historical Database.

2.2 Data Collection

From the end of June 2017 to the beginning of October 2018, the process of data collection and entry was undertaken. Notification data for measles, pertussis and scarlet fever were extracted from the Registrar-General's *Weekly Return*, for a thirty-year period. This is equivalent to 1,560 weeks and therefore a disease record spanning 1,560 copies of the *Weekly Return*, from the week ended 6th January 1940 to the week ended 2nd January 1970. Notification data was collected for 125 LGDs in the administrative (historic) county of Lancashire, and a total of 74 LGDs for four Welsh administrative counties (Glamorgan, Monmouthshire, Pembrokeshire and Carmarthenshire), which together compose the region of South Wales. Measles, pertussis and scarlet fever notifications were recorded for each week and each LGD separately. The vast majority of the *Weekly Returns* were accessed from the Documents Division at the Hallward Library on the University Park Campus, Nottingham. Annually recorded demographic data on population size, population density (persons per acre) and the absolute number of births for Lancashire and South Wales LGDs was abstracted from thirty annual copies of the Registrar–General's *Statistical Review* in the autumn of 2017. These reports were also accessed from the Document Division at Hallward Library.

For each copy of the *Weekly Return*, all pages containing case reports for measles, pertussis and scarlet fever in all LGDs for both regions were photographed to construct an image library of disease notification data, from which case numbers could be entered into .csv file format on Microsoft Excel ready for manipulation. Due to the significant number of photos captured to accurately catalogue the relevant notification data from the *Weekly Returns*, it took several weeks to organise the photos to ensure they were in the correct chronological and geographical order as organised in the reports. This was important to minimise the risk of data entry errors that could adversely affect the results of analyses undertaken later in the research process. In total, 12,774 photographs were captured to

assemble the notification data for LGDs in Lancashire and the four counties of South Wales from the *Weekly Return* across the study period. The same approach was adopted when collecting annual district-level demographic data (population size, birth rates, population density per acre) from copies from the *Statistical Review.* Considerable time was spent editing the images to improve their clarity to prevent data entry errors, since many of the reports, particularly those published during the wartime years, were not in prime condition, due to the presence of ink smudges, creases and discolouration over time.

The original decision to photograph pages with infectious disease data and subsequently organise the images for data entry was adopted after experimenting with the use of optical character recognition (OCR), utilising software such as *Abbyy FineReader 14* and *Nuance OmniPage Ultimate*. OCR is both a process and software technology that converts a hard copy of a printed, typed, or handwritten document into an electronic form that can be read and edited in separate word-editing and data handing software programs (Chaudhuri et al., 2017). In theory, the use of OCR should remove the need to manually perform data entry.

Unfortunately, OCR software proved to be extremely ineffective in accurately processing and rendering the data contained within scanned copies of *Weekly Return*, principally due to the age of the documents which posed a range of issues that ultimately reduce the likelihood of accurate letter and number recognition. For instance, the oldest copies of the *Weekly Return* from which data was abstracted are 80-years old. Over time, pages have suffered from discolouration, been afflicted by stains and creased due to repeated folding. Printing noise from the time of publication is also evident, characteristic of historic documents (Milligan, 2013). Together, these issues have contributed to the degradation of the printed text contained within the reports and prevent clean images of the *Weekly Returns* from being produced when scanned. Another issue regarded the non-standard typeface used in the *Weekly Returns*. This resulted in numerous recognition errors. OCR routines are often unable to detect the typeface used in historical documents unless the

specific typeface has been programmed into the software (Pal and Dash, 2014). Generally, OCR routines only recognise commonly used typefaces preloaded in a library stored within the OCR program.

2.3 Data Quality

The *Weekly Return* represents a remarkably complete and lengthy time-series of geographically aggregated infectious disease data at a fine temporal scale since the incidence of notifiable infections were recorded separately for each local government district every week. The availability of parallel information at the same geographical level, in the form of annual statistical reports produced by the Registrar-General and local district medical officers, provides a wealth of data on host demography and immunisation practices (Grenfell et al., 2001). This information is invaluable for placing observed epidemiological patterns within their wider ecological context. It is worth noting that the *Weekly Return also* provided space for the amendment of notifications in previous returns, acknowledging the issue of erroneous returns, as well as highlighting their provisional nature.

However, the morbidity data provided by the *Weekly Returns* suffers from the same limitations which affect many other historical, observational time-series data for infectious diseases. Notified cases are based on clinical diagnoses made and recorded by GPs. In some cases, diagnoses of childhood diseases may be made erroneously due to confusion with other diseases with similar symptomology. This is a greater issue among rural practitioners less accustomed to cases of diseases that are predominantly associated with densely populated urban areas (Smallman-Raynor et al., 2003). Additionally, sub-clinical cases of infection may go undiagnosed and escape notification (Noah, 2006), another source of error within the infectious data reported by the *Weekly Returns*. Consequently, the Weekly Returns do not present the full magnitude of disease incidence within a local

authority area. As Clarkson and Fine (1985) explain, there are a series of actions that must take place for an infectious disease case to be correctly notified. Firstly, the infected individual must suffer from a clinical version of the disease, exhibiting diagnosable symptoms. Second, the infected individual must be seen by a medical practitioner such as a GP who is responsible for notifying diseases. Third, the physician/medical practitioner needs to make an accurate diagnosis. Differences between these actions are largely responsible for explaining the inefficiencies in disease notification and thereby differences in data quality for individual childhood infections. The greatest difference between infections arguably lies with the exhibition of clinical disease in infected individuals. For instance, the symptomology of measles is almost universal and typically characterised by a clinical expression of the illness centred around the distinctive rash it produces. This contrasts considerably with pertussis, which has a much greater frequency of asymptomatic cases, and a range of mild forms of the disease are known to circulate among populations (Cherry, 1998).

Data quality is an important factor to consider when analysing spatiotemporal changes in endemic threshold populations since incomplete observations have the potential to obscure dynamical processes such as local extinction of infection, complicating estimates of endemic threshold size (Gunning et al., 2014). Under-reporting of cases can cause the duration of periods of zero case reports to increase, which leads to the endemic threshold size being overestimated. On the other hand, over-reporting may cause infections to appear more persistent in time and space than would be expected, but the effect of over-reporting is generally more difficult to quantify with existing data.

Correcting for incomplete observation poses a range of challenges. For instance, identifying the difference between under-reporting and stochastic extinction can be a complicated task for settlements where the disease is often on the verge of stochastic extinction (Gunning and Wearing, 2013). This issue arises from utilising a disease record that captures not only

information on disease persistence post-epidemics but also the invasion dynamics associated with the reintroduction of infection from neighbouring districts (Conlan et al. 2009). Bartlett (1957) recognised that the delays in disease reporting and under-reporting of infections needed to be considered when examining the persistence of infection since it cannot be assumed that a single week with no reported cases of infection within a district is a reliable indication that the chain of disease transmission has been broken locally. To remove ambiguity concerning whether stochastic extinction within a district has truly taken place or whether a disease has simply gone unreported, Bartlett introduced fadeouts as a measure of persistence.

An important factor that may influence the quality of disease reporting for all three infections under analysis during the early years of the study period is the disruptive effect of wartime. The first five years of the study period are during World War II, and the caveat of complications arising from conflict must be taken into the account with regards to the quality of the surveillance data collected during this period. For instance, medical statisticians from the period suggested that widespread disruption to existing public health systems in industrial centres vulnerable to aerial bombardment may have increased the likelihood of under-reporting, reporting delays and misclassification issues in provisional case reports (Smallman-Raynor and Cliff, 2015). This could be the result of school closures, and the evacuation of children, whether that be official or unofficial (Stocks, 1941). However, it has also been postulated that the 1102 ‘safe’ local governments districts to which evacuees were sent, referred to as reception areas, may have experienced inflation in clinical diagnosis and notification rates compared to what would have been expected in peacetime (Smallman-Raynor et al., 2003). This resulted from the provision of free home visits by GPs for evacuees, organised by the Ministry of Health, the vigilance of anxious foster parents and the attentiveness of teachers who closely watched evacuees of school-age (Stocks, 1941: 337). It is difficult to gauge the effect of evacuation on wartime disturbances on the quality of measles and pertussis surveillance data since statuary notification for the

diseases across all of England and Wales did not begin until two months after the outbreak of war (Stocks, 1942).

An assessment of the quality of pertussis, measles and scarlet fever notification data recorded in the *Weekly Returns* across the study period, based on the findings of previous empirical studies, will now be provided.

2.3.1 Pertussis

According to Clarkson and Fine (1985), only an estimated 5 to 25% of the actual number of pertussis cases are believed to have been notified in the surveillance data contained within the *Weekly Returns*. This is despite detailed surveillance procedures put in place for pertussis since the end of 1939 by the General Registrar Office, across all LGDs in England and Wales. There is some empirical evidence to support the veracity of Clarkson and Fine (1985)'s estimate. A past study revealed only 18.7% of pertussis cases clinically diagnosed by doctors in Nottingham Area Health Authority in 1982 were notified to the OPCS (Jenkinson, 1983). Clarkson and Fine (1985) also highlight that, before 1976, the reporting efficiency for pertussis may have been up to five times greater among spotter practices, who notify pertussis cases to the Royal College of General Practitioners, than by the average GP. The differential between those who were part of a sentinel reporting system and clinicians in standards practices may emphasise the difficulties involved with diagnosing pertussis, since it is not uncommon for the characteristic symptoms, such as the distinctive cough, to be absent. Although notification inefficiency affects the magnitude of case reports, it does not affect the fundamental large-scale spatiotemporal trends that can be observed within the surveillance data. For instance, Rohani et al. (1999) state that previous empirical work on pertussis surveillance has exhibited a significant correlation between notified cases reports, such as the Registrar General's Weekly Returns,

serological isolation of *Bordetella pertussis* by the Public Health Laboratory Service, and independent notification data for pertussis held by the Royal College of General Practitioners (Fine and Clarkson, 1982; Miller et al., 1992). It is worth noting that the issue of waning immunity among those infected by pertussis later in life has been overlooked regarding the under-reporting of pertussis cases in England and Wales during the mid-twentieth century. During the pre-vaccine era, pertussis reinfection in adults commonly went undiagnosed and was not well-characterised (Gunning et al., 2014), with pertussis notifications in the *Weekly Return* consisting almost exclusively of cases of infection amongst children (Cherry, 1998).

2.3.2 Measles

Measles records in England and Wales, compared to surveillance records on other childhood infections, are generally considered to be of excellent quality, partly explaining the prevalence of measles studies within the fields of medical geography and spatial epidemiology. This is due to a combination of factors. Firstly, measles is generally considered to be far more straightforward to diagnose than other childhood infections (Bjørnstad and Grenfell, 2008), resulting in more accurate diagnoses of cases and reliable surveillance data. This is due to the characteristic rash routinely associated with the disease, and the presence of Koplik's small spots; bluish-white spots located on the inside lining of the cheek (Black, 2013). Although diagnoses of measles cases may have been more accurate historically compared to other diseases, Conlan et al. (2009) note that in practice contemporary clinical diagnoses rates can be relatively low, with the then-Health Protection Agency (now Public Health England) reporting a case confirmation rate of around 20%. Anecdotally, an association between the accuracy of measles diagnoses and abundance of the infection has been discussed, with case confirmation rates as low as 1% observed during periods of low measles incidence in England and Wales throughout the

2000s (Conlan et al., 2009). Higher case reporting rates for measles in LGDs during the study period would therefore be expected, given the scale and frequency of measles epidemics before the introduction of mass vaccination in 1968. Secondly, the disease has long constituted a major public health threat and therefore was subjected to mandatory notification nationwide in England and Wales by 1940.

However, under-reporting of measles cases in England and Wales has received attention in past studies that have attempted to quantify the scale of this issue. According to previous empirical studies (Clarkson and Fine, 1985; Finkenstädt and Grenfell, 2000; Finkenstädt et al., 2002), the under-reporting bias for measles ranges from 40 to 60% between 1940 and 1969, and it has been previously stated that measles notifications would need to be multiped by a factor of 1.5 to 2 to provide a truer reflection of the magnitude of cases during this period (Cliff et al., 1981). Gunning et al. (2014) provide an extensive analysis of measles reporting rates in England and Wales, utilising a dataset formed of weekly measles notifications for sixty towns and cities between 1944 and 1968. This is a subset of a much larger dataset of measles case reports for all LGDs in England and Wales as recorded by the *Weekly Returns* over the same period (Grenfell et al., 2001). Their study found significant variability across geographical areas for measles reporting rates in England and Wales, partly explained by the idiosyncratic nature of notification and data collection by public health officers and practitioners at the local level. Moreover, demographic factors such as school attendance explain a non-trivial proportion of variation in reporting rates.

2.3.3 Scarlet fever

Despite many thorough literature searches, to date, there are no studies that indicate the scale of under-reporting associated with scarlet fever notifications during the mid-twentieth century in England and Wales. This parallels the paucity of empirical studies on scarlet fever morbidity for this period in medical geography and epidemiological literature. This is

unsurprising as scarlet fever has largely faded from public consciousness and the minds of medical experts, due to the power of antibiotics to treat infection and the collapse in disease morbidity. The significantly greater availability of spatiotemporal datasets already primed for analysis for measles and pertussis, which have continued to pose challenges to public health even after the introduction of mass vaccination, has provided much more fertile ground for the geographical and epidemiological study of childhood infections.

Based upon what is understood regarding the quality of pertussis and measles notifications throughout the study period in England and Wales, and the similarities in symptoms exhibited in scarlet fever cases compared with measles, it would be fair to assume there would be a substantial underreporting bias across Lancashire and South Wales. There is some indication of the quality of scarlet fever case reports in the *Weekly Returns* during the early years of the study period, during the 1940s, albeit limited. A study undertaken on sickness in the population of England and Wales during the mid-1940s by Dr Percy Stocks, Chief Medical Statistician of the GRO, determined that the notification data for scarlet fever collated in the *Weekly Return*s were 'fairly complete' in nature (Smallman-Raynor et al., 2003; Stocks, 1949). An earlier paper noted that a significant observation error had been identified in the notification of scarlet fever cases around the late-1930 (Stocks, 1941).

2.4 Database Formation

For the study period, weekly disease counts, annual mid-point population estimates, number of births, population density (persons per acre) for local government districts were abstracted from the *Weekly Returns* to form a 125 (geographical unit) × 1560 (week) space-time matrix of case notifications for pertussis, measles and scarlet fever in Lancashire, and a 74 (geographical unit) × 1560 (week) space-time matrix for South Wales. These datasets facilitate the various quantitative analyses laid out and presented in this chapter and thesis.

Since population density is measured in person per acre in the *Statistical Review*, a decision was made to transform the measurement of population density to persons per square kilometre (per km^2). This transformation of the arithmetic density from persons per acre to square kilometre allows for much greater variation in density levels between geographical units, providing a more nuanced perspective of settlement patterns and relative densities in terms of urban vs rural, thus painting a more accurate picture of the nature of population density within a region composed of numerous subpopulations.

2.5 Data Issues: Measuring Uncertainty

Uncertainty estimates are used to examine how the frequency at which data is sampled affects the estimation process (Capaldi et al., 2012). To date, within the field of disease ecology and epidemiology, uncertainty has often been considered primarily in terms of inadequate surveillance (of either hosts or pathogens) or the often accidental misclassification of cases of infection (see Section 2.3). Although absent here, an uncertainty analysis that attempts to formally quantify the limitations of the available data utilised to produce the datasets subject to analysis in this thesis could have been conducted. Estimates of disease persistence and of risk factors, which extrapolate from specific data sources to population-level measures, are subject to a broader range of uncertainty because of the combination of multiple data sources and value choices. One should consider all sources of uncertainty, including those occurring from measurement error, systematic biases, and extrapolation to compensate for incomplete data. Fine and Clarkson (1983) detail a relatively straightforward method for estimating the efficiency of historic disease notification with specific reference to England and Wales. Crude estimates are obtained from a comparison of annual numbers of births and notifications, modified to include detailed age-specific data. These analyses provide evidence for a strong positive correlation between notification efficiency and incidence for pertussis and measles, detecting a dramatic fall in the notification efficiency for pertussis between 1957 and 1976.

Part Two: Data Analysis

2.6 Exploratory Data Analysis

Empirical time-series plots of incidence rates (cases per 100,000 population) for the Lancashire and South Wales regions were produced to analyse the temporal trends of measles, pertussis and scarlet fever incidence throughout the study period. Time-series plots of a tiny subset of three districts of varying population size for Lancashire (Manchester CB, St Helens CB and Little Lever UD) and South Wales (Cardiff CB, Merthyr Tydfil CB, and Fishguard & Goodwick UD), were produced to illustrate the epidemic behaviour of Type I, Type II and Type III communities within the regional populations.

Choropleths maps are utilised to visualise geographical and temporal changes in patterns of measles, pertussis and scarlet fever persistence across the study period, as measured by percentage endemicity, using the 'moving window' approach. (see Section 4.6.1). Choropleth maps are also utilised to explore overall geographical patterns and compare rates of measles, pertussis and scarlet fever incidence for each of the nine time–windows across the study period. Due to the fall in the magnitude of cases for scarlet fever and pertussis in each region as visualised by the time series plots, it was necessary to define classes using the manual classification method to ensure consistent data intervals are available to allow comparison across all time–windows. This also applied to measles since the magnitude of cases can vary considerably depending on the presence of an epidemic outbreak within a time–window.

Analyses of sample correlation coefficients are performed to assess the level of spatial synchronicity between individual districts and the overall regional pattern of epidemic activity for all three diseases in the two regions. This was achieved by calculating the correlation between the reported annual counts for each district and the mean average over

the remaining number of districts in the region. The frequency distributions of sample correlation coefficients for individuals LGDs by disease and region are visualised using histograms and geographically displayed using thematic choropleth maps. All choropleth and proportional symbol maps were produced using QGIS 3.12 software 'Bucreşti', an open-source desktop geographic information system application. Time-series plots and heatmaps were produced using R package `ggplot2`.

2.7 Endemic Threshold Estimation

Following the approach of Cliff et al. (2000), a dynamic 'moving widow' empirical regression approach is pursued to evaluate how the endemic threshold size of three diseases in two unique geographical regions responds to demographic changes over time, as well the introductions of disease interventions and evolution of spatial relationships between local populations over the same period. A full description of the 'moving window' regression approach for endemic threshold estimation is presented below.

2.7.1 Time–Windows

Building on the empirical regression approach outlined by Cliff et al. (2000), the thirty-year time series of weekly pertussis, measles and scarlet fever notifications was broken down into nine 72-month time–windows to track temporal changes in the endemic threshold value in Lancashire and South Wales across the study period, yielding nine threshold size estimates. The purpose of employing this method was to explicitly monitor any systematic time changes across the study period that may affect the endemic threshold population size. This dynamism is important since the study period spans a length of time that saw transformative demographic events, advancements in public health and disease control and socio-economic changes that would fundamentally affect both regional populations, the

nature of and conditions for disease persistence. These changes over time are discussed in greater detail in various sections of Chapter 3.

A time–window of 72 months duration was selected to ensure a satisfactory number of windows to sufficiently detect temporal changes in threshold population size in a relatively short study period, whilst allowing a sufficient number of months in each time window to produce estimates from model parameters without unreasonably large standard errors. Time-series analyses of pertussis, measles and scarlet fever in Lancashire and South Wales reveal epidemic outbreaks occurred approximately every two to five years during the study period. Consequently, a 72-month window ensures more than one epidemic for each disease is captured in all nine time–windows. There is a 36-month overlap to ensure a smoothing effect between preceding and successive windows. Potential issues with this approach are discussed in section 4.6.3. 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 six time–windows constitute the pre-vaccine era of the study period, with the latter three time–windows forming the vaccine era. The population size of each local government district (LGD) for each time window was defined as the mean population size for the 72 months. Similarly, the number of susceptibles input (birth rate) for each LGD is defined as the mean number of susceptibles input across the length of the time window, and population density (number of persons per km^2) for each LGD in each time window was defined as the mean population density across each time–window.

2.7.2 Modelling Procedure

Weekly case notification data for all three diseases in each time–window was transformed into an absence/presence dataset, using a binary code (where 1 = infection present, and 0 = infection absent). This enabled the calculation of the proportion of weeks in each period that disease was present, measuring the level of disease persistence in each district. By

multiplying the proportion of weeks by 100, one obtains the percentage endemicity for each infection for all districts. The percentage endemicity variable was combined with annual mid-point population estimates to produce detailed datasets for each time–window; the annual number of births and population density (persons per acre) for local government districts were abstracted from the *Registrar General's Statistical Review*.

Following this, a linear regression was fitted for each time–window, with percent endemicity for each LGD against mean population size over the same period. To estimate the endemic threshold size for each window, the approach of Cliff et al. (2000), building upon the approach employed by Black (1966) in his study of measles endemicity in insular populations, is adopted, with the threshold population estimate being determined initially by using a simple linear regression of the form:

$$(\text{percentage endemicity}) = \hat{b}_0 + \hat{b}_1\,(\text{mean population size}), \qquad \text{(eq. 2.1)}$$

Since variables range over many orders of magnitude, and it is unknown what form the regression relationship may take given the nature of the data under analysis, with a large number of districts in both regions under study with significant variation in population sizes. Endemic threshold size estimates are calculated additionally using a simple linear regression of the form log-linear:

$$(\text{percentage endemicity}) = \hat{b}_0 + \hat{b}_1 \log(\text{mean population size}), \qquad \text{(eq. 2.2)}$$

And of the form log:

$$\log(\text{percentage endemicity}) = \hat{b}_0 + \hat{b}_1 \log(\text{mean population size}), \qquad \text{(eq. 2.3)}$$

2.7.3 Analysing Connectivity

To analyse the effect of geographical connectivity and isolation on the estimation of endemic threshold size, a dummy variable was introduced to the regression equation utilised to calculate the estimate. In the absence of weekly or annual data on migration/population flows between LGDs in Lancashire and South Wales, a binary distinction between the most connected and isolated LGDS was made according to distance from the nearest endemic centre, which acts as a crude connectivity index. To define the distance from the nearest endemic centre, a Euclidean distance matrix was calculated from the centroids of each LGD polygon in a shapefile of the administrative district level boundaries for Lancashire and South Wales. Liverpool CB and Manchester CB in Lancashire and Cardiff CB in South Wales were defined as endemic centres, since all three LGDs consistently report, or very close to, 100% endemicity across all nine time–windows. To produce the dummy variable, distance from the endemic centre was encoded, with LGDs below the median distance from the nearest endemic centre coded as 0, and LGDs greater than the median distance from the nearest endemic centre as 1. Utilising the models above and the estimated values of parameters, the endemic threshold is again calculated, but for two circumstances; when the dummy variable is 1 and when the dummy variable is 0.

2.7.4 Analysing Density

Black (1966) observed an inverse relationship between population density and duration of individual epidemics, resulting in variation in the prevalence of measles cases. Black argues implicitly that the epidemic infection persists longer in dispersed populations rather than crowded populations, affecting the number of months in which the disease is reported and thus complicate a simple population-based estimate of critical community size. In other words, greater geographical dispersion results in a damping effect that depresses the calculated endemic threshold value. Following the same approach as described to analyse

the relationship between distance formed endemic centre and endemic threshold size, 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. Population density per square kilometre acts as a simple proxy for susceptible density in the endemic threshold calculation assuming a linear relationship between the density of susceptibles and most densely populated districts.

The 'moving window regression analyses yielded a total of nine threshold estimates across the study period for the full population-based sample of Lancashire and South Wales districts for each disease, as well nine threshold estimates for low and high connectivity districts and low and high-density districts for each region. The threshold estimates were plotted, using a simple scatterplot approach with circles denoting the estimated threshold size in each window and a LOWESS smoother fitted to show time trends. In the case of pertussis, scatterplots feature a shaded area indicating the period when successful preventive measures for disease elimination were available and applied nationwide.

2.7.5 Modelling Issues

In the case of Lancashire, a key issue was the swamping effect of two outliers which persisted in each of the nine-time–windows. These outliers were Liverpool CB and Manchester CB, both possessing populations roughly three to four times greater than the next largest district, Salford CB. Consequently, these urban centres yielded excessive influence over the fitting of the model to the case notification data. 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 the models before estimation. To evaluate the extent to which Liverpool CB and Manchester CB affect the calculation of endemic threshold

estimates during the study period, preliminary analyses were performed with empirical regression models for each disease across all time–windows fitted with Liverpool CB and Manchester CB included in the modelling procedure and comparing the estimates with those calculated with the two districts omitted. It was found that the omission of Liverpool and Manchester CBs resulted in only marginal increases in the regional threshold estimates across the time–windows. Consequently, Liverpool CB and Manchester CB were omitted from the models before estimation. A fuller discussion of the problem of using percentage data in regression is provided in Appendix II.

Another issue is that the 36-month overlap to ensure a smoothing effect between preceding and successive windows poses a challenge to the assumption of stationarity. Stationarity is an important property and issue in time series. However, real-world time series are often non-stationary, with significant properties such as mean, frequency, variance and kurtosis changing over time. Often such time-series possess high volatility, trend and are frequently characterised by heteroskedasticity. Although not employed here, one remedy would have been to employ a moving average technique to produce a smoothing effect, remove noise fine-grained variation between each period represented by the nine time–windows. The time–window size could have been specified by defining the window width and the number of raw observations used to calculate the moving average value. The moving window, defined by the window width, would then slide along the time series to calculate the average values in the new series.

2.8 Hotspot Analysis

The purpose of the hotspot analysis is to investigate the effects of spatial coupling, connectivity and impact on mass vaccination on endemic persistence of pertussis in Lancashire and South Wales, by analysing differences in the number and geographical distribution of hotspots between pre-vaccine (1946–1957) and vaccine eras (1958–1969) for the disease. To this end, two time series of 144-months duration, representing the two eras, are studied. The onset of vaccination midway through the study period serves as a natural experiment to analyse the two mechanisms which can produce a lower than expected number of fadeouts:

I. increased chains of transmission between and within subpopulations due to high population density, and
II. higher rate of disease re-introduction due to high geographical connectivity between subpopulations and endemic centres.

Here, a fadeout is defined as a period of four weeks or more without reported cases of infection. The detection of potential hotspots in the vaccine era can reveal important information concerning potential endemic reservoirs and the location of corridors of infection, where regional movement patterns between subpopulations are concentrated and the spatial import or infection is a frequent occurrence (Cliff et al., 1993; Xia et al., 2004).

2.8.1 Total Fadeouts & Population Size (Pre-vaccine era)

Following the method of Bharti et al. (2010), the hotspot analysis is centred on an evaluation of the residuals detected by the OLS linear regression models. These residuals can potentially reveal areas of key epidemiological importance for the regional persistence of pertussis in Lancashire and South Wales. LGDs with negative residuals would 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.

OLS linear regression models were fitted to analyse the association between the total annual number of fadeouts and mean population size for both regions to estimate reintroduction events. The total number of fadeouts should scale inversely with population size, due to the increased likelihood of transmission events (Conlan et al., 2007). An overall negative correlation between the number of fade-outs and population size is well-established in previous research empirical research on measles and, to a lesser extent, pertussis. This is due to stochastic fluctuations in birth, death and migration rates (Bartlett, 1957; Black, 1966; Bjørnstad et al., 2002) alongside the natural dynamical activity of childhood disease. Consequently, a negative relationship between local population size and the total number of annual fadeouts in each 144 month-long time series is expected.

To provide an initial assessment of the impact of spatial proximity and human mobility on pertussis persistence in Lancashire and South Wales, the OLS regression residuals were tested for spatial autocorrelation by performing a Moran's *I* test using the R package `spdep`. Since detailed data on the movement of individuals between LGDs in Lancashire and South Wales during the study period is not available, a geographical proxy for mobility and interaction between districts was utilised. A conventional contiguity-based spatial weighting that could capture characteristics of contagious diffusion was incorporated in Moran's *I* tests for both regions, with districts defined as neighbours where they share administration boundaries with common borders (Moran, 1950). Based on the rooks-contiguity relationship, an individual Lancashire LGD has an average of 4.91 neighbouring districts, whilst an individual LGD in South Wales has an average of 3.95 neighbouring districts.

2.8.2 Reported Cases (Vaccine era)

Pertussis hotspots in the vaccine era are defined as LGDs with a greater number of cases notifications for their population size than the mean regional total of cases reported (Bharti et al., 2010). These LGDs were identified after calculating the total number of weekly cases notified in each reporting district in the 144-month period following the introduction of routine pertussis vaccination. To test for spatial autocorrelation amongst 'hotspots', a Moran's *I* test is performed as described previously, only with hotspots treated as a binary variable (i.e., 1 = district identified as a hotspot; 0 = district not identified as a hotspot). To facilitate the identification of geographical patterns and clustering of regional pertussis hotspots in Lancashire and South Wales, hotspots were visualised by producing simple thematic maps. Thematic maps were created using QGIS 3.12 "București".

2.9 Survival Analysis

Rates of re-introductions

Comparing the length of inter-epidemic periods in the pre-vaccination and vaccine-era time series provides additional insight on spatiotemporal changes in disease persistence at the local level (Ferrari et al., 2008; Grenfell and Anderson, 1989), allowing an assessment of the impact of vaccination on the strength of spatial coupling to be assessed as well as the identification of districts of particular epidemiological importance to maintaining the regional circulation of childhood diseases. In the absence of vaccination, long inter-epidemic periods indicate low geographical connectivity, usually coupled with small population size and low susceptible input via births, resulting in infrequent to rare re-introduction of disease (Bartlett, 1957). Short inter-epidemic periods suggest frequent disease introduction due to the presence of 'rescue effects' and external transmission, as a consequence of high geographical connectivity, or larger populations (Grenfell et al., 2001; Wearing and Rohani,

2009). Building on the work of Bjornstad and Grenfell (2008) and Bharti et al. (2010), a survival analysis was conducted by fitting a Cox proportional hazard regression model, with inter-epidemic period length serving as the waiting time, to determine the survival probability of disease fadeouts. The waiting time represents the number of weeks without reported cases until a reintroduction event.

In the survival analysis presented here, the outcome is a disease reintroduction event. The time to pertussis reintroduction represents the length of fadeout duration. It is expected, based on previous research and epidemiological theory concerning the persistence of infection and critical community size, that districts with larger populations, greater population densities, the input of susceptible individuals via birth and a high degree of spatial coupling will experience much higher rates of disease reintroduction, resulting in much shorter times to disease reintroduction than less populated and dense districts which are more geographically isolated and possess lower rates of susceptible recruitment.

The survival analyses of pertussis endemicity in the pre-vaccine and vaccine eras for the Lancashire and South Wales region were performed using two R packages: `survival` for computing survival analyses and `survminer` for visualizing survival analysis results.

2.9.1 Cox Proportional Hazards Model

To successfully perform the survival analysis, a cox proportional hazards model must be fitted. The Cox proportional-hazards model is one of the most important methods used for modelling survival analysis data and can be applied to both quantitative predictor variables and for categorical variables. The model is essentially a regression model commonly used for investigating the association between the survival time of a group, such as infected individuals or patients, and one or more predictor variables (Cox, 1972). The objective of

using a cox model is to simultaneously evaluate the effect of known or hypothesised several factors on survival. In other words, a Cox model enables the examination of how specified factors influence the rate of a particular event happening such as infection or death at a particular point in time. This rate is commonly referred to as the hazard rate (Sedgwick, 2012). Predictor variables are usually termed *covariates* in the survival-analysis literature. Fitting a cox regression model enables the visualisation of the predicted survival proportion at any given time point for a particular group under investigation.

The Cox proportional hazards model is conveyed by the *hazard function* denoted by h_t. Briefly, the hazard function can be interpreted as the risk of expiring at time t. In the present analysis, the hazard function is the risk of disease reintroduction at time t. It can be estimated as follows:

$$h_t = h0_t \times \exp\left(b_1 x_1 + b_2 x_2 + \cdots + b_p x_p\right), \qquad \text{(eq. 2.4)}$$

where t represents the survival time $\mathrm{h_t}$ is the hazard function determined by a set of p covariates (x_1, x_2,..., x_p) and the coefficients (b_1, b_2,..., b_p) measure the effect size of the covariates. The term $\mathrm{h0}$ represents the baseline hazard. It corresponds to the value of the hazard if all the covariates are equal to zero i.e., the quantity $\exp(0)$ equals 1. A cox proportional hazards model can also be expressed as a multiple linear regression of the logarithm of the hazard on the variables x, with the baseline hazard being an intercept term that varies with time. The quantities $exp(b)$ represent hazard ratios. A value of b_p greater than zero, or equivalently a hazard ratio greater than one, indicates that as the value of the *n*th covariate increases, the probability of a hazard event increases, thus reducing the length of the survival time.

A hazard ratio is a measure of an effect of an intervention or covariate on an outcome I.e., dependent variable over time. Hazard ratios are most frequently reported in time-to-event

analysis or survival analysis when the aim is to establish the length of time required for a particular event or outcome to occur (Sedgwick, 2012). The outcome could be a negative outcome such as the time until death or a positive outcome, such as time to disease-free survival. When hazard ratios are used in survival analysis this reflects the analysis of time survived to an event (Altman and Bland, 1998). Alongside hazard ratios, confidence intervals are reported, providing the range of values that is likely to include the true population value, measuring the precision of the hazard ratio. The narrower the confidence interval, the more precise the estimate. The precision of any estimate will be influenced by the sample size to some extent. If the confidence interval includes 1, a hazard ratio is not significant. A hazard ratio above 1 indicates a covariate that is positively associated with the event probability (Spruance et al., 2004).

2.9.2 Kaplan-Meier Survival Curves

Kaplan-Meier survival curves were constructed for hotspots, and other districts in the pre-vaccine and vaccine eras, to assess the impact of disease intervention in the form of mass vaccination on the rate of disease reintroductions and duration of fadeout events. The Kaplan-Meier (KM) method is a non-parametric method used to estimate the survival probability from observed survival times (Kaplan and Meier, 1958). The method is graphical, displaying survival data or time-to-event analysis and is commonly drawn as a step function. The survival probability at time t, $S_{(t)}$, is calculated as follow:

$$S_{(t)} = S_{(t-1)}\left(1 - \frac{d_t}{n_t}\right), \qquad \text{(eq. 2.5)}$$

where $S_{(t-1)}$ equals the probability of being alive at $(t-1)$, n_t represents the number of patients alive just before t, and d_t is the number of events at t. The estimated probability $S_{(t)}$ is a step function that changes value only at the time of each event (Bland and Altman, 1998).

The Kaplan-Meier survival curve, a plot of the Kaplan-Meier survival probability against time, provides a useful summary of the data that can be used to estimate measures such as median survival time (Sedgwick, 2014). The Kaplan-Meier curve is a form of univariate analysis, describing the length of survival according to one factor under investigation. Additionally, Kaplan-Meier curves are useful when the predictor variable is categorical, effectively making comparisons between groups. Log-rank tests are utilised to test whether differences between hotspots and other districts in terms of fadeout survival probability are statistically significant, and to compare across the pre-vaccine and vaccine eras.

2.9.3 Log-rank Test

The log-rank test is a non-parametric test and the most widely used method of comparing two or more survival curves (Bland and Altman, 2004). The null hypothesis is that there is no difference in the overall survival distributions between the groups in the population (Mantel, 1966). The log-rank test makes no prior assumptions about survival distributions and compares the observed number of events in each group to what would be expected if the null hypothesis were true (Clark et al., 2003). The log-rank statistic is approximately distributed as a chi-square test statistic and thus, to test the null hypothesis, the log-rank test calculates a chi-square (χ^2) statistic, which is compared to a χ^2-distribution. If the p-value <0.05, then the result of the test is statistically significant; survival distributions of the different groups are not equal within the population. The R package survival was used to perform the log-rank test comparing the two survival curves of hotspots and other districts within the regional populations of Lancashire and South Wales in the pre-vaccine and vaccine eras. The test produces a weighted observed number of events in each group, a weighted expected number of events in each group, and a chi-square statistic for a test of equality.

2.10 Endemic–Epidemic Modelling

Weekly pertussis, scarlet fever and measles notification data for LGDs in Lancashire and South Wales obtained from the Weekly *Returns* are classic examples of infectious disease counts data collected by public health surveillance systems. Datasets of this nature are invaluable resources for extrapolating temporal and spatial parameters to improve our understanding and prediction of how infectious disease spreads geographically. Infectious disease data tends to be the product of inherently spatiotemporal processes which are only partially observable, and observations are not independent (Becker and Britton, 1999). Infectious disease data also tends to feature autoregressive, self-exciting behaviour, as a result of demographic stochasticity. To effectively model the endemic-epidemic dynamics of childhood infection, one must use a statistical approach that can effectively capture the autoregressive, spatial and temporal components of infectious disease data.

2.10.1 Modelling Rationale

The statistical analysis of infectious disease data has been predominantly dictated by individual-based mechanistic modelling of the epidemic process (Becker, 1989; Daley and Gani, 1999). In particular, continuous-time models, such as the susceptible-infected-removed (SIR) model, have been applied to estimate relevant parameters from detailed data on the infection process (Anderson and Britton, 2000). Mechanistic models diverge from simple empirical models such as regression models because their structure demands the formulation of explicit hypotheses about the potential biological mechanisms that influence infection dynamics (Lessler and Cummings, 2016). Such hypotheses range from straightforward representations of the time until parts of the disease process are complete, such as the incubation period, to complex agent-based models that attempt to explicitly characterise social interactions and networks (Eubank et al., 2004; Sartwell, 1966).

However, modelling of this nature is too ambitious for routinely collected surveillance data for areal units, which do not possess data on individual cases. The absence of detailed information on susceptibles essentially makes detailed mechanistic modelling of the infection process impossible.

Another issue that blights historical infectious disease data is underreporting, due to subclinical cases or misdiagnosis, and delays in case reporting (Diggle et al., 2002). Mechanistic models often assume that the time unit in which data are collected equals the generation time of the disease under analysis (Daley and Gani, 1999), yet this is rarely the case in practice; the generation time of measles and pertussis is almost twice as long as the time unit in which cases are reported in the *Weekly Returns*. This issue can result in significant overdispersion. Mechanistic modelling techniques possess other imitations which inhibit efforts to construct realistic stochastic models for the statistical analysis of historical time-series data of infectious diseases. The parameters in a traditional SIR model do not allow for the quantification of uncertainty. Calculating SIR models over a limited number of potential values for each parameter results in a range of future trajectories but does not quantify uncertainty in the predictions. Another significant limitation of the SIR model is the simple assumptions made about the population which underpin the model. It assumes homogeneous mixing of the population, assuming in individuals within a population have an equal probability of coming into contact with each other. This does not reflect the social structures which dictate human activity, which concentrate most contact between individuals within restricted networks (Tolles and Luong, 2020). The SIR model also assumes a large, closed population with no migration, births, or deaths from causes other than the epidemic. This fails to consider epidemiological coupling at the local level and the role of stochastic effects which are key to understanding disease incidence in small populations (Huppert and Katriel, 2013). Development of a greater understanding of stochastic effects on disease occurrences over time across local populations of varying

sizes is critical to understanding changes in endemic threshold populations and the spatiotemporal persistence of childhood diseases.

Alongside mechanistic approaches, empirical modelling has long been a feature of statistical analysis of infectious disease data. Unlike mechanistic modelling, the main requirement of an empirical model is to explain the variability in the observed data, rather than the underlying mechanism. Generalised linear models (GLMs) are a common method of empirical modelling, representing an extension of standard linear regression which is utilised for explaining and predicting count data, and are a class of fixed effects regression models that can accommodate non-normal responses and non-linear relationships between the response variable and covariates (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989; Diggle et al., 2002). GLMs allows the relationship between covariates and the response variable to be expressed additively in a linear formulation, and generally assume either Poisson or Binomial probability distributions for the response. Similar to conventional linear models, which is a particular type of GLM, standard GLMs assume independence between observations and that they are equally distributed. In the study of infectious diseases, it is often the case that disease counts are aggregated over geographical units and are compared to aggregated covariate summaries, using models such as log-linear Poisson models. This class of models is known as parameter-driven. Similar parameter-driven formulations with suitable prior distributions on latent parameters are utilised in the study of non-infectious diseases, such as counts of cancer incidence (Held et al., 2005).

Poisson regression is the standard method used to analyse count data. However, many real-life data situations violate the assumptions upon which the Poisson model is based. For instance, the Poisson model assumes that the mean and variance of the response are equal. If the variance is greater than its mean, there is likely heterogeneity in the data indicating the Poisson model is overdispersed. The potential causes of overdispersion when analysing large spatiotemporal disease count datasets are numerous; zero inflation (Deng

and Paul, 2005); misspecification of the probability model (i.e., selecting a Poisson model when a negative binomial distribution would capture a great amount of variation); the presence of spatial autocorrelation in datasets with geographical neighbours often tending to display residual spatial dependence (Haining et al., 2009). The violation of assumptions of independence among observations is a particular issue for the maximum likelihood estimation of both Poisson and negative binomial regressions, which requires independent observations (Barron, 1992) since disease counts tend to be clustered or aggregated. Model overdispersion can be checked by observing the Deviance-based dispersion statistic of a Poisson or negative binomial model; a dispersion value greater than unity (i.e., >1) indicates overdispersion.

It has been recognised that purely parameter-driven models such as the Poisson and Binomial GLMs described are often unable to describe epidemic activity at the local scale (Held et al., 2006). More realistic models with extensions are often required to consider and assess the influence of unobserved covariates that may affect disease incidence, reducing the presence of significant levels of overdispersion (Held et al., 2005). Additionally, empirical models often fail to adequately capture periodic epidemic outbreaks one tends to see in infectious disease data and no allowance is made for these outbreaks (Held et al., 2005: Paul and Held, 2011).

To avoid the limitations of parameter-driven empirical and mechanistic modelling approaches, Held et al. (2005, 2006) draw upon the branching process model with immigration (Bartlett, 1956; Guttorp, 1995). Branching processes are stochastic individual-based processes that play a fundamental role in epidemiological theory, particularly with regards to the threshold behaviour of epidemics and the calculation of mechanistic critical vaccination thresholds (Farrington et al., 2003). In a branching process model in the absence of immigration, a closed population of individuals developing under the usual laws for branching processes either increase indefinitely with time or become extinct (Heathcote,

1965). However, with the introduction of immigration to a closed population, a stationary distribution for population size will exist for processes in continuous time when the immigration distribution is Poisson (Bartlett, 1956).

Held et al. (2005, 2006) proposed a model framework that acts as a compromise between mechanistic and empirical modelling approaches, aiming to provide a realistic model capable of handling infectious disease count data from historical disease surveillance records which feature seasonal variation, periodic epidemic outbreaks and areas with low counts. The key feature of the model outlined by Held et al. (2005, 2006) is the additive decomposition of mean incidence, i.e., disease risk, into two components: endemic and epidemic. In dynamical models of infectious disease counts, the distinction between endemic and epidemic incidence is often made (Finkenstädt et al., 2002). The endemic component is parameter-driven, relating disease incidence to latent parameters such as the seasonal endemic rate, whilst also describing the risk of new cases to covariates independent of the history of the infection process. These covariates can include population density, socio-demographic variables, and vaccination coverage, which can all vary geographically and temporally. The epidemic component is observation-driven, allowing for explicit temporal dependence on the number of cases beyond parametric seasonal patterns, with the autoregressive parameter also allowing for periodic epidemic outbreaks in the data. Since the model is not mechanistic and does not assume the time unit for which data is collected as ‘generation time’, cannot be interpreted as the basic reproduction number. The two-component model for disease counts (see Appendix II) forms the methodological basis for the HHH model.

2.10.2 The HHH Model

Endemic-epidemic multivariate negative binomial time-series models, referred to as HHH models (Meyer et al., 2017), are applied to model pertussis, scarlet fever and measles areal data in each time window for the Lancashire and South Wales regions. This regression framework is appropriate for analysing disease counts aggregated by time and period (Held and Paul, 2012), and has been developed by building upon the Poisson branching process with the immigration approach outlined by Held et al. (2005).

The HHH 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). It is a stochastic model able to capture space-time dependence caused by the geographical spread of disease across time, by bringing the number of cases in different geographical units into consideration. Additionally, the model allows for overdispersion to adjust for unobserved covariates that affect disease incidence and the heterogeneity presented due to spatial correlation and temporal dependence of cases. The model permits non-stationarity, considering interventions on disease counts over time such as vaccination coverage and improvements in hygiene, as well as seasonality and the effect of extrinsic events on key demographic parameters which affect disease persistence, such as explosive birth rates in short periods of time. Extensions can be incorporated into the basic model formulation to assess the effect of seasonality, socio-demographic characteristics, temporal trends and localised disease dynamics on the endemicity and epidemicity of infection.

The moving window approach allows estimates of regional endemic threshold populations to be calculated over time 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 HHH 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 hotpots, potentially influencing the temporal changes in endemic threshold populations. The HHH modelling procedure was performed using the methodological tools provided by the *R* package `surveillance` (Meyer et al., 2017). A glossary of notation for key HHH model parameters can be found in Appendix II.

2.10.3 Model Formulation

The HHH model applied in the analysis for pertussis, measles and scarlet fever counts Y_{it} from geographical units $i = 1, \ldots, I$ during periods $t = 1, \ldots, T$, first outlined in its most simple form by Held et al. (2005, 2006) (see Appendix II), and extended in a series of later papers (Paul and Held 2011; Held and Paul, 2012; Meyer and Held, 2014), 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

$$Y_{it}|\boldsymbol{Y}_{\boldsymbol{t-1}} \sim \text{NegBin}(\mu_{it}, \psi), \qquad \text{(eq. 2.6)}$$

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

$$\mu_{it} = e_{it}\nu_{it} + \lambda_{it}Z_{,t-1} + \phi_{it}\sum_{j \neq i}\omega_{ji}\,Y_{j,t-1}\,, \qquad \text{(eq. 2.7)}$$

where e_{it} is the offset of known counts reflecting population at risk and ω_{ji} is the weight for the neighbourhood component reflecting the 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

$$\log(v_{it}) = \alpha_i^{(v)} + \beta^{(v)^T} z_{it}^{(v)}, \quad \text{(eq. 2.8)}$$

where v is the unknown endemic parameter. The second (autoregressive) component accounts for autoregressive effects; the reproduction of disease within district i

$$\log(\lambda_{it}) = \alpha_i^{(\lambda)} + \beta^{(\lambda)^T} Y_{it}^{(\lambda)}, \quad \text{(eq. 2.9)}$$

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

$$\log(\phi_{it}) = \alpha_i^{(\phi)} + \beta^{(\phi)^T} Y_{it}^{(\phi)}, \quad \text{(eq. 2.10)}$$

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 $\alpha_i^{(\phi)}$ are component-specific intercepts and $\beta^{(\lambda)^T}$, $\beta^{(\lambda)^T}$ and $\beta^{(\lambda)^T}$ 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. Each parameter is also allowed to vary over time to reflect situations where the infectiousness of disease changes over time. For instance, this may be due to the implementation of immunisation programmes or other public health interventions, seasonality or through external factors which influence the spread of infection. Other scenarios include a declining number of susceptibles over time,

which would effectively decrease λ, and sudden outbreaks where $\lambda_t > 1$ for a limited time period, which is allowed to be estimated from infectious disease data.

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 HHH models fitted for each time window. The significance of the two epidemic components is assessed using dominant eigenvalues ($maxEV$), 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, maximum likelihood estimation is used to estimate parameters and standard errors, by maximising the negative binomial log-likelihood of the model. The HHH model framework allows for covariate effects on either the endemic or epidemic components of disease incidence to be included using model extensions.

Weekly case notifications for measles, pertussis and scarlet fever obtained from the *Weekly Returns*, for all administrative districts in Lancashire and South Wales, take the form of aggregated counts by region and period, which can be loaded into *R* from external data sources (Höhle and Mazick, 2010). For count data to be analysed via the application of HHH models in the surveillance package, a data object of the class *sts* (surveillance time series) must be created. Essentially an object of the class *sts* involves three data matrices: observed counts z_{it} are stored in the *T x I* matrix (observed), a corresponding matrix with time-varying population numbers (or fractions), and an *I* × *I* neighbourhood matrix quantifying the spatial coupling between the *I* units. The observed and population matrices are 312 × 125 for each Lancashire time window and 312 × 74 for each South Wales time window. 312 is equivalent to the total number of weeks/observations for each time window. To incorporate spatial interaction in the model, the neighbourhood matrix consists of

adjacency orders o_{ji} between the districts. (Bivand et al., 2013). Data of the *sts* class can be visualised in four distinct ways: an individual unit time-series of weekly counts, an overall time series plot, a choropleth map of disease incidence by district, or animated maps. The default plot type is *observed ~ time*, producing an overall time series of count data. When this is modified to *observed ~ time | unit*. it is possible to visualise the district-specific time series of disease counts.

2.10.4 Model Extensions

Spatial Interaction

With multiple geographical units under analysis (125 LGDs in Lancashire, 74 LGDs in South Wales), spatiotemporal dependence is adopted by the third component in eq. 4.10. Weights w_{ji} in the neighbourhood component reflects the strength of transmission from region j to region i, collected into an $i \times i$ weight matrix (w_{ji}):

$$w_{ji} = \begin{cases} \frac{1}{n_{j,}}, \\ 0, \end{cases} \qquad \text{(eq. 2.11)}$$

for $i \sim j$, where the symbol $\sim$ denotes *is adjacent to* and n_j is the number of first-order neighbours of district *j*. This equation represents a normalised version of the binary and symmetric adjacency indicator matrix $\mathbf{A} = (1(i \sim j))j_{j,i=1,\ldots,I}, i = 1,\ldots,I$. The purpose of normalising the adjacency matrix is to ensure district j distributes its cases uniformly to n_j neighbours (Paul et al., 2008). Due to normalisation, the weights w_{ji} for transmission from district j to district i is determined not only by the districts' neighbourhood o_{ji} but also by the total amount of neighbourhoods of district j in the form of $\sum k \neq j\, o_{ji}^{-d}$. This results in variation amongst transmission weights for a specific order of adjacency (Meyer and Held, 2014).

In its basic formulation, a HHH model assumes the spread of infection is restricted to first-order neighbours; all districts have the same epidemic potential for importing cases from adjacent units (Meyer et al., 2017). Disease transmission only takes place between neighbouring districts during the period $t \rightarrow t + 1$, with the single exception to this is the independent importation of disease cases via the endemic component (Meyer and Held, 2014). 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 the 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 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). Ideally, transmission weights would be calculated using existing movement network data (Paul et al., 2008; Geilhufe et al., 2014), yet detailed data for the movement of individuals between the numerous LGDs in Lancashire or South Wales between January 1940 and December 1969 is unavailable. The HHH modelling framework can be extended to account for short-range, commuter-driven spread and long-range transmission of cases between districts by incorporating a gravity model of spatial interaction and power-law extension.

Gravity Model

The gravity model is the most commonly used formulation of spatial interaction analysis (Gatrell and Bailey, 1996), and has been widely applied in a wide variety of fields, such as migration, trade and commodity flows, transportation theory and, increasingly, spatial

epidemiology. Its name is derived from its resemblance to Newton's law of gravity (Barrios et al., 2012). According to its most generic form, the gravity model states that the attraction a_{ij} between two objects i and j is directly proportional to their mass, m_i and m_j and inversely proportional to the distance separating them, d_{ij}, as follows:

$$\alpha_{ij} = m_i m_j d_{ij}^{-2} \quad \text{(eq. 2.12)}$$

This formulation can be adapted to reflect this basic assumption about spatial interaction, as follows: The movement between two communities k and j is directly proportional to their community size, n_k and n_j, and inversely proportional to the distance between the two areas, d_{kj} (Xia et al., 2004).

In recent years, there has been an increasing number of applications of the gravity model in the analysis of spatiotemporal patterns of disease persistence, particularly in metapopulation models of infections such as measles (Xia et al., 2004; Jandarov et al., 2014) and influenza (Viboud et al., 2006; Eggo et al., 2010). Gravity models allow realistic connectivity structure, which might be distance-dependent, to be incorporated into stochastic modelling frameworks (Barthélemy, 2011), enabling the study of spatial coupling patterns between subpopulations and hierarchical transmission between cities, towns and villages (Xia et al., 2004). The addition of a gravity model of spatial interaction to the HHH model allows the model to describe spatiotemporal patterns of endemic and epidemic incidence are affected by the network of spatial spread of the disease (Cliff et al.,1993).

To reflect commuter-driven spread, the model is extended to account for the district-specific population in the spatiotemporal epidemic component. District susceptibility to infection is scaled according to population size, multiplying the neighbourhood parameter (Φ) by district population size ($e_i^{\beta pop}$).

Power-law decay of Spatial Interaction

To reflect the long-range transmission of cases and epidemiological coupling between geographical units (Keeling and Rohani, 2008), is 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).

To apply the power-law principle in the complex network of geographical districts in the Lancashire and South Wales regions, the measure of distance on which the power-law acts first needs to be defined. It could be defined according to one of the following methods: Euclidean distance between district centroids calculated according to their coordinates, and the order of neighbourhood. Euclidean distance follows a continuous power-law, whereas the second one is discrete. Yet calculating Euclidean distance via centroid coordinates is problematic when applied to contiguous data, due to the variation in shape and area of individual districts. Specifically, a very small neighbouring region would be attributed a stronger link than a large neighbour with a centroid further apart, even if the latter shares more boundaries than the former. However, using common boundary length as a measure of spatial coupling would also be limited since this would only cover adjacent geographical districts (Keeling and Rohani, 2002). For these reasons, the power-law extension will be defined according to the discrete measure of neighbourhood order.

Formally, two districts, district j and district i, are kth-order neighbours if the shortest distance between them has k steps across distinct districts. The network of districts thus features a symmetric $I\ x\ I$ matrix of neighbourhood orders with zeroes on the diagonal by convention. According to this discrete distance measure, the previously used 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 to be estimated.

The raw power-law weights w_j can be normalised to

$$w_{ji} = \frac{o_{ji}^{-d}}{\sum_{k=1}^{I} o_{jk}^{-d}}, \quad \text{(eq. 2.13)}$$

such that $\sum_{k=1}^{I} w_{jk} = 1$ for all rows j of the weight matrix. As the decay parameter d increases, Higher-order neighbours diminish in importance. The limit $d \to \infty$ corresponds to the previously used first-order dependency, whereas d = 0 assigns equal weight to all districts.

Random Effects

In statistical modelling of infectious diseases, random effects are a common approach to accounting for unobserved heterogeneity. For instance, using a much-analysed dataset on measles outbreaks in Providence, Rhode Island, Li et al. (2003) proposes a random effects model which assumes the probability of avoiding infection varies randomly within households, allowing for household-dependent heterogeneity in measles transmission rates. This provided a better fit to the Providence measles data, compared to a standard modelling approach fitting a chain binomial model. The use of generalised linear mixed models for the analysis of longitudinal infectious data which explicitly accounts for heterogeneity across subjects has also been applied regularly to determine the influence of random effects on disease incidence (e.g., Kleinman et al., 2004; Lee et al., 2018; Zeger and Karim, 1991). The spatial analysis of disease risk has also been an area where random effects are routinely considered. For example, when observed data are scarce, a maximum likelihood approach can lead to very unstable estimates of area-specific risk and linear trends due to significant random variation in area effects, i.e., extra-Poisson variation (Breslow, 1984).

In the case of HHH modelling, random effects can be incorporated to account for the influence of unobserved covariates in regions with a large number of districts that exhibit heterogeneity in epidemic and endemic disease incidence (Paul and Held, 2011). This approach is preferable to allow for variation in the transmission probability of an infectious disease agent across regions by simply including region-specific autoregressive parameters into the model (Paul et al., 2008). If the time series is highly multivariate with a high number of districts included, the estimation procedure can get unstable and identifiability issues can occur, and therefore the aforementioned approach is limited to a small to moderate number of regions.

For surveillance data on childhood infections, a common example of unobserved heterogeneity is under-reporting (Broutin et al., 2005; Gibbons et al., 2014). The second source of unobserved heterogeneity is edge effects, with districts on the borders of the two regions missing potential sources of infection from across the border, with districts in other counties and regions with which they share significant spatial interaction.

To improve the fit of previous HHH model formulations, district-specific intercepts in the endemic or epidemic components are allowed, and the following independent random effects in all three components are included:

$$\alpha_i^{(v)} \sim^{iid} \mathrm{N}\left(\alpha^{(V)}, \sigma_V^2\right), \quad \text{(eq. 2.14)}$$

$$\alpha_i^{(\lambda)} \sim^{iid} \mathrm{N}\left(\alpha^{(\lambda)}, \sigma_\lambda^2\right), \quad \text{(eq. 2.15)}$$

$$\alpha_i^{(\phi)} \sim^{iid} \mathrm{N}\left(\alpha^{(\phi)}, \sigma_\phi^2\right) \quad \text{(eq. 2.16)}$$

It is possible to assume correlations between the three intercepts, by specifying ri(corr = "all") in the component formulae. Accounting for the district-specific population (random effect) in the spatiotemporal component can further improve the HHH model's ability to describe the role of geographical connectivity on disease persistence alongside the

inclusion of a power-law extension, as an additional parameter to describe the strength of spatial dependence between districts (Leroux et al., 2000). Since travel would be expected to be concentrated towards districts with significant economic activity and greater populations, such as large regional conurbations, it would also be expected that a higher number of cases would be imported from surrounding areas (Bartlett, 1957).

2.10.5 Likelihood Inference

For HHH models, likelihood inference is performed using generic numerical optimisation routines (Paul and Held, 2011). ML estimates of parameters and standard errors are obtained by numerically maximising the respective Poisson log-likelihood or negative binomial log-likelihood of the model. For data with overdispersion, the Poisson log-likelihood is replaced by the negative binomial log-likelihood (Paul et al., 2008). Even with a large number of parameters, convergence is usually achieved quickly using optimisation routines. Numerical approximations of the score function and the generation of a fisher information matrix are used to fit the model (Held et al., 2005, 2006).

2.10.6 Model Assessment

Goodness-of-fit: AIC

Classical model choice based on information criteria such as Akaike's Information Criterion (AIC) which correspond well to fixed effects likelihoods cannot be used straightforwardly in the presence of random effects and is often problematic. The inclusion of random effects in the HHH model results in a more complex inference process, requiring a penalised likelihood approach for obtaining parameter estimates (Breslow and Clayton, 1993; Kneib and Fahrmeir, 2007). This approach treats variance components as fixed when estimating both fixed and random effects. Variance components are estimated by maximising the

marginal log-likelihood until convergence is reached, after integration of the fixed and random effects (Paul and Held, 2011). When a HHH model is performed with the inclusion of random effects, the following alternating algorithm for the estimation of all parameters is utilised: regression coefficients are updated according to the given current variance parameters, followed by the variance components being updated according to the given current regression coefficients via Newton steps. These steps are iterated until parameter estimates remain constant and no longer alter, reaching convergence.

Likelihood inference for the regression parameters $\beta, b,$ given known variance components, is based on penalized log-likelihood:

$$l_{\text{pen}}(\beta, b;\, \Sigma) = l(\beta, b) + \log p(b|\Sigma)\,, \qquad \text{(eq. 2.17)}$$

The inclusion of random effects nullifies simple AIC based model comparisons; AIC values are no longer obtained in the model summary. To make a quantitative assessment of the predictive performance of the HHH models and inclusion of various covariate extensions, the models must be compared using successive one-step-ahead predictions assessed by strict proper scoring rules (Czado et al., 2009).

Predictive Performance: Proper scoring rules

Scoring rules have previously been recommended for evaluating the probabilistic predictions and predictive performance of models which utilise count data (Gneiting and Rafferty, 2007; Czado et al., 2009). For instance, Meyer and Held (2014), utilising the scoring rules approach, found that power-law transmission weights more appropriately reflect the spread of influenza than previously used first-order weights, which originally only enabled epidemic transmission of the disease to directly adjacent, neighbouring districts within a one-week period.

Scoring rules measure the predictive quality by assigning a numerical score, *S*(*P, y*), based on the probability distribution *P* from a fitted model and the later observed true value y (Paul and Held, 2011). Lower scores correspond to better predictions (Meyer et al., 2017). For a scoring rule to be considered proper, the expected value of the score is minimised under the predictive distribution *P*, with the observed value *y* is realised from *P* (Wei and Held, 2014). A scoring rule is classed as strictly proper if the minimum realised is unique. Strict propriety is key to ensuring that a scoring rule simultaneously addresses sharpness, the concentration of the predictive distribution, and calibration, the statistical consistency between the predicted probabilistic and observed distributions (Paul and Held, 2011). Unless there is a clearly defined underlying decision problem that requires a specific scoring rule, there is no automatic choice of a proper scoring rule to be applied in any given context. Applying a variety of scoring rules to assess the predictive performance of the same models is often appropriate since they have different strengths and emphasis, and probabilistic predictions have multiple, simultaneous uses (Harville, 1997).

Scoring rules are utilised to assess the predictive performance of the HHH models with extensions applied for the childhood infections under analysis in Lancashire and South Wales across the nine time–windows. Strictly proper scoring rules such as the logarithmic score (logS) or the ranked probability score (RPS) consider the whole predictive distribution to assess calibration and sharpness simultaneously. The most common strictly proper scoring rule used to assess model performance is the logarithmic score (Good, 1952):

$$\mathrm{logS}(P, y) = -\log\big(\mathrm{P}(Y = y)\big), \qquad \text{(eq. 2.18)}$$

where $P(Y = k)$ denotes the probability mass function, and y denotes the count that materializes. The logS is the negative log-likelihood evaluated at the actual observation, providing no credit for assigning high probabilities to values close, but not identical to, count y. Note that the log-score for a HHH model's prediction in district i in week t equals the

associated negative log-likelihood contribution. The logS is also highly sensitive to extreme cases and outliers as it strongly penalises low probability events (Paul and Held, 2011).

Less sensitive to extreme events (Czado et al., 2009), another strictly proper scoring rule utilised is the ranked probability score (Epstein, 1969), originally introduced for ranked categorical data. It can be simply adapted for count data with the following formulation:

$$\mathrm{RPS}(P, y) = \sum_{k=0}^{\infty} (\mathrm{P}(Y \leq k) - 1(y \leq k))^2, \qquad \text{(eq. 2.19)}$$

where $P(Y \leq k)$ denotes the cumulative distribution function. The RPS is the only proper scoring rule that only depends on the predictive probability distribution P through the $P(Y = k)$ at the observed count (Good, 1952). This scoring rule also blows up score differentials between competing models, adding extra weight to situations with unusually high observed or predicted counts (Czado et al., 2009).

In the HHH framework, predictive model assessment is undertaken by computing predictions that correspond to the fitted values of a test period. Mean scores are used to rank and compare different models. This is mostly done informally by ordering the obtained mean scores, essentially conducting a goodness-of-fit assessment. The assessment evaluates true one-week-ahead predictions, refitting the HHH model up to week t to get predictions for week $t + 1$, across the time-series. Statistical significance of the differences in mean scores can be investigated by a paired *t*-test (Meyer et al., 2017). An evaluation of the strength of calibration of the predictions of the HHH models can be performed by using unconditional and regression tests based on proper scoring rules (RPS, LogS and DSS) (Wei and Held, 2014). The application of these methods to weekly data on meningococcal disease incidence in Germany, 2001-2006, demonstrated the tests to be powerful tools for detecting miscalibrated predictions, their implementation is straightforward, and they capably facilitate model comparison and selection (Gneiting et al., 2007). Specifically, for

count data, a regression approach based on proper scoring rules work most effectively for detecting miscalibration for over-dispersed predictions (Wei and Held, 2014).

2.11 Chapter Summary

This chapter has provided a detailed discussion of the historical, archival data sources used and data collected to facilitate the research project, considering issues of data quality as well as the logistical challenges faced when handling and working with a significant quantity of quantitative data. The Registrar–General's *Weekly Return* provides an extraordinarily rich time-series of geographically aggregated infectious disease data at fine spatial and temporal scales, an ideal source of data for studying temporal changes in disease persistence in different regional metapopulations of England and Wales. The *Weekly Returns* are complemented by similarly high-quality, annual, Registrar–General's *Statistical Review* reports, which provide vital information on demographic stochasticity and susceptible dynamics, enabling a detailed analysis of observed epidemiological patterns within their ecological context. These data sources enabled the construction of detailed datasets with which one could perform a wide-ranging analysis of disease persistence for multiple childhood infections over space and time.

The chapter details the six major elements of the data analysis featured in this thesis:

1) A time-series analysis of regional and district-level disease incidence;
2) a district-level correlation analysis of spatial synchrony of regional disease activity;
3) endemic threshold estimation using a "moving window" empirical regression approach;
4) a hotspot analysis centred around an analysis of OLS regression residuals and performance of tests for spatial autocorrelation to compare and contrast the frequency and spatial distributions of regional pertussis hotspots before and after vaccination;

5) a survival analysis of rates of disease reintroduction before and after vaccination among pertussis hotspots and non-hotspots, involving the use of Cox proportional hazards models and application of Kaplan-Meier survival curves;
6) endemic–epidemic “HHH” modelling: multivariate negative binomial time-series models were fitted to each time window to analyse disease counts aggregated by time and period and investigate the role of spatial spread on the persistence and re-introduction of disease.

The chapter also provides rationales for the quantitative methods selected to perform analyses while also acknowledging limitations, modelling issues and potential remedies to address these issues if such a study was to be repeated in future. Together, the array of quantitative methods described and utilised in this body of work are designed to generate findings that will address the research questions stated in the introductory chapter in Section 1.3. All quantitative analyses were performed using R, a programming language and free software environment for statistical computing. Many of the statistical outputs from the aforementioned analyses listed were visualised using QGIS. The results of the statistical analyses performed are presented in Chapters 4, 5, 6 and 7. The next chapter will now provide a detailed overview of the study period and geographical settings subject to analysis in this study.

Chapter 3: Study Period & Settings

3 Introduction

This chapter composed of three parts. Part one describes the period in which the study is set, with particular focus on a key demographic event during the period that would significantly influence the disease dynamics of the three childhood infections under analysis: the 'Baby Boom' era (1945–1964). Part two describes key administrative geographies of relevance to this study, provides a detailed account of the regional and sub-regional geographies of Lancashire and South Wales, as well as the networks of geographical connectivity in each region. Part three provides summaries of wider demographic changes during the study period in the Lancashire and South Wales regions.

Part One: The Period

3.1 The Study Period

The study period extends over a thirty-year interval from 6th January 1940 to the week ended 2nd January 1970. The year 1940 represents the first complete year of statuary 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, measles and scarlet fever notifications suitable for analysing temporal changes in endemic threshold size. The geographical detail of the reports is also exemplary, presenting an excellent opportunity to analyse regional disease persistence, with the consistent quality of case reports and comparability across districts throughout the study period. This was lost with the boundary changes of the 1972 Local Government Act, which abolished hundreds of districts, significantly reducing the resolution of spatiotemporal data collected from 1974 onwards. Lancashire and South Wales reported a high number of cases for all three childhood infections across the thirty-year period, allowing changes in endemic threshold

size to be detected across time. The length of the study period facilitates the analysis of spatiotemporal changes in endemic threshold populations for measles, scarlet fever and pertussis across nine-time–windows with an equal number of reporting weeks, before the significant boundary changes and abolishment of local administrative units on 1st April 1974. The period also represents a unique point in the epidemiological history of mid-twentieth century England and Wales, with the demographic upheaval of the post-war baby boom and the introduction of routine mass vaccination for childhood diseases.

3.2 The 'Baby Boom' Period (1945–1964)

The study period chosen incorporates a demographic event with profound consequences for the endemicity and epidemicity of childhood infectious diseases: the post-war 'baby boom'. Loosely defined, the 'baby boom' era in England and Wales broadly spanned the period 1945 to 1964 (Hobcraft, 1996), characterised by two major peaks in birth rate, between 1945–47 and 1962–1964 (Van Bavel and Reher, 2013). Baby boom events cause abrupt temporal changes in susceptible recruitment (Grenfell et al., 2001), thus representing extrinsic variations in the recruitment rate of susceptibles that would not normally be expected (Earn et al., 2000). During baby booms, the pool of susceptibles swells, supplying substantial fuel for epidemic outbreaks. The growth of susceptible populations also drives disease spread beyond the confines of local communities, compensating for the geographical unevenness in the distribution of individuals and households characteristic of regional populations (Grenfell et al., 2002). This unevenness can hinder the transmission of infection with increasing distance away from centres of epidemic activity with the absence of a sufficient number of susceptibles to maintain chains of infection. More subtly, spatial heterogeneities in population density or demography (Blasius et al., 1999; Grenfell et al., 2001), and temporal changes in parameters due to the introduction of interventions such as vaccination, which discount changes in birth rates depending on the level of vaccine uptake, can significantly influence the spatial dynamics of waves of infection (Earn et al., 2000;

Bjørnstad et al., 2002). These local spatial heterogeneities are exacerbated by the degree of isolation of individual towns and cities and their connectivity within the broader regional metapopulation (Bolker and Grenfell, 1996; Broutin et al., 2005). Birth rate and population size represent readily available measures of dynamically relevant local characteristics that can capture the non-linear dynamics of childhood epidemics as a function of local population size (Bartlett, 1956). Another factor in the non-linear dynamics of childhood disease activity is the impact of environmental forcing, which mainly comprises seasonality in transmission as a consequence of term-time schooling (Finkenstädt and Grenfell, 2000; Metcalf et al., 2009).

Compared to other western nations, England and Wales experienced a relatively mild baby boom with around 70 extra births per 1,000 (Van Bavel and Reher, 2013), and recorded twin peaks in the birth rate. In demography circles, baby booms can be captured by a single measure; the total fertility rate (TFR). The total fertility rate is the average number of children born to a woman if they survived to the end of their reproductive life if she was to experience contemporaneous age-specific fertility rates throughout that period (Alkema et al., 2011). Total fertility is calculated by the sum of age-specific fertility rates weighted by the number of years in each age group, divided by 1,000 (Hobcraft, 1996).

Hobcraft (1996) provides a detailed account of the post-war baby boom in England and Wales, tracking the annual TFR from 1938 to 1972. This is presented graphically in Figure 3.1 on the following page. The lowest TFR during World War II was 1.73 in 1941, slightly below the average pre-war levels. The fertility rate steadily grew, falling briefly to 2.04 in 1945. In the immediate post-war years, the TFR dramatically increased, rising to 2.68 by 1947 (see Fig 3.1) the highest fertility rate in England and Wales since the end of World War I (Hobcraft, 1996). This represents the first of two peaks during the baby boom era, reflecting demand from delayed childbearing of older couples. This initial boom largely subsided by 1950, after several years of post-war austerity. The TFR fell to just under 2.20

compared with the pre-war level of about 1.80 (see Fig 3.1). However, the immediate post-war period saw the establishment of the welfare state, with the creation of the National Health Service, maternity care and family allowances substantially reducing the costs of bearing and raising children (Hobcraft, 1996). These improvements in the conditions for parenthood laid the foundations for a recovery in the fertility rate in the 1950s, culminating in a second peak in 1964 (Van Bavel and Reher, 2013), with the TFR reaching a post-war high of 2.93 (see Fig 3.1). However, 1964 marked the end of the baby boom, with fertility and birth rates declining precipitously in the following years. The 1960s was as much a decade of contraceptive revolution as it was of the sexual revolution, marking a significant shift in the proximate determinants of fertility.

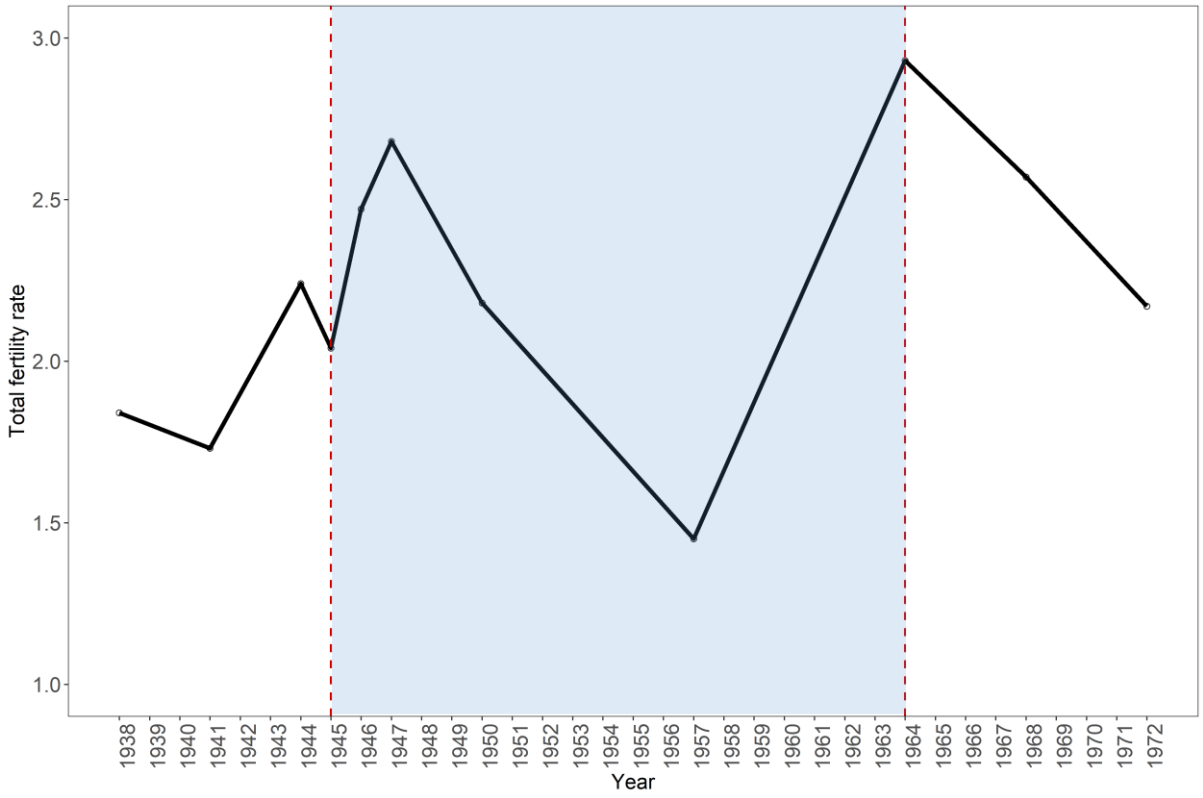

Figure 3.1 *Total fertility rate in England and Wales, 1938–1972. The baby boom era between 1945 and 1964 is shaded in blue (Figure adapted from Table 1, Hobcraft 1996: 494). Notable spurts in the total fertility rate are visible in 1947 and 1964; these peaks are considered to bookmark the start and end of the postwar baby boom experienced in the United Kingdom. The total fertility rate reached a nadir in 1957.*

Part Two: Regional Geography

3.3 Administrative Geography of Reporting Units

The Registrar-General's *Weekly Return* collates disease counts for statutorily notifiable infections for administrative districts in England and Wales. During the study period, weekly measles, pertussis, and scarlet fever notification data were reported for 945 cities and towns and 457 rural districts nationwide. Local government districts (LGDs) were divided into four sub-categories: county boroughs, municipal boroughs, urban districts, and rural districts.

3.3.1 County Boroughs

County boroughs of England and Wales were established in 1889 by the 1888 Local Government Act, with settlements with populations greater than 50,000 given powers to manage their administrative affairs. County boroughs were thus treated as administrative counties in their own right. The threshold was raised to 75,000 in 1926 and 100,000 in 1958. After WWII, the creation of new county boroughs in England and Wales ended; no county boroughs were created in the postwar period before their eventual abolition.

3.3.2 Municipal Boroughs

The earliest of the reporting units featured in the Weekly Returns, municipal boroughs in England and Wales were initially created following the passage of the 1835 Municipal Corporations Act. Unlike county boroughs, municipal boroughs had more limited powers of self-government, as well populations below 50,000. Municipal boroughs were abolished in April 1974 and succeeded by districts in Wales and by metropolitan or non-metropolitan districts in England.

3.3.3 Urban & Rural Districts

Alongside municipal boroughs, administrative counties were subdivided by Urban and Rural Districts for most of the twentieth century until the reforms introduced by the 1972 Local Government Act. Following the passage of the 1875 Public Health Act, urban and rural sanitary districts were established in England and Wales and divided into two categories based on existing structures. In 1894, these local government bodies were superseded by urban and rural districts, introduced by the Local Government Act passed that year.

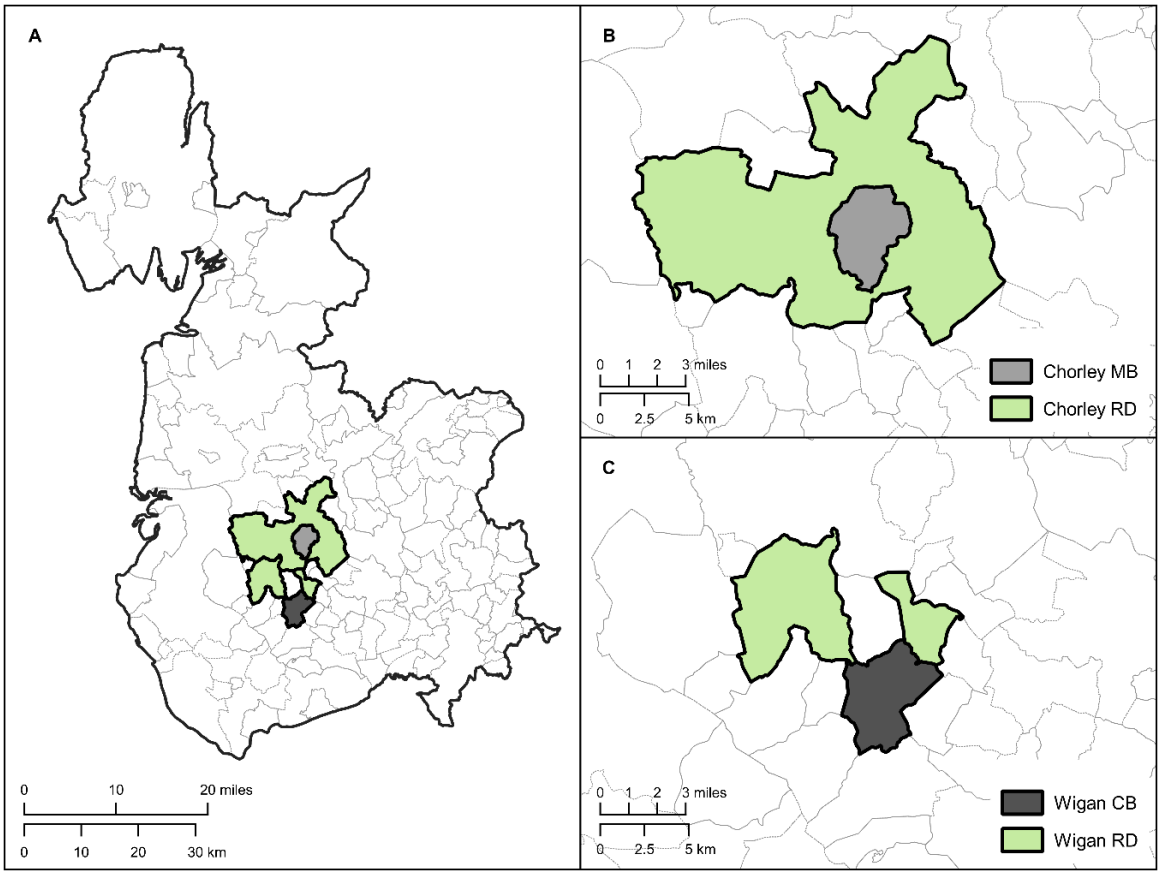

Figure 3.2 *The divergent administrative geographies of Rural Districts, using the examples of Chorley RD and Wigan RD, Lancashire.*

Urban districts generally comprise small towns, usually with less than 30,000 inhabitants, and often with rural surroundings (Hasluck, 1948). In contrast, rural districts usually contained villages and market-towns of differing sizes, which in the odd instance could be

larger than settlements forming urban districts. Some rural districts in England and Wales were redefined as urban districts or merged with existing urban districts or boroughs due to steady growth in urbanisation which accompanied industrial development over successive decades (Jackson, 1966). Rural districts typically had a rounded shape like a doughnut, surrounding either an urban district or a municipal borough. However, the shape would not always take a perfect form (see Fig. 3.2, panel A). For instance, in the context of Lancashire, Chorley Rural District in Central Lancashire surrounds the town and municipal borough of Chorley (see Fig. 3.2, panel B). However, it was not unusual for Rural Districts to be fragmented, consisting of several detached parts. A prime example of this is the Wigan Rural District (Fig. 3.2, panel C).

3.4 Geography of Lancashire

The Lancashire region consists of the administrative county of Lancashire, formed in 1889, occupying the area of the historic county and eighteen county boroughs. A breakdown of the names and locations of local government districts that compose the administrative county can be found in Appendix I (Table I.1). The historic county of Lancashire is located in North-west England, bounded by the historic counties of Cumberland and Westmoreland to the north, Yorkshire to the east, Derbyshire and Cheshire to the south (see Fig. 3.3). At its greatest expanse, historic Lancashire had a breadth of 45 miles (72.5 km), and length of 76 miles (122.3 km), and an area of 1,208,154 acres, roughly 4,890 square kilometres.

Lancashire is a region of geographical contrasts. South Lancashire is densely populated, dominated by the two urban, economic centres of Manchester CB and Liverpool CB, and their metropolitan areas (see Fig. 3.4). The region is polycentric with a complex settlement hierarchy, featuring smaller urban settlements that are partially autonomous and partially dependent on the larger metropolitan cores, at the heart of Greater Manchester and Merseyside conurbations (Fig. 3.4). The region can be loosely broken down into 12 distinct

sub-regional areas: Greater Manchester, Merseyside, Rossendale, Pennine Lancashire, West Lancashire, Central Lancashire, South Ribble, Ribble Valley, Wyre Valley, The Fylde, North Lancashire and Furness. The locations of these geographical areas are displayed in Figure 3.5.

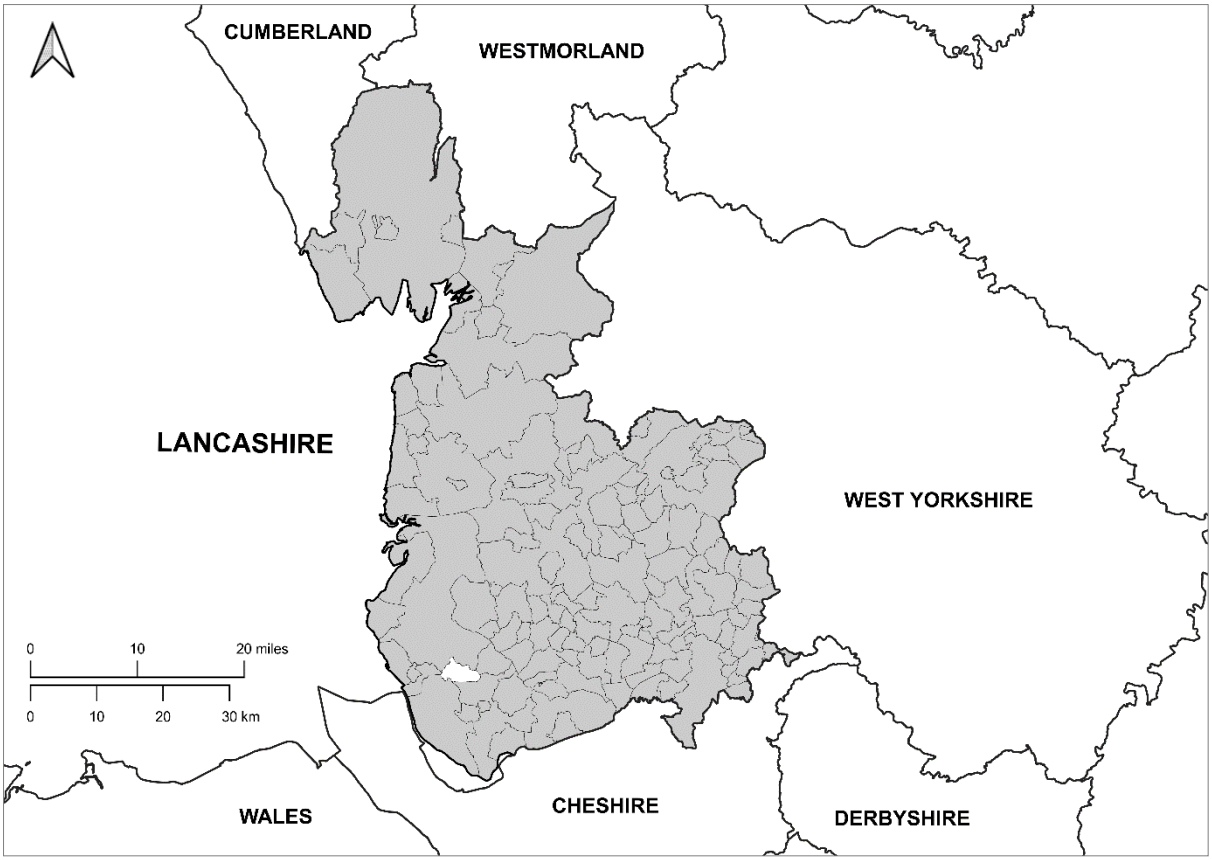

Figure 3.3 *Administrative, county, and district boundaries for Lancashire and North-West England, before the implementation of local government reforms in England and Wales on 1st April 1974. The districts which provide the geographical underpinning of the analysis presented in this thesis are shaded in grey.*

The West Lancashire coastal plain is home to an urbanised landscape, while East Lancashire represents the urban periphery, consisting of many small local population centres located in industrialised valleys, surrounded by agricultural land and bordered on its eastern edges by the Pennine hills, a continuous range of uplands separating Lancashire and the neighbouring region of Yorkshire.

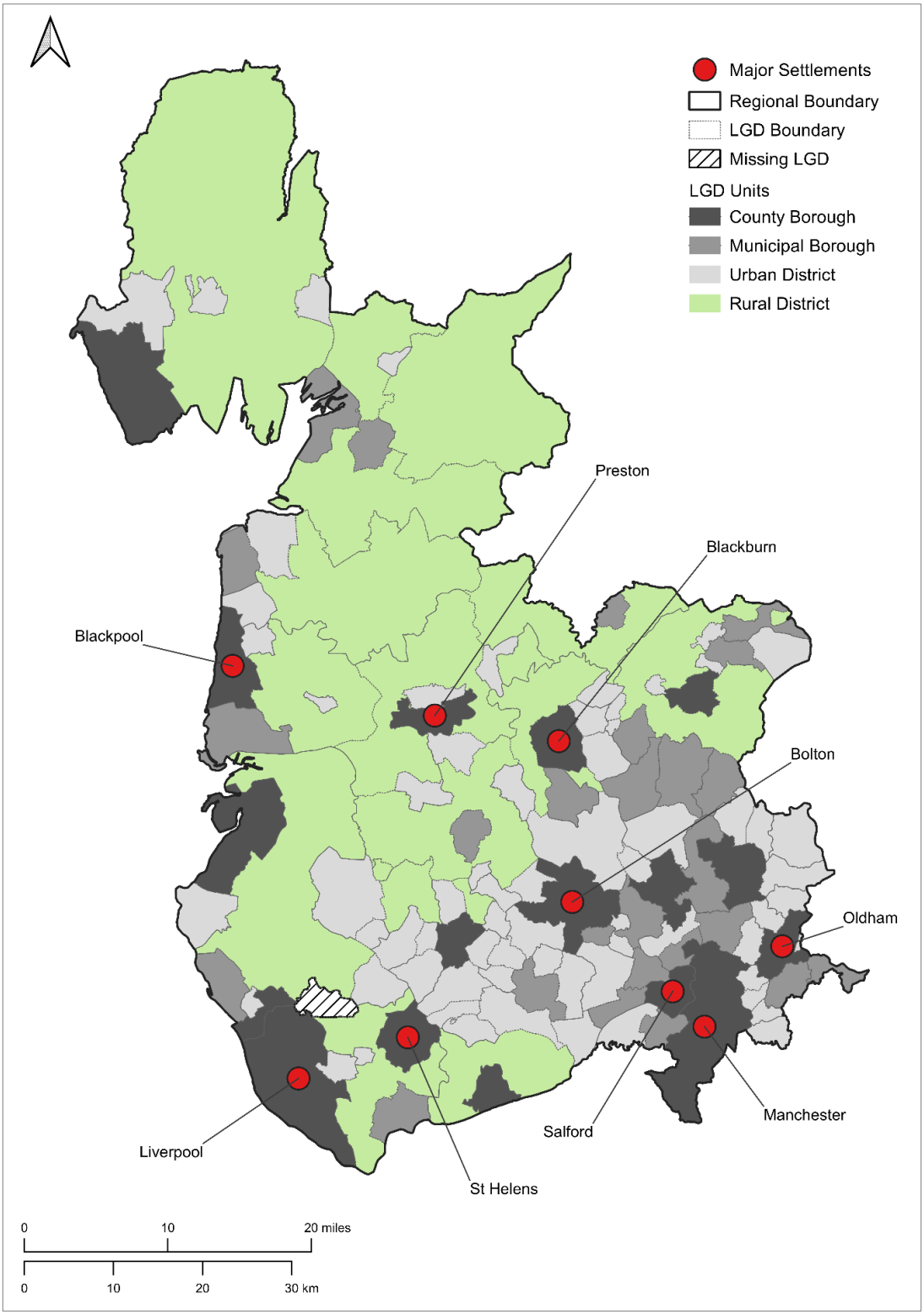

Figure 3.4 *Geographical make-up of local government districts in the Lancashire region. Red circles represent major towns and sub-regional centres with a mean population greater than 100,000 during the study period (1940–1969). The majority of the regional population is concentrated within urbanised areas located in the southern portion of the region.*

To the north lies the rural periphery, occupied mainly by agricultural land (Fig. 3.4), populated by villages, hamlets and a few long-established market towns, such as Lancaster and Ulverston. In the North-west, prominent metropolitan areas are situated on the coastal edge, such as the large resort towns of Blackpool and Morecambe (Fig. 3.4).

South of the Lancashire Pennines lies the South Lancashire Coalfield, which spans what now comprises the modern-day metropolitan counties of Greater Manchester and Merseyside (see Fig 3.5). The Southern part of Lancashire has historically been the most urbanised area of the region (Carter, 1962). South-East Lancashire is the heart of the industrial region, focused on the Manchester conurbation. Several large towns surround the city, including Salford, Bolton, and Stockport, and between them lie smaller towns, suburban to both the regional centre of Manchester and other major town centres. In 1951, the General Register Office for England and Wales began reporting on Southeast Lancashire as a homogeneous conurbation in the decennial census (Frangopulo, 1977). By the end of the decade, Manchester and its surrounding urban area represented the most complex polycentric functional urban region in the United Kingdom, outside of London (Freeman and Snodgrass, 1959). An amalgamation of seventy former LGDs, nearly 60% of all LGDs in the historic, administrative county of Lancashire, including eight county boroughs and sixteen municipal boroughs, would form the metropolitan county of Greater Manchester in 1974. In South-West Lancashire, Liverpool and its surrounding urban area, now the Merseyside region, are also polycentric (Pollard et al., 2006). Liverpool is neighboured by large towns such as Bootle, Southport and St. Helens, each of which possesses outlying suburbs composed of smaller towns and villages (Pollard et al., 2006). The North-West Green Belt is interspersed throughout Southern Lancashire, encompassing many districts throughout the Ribble Valley, West Lancashire and The Fylde coastal plain, and extending as far south as Northern Cheshire. The green belt was first drawn up in the 1950s, to prevent the many cities, towns and villages which form the Greater Manchester and Merseyside conurbations from merging (Hebber and Deas, 2000).

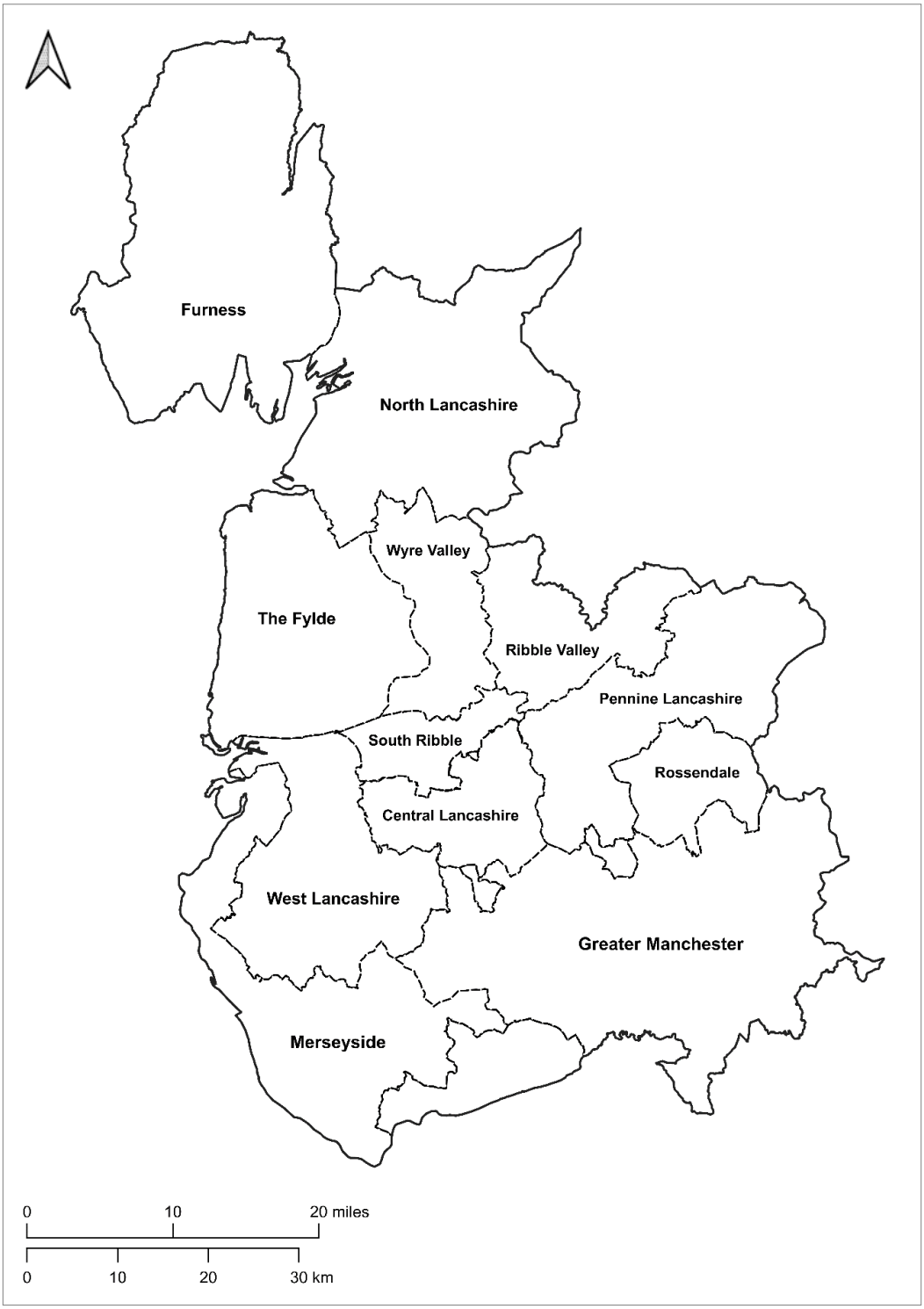

Figure 3.5 *Key geographical and metropolitan areas within the Lancashire region during the twentieth and twenty-first centuries. Greater Manchester and Merseyside became metropolitan counties in 1974. The areas of Rossendale, South Ribble, West Lancashire, Ribble Valley, The Fylde and Wyre Valley are now non-metropolitan district with borough status in contemporary Lancashire.*

3.5 Connectivity in Lancashire

In Lancashire, particularly in the eastern and southern portions of the region, commuting flows have historically been strongly shaped by local geography, which in the study area features many similarly sized, closely spaced towns and cities with strongly localised identities (Coombes, 2019). Analysis of census data on travel to work movement which became available at the end of the twentieth century provides some picture of the intra-regional movements in terms of commuting patterns in Lancashire (Baker and Hebbert, 1995). Perhaps unsurprisingly, travel to work patterns show high levels of connectivity between Liverpool and Manchester, with virtually all significant flows of commuters to the main urban centres originating from districts located in the periphery of the metropolitan hinterland, such as Rossendale and the Ribble Valley (Baker and Hebbert, 1995). Similarly, commuting patterns to and from the metropolitan areas of Merseyside and Greater Manchester extend across the region via a trans-Pennine corridor that travels westwards across the Pennines, linking with urban centres in Yorkshire such as Leeds and Bradford (Ravetz and Warhurst, 2013). This corridor passes through large towns in the east of the Lancashire region, settlements such as Blackburn, Burnley and Rochdale. However, these inter-regional commuter linkages are not as strong as those between the Liverpool and Manchester city regions.

In terms of geographical connectivity, the study period was a time of monumental change for the Lancashire region, which experienced a revolution in regional transport infrastructure. Before the post-war period, a spider's web of rail lines covered the region radiating from urban centres (see Fig. 3.6), with suburban passenger traffic concentrated in the southern portion of the region (Patmore, 1964). Lancashire's railway network shrank significantly between 1940 and 1970, with the loss of 239 stations closing across the region, over half of which were located in Manchester, Liverpool and surrounding towns.

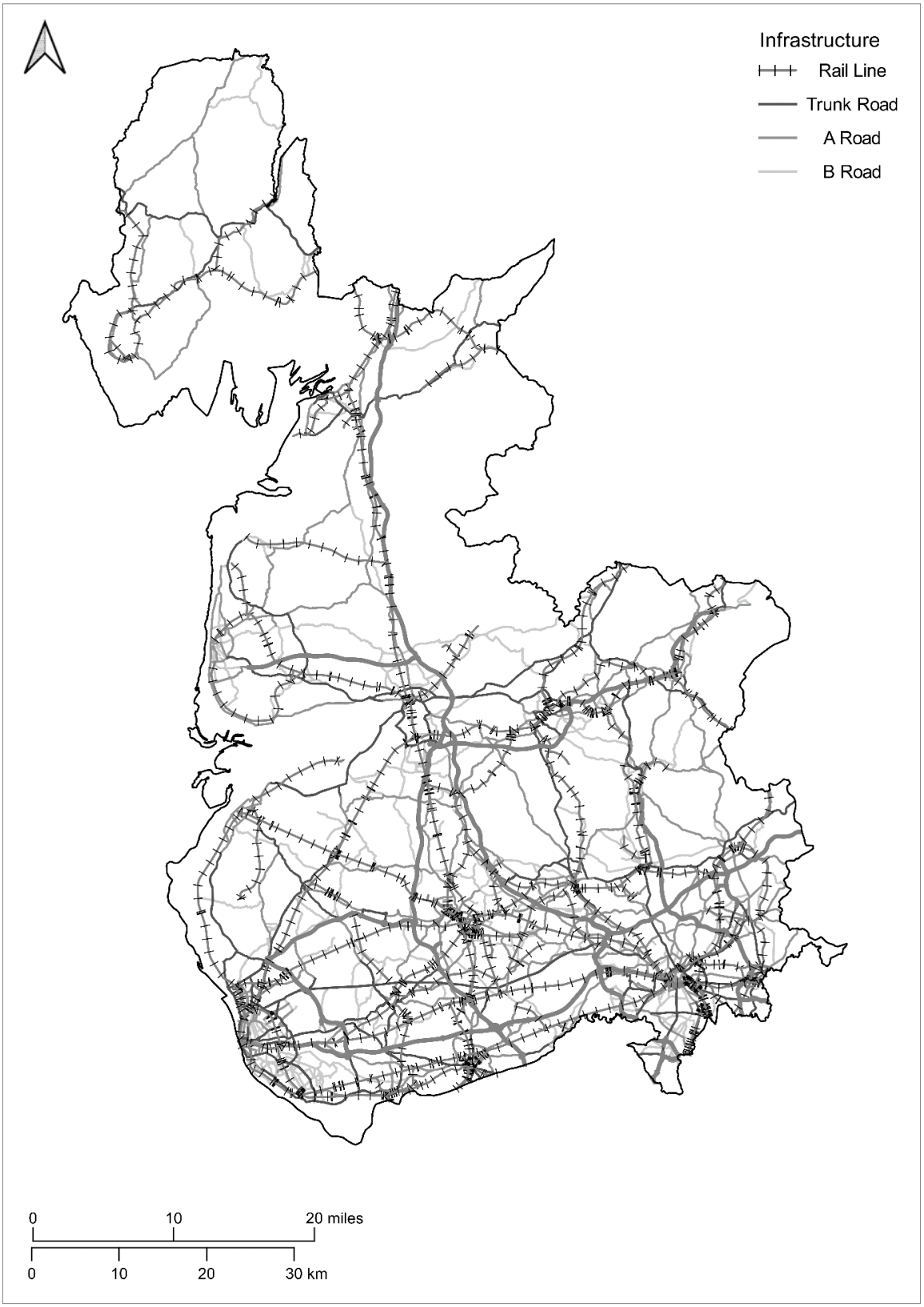

Figure 3.6 *Major Road and rail connectivity networks within the Lancashire region, 1940–1969. Rail and road links are concentrated in the southern portion of the Lancashire region, specifically around the Manchester and Liverpool conurbations.*

Many rural areas and small urban settlements swallowed up by suburbanisation and urban overspill from larger towns and cities were left without ready access to the rail network, inhibiting long-distance population flows. Lancashire was left with a skeletal system, primarily serving the two conurbations of Merseyside and Greater Manchester, majors towns and commuters.

Car ownership continued to grow at a great pace throughout the first half of the twentieth century, leading to the eventual post-war construction of motorways and a dense road network that would supplant the rail network. In 1946, a new national road construction programme included a North-South Lancashire route and an East-West, Yorkshire - Manchester- Liverpool route (Mackie et al., 1995). These would eventually become the M6 and M65 motorways respectively, which were opened in stages through Lancashire until completion in 1965 (Mackie et al., 1995). The M6 acts as the main road artery through Lancashire, connecting Manchester and Liverpool and its surrounding urban areas with the north of the region past Lancaster, via Preston, which is a key transport hub in the county for both road and rail, acting as an interface for Lancashire's numerous geographical areas.

Despite the significant changes in the provision of rail and road infrastructure during the mid-twentieth century and period under study, previous empirical study has shown that changes in personal mobility in Lancashire have changed very little over the last fifty years. For instance, Pooley et al. (2010) analysed the everyday travel of children aged 10 and 11 from the 1940s to 2010. They found that in 1940s Manchester, children aged 10 and 11 travelled around 3,500 km annually, with the mean distance travelled only 1.3 km. By the 2000s, children of the same age in Manchester travelled further in total, around 1000km more than in the 1940s, but the mean trip distance continued to be very low, at 1.5 km. According to Pooley et al. (2010), data for Lancaster reveals very similar trends, although there is evidence of a greater increase in everyday travel for those residing in more remote rural areas and settlements with the concentration of service and economic activity in towns.

3.6 Geography of South Wales

South Wales is generally considered to be formed of four historic counties: Glamorgan, Monmouthshire, Carmarthenshire and Pembrokeshire (see Fig. 3.7; Jenkins, 2014). Fundamental to the economic development and demography of South Wales over the last three centuries has been the presence of the largest continuous coalfield in the United Kingdom within the region (Davies et al., 2008). The South Wales coalfield, now largely exhausted, covers an area of roughly 1,000 square miles, stretching across Glamorgan, Monmouthshire, Carmarthenshire, and a small portion of Pembrokeshire. Numerous, narrow, peri-urban valleys criss-cross the coalfield, mostly running North-South parallel to each other (Jenkins, 2014). This area is widely known as 'The Valleys' (Dicks, 1999).

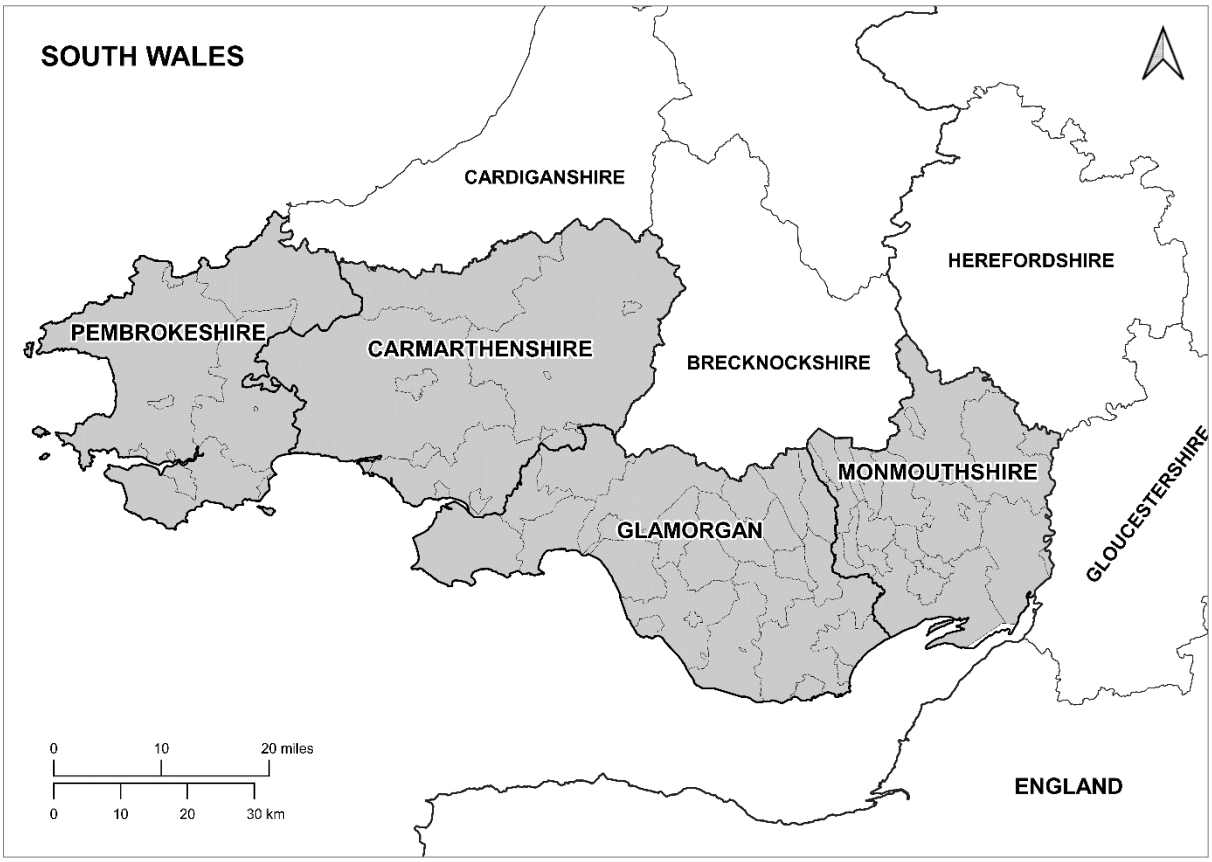

Figure 3.7 *County and administrative boundaries of the South Wales region before the enactment of The Local Government Act 1972. The local government districts which provide the geographical underpinning of the analysis presented in this thesis are shaded in grey.*

Glamorgan

The historic administrative county of Glamorgan is bordered to the north by Brecknockshire, east by Monmouthshire, and to the west by Carmarthenshire (see Fig. 3.7). The county is home to the two largest settlements in South Wales: Cardiff, the county town and capital of Wales from 1955 onwards, and Swansea (Davies et al., 2008). Glamorgan can be separated into three distinct, contrasting geographical areas. Firstly, to the south lies the Vale of Glamorgan, a low-lying, primarily agricultural, landscape formed on a gently undulating limestone plateau (Conduit, 1997). The Vale is predominantly rural, characterised by numerous villages, hamlets and extensive farmland (Newman et al., 1995). Major urban settlements found in the Vale of Glamorgan include Cardiff (See Fig. 3.8), the largest city in Wales that is situated at the mouth of the River Taff, and the coastal towns of Barry, Penarth and Bridgend. The second sub-region covers the northern portion of the county, an area dissected by deep narrow valleys that run approximately parallel to each other in a north-south direction. This area has historically been home to multiple urbanised mining districts, possessing either urban district or municipal borough status (Fig. 3.8). Finally, to the west of Glamorgan is the Gower Peninsula and Swansea Bay, a coastal sub-region. Swansea is the largest settlement, functioning as the dominant urban centre in Southwest Wales and serving as a gateway for the more distant rural areas of Carmarthenshire and Pembrokeshire to the west (Edwards, 1980).

Carmarthenshire

Carmarthenshire is bordered by historic counties of Cardiganshire to the north, Glamorgan to the east, and Pembrokeshire to the west (Fig. 3.7). The market town of Carmarthen, 12 miles northwest of Llanelli, is the county town and regional administrative centre. The most urbanised area of the county is located in the southeast portion concentrated around Llanelli, the largest settlement in Carmarthenshire and a former manufacturing centre (Phillips, 1994).

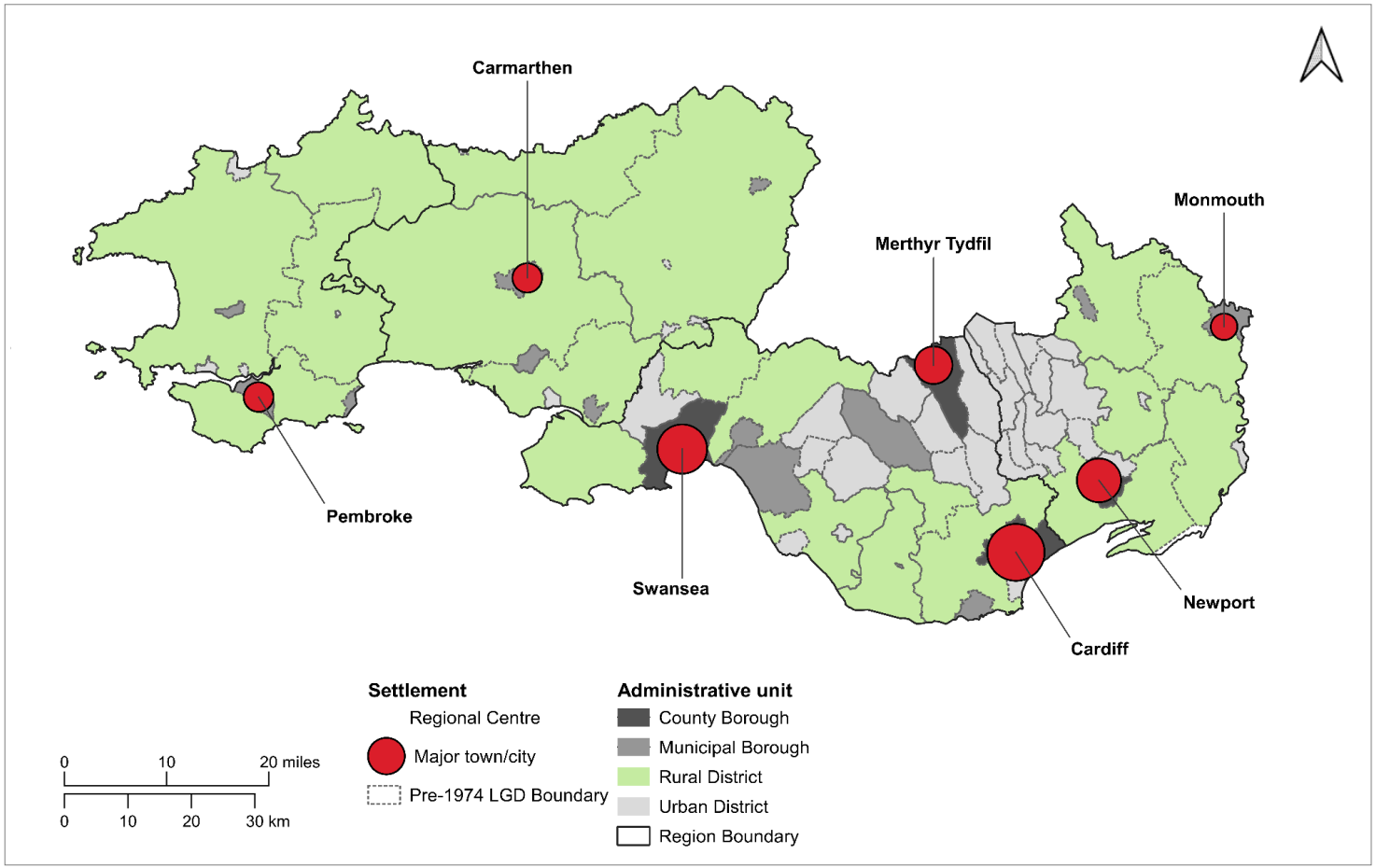

Figure 3.8 *Geographical make-up of local government districts in the South Wales region. Red circles represent major settlements and county towns of varying population sizes, 1940–1969. Urbanised districts are concentrated in the eastern portion of the region, in the county of Glamorgan and the western portion of Monmouthshire. Significantly less populated rural districts are found in Carmarthenshire, Pembrokeshire, East Monmouthshire and the Vale of Glamorgan, the coastal plain which spans the distance between Cardiff and Swansea Bay.*

This urban and industrial growth in the southeast of the county was nurtured by the area's proximity to Swansea and surrounding towns which form the Swansea Urban Area, located next to the Carmarthenshire-Glamorgan border. The county is primarily rural in nature and agricultural in function and flanked by mountainous areas to the North and East (Lloyd et al., 2006).

Monmouthshire

Monmouthshire is one of thirteen historic counties of Wales. The River Wye serves as the county's eastern boundary separating it from England. The eastern portion of the county largely consists of agricultural lowland, whilst the western area of the county, with its rich iron and coal deposits (Jones, 1969), forms part of the South Wales Valleys area. Newport, located to the south of the county on the Severn Estuary, was the largest population centre in Monmouthshire, joining Cardiff and Swansea with county borough status in 1891 (Youngs, 1991). Other settlements of importance include the county town of Monmouth, Abergavenny, and the border town of Chepstow. Monmouthshire's status as a Welsh county was ambiguous between the sixteenth and mid-twentieth centuries, with many unofficially considering it to be part of England during this period (Williams, 2011). In the *Weekly Returns*, Monmouthshire was classed as an administrative county of Wales. The county's legal inclusion in Wales was clarified by the 1972 Local Government Act, which simultaneously abolished the historic county as an administrative area.

Pembrokeshire

Pembrokeshire is located in the far south-west of Wales. It is bordered by Carmarthenshire to the east and Cardiganshire to the northeast (Fig. 3.7). The Celtic Sea lies adjacent to Pembrokeshire, with numerous ferry connections between the county and Ireland (Davies et al., 2008), enabling a consistent external population flow into the South Wales region. Pembrokeshire is composed of small, primarily coastal, market towns with populations below 20,000, separated by extensive sparsely populated rural areas (Fig. 3.8), featuring a

significant number of widely dispersed villages and hamlets (Jenkins, 2014). A significant proportion of the county population is concentrated on an urban agglomeration surrounding the Milford Haven waterway. Here, a succession of distinct yet inter-dependent settlements, such as Pembroke, Milford Haven, and Neyland, form the third most populous urban area in South Wales. The administrative centre of Pembrokeshire and the most populous settlement in the county is the market town of Haverfordwest.

The Valleys

The topography of the South Wales Valleys has strongly influenced the form of urban development adopted following the growth of metallurgical industries in the latter half of the nineteenth and early twentieth centuries (Jenkins, 2014). With the rapid expansion of the iron and coal industries in towns at the northern heads of the South Wales Valleys, such as Rhymney, Merthyr Tydfil and Ebbw Vale, a huge influx of migrants surged into the region over successive decades during this period (Davies et al., 2008). Numerous rows of densely packed terraced housing were constructed on the lower sides of valleys to accommodate workers, sometimes extending precariously over large slopes and spurs (Bowen, 1960). Urban settlements are concentrated near the site of collieries and other centres of industry (Jones, 1969), and are linearly organised along routes of communication such as railway lines, canals and roads running parallel to the various tributaries and rivers on the valley floor (see Fig 3.9). This has been described as a 'hand and fingers' pattern of urban development (Welsh Assembly Government, 2008) but is more commonly referred to as ribbon development.

The 'hand and fingers' pattern of urbanisation is best exemplified by the development of the Rhondda. The area is home to two valleys which are the major tributaries of the River Rhondda the larger Rhondda Fawr and smaller Rhondda Fach (Lewis, 1959). In 1851, the population of the Rhondda was around 1,000 inhabitants before soaring to 153,000 by 1911, as the area grew to become the centre of the South Wales coal industry (Jenkins,

2014). However, despite the area becoming heavily populated, the Rhondda remained a collection of villages and hamlets dispersed among its many valleys (Lewis, 1959). The Rhondda is composed of 16 communities; in Wales, a community is the lowest tier of local government, equivalent to parishes in England (Davies et al., 2008). The geographical obstacles posed by the land, such as the deep entrenchment of the valleys, the narrow valley floors and numerous ridges, have significantly inhibited the physical formation of larger towns from existing communities in the Rhonda (Lewis, 1959), with settlements historically clinging on to the valley sides above the main communication route that runs along the valley floor (Jones, 2014).

Beyond the rapid growth of 'Heads of the Valleys' towns, urban development followed the expansion of the coal and steel industries within the Valleys and on the coastal plain, where Cardiff and Swansea became significant ports for exports of coal and steel (Jones, 2014). In terms of rural settlement, nucleated villages only became significant in the eastern and southern peripheries of the South Wales region, in the Vale of Glamorgan and Monmouthshire (Davies et al., 2008). During the mid-twentieth century, Carmarthenshire and Pembrokeshire were, and continue to be, largely characterised by minor urban settlement on the coast around natural harbours, whilst isolated village communities are widely dispersed further inland (Lloyd et al., 2006; Davies et al., 2008).

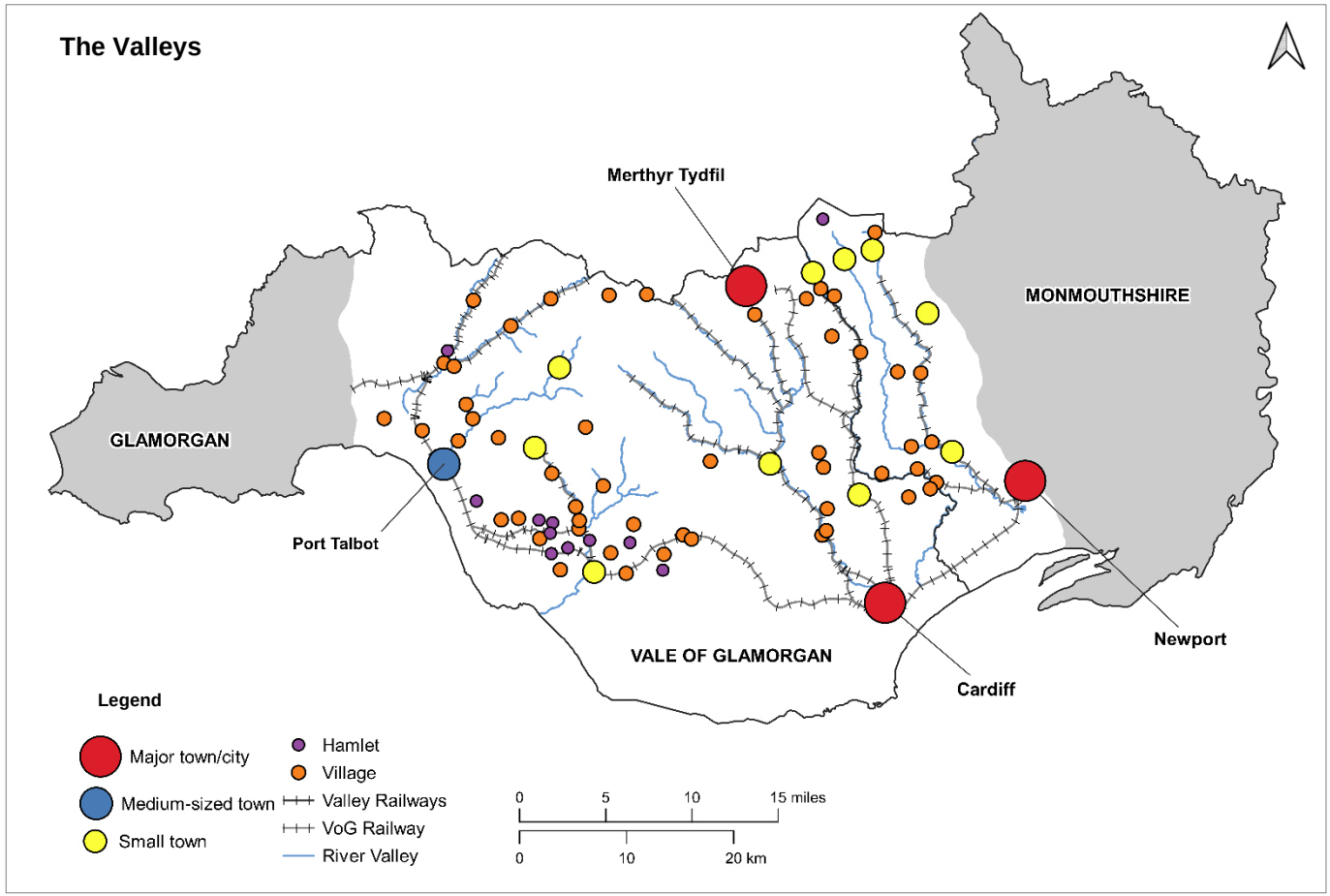

Figure 3.9 *Geography of the 'The Valleys', South Wales. Red circles denote major urban settlements (population > 100,000). Blue circles denote medium-sized towns (population ~50,000–100,000). Yellow circles denote small towns (population < 50,000). Orange and purple circles denote villages and hamlets, respectively. Hamlets and villages follow a linear settlement pattern along the valleys. Larger population centres are located either at the head or bottom of the valleys, where physical restrictions no longer constrain settlement development.*

3.7 Connectivity in South Wales

Connectivity between communities within the South Wales region is heavily influenced by the topography of the region, particularly in the counties of Glamorgan and Monmouthshire where the Valleys sub-region is located. The obstacles posed by the intervening topography of the Valleys has historically made direct movement between valleys very difficult (Welsh Office, 1967). In the Valleys, both roads and rail lines run parallel to rivers along the floor of the valleys, with settlements in a valley becoming more closely associated with each other than with settlements in neighbouring valleys, although they might be far closer geographically when measured by distance from point-to-point. For instance, in the Rhondda, the original road layout followed the valleys with limited connections between them, with separate roads running through the Rhondda Fawr and the Rhondda Fach (Lewis, 1959). There are a limited number of roads that climb steeply over the high passes between valleys (Burges and Moles, 2015), the best example being the A465 'Heads of the Valleys' trunk road, the east-west corridor which runs across the northern edge of the Valleys. Many existing roads were constructed on high passes during the 1920s, as part of a major unemployment relief programme for former miners (Coombs and Hinch, 1969), which aimed to connect the isolated hamlets and villages dotted across the valleys, transforming them from dead-end communities. The physical orientation of valleys in a north-south direction focuses movement on the better connected urban cores of Cardiff CB, Swansea CB and, to a lesser extent, Newport CB located on the valley mouths of the flatter coastal belt (Burges and Moles, 2015).

In the eastern portion of the valleys, the towns of Pontypridd, Caerphilly and Pontypool together represent a connections corridor across the Valleys sub-region, with each settlement acting as an interface at the bottom of the valley in which they lie between towns on the coastal plain and the 'Heads of the Valleys' towns of Ebbw Vale UD, Merthyr Tydfil MB and Aberdare UD. Rail links to the valleys, such as the Rhymney, Rhondda and Taff

Vale Lines (Barrie, 1982; Kidner, 1995), were vitally important to intra-regional connectivity from the late-nineteenth to the mid-twentieth century, providing regular, direct communication with Cardiff. Chepstow, which lies on the border between Monmouthshire and England, is a key settlement in terms of national connectivity, functioning as a gateway to the South Wales region Wales (Welsh Assembly Government, 2008). Both the South Wales mainline and the M4 run through Chepstow, linking South Wales to regional centres in Southwest England, such as Bristol and Bath, as well as the Greater London region.

Despite the remoteness of much of Pembrokeshire and Carmarthenshire, both counties have been accessible via long-established rail links with major settlements in South Wales and beyond (see Fig 3.10). The market town of Bridgend, which lies equidistant between Cardiff and Swansea, has historically been an important centre of regional connectivity, acting as a gateway to Southwest Wales (Welsh Assembly Government, 2008). The group of railway lines that run through Bridgend to Carmarthenshire and Pembrokeshire from Swansea are collectively known as the West Wales Lines (Page, 1988). However, with increasing dependence on the car and declining profitability of British Rail in post-war England and Wales, the post-war years saw the significant loss of many rural railway services across South Wales (Hunt, 2011).

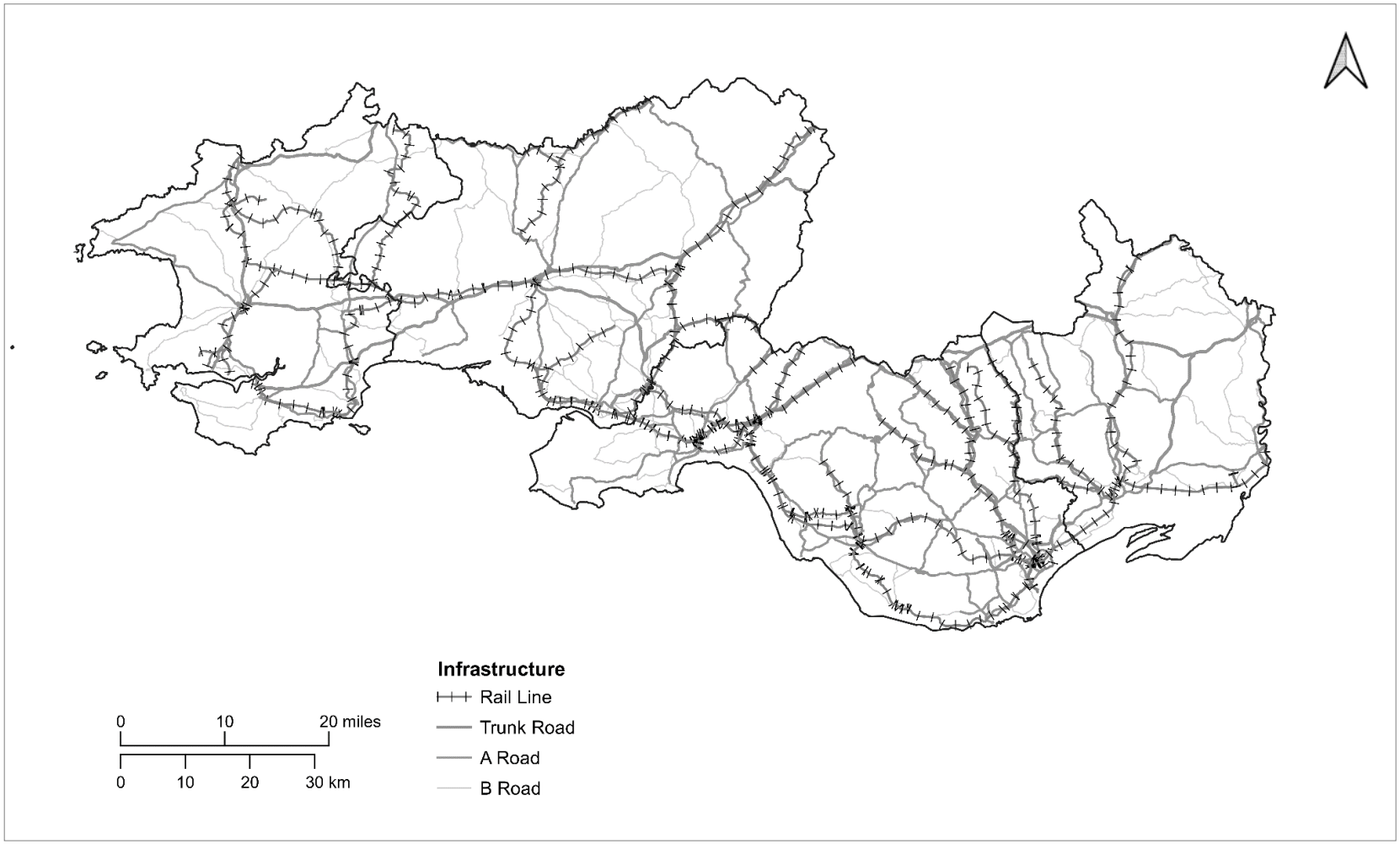

Figure 3.10 *Major road and rail connectivity networks within the South Wales region, 1940–1969. Despite a significant portion of the region being lightly populated and rural, extensive rail and road connections existed beyond the urbanised areas found in Glamorgan and West Monmouthshire. A significant quantity of South Wales's railway was lost with the fall of the Beeching Axe in the mid-1960s.*

Part Three: Regional Demography

3.8 Demography of Lancashire

The population of Lancashire experienced modest growth throughout the study period. At the beginning of the study period in 1940, the county population was 4,628,013. By the end of the study period, the total population had grown by approximately 10%, reaching a total of 5,110,723. The administrative county of Lancashire, excluding county boroughs, was also the most populous of its kind outside of London during the mid-twentieth century, with a population of 2,280,359 inhabitants in 1961. By the 1971 census, the population of Lancashire and its county boroughs represented the most populous geographical county in the United Kingdom.

3.8.1 Population Size

Regional Summary

The geographical distribution of the mean population size of the 125 local government districts in the Lancashire region for the study period is presented in Figure 3.11. The bulk of the regional population is concentrated within districts located in the Southeastern portion of Lancashire, in the areas surrounding the urban centre of Manchester CB. A detailed description of changes in district population size across the study period, broken down by sub-category of local government district, will now be provided.

County Boroughs

The absolute number of the regional population residing in county boroughs (CBs) fell between 1940–1969, from 2,901,260 to 2,725,100, a fall of approximately 6%. The number of people living in CBs grew in the first decade of the study period, peaking in 1951, with the number standing at 3,070,280. However, the population in CBs fell steadily across the

1950s reverting to 1940 levels at the beginning of the 1960s. Between 1961 and 1969, the total population residing in CBs fell from 2,931,297 to 2,725,100, a fall of ~7.0%, and it was this period of decline that explains the overall fall across the study period. The share of the regional population residing in CBs was 60.6% in 1940, declining to 53.3% by 1969.

Municipal Boroughs

The total population residing in Lancashire's municipal boroughs (MBs) increased from 840,910 in 1940 to 902,228 in 1969, a rise of +6.8%. The period with the largest proportional growth was between 1940 and 1951, coinciding with the first spurt of the baby boom period, with the population growing to 884,824 in 1951 a rise of 5.0% from 1940. The proportional share of the overall regional population residing in MBs remained constant across the study period, ranging from 17.55% in 1940 to 17.60% in 1969, peaking in 1961 at 17.76%.

Urban Districts

In absolute terms, the total population in Lancashire living in urban districts increased from 788,930 in 1940 to 1,063,150 in 1969, a growth rate of +25.8%. The proportional share of the regional population residing in UDs grew from 16.5% at the beginning of the study period to 20.81% in 1969. The 1960s was the period which experience the most significant growth, with the overall population found in UDs rising by 2.6% between 1961 and 1969, compared to an increase of +1.6% between 1951 and 1961, and 0.7% between 1941 and 1951.

Rural Districts

Rural districts (RDs) in Lancashire experienced significant growth across the study period. The absolute population residing in RDs rose from 260,470 in 1940 to 422,300 in 1969. During this period alone, the population living in RDs increased by ~ 100,000, compared to a total increase of around 60,000 in the previous two decades. As a share of the Lancashire region's overall population, the proportion living in RDs increased from 5.4% in 1940 to 8.3% in 1969, with 2% of this growth taking place between the period 1961 and 1969.

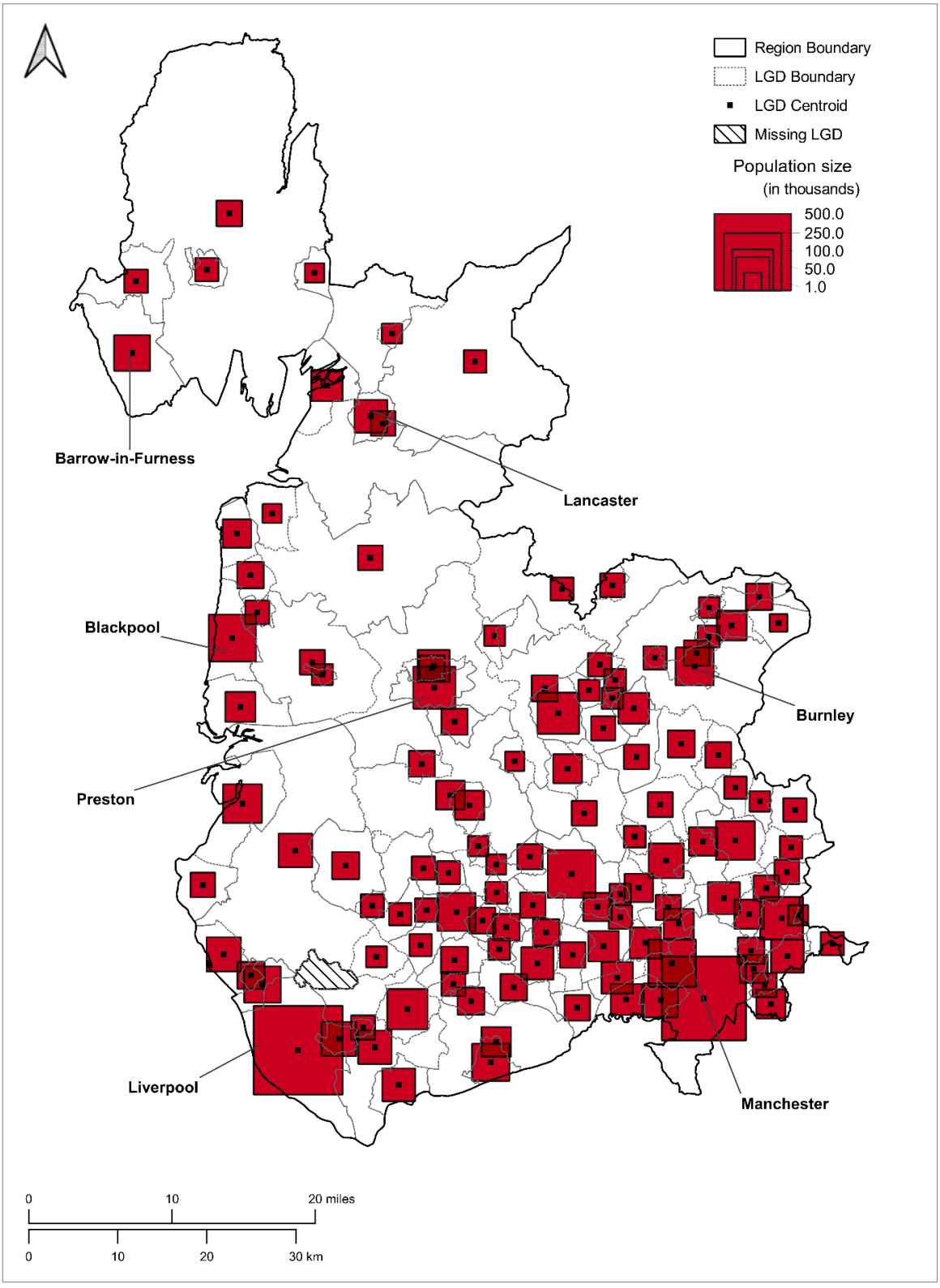

Figure 3.11 *Geographical distribution of the population of Lancashire, by local government district, 1940–1969. The mean population size for each district throughout the study period is used to produce the proportional symbols. Squares represent the mean population size of each local government district during the study period.*

3.8.2 Birth Rate

Regional Summary

Regarding the birth rate, which can be viewed as the rate of annual susceptibles input into the population, the birth per 1,000 persons in Lancashire increased from 14.54 per 1,000 in 1940 to 18.84 per 1,000 in 1969.

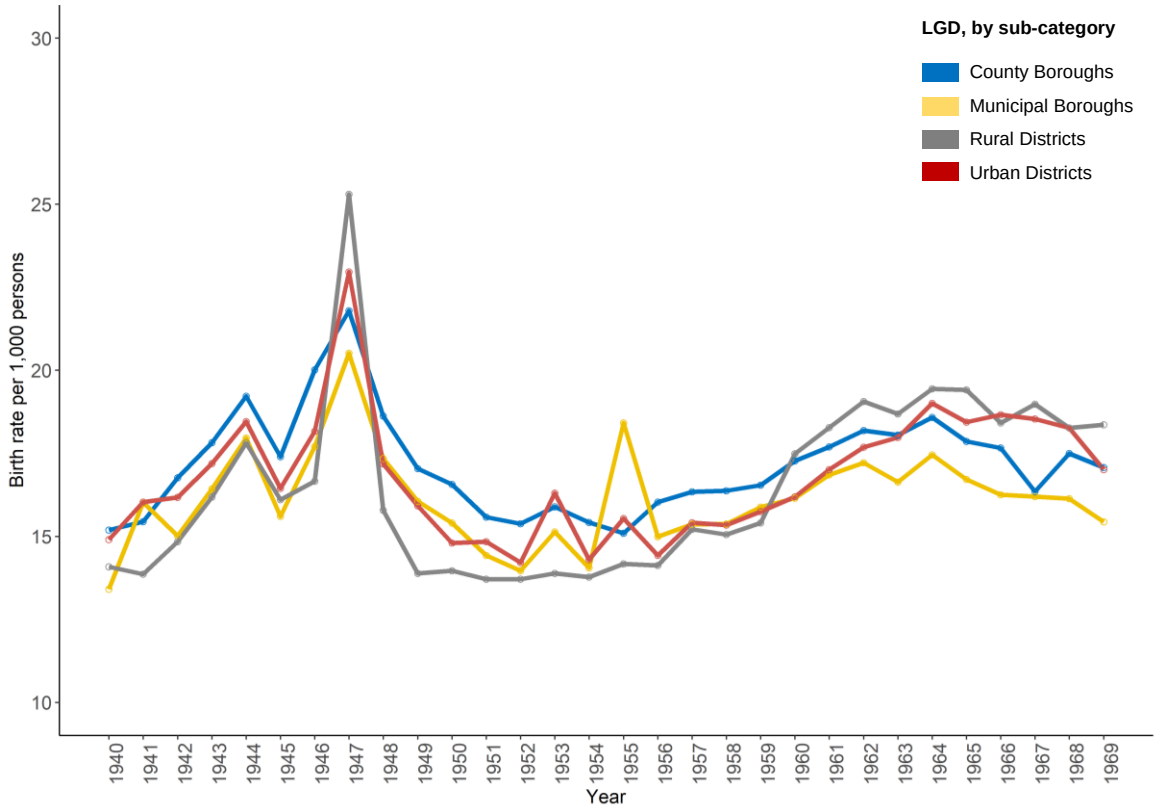

Figure 3.12 *Mean annual birth rate per 1,000 persons by sub-category of local government district in the Lancashire region, 1940–1969. The mean annual birth rates peaks dramatically for all four categories of LGD in 1947, the beginning of the baby boom era. After a fall in the early 1950s, a steady rise in the mean birth rate is visible in later years.*

County Boroughs

Between 1940 and 1961, both Liverpool CB and Manchester CB experienced a steady birth rate; Liverpool CB's birth rate grew from ~20 births per 1,000 in 1940 to 22.39 in 1961, whilst the birth rate increased from 16.69 per 1,000 in 1940 to 19.88 per 1,000 in

Manchester CB during the same period. Both urban centres experienced a peak in their birth rates in 1947, coinciding with the baby boom. Between 1961 and 1969, both cities experienced a significant fall in birth rates. In Liverpool CB, the birth rate fell by approximately one-third from between 1961–1969, falling to 16.63 births per 1,000. Similarly, Manchester CB's birth rate fell below wartime levels to 16.84 births per 1,000. The 16 other county boroughs saw an overall increase in mean birth rate, from 15.19 per 1,000 in 1940 to 17.69 per 1,000 persons in 1961. Between 1961–1969, the mean birth rate declined from 17.69 to 17.07 births per 1,000.

Municipal Boroughs

Among Lancashire's municipal boroughs, the birth rate per 1,000 grew from 13.41 in 1940 to a high for the study period of 16.85 birth per 1,000 in 1961, before declining to 15.43 births per 1,000 in 1969. The most fruitful period of growth in the mean birth rate came between 1951 and 1961, increasing by 2.4 births per 1,000 compared to an increase of 1.00 births per 1,000 in the previous period (1940-1950).

Urban Districts

The mean birth rate among urban districts grew drastically from 14.9 births per 1,000 in 1940 to 20.7 births per 1,000 by 1969. Between 1940 and 1951, the mean birth rate fell to a low of 14.84 but recovered significantly across the 1950s, reaching 17.0 births per 1,000 in 1961 and the rise in birth rates remained constant reaching a high of 20.7 births in 1969.

Rural Districts

The most dramatic growth in birth rates was found among rural districts. Between 1940–1951 the birth rate fell slightly from 14.08 births to 13.71 births per 1,000. The birth rate grew substantially increasing from 13.71 births to 18.27 births by the beginning of the 1960s. From 1961 to 1969, the birth rate stabilised with limited growth across the decade, rising marginally to 18.36 births by the end of the study period.

3.8.3 Population Density

Table 3.1 *Temporal changes in population density (per km^2) by major urban centre and district sub-category, Lancashire, 1940–1969.*

Administrative unit	Population density (per square kilometre)			
	1940	**1951**	**1961**	**1969**
All Districts	1,684.0	1,737.2	1,744.6	1,762.5
Liverpool CB	7,453.2	7,113.6	6,619.6	5,659.3
Manchester CB	6,685.1	6,372.6	6,002.1	5,142.5
CBs	4,613.5	4,329.8	4,046.4	3,741.0
MBs	2,069.6	2,110.9	2,092.9	2,004.2
UDs	1,203.0	1,304.4	1,328.7	1,500.9
RDs	120.3	148.6	164.1	227.7

Regional Summary

The mean population density in the Lancashire region rose from 1,684 persons per km^2 in 1940 to 1,763 persons km^2 in 1969. A significant proportion of this growth in density came between 1940 and 1951, primarily driven by the substantial population growth in the immediate post-war years. A general but noteworthy trend is the decline in population density throughout the study period in major urban centres and large towns and growth in density in smaller satellite settlements and rural districts, reflecting wider social-economic changes during the period associated with suburbanisation, deindustrialisation, and slum clearance. Changes in population density are detailed in Table 3.1.

3.9 Demography of South Wales

South Wales's population is unevenly distributed, with the bulk situated in the congested districts of the Valleys to the Southeast and densely populated urban centres on the Glamorgan coast. In contrast, the sparsely populated rural areas are located predominantly in the Southwest, in Carmarthenshire and Pembrokeshire. The industrial revolution was essential for creating the uneven demographic profile of South Wales during the twentieth century that continues to persist in the twenty-first century (Jenkins, 2014). The staggering scale of industrial activity born from the South Wales coalfield's exploitation attracted considerable, sustained migration from the rural counties of Wales to Glamorgan and Monmouthshire's valleys where iron and coal extraction were centred (Gareth Evans, 1989). The rapid expansion of employment and in-migration from the rural periphery to the industrial south led to two-thirds of the Welsh population residing in Glamorgan and Monmouthshire by 1911 (Jenkins, 2014). Glamorgan saw the largest growth of any county in Wales during the industrial revolution, becoming the most populous and industrialised county in the region, with Cardiff at its centre (Davies, 2008). The county population rose fifteenfold, from 70,879 in 1801 to 1,120,910 in 1911.

By the time of the Great Depression of the 1930s, immediately before the period under study, South Wales began to experience a marked decline in birth rates and a growing trend of depopulation. The industrial decline, which would go on to blight the region economically for much of the twentieth century, had begun and led to significant migration from South Wales, particularly of young adults (Davies, 2007; Jenkins, 2014). In Glamorgan, population growth in absolute terms was notably much slower than before the beginning of the twentieth century; the population rose to 1,229,728 in 1961, a limited increase of 9.7% compared to the 1911 census population of 1,120,910. Meanwhile, the predominantly agricultural Carmarthenshire experienced little growth in its population, growing from 160,406 in 1911 to 168,008 in 1961, an increase of just 4.5% (GRO, 1964a).

Pembrokeshire experienced an almost identical growth rate over the same period, from 89,956 in 1911 to 94,124 in 1961, an increase of 4.6%; 50% of the county's population resided within Pembrokeshire's three MBs (GRO, 1964b). However, during the study period, Glamorgan and Monmouth remained densely populated in comparison, with the Valleys home to around 30% of the Welsh population (David et al., 2004), and a significant concentration of the regional population in the urban cores and the metropolitan hinterland situated on the coastal plain.

3.9.1 Population Size

Regional Summary

In absolute terms, the total population size of the South Wales region increased from 1,832,790 in 1940 to 1,987,820 in 1969. This population growth was relatively gradual in contrast to Lancashire, rising by +8.5%. The geographical distribution of the mean population size of the 74 local government districts in the South Wales region for the study period is presented in Figure 3.13. The majority of the regional population is concentrated within urbanised districts located in the valleys areas of Glamorgan and the urban cores of Cardiff CB and Swansea CB. An in-depth description of changes in district population size across the study period, broken down by sub-category of local government district, will now be provided.

County Boroughs

Throughout the study period, South Wales's population increased by around 40,000–50,000 every ten years. Cardiff CB, the largest local government district in South Wales, grew from a population of 226,100 at the beginning of the study period to 285,860 in 1969, a percentage increase of 26.6%. The most considerable growth period came between 1961 and 1969, increasing by nearly 30,000, equal to the total increase in absolute terms over

the two previous decades. Cardiff CB's population had grown by 17,000 between 1940 and 1951 and 13,000 between 1950 and 1961, respectively. Between 1940 and 1969, Swansea CB's population increased by a similar margin of +20.1%, whereas Merthyr Tydfil CB's population declined by -6.7% over the same period. The combined population residing in the four county boroughs of South Wales increased approximately +14%, from 541,120 in 1940 to 625,540 in 1969.

Municipal Boroughs

The total population residing in municipal boroughs between 1940 and 1969 declined in absolute terms from 321,018 to 309,470, a fall of -8.5%. Rhondda MB experienced the most significant decline in population size proportionally among MBs during this period, which suffered a fall of -17.21%. At the beginning of the study period, Rhondda MB had previously been the most densely populated of the municipal boroughs situated in South Wales, second only in population size to Cardiff CB. Between 1940–1969, only four of the 14 municipal boroughs in South Wales experienced a decline in population size. Significant population growth among MBs was sustained throughout the period in market towns, such as Haverfordwest MB (+37.7%), Pembroke MB (+32.7%), and Abergavenny MB (+15.6%).

Urban Districts

The total population increased from 598,152 to 613,1120 between 1940 and 1969, growing by +9.2%. The largest percentage increase in population size came in Cwmbran UD, rising by +142.6% during the study period. The settlements designation as a 'new town' in 1949 primarily drove this explosive growth. Seventeen UDs experienced growth in population size during the study period. However, of these districts, 11 districts witnessed an increase of +10% or greater. Of the 20 urban districts which experienced depopulation throughout the study period, a significant number are located within the Valleys. These districts include Maesteg UD (-6.6%), Ebbw Vale UD (-6.7%), Pontypool UD (-8.4%), Pontypridd UD (-8.8%), Blaenavon (-20.3%) and Abertillery UD (-17.1%).

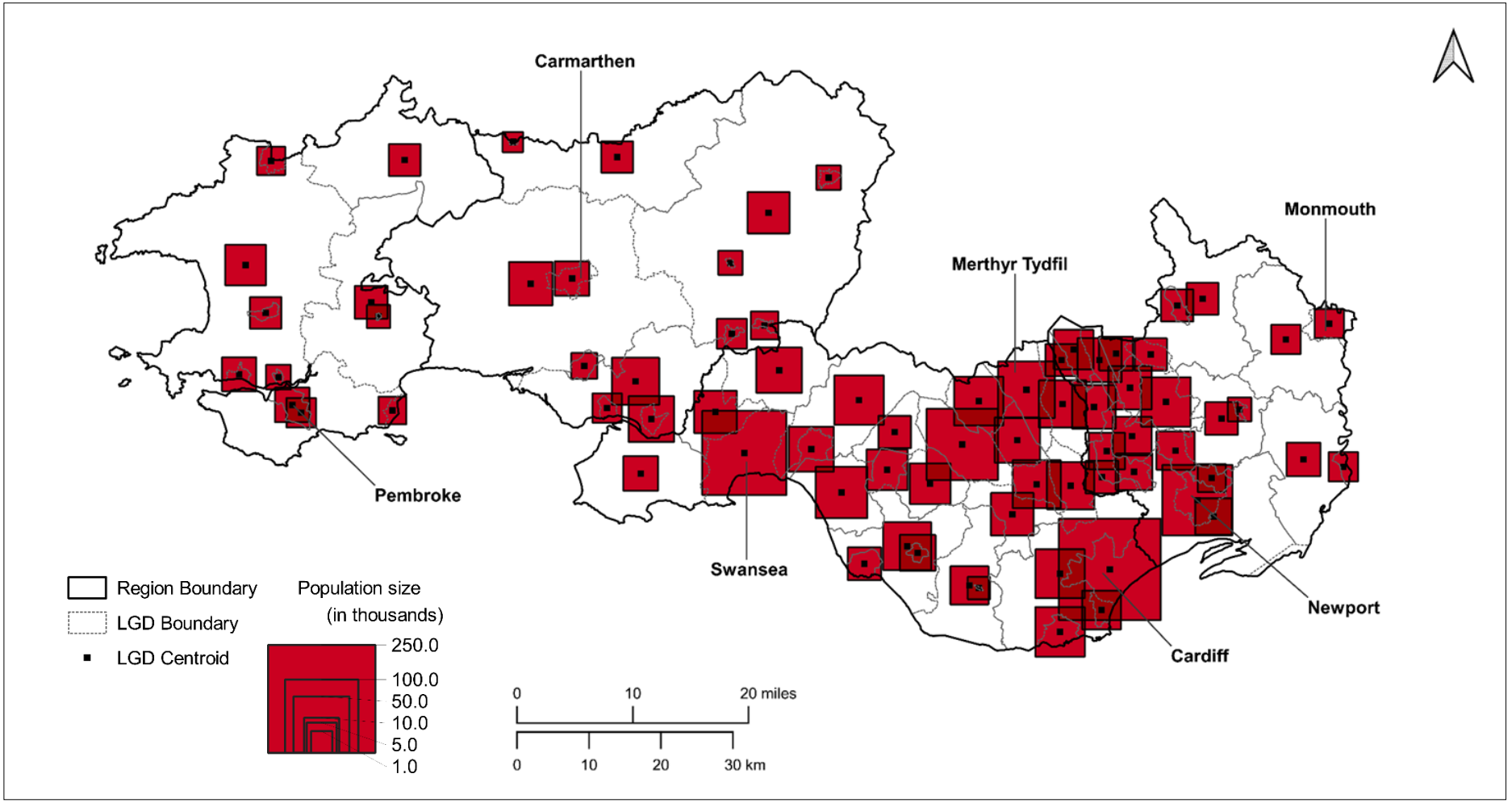

Figure 3.13 *Geographical distribution of the population of South Wales, by local government district, 1940–1969. The mean population size for each district for the study period is utilised to produce the proportional symbol. Squares represent the mean population size of each local government district during the study period.*

Rural Districts

The total population inhabiting rural districts in South Wales increased by 26.3%, from 372,500 in 1940 to 439,690 in 1969. Of the 20 rural districts in South Wales, 14 experienced population size growth, with nine RDs experiencing an increase of 20% or greater across the study period. Rural communities across the four counties of South Wales experienced significant population growth.

3.9.2 Birth Rate

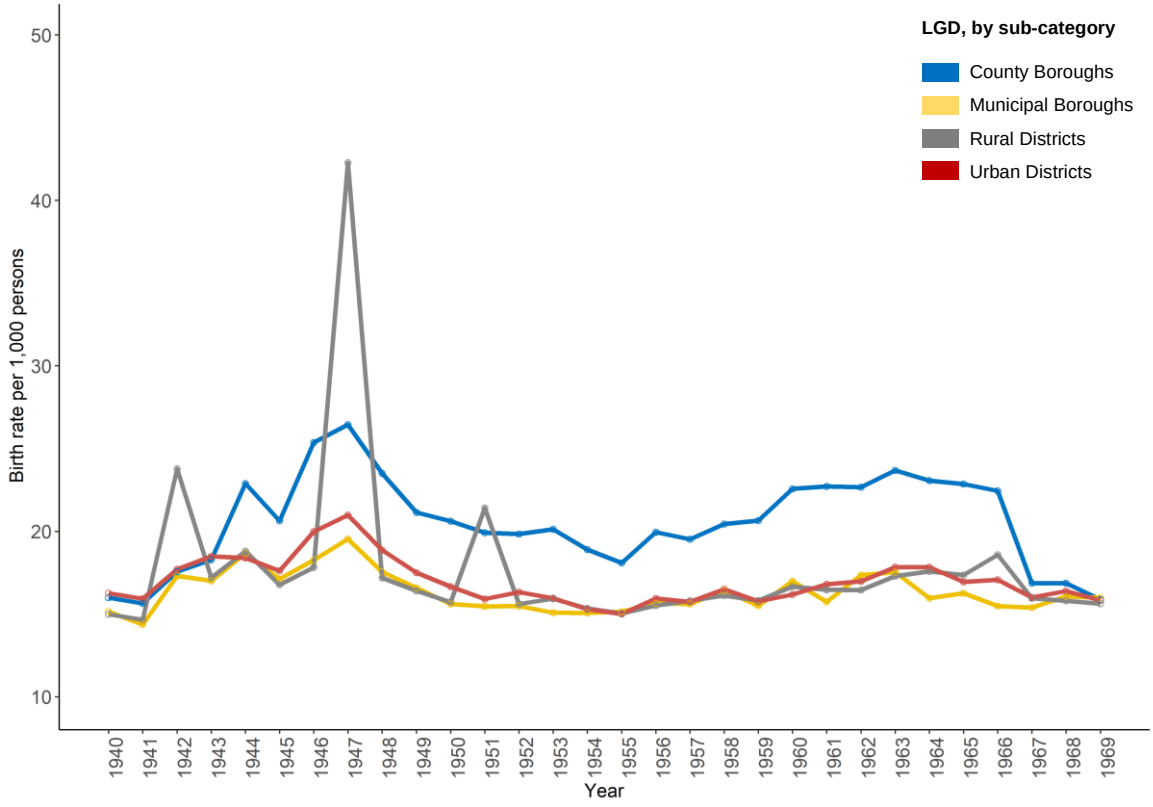

Figure 3.14 *Mean annual birth rate per 1,000 persons by sub-category of local government district in the South Wales region, 1940–1969. The mean annual birth rates peaked dramatically for rural districts in contrast to other districts in 1947-1948. After this moment in time, birth rates remained relatively constant across all districts, with some growth in the early 1960s before another fall in the latter half of the decade. For all districts categories, the mean birth rate had stabilised around approximately 15 births per 1,000 persons by 1969.*

Regional Summary

In absolute terms, the annual number of births in South Wales steadily increased between 1940-1961, before declining between 1961 and 1969. Across all regional districts, the annual number of births peaked in 1947, coinciding with the first baby boom, after which the number of births fell sharply, reaching a nadir in 1955. From the mid-1950s onwards, the birth rate recovered dramatically, peaking in 1963 and 1964. This rise coincided with the second baby boom spurt during this period across England and Wales.

County Boroughs

Between 1940 and 1961, Cardiff CB's birth rate increased from 16.57 births to 19.53 births per 1,000, with the peak birth rate recorded in 1947 (22.9 births per 1,000). A second, lesser peak was reached during the 1960s; 20.26 births per 1,000 in 1963. In subsequent years until the end of the study period, the birth rate fell dramatically, below wartime levels, to 15.87 births per 1,000 in 1969. The mean birth rate per 1,000 for the combined CBs population was almost identical in the first and final years of the study period; 15.99 births per 1,000 in 1940 compared to 15.98 births per 1,000 in 1969 (see Fig. 3.14).

Municipal Boroughs

The mean birth rate across municipal boroughs increased by approximately 25% in the lead-up to the beginning of the baby boom period. The birth rate rose from 15.1 births per 1,000 in 1940 to 19.5 births per 1,000 by 1947. From 1947 onwards, the birth rate fell to 15.07 per 1,000 in 1954. In subsequent years, the birth rate rose briefly, rising to 17.53 births per 1,000 in 1963. From 1963 to 1969, the birth rate declined and more-or-less plateaued (see Fig. 3.14), falling to 15.8 per 1,000 by 1969.

Urban Districts

The mean birth rate fell marginally between 1940 and 1969, from 16.26 births per 1,000 to 15.62 births per 1,000. Birth rates initially peaking during the first baby boom in 1947,

reaching 20.98 births per 1,000. The birth rate fell over 25% in subsequent years, reaching a low of 15.00 per 1,000 in 1955. The birth rate recovered in the late 1950s and early 1960s, spiking at 17.83 births per 1,000 in 1964. This growth was succeeded by a decline, similar to the one observed among municipal boroughs, falling to 15.60 births per 1,000 in 1969.

Rural Districts

Birth rates in rural districts experienced steady growth across the study period, rising from 14.99 births per 1,000 in 1940 to 15.62 per 1,000 in 1969. The mean birth rate across South Wales RDs peaked in 1947 (see Fig. 3.14), registering 42.28 births per 1,000, before falling to 15.03 in 1955. The birth rate recovered slightly in the late 1950s, rising to 17.57 per 1,000 in 1964 in the final year of the baby boom period, before slowly falling and plateauing.

3.9.3 Population Density

Table 3.2 *Temporal changes in population density (per km^2), by major urban centre and district sub-category, South Wales, 1940–1969.*

	Population density (per square kilometre)							
	1940	**1951**	**1961**	**1969**	**Peak**	**Year**	**Trough**	**Year**
All Districts	970.5	889.2	872.4	871.0	970.5	1940	871.0	1969
Cardiff CB	4,341.0	4,001.4	4,199.0	3,629.0	4,341.0	1940	3,629.0	1969
CBs	2,841.7	2,556.5	2,599.7	2,265.6	2,841.7	1940	2,265.6	1969
MBs	1,427.2	1,321.5	1,316.2	1,314.6	1,427.2	1940	1,314.6	1969
UDs	1,069.1	973.3	931.9	954.4	1,069.1	1940	954.4	1969
RDs	94.03	97.6	106.2	127.2	127.2	1969	94.0	1940

Regional Summary

The mean population density in the South Wales region fell consistently across the study period, declining from 970 persons per km^2 in 1940 to 871.01 persons per km^2 in 1969. The most significant fall in population density during the study period was registered between 1940–1951 (see Table 3.2). This decline was felt chiefly in the immediate post-war years, falling by ~ 10% to 889.20 persons per km^2 in 1951 compared to 1940 levels.

3.10 Explaining Regional Demographic Change

The post-war period witnessed a significant redistribution of the population of England and Wales. According to the model of growth and change for urban areas in England forwarded by Hall et al. (1974), before the twentieth century, people and employment were predominantly concentrated in urban cores. Urban cores are defined as the old city within a metropolitan area, with high populations and population densities. For example, the core of the Greater Manchester conurbation is Manchester CB. Between 1900 and 1950, this had slowly begun to shift, with populations starting to decentralise from urban centres into smaller settlements on the metropolitan margins (Robert and Randolph, 1983). However, employment remained firmly centralised in the core. Of the total metropolitan population of England and Wales, 71% resided in urban cores in 1931 (Hall, 1974). By the 1960s, it was clear that the process of decentralisation, of both population and employment, had begun (Hamnett and Randolph, 1982), with the frontier of population decentralisation advancing far into the inter-metropolitan hinterlands. By 1971, the proportion of the English and Welsh population residing in major urban centres had declined to 59% (Hall, 1974).

This growing movement of the population from the old urban cores into largely autonomous centres of employment had become one of the significant social changes of the post-war years (Hamnett and Randolph, 1982). The final stage of Hall (1974) model was the overall loss of population and employment in the large metropolitan centres to an outer ring of

peripheral urban areas, which, in turn, would begin to experience some level of population loss to smaller satellite towns. This stage results in greater regional dynamism (Hall, 1974). During the study period, Manchester and Liverpool experienced this significant decentralisation of population and employment. Consequently, both cities would register some of the most significant population losses in absolute terms outside of London during the 1960s. This process was accompanied by the growth of remote, historically sparsely populated, rural areas lying beyond the immediate metropolitan hinterlands, representing a significant departure from previous movement patterns (Vining and Kontuly, 1978).

The process of decentralisation evident in Manchester CB and Liverpool CB during the latter half of the study period, as well as in free-standing urban areas in South Wales, such as Cardiff CB, where decentralisation had already affected medium-sized centres in the less prosperous areas (Hall, 1974), was fuelled by post-war slum clearance and significant geographical shifts in industrial activity. Between 1955 and 1985, approximately 1.5 million houses were demolished due to slum clearance in England and Wales, displacing over 3.5 million people (Yelling, 2000). The majority of clearances occurred in conurbations and older industrial towns, in areas such as Merseyside and Greater Manchester. The North West had played an essential part in the early stages of the industrial revolution and accounted for 32% of all unfit houses (Yelling, 2000). In 1954, council returns recorded 88,000 houses in Liverpool CB and 68,000 in Manchester CB as unfit for habitation and in need of demolition (Yelling, 2000). During a speech in the House of Commons by Alf Morris MP in 1965, it was noted that 20% of the country's most inadequate dwellings were in the North West, with approximately three-quarters of the region's poorest residences located in a belt of land dominated by Manchester and Liverpool. In Manchester, around 27% of the total housing stock was considered unfit for habitation (*Hansard, 22 November 1965,* col *202-14*). Between 1955 and 1965, 12,200 houses had been cleared in Liverpool, resulting in significant population movement away from the core (Yelling, 2000).

Many urban districts located in South Lancashire experienced significant growth due to the redistribution of populations from overly crowded cities due to housing clearances, such as Salford, St. Helens and Wigan. For instance, Worsley UD's population grew by 91% during the study period after Salford CB was forced to rehouse many of its inhabitants in overspill estates due to a lack of available land (Pratt, 1977). Around Manchester CB and Liverpool CB, there was little appetite for new towns. Many free-standing towns situated within the urban hinterlands, such as Middleton MB, Crosby MB and Radcliffe MB, experienced considerable growth as the densely crowded urban centres decentralised. This was also true of urban districts representing satellite towns in the wider metropolitan areas of the region and reflected by the surge in population density among urban districts as population density among CBs and the two major conurbations fell significantly during the study period (see Table 3.1). One Lancashire urban district designated as a new town was Skelmersdale UD in 1961 (Field, 1968). This existing town was substantially expanded to accommodate the 'overspill' population from Liverpool. Whilst Liverpool CB registered a population decline of 9.2% between 1961 and 1969, according to population size statistics obtained from the Registrar–General's *Statistical Review*, Skelmersdale experienced a population boom, with its population rising from ~6,300 in 1961 to ~23,000 in 1969, an increase of almost 375%. In terms of the proportion of 1955 housing stock cleared by 1985, the proportionate impact was most significant among Lancashire cities. Liverpool, Manchester, and Salford all cleared more than a quarter of their 1955 housing stock over the next three decades (Yelling, 2000).

In contrast to Lancashire, South Wales did not experience clearances on a similar scale, despite the many densely populated towns that inhabited the valleys born from the industrial revolution and economically dependent on the fortunes of heavy industry. Throughout the early and mid-twentieth century, South Wales experienced a steady rate of depopulation, even in the rural periphery where a considerable proportion of the region's population lived, due to a continual decline in economic opportunities throughout the early and mid-twentieth

century. This was reflected by a gradual decrease in mean population density across all urban sub-categories of local government districts, and a relatively low rate of growth in population density within rural areas compared to the Lancashire region (see Table 3.2). Cwmbrân UD on the Monmouthshire-Glamorgan border was designated a 'New Town' in 1949, as part of the UK Government's New Town Act, to stem depopulation in the region. It was hoped the new town would provide new employment opportunities in the Southeastern portion of the South Wales coalfield (Wannop, 1999) rather than provide housing for those left homeless by wartime bombing or due to slum clearances. The scale of inter-regional industrial migration during the post-war years in England and Wales should also be considered a driver of population decentralisation. Between 1945 and 1965, 2,756 migrant factories were set up by manufacturing firms in areas of the country other than the towns in which they formerly operated. Keeble (1972) notes that these fundamental geographical shifts in industrial activity had a significant impact upon more peripheral counties and regions in England and Wales, where factory migration provided employment opportunities in the economically stagnant development areas.

3.11 Chapter Summary

During the study period, the overall population of the Lancashire and South Wales grew steadily, in spite of dramatic spurts in fertility and birth rates in the late 1940s and early 1960s; in the case of the former by approximately 11%, from 4,630,000 in 1940 to 5,130,000 in 1969, and in the latter by roughly 8.5%, from 1,833,000 in 1940 to 1,988,000 in 1969. But this is far from the full story, as the detailed account of the shifting demography of the regions provided in Sections 3.8 and 3.9 reveals. In Lancashire, ever greater numbers moved from the traditional urban cores of Southern Lancashire, either forced by the wind of change epitomised by postwar slum clearance and the construction of large social housing estates on the fringes of large cities or in neighbouring small towns, or more willingly, to smaller satellite settlements and rural areas as part of a growing process of suburbanisation

which took root in the 1950s, nurtured by the explosive growth in private car ownership in post war England and Wales. This was evidenced by a considerable fall in the number of inhabitants residing in county and municipal boroughs in Lancashire over the course of the study period, whilst the populations of urban and rural districts grew markedly (see Section 3.8.1). In South Wales, different trends were observed. Unlike Lancashire, major urban centres (Cardiff CB, Swansea CB and Newport CB) all experienced significant population growth during the study period. As the labour-intensive metallurgical industries of the Valleys fell into steep decline during the 1950s and 1960s, there was a process of population centralisation, contrary to the population decentralisation in England described by Hall (1974), with the population size and density of communities in the valleys falling as people migrated in search of greater economic opportunity in the metropolitan regional centres on the coastal plain of Glamorgan and West Monmouthshire.

It is important to recognise the shifting demography of the two regional metapopulations during the study period, as the traditional spatial structure of the regions' subpopulations, shaped by industrial forces in the eighteenth and nineteenth centuries, became obsolete in the postwar period as dramatic structural changes in employment and transport took hold. Sustained intra-regional migration inevitably results in a significant relocation, or indeed reallocation, of susceptible individuals and populations. In some districts, this may serve to significantly reduce the spacing between susceptibles, creating optimum conditions for epidemic outbreaks to emerge and for disease to persist. In other districts, the geographical redistribution of significant numbers of individuals during the study period only serves to further diminish the likelihood of disease persistence, removing the human tinder wood that would otherwise fuel intrinsic sparks of epidemic activity, potentially driving the spatial import of infection. In the following chapter, a descriptive analysis of disease incidence and spatial synchrony of measles, pertussis and scarlet fever activity in Lancashire and South Wales at the regional and district level will be undertaken, exploring the effects of the demographic events, trends and processes detailed in this chapter on disease persistence.

Chapter 4: Exploratory Data Analysis

4 Introduction

The following chapter presents an exploratory and descriptive analysis of time-series of monthly notification data for measles, pertussis and scarlet fever in Lancashire and South Wales (January 1940–December 1969). Time series plots are presented and discussed, visualising regional trends in incidence of the three infections, as well as time series trends for two regional subsets of three districts of varying population size over the course of the study period. Choropleth maps visualising temporal changes in disease incidence (cases per 100,000 population) for nine time–windows in Lancashire and South Wales are presented and discussed. Finally, a statistical examination of the spatial synchrony of epidemic behaviour at the local and regional level for Lancashire and South Wales, using sample correlation coefficients, is presented. For reference, further details on the districts analysed and discussed in this chapter, principally their geographical location within their wider region, and in relation to other districts, can be found in Appendix I (Figures I.1 and I.2).

4.1 Time Series Plots: Measles

4.1.1 Regional Trends

Across the study period, both major and minor measles epidemics show a biennial pattern (see Fig. 4.1). Major epidemic outbreaks are often preceded by minor epidemics and followed by an inter-epidemic trough. Major epidemic years in Lancashire tend to exceed 200 cases per 100,000 population. The largest epidemic peak is observed in 1957, with an incidence rate of ~300 cases per 100,000 population. Notably large epidemic outbreaks are observed in 1947 and 1963, coinciding with the beginning and end of the baby boom period.

After 1963, major epidemics in successive years do not rise above 200 cases per 100,000 population at their peak, as the fertility rate declined throughout the mid- and late-1960s; minor epidemics also fall significantly in scale.

Major epidemics in South Wales are greater in magnitude but more irregular in their cyclical pattern over the course of the study period. Measles epidemics mostly follow a biennial pattern, from 1940 to 1955, but the odd triennial epidemic is evident (see Fig. 4.1). Similar to Lancashire, major epidemic outbreaks are often followed by an inter-epidemic trough lasting one year and preceded by minor epidemics. Major measles epidemics in South Wales tend to be more substantial in magnitude than epidemic outbreaks in Lancashire, regularly exceeding 200 cases per 100,000 population. The most significant epidemic outbreak in South Wales was observed in 1955, with the epidemic exceeding a notification rate of 400 cases per 100,000 population. Intriguingly, the 1955 measles epidemic in South Wales was not prefaced but succeeded by a minor epidemic in the following year.

The 1955 epidemic outbreak marks a major departure in terms of the cyclical activity of the measles epidemics, with a five-year gap until the next major measles epidemic observed with a notification rate in excess of 200,000 cases per 100,000 population. In 1961, there was a major measles epidemic exceeding 200,000 cases per 100,000 population for the first stime since 1955, followed by annual epidemics between 1963 and 1965. The magnitude of measles outbreaks falls substantially after 1961 to the end of the study period, with no epidemic exceeding 200,000 cases per 100,000 population. With reduced magnitude, the epidemic cycle slowly reverts back to a biennial cycle across the 1960s. The introduction of routine measles vaccination in 1968 signifies a dramatic decline in notifications of measles cases in the final two years of the study period.

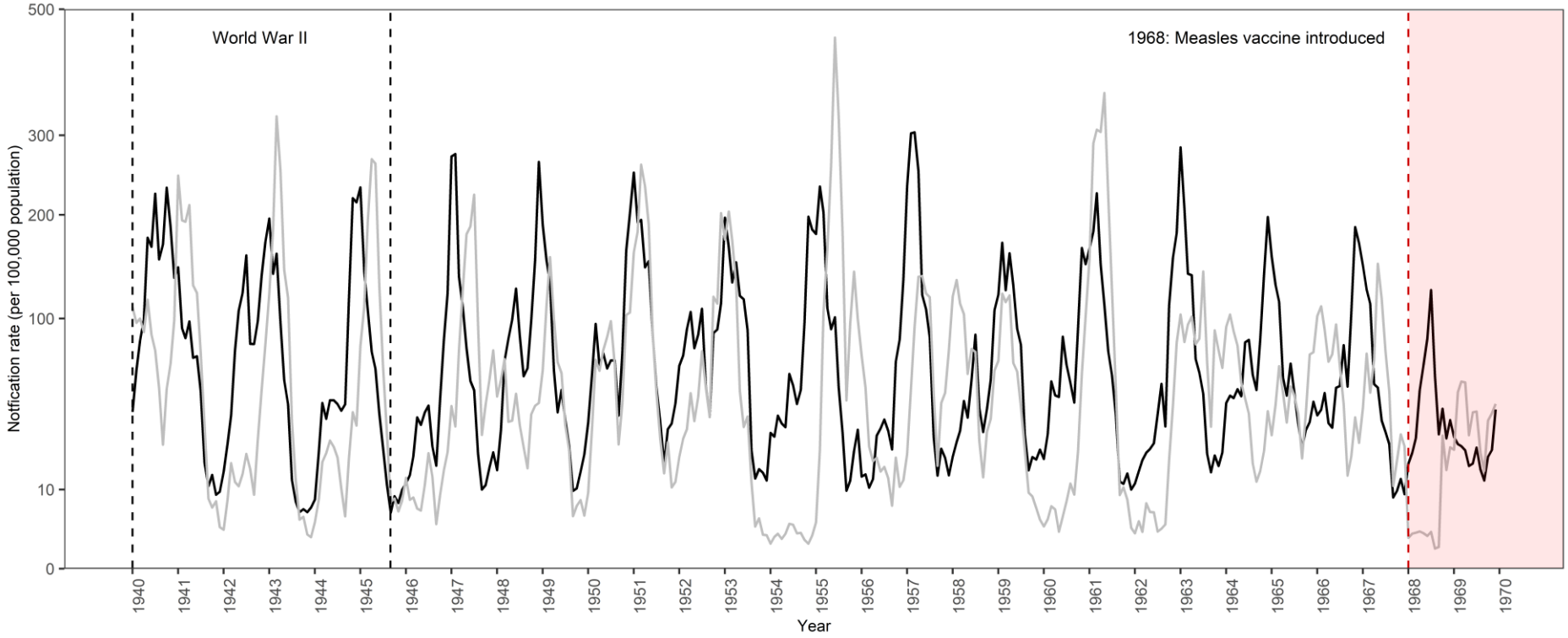

Figure 4.1 *Monthly time series of measles notification rates per 100,000 population in Lancashire and South Wales, January 1940–December 1969.* ***Black*** *= Lancashire,* ***grey*** *= South Wales. Red dashed line marks the onset of mass vaccination; red shaded area represents vaccine era. Source: Registrar–General's Weekly Returns (HMSO: London).*

4.1.2 District Trends: Lancashire Subset

Manchester CB exhibits a regular biennial pattern of major measles epidemics throughout the study period (see Fig. 4.2B), characteristic of Type I waves of infection. The largest epidemic was observed in 1957, with over 3,000 measles cases reported. Major epidemics in 1959, 1961 and 1963 reported between 1,400–1,600 cases, with the magnitude of cases continuing to fall across the 1960s. Prior to the introduction of measles vaccination in 1968, the incidence rate fell below 1,000 cases during major outbreaks in 1965 and 1967, reflecting the rapid decline in birth and fertility rates after the end of the baby boom period in 1964. Despite the fall in case magnitude, Manchester CB maintains a strongly biennial cycle of epidemic outbreaks until the end of the study period.

St Helens CB (Fig. 4.2C), located 33km to the west of Manchester CB, exhibits a more irregular biennial epidemic cycle suggestive of differences in the amplitude of disease activity, with more frequent minor peaks in measles activity between 1951 and 1957. This irregular cyclical activity is indicative of Type II waves of infection. Additionally, the largest epidemic outbreak coincides with the end of the baby boom in 1963. Major epidemic outbreaks in St Helens CB tend to exceed 500 annual cases. After the major epidemic in 1963, there was a significant fall in the number of cases in minor epidemic years. Despite the introduction of measles vaccination in 1968, a major outbreak was observed in 1969.

In the least populated district of the subset, Little Lever UD displays an irregular pattern of annual and triennial epidemics is evident between 1940 and 1965 (Fig. 4.2D). Little Lever UD is located 4km south-east of the major town of Bolton and 16km north-west of Manchester CB. After 1965, there is extremely limited measles activity and long periods of disease fadeout, indicative of Type III waves, due to the small population size and lack of sufficient re-introduction event, together preventing the persistence of epidemic outbreaks.

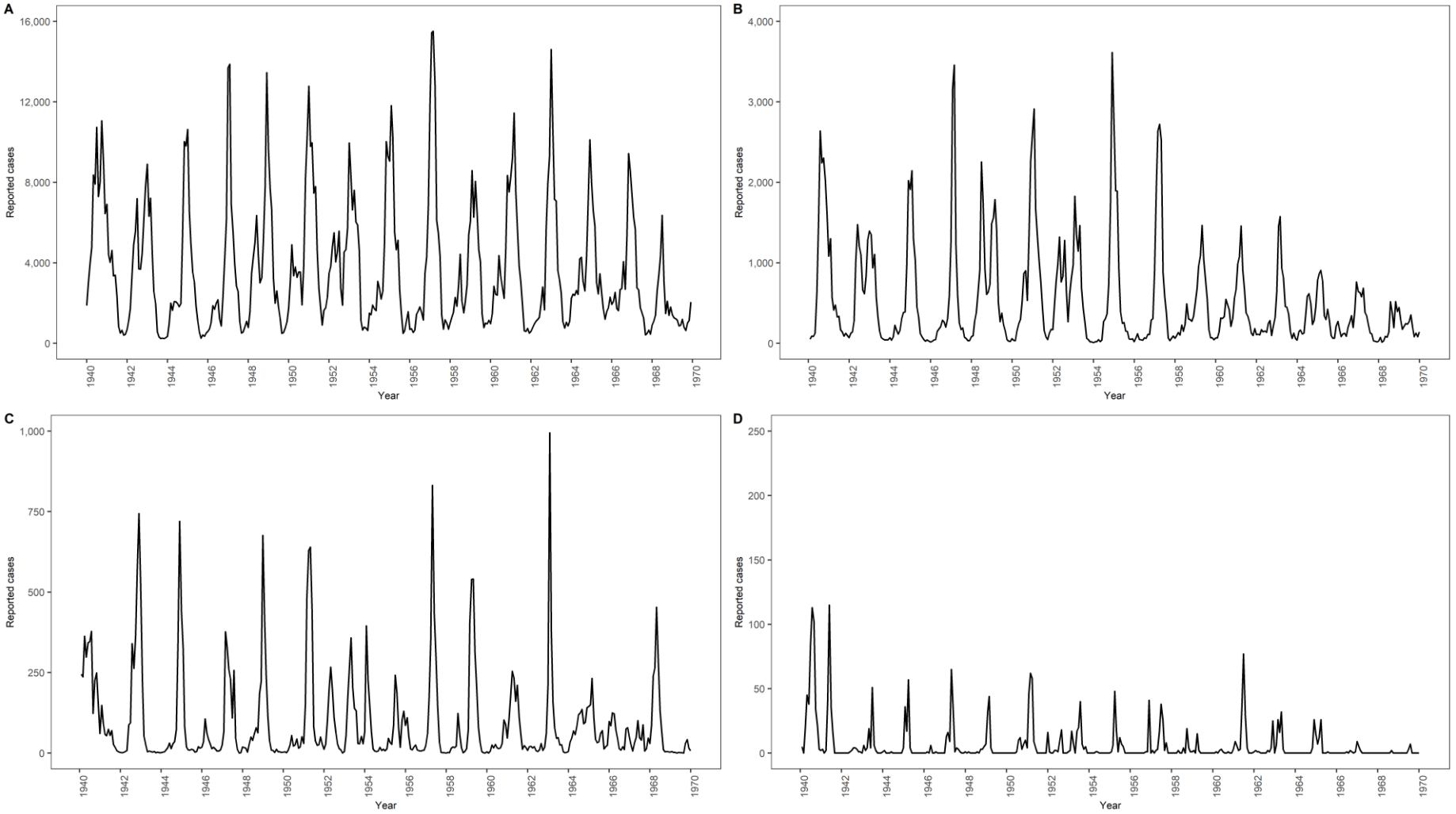

Figure 4.2 *Monthly time-series of reported measles cases for Lancashire and subset of districts of varying population size, January 1940–December 1969.* ***(A)*** *Regional trend,* ***(B)*** *Manchester CB (pop: ~650,000),* ***(C)*** *St Helens CB (pop: ~100,000),* ***(D)*** *Little Lever UD (pop: ~10,000).*

4.1.3 District Trends: South Wales Subset

Cardiff CB (Fig. 4.3B) mostly exhibits a mixture of biennial and triennial epidemic cycles across the period of study with no period of fadeout, characteristic of Type I waves of infection. This is with the exception of the late 1940s, where a series of annual minor epidemics are observed between major outbreaks in 1945 and 1949. The largest epidemics are observed in 1955 and 1961, with approximated 2,500 cases of infection reported. A biennial pattern of measles outbreaks is present throughout the 1950 but after the 1961 measles epidemic, the depletion of the susceptible pool coupled with falling rate of susceptible recruitment forces a shift in the cyclical nature of measles activity, with minor epidemics separating major outbreaks following a triennial cycle.

Merthyr Tydfil CB (Fig. 4.3C) exhibits a mixture of annual and biennial epidemic cycles from 1940 to the mid-1950s, with an increase in the magnitude of cases and amplitude of epidemic outbreaks as time progresses over this period. A major measles epidemic in 1956 is followed by a series of annual large annual epidemics between 1957 and 1959. Across the 1960s, irregular cycles of annual and biennial epidemics of reduced magnitude are evident bur characteristic of Type II waves of infection.

In the least populated district of the subset, located on the coastal edge of West Pembrokeshire, Fishguard & Goodwick UD (Fig. 4.3D). exhibits a highly irregular pattern consistent with a Type III community: sporadic annual epidemic outbreaks often followed by long periods with no reported cases, with these inter-epidemic periods sometimes lasting several years. The largest outbreaks are observed in 1943, and during the 1950s, when there seems to be a sustained period of elevated activity. Due to its small population size and significant distances between Fishguard & Goodwick UD and more populous, major population centres, there is extremely limited measles activity and long periods of disease fadeout, indicative of Type III waves with minimal re-introduction events.

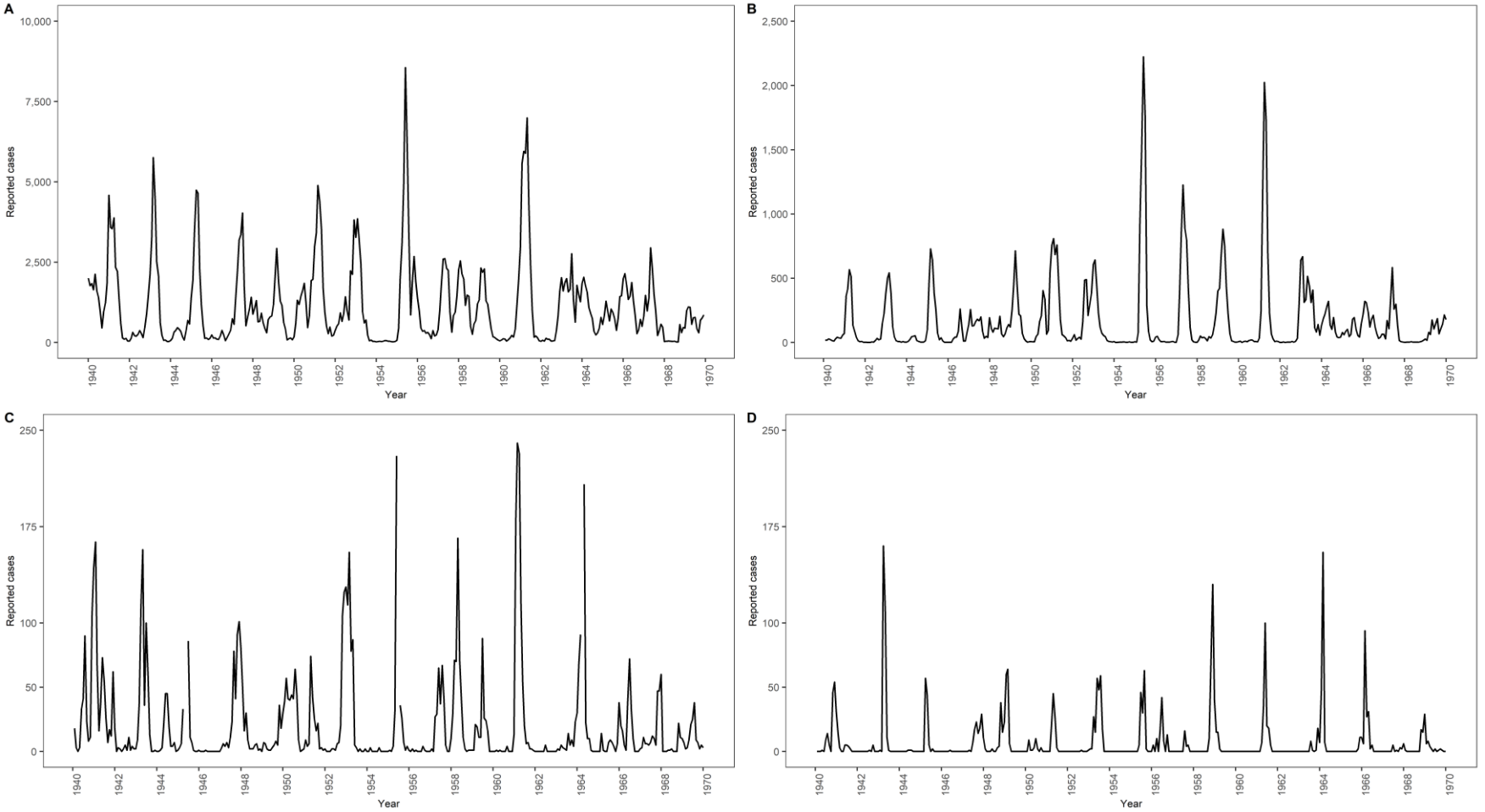

Figure 4.3 *Monthly time-series of reported measles cases for South Wales and subset of districts of varying population size, January 1940–December 1969.* ***(A)*** *Regional trend* ***(B)*** *Cardiff CB (pop: ~250,000),* ***(C)*** *Merthyr Tydfil CB (pop: ~100,000),* ***(D)*** *Fishguard & Goodwick UD (pop: ~5,000).*

4.2 Time Series Plots: Pertussis

4.2.1 Regional Trends

Across the study period, major pertussis epidemics in Lancashire tend to follow a cycle of three to five years (see Fig. 4.4). The largest epidemic outbreak is observed in 1941 at the height of World War II, with an incidence rate of 80 cases per 100,000 population. Large outbreaks in the pre-vaccine era prior to 1957 are characterised by an incidence rate above 40 cases per 100,000, and some irregularity in the cyclical pattern of epidemic activity is visible in the pre-vaccine period. With the onset of routine mass vaccination in 1957, the magnitude of cases and amplitude of pertussis epidemics declined substantially, with minor outbreaks emerging following regular, smooth five-year epidemic cycles. The incidence rate during these minor outbreaks does not rise above 20 cases per 100,000 population.

Pertussis epidemics in South Wales (Fig. 4.4) mirror the pattern observed in Lancashire, with major outbreaks observed approximately every two to three years during the pre-vaccination period, although the magnitude of cases is lower in South Wales, as evidenced by the size of the epidemic outbreak in 1941, far below the rate observed in the Lancashire region that year. During the mid-to-late 1940s, a series of minor epidemics are observed before three biennial outbreaks between 1949 and 1954. 1957 marks the only year in which the notification rate for an epidemic outbreak in South Wales exceeds an epidemic outbreak in the Lancashire region in the same year (Fig. 4.4); 40.16 cases per 100,000 population in the former compared to 29.97 cases per 100,000 population in the latter. During the vaccine era (1957-1969), major pertussis epidemics in South Wales follow a four-to-five year cycle with a reduced frequency of reported cases. Large outbreaks rarely exceed an incidence rate of 20 cases per 100,000 population. A visual comparison of the monthly notification time-series for Lancashire and South Wales suggests there is a time lag, with peaks in epidemic activity in the latter slightly behind the former.

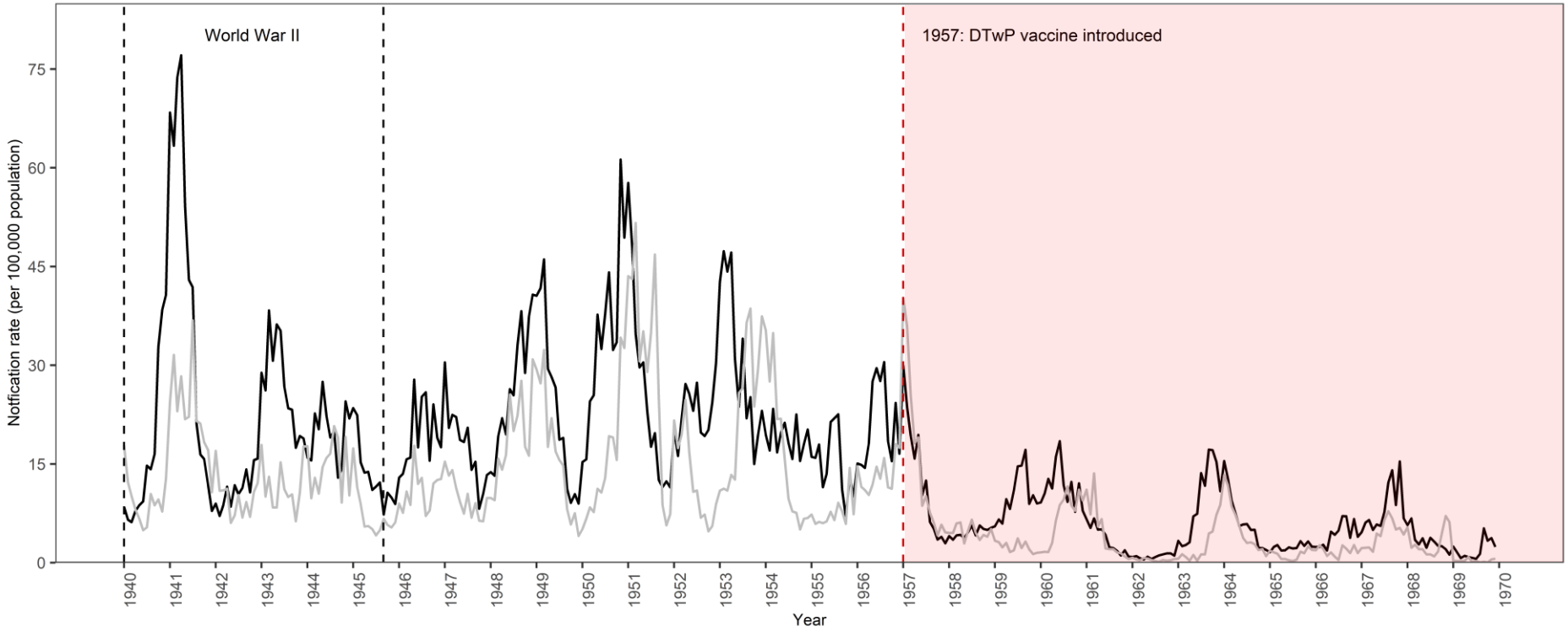

Figure 4.4 *Monthly time series of pertussis notification rates per 100,000 population in Lancashire and South Wales, January 1940–December 1969.* ***Black*** *= Lancashire,* ***grey*** *= South Wales. Red dashed line marks the onset of mass vaccination; red shaded area represents vaccine era. Source: Registrar–General's Weekly Returns (HMSO: London).*

4.2.2 District Trends: Lancashire Subset

Manchester CB (See Fig. 4.5B) exhibits a regular biennial pattern of major epidemics throughout the pre-vaccine era with no fade-out of disease, with minor epidemics preceding major outbreaks. The largest pertussis epidemic in Manchester CB was in 1941. From 1952 onwards, preceding the onset of mass vaccination, the magnitude of pertussis cases declines during a period of falling birth rates and thus reduced rate of susceptible recruitment. This is exacerbated by the advent of the vaccine era, with the introduction of vaccination in 1957 further curtailing the magnitude of reported case of pertussis but also reducing the amplitude of epidemic activity. During the vaccine-era, biennial epidemics have been replaced by a four-year cycle, but no fade-out of disease is reported. Manchester CB exemplifies a Type I settlement in which disease is endemic.

St Helens CB (Fig. 4.5C) exhibits a more irregular biennial epidemic cycle than Manchester CB, with fewer minor outbreaks between major epidemics during the pre-vaccine era. The largest pertussis outbreak in the district during the study period was in 1941. In the vaccine era, minor outbreaks are extinguished, with the exception of 1961 and 1964. There is a significant decline in the magnitude of cases in St Helens CB with vaccination serving to accelerate the trend southwards which began in the early 1950s with falling fertility rates. The irregular cyclical activity is indicative of Type II waves of infection.

Little Lever UD (Fig. 4.5D) exhibits a highly irregular pattern of epidemics in the pre-vaccine era, in contrast with the two more populous districts, with inter-epidemic periods between outbreaks varying in length from two to four years depending on the rate of disease reintroduction, characteristic of a Type III community. There is no major outbreak of pertussis in 1941, suggesting pertussis infection is limited in small local populations beyond densely populated, urban centres. The introduction of mass vaccination results in the almost complete eradication in Little Lever UD between 1957 and 1969, with extensive periods of fadeout and no further epidemic outbreak detected.

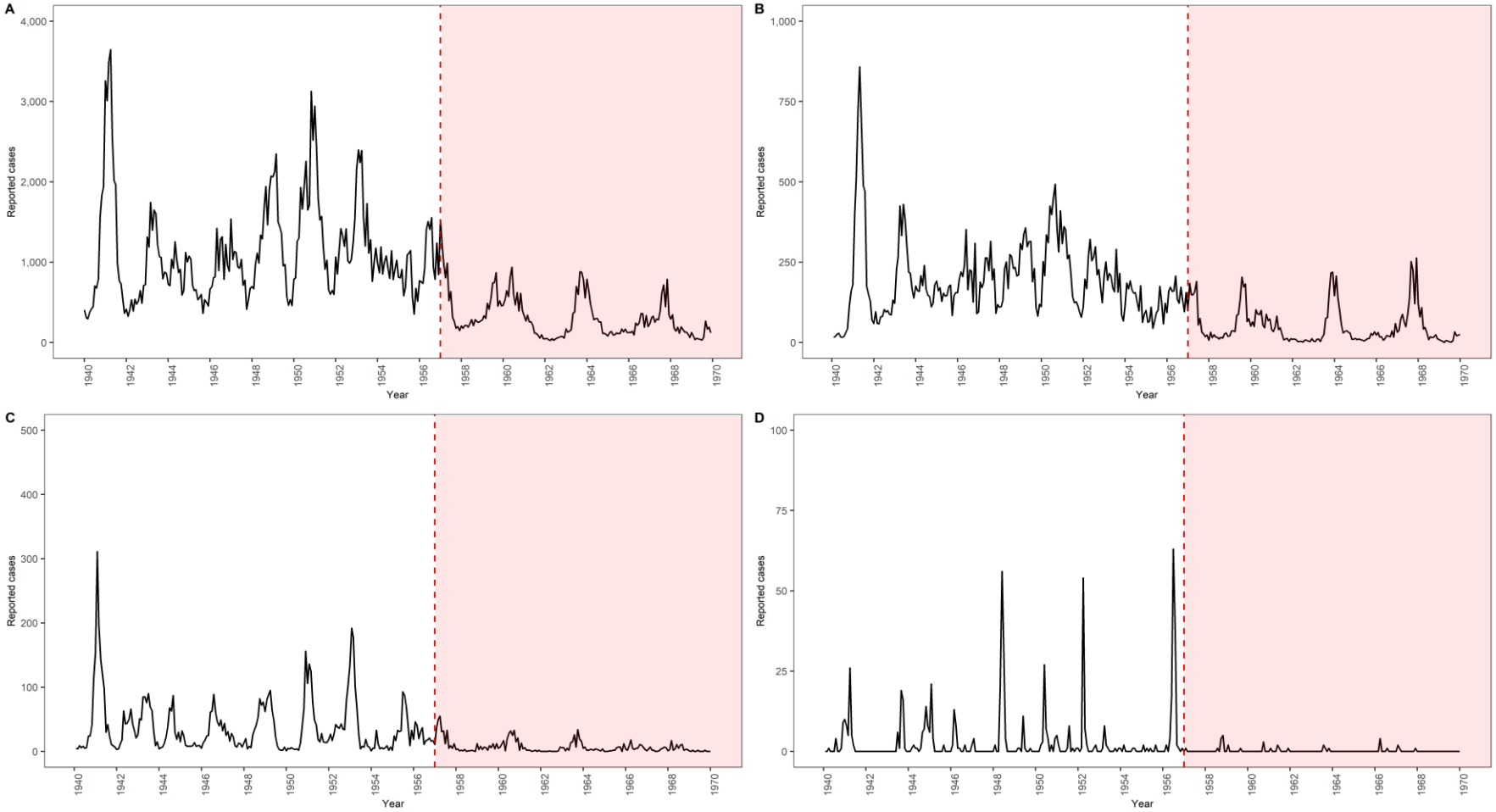

Figure 4.5 *Monthly time-series of reported pertussis cases for Lancashire and subset of districts of varying population size, January 1940–December 1969.* ***(A)*** *Regional trend* ***(B)*** *Manchester CB (pop: ~650,000),* ***(C)*** *St Helens CB (pop: ~100,000),* ***(D)*** *Little Lever UD (pop: ~10,000). Red dashed line marks the onset of mass vaccination; red shaded area represents vaccine era.*

4.2.3 District Trends: South Wales Subset

In the pre-vaccine era, Cardiff CB exhibits two distinct epidemiological patterns (See Fig. 4.6B). From the beginning of the study period to 1950, there are series of annual minor outbreaks slowly increasing in magnitude before a more stable, regular pattern of major biennial epidemics emerges in the 1950s. The largest pertussis epidemic in Cardiff CB was observed in 1951, coinciding with the largest regional pertussis epidemic during the study period. Between 1951 and 1957 there is much more pronounced epidemic cycle, with outbreaks occurring roughly every two years. During the vaccine-era (1957-1969), biennial epidemics are replaced by a four-year cycle, but no fade-out of pertussis infection is reported. Cardiff CB exemplifies a Type I settlement.

Merthyr Tydfil CB exhibits a much more pronounced epidemic cycle than Cardiff CB during the pre-vaccine era (Fig. 4.6C), with major pertussis epidemics detected every three-years. In the vaccine era, the pre-vaccine epidemic pattern is replaced initially by a series of minor outbreaks every 1.5 years initially, and Merthyr Tydfil CB experiences its largest pertussis epidemic during the study period in 1959, two years after the nationwide rollout of the pertussis vaccine, suggesting low vaccine uptake in the district in the early year of the immunisation campaign. After the 1959 outbreak, pertussis notifications fall significantly and outbreaks are highly irregular, with evidence of frequent fadeout before minor outbreaks. Merthyr Tydfil CB This irregular cyclical activity is indicative of a Type II community.

Fishguard & Goodwick UD exhibits a highly irregular pattern of sporadic epidemics in the pre-vaccine era (Fig. 4.6D), with the duration of inter-epidemic periods between outbreaks varying considerably. During the vaccine era, there are long periods of disease fadeout with the occurrence of short-lived outbreaks highly dependent on import of infection from external sources. This is indicative of Type III waves of infection.

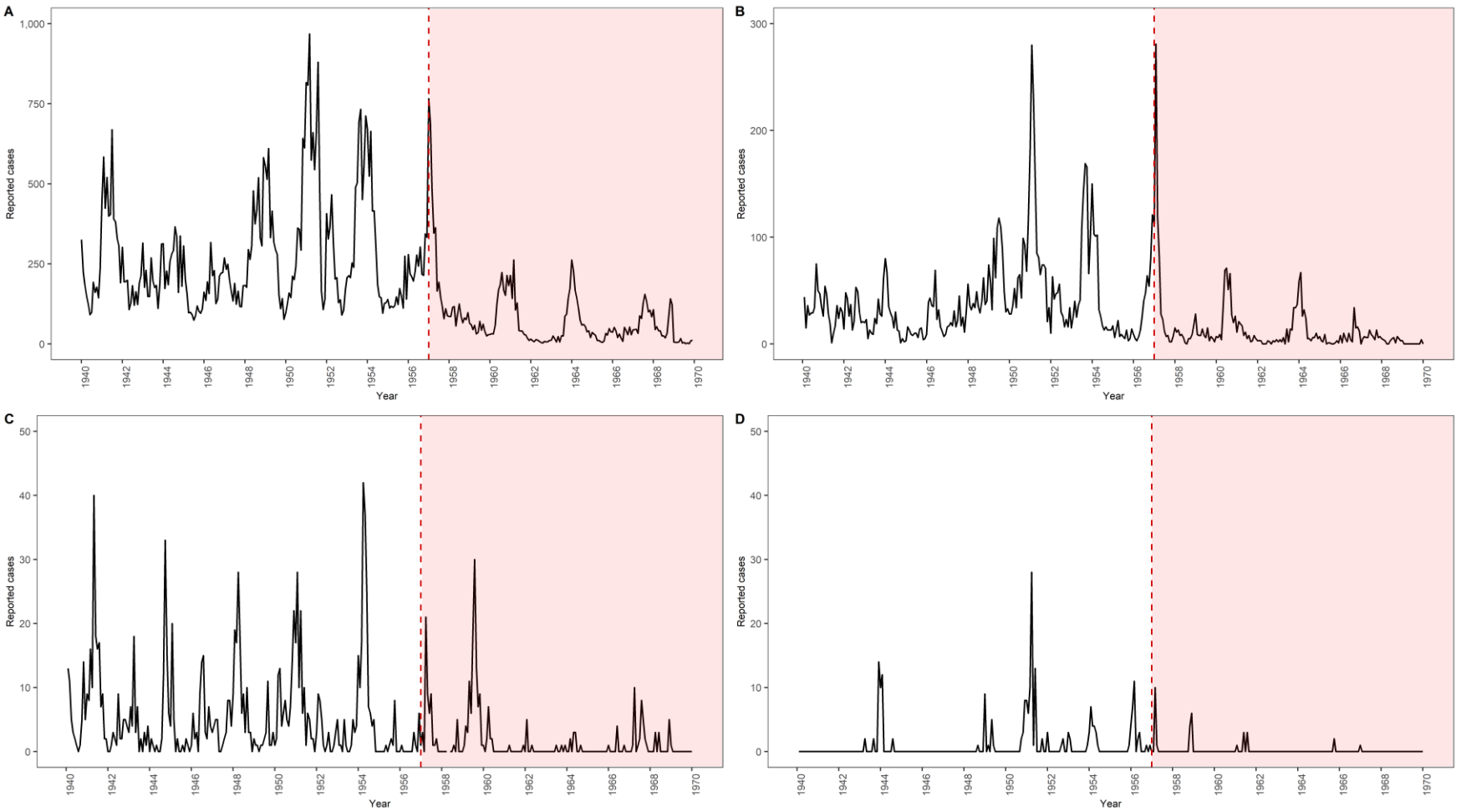

Figure 4.6 *Monthly time-series of reported measles cases for South Wales and subset of districts of varying population size, January 1940–December 1969.* ***(A)*** *Regional trend,* ***(B)*** *Cardiff CB (pop: ~250,000),* ***(C)*** *Merthyr Tydfil CB (pop: ~100,000),* ***(D)*** *Fishguard & Goodwick UD (pop: ~5,000). Red dashed line marks the onset of mass vaccination; red shaded area represents vaccine era.*

4.3 Time Series Plots: Scarlet fever

4.3.1 Regional Trends

The Lancashire and South Wales regions (see Fig. 4.7) exhibit a relatively strong degree of synchronicity with regards to both notification rates and patterns of epidemic activity. In the wartime period, there is a gradual climbing trend with the notification rate rising from approximately 20 cases per 100,000 population in 1940 to a high of 47 cases per 100,000 population in 1945, with significant, annual epidemic outbreaks in the winters of 1943-44 and 1944-45. From the summer of 1945 to the autumn of 1947 there is a significant fall in the notification rate in both regions, declining to approximately 10 cases per 100,000 population. However, there is a significant recovery in scarlet fever incidence between 1948 and 1950 coinciding with the beginning of the baby boom, with the notification rate reaching 40 cases per 100,000 in South Wales in early 1948, and a similar total in Lancashire and South Wales in 1950. From the beginning of the 1950s, the incidence of scarlet fever cases declines gradually and consistently in both regional populations, despite a period of high birth rates culminating in significant growth in the density and size of susceptible populations across Lancashire and South Wales, particularly concentrated in urban and metropolitan areas. Scarlet fever remains endemic in both regions during the study period. The epidemic cycle for the disease follows a mixture of annual and biennial outbreaks in Lancashire and South Wales. In the 1960s, there is a divergence between the two regional time-series; in South Wales, a large epidemic reaching approximately 30 cases per 100,000 population is recorded in the autumn and winter of 1968. In Lancashire, during this period, the notification rate is around five cases per 100,000 population. A period of heightened epidemic scarlet fever activity is recorded in Lancashire over a three-year period, from the end of 1964 to mid-1967.

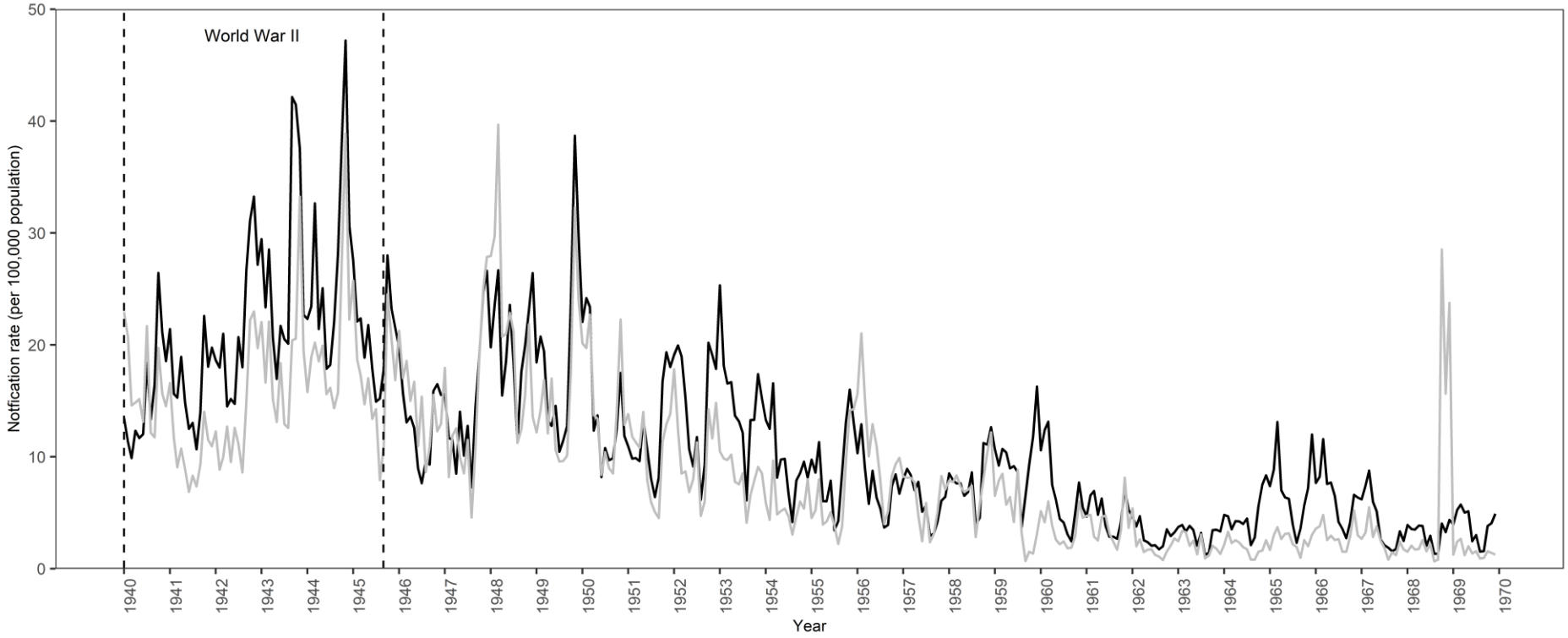

Figure 4.7 *Monthly time series of scarlet fever notification rates per 100,000 population in Lancashire and South Wales, January 1940–December 1969.* ***Black*** *= Lancashire,* ***grey*** *= South Wales. Source: Registrar–General's Weekly Returns (HMSO: London).*

4.3.2 District Trends: Lancashire Subset

In Manchester CB, the largest annual outbreak of scarlet fever is observed during the wartime period in 1943 (see Fig. 4.8B). There is a period of gradual decline in the magnitude of cases in reported in Manchester CB following the epidemic of 1943 which lasts to the summer of 1947, possibly due to the slow exhaustion of the supply of susceptibles as a consequence of annual epidemic activity. From the end of 1947 to 1951, there is a significant recovery in the number of reported cases of scarlet fever infection and a large outbreak in 1950. This outbreak is approximately 1.5–2 years after the first major peak in birth rates marking the beginning of the baby boom period in 1948. 1950 marks the last major outbreak in Manchester CB, with the number of scarlet fever cases reported falling steadily over the course of the 1950s and 1960s. With the exception of slightly heightened activity in 1965, there is no pattern of major epidemic outbreaks in Manchester CB during the period 1964-68, despite being one of the two major regional endemic centres of infection.

St Helens CB (see Fig. 4.8C). exhibits a clearly pronounced series of three annual epidemic outbreaks in the years 1944, 1945 and 1946, with no major outbreak in 1943, unlike that observed in Manchester CB. Unlike Manchester CB, St Helens CB exhibits a mixture of epidemic cycles; biennial major outbreaks between 1946–1952 succeeded by a seemingly quadrennial pattern of large outbreaks in the latter half of the study period, with minor outbreaks in intervening years. This irregular pattern of disease activity is characteristic of type II waves. Most notably, St Helens experiences its largest epidemic outbreak in 1956, exceeding the number of scarlet fever cases reported In Manchester CB that year despite possessing one-third of the population size. Visual comparison of the time series of scarlet fever incidence for Manchester CB and St Helens CB reveals very different epidemic cycles and outbreak patterns.

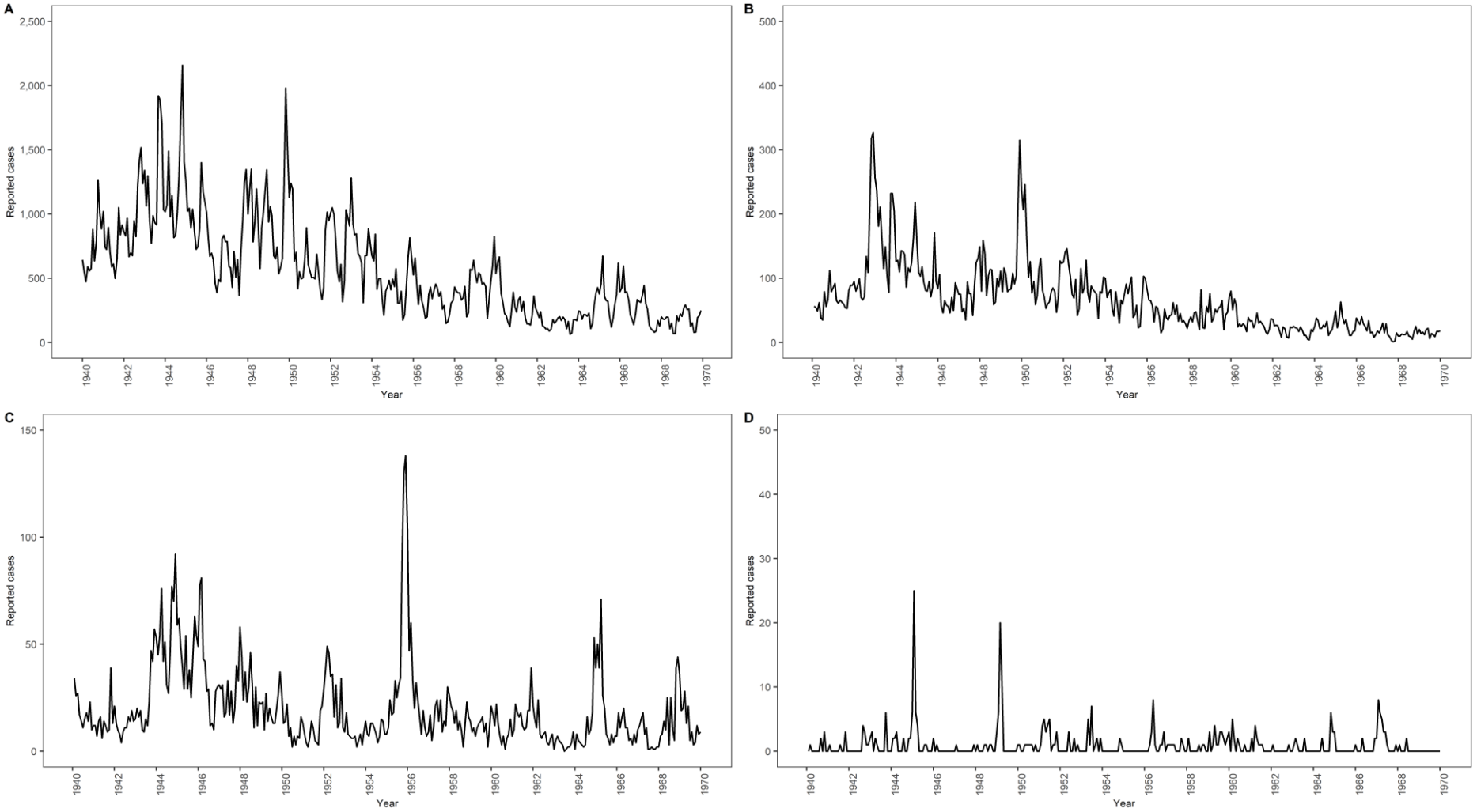

Figure 4.8 *Monthly time-series of reported scarlet fever cases for Lancashire and subset of districts of varying population size, January 1940–December 1969.* ***(A)*** *Regional trend* ***(B)*** *Manchester CB (pop: ~650,000),* ***(C)*** *St Helens CB (pop: ~100,000),* ***(D)*** *Little Lever UD (pop: ~10,000).*

St Helens CB is located on the edge of the Merseyside metropolitan area, situated only 17 km east of Liverpool CB. As has been shown in previous research (Grenfell et al., 2001), Liverpool CB has a more erratic, irregular cycle of disease activity that is out of sync with Manchester CB, due to extremely high birth rates, and these divergent dynamics may play an influential role in shaping the epidemic pattern observed in St Helens CB.

Similar to the outbreak of 1956, there is another large epidemic outbreak in 1965 that is not reflected in the Manchester CB time-series, with the number of scarlet fever cases reported exceeding the number notified in Manchester CB that year.

Little Lever UD (Fig. 4.8D). displays an irregular pattern of epidemic outbreaks, with inter-epidemic periods varying in length from one year to four years over the course of the study period. The largest epidemic outbreak was in observed at the conclusion of World War II in 1945. A second very noticeable epidemic surge in cases was observed in 1949, slightly preceding the major epidemics observed in Manchester CB and on a regional scale in 1950. After the outbreak of 1949, minor outbreaks are observed on an annual or biennial basis, and there are unusually short periods of disease fadeout, despite Little Lever's UD small population size, which may be explained by its location within the highly-connected, metropolitan hinterland of Manchester CB. One would therefore expect to see regular spatial import of infection from external sources (i.e., neighbouring satellite settlements), precluding the disease patterns associated with small urban or rural communities in the form of type III waves.

4.3.3 District Trends: South Wales Subset

Cardiff CB (see Fig. 4.9B). exhibits a three-to-five year cycle of major epidemics, with minor outbreaks in intervening years during inter-epidemic troughs. Significant outbreaks of

similar magnitude were observed in 1943, 1945, 1948 and 1953. From 1953 onwards, there is a gradual decrease in scarlet fever incidence, with a drastic drop in the number of reported cases clearly observable from 1959 onwards, incidentally coinciding with the introduction of mass vaccination for pertussis in England and Wales in 1957. A relatively large outbreak of scarlet fever is observed in 1969, mirroring the major epidemic outbreak detected in the regional time-series (Fig. 4.9A). Prior to the final years of the 1950s, the time-series of scarlet fever incidence reveals no period of fadeout, characteristic of Type I waves of infection.

Merthyr Tydfil CB (Fig. 4.9C) exhibits a pattern of major outbreaks approximately every 1.5 years up until the end of the 1950s, coinciding with the significant fall in incidence of scarlet fever infection witnessed across other districts in South Wales as well as Lancashire. The largest outbreak observed is in 1951, out of sync with the outbreak pattern of Cardiff CB. From the early 1950s onwards, the major epidemic cycle adopts a biennial pattern. A key feature of the scarlet fever incidence time-series in Merthyr Tydfil is the regularity of fadeout and reintroduction events characteristic of Type II waves, with frequent minor outbreaks between major events.

Fishguard & Goodwick UD (Fig. 4.9D) exhibits a time-series of disease incidence consistent with a Type III community: extremely sporadic epidemic outbreaks succeeded by long periods of fadeout, with no reported cases. Inter-epidemic periods last for several years; no scarlet fever cases were reported for over a period of approximately seven years (72 months), between 1962–1969. Due to its small population size and remoteness, Fishguard & Goodwick UD experiences extremely limited disease activity and long periods of disease fadeout, indicative of Type III waves with minimal re-introduction events.

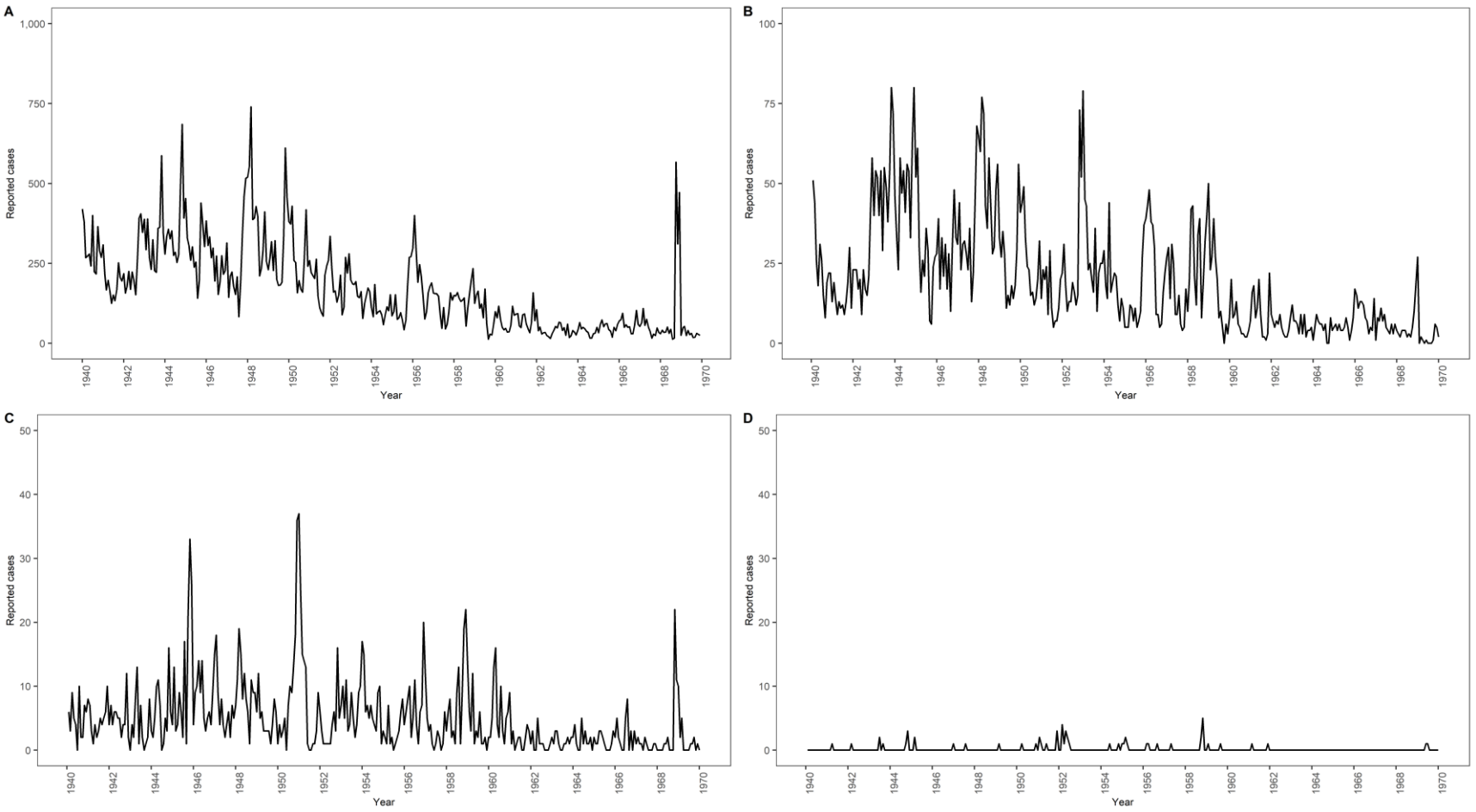

Figure 4.9 *Monthly time-series of reported scarlet fever cases for South Wales and subset of districts of varying population size, January 1940–December 1969. (A) Regional trend, (B) Cardiff CB (pop: ~250,000), (C) Merthyr Tydfil CB (pop: ~100,000), (D) Fishguard & Goodwick UD (pop: ~5,000).*

4.4 Incidence Mapping: Lancashire Region

4.4.1 Measles

Measles incidence in the Lancashire region is predominantly concentrated within the endemic centres of Liverpool CB and Manchester CB and their surrounding metropolitan hinterlands, as well as growing commuter districts lying on the urban/rural periphery, such as in the Rossendale Valley. Generally, measles incidence is found to be greater in Manchester's surrounding metropolitan area compared to Merseyside in Southwest Lancashire (see Fig. 4.10). This may be due to Greater Manchester representing a larger polycentric area and thus possessing greater levels of accessibility and connectivity between subpopulations, fuelling disease spread. However, districts that exhibit high measles incidence rates are widely and evenly distributed across the region in successive time–windows. Persistent hotspots across all nine of the time–windows include Middleton MB in Greater Manchester, Crompton UD, situated between Manchester CB and Oldham CB, and Barrowford UD in East Lancashire (see Fig. 4.10). Littleborough UD exhibits high rates of measles incidence between time–windows one and six.

Consistently higher rates in these districts, despite small or medium-sized populations, may be explained by their location on the regional border or in between major towns, and their function as key 'commuter' settlements, primarily residential with high connectivity with several local and large population centres. Lower rates of measles incidence tend to be found in sparsely populated districts located in the rural areas of east Pennine Lancashire and the north-west portion of the region, districts in the Furness peninsula, as well as small urban areas surrounding Lancaster MB and Morecambe Bay (see Fig. 4.10).

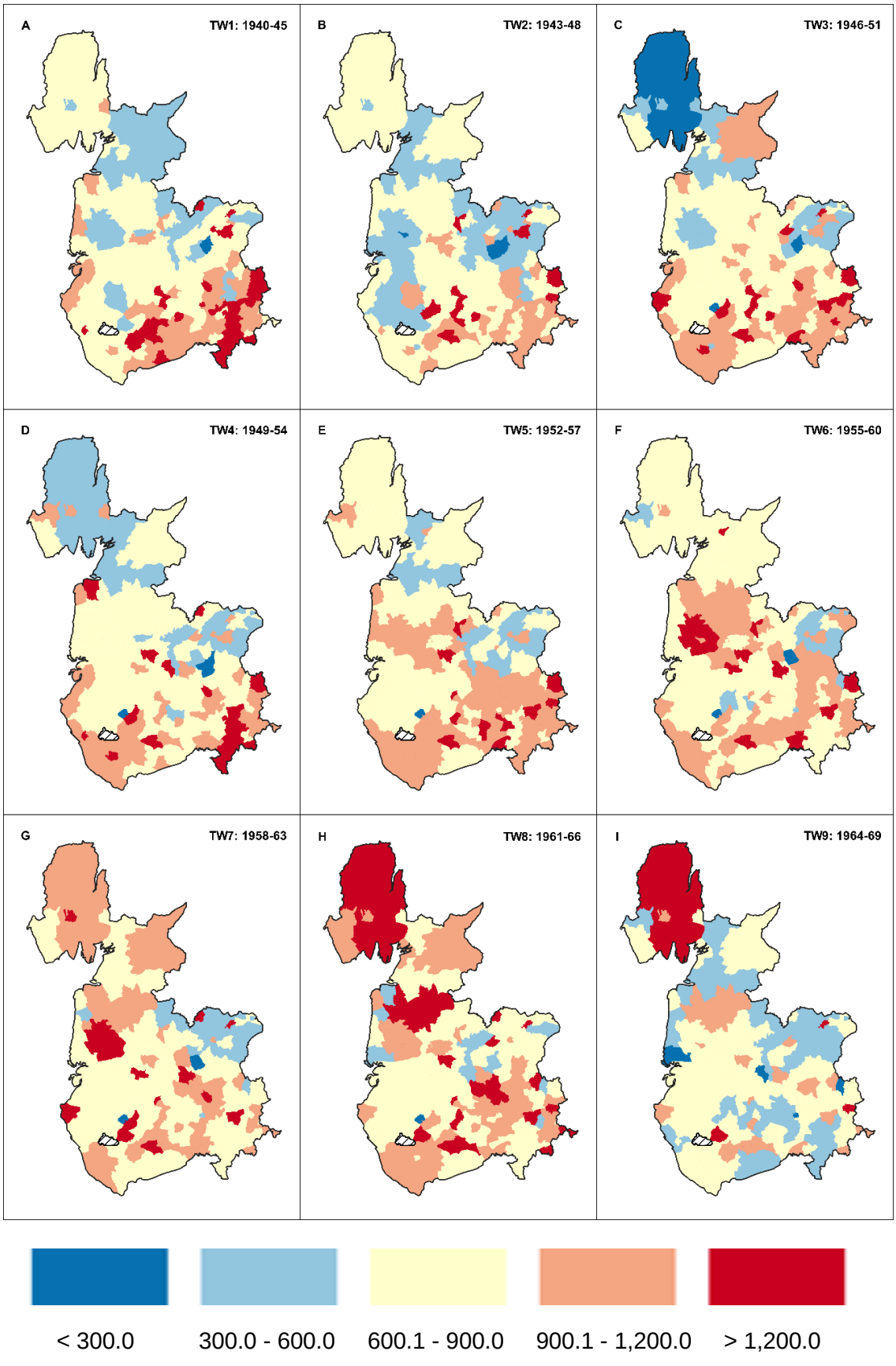

Figure 4.10 *Mean measles incidence (cases per 100,00 population) in the Lancashire region across nine 72-month time–windows, by local government district, 1940–1969.*

The lowest rates of measles incidence are most consistently observed across the nine time–windows in Pennine Lancashire in the east of the region (see Fig. 4.10), an area which suffered significantly from depopulation in the postwar period due to the collapse of local textile-based industries. In these geographical areas identified, the incidence rate for measles is typically lower than 300 cases per 100,000 population across successive time–windows. However, in the final three time–windows, notably elevated incidence rates for measles are observable in the north-west of the region, in districts such as Ulverston UD and Barrow-in-Furness CB. This may reflect violent epidemic episodes, and the spatial import of infection from neighbouring counties and population centres, such as the large town of Carlisle to the north, or by individuals moving into or visiting the Lake District, as population decentralisation and growing desire to move to the countryside gathered pace in the 1960s.

It should be noted that the fall in measles incidence rates below 900 cases per 100,000 population in many districts in the final time–window (see Fig. 4.10) partially reflects the effectiveness of the measles vaccine, introduced in 1968, in removing susceptibles from the population and increasing the spacing between remaining susceptible individuals.

4.4.2 Pertussis

In the pre-vaccine era (1940–1957; time–windows one to six), pertussis incidence is predominantly concentrated within the urban cores of Liverpool CB and Manchester CB and their surrounding metropolitan hinterlands which form the urban agglomerations of Merseyside and Greater Manchester. Low rates of pertussis cases are generally found in North Lancashire and the far north-west, Pennine Lancashire on the eastern edges of the region and Central Lancashire. These areas are more remote, generally populated by small and often isolated settlements. These areas of the Lancashire region also tend to have much greater spacing between susceptibles, with limited pools of susceptibles available.

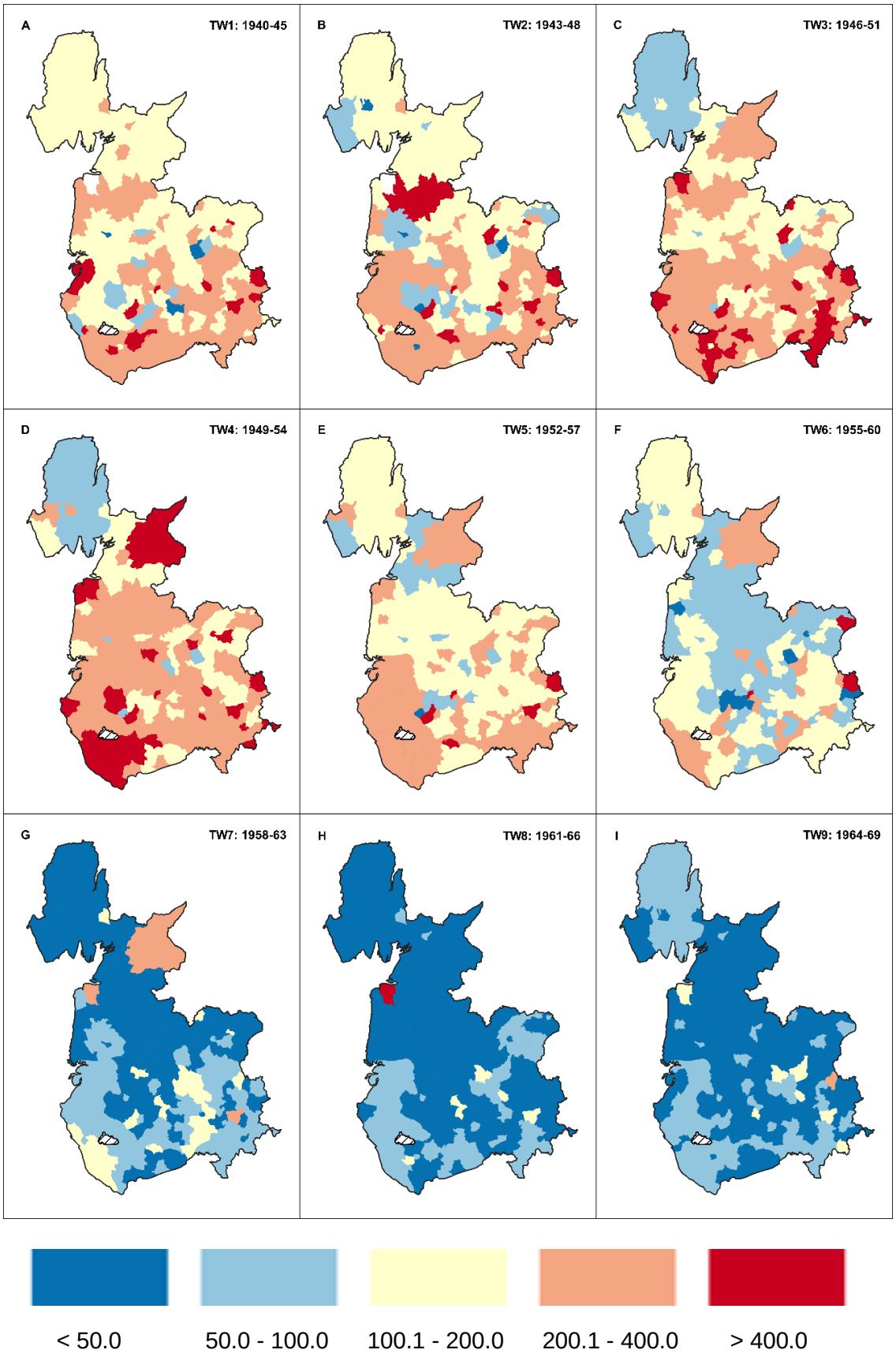

Figure 4.11 *Mean pertussis incidence (cases per 100,00 population) in the Lancashire region across nine 72-month time–windows, by local government district, 1940–1969.*

In urban districts which populate the hinterlands of South Lancashire, incidence rates range between 200 and 400 cases per 100,000 population (see Fig. 4.11). Relatively few districts possess an incidence rate consistently greater than 400 cases per 100,000 population, which may indicate potential hotspots of disease transmission. One such district is Littleborough UD, located on the urban-rural periphery of Greater Manchester and East Lancashire, approximately 20km northeast of Manchester CB, and 4km north of Rochdale CB, which consistently displays a rate greater than 400 cases per 100,000 population from time–windows one to six. Other urban districts which fit in this category include Upholland UD, which is situated on the fringes of West Lancashire bordering Merseyside, 6.5km west of Wigan CB and 20 km Northeast of Liverpool CB. Upholland UD exhibits high incidence rates consistently across the nine time–windows (Fig. 4.11). Middleton MB and Tottington UD, which form part of the hinterland and network of local population centres that surround Manchester CB, also exhibit heightened rates of pertussis incidence. High pertussis rates are also observed localised in small population centres located in North Lancashire, in districts which grew as a consequence of population decentralisation during the study period, as deindustrialisation gathered pace and fuelled internal migration in the region in search of better economic opportunities. Two such examples are Preesall UD, and neighbouring Fleetwood MB located in the Fylde Peninsula, which display high rates of pertussis incidence across the study period.

With the onset of routine pertussis immunisation for infants and children in 1957, after the introduction of the whole-cell vaccine during time–window six (1955-60), pertussis is in retreat across the Lancashire region, with a significant reduction in the magnitude of cases over the course of three successive time–windows. The incidence rate for all but 11 of Lancashire's 125 districts falls below 50 cases per 100,000 population by the ninth and final time–window (1964-69). Districts in which pertussis infection stubbornly persists are, for the most part, located within the metropolitan hinterlands of the Liverpool and Manchester conurbations, with isolated pockets of heightened pertussis activity found in Central

Lancashire, Pennine Lancashire, and even in the more remote northern portion of the region. High rates in these districts may be fuelled by localised outbreaks triggered by reintroduction of disease from external sources, by favourable demographic conditions resulting in a steady supply of susceptibles or due to high levels of accessibility and connectivity nurturing disease transmission.

4.4.3 Scarlet fever

In contrast to pertussis and measles, incidence of scarlet fever infection in Lancashire declines consistently and progressively across the study period. Time–windows one to three see higher rates of scarlet fever incidence primarily situated within the endemic centres of Liverpool CB and Manchester CB and their surrounding metropolitan hinterlands (see Fig. 4.12). This is also found to be the case regarding Pennine Lancashire, in districts neighbouring large towns such as Burnley CB and Blackburn CB, and in South Ribble and Central Lancashire, on rural periphery of the north-western edge of Greater Manchester. Higher incidence rates are also visible on the coastline of North Lancashire (Fig. 4.12), in the resort town districts of Blackpool CB, Lytham St Anne's MB and Fleetwood MB. Areas of the Lancashire region with noticeably low rates of scarlet fever incidence include districts in the Furness Peninsula and West Lancashire. This pattern is most prominent in the first two time–windows (Fig. 4.12).

By time–window three, scarlet fever incidence has already begun to fall considerably, with falling incidence rates in significant portions of the hinterland surrounding Manchester, in Central Lancashire and Pennine Lancashire. The fall in incidence was more gradual in Merseyside and in South Lancashire districts situated between the metropolitan areas of the two conurbations. Higher rates of scarlet fever incidence also persist in Central Lancashire, such as in the South Ribble area surrounding Preston CB (Fig. 4.12).

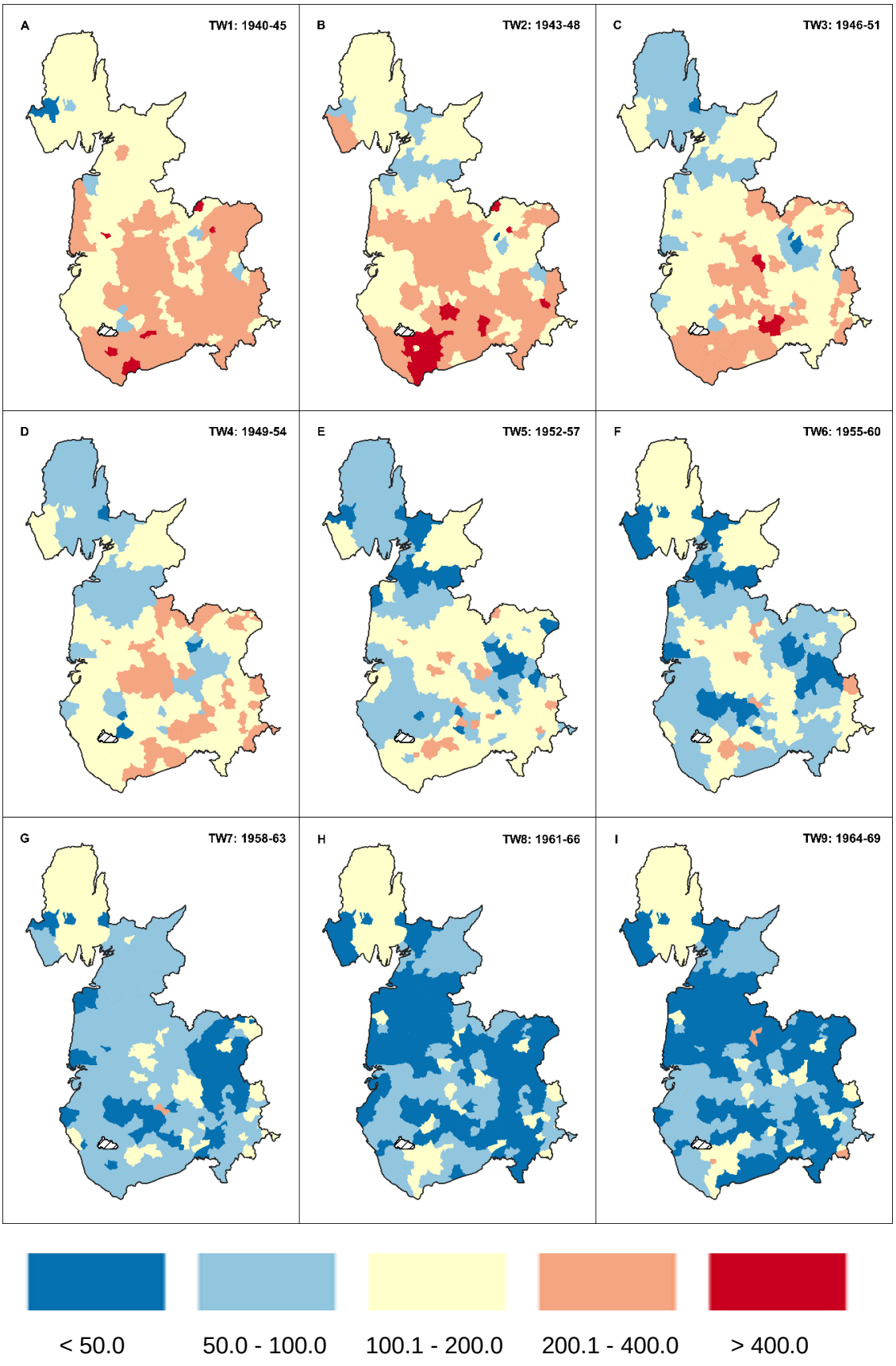

Figure 4.12 *Mean scarlet fever incidence (cases per 100,00 population) in the Lancashire region across nine 72-month time–windows, by local government district, 1940–1969.*

As the study period progresses across successive time–windows, the retreat of scarlet fever infection continues to gather pace; the incidence rate falls from approximately ~200 cases per 100,000 population for the majority of districts located in the densely populated, urbanised southern portion of the region in the first two time–windows to below 100 cases per 100,000 population by time–window seven (1958–63). Districts with an incidence rate below 50 cases per 100,000 population are located within the Fylde area of North Lancashire, the Furness Peninsula, Pennine Lancashire, and pockets around Central Lancashire and the Greater Manchester metropolitan area.

By the final time–window, scarlet fever incidence is concentrated within the endemic centres of Liverpool CB and Manchester CB, and there are isolated pockets of infection dotted across the region, specifically ‘gateway’ districts located on the borders of the region, such as Littleborough UD in East Lancashire and Ulverston/North Lonsdale RD in North-west Lancashire, on the edge of the Furness peninsula. More persistent areas of scarlet fever infection such as Burnley CB, Blackpool CB, Chorley UD in Central Lancashire, as well as St Helen CB and Whiston RD in Merseyside all experience a significant drop in scarlet fever incidence by the end of the study period

4.5 Incidence Mapping: South Wales Region

4.5.1 Measles

Districts in South Wales with the highest incidence rates of measles infection (greater than 2,000 cases per 100,000 population) represent local epidemic outbreaks or persistent measles hotspots, with consistently high notification rates across successive time–window (see Fig. 4.13). These tend to be market towns and isolated urban areas within largely rural, sparsely populated or peripheral areas of the county in which they are located. For instance,

Llandeilo UD and Llandovery MB in north-east Carmarthenshire are two such market town districts that are relatively remote yet harbour substantial incidence rates. Other market towns with rates above 2,000 cases per 100,000 include Ammanford UD on the Carmarthenshire-Glamorgan border, Kidwelly MB in South Carmarthenshire, and the market towns of Pembroke MB and Haverfordwest MB in Pembrokeshire. The coastal towns settlements of Penarth UD and Porthcawl UD in Glamorgan, urban districts neighbouring the Glamorgan–Monmouth border in the eastern portion of the valleys (Risca UD, Bedwas & Machen UD and Rhymney UD) and the market towns of Usk UD, Monmouth MB and Caerleon UD in Monmouthshire also register very high incidence rates across the study period (Fig. 4.13). This may be due to the high intrinsic birth rates in the districts, combined with relatively high population densities and migration from surrounding districts or further afield helping to fuel significant, if sporadic, outbreaks which result in elevated notification rates across successive time–window.

Outside of the relatively densely populations located in market towns or mining communities in the Valleys, large portions of the South Wales region observe very low rates of measles notifications (Fig. 4.13), most of which are below 375 cases per 100,000 population. This reflects the primarily rural and often remote nature of much of the region, particularly the counties of Pembrokeshire and Carmarthenshire, which are largely pastoral and with large, sparsely populated rural districts. High levels of population dispersion coupled with small population sizes, which tend to be scattered among hamlets and small villages, results in the limited transmission of infection and very low rates of susceptible recruitment. This results in long periods of disease fadeout, and as a consequence, vastly reduce rates of infection, in contrast to the vastly more connected and densely populated market town settlements with urban district or municipal borough status.

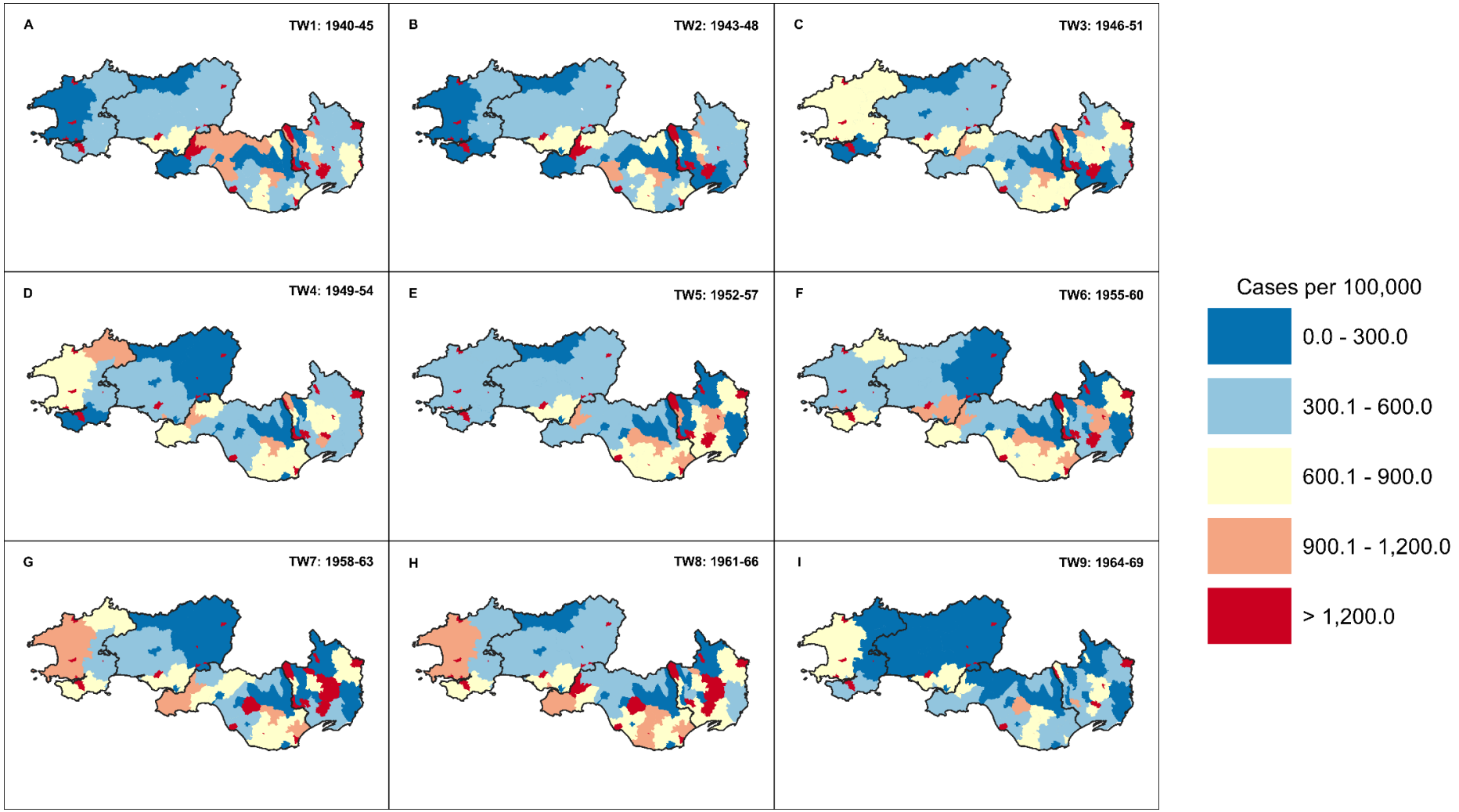

Figure 4.13 *Mean measles incidence (cases per 100,00 population) in the South Wales across nine 72-month time–windows, by local government district, 1940–1969.*

Elevated levels of measles activity shaded in yellow are visible in the districts of in West Glamorgan, surrounding Llanelli MB, Swansea CB, and districts immediately neighbouring Cardiff CB. There is also periodically heightened activity in Pembrokeshire, around the districts of Pembroke and Haverfordwest MBs, particularly in time–window seven and eight. The final time–window (1964-69) witnesses the lowest magnitude of measles cases across the nine time–window, reflecting the introduction of measles vaccine in 1968.The highest incidence rates are found in small urban areas and market towns, reflecting the rapid boom and bust epidemic outbreak dynamics in such communities after the disease is introduced to these districts from outside sources.

4.5.2 Pertussis

Pertussis incidence in South Wales is not spatially concentrated within the county of Glamorgan and the valleys sub-region, as noted for the two other childhood diseases, with 'market town' districts reporting higher rates across the wider region (see Fig. 4.14). From time–window one to six, the general geographical pattern of pertussis incidence is as follows: higher rates across Glamorgan and districts immediately neighbouring the Valley and the Cardiff area (Fig. 4.14). High rates are notable around Pontypool UD and the West Monmouthshire, in the eastern portion of the Valleys. Lower rates in infection are visible consistently in the densely populated and populous district of Rhondda MB in Glamorgan. In the more rural, peripheral and less populous counties of Carmarthenshire and Pembrokeshire, pertussis infection tends to be localised, concentrated within urbanised areas which typically take the form of small market towns (Fig. 4.14), such as the districts of Kidwelly MB, Llandeilo UD and Lllanelli MB, located in South and West Carmarthenshire (Fig. 4.14). However, districts which serve as entry points to the wider region may also feature relatively high rates of pertussis infection. One such example is Cemaes RD in North Pembrokeshire, lying on the county edge neighbouring Ceredigion, a county in mid-Wales situated close to the popular resort area of Cardigan Bay. Key road and railway networks

which connect South Wales with mid-Wales and the major town of Aberystwyth also pass through Cemaes RD.

Notably, districts in Pembrokeshire sees more elevated rates of pertussis incidence compared to Carmarthenshire, despite being a further distance away from the large towns and cities situated in Glamorgan where pertussis in more endemic (Fig. 4.14). This maybe the result of Pembrokeshire serving as both a key gateway and exit point for the wide region by sea, resulting in significant ferry traffic with the Republic of Ireland and other areas, but also the popularity of the county as a tourist destination for family holidays, in picturesque resort towns such as Tenby MB and Pembroke MB.

Across successive time–windows, the slow retreat of pertussis infection is evident across much of South Wales (Fig. 4.14), with the onset of mass vaccination serving to force the significant reduction in number of cases across all districts in the region, as chains of transmission are severely disrupted or broken. In response, maps for time–window six to nine exhibit a notable transformation in the colour shading of many districts, as the red patches of high incidence fade to yellow and blue as the years with routine vaccination against pertussis begin to mount, and cases fall dramatically in number. A steep decline in pertussis incidence is observed in Carmarthenshire during the vaccine era, with the combination of effective public health interventions and the sparely populated, rural nature of the county serving to eliminate chains of pertussis transmission across urban and rural districts alike. Stubborn, localised pockets of relatively high pertussis activity are notable in districts such as Pembroke MB and Haverfordwest MB in Pembrokeshire, Llanelli MB in South-East Carmarthenshire near the Swansea urban area, Glamorgan. These districts also feature as areas of relatively high scarlet fever incidence in later time–windows during the study period.

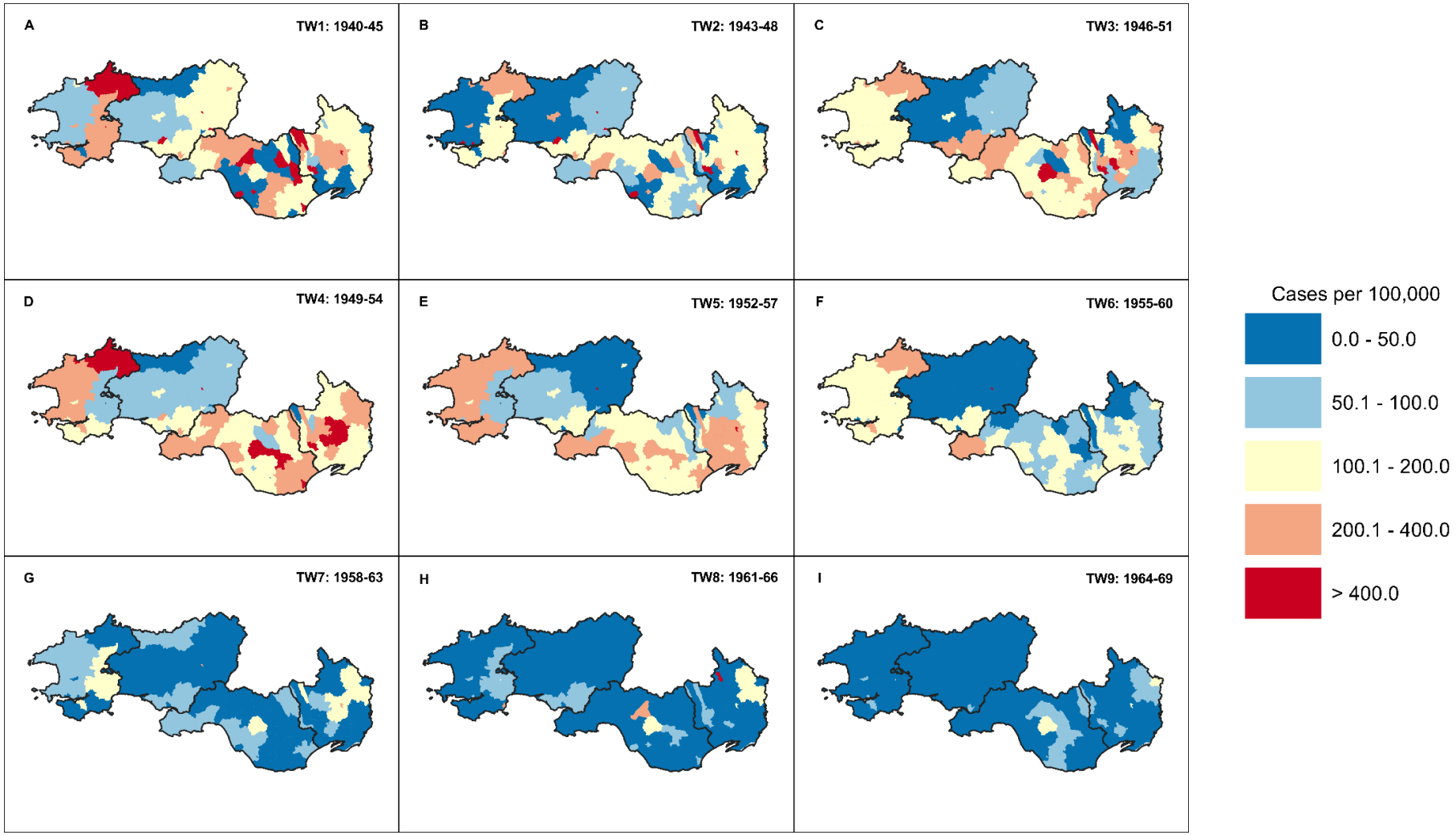

Figure 4.14 *Mean pertussis incidence (cases per 100,00 population) in the South Wales across nine 72-month time–windows, by local government district, 1940–1969.*

4.5.3 Scarlet fever

The most notable trend in scarlet fever incidence is the gradual yet consistent retreat of the disease across the region over successive time–windows (see Fig. 4.15). From time–windows one to three, scarlet fever activity is primarily concentrated within the more dense, populous peri-urbanised valleys of Glamorgan and Monmouthshire directly north of Cardiff, in districts such as Rhondda MB, Caerphilly UD, Pontypridd UD and Mountain Ash UD. Elevated levels of scarlet fever incidence are also found in some of the more rural, sparsely populated districts of Carmarthenshire and Monmouthshire on either side of the Glamorgan border (Fig. 4.15). As the study period progresses, the gradual fall in scarlet fever incidence across the region begins to accelerate by time–window six, with notable drops in scarlet fever cases in urban districts located in the Valleys sub-region, as well as districts situated on the coast of the Vale of Glamorgan. Relatively high levels of incidence are found in the North Pembrokeshire area between time–windows one to six (Fig. 4.15), specifically in Cemaes RD, despite its remoteness and high level of population dispersion. This might be due to the districts location on the route to mid-Wales, the town of Aberystwyth and the popular Cardigan Bay, resulting in significant edge effects and spatial import of infection from the neighbouring county of Ceredigion.

Coinciding with the vaccine-era for pertussis after the introduction of mass vaccination for that disease in 1957, scarlet fever activity continues to fall substantially from time–windows seven to nine (Fig. 4.15), with isolated pockets found most notably in Pembroke MB, which served as a ferry port and thus a gateway to and from the region, Ogmore and Garw UD in central Glamorgan, neighbouring Rhonda MB and slightly higher rates in the nearby Caerphilly and Pontypridd UDs, all districts tightly connected through commuter links by rail and road via Cardiff CB. In Northwest Pembrokeshire, scarlet fever incidence fell dramatically in the final three time–windows, with the vast majority of districts observing an incidence rate of below 50 cases per 100,000 population by time–window nine.

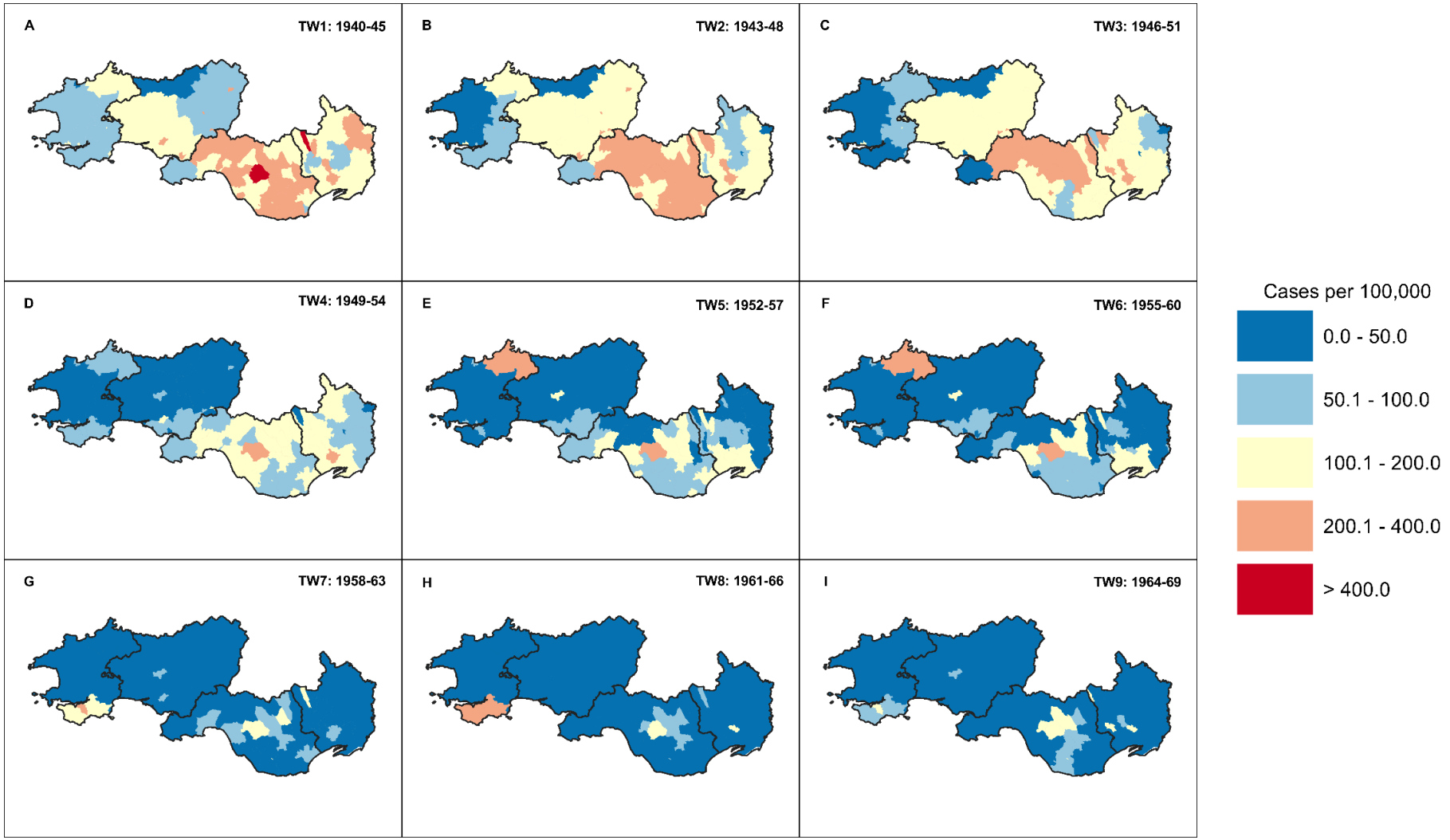

Figure 4.15 *Mean scarlet fever incidence (cases per 100,00 population) in the South Wales across nine 72-month time–windows, by local government district, 1940–1969.*

4.6 Spatial Synchrony: Lancashire Region

4.6.1 Measles

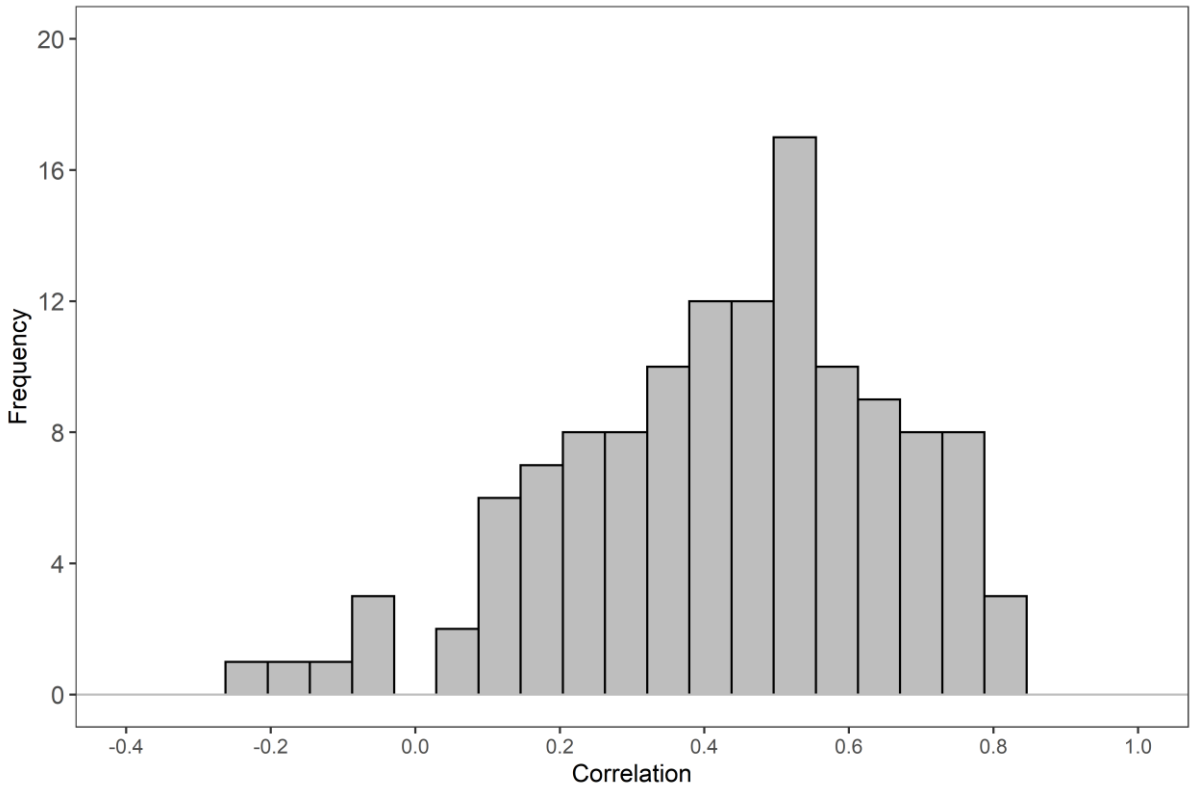

Figure 4.16 *Frequency distribution of the ordinary sample correlation coefficient between annual counts of measles notifications for each Lancashire district and the regional mean over the remaining 124 districts, 1940–1969. The mean average correlation is 0.48, indicating an intermediate level of synchronicity across the region. Most correlation coefficients for individual districts are evenly spread between the values of 0.20 and 0.60.*

Sixty-seven of Lancashire's 125 districts were found to have a correlation coefficient greater than the regional mean of 0.48 (see Fig. 4.16), indicating relatively higher levels of synchronicity. Several districts diverge from the pattern of regional measles activity despite their close proximity to endemic centres tightly correlated with the regional disease pattern. For instance, Billinge and Winstanley UD possesses an extremely weak correlation with the

regional dynamics at 0.07 while St Helens CB, which neighbours Billinge & Winstanley UD, has a relatively high correlation of 0.74. The largest sample correlation coefficient was found in Farnworth MB (0.83), located approximately 3.7km southeast of Bolton CB, the latter of which also recording a very strong positive correlation (0.82). Other districts which indicate a strong correlation with the regional average include Salford CB (0.81), Manchester CB (0.77), and Royton UD (0.77) (see Fig. 4.17). The analysis of spatial synchrony in measles activity in Lancashire indicates a high degree of synchronicity in districts of varying population sizes centred around Bolton CB, which forms part of Greater Manchester, with similar activity located in surrounding satellite settlements and in Manchester and Salford CBs with which Bolton has strong economic ties.

In great contrast to Manchester CB and districts situated within the metropolitan hinterland of Greater Manchester, Liverpool CB was found to have a much weaker correlation with the regional average (Fig. 4.17), standing at 0.41. Liverpool CB is unusual among the more populous districts in the Lancashire region since successive major and minor epidemics were of almost the same height during the 1940s and 1950s. This difference has previously been attributed to Liverpool's relatively high birth rate and resultant rapid replacement of susceptibles after major epidemics (Grenfell et al., 1995; Grenfell et al., 2001), and goes some way to explaining why measles dynamics in Liverpool CB are notably less correlated with the regional disease pattern.

It is therefore not so much the shape of the main epidemic but the pattern of intervening minor epidemics which varies among the cities. Differences between Liverpool and Manchester CBs in their phase relationship and epidemic behaviours reveals even geographically close communities of similar population sizes can have a completely different pattern during the minor and major epidemic years, due to within community disease dynamics and variations in levels of susceptible recruitment.

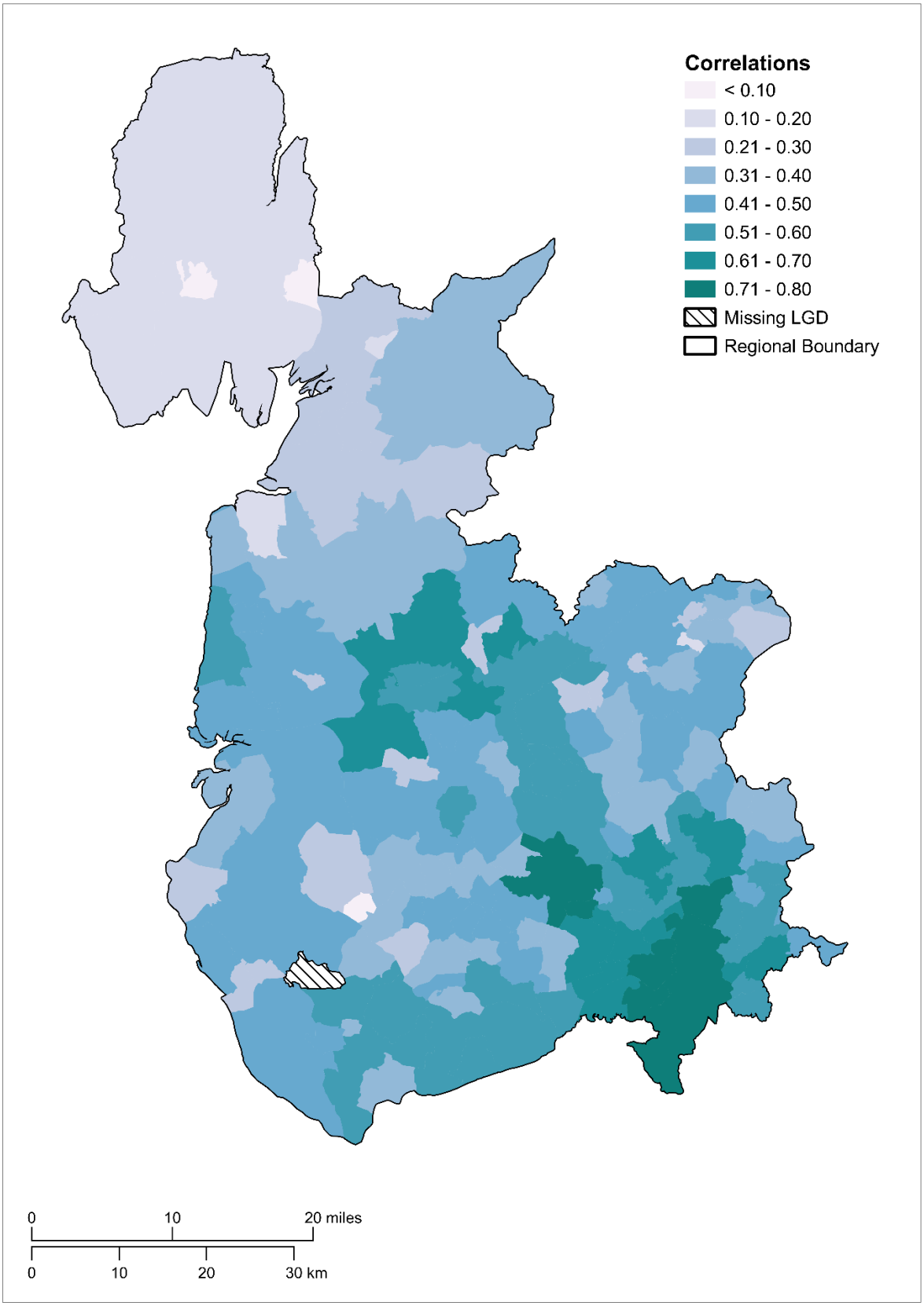

Figure 4.17 *Geographical distribution of the ordinary sample correlation coefficient between annual counts of measles notifications for individual Lancashire districts and the regional average over the remaining 124 districts, 1940–1969.*

Skelmersdale UD, on the edge of the Merseyside area north-east of Liverpool CB, has a notable negative correlation (-0.25), indicating a strongly asynchronous, divergent pattern of measles epidemics during trough periods for surrounding districts. As discussed in Section 3.10, the huge population growth experienced in Skelmersdale UD during the 1950s and 1960s, with significant numbers arriving primarily from the overspill population of Liverpool CBs all-year round, substantial growth in susceptible recruitment year-on-year accompanied by significant levels of commuting to neighbouring population centres and centres of employment created a unique stochastic dynamic resulting in a heavily localised pattern of measles activity without parallel among districts in the rest of the region. Apart from Skelmersdale UD, five other districts have a negative correlation. Two of these districts neighbour Preston CB (Fig. 4.17); Kirkham UD (-0.16) and Longridge UD (-0.10). These districts represent small, more remote urban communities which, due to their rurality and subsequent lower levels of connectivity and low population size facilitates the isolation of the local disease activity found in these districts and average pattern of measles activity found across the regional metapopulation.

The other three districts with negative correlations districts are found in Northwest Lancashire (Fig. 4.17), notably small settlements in sparsely populated rural areas representing some of the most remote communities in the region. These are Grange UD (0.07) and Ulverston UD (0.09). These local populations find themselves cut-off from the greater regional metapopulation dynamics, acting as 'islands' adrift from the 'mainland' metapopulation where disease activity is primarily concentrated within and surrounding Manchester and Liverpool CBs. The high degree of isolation, sparse population and lack of connectivity leads to highly irregular disease introduction events and significant spacing between epidemic outbreaks, resulting in a pattern of disease incidence highly divergent from the pattern found in more densely populated and better connected local populations in the more heavily urbanised areas of Lancashire.

4.6.2 Pertussis

The mean sample correlation coefficient is 0.61, indicating a relatively high degree of synchronicity across the region's districts (see Fig. 4.18), and a higher level than that found in relation to measles activity. Over half of Lancashire's districts have a correlation coefficient greater than the regional average with strong synchronicity in disease patterns at a local level found across the region. Examples include Bootle CB, South-west Lancashire (0.76), Burnley CB, North-east Lancashire (0.78), Ulverston/North Lonsdale RD, North-west Lancashire (0.74), and Salford, South-east Lancashire (0.88).

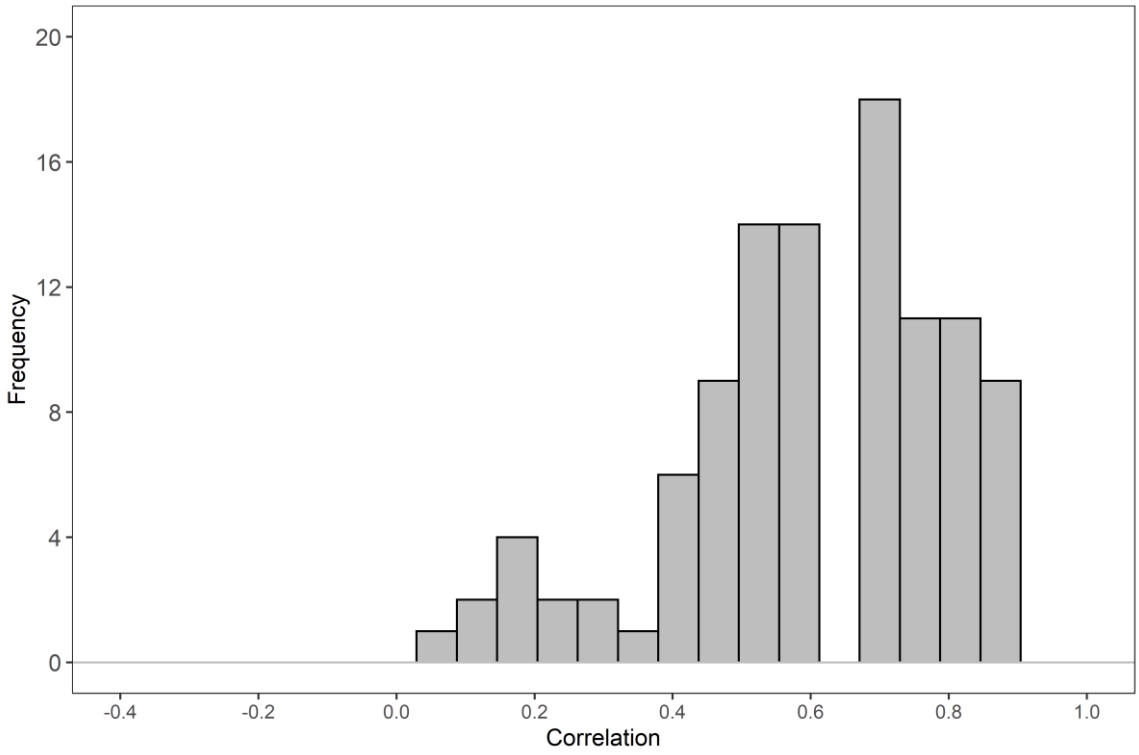

Figure 4.18 *Frequency distribution of the ordinary sample correlation coefficient between annual counts of pertussis notifications for each Lancashire district and the mean average over the remaining 124 districts, 1940–1969. The results indicate the majority of districts experienced a high degree of synchronicity with the regional disease pattern, with 71 districts across the region having a correlation coefficient greater than the regional mean average (0.61).*

Significant positive correlations exist between large towns, cites and the regional average, despite the extensive distance between some of these settlements and notable differences in population size. For instance, Manchester CB (~650,000) has a correlation of 0.89 while Preston CB (~100,000) has a correlation of 0.87, despite the two districts being separated by a distance of 56km. Overall, there is a high degree of correlation among major towns including Salford CB, Rochdale CB (0.86), St Helens CB (0.85) and Warrington CB (0.84) (see Fig. 4.19). The strongest correlation with the regional disease pattern is found in Manchester CB and the neighbouring district of Droylsden UD (Fig. 4.19), both sharing a correlation of 0.89. This indicates the centre of regional disease activity appears to be located in Manchester and its surrounding metropolitan area.

Unlike with measles activity, the demographic stochasticity within Liverpool CB does not result in an asynchronous cycle of pertussis epidemic activity, with the district registering a strong correlation of 0.86. This suggests a regular pattern of disease waves and epidemic activity common to both Liverpool CB and districts within and surrounding Greater Manchester (Fig. 4.19). None of the 125 districts in Lancashire were found to have a negative correlation (Fig. 4.19). The weakest sample correlation coefficient in relation to the regional average is found in Kirkham UD (0.04), located midway between Blackpool and Preston, while Longridge and Skelmersdale UDs were found to have a slightly higher, yet nevertheless weak, correlation of 0.14. Both Kirkham and Longridge UDs are small urban settlements with populations below 10,000, located within the predominantly rural areas of North and Central Lancashire (Fig. 4.19). The lack of connectivity of these small towns with the wider metapopulation dynamics of the region and nearby large towns alongside low population sizes may explain the lack of synchronicity with the regional disease pattern, with a higher number of fadeouts, of longer duration, and asynchronous outbreaks. The weak correlation found in Skelmersdale UD is expected given the demographic upheaval experienced during the study period, with significant numbers of Liverpool's overspill population migrating to the Skelmersdale new town in the late 1950s and 1960s following

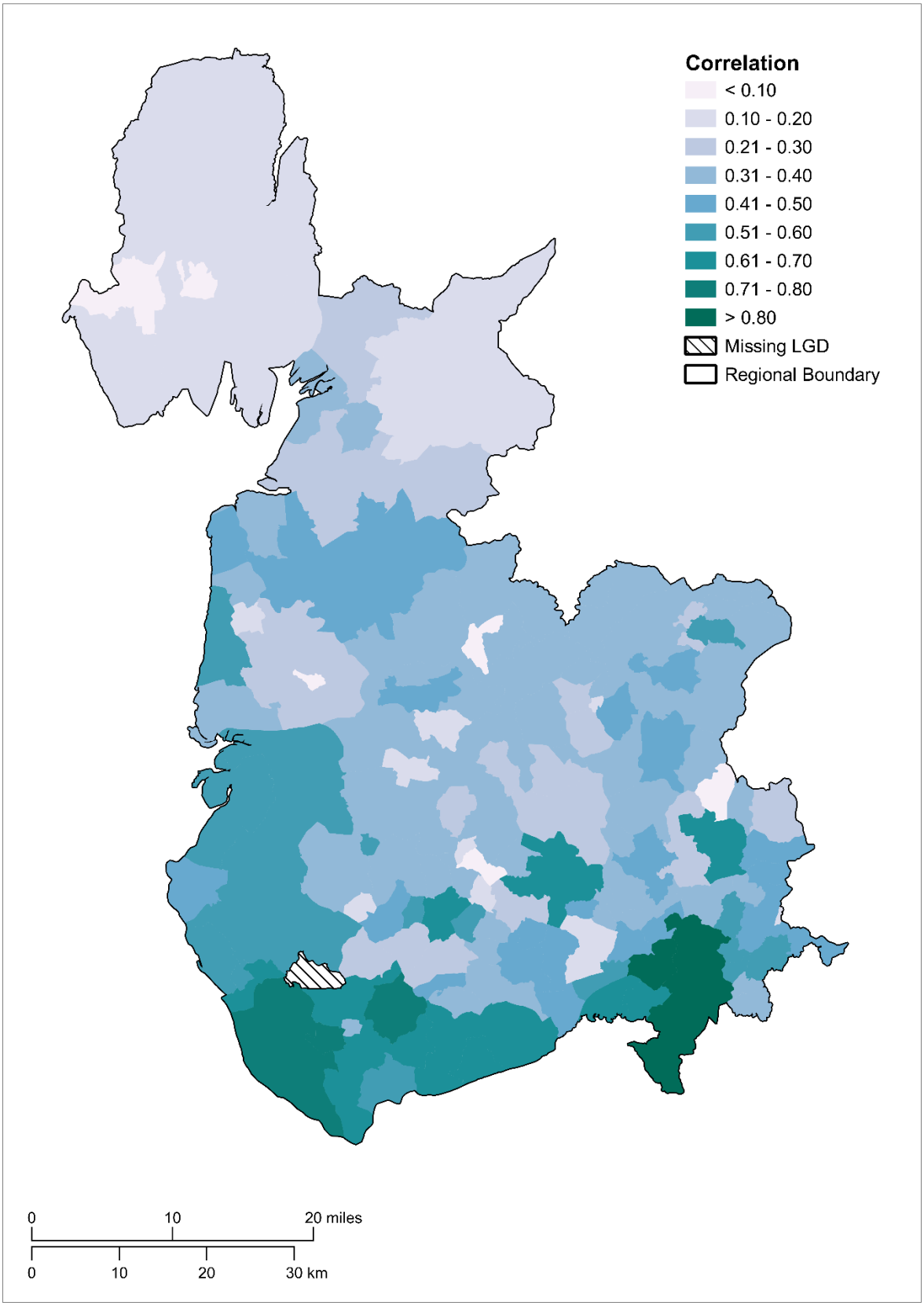

Figure 4.19 *Geographical distribution of the ordinary sample correlation coefficient between annual counts of pertussis notifications for individual Lancashire districts and the regional average over the remaining 124 districts, 1940–1969.*

the extensive slum clearance in the city, coupled with the high extrinsic variation in birth rates during the baby boom period. Such trends nurture highly asynchronous, localised disease activity, and drag Skelmersdale UD away from the regional disease pattern. Other weak correlations are found somewhat unexpectedly in districts situated in the Greater Manchester metropolitan area, such as Aspull UD (0.19) and Worsley UD (0.20), satellite settlements with small population sizes insufficient to maintain the chains of transmission for the disease. It is unlikely that distance is a causal factor in asynchronous outbreaks, since Worsley UD is situated only 9.25km from Manchester CB, while Aspull UD is located 5.5km from Wigan CB, which has a strong correlation (0.66) with the regional disease pattern.

In the rural periphery of the region in north-west, weak correlations are detected in multiple districts including Dalton-in-Furness UD (0.27), Poulton-le-Fylde UD (0.25) and Carnforth UD (0.20). These districts are characterised by highly dispersed, small populations surrounded by extensive rural areas. They are also remote communities, far removed from major industrial towns of South Lancashire, and the centre of the disease activity. These factors combine to ensure the less populous districts of Northwest Lancashire are only weakly correlated with the regional disease pattern.

4.6.3 Scarlet fever

The mean sample correlation coefficient between the annual counts of reported cases of each district and the remaining 124 districts is 0.60, indicating a relatively high degree of synchronicity in regional disease activity (see Fig. 4.20). Strong correlations exist between the most populous districts featuring large towns and cities across the Lancashire region and the average regional pattern of disease activity.

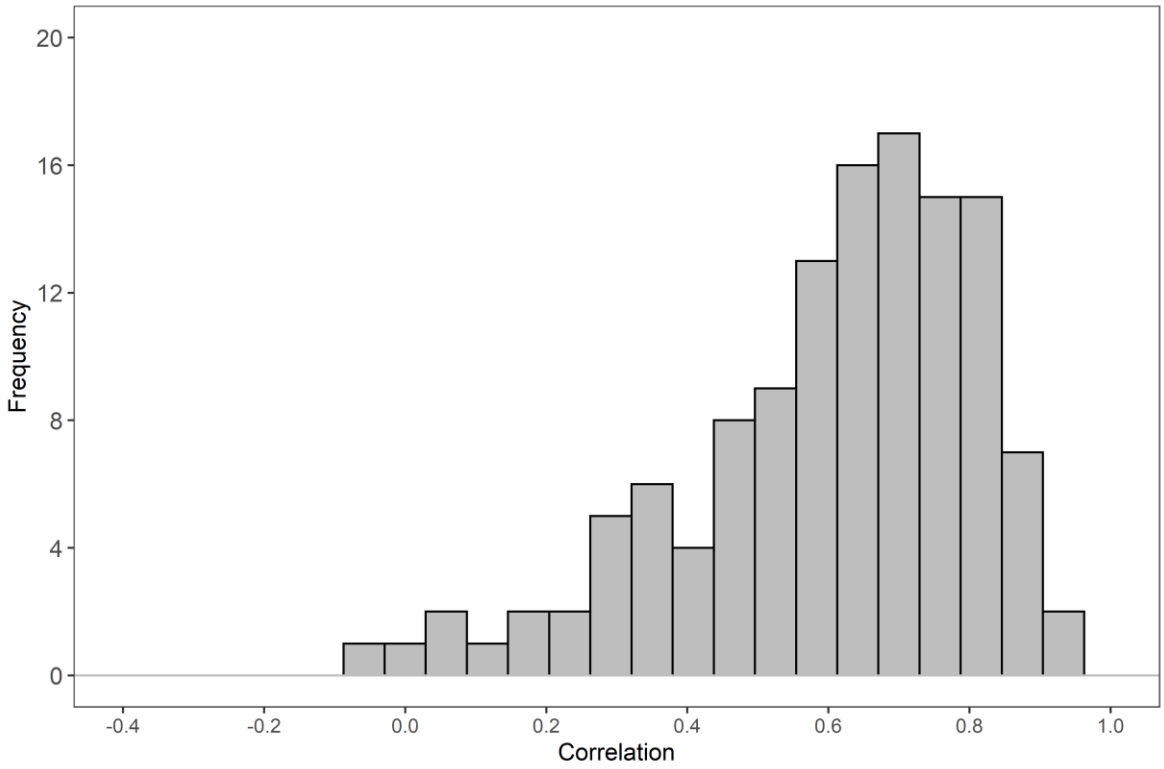

Figure 4.20 *Frequency distribution of the ordinary sample correlation coefficient between annual counts of scarlet fever notifications for each Lancashire district and the mean average over the remaining 124 districts, 1940–1969. The results indicate a large proportion of districts experience a high degree of synchronicity with the regional disease pattern, with 73 districts registering a correlation coefficient higher than the mean average for the region*.

For instance, Blackpool and Oldham CBs both have a correlation of 0.87 despite being separated by a distance of ~92km (see Fig. 4.21). However, both districts have significant spatial interaction with the endemic centre of Manchester CB. The strongest correlations among districts in Lancashire suggests that the regional epidemic centre of activity lies within the Greater Manchester area (Fig. 4.21), with Salford CB reporting the strongest correlation (0.93) closely followed by Manchester CB (0.92) and the much less populous suburban district of Prestwich UD (0.87), the latter of which is situated only 5km north of Salford CB and 5.3km north of Manchester CB**.** Neighbouring the aforementioned Oldham district, Chadderton UD located one mile west of Oldham and situated between the town

and Manchester CB, has the third strongest correlation in the region standing at 0.90. This indicates a corridor of disease spread westwards from Salford and Manchester CBs to the metropolitan hinterland. Other strong correlations among major towns and population centres include Liverpool CB (0.86), Rochdale CB (0.85), as well as smaller coastal towns within the Fylde area in the north-west of the region (Fig. 4.21), including Lytham St Anne's MB (0.84) and Morecambe MB (0.86) further north. The strong correlations in the various popular coastal resort towns located in the Fylde area and Blackpool CB indicates a high degree of synchronicity with other large towns further inland in the industrial heartlands of the Lancashire region, with disease patterns influenced by the repeated introductions of disease from visitors, commonly families with young children, from those industrial towns.

Only one district was found to have a negative correlation: Skelmersdale UD (-0.06). Notably, this district is the only one of the 125 districts in the Lancashire region that reports a consistently negative correlation with the mean regional disease pattern for all three childhood infections (see Sections 4.6.1 and 4.6.1). This is indicative of highly asynchronous, localised disease activity, nurtured by a number of factors: rapid urban expansion and explosive population growth in the 1950s and 1960s, with Skelmersdale's designation as a second wave new town for Liverpool CB and the wider North Merseyside conurbation's overspill population, the high birth rates associated with the baby boom during this period, and consistent populations flows to nearby towns and cities to which inhabitants of Skelmersdale regularly commute These include Wigan and Southport CBs, as well as Liverpool CB. Together, these factors ensured almost constant asynchronous localised disease activity in Skelmersdale UD, and the district's regular departure from the regional mean.

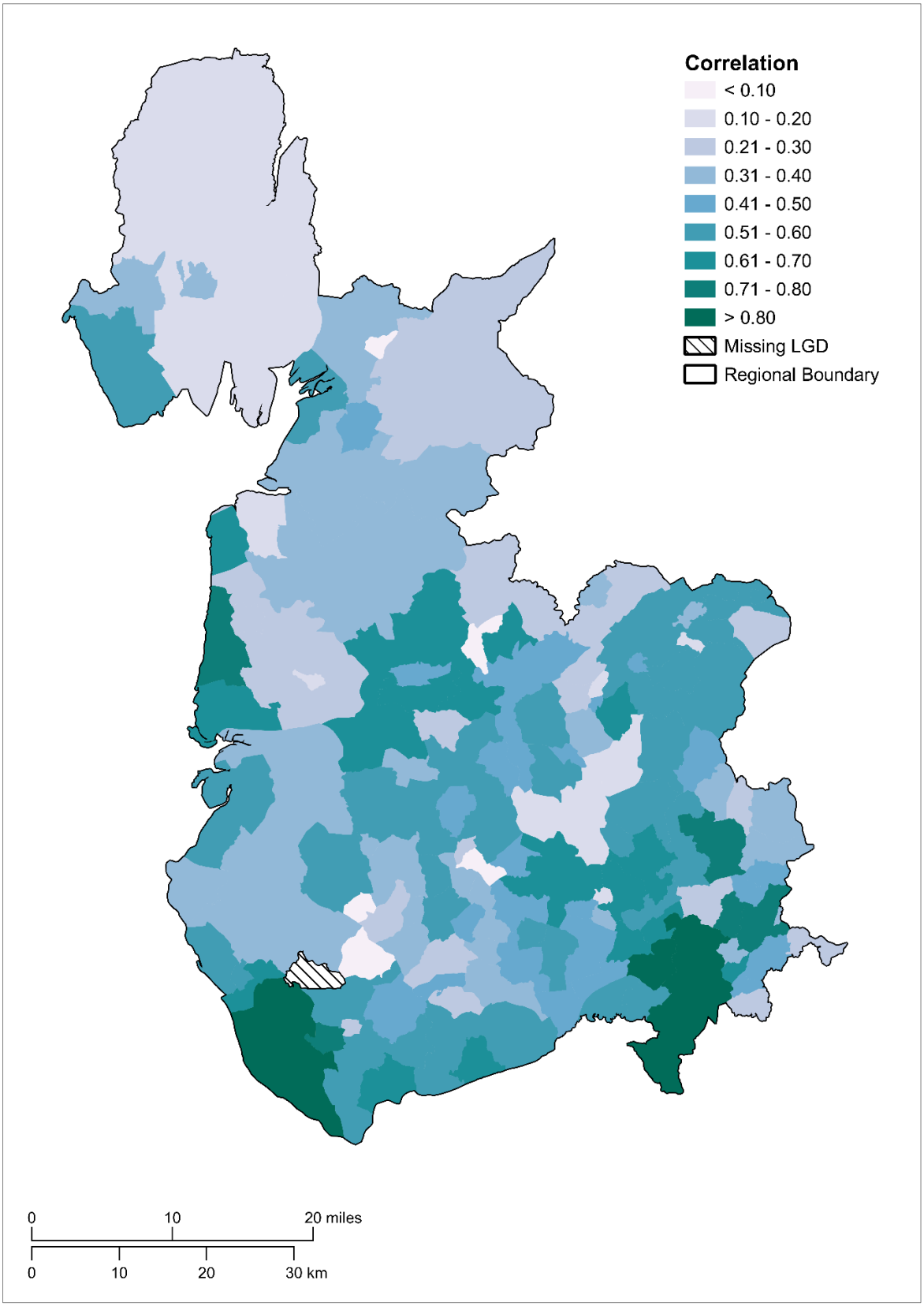

Figure 4.21 *Geographical distribution of the ordinary sample correlation coefficient between annual counts of scarlet fever notifications for individual Lancashire districts and the regional average over the remaining 124 districts, 1940–1969.*

Despite their close proximity to Liverpool and Manchester CBs respectively, the districts of Rainford UD (0.07) and Blackrod UD (0.08) have virtually no correlation with the regional disease pattern observed in major towns and other less populous districts located within the Merseyside and Greater Manchester conurbations (Fig. 4.21). This may be partly explained by their small population sizes, as villages located on the very edge of the urban periphery, both districts lack a susceptible population of sufficient size to maintain the pattern of disease activity observed in more populous nearby settlements, highly prone to regular fadeout events. The proximity of Rainford UD and Blackrod UD to multiple large towns also increases the likelihood of regular disease introduction from neighbouring areas in different phases of disease activity, fuelling asynchronous disease activity. For instance, Blackrod UD is closely situated to the populous districts of Wigan CB (6.3km) and Bolton CB (10.6km), while Rainford UD lies 5.6km north of S Helens CB and 11.3km from Wigan CB.

Other districts in Lancashire which have notably weak correlations include Ulverston/North Lonsdale RD (0.07) and Carnforth UD (0.14). These districts are located in the region's periphery, in the remote and sparely populated north-west of the region. With the limited connectivity of these districts coupled with highly dispersed local populations, an irregular disease pattern out of sync with the wider regional metapopulation dynamics are expected. Intriguingly, Middleton MB (0.27) has a low positive correlation despite its proximity to Manchester CB (Fig. 4.21), with the two districts sharing boundaries and Middleton serving as a major suburb of Manchester. Since Middleton MB is equidistant between Manchester and Rochdale CBs, situated approximately 8km from the centre of each district, the high degree of connectivity with both may result in a significantly elevated rate of re-introduction events. This would fuel more irregular minor epidemic outbreaks and overall disease pattern when coupled with the disease activity driven by intrinsic stochasticity. Notably, there are smaller local urban populations within the Greater Manchester conurbation which possess much stronger correlations with the regional pattern of scarlet fever activity than Middleton MB, such as Huyton-with-Roby UD (0.73) and Chadderton UD (0.90).

4.7 Spatial Synchrony: South Wales Region

4.7.1 Measles

The average mean correlation is 0.56 indicating a relatively high degree of synchronicity in measles incidence between districts and the mean regional pattern (see Fig. 4.22). This indicates a notably higher level of synchronicity in measles activity in South Wales compared to the Lancashire region.

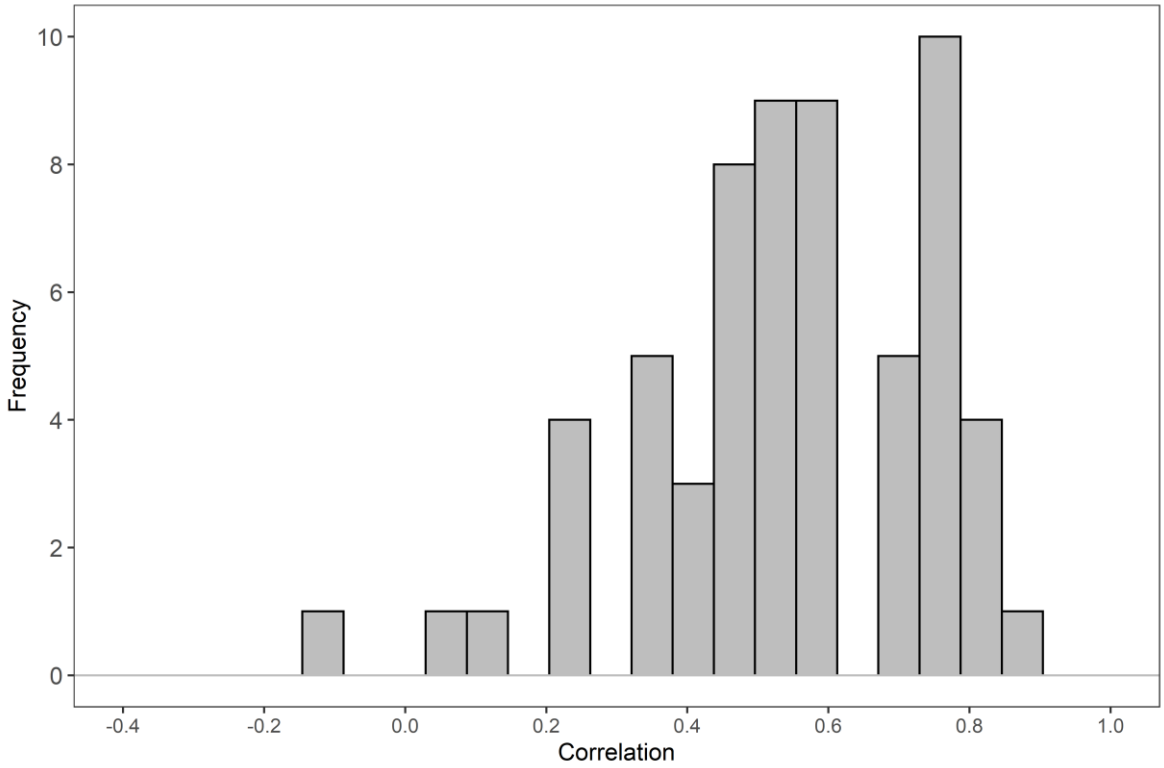

Figure 4.22 *Frequency distribution of the ordinary sample correlation coefficient between annual counts of measles notifications for each South Wales district and the average over the remaining 74 districts, 1940–1969. The histogram indicates the majority of districts with a positive correlation above the mean calculated across all districts. Mean sample correlation coefficient = 0.56.*

Districts most strongly correlated with the average pattern of measles activity in South Wales tend to be concentrated in the urban districts neighbouring and in close proximity to

Cardiff CB (see Fig. 4.23), particularly to the northeast of Cardiff in the mining communities lying In the Rhymney Valley, on the Glamorgan–Monmouthshire county border close to the towns of Caerphilly and Newport. Gelligaer UD, located in the Rhymney River valley, 9.65km north of Caerphilly UD, has the strongest positive correlation of the 74 regional districts, at 0.85, indicating a potential epicentre of measles activity. The district is closely followed by Newport CB (0.79), Caerphilly UD (0.79), Cardiff RD (0.79) and Cwmbran UD (0.75). Many of the high positive correlations with the mean regional pattern of measles activity are among districts situated within the South Wales valleys. These include Mountain Ash UD (0.75), Ogmore and Garw UD (0.77), Rhondda MB (0.74) and Merthyr Tydfil CB (0.73). Cardiff CB also exhibits a strong correlation at 0.74, compared to a slightly lower correlation in Swansea CB (0.69). Intriguingly, Carmarthen RD also boasts a very strong positive correlation (0.84), second only to Gelligaer UD. This could indicate a second potential source of measles incidence or a potential outlier, with the district brought into greater synchrony with the average regionals pattern due to the sheer number of timely reintroduction events from the more heavily urbanised, neighbouring districts in south-east Carmarthenshire and south-west Glamorgan and the major towns of Llanelli and Swansea.

Other districts with low positive correlations include Tredegar UD (0.24), Blaenavon UD (0.26) and Porthcawl UD (0.33). All three districts are located close to key population centres such as Rhonda MB, Merthyr Tydfil MB and Cardiff CB (see Fig. 4.23). Both Tredegar and Blaenavon are located at the head of their valleys, the Ebbw and Llywd valleys respectively. Travel in valley communities is often linear in fashion and constrained by topographical obstacles, thereby limiting connectivity between settlements across larger geographical spaces, with movement often constrained to travelling northwards or southwards. For settlements at the head of valleys, such as Tredegar and Blaenavon, there is a greater degree of isolation with population flows from other key local populations limited. This phenomenon may isolate communities in valleys from heightened disease activity found in other valleys and the coastal plain where corridors of infection may be found.

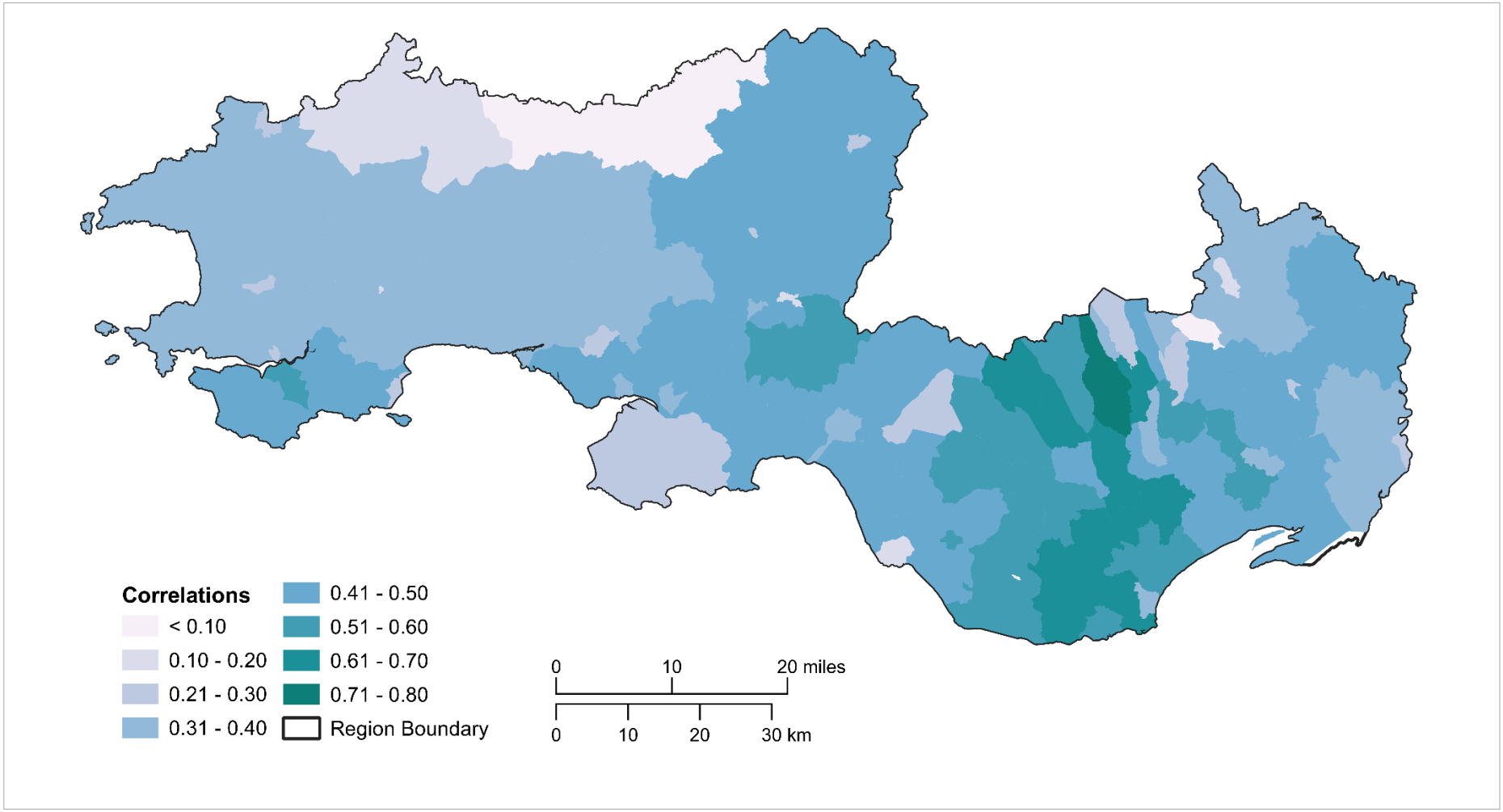

Figure 4.23 *Geographical distribution of the ordinary sample correlation coefficient between annual counts of measles notifications for individual districts in South Wales and the regional average over the remaining 74 districts, 1940–1969.*

However, the smaller local population sizes coupled with limited population flows and intrinsic variation in susceptible recruitment are likely to significant factors in reduced correlations with the regional disease pattern compared to more populous and densely populated valley towns and districts, such as Merthyr Tydfil and Caerphilly UD.

No district in South Wales has a negative correlation with the average regional disease pattern (Fig. 4.22). The weakest correlation was detected in Newcastle Emlyn RD (0.04), a very remote, rural district on the edge of north Carmarthenshire and the wider regional metapopulation as whole (Fig. 4.23). The population size is low and highly dispersed, resulting in limited susceptible recruitment with chains of transmission within the district highly vulnerable to collapse resulting in long periods of fadeout without consistent introduction of measles from external sources and neighbouring districts. Cowbridge MB, despite its close proximity to Cardiff CB only 19.3km away, was also found to have an extremely low correlation (0.09). One may expect the disease pattern in Cowbridge MB to be similar to that found in Cardiff CB, with a similar strength of correlation with the overall regional disease pattern. Yet the disease pattern in Cowbridge is highly affected by the district's extremely small population (~4,000) coupled with its location within the more pastoral areas of the Vale of Glamorgan, where the population is more highly dispersed in rural areas and suffers from poor connectivity with Cardiff. For instance, Cowbridge does not have a railway link with the city. These factors result in highly frequent fadeouts of measles and limited reintroduction events, with disease activity highly localised due to the poor accessibility of the local population. Due to the very small population size and low density, chains of transmission are not supported, with low transmission probability amplified by very limited susceptible recruitment.

Districts detected with low positive correlations include Fishguard & Goodwick UD (0.26), a small local population located on the western edge of the South Wales region in Pembrokeshire, surrounded by sparsely populated rural districts and lying a significant

distance away from the major population centres. Another remote district featuring a low correlation with the region disease pattern is Llandovery MB (0.34), an insignificant population of around 2,000, located in the high rural, thinly populated rural area of North-east Carmarthenshire. In Monmouthshire, the market town of Abergavenny MB has a correlation of 0.23, despite being much more closely located to the more significantly populated settlements in Glamorgan, and the major population centre of Newport CB located in South Monmouthshire. The limited correlation between Abergavenny MB and the regional disease pattern may reflect the asynchronous outbreaks associated with minor populations unable to support chains of transmission without introduction of disease from larger population centres. Additionally, such reintroduction events are constrained by the limited connectivity and remoteness of the predominantly rural, sparsely populated area of North Monmouthshire in which Abergavenny MB is situated.

4.7.2 Pertussis

The average correlation is 0.54 indicating a relatively high degree of synchronicity in pertussis incidence/disease patterns between districts and the average pattern in South Wales. Of the 75 districts in South Wales, 34 districts have larger correlation coefficients than the regional average (see Fig. 4.24).

The strongest correlations with the average among districts in South Wales are found in Cardiff CB (0.81) and Swansea CB (0.79). Strong positive correlations are also found in less populated districts surrounding the urban centres (see Fig. 4.25). For instance, Neath RD (0.72) and Neath MB (0.71) which lie 9.65km north-east of the centre of Swansea CB, and Pontardawe RD (0.71) neighbouring Swansea CB 16.1km to the north all display strong positive correlations.

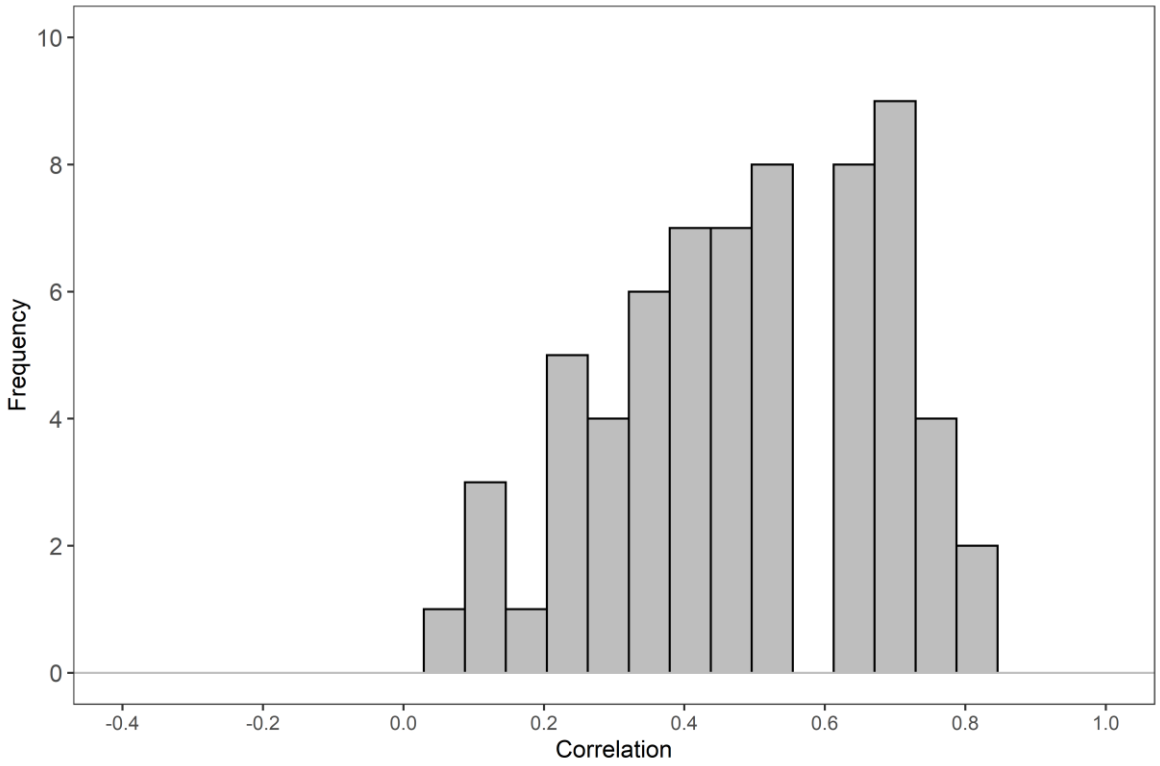

Figure 4.24 *Frequency distribution of the ordinary sample correlation coefficient between annual counts of pertussis notifications for each South Wales district and the average over the remaining 74 districts, 1940–1969. Mean sample correlation coefficient = 0.54.*

Significant positive correlations are also found in the more urbanised districts of Carmarthenshire (Fig. 4.25), located in the southeast of the county, within 24km of Swansea CB, such as Llanelli MB (0.69) and Llanelli RD (0.67). Carmarthen RD has a strong positive correlation of 0.76, despite being very sparely populated and the population within districts being high dispersed, spread across numerous small villages and hamlets. This may be due to the spread of disease from the multiple districts, with more substantial urbanised populations which neighbour the district, and its close proximity to Swansea CB and its wider metropolitan area.

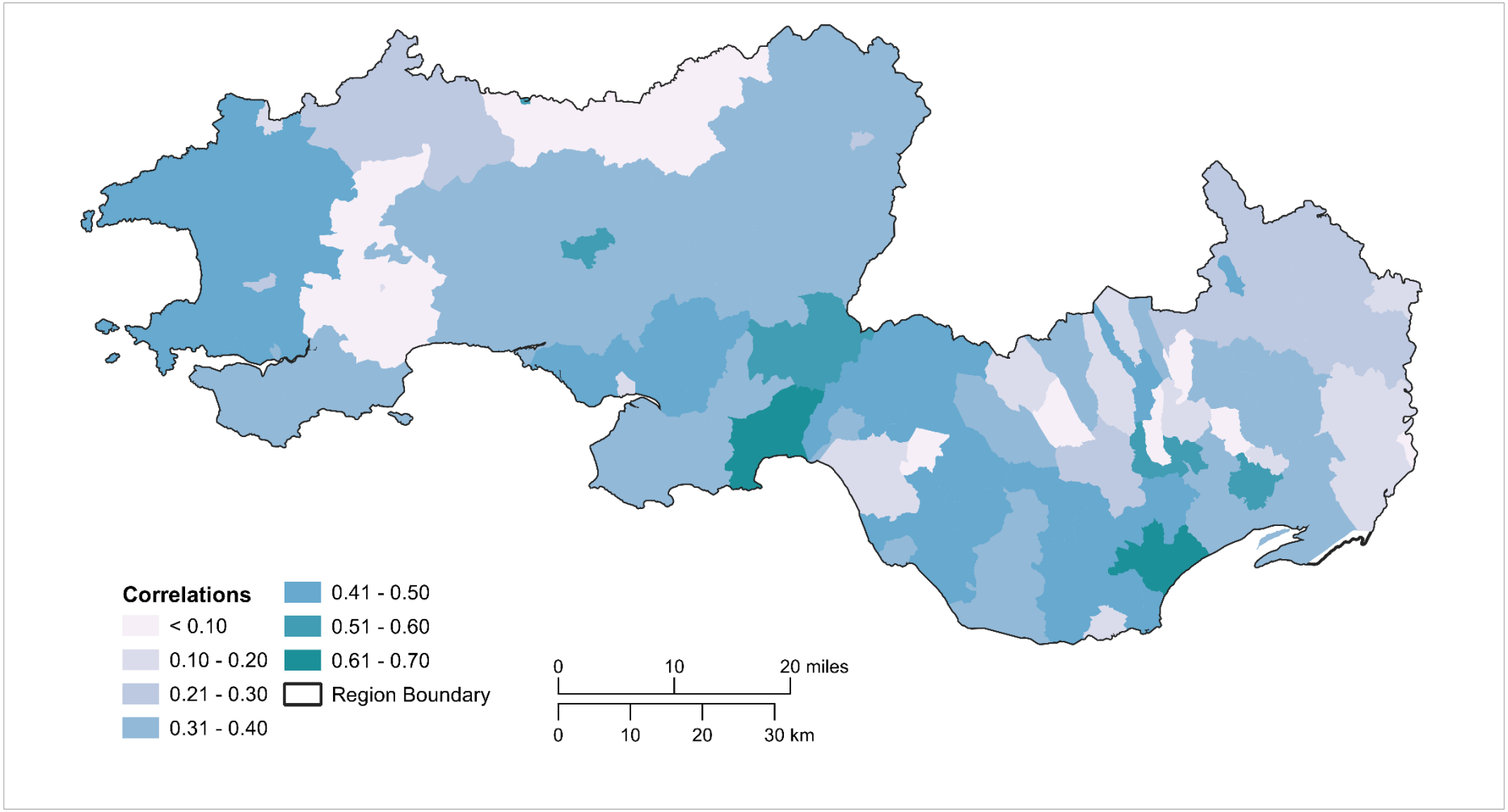

Figure 4.25 *Geographical distribution of the ordinary sample correlation coefficient between annual counts of pertussis notifications for individual districts in South Wales and the regional average over the remaining 74 districts, 1940–1969.*

Pontypridd UD (0.76), Abercarn UD (0.74), Risca UD (0.73), all closely neighbouring districts located in the south-eastern edge of the Valleys area are found to have relatively strong correlations with the average disease pattern in the region (Fig. 4.25). However, in contrast to measles, pertussis incidence in many districts in the Valleys area ae weakly correlated with the average. These include Rhondda MB (0.29), Caerphilly UD (0.29), Aberdare UD (0.25) and Gelligaer UD (0.37). The weak correlations in Rhondda MB and Caerphilly UD are of a surprise given their proximity and strong transport links with Cardiff CB as well as their status as key local population centres in the valleys, especially the heavily and densely populated Rhonda MB, which is only behind Cardiff and Swansea CBs in terms of population size, susceptible input and density. This finding suggests the impact of regular reintroductions resulting in minor epidemics and asynchronous disease activity, with infection being transmitted from Cardiff CB as well as surrounding local population centres.

None of the 75 districts had a negative correlation (Fig. 4.24). Llwchwr UD (0.04) had the weakest correlation of all districts, a somewhat surprising finding since it neighbours Swansea CB, with the community located only 8.9km from the centre of Swansea. One might expect the district to be closely coupled with Swansea CB like its neighbouring districts of Llanelli RD and Pontardawe (RD), and more populous urbanised districts close by such as Llanelli MB and Neath MB. This may be explained by its small population and divergent rate and size of susceptible recruitment, resulting in asynchronous outbreaks despite disease reintroduction from neighbouring districts. Another potential factor could be significant underreporting in the district during the study period. Very weak correlations are noted in isolated and lightly populated market towns dotted across the region. These include Llandovery MB (0.11) in North-east Carmarthenshire, Narberth UD (0.19) in Pembrokeshire and Monmouth MB (0.24) in North Monmouthshire (Fig. 4.25). All four districts have a mean population size below 6,000 inhabitants throughout the study period. Very weak positive correlations were also found in Newcastle Emlyn RD (0.09) and Narberth RD (0.22), both

districts with highly dispersed rural populations; the mean population size is approximately 10,000 for the districts, both of which are located many dozens of kilometres away from the nearest large town or prominent population centre. Newcastle Emlyn RD is approximately 64km north-west of Swansea CB, while Narberth RD is around 72.5km west.

4.7.3 Scarlet fever

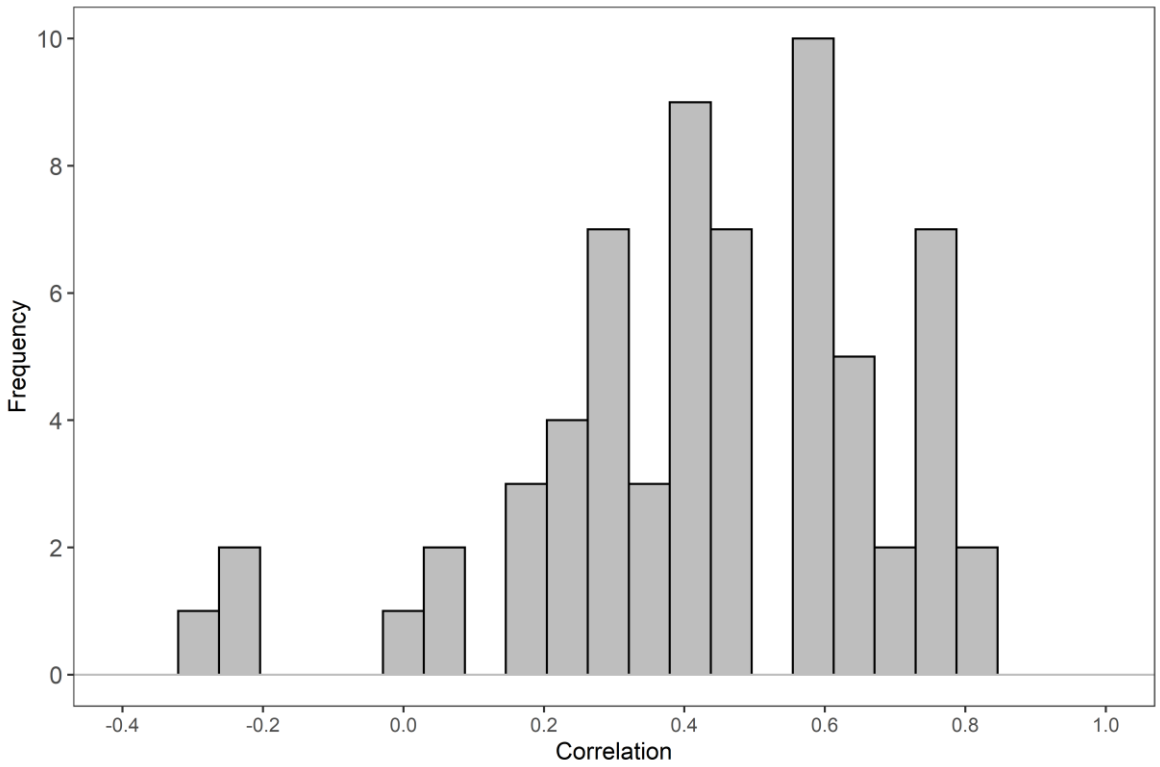

Figure 4.26 *Frequency distribution of the ordinary sample correlation coefficient between annual counts of scarlet fever notifications for each South Wales district and the average over the remaining 74 districts, 1940–1969. Mean sample correlation coefficient = 0.45.*

The mean correlation is 0.45, indicating a medium degree of synchronicity in scarlet fever disease patterns between districts and the average pattern in South Wales. Of the 75 local government districts in South Wales, 41 districts have larger correlation coefficients than the regional average (see Fig. 4.26).

The strongest correlations with the regional average among districts in South Wales are found in Rhondda CB (0.82) and Llanelli MB (0.81). High positive correlations are found primarily in districts which populate the Valleys (see Fig. 4.27), focused on the Rhondda Fawr, Rhondda Fach and Ebbw Fach Valleys. The former includes the aforementioned Rhondda MB, Llantrisant & Llantwitfardre RD (0.74) and Mountain Ash UD (0.70), while the latter contains Abertillery UD (0.77), Tredegar UD (0.77) and Ebbw Vale UD (0.75). Strong correlations are also present in Caerphilly UD (0.78) and Cardiff RD (0.77) at the foot of the Valleys area (Fig. 4.27). The high degree of synchronicity of these districts with the regional average suggests scarlet fever activity may originate and propagate from the Valleys. This notion is reinforced by the weaker correlations in the three major urban centres on the coast assumed to be centres of disease activity, Cardiff CB (0.59), Swansea CB (0.26) and Newport CB (0.32). Cardiff CB expectedly has a stronger correlation than the other two urban centres, due to the greater strength of interaction between the city and the valley communities, which are more tightly bound together due to strong social and economic links through industrial activities and Cardiff's role as a key gateway to the Valleys area.

Two of the 75 districts have a negative correlation (Fig. 4.26): Pembroke MB (-0.28) and Pembroke RD (-0.22). This may be partially explained by Pembroke's status as a ferry port, resulting in population flows from all over the South Wales region and further afield. When combined with small populations of approximately 12,000 and 6,000 respectively, high birth rates and the districts' remoteness in relation to the Valleys area and major population centres in Glamorgan, this creates fertile ground for significant asynchronous diseases activity, resulting in highly irregular disease patterns compared to the regional average. Two other districts located in Pembrokeshire, Cemaes RD (0.00) and Fishguard & Goodwick UD (0.04) report no correlation or almost non-existent correlation with the average (Fig. 4.27). Both districts are remote, with the former home to sparsely populated communities in the form of highly dispersed small villages and hamlets while the latter is on the western coastal edge of the region, lying approximately 138 km away from Cardiff CB.

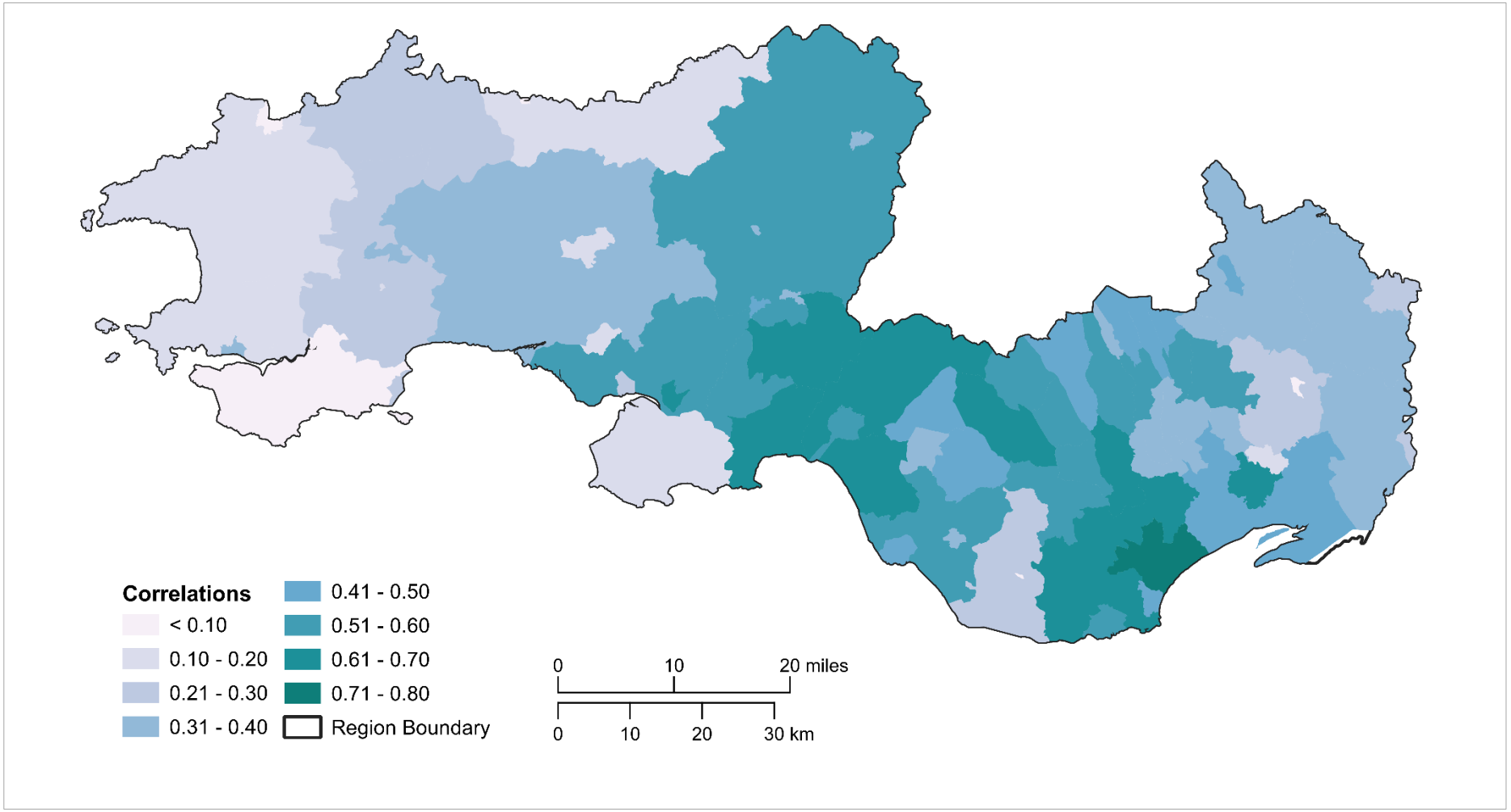

Figure 4.27 *Geographical distribution of the ordinary sample correlation coefficient between annual counts of scarlet fever notifications for individual districts in South Wales and the regional average over the remaining 74 districts, 1940–1969.*

Very weak correlations are noted in isolated market towns with small populations across the region, such as Cowbridge MB (0.06) in the Vale of Glamorgan and Usk UD (0.17) in Monmouthshire. Both districts have extremely small populations, ranging between 1,000 to 2,000 inhabitants. Several other rural and remote districts in Monmouthshire such as Magor & St. Mellons RD (0.18), Pontypool RD (0.19), and Mynyddislwyn UD (0.21) possess very weak correlations, highly suggestive of irregular disease patterns.

4.8 Chapter Summary

This chapter has provided a detailed exploratory and descriptive analysis of measles, pertussis and scarlet fever incidence in the Lancashire and South Wales regions across the study period. Several key trends are revealed. In the case of pertussis and scarlet fever, the incidence rates for both diseases decline gradually in both regions over the course of the study period, with notable accelerations in the fall of notification rates from the late 1950s onwards. The introduction of routine mass vaccination for pertussis in England and Wales in 1957 is followed by a dramatic fall in the magnitude of pertussis notifications in both regional metapopulations. However, the notification rate for measles remains relatively stable across the majority of the study period, with regional epidemic patterns found to be roughly synchronous. A significant fall in the notification rate is only noticed in the final two years of the study period, after the introduction of mass vacation for measles in 1968.

Disease activity is overwhelmingly, and consistently, concentrated in the heavily urbanised, industrial areas of each region. In the case of Lancashire, Manchester CB takes precedence closely followed by Liverpool CB, whereas in South Wales, visualisations of incidence patterns and sample correlations between individuals districts and the mean regional pattern of disease activity indicate Cardiff CB and the peri-urban industrialised valleys of Glamorgan are primarily the centre of disease activity. The time-series analysis of subsets of Lancashire and South Wales districts highlights the significant differences in wave-like

epidemic activity of childhood diseases among districts of varying population size. This trend is reflected in the analysis of spatial synchrony of epidemic behaviour at the local and regional level.

For instance, Manchester CB exhibits a regular biennial pattern of major measles epidemics throughout the study period (see Fig 4.2B), characteristic of Type I waves of infection found in settlements above the endemic threshold. Manchester CB consistently reports the strongest correlation with the mean regional pattern of disease activity for all three diseases among the 125 Lancashire districts (see Figs. 4.17, 4.19 & 4.21). As a major urban population centre and dominant economic settlement within the Lancashire metapopulation, waves of infection radiate from the Manchester conurbation and inevitably travel down the population hierarchy, first affecting intermediate sized towns below the endemic threshold value with regular outbreaks of epidemic disease. These Type II waves are reflected by large town districts which neighbour or surround Manchester CB, such as Rochdale CB, Bolton CB and Middleton MB which, despite being significantly less populated than Manchester CB, nevertheless report strong correlations with the mean regional disease pattern, most visibly for measles and scarlet fever (Figs. 4.17 & 4.21). With regards to pertussis, a similar pattern of disease spread is noted moving outwards from Liverpool CB, the other large population centre in the region, to surrounding districts in the Merseyside metropolitan area (Fig. 4.19). As one moves further away from the urban cores and beyond the metropolitan hinterlands where small satellite towns are well-connected with major population centres, irregular epidemics with extensive periods of disease fade-out, type III waves of infection, are the norm. As visualised and discussed in Sections 4.6 and 4.7, in both regional metapopulations, rural districts with highly dispersed, remote populations and limited urban settlement tend to report the weakest correlations with the regional disease pattern, indicative of communities that experience type III waves of infection. These districts are primarily found in northern portion of the Lancashire region, and the rural counties of Pembrokeshire and Carmarthenshire in South Wales (e.g., Figs. 4.19 & 4.27).

Overall, the exploratory analysis presented in this chapter indicates that disease persistence is shaped by factors which are intimately geographical in nature and are not purely driven by intrinsic demographic events or stochastic behaviour. The findings suggest that the spatial structure and geographical distribution of subpopulations within the wider regional metapopulation play a key role in dictating the spatial synchrony of disease activity. Even in a region in Lancashire, where the Manchester and Liverpool conurbations operate as major endemic reservoirs for childhood infections, there are nevertheless pockets of epidemic activity and disease persistence which are driven by more localised relationships and dynamics that do not fit the classic Bartlett model of disease spread according to varying population size. This is indicated by the correlations between individual districts and the regional pattern of pertussis activity in South Wales, with strong correlations observed in urban districts situates in lightly populated, pastoral areas such as central Carmarthenshire and West Pembrokeshire. These areas are home to market towns such as Haverfordwest MB and Pembroke MB, settlements with populations of approximately 10,000 or less. One would expect these districts would be subject to Type III waves, sporadic epidemic outbreaks with long periods of disease fadeout, due to the small susceptible populations which exist in these communities and their remoteness, many dozens of miles away from the major urban centres in the wider region. Thus, one would also expect to detect weak correlations with the regional pattern of disease activity in these districts, but that is not what is observed (see Fig. 4.25).

In the following chapter, endemic threshold population estimates will be calculated and analysed to further explore the persistence dynamics in the two regional metapopulations beyond explanations centred on demographic stochasticity, instead focusing on the geography of the two regions which both possess complex hierarchical spatial structures and varying levels of connectivity between subpopulations.

Chapter 5: Spatiotemporal Changes in Endemic Thresholds

5 Introduction

This chapter presents the results of the 'moving window' empirical regression approach for estimating and tracking spatiotemporal changes in endemic threshold populations for measles, pertussis and scarlet fever in the Lancashire and South Wales regions. This approach was chosen to monitor systematic time changes in endemic threshold size in response to wider social, economic and demographic changes affecting the regional metapopulations throughout the study period (these changes are discussed in detail in chapter three). After describing the temporal changes in endemic threshold populations in each regional metapopulation, among low/high density and low/high connectivity districts, spatiotemporal patterns of percentage endemicity for each infection over nine time–windows are visualised using choropleth maps and analysed. Finally, a discussion of the findings is presented, exploring factors which potentially shape regional differences in spatiotemporal changes in endemic threshold populations over time. A full breakdown of the statistical outputs of empirical regressions fitted, along with accompanying endemic threshold estimates for each time–window, can be found in Appendix III (Tables III.1–15). For pertussis and scarlet fever, endemic threshold estimates calculated using the best fitting regression model, an empirical log-log regression, are presented. For measles, endemic threshold estimates obtained from empirical log-linear regression models are presented.

5.1 Endemic Threshold Estimates: Measles

5.1.1 Regions

Lancashire: A mean regional endemic threshold population of ~292,000 is estimated for the Lancashire region, using a sample of 123 local government district populations.

Endemic threshold estimates range between 209,000–378,000 over the course of the study period (see Fig. 5.1). The largest endemic threshold population estimates were calculated in time–window two (378,000, 95% CI: 295,000–484,000) and time–window six (368,000; 95% CI: 311,000-436,000). From time–window six onwards, the endemic threshold fell significantly, almost halving by time–window nine (Fig. 5.1). This final estimate is somewhat surprising given the introduction of mass, routine measles vaccination in 1968, which hypothetically should shrink the pool of susceptibles and fuel an increase in the endemic threshold population, albeit it a small increase given its introduction late in the time–window.

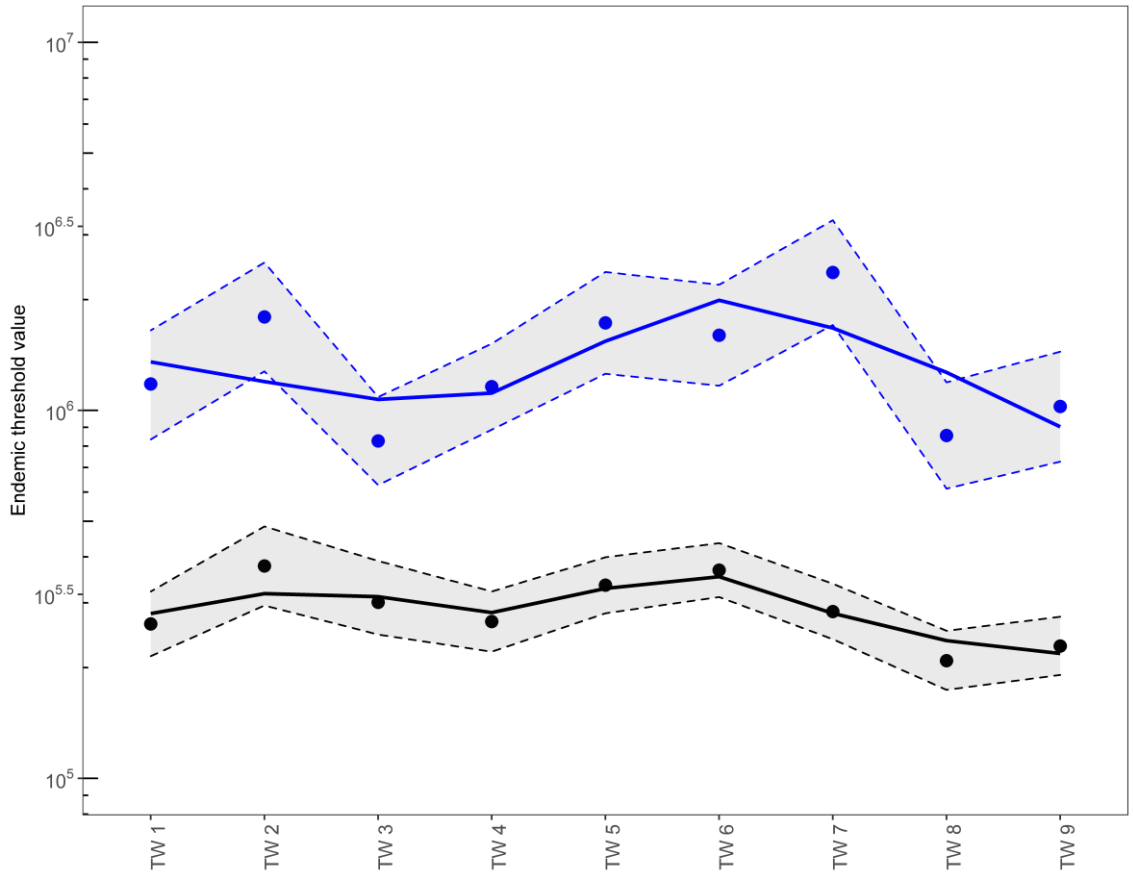

Figure 5.1 *Regional endemic threshold size estimates for measles in Lancashire and South Wales for nine time–windows, 1940–1969. Dots represent the endemic threshold estimate for each window. To show the time trend across time–windows, solid LOESS lines have been fitted to the points for each region. The pink shaded area denotes the vaccine era. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals.*

South Wales: The mean endemic threshold value for South Wales across the nine time–windows is ~1,390,000. By time–window two, the estimated threshold population had risen to ~1,795,000 (95% CI: 1,277,000-2,523,00), falling markedly to 826,000 (95% CI: 627,000-1,088,000) in the next window. Over the course of successive time–windows, the endemic threshold population estimates grew substantially (Fig. 5.1), peaking at 2,370,000 (95% CI: 1,706,000-3,288,000) in time–window six. In the penultimate time–window, the endemic threshold population fell significantly once more, to 855,000 (95% CI: 613,000-1,1,92,000). This was followed by a slight recovery in the final time–window (Fig. 5.1).

5.1.2 Low Density Districts

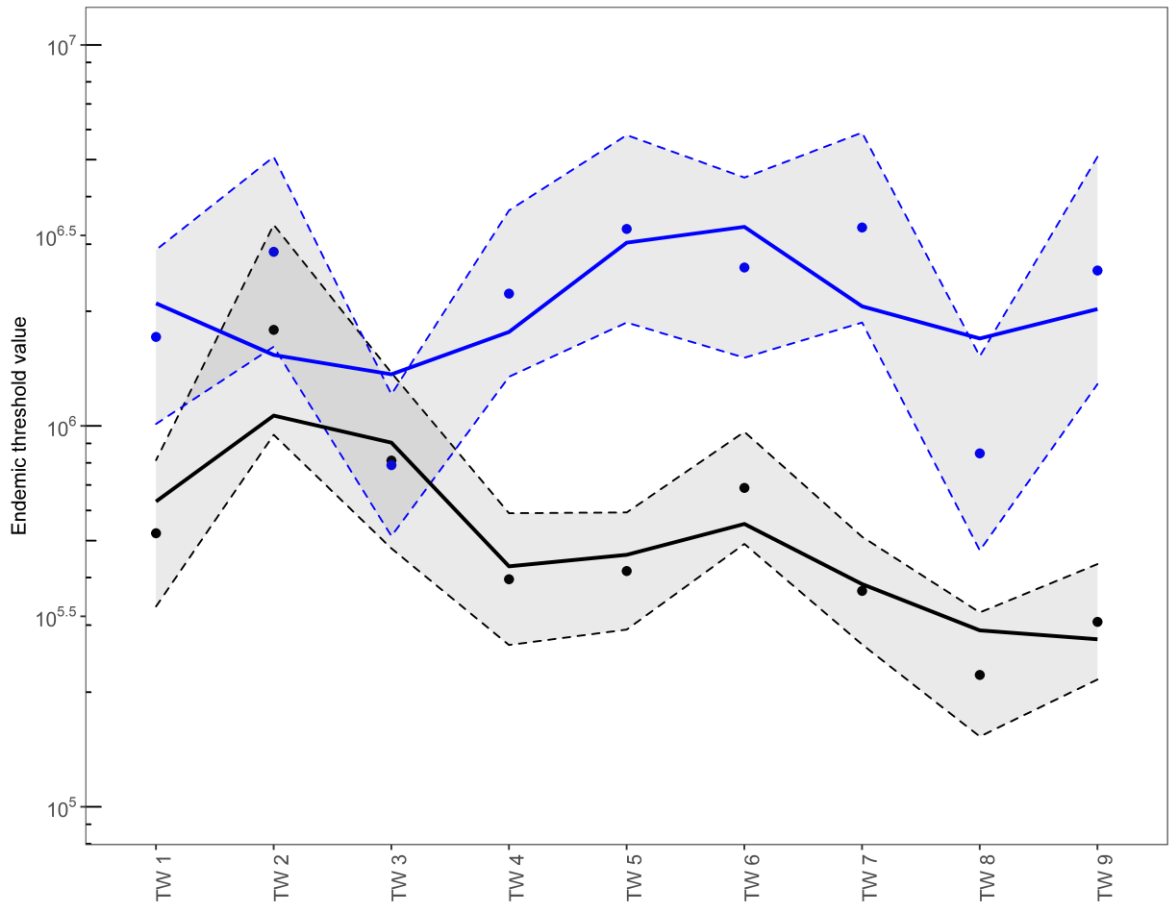

Figure 5.2 *Endemic threshold size estimates for measles in low density districts in Lancashire and South Wales for nine time–windows, 1940–1969. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals.*

Lancashire: Among low density districts in Lancashire, the mean endemic threshold population size across the nine time–windows is 614,000. The estimated endemic threshold value in time–window one is 523,000, increasing dramatically and peaking at 1.8 million in the following window (1943-48). Across successive time–windows, the endemic threshold fell substantially in value (see Fig. 5.2), to 416,000 by time–window five, before staging a modest recovery in time–window six. By the ninth and final time–window, the endemic threshold population had more than halved to 306,000 (Fig. 5.2).

South Wales: The mean threshold population size among low density districts in South Wales across the nine time–windows is approximately 2,250,000. After the opening time–window, precipitous drops in endemic threshold size are observed in time–window three (789,000; 1946-1951) and time–window eight, falling to (847,000; 1961-66). In the ninth and final time–window, the endemic threshold population stages a substantial recovery (Fig. 5.2), increasing to 2,560,000.

5.1.3 High Density Districts

Lancashire: Among high density districts in Lancashire, the mean endemic threshold population size across the study period is ~270,000. Across the nine time–windows, endemic threshold estimates remain relatively constant (see Fig. 5.3), with the lowest endemic threshold size observed in time–window eight (215,000; 95% CI: 165,000-281,000), and the highest estimate in time–window six (330,000, 95% CI: 251,000-434,000).

South Wales: Among high density districts, the mean threshold population size over the full duration of the study period is 1,150,000. Similar to low density districts, significant falls in the size of the endemic threshold population are observed around time–window three

and time–window eight (Fig. 5.3). However, these falls in endemic threshold size are not met by sizable recoveries in the following time–window. For instance, after falling from 2,010,000 (95% CI: 1,325,000-3,024,000) in time–window seven to 864,000 (95% CI: 578,000-1,290,000) in time–window eight, the endemic threshold population of high density districts stages no recovery, only falling further to 690,900 (95% CI: 468,000-1,016,000).

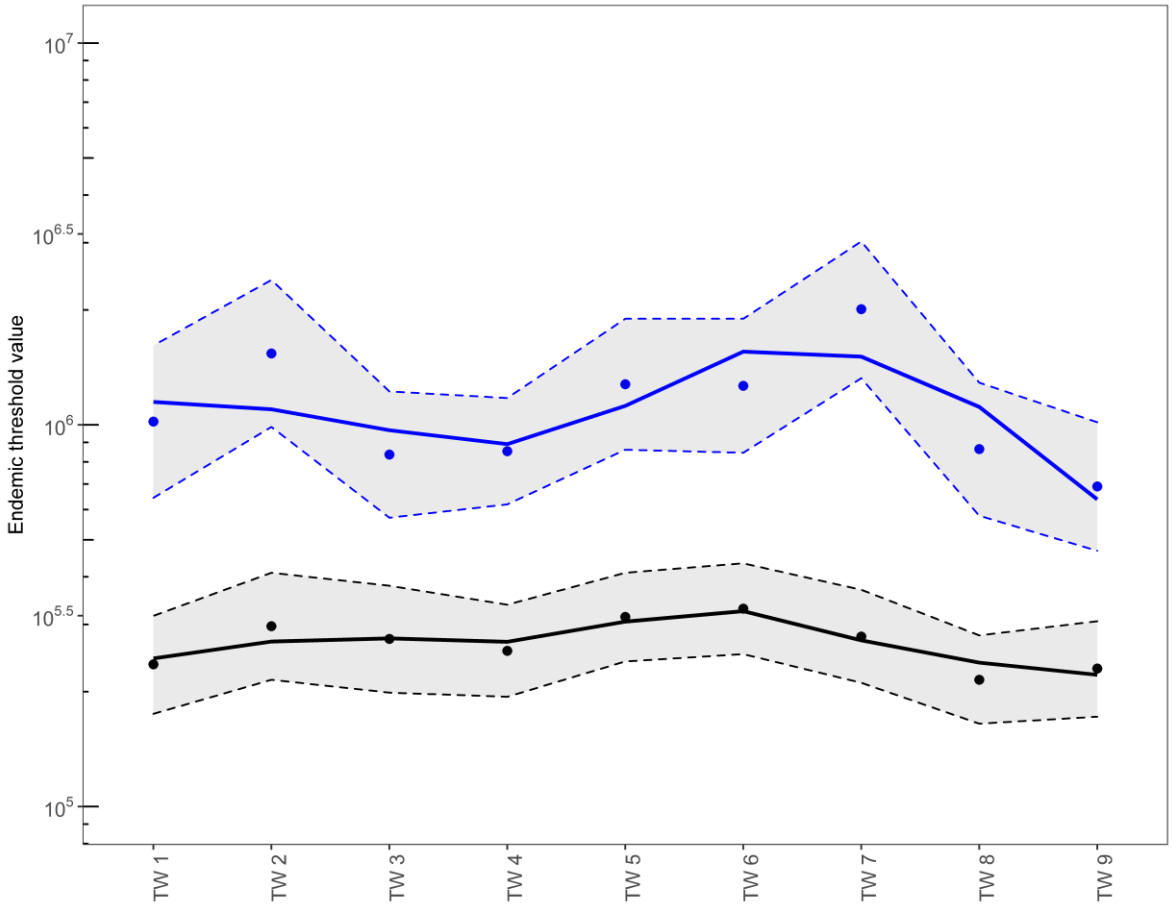

Figure 5.3 *Endemic threshold size estimates for measles in high density districts in Lancashire and South Wales for nine time–windows, 1940–1969. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals.*

Reporting of finding continues on the next page.

5.1.4 Low Connectivity Districts

Lancashire: Among low connectivity districts, the mean endemic threshold size is 466,000 over the course of the study period. In the first time–window (1940-45), the endemic threshold size was approximately 377,000 (95 CI%: 281,000-506,000), jumping to 745,000 (95 CI%: 513,000-1,084,000) in the following time–window before declining across successive time–windows (see Fig. 5.4), falling to 412,000 (95 CI%: 307,000-553,000) by time–window five (1952-57)., Endemic threshold estimates continue to gradually fall, with a final estimate of 374,000 (95 CI%: 271,000-518,000) in time–window nine.

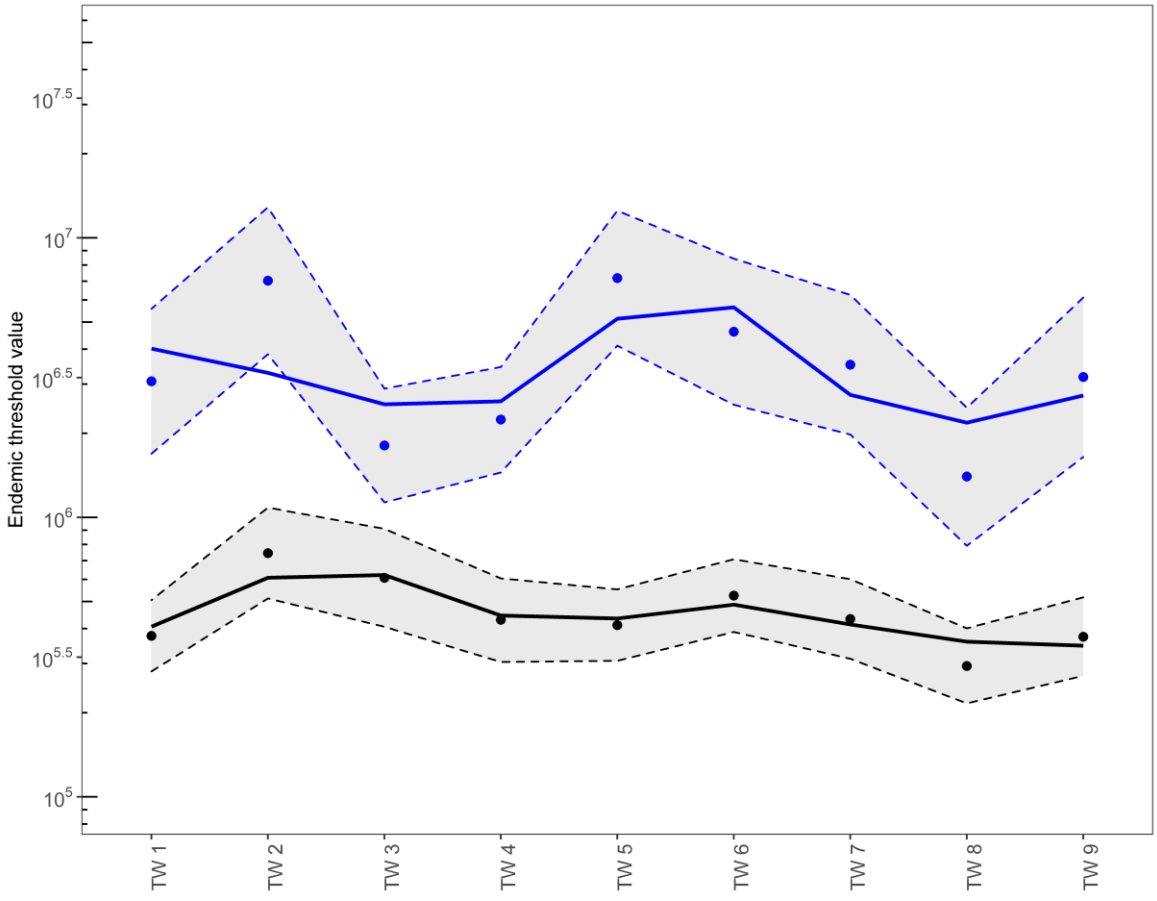

Figure 5.4 *Endemic threshold size estimates for measles in low connectivity districts in Lancashire and South Wales for nine time–windows, 1940–1969. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals.*

South Wales: Among low connectivity districts, the mean regional threshold population size across the nine time–window is 3,800,000. In the opening time–window (1940-45), the estimated endemic threshold value is 3,070,000, doubling in time–window two to ~7,000,000 before falling substantially in time–window three (Fig. 5.4), mirroring the pattern among low density districts (see Fig. 5.2). In this vein, there was a recovery in the endemic threshold population across successive windows, before another dramatic decline in time–window eight, reaching a low of 1,400,000 (95% CI: 792,000-2,470,000). A sizable recovery in the endemic threshold population is observed in the final time–window (Fig. 5.4).

5.1.5 High Connectivity Districts

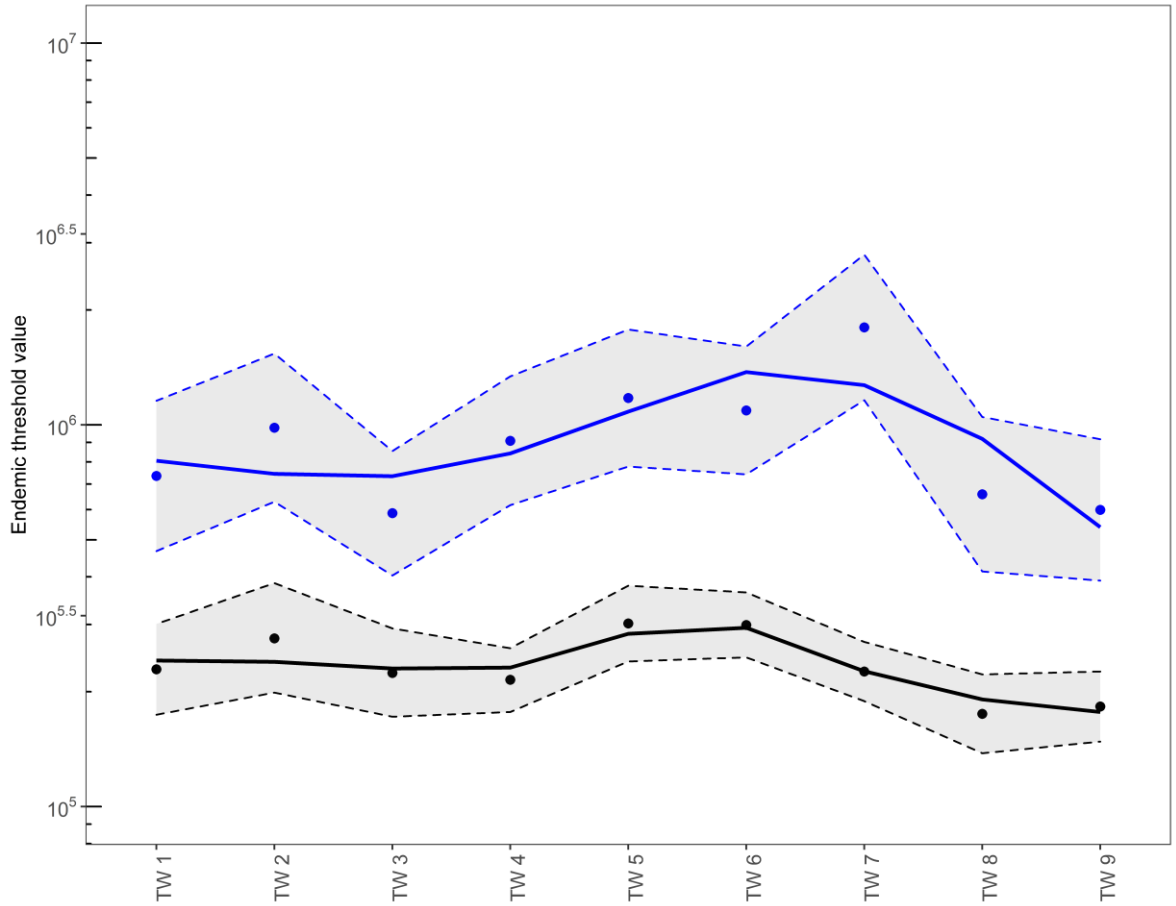

Figure 5.5 *Endemic threshold size estimates for measles in high connectivity districts in Lancashire and South Wales for nine time–windows, 1940–1969. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals.*

Lancashire: Among high connectivity districts in the Lancashire region, the mean endemic threshold population size across the study period was 236,000, ~35,000 lower than high density districts. Between time–windows one and seven, endemic threshold estimates ranged ~200,000–300,000 (see Fig. 5.5), before falling slightly in the latter time–windows, to a low of 183,000 (95% CI: 148,000-226,000) in time–window nine.

South Wales: Among high connectivity districts, the regional mean threshold population size across the nine windows is 950,000. In time–window two, the endemic threshold population reaches 983,000 (95% CI: 629,000-1,537,000), before almost halving in size the following time–window, to 587,000 (95% CI: 403,000-854,000). This significant fall in time–window three (1946-51) is followed by a substantial growth in threshold estimates across successive windows (Fig. 5.5), before declining markedly once more in time–window eight, falling by approximately two-thirds, from ~1,800,000 (95%CI: 1,1,60,000-2,792,000) in time–window seven, to 658,000 (95% CI: 413,000-1,049,000).

5.2 Endemic Threshold Estimates: Pertussis

5.2.1 Regions

Lancashire: Among the full sample population of 123 districts in Lancashire, the mean endemic threshold population size in the pre-vaccine era (1940-1957) is approximately ~156,000. After the onset of routine mass vaccination for pertussis, the mean endemic threshold value for the vaccine era (1958-1969) almost increased by a factor of three, to ~431,000 (see Fig. 5.6). In time–window one, endemic threshold estimate is 125,000 (95% CI: 92,000-169,000); endemic threshold size remained relatively consistent in size until time–window six. From time–window six onwards, the introduction of vaccination saw the regional threshold population gradually rise across successive windows until peaking in the final time–window (Fig. 5.6), with an estimate of 461,000 (95% CI: 332,000-641,000).

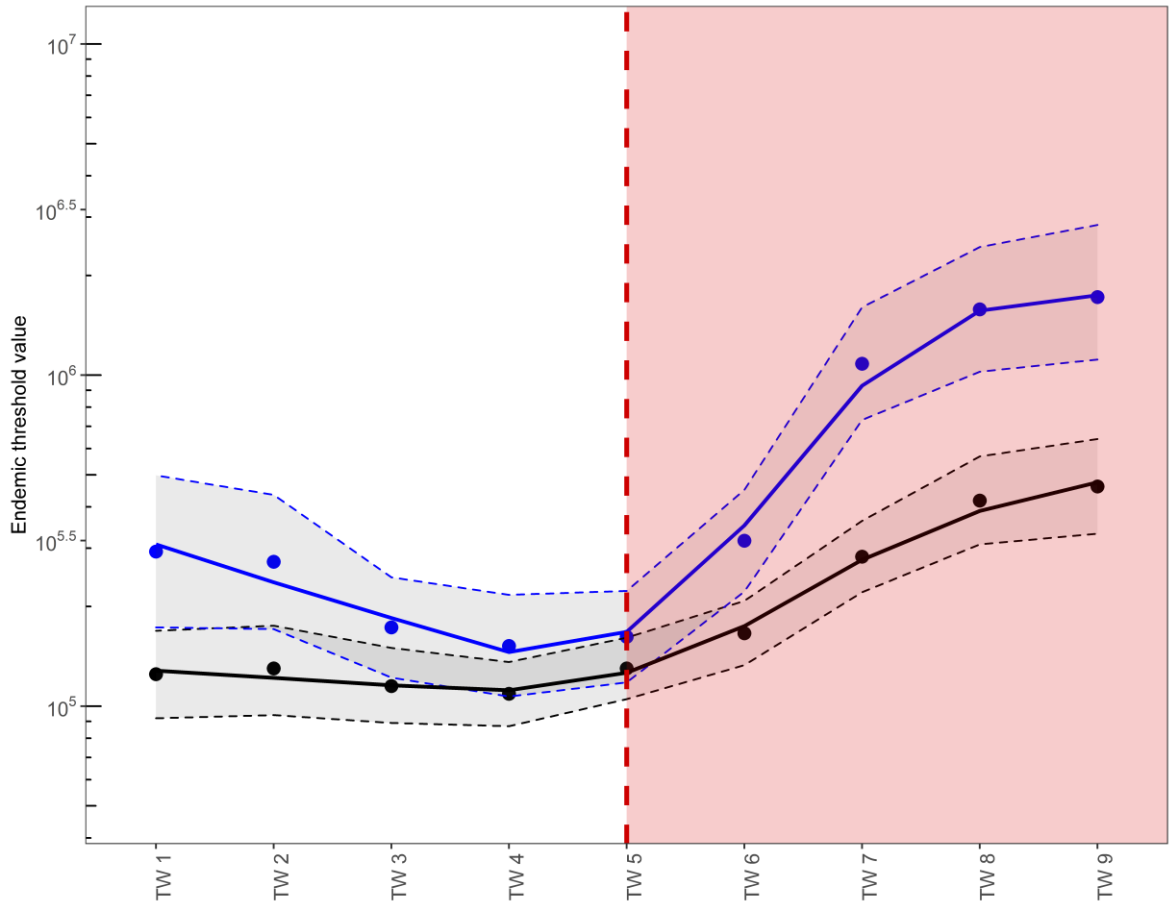

Figure 5.6 *Regional endemic threshold size estimates for pertussis in Lancashire and South Wales for nine time–windows, 1940–1969. The pink shaded area denotes the vaccine era. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals. The pink shaded area denotes the vaccine era.*

South Wales: Among the full sample population of 74 districts in South Wales, the mean endemic threshold population size in the pre-vaccine era (1940-1957) is estimated at 228,000. The mean endemic threshold value for the vaccine era (1958-1969) is an estimated population size of 1.46 million (see Fig. 5.6). In time–window one, the endemic threshold is 293,000. This value fell across successive time–windows to a low of 152,000 by time–window four and recovered only marginally to 162,000 by time–window five (1952-57). In the vaccine era, the endemic threshold population grew almost sixfold between time–windows six and nine (Fig. 5.6), rising from 316,000 (95 CI%: 222,000-450,000) to a peak of 1,720,000 (95% CI: 1,114,000-2,483,000).

5.2.2 Low Density Districts

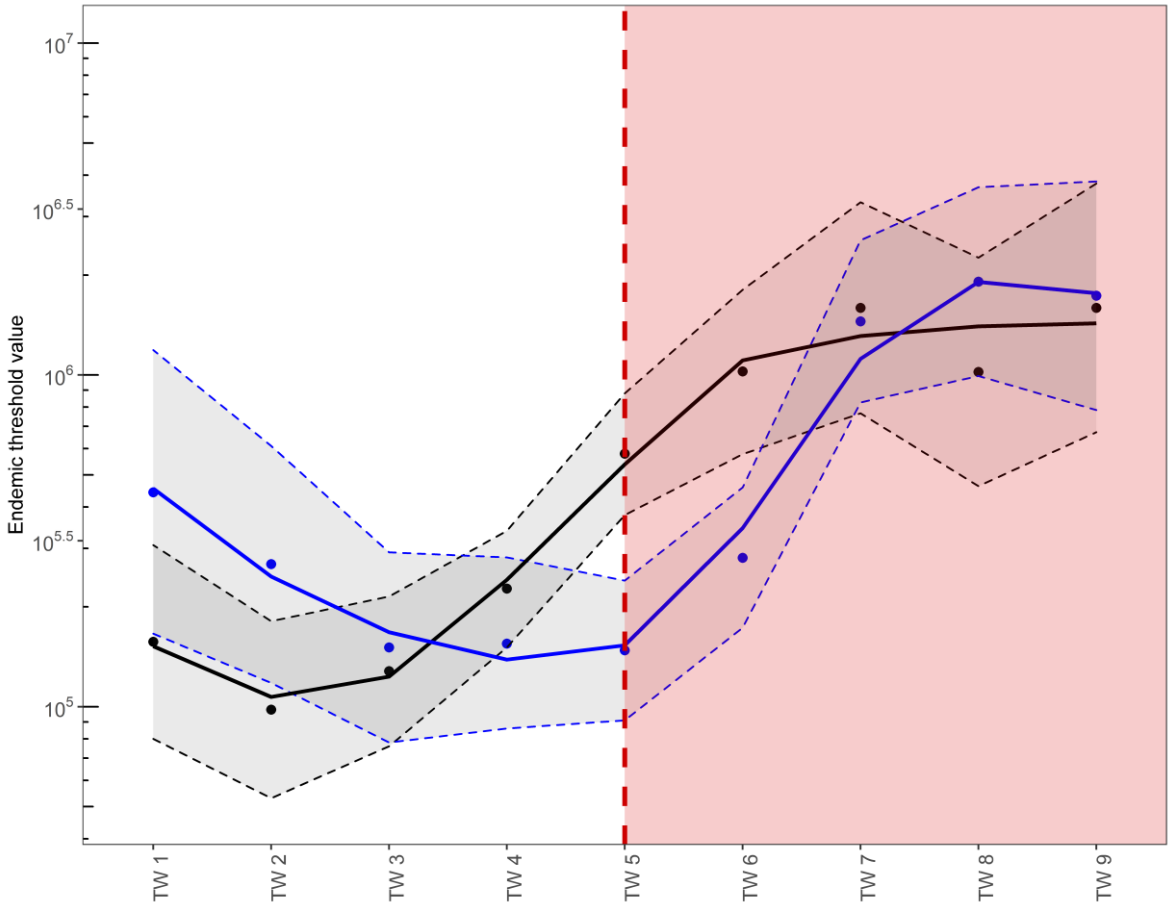

Figure 5.7 *Endemic threshold size estimates for pertussis in low density districts in Lancashire and South Wales for nine time–windows, 1940–1969. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals. The pink shaded area denotes the vaccine era.*

Lancashire: The mean endemic threshold population size in pre-vaccine era Lancashire, among low density districts, is estimated at ~171,000. After the onset of mass vaccination, this increased dramatically, almost tenfold, to a mean threshold size of approximately ~1,070,0000 (see Fig. 5.7). The lowest estimate was observed in time–window two (98,000, 95% CI: 53,000-181,000), concurrent with rising birth rates at the beginning of the baby boom (Fig. 5.7). Mass vaccination marks the beginning of an exponential increase in endemic threshold population size, rising from 579,000 (95% CI: 378,000-880,000) in time–window five to 1,593,000 (95% CI: 672,000-3,774,000) in time–window nine.

South Wales: Among low density districts in South Wales, the mean endemic threshold population size in the pre-vaccine era is estimated at 241,000. In the vaccine era, the mean endemic threshold population had risen to ~1,700,000 (see Fig. 5.7). In time–window one, the endemic threshold value is 443,000 (95% CI: 166,000-1,188,000), before falling gradually to a low of 148,000 (95% CI: 91,000-240,000) by time–window five. Vaccination marks a period of significant growth, with the endemic threshold rising exponentially across successive time–windows (Fig. 5.7), peaking in time–window eight at approximately 1,911,000 (95% CI: 993,000-3,828,000).

5.2.3 High Density Districts

Lancashire: Among the subset of high density districts in Lancashire, the mean endemic threshold population size in the pre-vaccine era is estimated approximately ~180,000. After the introduction of vaccination, the threshold population more than doubled, rising to a mean estimated threshold value of ~470,000. Size estimates for the endemic threshold population remain consistent across time–windows one to five (see Fig. 5.8). The onset of vaccination sees the endemic threshold population double in size between time–windows six and nine, rising to 376,000 (95 CI%: 256,000-552,000) in the final time–window.

South Wales: Among the subset of high density districts in South Wales, the mean endemic threshold population size in the pre-vaccine era is estimated at ~224,000. In the vaccine era, this figure rises substantially, to approximately ~1.360,000 (Fig. 5.8). In time–window two observes a pre-vaccine era high estimate of 261,000 (95% CI; 147,000-471,000), before falling in successive time–windows to a low of 148,000 (95% CI: 96,000-230,000) in time–window four. Introduction of mass vaccination tremendous growth in the size of the endemic threshold population from time–windows six to nine, reaching a peak of 1,726,000 (95 CI%: 888,000-3,356,000) in the final time–window.

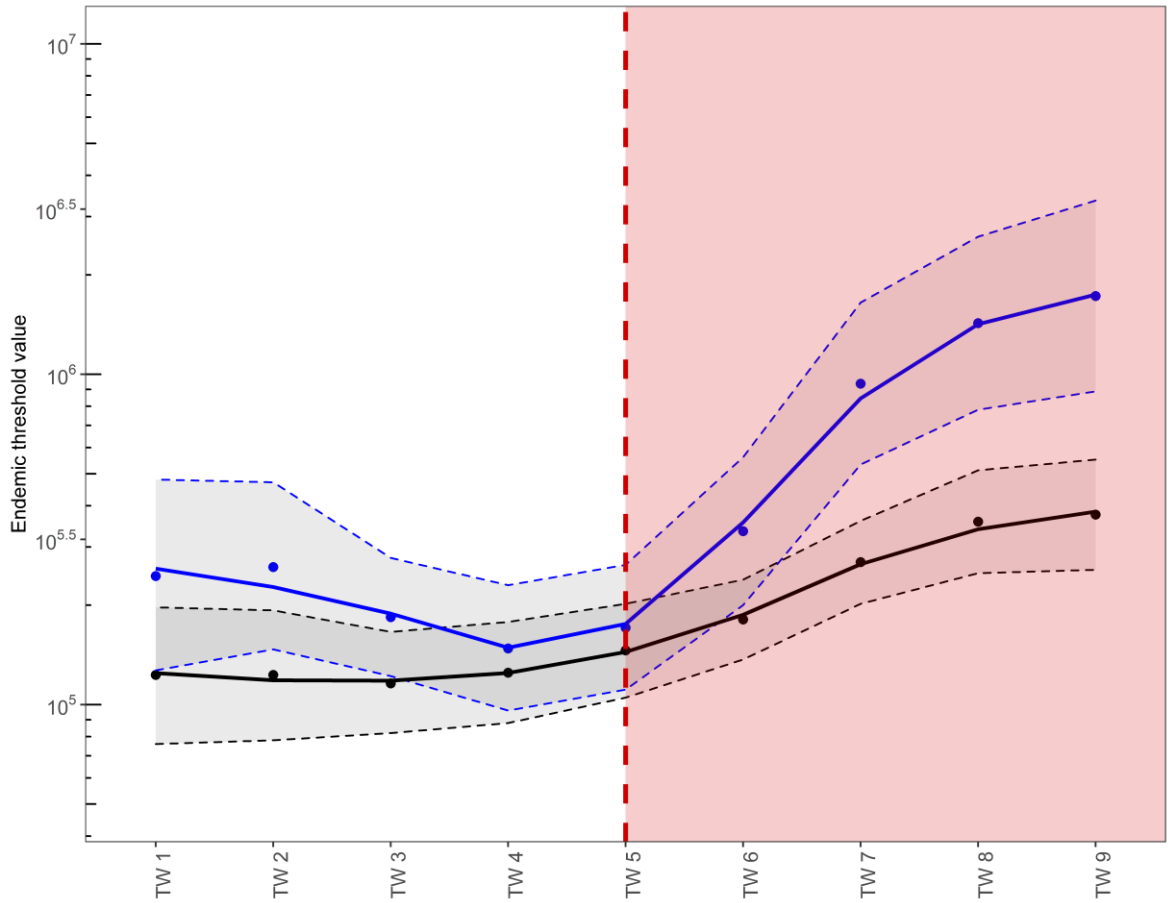

Figure 5.8 *Endemic threshold size estimates for pertussis in high density districts in Lancashire and South Wales for nine time–windows, 1940–1969. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals. The pink shaded area denotes the vaccine era.*

5.2.4 Low Connectivity Districts

Lancashire: The mean endemic threshold population size in the pre-vaccine era is ~157,000. In the vaccine era, this increases substantially to a mean threshold population size of approximately ~1,150,000 (see Fig. 5.9). In time–window one, the endemic population estimate is 139,000 (95% CI: 97,000-199,000), with values ranging between 120,000–160,000 in the pre-vaccine era. In the vaccine era, there is substantial threshold growth from time–windows six to nine (Fig. 5.9), increasing from 231,000 (95% CI: 168,000-317,000) to a peak of 1,681,000 (95% CI: 950,000-2,976,000) by time–window nine.

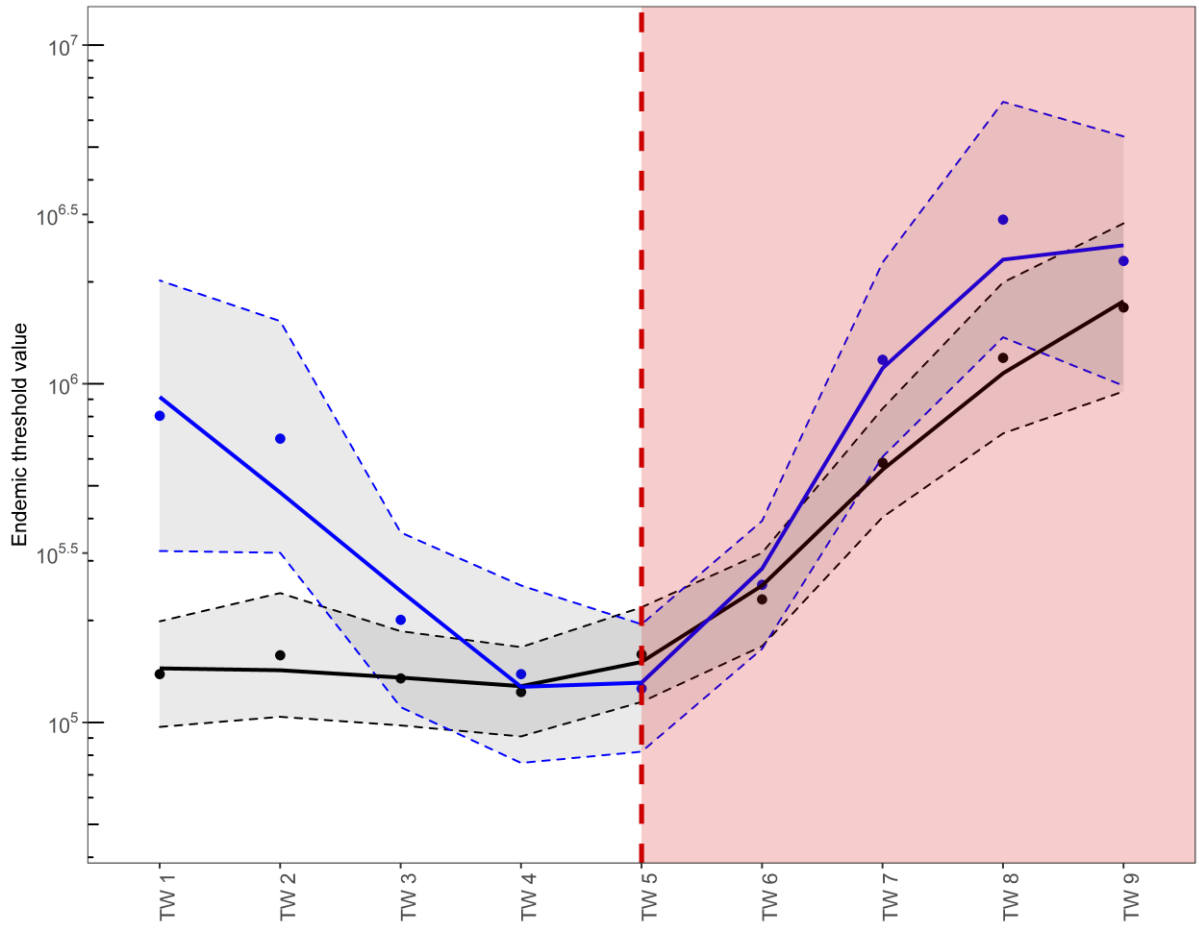

Figure 5.9 *Endemic threshold size estimates for pertussis in low connectivity districts in Lancashire and South Wales for nine time–windows, 1940–1969 Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals. The pink shaded area denotes the vaccine era.*

South Wales: The mean endemic threshold population size in the pre-vaccine era is estimated at 369,000. After the introduction of vaccination, this increased substantially to an average threshold population size of ~2,180,000 (see Fig. 5.9). In time–window one, the endemic threshold value was estimated at 850,000 (95% CI: 321,000-2,020,000) falling significantly across successive time–windows to a low of just 126,000 (95% CI: 82,000-195,000) in time–window five. However, the introduction of vaccination is followed by an exponential growth in endemic threshold population size (Fig. 5.9), peaking at 3,054,000 (95% CI: 1,372,000-6,802,000) in the penultimate time–window.

5.2.5 High Connectivity Districts

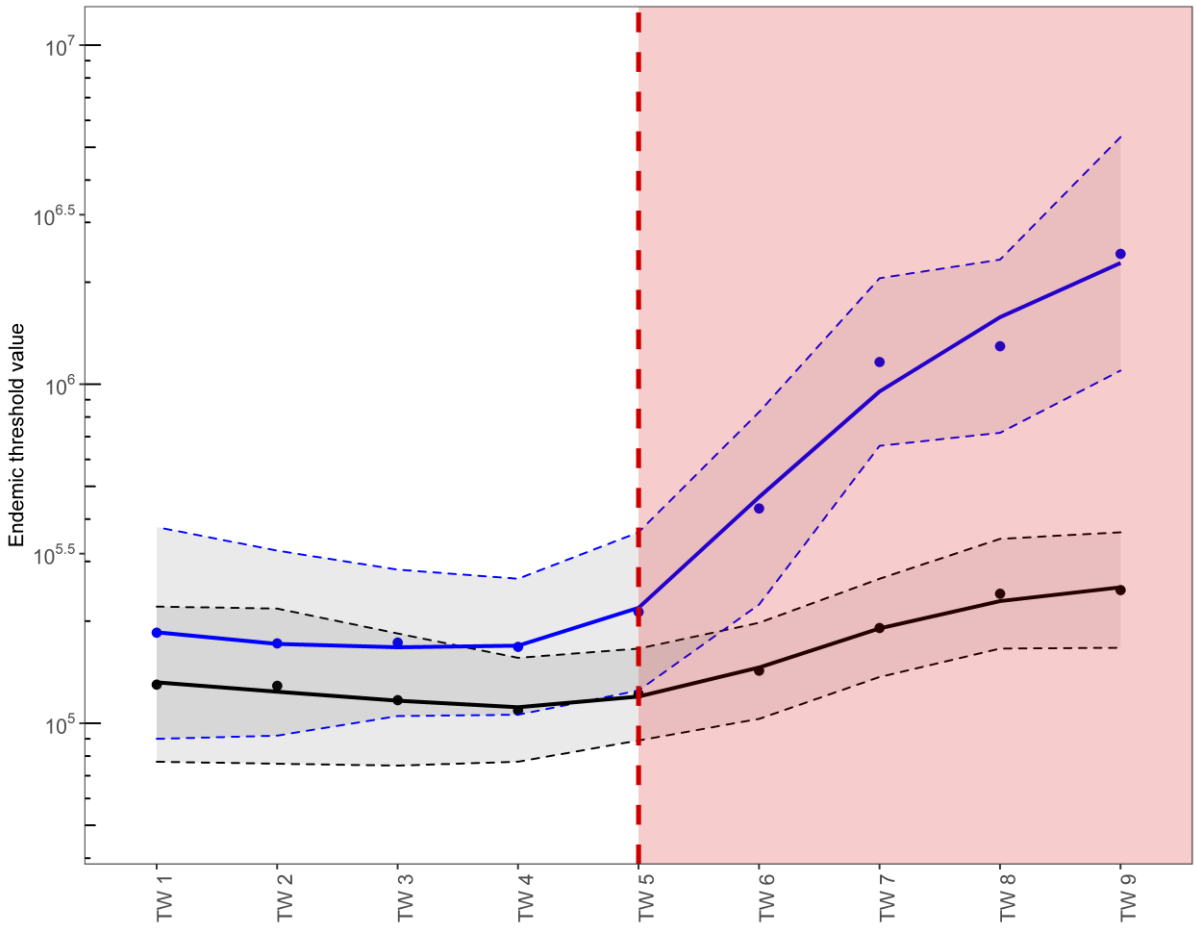

Figure 5.10 *Endemic threshold size estimates for pertussis in high connectivity districts in Lancashire and South Wales for nine time–windows, 1940–1969. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals. The pink shaded area denotes the vaccine era.*

Lancashire: The mean endemic threshold estimate during the pre-vaccine era is ~180,000, almost doubling in size during the vaccine era, rising to approximately ~350,000. The endemic threshold population in time–window one is 130,000 (95% CI: 77,000-221,000). There is minimal variation in threshold size between time–windows one to six (see Fig. 5.10). The introduction of vaccination only sees a relatively small increase in the endemic threshold estimates compared to groupings of high density, low connectivity and low density districts in Lancashire, rising to 247,000 (95% CI: 167,000-366,000) by time–window nine.

South Wales: The mean endemic threshold population in the pre-vaccine era is estimated at ~223,000. In the vaccine era, this figure rose substantially to ~1.620,000. From time–windows one to five, endemic threshold population estimates ranged between 168,000–213,000 (Fig. 5.10). In the vaccine era, the size of the endemic threshold population increases almost by a factor of five, rising from 430,000 (95% CI: 224,000-828,000) in time–window six to 2,425,000 (95% CI: 1,097,000-5,359,000) by the final time–window.

5.3 Endemic Threshold Estimates: Scarlet fever

5.3.1 Regions

Lancashire: Among the full sample of 123 districts in Lancashire, the mean regional endemic threshold across the nine time–windows is ~161,000. In time–window one, the regional endemic threshold population size is estimated to be 71,000 (95% CI: 57,000-89,000). Over the course of the study period, the size of the endemic threshold population monotonically increases (see Fig. 5.11), with no fall observed in the size of estimates across successive time–windows. The endemic threshold population peaks in the ninth time–window, at 213,000 (95% CI: 154,000-294,000).

South Wales: The mean regional endemic threshold population size for South Wales across the nine time–windows is ~283,000. A similar pattern of growth in endemic threshold population is observed in South Wales as the one previously described in Lancashire over the course of the study period. In contrast to Lancashire, the growth in regional endemic threshold size increases exponentially in South Wales in from time–window six onwards (Fig. 5.11), rising from 192,000 (95% CI: 144,000-256,000) to a peak of 708,000 (95% CI: 440,000-1,137,000) in time–window eight, before falling slightly in the final time–window

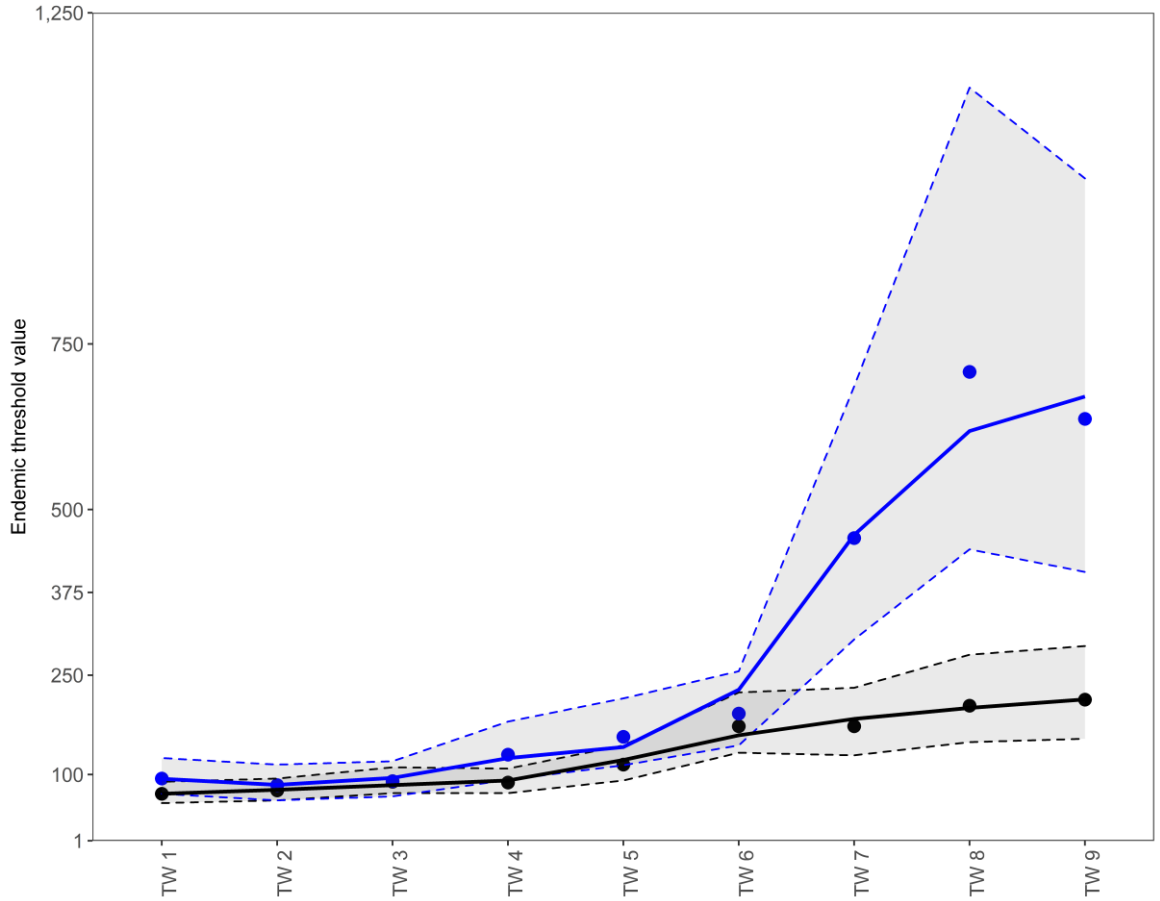

Figure 5.11 *Regional endemic threshold size estimates for scarlet fever in Lancashire and South Wales for nine time–windows, 1940–1969. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals*

5.3.2 Low Density Districts

Lancashire: For low density districts, the mean endemic threshold population across the nine time–windows is ~163,000. In time–window one, endemic threshold size is estimated to be approximately 89,000 (95% CI: 66,000-122,000). The size of the endemic threshold population gradually increases across successive time–windows (see Fig. 5.12), peaking at 293,000 (95% CI: 166,000-517,000) in the final time–window. Overall, the endemic threshold population increases by approximately 200,000 over the course of the study period.

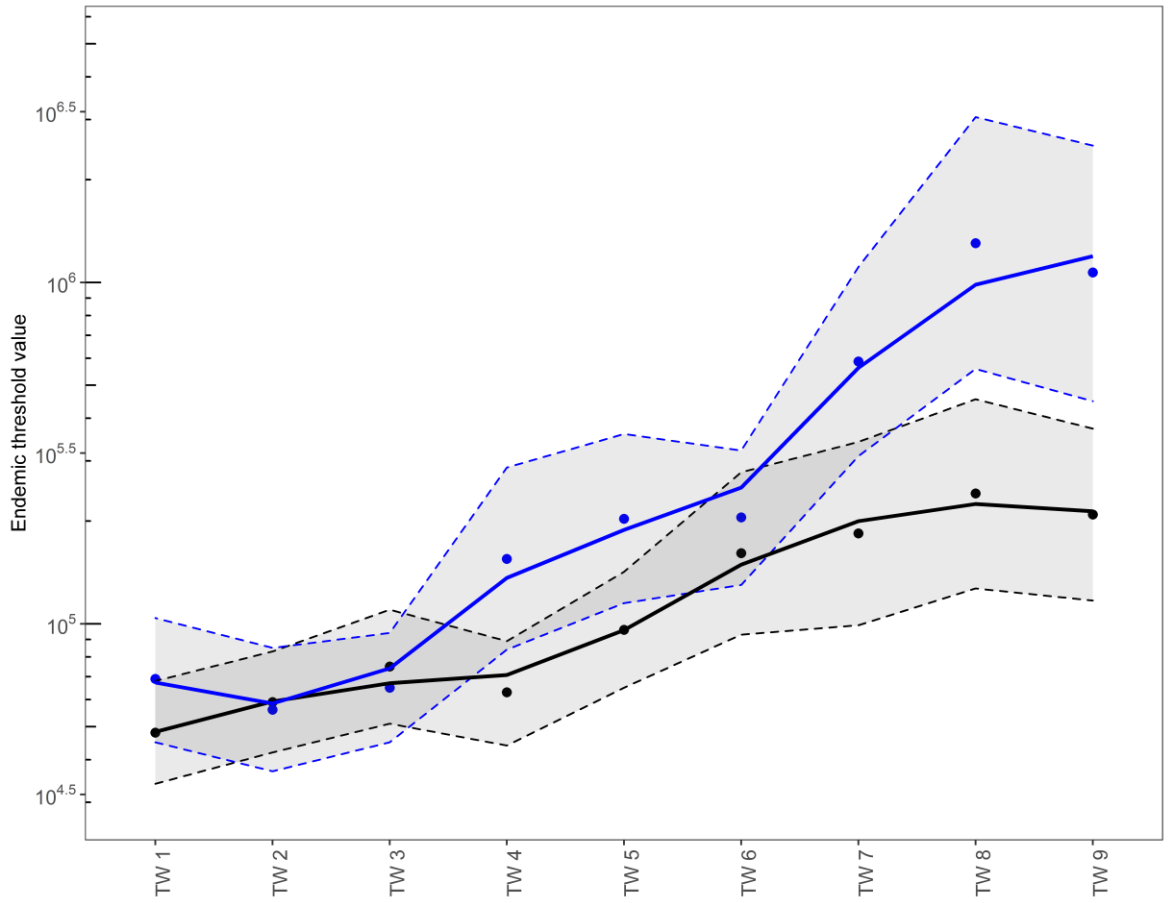

Figure 5.12 *Endemic threshold size estimates for scarlet fever in low density districts in Lancashire and South Wales for nine time–windows, 1940–1969. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals.*

South Wales: The mean endemic threshold population size for low density districts across the study period is ~412,000. Reflecting the overall regional trend of growth in endemic threshold population size, estimates gradually rise across the time–windows, from 69,000 (95% CI: 45,000-104,000) in time–window one to 205,000 (95% CI: 130,000-322,000) in time–window six (Fig. 5.12). In following time–windows, the endemic threshold population increases exponentially, peaking at 1,303,000 (95% CI: 558,000-3,050,000) in time–window eight, before falling in the final time–window, to 1,070,000 (95% CI: 449,000-2,517,000).

5.3.3 High Density Districts

Lancashire: The mean endemic threshold population across the nine time–windows is ~126,000. In time–window one, the estimated endemic threshold population size is approximately 48,000 (95% CI: 34,000-68,000). The size of the endemic threshold gradually increases across successive time–windows (see Fig. 5.13), peaking at 241,000 (95% CI: 127,000-455,000) in time–window eight before falling slightly in the final time–window (see Fig. 5.13). Overall, the endemic threshold population increases by ~ 150,000 over the course of the study period, with the greatest increases observed in later time–windows.

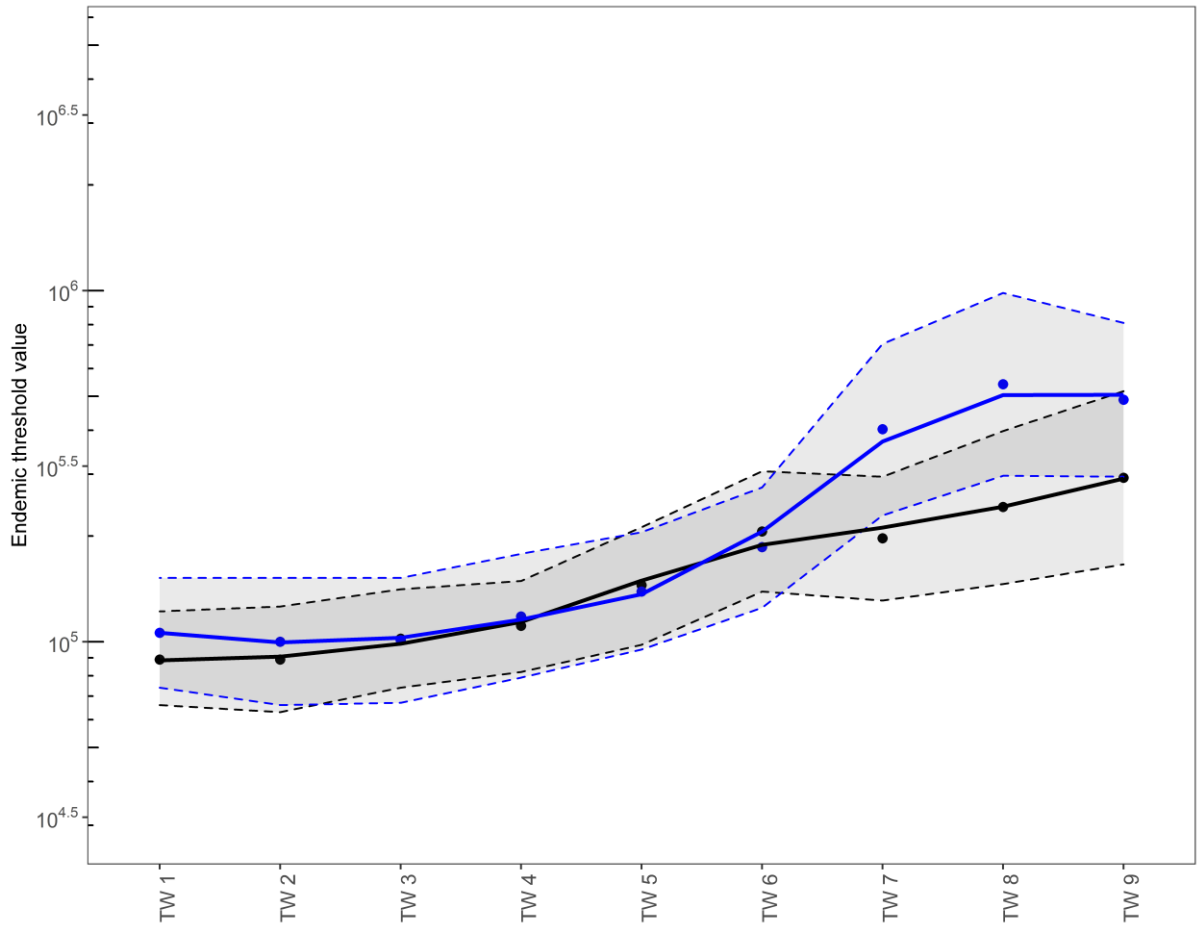

Figure 5.13 *Endemic threshold size estimates for scarlet fever in high density in Lancashire and South Wales for nine time–windows, 1940–1969. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals.*

South Wales: Among high density districts, the mean endemic threshold size is ~147,000. Estimates of endemic threshold population size rise monotonically across the time–windows, from 106,000 (95% CI: 74,000-152,000) in time–window one to 186,000 (95% CI: 130,000-322,000) in time–window six. In time–windows seven to nine, the rate of growth in the endemic threshold size increases (Fig. 5.13), peaking at 541,000 (95% CI: 297,000-985,000) in time–window eight, before decreasing marginally in the final time–window.

5.3.4 Low Connectivity Districts

Lancashire: Among low connectivity districts, the mean endemic threshold population across the nine time–windows is ~157,000. In time–window one, the estimated endemic threshold population size is approximately 74,000 (95% CI: 57,000-97,000). The endemic threshold population size steadily increases across successive time–windows (see Fig. 5.14), peaking at 257,000 (95% CI: 163,000-403,000) in time–window nine.

South Wales: The mean endemic threshold population size is ~426,000. Growth of estimates of endemic threshold population size in low connectivity districts follow a similar trajectory as low density districts (see Fig. 5.12) and the wider regional endemic threshold population (Fig. 5.14). In time–window one, the endemic threshold size is estimated to be 149,000 (95% CI; 101,000-220,000). The endemic threshold population falls marginally in time–window two, before growing steadily to 260,000 (95% CI: 158,000-428,000) in time–window six. In time–windows seven to nine, the rate of growth in the endemic threshold size is notably greater and peaks at 1,195,000 (95% CI: 485,000-2,965,000) in time–window eight, before falling in the final time–window.

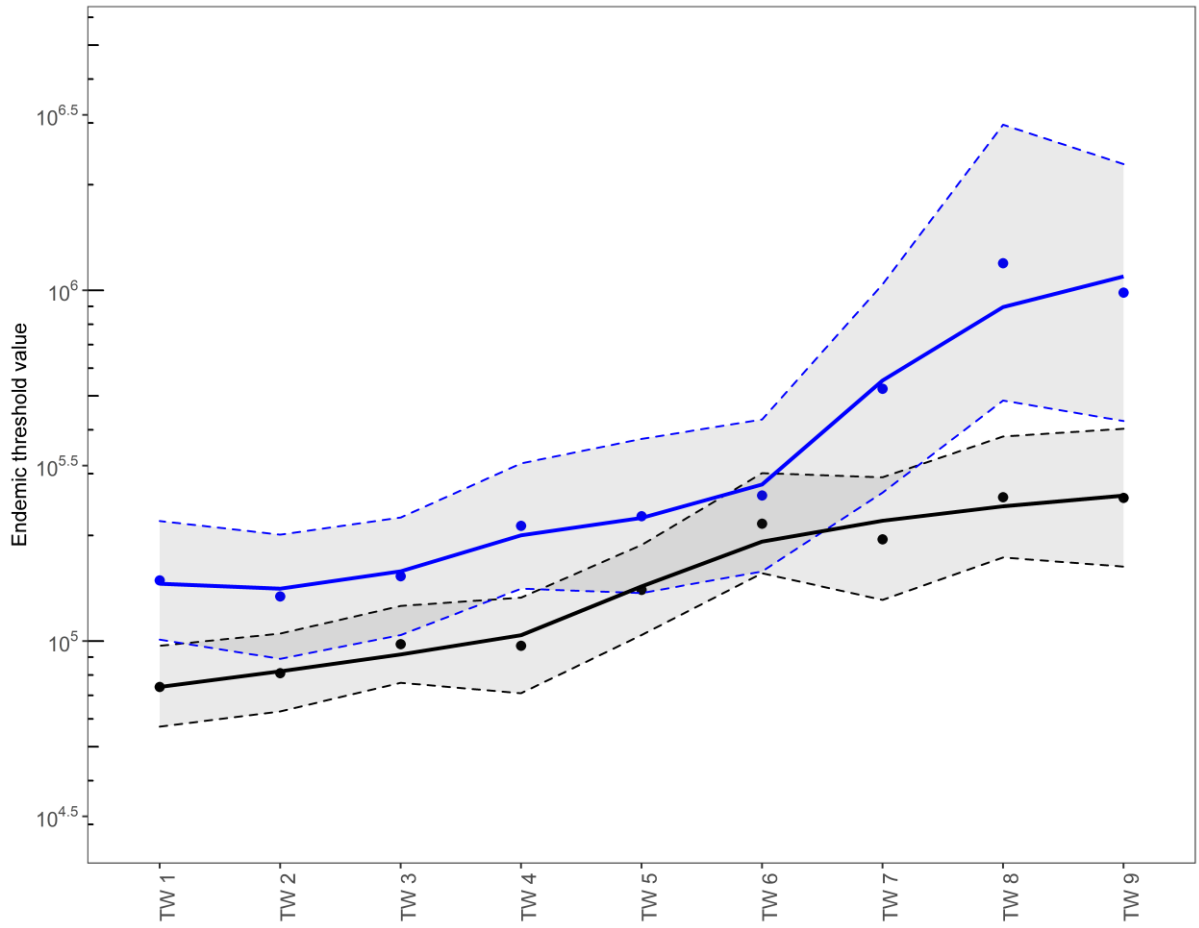

Figure 5.14 *Endemic threshold size estimates for scarlet fever in low connectivity districts in Lancashire and South Wales for nine time–windows, 1940–1969. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals.*

5.3.5 High Connectivity Districts

Lancashire: The mean endemic threshold population across the nine time–windows is ~130,000. In time–window one, the estimated endemic threshold population size is approximately 74,000 (95% CI: 50,000-109,000). The endemic threshold population size steadily increases across successive time–windows (see Fig. 5.15), with only a negligible fall in time–window four, peaking at 218,000 (95% CI: 127,000-374,000) in the final time–window.

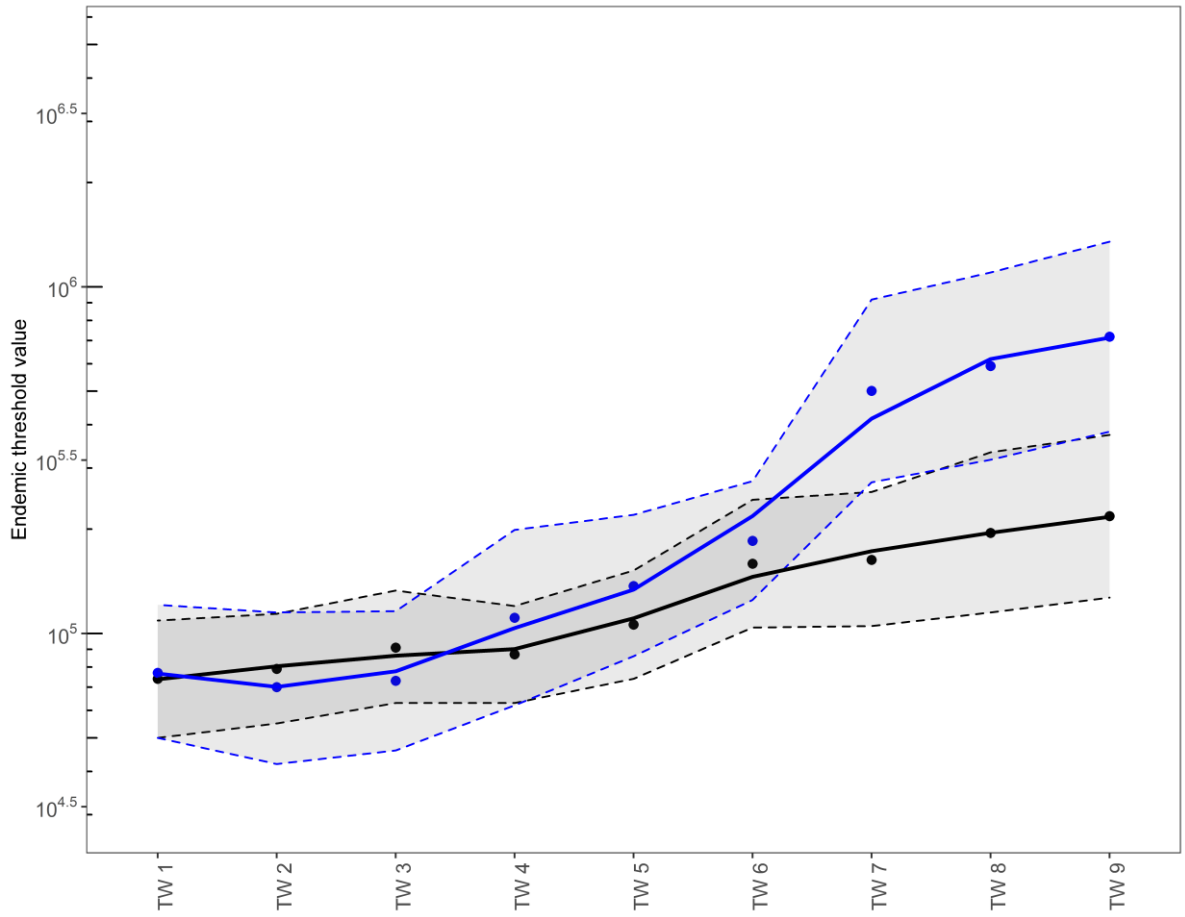

Figure 5.15 *Endemic threshold size estimates for scarlet fever in high connectivity districts in Lancashire and South Wales for nine time–windows, 1940–1969. Black = Lancashire, blue = South Wales. Dotted lines represent the 95% confidence intervals.*

South Wales: Among high connectivity districts, the mean endemic threshold population size is ~274,000. Estimates of endemic threshold population size rise monotonically across the time–windows, from 77,000 (95% CI: 50,000-121,000) in time–window one to 185,000 (95% CI: 125,000-275,000) in time–window six. In time–windows seven to nine, the rate of growth in the endemic threshold population size increases substantially (Fig. 5.15), peaking at 718,000 (95% CI: 382,000-1,350,000) in time–window nine.

On the following page, Tables 5.1 and 5.2 provide a concise statistical summary of the temporal changes in the estimated size of endemic threshold populations for the three childhoods infections in Lancashire and South Wales over the course of the study period.

Table 5.1 *Temporal changes in the endemic threshold population size for measles, pertussis and scarlet fever in the Lancashire region, 1940–1969.*

Disease	Time window one (1940-45)			Time window nine (1964-69)			[+/-]
	Estimate (in 000s)	95% CI		Estimate (in 000s)	95% CI		(in 000s)
		Lower	Upper		Lower	Upper	
Measles	263	215	322	229	191	275	-34
Pertussis	125	92	169	461	332	641	+336
Scarlet fever	71	57	89	213	154	294	+142

Table 5.2 *Temporal changes in the endemic threshold population size for measles, pertussis and scarlet fever in the South Wales region, 1940–1969.*

Disease	Time window one (1940-45)			Time window nine (1964-69)			[+/-]
	Estimate (in 000s)	95% CI		Estimate (in 000s)	95% CI		(in 000s)
		Lower	Upper		Lower	Upper	
Measles	1,180	834	1,650	1,025	726	1,444	+155
Pertussis	293	173	498	1,720	1,114	2,843	+1,427
Scarlet fever	94	71	125	637	406	1,000	+543

5.4 Geographical Patterns of Endemicity: Lancashire

For reference, details on the districts discussed in this section, their geographical location within their wider region and in relation to other districts, can be found in Appendix I (Figures I.1 and I.2).

5.4.1 Measles

Across all nine time–windows, the plots of the regional endemic threshold values reveal indicate relatively stable estimates for measles in Lancashire, with slight downward trend in later windows, particularly in high density and high connectivity time–windows. Fig. 5.16 provides a visual display of percent endemicity for each individual district in the region, with a single choropleth map for each time–window. The figure paints a picture of the regional dynamics metapopulation dynamics which aids interpretation of the relative stability of the endemic threshold population for measles across the nine time–windows.

The largest urban centres, Manchester CB and Liverpool CB, are fully endemic across the study period, experiencing no fadeout events. Salford CB, which immediately neighbours Manchester CB, and the districts of Bootle CB, Huyton-with-Roby UD and Crosby MB, that lie adjacent to Liverpool CB, all exhibit persistence levels of 80% and above. Other districts which display constantly high levels of measles persistence are the regional centres of Burnley and Blackburn CBs in Pennine Lancashire, Preston CB lying in the centre of the region, Blackpool CB on the North Lancashire coast, and Oldham CB on the edge of South Lancashire.

From the plots in Fig. 5.1, a slight increase in the endemic threshold population for measles is detected between time–windows one and two, with the increase greatest among low density, low connectivity and high density districts (see Figs. 5.2, 5.3 and 5.4). The highest

levels of measles endemicity re concentrated within the urban centres of the Greater Manchester and Merseyside conurbations across all nine time–windows according to the patterns displayed in Fig. 5.16, with Liverpool CB and Manchester CB functioning as endemic reservoirs of infection.

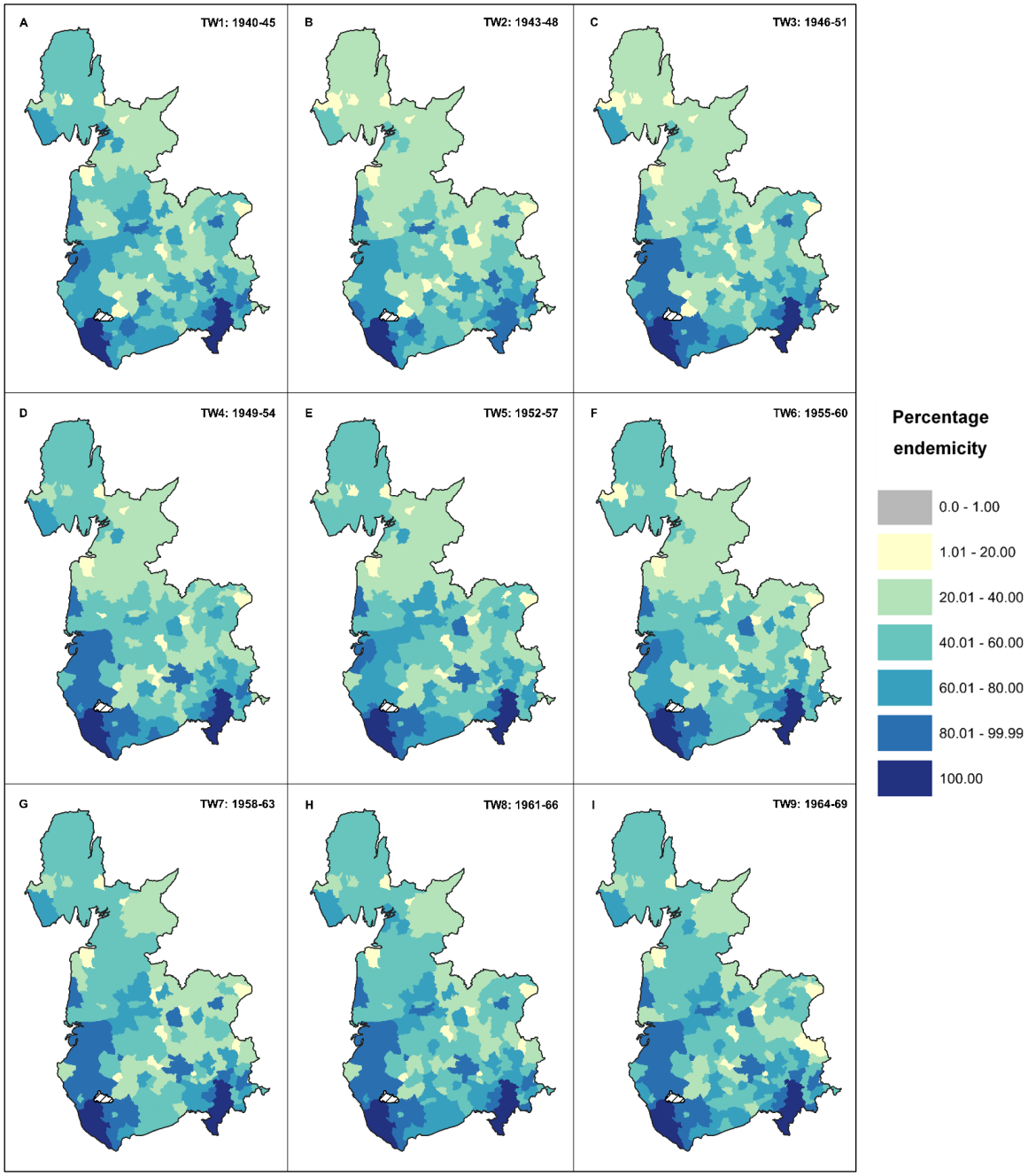

Figure 5.16 *Temporal changes in percentage endemicity of measles in the Lancashire region, by district population, across nine time–windows (1940–1969).*

Only 11 of the 125 districts in the Lancashire region during time–window two had a percentage endemicity of greater than 80%. The densely populated municipal boroughs and urban districts surrounding these settlements, such as Ashton-in-Makerfield UD, Eccles MB and Droylsden UD all saw a slight fall in endemicity. Compared to the first-time–window, there was a reduced level of persistence in districts across much of the northern half of the region, with a decline in measles endemicity in Lancaster MB, Morecambe & Heysham MB, Fleetwood MB in the north-west, and in districts situated in Pennine Lancashire in the eastern portion of the region, such as Colne MB and Burnley RD.

From time–window three onwards (Fig. 5.16C), there is increased measles persistence is visible in the Merseyside area, with percentage endemicity rising above 80% in numerous districts surrounding Liverpool including West Lancashire UD and RD, Crosby MB, Bootle CB, Southport CB and Whiston RD. Across successive time–windows, most notably time–windows seven to nine, measles persistence is consistently high, above 80%, in Southport CB and West Lancashire RD, south to Liverpool CB and immediately adjacent districts of Bootle CB and Whiston RD and stretching out to St Helens CB at the Eastern edge of the Merseyside metropolitan area. Like the higher levels of measles persistence in the Merseyside area over the course of the study period, the percentage endemicity of measles in districts in the Greater Manchester area rose steadily across the nine time–windows.

The rise in measles endemicity in districts located in around the Merseyside area is mirrored by increasing levels of endemicity in the more sparsely populated and pastoral districts of North and Central Lancashire, specifically in the Fylde peninsula surrounding Blackpool CB, and South Ribble, in the districts neighbouring Preston CB. The gradual increase in measles persistence in the more sparsely populated and distant districts of Central and North Lancashire may explain the fall in the endemic threshold population for measles from time–window six onwards in low density and low connectivity districts, coinciding with the sharp rise in birth rates and population growth in the latter half of the study period.

In the north-west corner of Lancashire lies Barrow-in-Furness CB, the most isolated sub regional population centre and county borough in the region, a travel distance of 81 km north of Liverpool CB, and 97km north-west of Manchester CB. Despite being surrounded by the coastline and rural, sparely populated districts, Barrow-in-Furness CB manages to sustain relatively high levels of measles persistence, with cases notified between 60-80% of reporting weeks in six of the nine time–windows. Lower levels are visible in time–windows two, five and six, coinciding with regional epidemic troughs and lower fertility rates.

5.4.2 Pertussis

In the pre-vaccine era time–windows (see Figs. 5.17A–E), pertussis endemicity tends to be concentrated within the two dominant urban centres of the Lancashire region, Liverpool CB and Manchester CB, and major sub-regional population centres including Blackpool CB, Burnley CB, Blackburn CB, and Preston CB. In rural districts and sparsely populated urban districts in North, Central and Pennine Lancashire on the eastern edges of the region, pertussis persistence is generally low, with less than 40% of reporting weeks notifying cases of the disease. Within these areas there are pockets of heightened disease activity and thus higher levels of persistence, normally found in mid-sized towns with populations of ~50,000, such as Barrow-in-Furness CB and Lancaster MB in the more isolated north-western portion of the region surrounding Morecambe Bay, and Colne MB in the Pennine Lancashire. Another notable feature is the intermediate levels of pertussis persistence in rural districts in West and Central Lancashire most likely sustained by frequent reintroduction events from sub-regional population centres with elevated epidemic activity, such as Preston CB, and overspill from the Merseyside conurbation and commuting between settlements in the hinterland and the larger population centres. Unsurprisingly, elevated levels of pertussis persistence were consistently found in the districts adjacent to the two endemic centres of Liverpool and Manchester CB during the pre-vaccine era.

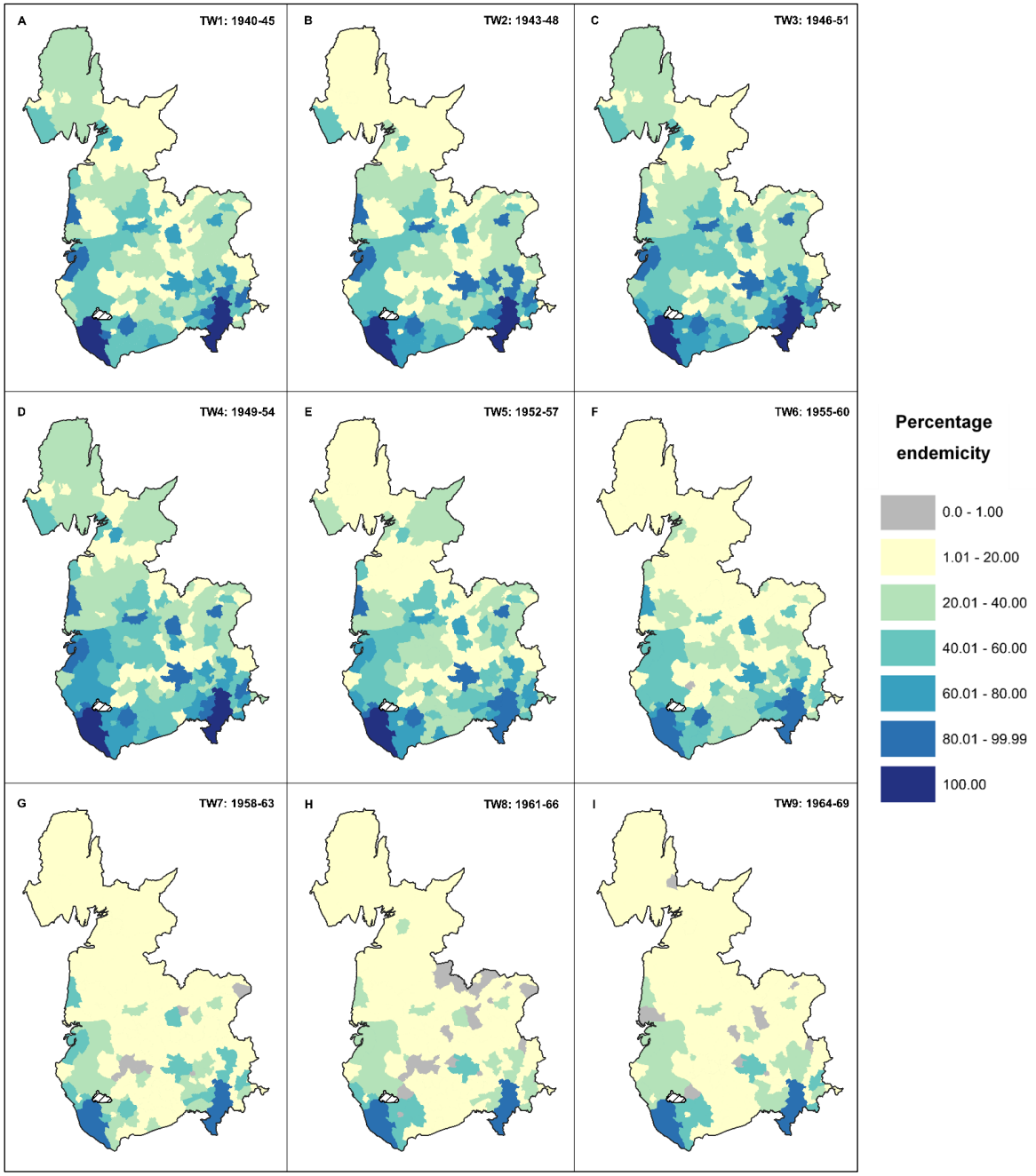

Figure 5.17 *Temporal changes in percentage endemicity of pertussis in the Lancashire region, by district population, across nine time–windows (1940–1969).*

Time–window four (Fig. 5.17D) saw the highest levels of pertussis endemicity overall across the region's 125 districts, coinciding with the high birth rates at the beginning of the baby boom period in the late-forties and early-fifties, with significant activity in Liverpool CB and surrounding districts, including West Lancashire RD, Bootle CB, and Whiston RD. This pattern was also mirrored in the Greater Manchester area, with high pertussis persistence

found in Manchester CB which acts as an endemic reservoir for the disease, and very high levels of endemicity in Salford CB and Stretford MB districts which share boundaries with Manchester CB to the east. There are also high levels of persistence in districts in the metropolitan internals of the greater Manchester and Merseyside conurbations.

With the advent of mass vaccination for pertussis by time–window six (Fig. 5.17F), the spatial impact of immunisation on pertussis persistence is almost immediate. Only two districts, outside the endemic centres of Liverpool and Manchester CBs report cases of infection in more than 80% of reporting weeks: Salford CB, which is essentially locked into the disease activity of its adjacent neighbour Manchester CB, and St Helens CB in Merseyside. In the previous time–window (Fig. 5.17E), six districts outside of the two city regions had endemicity levels greater than 80%. Relatively high level of persistence remained in Blackpool, Blackburn, Bolton and Bootle CBs, as well as smaller populations such as Middleton MB, Eccles MB and Huyton-with Roby UD where disease spread is heavily influenced by close proximity and consequent commuting to Liverpool and Manchester CBs. However, in the very next time–window (Fig. 5.17F), the pattern of pertussis persistence which defines the vaccine-era has already taken shape. There is a drastic fall in incidence and persistence across the region as immunisation removes significant numbers of susceptibles from local subpopulations. Vaccine uptake has been sufficiently high to vastly reduce endemicity of pertussis infection in major towns beyond South Lancashire that tend to play crucial roles in transmitting infection to the most peripheral districts on the northern and eastern edges of the region, such as Blackpool CB, Preston CB and Burley CB.

Despite a relatively large population of around ~60,000 during the vaccine-era, Barrow-in-Furness CB reports rare reintroduction events with significant periods of disease fadeout, with less than 20% of reporting weeks notifying cases. In the case of pertussis, the isolation of the large town of Barrow from the metapopulation network of large towns within the

southern and central portion of the region has an analogous effect as the small populations in rural and urban districts in the rest of the region which are much closer in travel distance to but do not neighbour or are not located within the Merseyside or Greater Manchester metropolitan areas. This lies in significant contrast to measles, with Barrow-in-Furness CB sustaining relatively high levels of measles persistence, with cases notified between 60-80% of reporting weeks in six of the nine time–windows (see Fig. 5.16). One possible explanation for these different behaviours is that pertussis is less transmissible than measles (Kretzschmar et al., 2010). Estimates of the basic reproduction number (R_0) for measles are very high, at around 20 (Wallinga et al., 2003), compared to 5-6 for pertussis (Kretzschmar et al., 2010), but the disease has a shorter infectious period than pertussis. Thus, infected individuals with pertussis produce fewer secondary infections during the infectious period on average, hence much lower levels of pertussis persistence are observed. Another possible explanation may be the underreporting of pertussis cases due to routine misclassification of cases of infection .

Although Manchester CB and Liverpool CB are of sufficient size and density to maintain susceptible populations large enough to sustain circulation of pertussis during the 1960s, vaccine uptake significantly reduce the scale of infection being exported to neighbouring areas with vaccination drastically reducing incidence of pertussis cases. Only four districts outside of Manchester and Liverpool CBs (Bolton CB, St Helens CB, Huyton-with-Roby UD, Whiston RD) consistently report intermediate levels of pertussis persistence, meaning more frequent reintroductions of disease and fadeout events with reduced duration, across the time–windows which fully cover the period of mass vaccination (Fig. 5.17G-I).

5.4.3 Scarlet fever

The general temporal pattern of scarlet fever endemicity in Lancashire districts across the study period has strong similarities with pertussis endemicity. Scarlet fever persistence is highest in the time–windows at the beginning of the study period in the 1940s, before falling progressively across succeeding time–windows, with a notably steep decline in endemicity from the late 1950s onwards. Geographically, the highest levels of scarlet fever persistence are found among districts situated within the Merseyside and Greater Manchester areas, with the urban cores of Liverpool and Manchester CB at the centre. Despite lower population densities and local populations of varying size, high levels of scarlet fever endemicity are found in districts immediately bordering Liverpool and Manchester, such as Crosby MB, Bootle CB and Salford CB, and other districts situated within a 25km radius of the two conurbations. Beyond the urban agglomerations found in the Southern portion of the Lancashire, scarlet fever persistence is found to be concentrated within the sub-regional centres of Burnley CB, Blackburn CB, Preston CB and Blackpool CB, stretching east-to-west across Pennine Lancashire, South Ribble and to the coast of the Fylde peninsula.

In time–windows one to four (see Fig. 5.18A, B), relatively high levels of scarlet fever persistence are notable within the sparsely populated, rural districts of Central Lancashire and West Lancashire, centred around the Ribble area, as well as in Barrow-in-Furness CB, the most distant county borough and Large town from the two endemic centres in Southern Lancashire, and around the Morecambe Bay area, in Morecambe and Heysham MB and Lancaster MB. Time–window two experiences the highest levels of scarlet fever persistence more generally across the Lancashire region; scarlet fever endemicity exceeds 80% of reporting weeks for all county boroughs with the exception of Barrow-in-Furness CB.

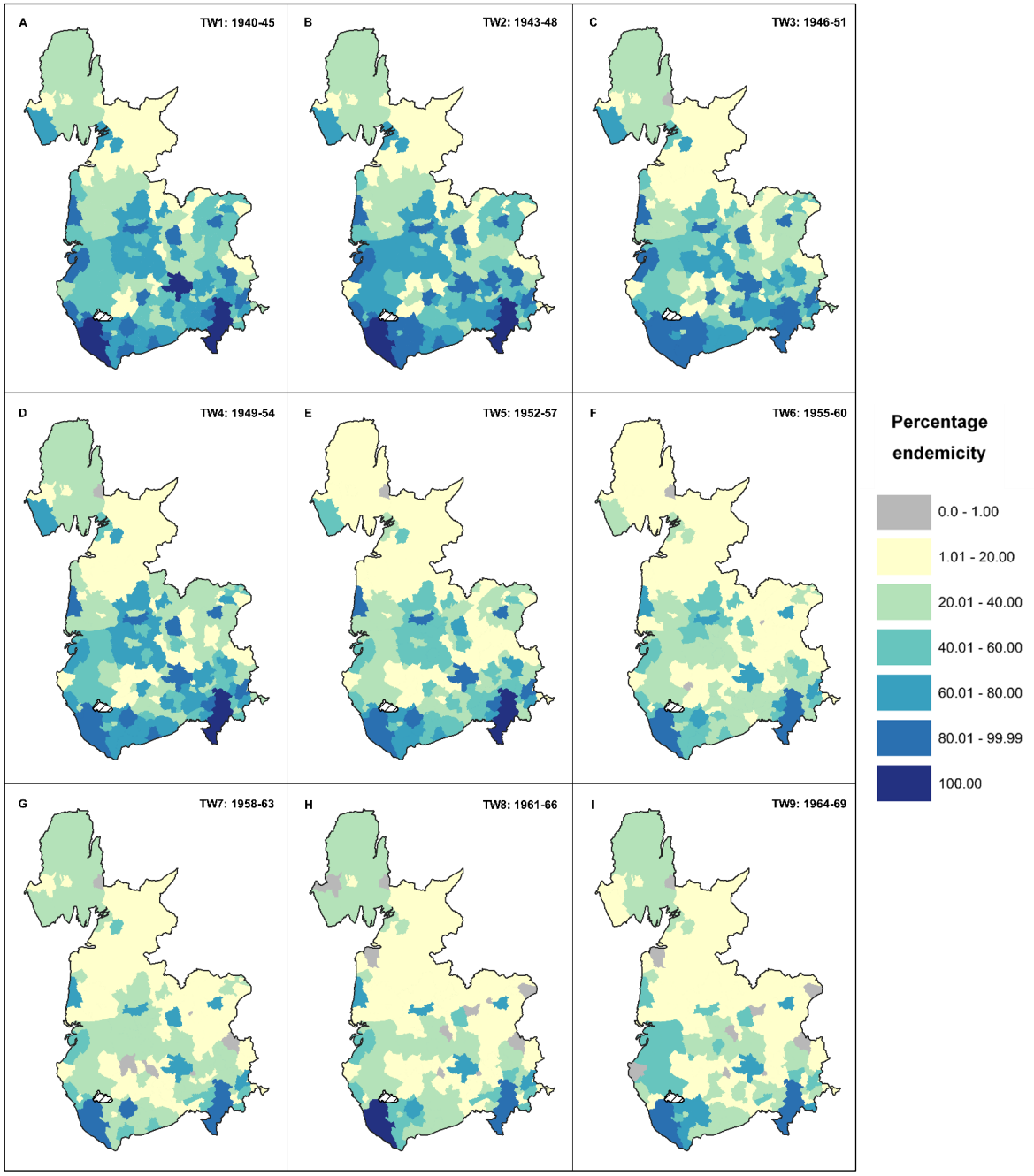

Figure 5.18 *Temporal changes in percentage endemicity of scarlet fever in the Lancashire region, by district population, across nine time–windows (1940–1969).*

In time–window three, there is a notable retreat in scarlet fever persistence in the immediate post-war period in rural districts with highly dispersed population located in the more pastoral areas of the Lancashire region, such as Pennine Lancashire, Central Lancashire, North Lancashire, and the Fylde area. Coinciding with the progressive decline in scarlet fever incidence across successive time–windows during the study period, scarlet fever

endemicity becomes increasingly confined within larger population centres, specifically the major industrial towns situated in South Lancashire.

By time–window six (1955-1960), scarlet fever endemicity begins to notably decline in dense, heavily-populated urbanised districts with over 100,000 inhabitants, with endemicity falling below 80% of reporting weeks for the first time in the study period, with the exception of Manchester and Liverpool CBs. These districts include Oldham, St Helens, Salford, Bolton, Burnley, Preston and Blackpool CBs. Scarlet fever persistence in Central Lancashire and the surrounding districts around Preston CB, which border Wigan and Bolton CBs, also begins to drop sharply over the course of the final four time–windows. In the final three time–windows (1958-63, 1961-66 and 1964-69), Liverpool and Manchester CBs remain endemic centres for scarlet fever, from which disease is reintroduced regularly to immediately neighbouring districts and more sporadically to districts which form the metropolitan hinterlands and extend to the urban periphery of West Lancashire, Central Lancashire and the Rossendale Valley.

With the exclusion of Liverpool and Manchester CBs, Bolton and Burnley CBs are the only districts which maintain relatively high levels of scarlet fever endemicity across the study period, with 60–80% of reporting weeks containing a case notification for the disease. Within the regional context, both districts are quite large, with populations of approximately 100,000 inhabitants. Nevertheless, the high level of scarlet fever persistence is unexpected when one considers the distance between these districts and the endemic centres. Despite Bolton CB lying 16km northwest of Manchester CB, its reports higher levels of scarlet fever persistence than more densely populated districts located closer to Manchester CB. One would expect that these districts would share a more significant degree of interaction with Manchester CB than Bolton CB. The heightened levels of scarlet fever endemicity in Bolton may be indicative of intrinsic variation in the recruitment rate of susceptibles at the local level. This sustains autoregressive epidemic behaviour, generating localised outbreaks.

In contrast to Bolton CB, Burnley CB is situated considerably further away from Manchester CB, located 34km north in Pennine Lancashire in the eastern portion of the region and one would assume that, following the logic of Tobler's first law of geography (Tobler, 1970), the enhanced distance between the endemic centre of Manchester CB and Burnley CB, would result in a much low level of scarlet fever endemicity, particularly given the very low incidence of scarlet fever infection by the final time–window and Burnley CB's population size, which at ~100,000 is well-below the estimated endemic threshold population for scarlet fever persistence in time–window nine (see Table 5.1). A possible explanation is Burnley's role as a key sub-regional population and economic centre, significant interactions with surrounding local populations from which people commute may result in a form of positive feedback loop. Additionally, Burnley CB's role as a gateway district in Pennine Lancashire may influence higher levels of scarlet fever persistence due to edge effects. This complex mosaic of spatial interaction between Burnley CB, surrounding local communities and large populations across the regional border could result in a greater volume of transmission events and fuel higher levels of endemicity within the district.

5.5 Geographical Patterns of Endemicity: South Wales

5.5.1 Measles

Fig. 5.19 displays the percent endemicity of measles in individual South Wales districts across the nine time–windows reveal the highest levels of disease persistence in the urban cores Cardiff CB, Swansea CB and Newport CB, as well as Rhondda MB, which consists of numerous hamlets, villages and small towns in two densely populated valleys. High levels of measles persistence are also visible in the urban periphery of the Cardiff and Swansea metropolitan areas, in districts such as Port Talbot MB, Penybont RD and Bridgend UD, which are situated between the two county boroughs. Across the districts in Pembrokeshire,

Carmarthenshire, Northwest Glamorgan and the western half of Monmouthshire, significantly lower levels of measles endemicity are present, alluding to the much more sparsely populated, isolated, and rural nature of the districts lying in these areas.

Cardiff CB, and to a lesser extent Swansea and Newport CBs, serve the South Wales region as endemic reservoirs of infection for all three childhood diseases across the nine time–windows. These densely populated districts act as persistence hotspots. In the context of measles, disease diffuses from these districts to immediately surrounding areas, most notably towns at the edge of the Vale of Glamorgan and the coastal plain such as Port Talbot MB, and settlements at the foot of the peri-urbanised Valleys such as Caerphilly UD, Pontypridd UD and Pontypool UD. Outside of the urban centres on the coastline, measles endemicity is most concentrated within the county of Glamorgan. Levels of measles endemicity appear to decline in line with population size along the settlement hierarchy.

Beyond the county borders of Glamorgan, there is a significant drop in measles endemicity in Pembrokeshire, Carmarthenshire, and eastern Monmouthshire. Measles persistence is most frequently recorded below 20% of the reporting weeks in time–windows for in Cemaes RD, Pembroke RD and Narberth RD in Pembrokeshire, as well as isolated urban districts in Carmarthenshire and Pembrokeshire which historically have functioned as small market towns and key sub-regional centres of connectivity in the more sparsely populated Western half of the South Wales region. These districts tend to have populations in the region of 10-15,000 inhabitants. These include Llandilo and Llandovery UDs in North Carmarthenshire, and the coastal settlements in Pembrokeshire, such as Tenby, Milford Haven and Neyland MBs. Periodic increases in measles endemicity are notable in time–windows one (Fig. 5.19A), two (Fig. 5.19B), and eight (Fig. 5.19H) in Carmarthen RD, neighbouring Llanelli RD, and in Haverfordwest RD, which may be due to an increase in the susceptible pool and replenishment rate during these time–windows, with time–windows two and eight coinciding with the two peaks of the baby boom period

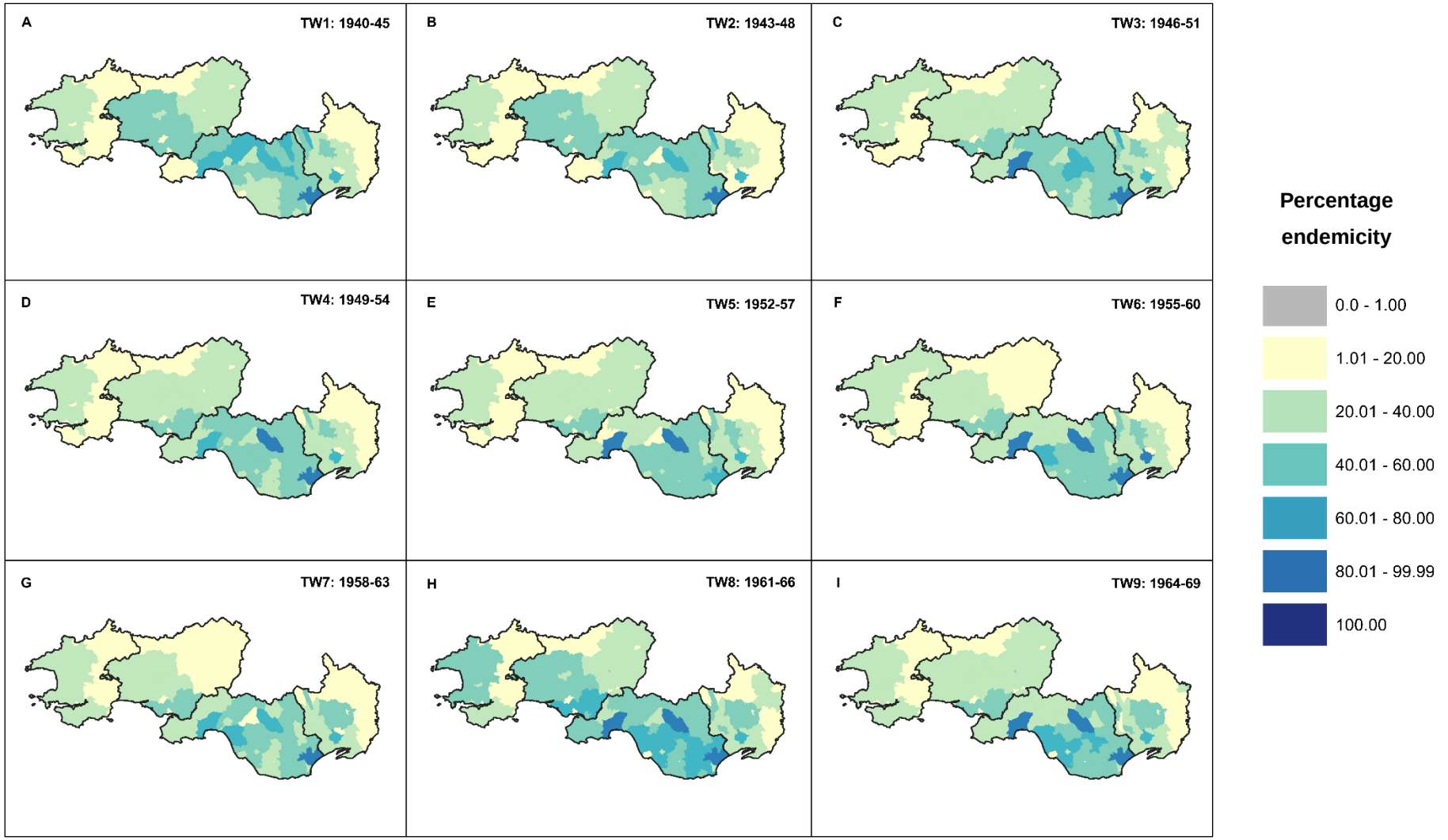

Figure 5.19 *Temporal changes in percentage endemicity of measles in the South Wales region, by district population, across nine time–windows (1940–1969).*

5.5.2 Pertussis

Levels of pertussis persistence are generally lower across all districts in the South Wales region during pre-vaccination era time–windows than observed for measles infection (see Fig. 5.20A–E), which may partially be explained by the greater level of underreporting associated with pertussis and lower infectivity of the disease compared to measles (Clarkson and Fine, 1985). Prior to the onset of mass vaccination, Cardiff and Swansea CBs are key endemic centres of infection in which pertussis is sustained by consistent rates of susceptible recruitment and density, ensuring chains of transmissions do not fade out and consistent mixing between infected and susceptible individuals. Newport CB also features as a prominent district in terms of pertussis endemicity, above 80% of reporting weeks with minimal disease fadeout during time–windows three, four and five (Fig. 5.20C–E). Newport CB is only 19km northeast of Cardiff, sharing a strong functional relationship with Cardiff CB and surrounding areas, such as Barry MB and Cardiff RD.

In the initial three time–windows (Fig. 5.20A–C), relatively high levels of endemicity are notable in the densely populated mining district of Rhondda MB and Ebbw Vale UD, at the head of the valleys in North Monmouthshire/Glamorgan, with pertussis cases notified between 60-80% of reporting weeks. Port Talbot and Neath MBs in Glamorgan and the districts of Llanelli MB and RD in Southeast Carmarthenshire, lying adjacent to and surrounding Swansea CB, show intermediate levels of pertussis endemicity with cases reported in 40-60% of reporting weeks in time–windows three, four and five (Fig. 5.20C–E). For Port Talbot and Llanelli MBs, these levels of persistence are an increase on the levels displayed in time–windows one and two (Fig. 5.20A & B). In all other districts which comprise the county of Carmarthenshire, very low levels of pertussis persistence are registered; below 20% of reporting weeks in each time–window in the pre-vaccination and vaccine eras. This is not unexpected since Carmarthenshire covers a large proportion of the rural and sparsely populated hinterland of the South Wales region.

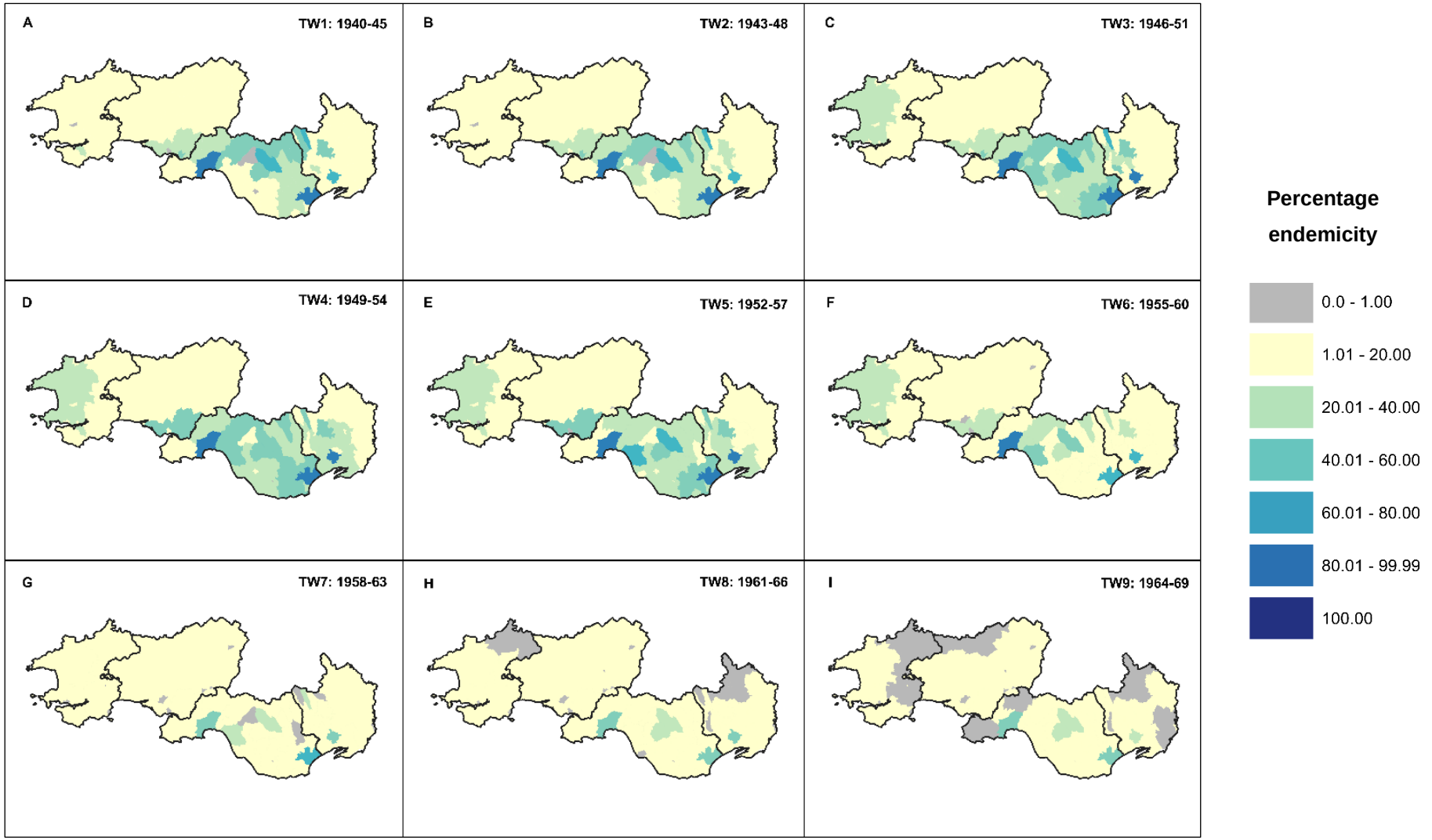

Figure 5.20 *Temporal changes in percentage endemicity of pertussis in the South Wales region, by district population, across nine time–windows (1940–1969).*

The largest settlements in the county, outside of Llanelli MB, tend to be small market towns with populations below 15,000, and many districts feature highly dispersed populations inhabiting small, isolated communities.

In Pembrokeshire, elevated level of pertussis endemicity are observed in Haverfordwest RD in time–windows three to six (Fig. 5.20C–F), and in Pembroke MB, in time–windows one, four, five and six (Fig. 5.20A, D– F). All districts in Monmouthshire, in the eastern portion of the South Wales region, record low levels of pertussis endemicity across the nine time–windows, with no district reporting pertussis endemicity in more than 20% of reporting weeks. The low levels of pertussis persistence in Monmouthshire districts, such as Usk UD, Abergavenny MB, and Monmouth UD may be associated with the significant rurality of the county, resulting in considerable population dispersion and isolation of households in hamlets and small villages. The small populations in Monmouthshire towns were not of a sufficient size to maintain chains of transmission or generate epidemic activity without introduction from an external source.

By time–window six (Fig. 5.20F), the effect of mass vaccination begins to be felt, with pertussis endemicity falling in the districts surrounding Swansea CB and Cardiff CB, in communities located in the Vale of Glamorgan and across the Valleys. In the remaining three time–windows (Fig. 5.20G–I), Pembroke MB is the only district in South Wales, outside of the county of Glamorgan, where pertussis endemicity levels are not below 20% of reporting week, despite its seemingly peripheral location on the western coastal edge of the region. This maybe largely due to the district's ferry port status, resulting in significant interaction between population flows from various locations and subpopulations, thus promoting spatial import of infection, and greater frequency of reintroduction events.

By time–windows eight and nine (Fig. 5.20H, I), both Cardiff and Swansea CBs begin to experience periodic fadeouts, indicating all districts in South Wales have populations below

the endemic threshold by the end of the study period. This is due to sufficient vaccine uptake to effectively break chains of transmission and deplete the susceptible pool from which epidemic activity is ignited and disease spread is sustained. Beyond the urban centres on the Glamorgan coast, Rhondda MB continues to exhibit intermediate levels of pertussis persistence with more frequent fadeout of infection. Other valley districts in Glamorgan such as Ogmore & Garw UD and Mountain Ash UD, neighbouring the more heavily populated Rhondda valleys and Merthyr Tydfil MB, also exhibit slightly higher levels of pertussis endemicity than the vast majority of districts in South Wales, although pertussis is still only notified in only 20-40% of reporting weeks, with vaccination effectively halving the proportion of weeks in which cases were reported in these areas during the pre-vaccine era. Of the 74 local government districts in South Wales, only seven districts register a percentage endemicity greater than 20% and are all located within the county of Glamorgan.

5.5.3 Scarlet Fever

Scarlet fever endemicity in South Wales follows a similar pattern of retreat across the nine time–windows of the study period (see Fig. 5.21), although the nature of decline is less dramatic than the trend thoroughly described for pertussis. Persistence of the scarlet fever infection is concentrated in the most densely and heavily populated country of the region: Glamorgan. In time–windows one to five (Fig. 5.21A–E), Cardiff CB, Swansea CB, Newport CB and Rhondda MB serve as key endemic centres for scarlet fever. From these population centres, waves of scarlet fever infection radiate out into the surrounding areas, such as the valley districts of Mountain Ash UD, Ogmore & Garw UD, and Merthyr Tydfil CB, alongside districts on the coastal belt of Glamorgan, including Barry MB and Port Talbot MB. Collectively these districts in time–windows one and two exhibit higher levels of scarlet fever persistence (Fig. 5.21A, B), with 60-80% of weeks reporting cases of scarlet fever. There are also elevated levels of endemicity in Central and North Carmarthenshire, in Llandilo RD

and Carmarthen RD, and the small towns of Carmarthen MB, Llandilo MB and Llandovery UD, indicating regular reintroduction of infection, most likely from the more populated, urban areas lying adjacent to the Southeast.

As time progresses, a gradual decline in scarlet fever persistence is evident in the Valley communities and districts immediately neighbouring urban centres of Cardiff CB, Swansea CB and Newport CB, and Rhondda MB. This parallels the falling incidence rates of scarlet fever across the study period (see Section 4.3, Chapter 4). Scarlet fever endemicity falls considerably following the first peak of the baby boom period in the late 1940s, between time–windows three and six. By time–window six (Fig.5.21F), high levels of persistence remain in the urban centres of Glamorgan along the coast, with diminishing levels of endemicity in surrounding districts and communities immediately north of Cardiff, a decline in persistence in Rhondda MB, and in Llanelli MB and RD in Southeast Carmarthenshire.

From time–window six onwards, there is a significant drop in scarlet fever persistence across Glamorgan and neighbouring districts in Carmarthenshire. There is an increased number of fadeout events in Cardiff CB and Swansea CB, suggesting a consistent increase in the endemic threshold population for scarlet fever across the study period. Nevertheless, persistence of the infection is still highest in the two urban centres, although percentage endemicity in time–windows seven to nine (Fig. 5.21G-I) for the two districts does not rise above 80%. Pembroke MB remains the one district outside of Glamorgan in South Wales where percent endemicity does not fall below 20% in the final three time–windows, which is also an increase in the persistence of scarlet fever in the district compared to the previous six time–windows, in which the percent endemicity did not rise above 20% (Fig. 5.21A-F). Across the districts that make up the rest of Pembrokeshire, Carmarthenshire and Monmouthshire, fadeout periods last many years, with only irregular and extremely short-lived reintroductions of disease, with the sparsely populated and highly dispersed populations in largely pastoral, rural districts unable to support chains of transmission.

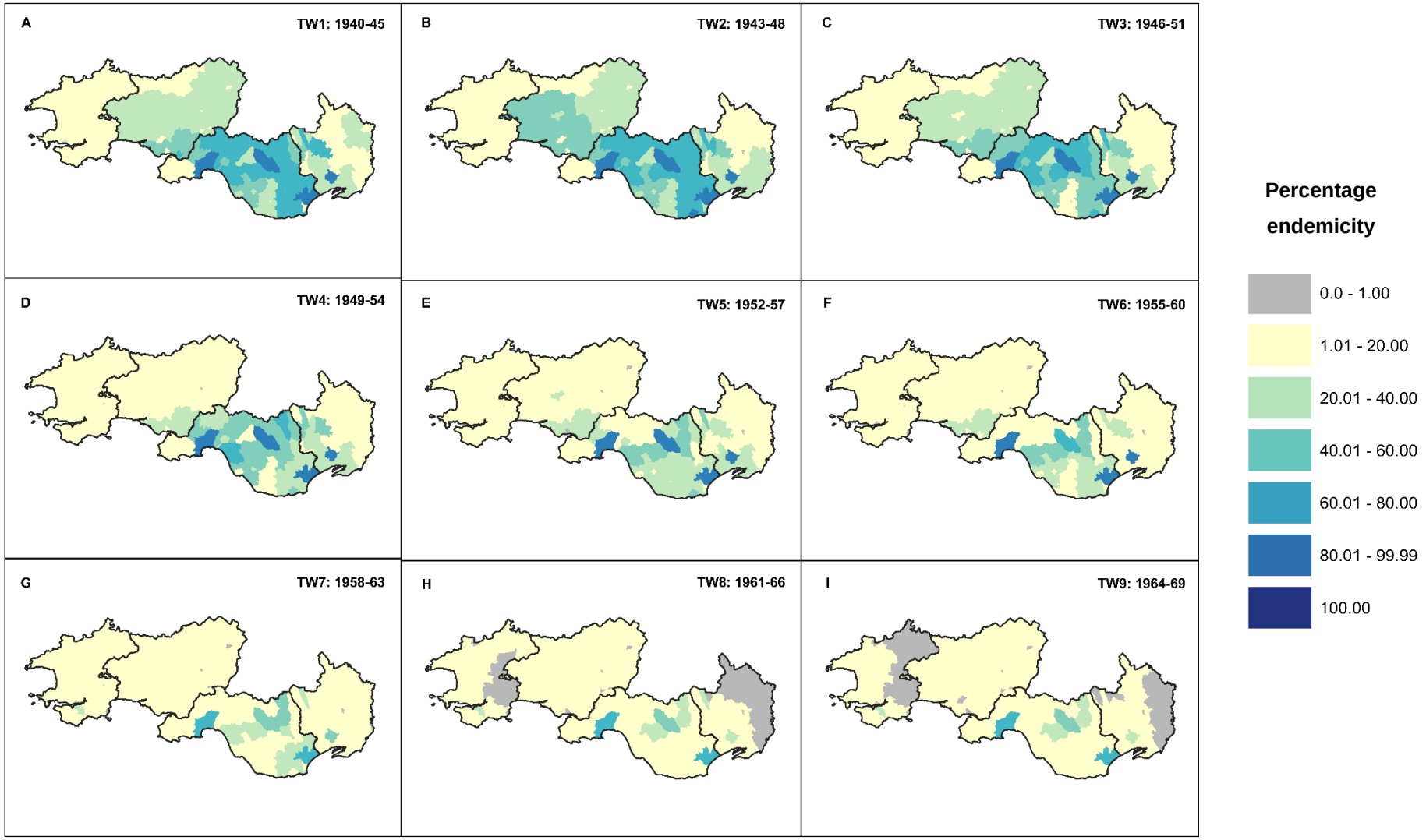

Figure 5.21 *Temporal changes in percentage endemicity of scarlet fever in the South Wales region, by district population, across nine time–windows (1940–1969).*

5.6 Discussion of Findings

In previous empirical work analysing endemic thresholds and the persistence of childhood infections, most prominently the study of measles endemicity nationally in England and Wales but also in island populations, a range of endogenous and exogenous factors have been cited as influencing levels of disease endemicity and thus playing an important role in shaping the size of endemic threshold populations.

With regards to endogenous factors, these include population density (Black, 1966; Cliff et al., 2000) and the size of the susceptible population (Bartlett, 1957, 1960), both of which affect the spacing of susceptibles within local subpopulations. The spatial distribution of susceptibles could fuel disease fadeouts if susceptibles are isolated in sparsely populated districts, often rural and peripheral within the context of a regional metapopulation, or nurture prolonged periods of epidemic activity, if individuals are spaced in way that enables the spread of infection without susceptibles being concentrated within one location. However, with term-time forcing, which concentrates children in settings such as primary schools (Hens et al., 2009; Rohani and King, 2010), one may expect the rapid exhaustion of susceptibles since a disease such as measles has a high level of infectivity (Anderson and May, 1991; Grenfell et al., 2002). Another key endogenous factor is birth rate, which fundamentally shapes the recruitment and replenishment rates of susceptible populations. Relatedly, fertility rate should also be seen as an important factor given the intimate link between the two, yet fertility rate is seldom considered explicitly as a measure for inclusion in stochastic modelling of childhood infections.

In terms of exogenous factors, of key importance with regards to limiting growth in endemic threshold populations is the potential for infection to be generated due to reintroduction, or ‘colonisation’ events from external sources. This is dependent on the scale of geographical connectivity within the wider regional metapopulation, as determined by networks of mobility

and the spatial distribution of local subpopulations in the wider region. Internal mobility and geographical connectivity within regions have been routing cited as an important factor behind disease persistence, serving as a driver of local disease spread but also of rescue effects. These have been previously cited as an important mechanism for maintaining sufficient levels of spatial coupling to maintain disease endemicity in metapopulations (Bolker and Grenfell, 1996; Metcalf et al., 2013b), thus restricting endemic threshold population growth. The effective implementation of disease intervention, either through vaccination, use of antibiotics or improvements in hygiene in local subpopulation is another important factor for curtailing disease spread, restricting infection of susceptibles and thus in-turn reducing disease persistence.

Many of these factors shall now be discussed within the context of the spatiotemporal changes observed in endemic threshold population size across the study period for the regional metapopulation of Lancashire and South Wales, and the changing nature of pertussis, measles and scarlet fever endemicity.

5.6.1 Stochasticity & Vaccination

The impact of extrinsic variation in birth rates on disease endemicity is demonstrable in the moving–window approach for both regional populations, with heightened levels of measles persistence coinciding with peaks in the fertility and consequent birth rate, most notably in time windows two and three (1943-48, 1946-51) and time windows seven and eight (1958-63, 1961-66) (see Figs. 5.1 & 5.2). Pertussis endemicity is notably greater at the beginning of the postwar baby boom period before falling sharply on local and regional scales towards the end of this period, coinciding with the second and final spurt in birth rate in the early 1960s (see Figs. 5.17 & 5.20). This reflects the onset of widespread, routine, vaccination which breaks down chains of transmission by diminishing the rate of susceptible recruitment, analogous to lowering birth rates, and caseloads, thus increasing the endemic

threshold. The introduction of vaccination and high levels of coverage sustained sees the geographical retreat of pertussis infection in both regions, falling back to the endemic centres of Liverpool CB, Manchester CB and Cardiff CB which maintain sufficiently large pools of susceptible to maintain infection.

In South Wales, the two largest urban centres, Cardiff CB and Swansea CB, are approximately one-third of the population size of the two main endemic centres in Lancashire during the study period, Manchester CB and Liverpool CB. Both Cardiff and Swansea lack sufficient size and reach to swamp and dictate the endemic dynamics of the whole region beyond the densely populated county of Glamorgan where the two urban centres are situated, on the coastal plain. It is worth noting that an important control on the spatial dynamics of childhood infection spread is the average size of the host populations and the degree of heterogeneity in size across settlements (Bartlett, 1957; Grenfell et al., 2001). Phase differences between subpopulations increase as peripheral districts within a regional metapopulation decrease in population size, which can result in asynchronous outbreaks which continues to fuel low levels of persistence in more rural areas. As demonstrated by consistently larger endemic threshold populations calculated for pertussis and measles in low connectivity districts in Lancashire and South Wales (see Figs. 5.4 & 5.9), demographic stochasticity coupled with a high degree of internal isolation in the deep troughs between epidemics ensures significant fadeout of infection in isolated populations, even when situated within a larger regional metapopulation in which childhood infections such as pertussis and measles are endemic (Conlan and Grenfell, 2007).

It is worth noting that there are other epidemiological parameters which may be important for the determination of the size of endemic threshold populations. These include the transmission term, the host life expectancy and the latent period of the infection (Dietz 1982), all parameters absent from the regression models fitted to estimate endemic threshold size in the nine time–windows in each region.

5.6.2 Population Decentralisation & Depopulation

A complicating factor in understanding local patterns of disease endemicity is undoubtedly the effects of population decentralisation, and internal migration within the Lancashire and South Wales regions. In the postwar period, these phenomena coincided with a significant decline in industrial activity which historically had been integral to structuring the distribution of the regional population. Furthermore, attempts to tackle significant overcrowding and slum housing in the urban cores of large towns and cities across region, by constructing municipal housing estates on newly acquired land at the outskirts of the town in cities in neighbouring districts, also had a profound impact on the demographic profile of multiple urban and rural districts across the region during the study period. These factors are important to consider within the context of disease endemicity, since they impact rates of susceptible recruitment, density and the spatial distribution of susceptible populations on a local and regional level, thus holding significant influence over the geographical persistence of infection.

Despite high birth rates across much of the study period ensuring a continuous rate of susceptible replenishment in the two large endemic centres, Liverpool CB and Manchester CB both experienced significant depopulation during the baby boom era. This did not affect the endemic persistence of measles, scarlet fever or pertussis (prior to vaccination) in Liverpool CB and Manchester CB (See Figs. 5.16 & 5.18), due to the large population size and susceptible replenishment rate in these settlements, as well as the constant feedback in disease transmission from their metropolitan hinterlands. However, the depopulation of these two urban cores did have a profound impact on neighbouring and surrounding districts. For instance, in the case of Liverpool CB, thousands of individuals and households were moved to growing suburbs in the metropolitan hinterland as slums were cleared, in the outer boroughs of Merseyside. Destination districts included Widnes MB, St Helens CB, Bootle CB, Southport CB as well as part of West Lancashire RD. Moreover, Liverpool City

Council built and owned large several ‘New Town’ council estates (Farmer and Smith, 1975), to which tens of thousands of people were moved to from Liverpool’s inner districts which form its core. Growing satellite districts in the metropolitan hinterlands of South Lancashire were also popular destinations for those leaving the economically devastated former mill towns of Central and East Lancashire. Internal migration in the Lancashire region during the 1950s and 1960s undoubtedly helped to change the spacing of susceptible individuals, redistributing susceptible populations in the wider regional metapopulation, sustaining high levels of persistence in the metropolitan districts of the South and eroding the endemicity of childhood infections within increasingly depopulated areas of North, Central and East Lancashire.

Mirroring the demographic fallout of industrial decline in Lancashire, South Wales witnessed significant depopulation in the historically more densely populated mining districts of the Valleys in Glamorgan and West Monmouthshire, such as Rhondda MB, Ebbw Vale UD, Merthyr Tydfil CB, Pontypool UD, and Aberdare UD, among others. Surplus labour migrated from the Valleys to the metropolitan hinterland consolidating around Cardiff, Swansea and Newport on the coastal plain or externally to England and places further afield. Concurrently, significant transformations in the nature of agricultural work in the postwar period (Bowen, 1960; Davies et al., 2008) and consequent decline in employment also fuelled rural depopulation in peripheral districts in South Wales. For instance, districts in the pastoral areas of East Pembrokeshire (such as Cemaes Rd and Narberth RD), and in North and West Carmarthenshire (Carmarthen RD, Newcastle Emlyn RD) all experienced significant declines in their population size, further diminishing susceptible pools, low rates of susceptible recruitment and increasing the levels of dispersion between susceptible individuals. Surplus labour migrated either to the metropolitan hinterland or urban cores located on the coastal plain of Glamorgan or to locations external to the region. These trends helped to fuel further reductions in the persistence of childhood infections in large areas of the region, with the gradual decline of already limited susceptible populations and

rates of susceptible replenishment, thus stimulating significant growth in the endemic threshold populations for measles and pertussis in low-density and low connectivity districts.

5.6.3 Mobility & Connectivity

It is notable that growth in endemic threshold populations for pertussis and measles is significantly more limited across the study period when contrasted with the overall regional pattern of growth in threshold estimates for the two diseases, in particular pertussis, due to the introduction of mass vaccination by 1958, as well as in comparison to low connectivity districts. Visual analysis of endemicity patterns for pertussis reveal higher levels in districts immediately neighbouring Manchester CB and Liverpool CB during the vaccine-era time windows, with pertussis endemicity mostly concentrated in the two endemic centres and elevated levels in satellite district which form part of the wider city regions, stretching into West Lancashire in the case of Liverpool CB, and in the Rosendale valley with regards to Manchester CB. In the Lancashire region, there is much greater connectivity between major population centres and satellite settlements found in the metropolitan hinterland and urban-rural periphery, which is reinforced in later time windows as the textile and manufacturing industries collapsed in many urban settlements in South and East Lancashire, which producing a transient population in search of economic opportunities within the Manchester CB and Liverpool CB conurbations.

Elevated levels of pertuusis endemicity in excess of 20% of reporting weeks during the vaccine era time –windows are observed in the sub-regional population centres of Preston CB, Burnley CB and Blackpool CB which act as gateway districts for sustained pertussis persistence in areas which are home to more rural and sparsely populated settlements and district. Large towns such as Burnley are of sufficient population size and density to maintain significant levels disease persistence once infection has been introduced from external sources, such as via long range transmission from the endemic centres of the South. In the

case of Blackpool CB, its popularity as a family coastal resort during the postwar period may act to inflate the persistence of measles and pertussis and encourage greater spatial import of infection from transient individuals. This would explain the high levels of measles endemicity, and to a lesser extent pertussis endemicity, across all nine time –windows.

Mobility in South Wales is limited by both the physical geography of the land but also by the significant spatial inter-dependence between valley communities in the regional hinterland and the urban centres situated on the coastal zone of Glamorgan. Much of the economic activity which drives the South Wales economy is concentrated within the two prominent urban centres on the coastal plain (Cardiff CB and Swansea CB) and in historic valley mining districts, which historically have been home to heavy industries centred on coal, copper and iron ore as well as steel manufacture. Although Cardiff CB and the Valleys were mutually dependent from the outset, the nature of this relationship had changed radically by the postwar period, with economic flows from the Valleys to urban centres on the coast decreasingly associated with products in search of an export market and increasingly of people in search of a labour market. Cardiff CB also has a close functional relationship with immediately surrounding towns, specifically the likes of Barry MB, Pontypridd UD and Caerphilly MB. Links between the city and the hinterland focused on the movement of labour hint at the importance of commuting not just as a key factor in population mobility but in the propagation of disease activity as carriers of infection.

Nevertheless, close functional relationships between Cardiff CB, Swansea CB and satellite settlements located in the Valleys or on the coastal plain of Glamorgan seemingly produce a highly connected, networked sub-regional population. This can be viewed as a ‘mainland’ population with the much more sparsely populated, lesser connected, and predominantly pastoral districts of Carmarthenshire, Monmouthshire and Pembrokeshire. This mainland grouping in the centre of Glamorgan is able to sustain higher levels of pertussis endemicity compared to districts in other countries at least until the introduction of mass vaccination

(see Fig. 5.20). In many ways, this historical and present economic co-dependency between communities of the Valleys of Glamorgan, and the major, urban, economic and administrative centres of Cardiff and Swansea on the county's coastal plain manifested itself epidemiologically, with the persistence of childhood infections fuelled by the high levels of spatial interaction between communities of varying size which surround, or are tightly connected to, the large populations that function as endemic centres for disease.

5.6.4 Spatial Structure

The more widespread distribution of medium-sized to large population centres, ranging in population size from 50,000 to 150,000, across the Lancashire metapopulation (in some cases over 100km away from the endemic centres), fuels a higher rate of disease introductions via capillary connections to surrounding local settlements with highly dispersed, minor populations and rural districts. A notable example across the study period is Barrow-in-Furness CB. Despite the district's high degree of isolation, situated in the far north-west of the study region approximately 150 km away from Manchester CB and 170km away from Liverpool CB in terms of travel distance, Barrow-in-Furness CB is still able to sustain relatively high levels of persistence for all three childhood infections, despite biological differences in infectious periods. Barrow-in-Furness is of sufficient population size sustain a rate of susceptible recruitment which allows the infections to propagate for prolonged periods in between transmission of diseases from external sources.

One would expect spatial coupling to be greatest around the most largely populated district as mobility and settlements spatially cluster within and around large cities; the size of the districts tends to decrease with distance from core cities (Grenfell et al., 2001). This is true of both the South Wales and regional metapopulations which explains the persistence of all three childhood infections in the urban centres of Liverpool CB, Manchester CB and Cardiff CB. Such coupling can also result in a commuter-driven agglomeration effect in which large

centres not only export disease to surrounding districts and settlements, but also import infections from surrounding local subpopulations (Bartlett, 1957; Viboud et al., 2006; Meyer and Held, 2014).

This is contrast to South Wales, where significant expanses of the region are heavily rural, agricultural and peripheral in nature, inhabited by isolated hamlets and villages with tremendously small populations which in some cases only number a handful of individuals (Jenkins, 2014). Notably, there are no population centres outside of the county of Glamorgan with a population greater than 50,000 inhabitants, with the exception of Newport CB, which clasps on to the county line between Glamorgan and Monmouthshire. For instance, no urban community or district in Pembrokeshire has a population greater than 15,000 inhabitants during the period of study. With much of the economic activity in Pembrokeshire, Carmarthenshire and significant portions of Monmouthshire unrelated to the industrial activities and heritage of district communities in Glamorgan, primarily centred on agriculture, settlements in this areas are more isolated and populations are generally less mobile, without the economic motivations that would encourage greater connectivity with the more highly populated districts which surround the urban centres of Cardiff and Swansea, and the densely populated communities located within the Valleys.

However, as noted in Sections 5.5.2 and 5.5.3, Pembroke MB seemingly shares a great deal in common with Barrow-in-Furness CB in Lancashire; two remote coastal towns with heightened rates of disease persistence beyond what would be expected in proportion to their population size and distance from the main regional urban centres from which epidemic waves of infection radiate. In the previous chapter, Sections 4.5.2 and 4.7.3, Pembroke MB stands outs as a pocket of stubborn localised persistence of pertussis infection, with disease activity negatively correlated with the average regional pattern observed. As previously mentioned in Section 4.5.3, Pembroke MB is home to a ferry port, which would have served as key gateway into South Wales and the United Kingdom from

Ireland during the mid-twentieth century before the arrival of cheap and accessible air travel, as well as an exit point from the region. Since the mean population size of Pembroke MB during the study period was ~10,000, the size at which one would expect long periods of disease fadeout and irregularly timed outbreaks, it seems very likely that the constant flow of individuals through Pembroke MB due to its status as a ferry port played a considerable role in elevating the levels of disease endemicity observed there.

This population structure results in a much more significant degree of spacing between susceptible individuals in the South Wales region compared to Lancashire, thus increasing the likelihood of breakdown for chains of disease transmission, reducing the duration of epidemic outbreaks and the number of weeks in which cases of childhood infections are reported. This ultimately leads to more prolonged fadeout periods in the majority of rural and urban districts in South Wales which are not located in Glamorgan and necessitates a larger threshold population to maintain disease persistence to compensate for the high levels of population dispersion and small local population sizes.

5.6.5 The Decline of Scarlet Fever

An almost identical upward trend in the size of the endemic threshold population estimates for both regions, districts of varying levels of population density and varying levels of connectivity, over the course of the study period is observed for one of the diseases: scarlet fever (see Figs. 5.11-15). Additionally, there is a progressive region-wide decline in scarlet fever endemicity in Lancashire and South Wales over the course of the nine time periods. Endemicity is still greater in Lancashire, with lower endemic threshold population estimate for the disease compared to South Wales, due to the greater levels of population density and higher rates of susceptible recruitment in the Lancashire region.

It is important to recognise both the morbidity and severity of scarlet fever had declined considerably by the beginning of the study period, compared to the start of the twentieth century when the disease was a leading cause of childhood mortality (Gale, 1945; McKinnon, 1946). The exact mechanisms for the diminishing caseload of scarlet fever up until the late 1950s, when the rate of decline suddenly becomes more precipitous, remain poorly understood. This is largely due to the dearth of research previously conducted in this area, despite the richness of morbidity data that is available in the Weekly Returns, perhaps reflecting the historical ack of public health attention paid to the disease after antibiotic treatments were introduced in the early postwar period. Perhaps one factor is improved nutrition, with the ending of rationing in the 1950s, recognising the findings of Duncan et al. (1996, 2000) which found the intensity of scarlet fever rose during periods of food shortages and high wheat prices during the latter half of the nineteenth century. Other factors include improvements in hygiene, decreased crowding as a result of population decentralisation and slum clearances in towns and cities, and introduction of antibiotics (Lamagni et al., 2018). It is possible that evolutionary drift facilitated the proliferation of less pathogenic strains may have been a deciding factor in the historical trajectory of the disease (Lamagni et al., 2018).

5.7 Chapter Summary

It is important here to recap the key findings presented in this chapter. For both regional metapopulations of Lancashire and South Wales, lower endemic threshold values are estimated across all nine time windows among high density and high connectivity districts, with the endemic threshold population consistently lower among these districts than the overall regional endemic threshold population values estimated for the full sample of Lancashire and South Wales districts. In the context of Lancashire, these low endemic threshold estimates in densely populated, highly-connected districts are indicative of high levels of spatial coupling between the Manchester and Liverpool conurbations and

surrounding urban districts, which fuels 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). Unsurprisingly, given the swamping effect of Manchester and Liverpool CB on the wider regional metapopulation, with a significant share of the overall regional population either residing or working within the two urban centres, endemic threshold population estimates are found to be consistently lower among all districts in the Lancashire region compared to South Wales, regardless of population density or connectivity.

Within the context of South Wales, the high endemic threshold values for time windows at the beginning of the study period visible for measles and pertussis in Sections 5.1 and 5.2 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 World War II, the post-war baby boom resulted in substantial growth in the number and density of susceptibles within districts of all sizes across South Wales, leading to a notable fall in the value of endemic threshold estimates.

As evidenced in Section 5.2, the introduction of routine pertussis vaccination nationwide from 1957 onwards is followed by significant growth in the size of the endemic threshold population growth in both regional metapopulations. 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. 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 urban districts in the Valleys or major population centres such as Cardiff CB. However, rescue effects restrict the growth of regional endemic threshold populations in the vaccine era (see Section 5.2), due to the constant reinfection of districts encouraging by tight spatial coupling between districts particularly in the southern portion of the region. 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. This finding emphasises the significant role spatial structure and the geographical distribution of subpopulations within a regional metapopulation can play in dictating the size of the endemic threshold population, even with the introduction of a powerful tool with which to breakdown chains of disease transmission. This finding also raises a key question. Namely to what extent does geographical connectivity affect the estimation of endemic threshold population size and impede efforts to control disease via vaccination? This question and aforementioned findings will be explored in greater detail in the following chapter, analysing hotspots of pertussis activity in the pre-vaccine and vaccine eras in the Lancashire and South Wales regions.

Chapter 6: Hotspot & Survival Analyses

6 Introduction

In the following chapter, findings are presented for a hotspot analysis of spatial coupling on pertussis persistence in the pre-vaccine and vaccine eras for Lancashire and South Wales. Survival analyses are conducted using a Cox regression model and Kaplan-Meier curves to analyse differences in fadeout survival probability and times of disease reintroduction for Lancashire and South Wales in the pre-vaccine and vaccine eras. A discussion of the findings and the insights they generate is provided at the end of the chapter. The analyses presented here has since been published, in the journal *Social Science & Medicine* (Munro et al., 2021).

6.1 Hotspot Analysis

6.1.1 Spatial Correlation & Coupling Patterns

Linear regression analysis reveals a negative association between TAFs and mean population size in Lancashire ($R^2 = 0.26$) (see Fig. 6.1A) and South Wales ($R^2 = 0.17$) (see Fig. 6.1B). 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. 6.2A, 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.

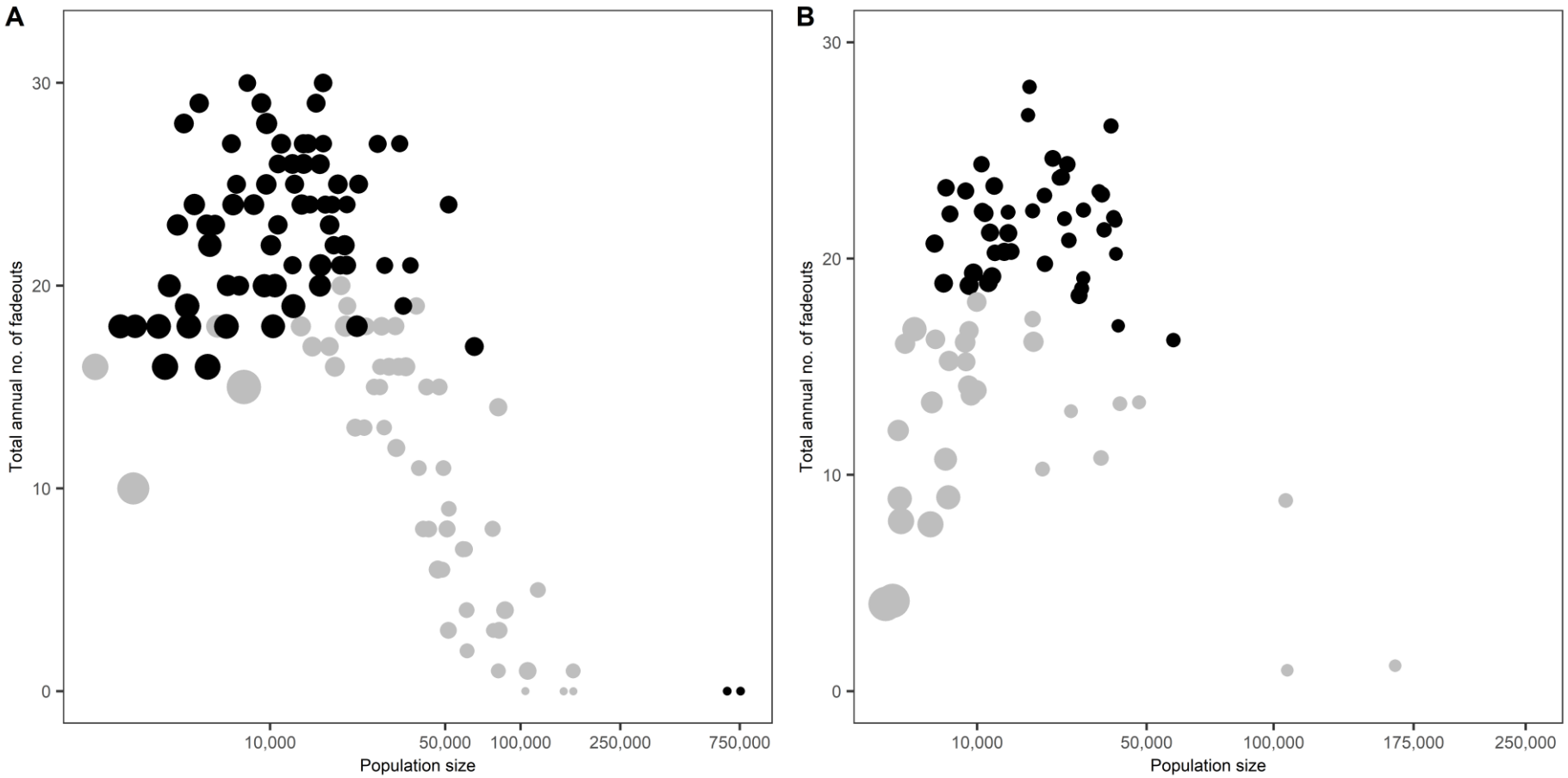

Figure 6.1 *Total annual number of fadeouts (TAFs) against mean population size for the pre-vaccine era in Lancashire and South Wales. Circle size increases with mean fadeout duration. Black = positive residuals, grey = negative residuals.* ***A****: Lancashire,* ***B****: South Wales.*

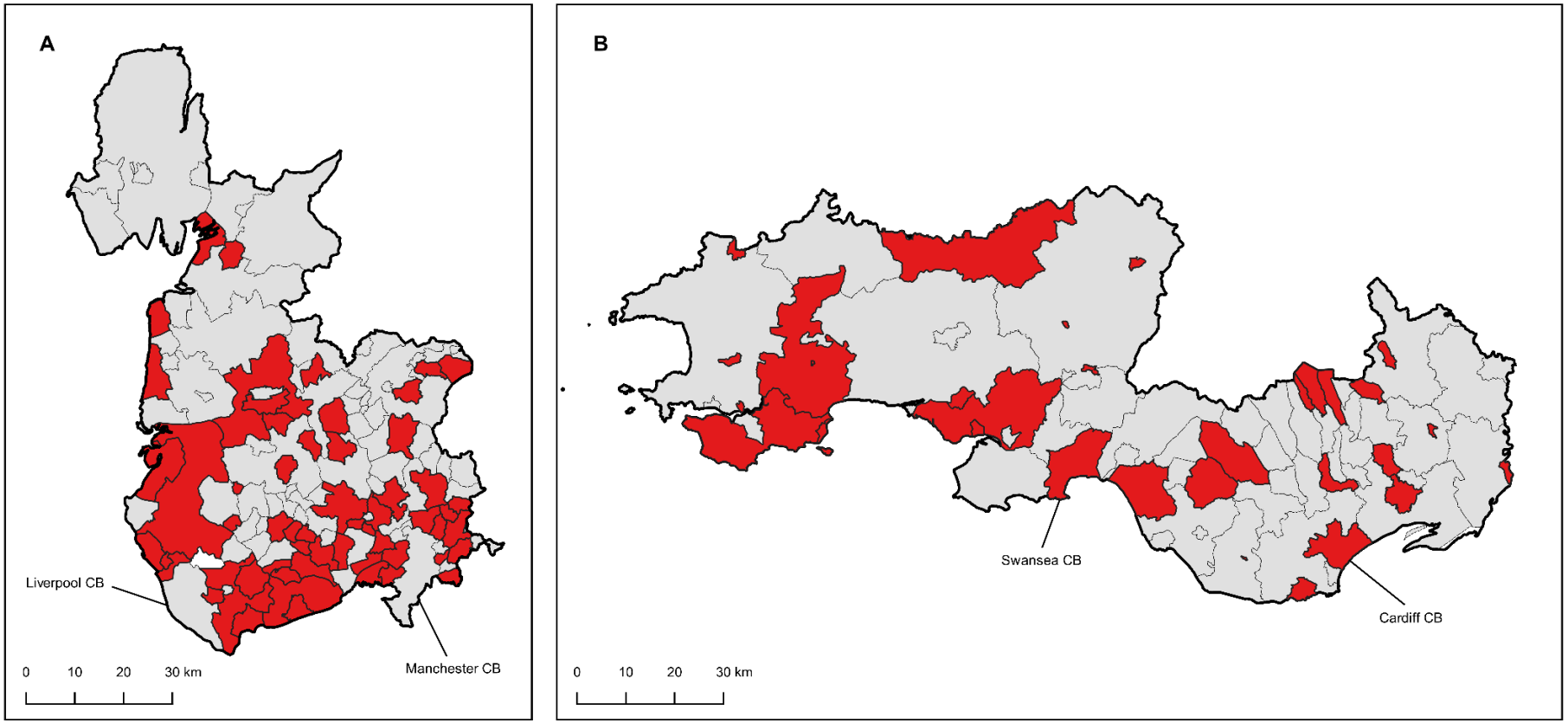

Figure 6.2 *Pre-vaccine era: Districts with higher rates of pertussis reintroduction in relation to population size in Lancashire and South Wales, identified by red shading ('hotspots'). **A**: Lancashire, **B**: South Wales.*

In South Wales, districts with inflated pertussis reintroduction rates are not concentrated spatially and are distributed unevenly across the region in Southeast Carmarthenshire, Southeast Pembrokeshire, and patches of 'the Valleys' in Glamorgan and Monmouthshire. In general, these districts represent small and medium sized urban centres and market towns (see Fig. 6.2B).

6.1.2 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 criterion. 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. 6.3A shows pertussis 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 in East and North Lancashire are also 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' area and include the largest population centres in South Wales (e.g., Cardiff CB, Swansea CB, Newport CB) and their neighbouring districts (Fig. 6.3B). Many of these districts contain key trunk roads and rail lines, important components of regional network connectivity.

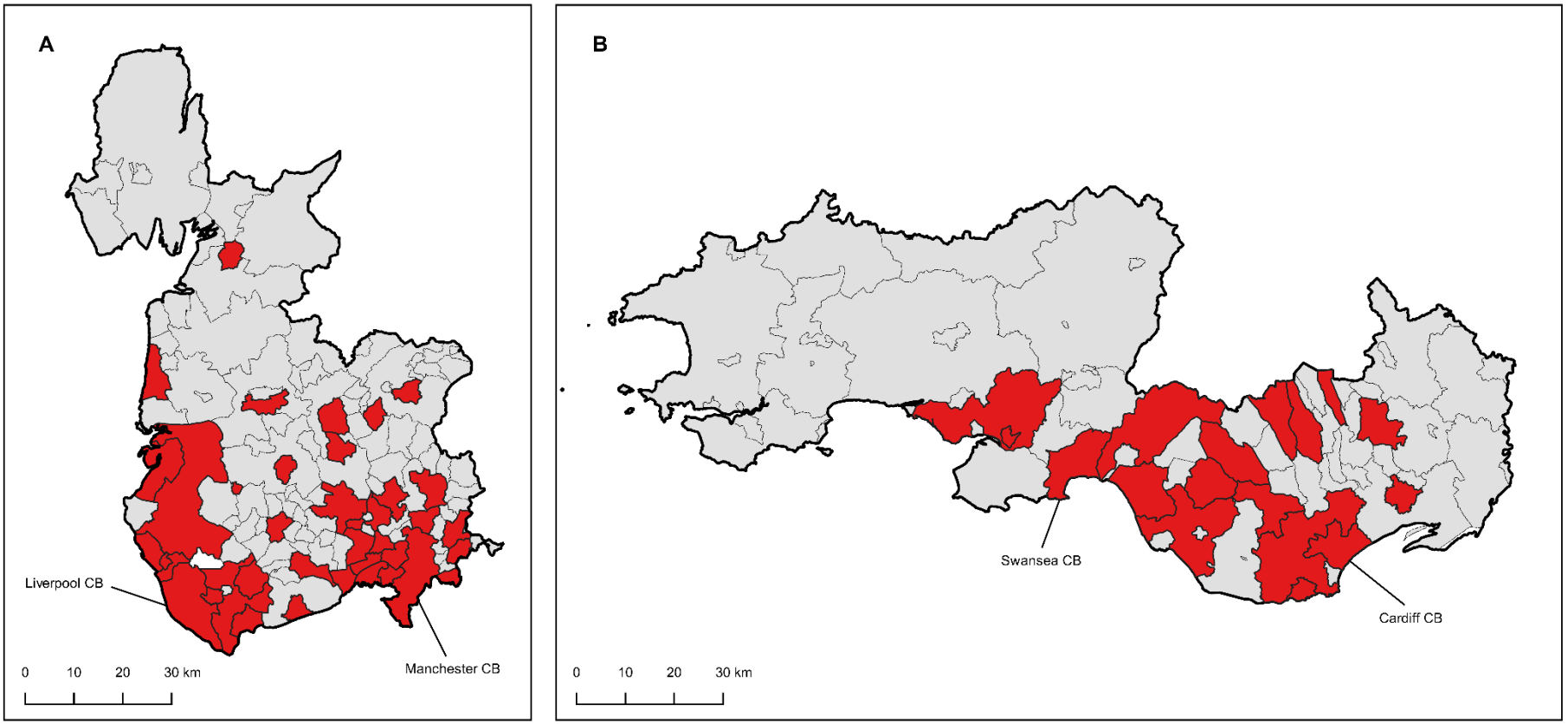

Figure 6.3 *Vaccine era: Districts reporting a higher number of pertussis cases than the regional mean in Lancashire and South Wales, identified by red shading (hotspots).* ***A****: Lancashire,* ***B****: South Wales.*

6.2 Cox Regression: 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 6.1). 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 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 6.1), 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).

Table 6.1 *Mean fadeout duration (in weeks) for hotspots and other districts in Lancashire and South Wales, pre-vaccine era (1946-1957) vs. vaccine era (1958-1969).*

	Mean fadeout duration (in weeks)				
	Pre-vaccine	**Vaccine**		**Pre-vaccine**	**Vaccine**
Lancashire			**South Wales**		
Hotspots (Mean)	1.96	2.73	**Hotspots (Mean)**	2.03	4.21
N	55	38	**N**	31	8
SD	1.4	1.4	**SD**	0.71	1.81
Other districts (Mean)	3.27	10.62	**Other districts (Mean)**	7.43	19.93
N	70	87	**N**	43	66
SD	2.21	7.91	**SD**	6.62	26.55

Table 6.2 *Cox regression model results: Rates of reintroduction in Lancashire and South Wales, pre-vaccine era (1946-1957) vs. vaccine era (1958-1969).*

		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**

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 6.2.

6.3 Survival Analysis: Fadeouts

6.3.1 Lancashire

The estimated mean time to disease reintroduction in Lancashire in the pre-vaccine era is 2.3 weeks for pertussis hotspots, as opposed to 3.2 weeks for other districts in the regional metapopulation. With the introduction of mass vaccination in 1957, there is marked change in the mean time to disease reintroduction in districts which are not defined as hotspots, rising to a mean of 9.6 weeks. However, there is only a negligible increase in the mean time to disease reintroduction in hotspots in the vaccine era, increasing to 2.8 weeks. The median survival time between the onset of a fadeout period and pertussis reintroduction is 2 weeks for hotspot districts compared to 2.6 weeks for districts in the pre-vaccine era. The difference in median survival time is much more prominent after the onset of vaccination,

growing three-fold, to 7.8 weeks, in districts other than hotspots. Similar to the trend in mean survival time, there was only a marginal increase in median survival time of fadeouts in hotspots in the vaccine era, rising to 2.5 weeks. A full breakdown of summary statistics for survival probability in hotspots versus other districts in Lancashire in the pre-vaccine and vaccine eras can be viewed in Table 6.3.

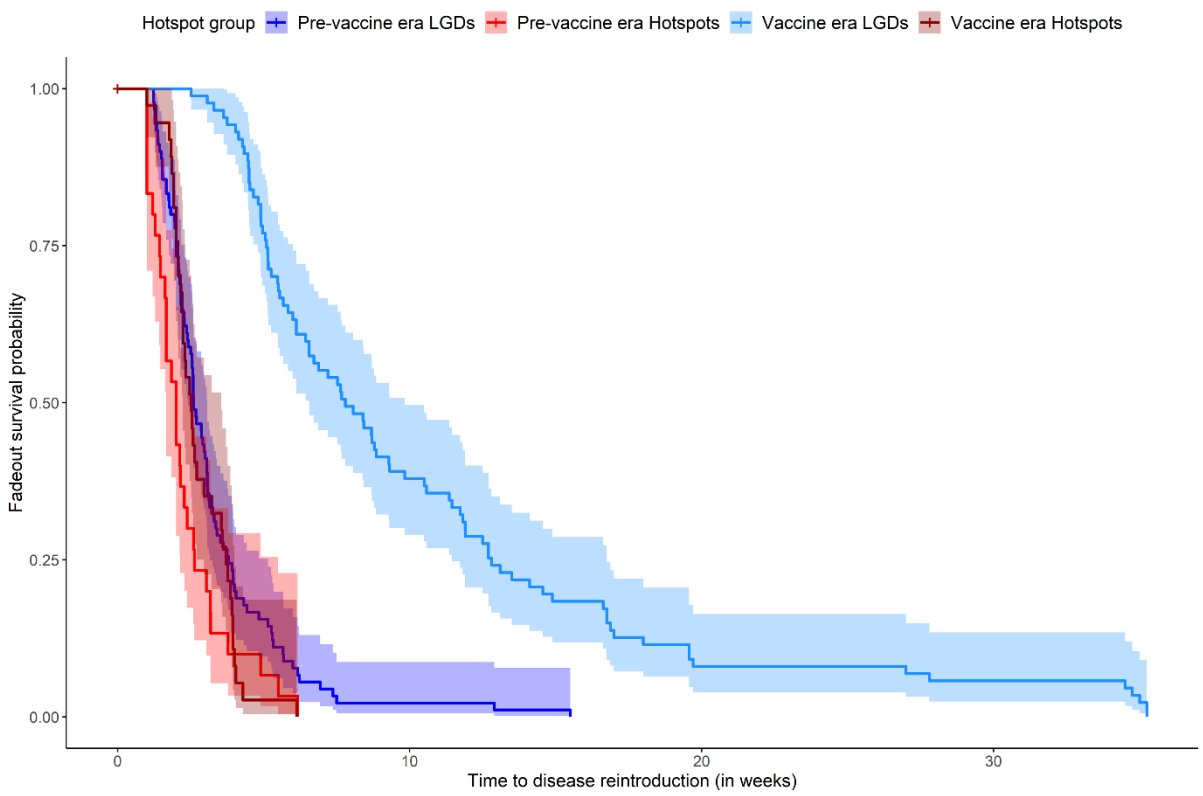

Figure 6.4 *Kaplan-Meier survival curves for the cumulative risk of reintroduction of pertussis in hotspots of infection and other districts in pre-vaccine era (1946-1957) and vaccine era (1958-1969) Lancashire. Shading indicates 95% confidence intervals (CI) on the Kaplan-Meier survival curves.*

Fig. 6.4 indicates clearly that fadeout survival probability is significantly greater among other districts in the vaccine-era that the pre-vaccine era, with a far lower time to diseases reintroduction. There is no difference in fadeout survival probability in hotspot districts after the introduction of pertussis vaccination in 1957.

Table 6.3 *Summary statistics for mean and median survival times for fadeout periods in pertussis hotspots and other districts in Lancashire, pre-vaccine era (1946-1957) vs. vaccine era (1958-1969).*

	Summary statistics					
					95% CI	
	N	Mean	S.E.	Median	Lower	Upper
Pre-vaccine era						
Hotspots	90	3.197	0.196	2.60	2.41	3.07
Other districts	35	2.292	0.237	2.00	1.67	2.60
Vaccine era						
Hotspots	87	9.643	0.581	7.80	6.56	9.83
Other districts	38	2.800	0.169	2.52	2.24	3.50

Log-rank tests were performed to determine if there are statistically significant differences in the survival distribution of hotspots and other districts in the pre-vaccine and vaccine eras in the Lancashire and South Wales regional metapopulations. The expectation was that the introduction of mass vaccination would significantly reduce the number of pertussis hotspots and increase the length of time until pertussis is reintroduced, once the disease has faded out, from an external source. This assume that mass vaccination effectively breaks down chains of transmission and thus increases the size of the endemic threshold population required to generate and sustain a sufficient susceptible population to maintain endemicity of infection (Broutin et al., 2005; Beyer et al., 2012).

The log-rank test is testing the null hypothesis that there is no difference in the overall survival distributions between the groups in the population. In this analysis, the following hypothesis is tested:

- H_0: There is no overall difference in the overall survival distribution of pertussis hotspots and other districts in the Lancashire region in pre-vaccine and vaccine eras.
- H_A: There is a statistically significant difference in the overall survival distribution of pertussis hotspots and other districts in the Lancashire region in pre-vaccine and vaccine eras.

Table 6.4 *Results of log-rank test for difference in survival between pertussis hotspots and other districts in Lancashire, pre-vaccine era (1946-1957) vs. vaccine era (1958-1969).*

	Pre-vaccine era			**Vaccine-era**		
	N	**Observed**	**Expected**	**N**	**Observed**	**Expected**
Other Districts	90	90	100.8	87	87	117.6
Hotspots	35	30	19.2	38	37	8.4
χ^2			7.4			144
df			1			1
***p*-value**			0.007			0.0005

The results of the log-rank tests in Table 6.4 indicate that there is a statistically significantly difference in survival between hotspots and other districts in the pre-vaccine era (χ^2= 7.4, $p < 0.05$). The number of observed hotspots exceeds the number expected by 11, from n =

19 to n = 30 hotspots. In the vaccine era, a statistically significantly difference in survival between hotspots and other districts is also detected (χ^2= 144, p < 0.05). Thus, the null hypothesis is rejected. The large size of the chi-square test statistic produced by the log-rank test indicates the data is a poor fit. Notably, only eight hotspots were expected in the regional metapopulation, yet 37 were observed.

6.3.2 South Wales

Table 6.5 *Summary statistics for mean and median survival times for fadeout periods in pertussis hotspots and other districts in South Wales, pre-vaccine era (1946-1957) vs. vaccine era (1958-1969).*

	Summary statistics					
					95% CI	
	N	**Mean**	**S.E.**	**Median**	**Lower**	**Upper**
Pre-vaccine era						
Hotspots	23	2.123	0.122	2.19	1.80	2.45
Other districts	51	6.806	0.607	5.32	4.55	6.78
Vaccine era						
Hotspots	17	4.471	0.378	4.39	4.23	6.06
Other districts	57	21.170	2.418	13.3	10.33	22.83

The estimated mean time to disease reintroduction in South Wales in the pre-vaccine era is 2.1 weeks for pertussis hotspots, as opposed to 6.8 weeks for other districts in the regional metapopulation. With the onset of vaccination, there is a significant change in the

mean time to disease reintroduction in non-hotspot districts, increasing in length by almost fourfold, to a mean of 21.2 weeks. The mean time to disease reintroduction in hotspots in the vaccine era doubles, but remains considerably shorter than other districts, increasing to approximately 4.5 weeks. The median survival time between the onset of disease fadeout and pertussis reintroduction is 2.19 weeks for pertussis hotspots, comparable to the median time for hotspots in Lancashire, while the median time for other districts in the pre-vaccine era is 5.3 weeks. In the vaccine era, the median survival time for fadeouts in pertussis hotspots is 4.39 weeks, compared to 13.3 weeks for other districts. A full breakdown of summary statistics for survival probability in hotspots versus other districts in South Wales in the pre-vaccine and vaccine eras can be viewed in Table 6.5.

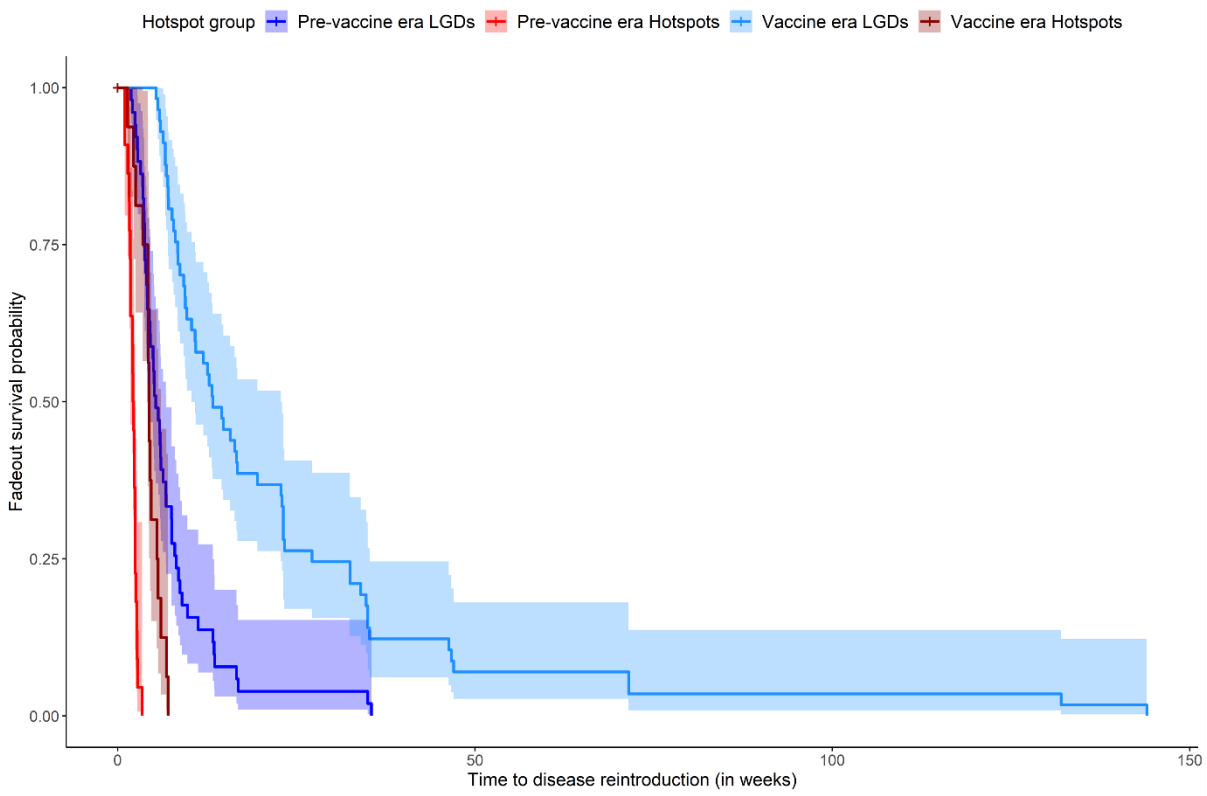

Figure 6.5 *Kaplan-Meier survival curves for the cumulative risk of reintroduction of pertussis in hotspots of infection and other districts in pre-vaccine era (1946-1957) and vaccine era (1958-1969) South Wales. Shading indicates 95% confidence intervals (CI) on the Kaplan-Meier survival curves.*

Fig. 6.5 clearly indicates that fadeout survival probability is significantly greater among other districts in the vaccine-era that the pre-vaccine era, with a far lower time to diseases reintroduction. Despite the introduction of routine mass vaccination, there is also minimal difference in fadeout survival probability in hotspot districts in the vaccine era (1958-1969).

Table 6.6 *Results of log-rank test for difference in survival between pertussis hotspots and other districts in South Wales, pre-vaccine era (1946-1957) vs. vaccine era (1958-1969).*

	Pre-vaccine era			**Vaccine-era**		
	N	**Observed**	**Expected**	**N**	**Observed**	**Expected**
Other Districts	51	51	68.3	57	57	70.5
Hotspots	23	22	4.7	17	16	2.5
χ^2			80.4			85.8
df			1			1
***p*-value**			0.0005			0.0005

Log-rank tests were performed to determine if there are statistically significant differences in the survival distribution of hotspots and other districts in South Wales in the pre-vaccine and vaccine eras. The following hypothesis was tested:

- H_0: There is no overall difference in the overall survival distribution of pertussis hotspots and other districts in the South Wales region in pre-vaccine and vaccine eras.
- H_A: There is a statistically significant difference in the overall survival distribution of pertussis hotspots and other districts in South Wales in pre-vaccine and vaccine eras.

The results of the log-rank tests in Table 6.6 indicate that there is a statistically significantly difference in survival between hotspots and other districts in the pre-vaccine era (χ^2= 80.4, $p < 0.05$). However, the significant size of the chi-square test statistic produced by the log-rank test indicates the data is a poor fit. The number of observed hotspots exceeds the number expected by 17, with 22 hotspots observed compared to an expected total of approximately five. A similar result is observed in the vaccine era, with a statistically significantly difference in survival between hotspots and other districts is also detected (χ^2= 85.8, $p < 0.05$), producing another large chi-square value. Notably, when rounded, only three hotspots were expected in the regional metapopulation, yet 16 were observed within South Wales during the vaccine era.

6.4 Discussion of Findings

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 results of the Cox regression analyses indicate vaccination serves to significantly reduce the effect of geographical coupling on disease persistence in South Wales, with increasing distance from endemic centres significantly associated with a fall in the rate of reintroductions among hotspots during 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 urbanised valleys. Limited connectivity across the wider region amplified the impact of

vaccination, with districts such as Haverfordwest MB and Llandeilo MB, key towns in Pembrokeshire and Carmarthenshire, becoming isolated from the broader metapopulation in the vaccine era and no longer presenting as pertussis hotspots.

A noticeable finding in South Wales is the presence of districts located in Glamorgan which do not register as hotspots despite being bordered on all sides by districts with high levels of pertussis persistence during the vaccine era (Fig. 6.3B). It is possible that these sparsely populated districts, with high population dispersal and low rates of susceptible recruitment effectively act as barriers to local diffusion of pertussis, restricting epidemic disease activity and circulation to urban centres and surrounding districts with which they are tightly coupled. The high level of pertussis persistence observed in valley districts, such as Merthyr Tydfil CB and Gelligaer UD, may reflect the valleys functioning as corridors of infection, with disease spread driven by population flows between communities that move in a linear fashion, since rail lines and roads that tie valley communities together are restricted to the valley floor, constricted by topographical obstacles that loom on either side of the river valleys. This serves perhaps to increase the density of transient infectious individuals within intra-valley population flows.

The high level of pertussis persistence in the Valleys detected by the hotspot analysis does not chime with the findings of the correlation analysis performed in Section 4.7.2, which indicated the mining districts in the Valleys were lightly correlated with the overall regional pattern of pertussis activity, with pertussis cases strongly correlated in either market towns or the larger urban centres located on the coastal plain in Glamorgan. However, it must be noted that the correlation analysis did not control for the spatial impact of vaccination, which would inevitably have significantly affected pertussis activity in the latter half of the time series. In future, an analysis of spatial synchrony of pertussis should be performed separately for the pre-vaccine and vaccine era time series, to facilitate a more informed comparison of findings between different methods of analysis.

The valleys of South Wales serve as an ideal candidate for future work exploring the epidemiological impact of regional transport connections and topography on the endemicity of childhood infections infection. The historic mining districts of Rhondda MB and Merthyr Tydfil CB represent divergent geographies valley districts, the former consisting of numerous hamlets, villages, while the latter represents a large, densely populated settlement at the head of the valleys. The geographical make-up of Rhondda MB may help to produce asynchronous pertussis activity at the district-level, between local subpopulations, ensuring high levels of persistence in the vaccine era, despite vaccination successfully serving to increase the endemic threshold population for pertussis by removing susceptibles from the wider population, as illustrated in Section 5.2, Chapter 5.

With the exceptions of Llanelli MB and RD in Southeast Carmarthenshire and a handful of districts in Monmouthshire, hotspots of pertussis persistence are absent outside of Glamorgan during the vaccine era. This contrasts sharply with the period without disease intervention, in which hotspots of pertussis activity included market towns in Pembrokeshire, Carmarthenshire and Monmouthshire, such as Haverfordwest MB, Llandeilo MB, and Abergavenny MB, important local population centres located on the edges of the region and further inland in the hinterland. These districts operated as pertussis hotspots despite their low population sizes, most likely due to the spatial import of infection from the multiple surrounding hamlets and villages economically linked to the towns. The loss of these hotspots in the vaccine era may reflect the geographical synchronisation of pertussis activity achieved through high vaccination coverage, with sufficient coupling between such towns and neighbouring districts to ensure a significant fall in pertussis endemicity countywide.

Moreover, vaccination is analogous to a significant reduction in birth rates. When coupled with declining fertility rates during the 1960s, this provides inhospitable conditions for the survival of pertussis infection, particularly in the relatively empty rural spaces beyond Glamorgan on the periphery of settlement, with highly dispersed populations. In these small,

isolated populations, demographic stochasticity ensures the elimination of pertussis in the deep troughs between epidemics outbreaks, which are dependent on reintroductions from other geographical areas. It is therefore unsurprising that the time to disease reintroduction extends considerably during the vaccine eras, as demonstrated by the analysis of Kaplan-Meier curves and fadeout survival probability. With high levels of remoteness and very low rates of susceptible recruitment with vaccination diminishing existing pools of susceptibles, pertussis retreats to the more densely populated urban districts and areas of the region; rescue effects are unable to sustain infection beyond urban centres and their metropolitan hinterlands over long distances due to the absence of sufficient density of contact between individuals from ‘donor areas’ and susceptibles in ‘recipients’ subpopulations (Xia et al., 2004). These conditions only serve to further the growth of endemic threshold populations over time.

In Lancashire, the densely-populated districts at the heart of the region’s conurbations, Manchester CB and Liverpool CB, serve as key endemic centres for pertussis, where the high, sustained levels of susceptible recruitment coupled with more limited spacing between susceptibles prevents fadeout of infection. This supports previous findings discussed in Section 4.6.2, which revealed Manchester CB to possess the strongest correlation with the regional pattern of disease activity, not just for pertussis, but also true to measles and scarlet fever. This suggests that these urban centres are the endemic reservoirs from which waves of infection radiate out into the wider region. Endemicity patterns for pertussis in Lancashire visualised and described in the preceding chapter, in Section 5.4.2, reveal that pertussis only persists endemically in Manchester CB and Liverpool CB after the onset of vaccination, with a drastic fall in incidence and persistence across the regional metapopulation. immunisation diminishes the potential for susceptible recruitment among infants and reduces the density of susceptible populations, sharply increasing the size of the endemic threshold population.

In previous research on measles metapopulations in England and Wales (Xia et al., 2004), it has been implied that large urban centres play a disproportionately important role in 'donating' infection to surrounding hinterlands since spatial coupling increases faster than linear with community size. This phenomenon may occur if the urban centres are disproportionately attractive in terms due to significant social or economic pull factors. Undoubtedly with the collapse of the textile industry across the Lancashire during the post-war period, inflicting significant job losses in smaller cotton mill towns and settlements associated with textile manufacture, the urban centres of Manchester CB and Liverpool CB gained even greater prominence as the industrial engines of the region and providers of employment opportunities.

The urban agglomeration of the metropolitan hinterlands of the Greater Manchester and Merseyside conurbations increases throughout the study period, with the growth of satellite settlements fuelled by families seeking economic opportunities in Manchester CB and Liverpool CB and by population decentralisation, with the clearance of thousands of unfit homes moving significant overspill populations from the urban cores to surrounding districts. For instance, districts such as Droylsden UD, Stretford MB and Middleton MB, which share contiguous borders with Manchester CB, became home to much of the city's overspill population during the late 1950s and 1960s. The impact of the development of new towns for overspill populations from the Liverpool and Manchester conurbations on disease persistence is discussed in Section 5.6.2. Significant numbers of Liverpool's overspill population migrated to the Skelmersdale new town in the late 1950s and 1960s, following the extensive slum clearance in the city. Coupled with the high extrinsic variation in birth rates during the baby boom period, the explosive growth of the and explosive growth of the district dragged disease activity in Skelmersdale UD away from the regional disease pattern, with highly asynchronous, localised epidemic outbreaks leaving the district very weakly correlated from the dynamics of the wider regional metapopulation. Within these contexts and complex inter-relationships between places, significant spatial interaction

exists between endemic reservoirs, large towns, and surrounding communities', fuelling rescue effects. Rescue effects restrict the growth of the regional endemic threshold population for pertussis in Lancashire in the vaccine era (see Section 5.2), with constant reinfection of districts as a result of tight spatial coupling. 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. This would explain the notable concentration of pertussis persistence within Liverpool and Manchester CB and their surrounding hinterlands in the vaccine era, and the much higher number of hotspots observed despite the relatively small population sizes of districts which neighbour the two endemic centres.

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 core-satellite 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 the 'core' districts of Manchester CB and Liverpool CB and 'satellite' districts, located within a distance of 24km of the two endemic centres, ensures pertussis persistence in less populated subpopulations due to high levels of geographical coupling and regular occurrence of rescue effects. In the South Wales regional metapopulation, the same dynamics are apparent between the core district of Cardiff CB, and to a lesser extent Swansea, and satellite districts situated within a 36km radius of the core.

6.5 Chapter Summary

The introduction of routine mass pertussis vaccination nationwide from 1957 onwards was largely successful in breaking chains of transmission by depleting the pool of susceptible individuals, significantly reducing disease incidence. Thus, this increased the population size requirements for maintaining chains of pertussis transmission. The findings of the hotspot and survival analyses laid out in this chapter build upon those described in Sections 5.2 and 5.5.2. In both the Lancashire and South Wales 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. 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.

The methods presented here enables the study of regional and local transmission patterns which are fundamentally rooted in human mixing behaviour across spatial scales. Identifying spatial interactions that contribute to recurring epidemic outbreaks or sustaining disease persistence are vital for understanding the wider geographical picture in which the growth of the endemic threshold population is shaped. It is clear that the spatial impact of vaccination was to isolate persistent hotspots of infection located in the most heavily urbanised areas often regional populations, districts which continue to possess rates and densities of susceptible recruitment that can main disease circulation in the presence of growing immunisation coverage. However, pertussis hotspots in the vaccine-era, beyond

Manchester CB and Liverpool CB in Lancashire and Cardiff CB in South Wales, are seemingly dependent on tight spatial coupling with such urban centres to ensure a much more regular reintroduction of disease to facilitate the persistence of pertussis within the district. In the next chapter, a rich multivariate regression framework in the form of the HHH model will be utilised to further study the spatial spread and persistence of the three childhood diseases under study in this thesis within the two regional metapopulations, to develop a greater understanding of the processes and factors which influence the persistence of such infections during the study period.

Chapter 7: Endemic–Epidemic Modelling

7 Introduction

This chapter presents the results of endemic-epidemic modelling of pertussis, measles and scarlet fever incidence across nine time–windows in the Lancashire and South Wales regions, using various sub-model formulations of the HHH model, a multivariate time series model for infectious disease count data. For a full description of the modelling procedure, model formulations fitted and analyses performed here, refer back to Section 2.10, Chapter 2. The presentation of results is followed by a discussion of findings, exploring what results obtained from the HHH models reveal about the spatial spread of childhood infections in Lancashire and South Wales across the study period. The purpose of this analysis is to elucidate greater understanding of potential mechanisms and factors behind long-term spatiotemporal changes in endemic threshold size in regional populations.

Firstly, two baseline models are fitted and described for pertussis, measles and scarlet fever counts in the regional populations of Lancashire and South Wales for each time–window: the endemic HHH model and the first-order HHH model. The simplest baseline model, the endemic HHH model only includes an endemic component, with an endemic mean and a multiplicative, district-specific offset in the form of population fractions that serve as the key demographic covariate accounting for the scaling of subpopulations. The log-linear predictor includes a linear trend and accounts for basic seasonal behaviour. A more complex baseline model, the first-order HHH model contains three components: the endemic, autoregressive, and spatiotemporal components. This spatiotemporal or ‘neighbourhood’ component, models the flow of infection from neighbouring districts, with transmission weights assumed to be known. Local districts distribute cases of infection uniformly to all first-order neighbours, limited by adjacency order. Thus, disease spread is considered a purely contiguous process.

The two central assumptions of the first-order HHH model formulation, that epidemic spread of infection only travels from directly-adjacent neighbouring districts and that all districts have the same potential for importing cases from neighbouring subpopulations, are overly-simplistic and problematic for several reasons. Firstly, these assumptions ignore wide variation in the size, density and replenishment rate of susceptible populations in local districts within a regional metapopulation. Secondly, it is well understood that humans have the ability to travel further and most often to metropolitan areas due to factors such as employment (Meyer and Held, 2014). Thirdly, it has been widely demonstrated that the spread of childhood diseases such as measles is a product not just of contiguous diffusion but a mixture of diffusion processes, including hierarchical diffusion of disease (Trevelyan, et al., 2005; Cliff et al., 1998). The first-order HHH model formulation was modified to embed more realistic networks of spatial interaction and disease spread into the modelling framework and analyse to what extent human mobility, viewed as an important driver of epidemic spread, influences the persistence of childhood infections in regional metapopulations with divergent spatial structures and demographic characteristics.

7.1 Results: Lancashire Region

7.1.1 Pertussis

Overdispersion

For the endemic HHH model, there is significantly greater levels of overdispersion across all time–windows compared to all other model formulations (see Fig. 7.1), with the least level of residual heterogeneity detected in time–window four (ψ = 2.07, 95% CI: 2.02–2.12). From time–window four onwards, the overdispersion parameter rises consistently (Fig. 7.1) to a high of ψ = 5.20 in time–window nine (95% CI: 4.94–5.45). The overall mean ψ parameter = 3.42 (95% CI: 3.30–3.55). The exclusion of epidemic components from the HHH endemic models results in a significant rise in residual heterogeneity and an inferior

fit to the count data, suggesting pertussis incidence is primarily driven by epidemic activity. First-order, gravity and power-law HHH models all report significantly lower overall levels of overdispersion than the endemic model formulation (see Fig. 7.1). The mean overdispersion parameters are ψ = 1.81 (95% CI: 1.72–1.89) for the first–order HHH model, ψ = 1.79 (95% CI: 1.70–1.87) for the gravity HHH model, and ψ = 1.76 (95% CI: 1.68–1.84) for the power-law HHH model.

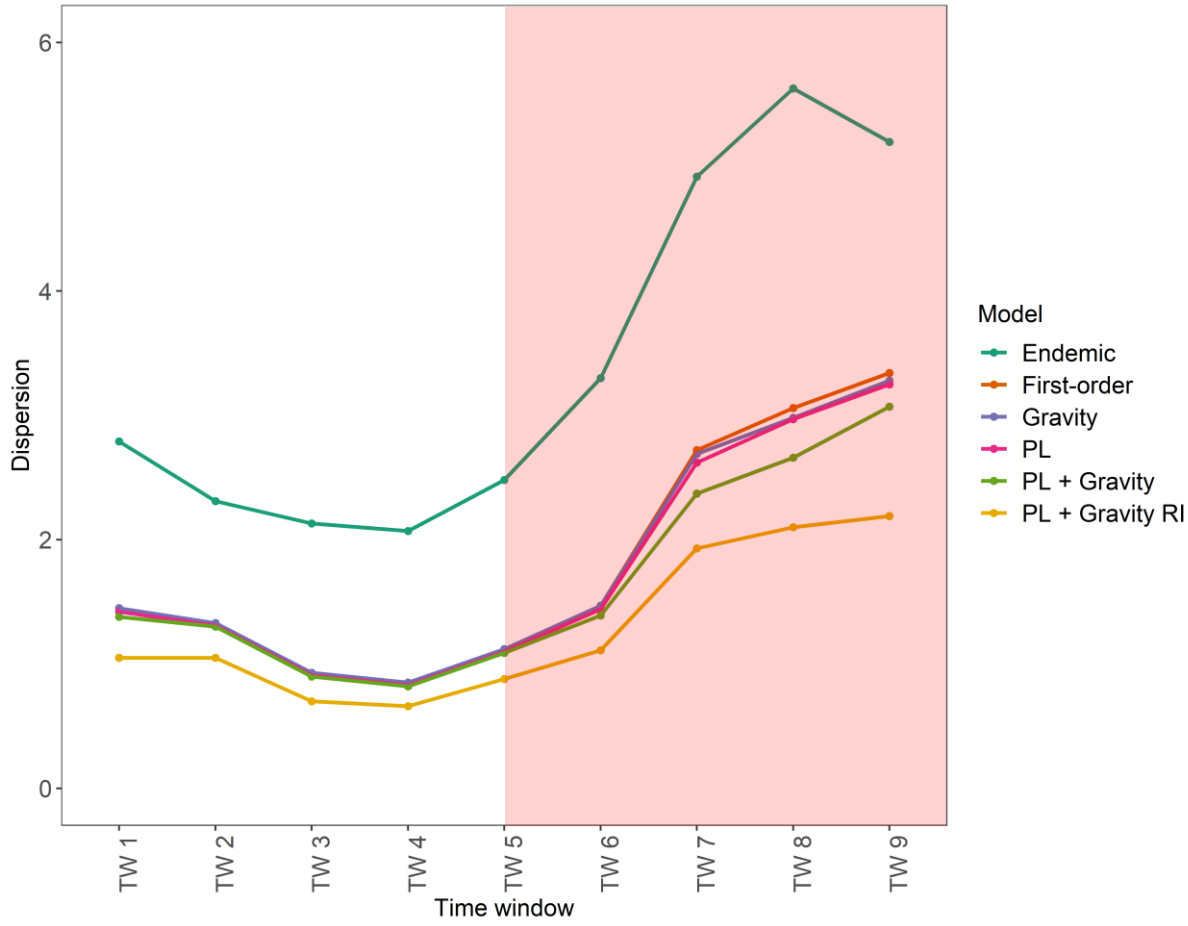

Figure 7.1 *Dispersion parameter estimates for five HHH model formulations analysing pertussis spread in Lancashire, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend. Pink shaded area denotes the vaccine era.*

The power-law + gravity HHH model provides the optimum fit in terms of residual heterogeneity, with the exclusion of random effects (Fig. 7.1). The mean ψ parameter is

1.67 (95% CI: 1.59–1.74). Underdispersion is observed in across time–windows one to four (Fig. 7.1), indicating less variation than would be expect based on a negative binomial distribution. This may be result of autocorrelation among adjacent districts, with clusters of heightened pertussis activity as well as clusters of sparsely populated, isolated districts with no cases dotted across the region. However, with the advent of vaccination, from time–window six onwards, overdispersion features rises considerably in successive windows (See Fig. 7.1). With the inclusion of random effects in the power-law + gravity HHH model formulation, the mean overdispersion rises to ψ = 1.30 (95% CI: 1.23–1.36).

Distance Decay

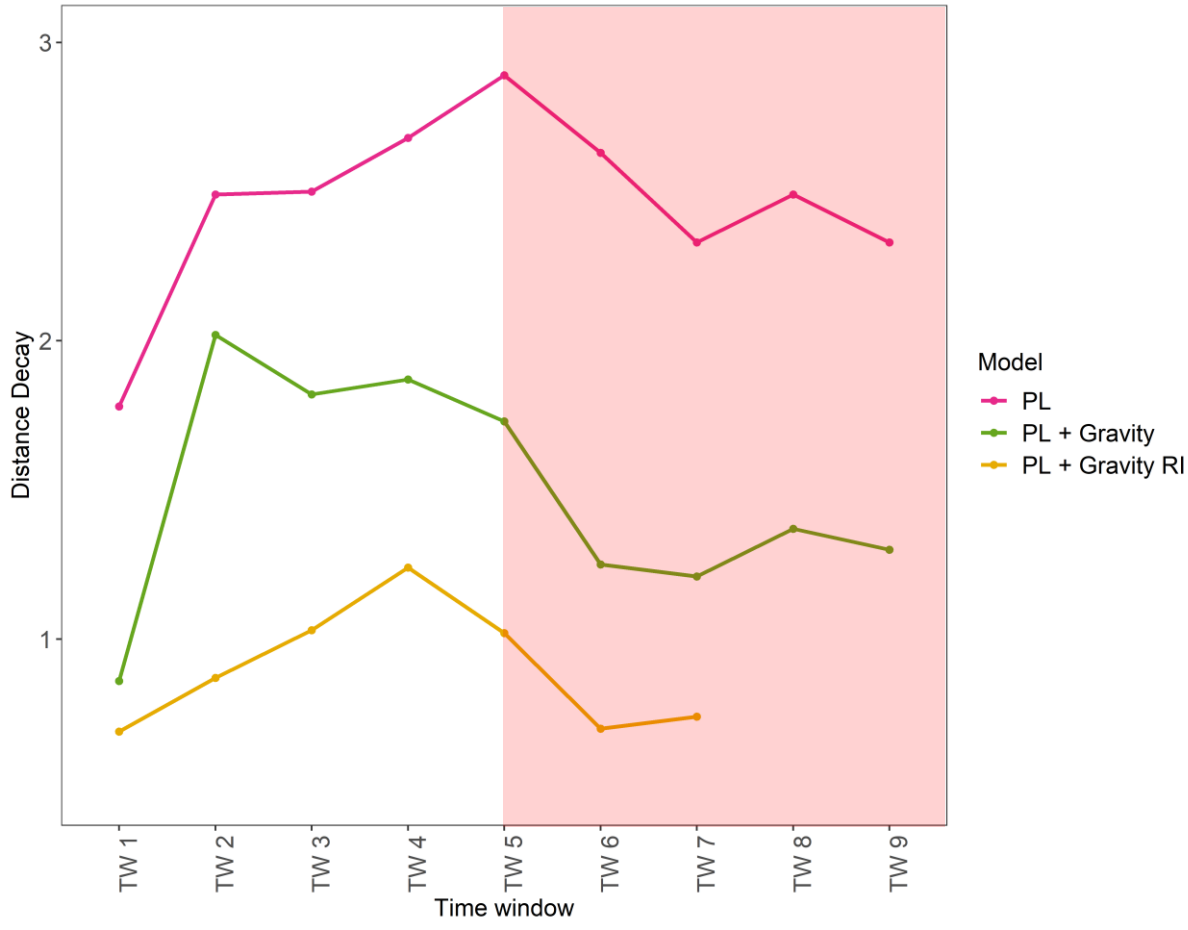

Figure 7.2 *Distance decay parameter estimates for three HHH model formulations analysing pertussis spread in Lancashire, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend. Pink shaded area denotes the vaccine era.*

For the power-law HHH model formulation, the mean distance decay parameter estimate is d = 2.46 (95% CI: 2.26–2.65). Inclusion of a gravity component leads to a significant fall in the mean distance decay parameter (d = 1.49, 95% CI: 1.32–1.66). Addition of random effects sees a substantial decline in the overall strength of the distance decay parameter in compared to previous HHH model formulations (see Fig. 7.2), with the greatest fall in the final two time–windows (Fig. 7.2). The extended power-law + gravity RI HHH model formulation has a mean distance decay parameter of 0.69 (95% CI: 0.58–0.81).

Gravity Parameter

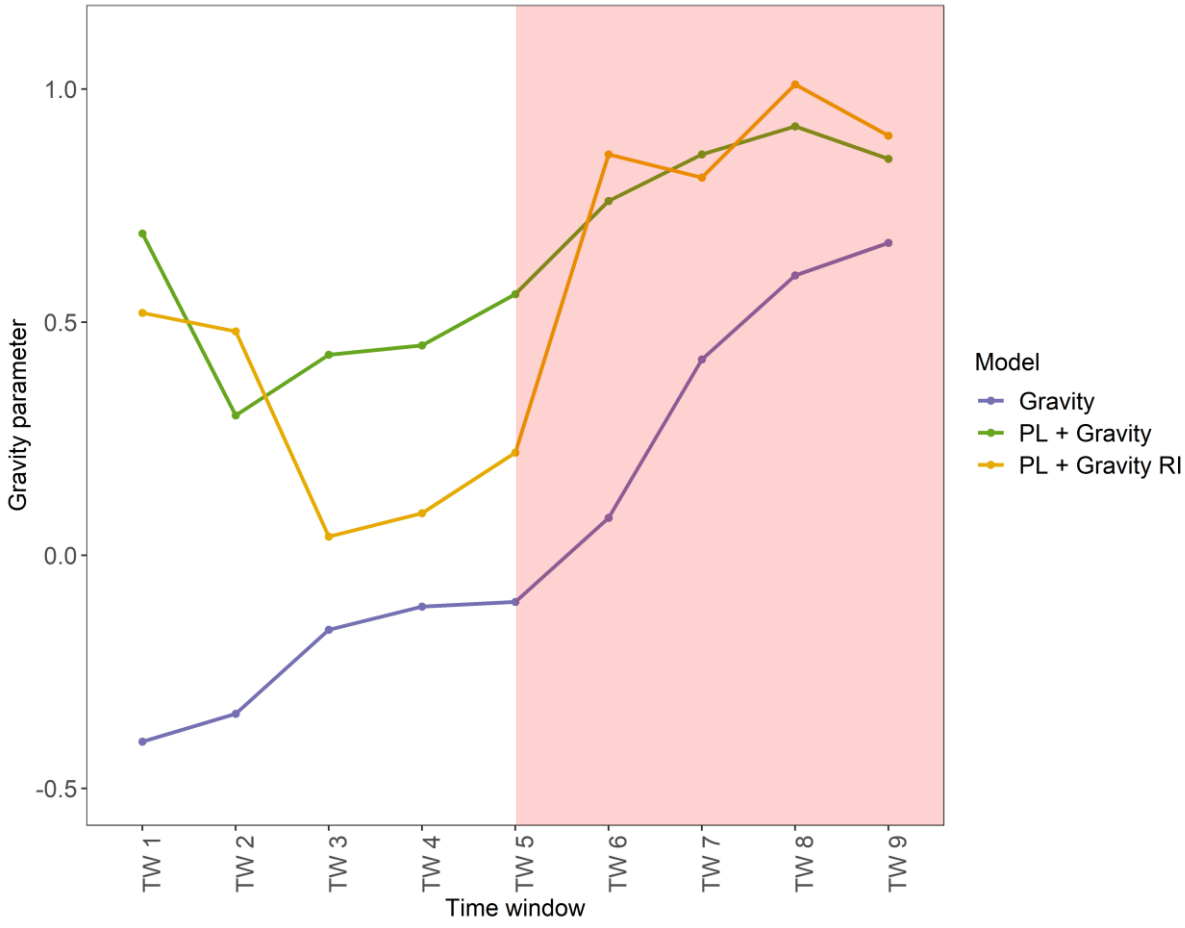

Figure 7.3 *Gravity parameter estimates for three HHH model formulations analysing pertussis spread in Lancashire, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend. Pink shaded area denotes the vaccine era.*

A mean $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$ parameter value of 0.65 (95% CI: 0.58–0.71) is calculated across the nine time–windows after extending the power-law HHH model to include a gravity component. The lowest parameter estimate is found in time–window two, at 0.30 (95% CI: 0.19–0.41). Despite an initially high parameter estimate in time–window one (0.69; 95% CI: 0.64–0.75), the strength of the estimates falls in time–windows two and three (see Fig. 7.3). From time–windows four onwards there is a gradual, consistent increase in the strength of the parameter across successive time–windows, peaking at 0.92 in time–window eight (95% CI: 0.86–0.98).

Using the gravity HHH model formulation, the $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$ parameter estimate is extremely weak across the first six time–windows, prior to vaccination. After negative values are observed for the $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$ parameter in the first five time–windows, from time–window six onwards there is a large jump in the strength of the parameter estimate (Fig. 7.3), from 0.08 (95% CI: -0.04–0.20) in time–window six to 0.67 (95% CI: 0.48–0.86) by time–window nine. The mean value of the $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$ parameter across the nine time–windows is 0.07 (95% CI: -0.06–0.20).

With the addition of random effects to the power-law + gravity HHH model, there is a slight fall in the overall mean $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$ value, to 0.55 (95% CI: 0.41–0.68) compared to 0.65 for the power-law + gravity HHH model. Yet, with the advent of vaccination, there is a significant rise in the strength of the $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$ estimate (Fig. 7.3), increasing from 0.22 in time–window five to 0.86 (95% CI: 0.75–0.97) in the next window). The strength of the $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$ parameter estimate continues to rise across the vaccine era in successive time–windows (Fig. 7.3).

maxEigenvalues

The strength of the maximum eigenvalue grows with the increasing complexity of the HHH formulation fitted, rising from $MaxEV$ = 0.61 for the first–order model to 0.76 for the power-

law + gravity HHH model (see Fig. 7.4). However, no HHH model formulation prior to the inclusion of random effects signify a major epidemic outbreak within the time–windows (Fig. 7.4), indicated by no maximum eigenvalue exceeding one in value. Heightened levels of epidemic activity are observed in the power-law + gravity HHH models, in time–windows one (0.82), six (0.86) and seven (0.80) respectively. Accounting for random effects leads to a further increase in the epidemic proportion of cases captured by the HHH models (Table V.32, Appendix V). The power-law + gravity RI HHH model formulation reduces the number of major epidemic outbreaks detected across the study period to two (in time–windows six and eight). The mean maximum eigenvalue is 0.95 (min = 0.87, max = 1.03).

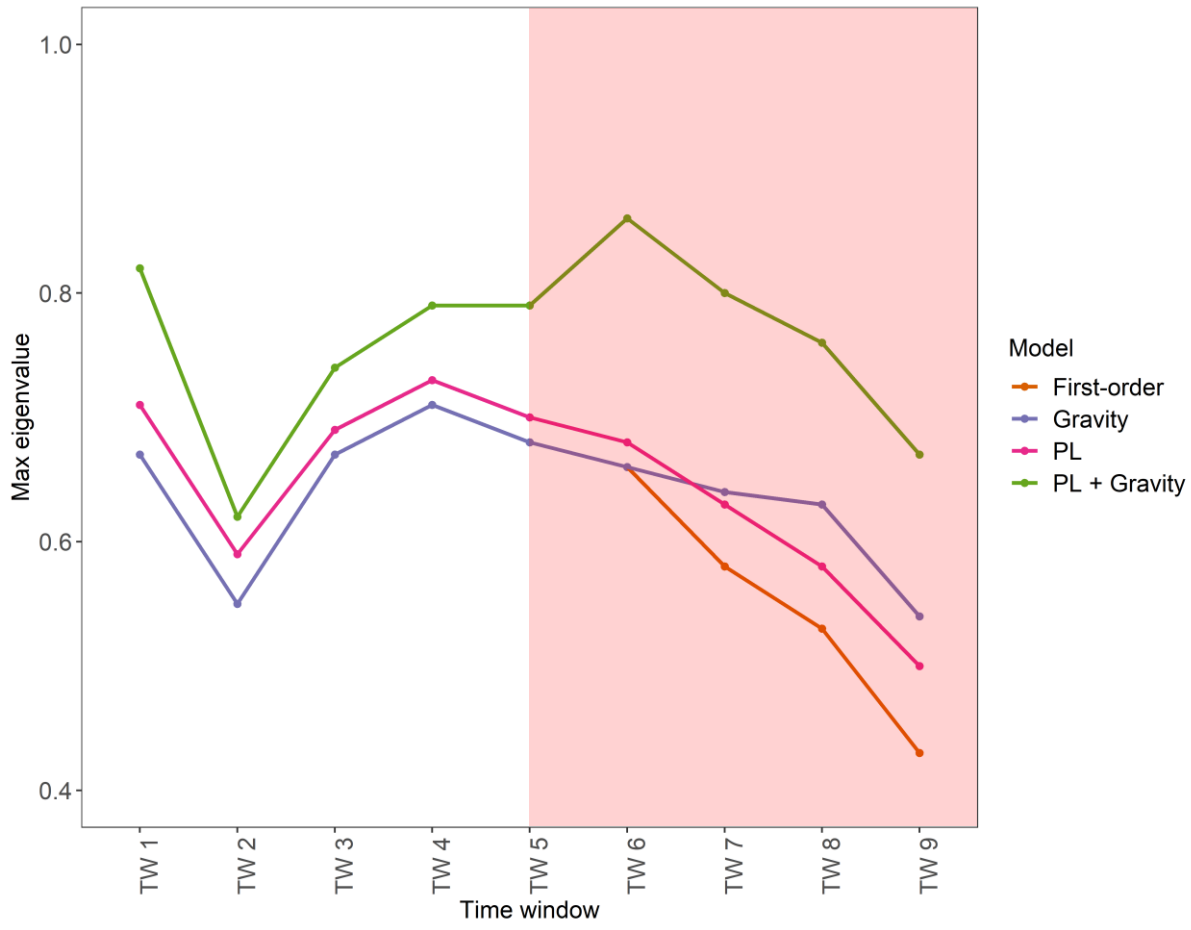

Figure 7.4 *Maximum eigenvalues for four HHH model formulations analysing pertussis spread in Lancashire, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend. Pink shaded area denotes the vaccine era.*

Predictive Assessment Scores

An assessment of the goodness–of–fit of the HHH model formulations based upon the AIC values reveals the HHH power-law + gravity model provides the best fit of the five model formulations (endemic, first–order, power-law, gravity and power-law + gravity), with a mean AIC value of 8.0844×10^5 (80844.6) across the nine time–windows Goodness–of–fit assessments using scoring rules suggests the power-law HHH model provides the best fit, without inclusion of random effects (*LogS* = 0.196, *RPS* = 0.202). This is despite the lack of convergence for HHH model formulations in time–window eight and lack of statistical significance. With inclusion of random effects, the power-law + gravity RI HHH model provides the best fit of the model formulations (*RPS* = 0.171). A full breakdown of AIC values and predictive mean scores by time–window and model fit for pertussis in Lancashire can be found in Appendix V (Tables V.36 and V.41).

7.1.2 Measles

Overdispersion

There is less overdispersion in the power-law + gravity models (see Fig. 7.5) compared to previous formulations, indicating reduced residual heterogeneity. The mean overdispersion parameter is 1.02 (95% CI: 1.00–1.07), with underdispersion detected in the time–window three (0.91, 95% CI: 0.88–0.94) and the highest level of overdispersion noted in time–window nine (1.12, 95% CI: 1.09–1.15). With the exception of the endemic HHH models, the overdispersion parameter ψ hovers around 1 for all other model formulations (Fig. 7.5). The higher levels of overdispersion in the endemic HHH models ($\bar{x}$ = 3.75, 95% CI: 3.68–3.82) indicates the importance of within-district and spatio-temporal epidemic components for providing a suitable fit for measles spread. With the addition of random effects, the fit for power-law + gravity RI HHH models indicates some underdispersion (Fig. 7.5), with a mean ψ of 0.92 (95% CI: 0.89–0.94) across the nine time–windows.

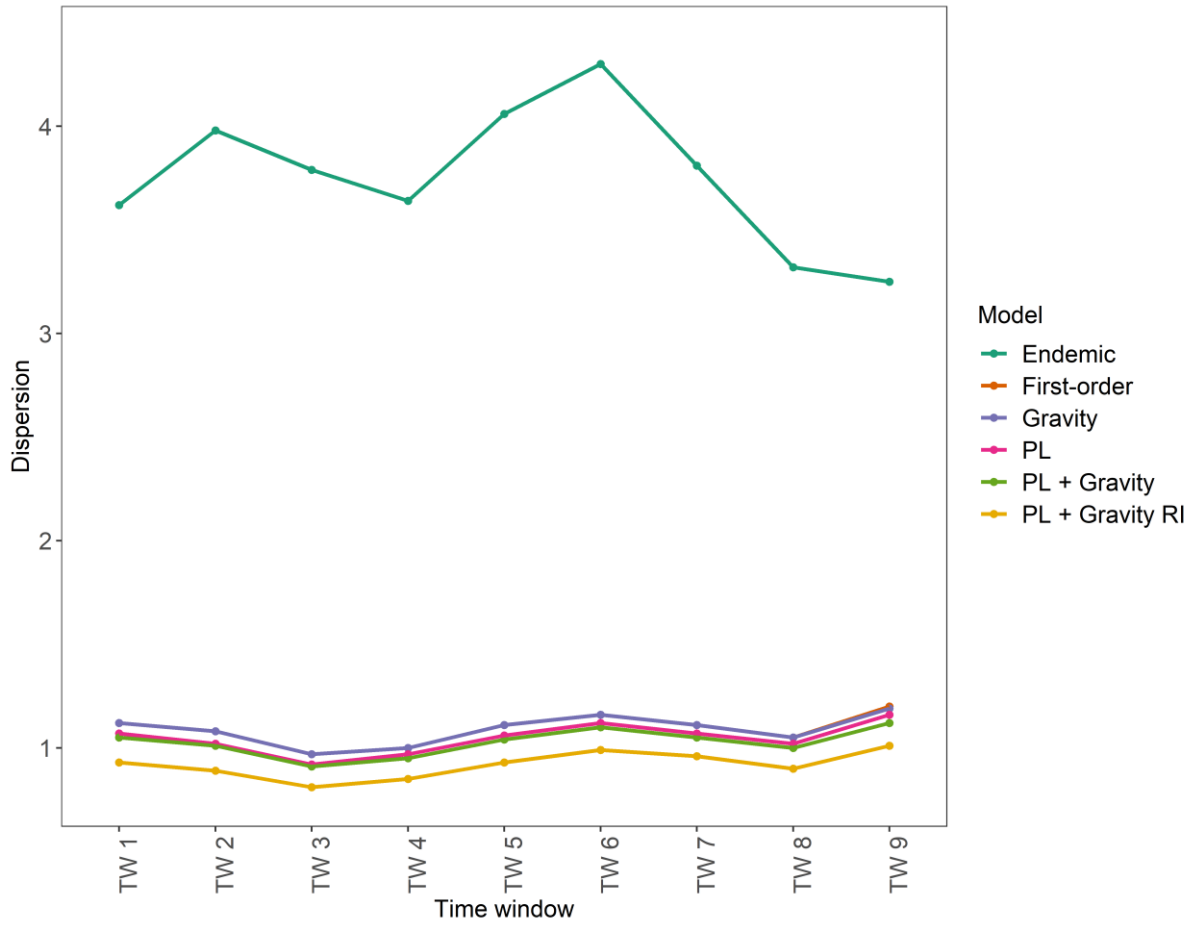

Figure 7.5 *Dispersion parameter estimates for five HHH model formulations analysing measles spread in Lancashire, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

Distance Decay

For the power-law HHH model formulation, the mean distance decay parameter estimate is d = 3.03 (95% CI: 2.90–3.17). The inclusion of a gravity measure leads to a notable fall in the mean distance decay parameter (d = 2.25, 95% CI: 1.32–1.66). With the inclusion of random effects in the power-law + gravity HHH model, there is a further decrease in the size of the distance decay parameter in each time–window compared to previous HHH model formulations (see Fig. 7.6). But like previous fixed-effect formulations, distance decay estimates remain relatively consistent in size over the course of the study period (Fig. 7.6).

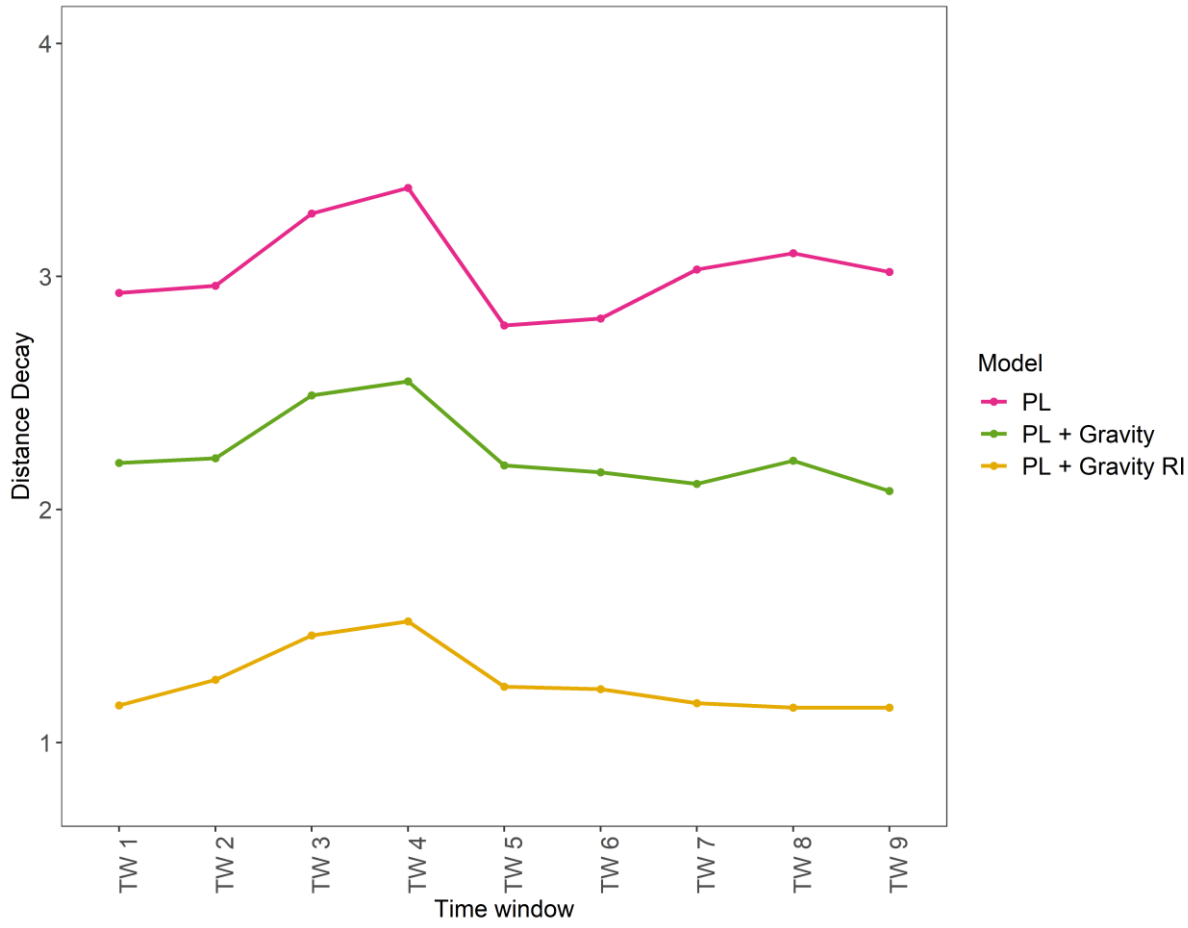

Figure 7.6 *Distance decay parameter estimates for three HHH model formulations analysing measles spread in Lancashire, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

Gravity Model

Extending the first–order HHH model formulation to include a gravity measure yielded a mean estimated coefficient of $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$ = 0.03 (95% CI: -0.03–0.09), suggesting no evidence of agglomeration effect. Models for time–windows one to three failed to produce an estimate with confidence intervals. However, the addition of a power-law to the model resulted in a drastic increase in the parameter $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$, rising to a mean of 0.46 (95% CI: 0.42–0.50), with a low 0.38 and a high of 0.57 in time–windows three and nine (see Fig. 7.7),

respectively. The power-law + gravity HHH model provides strong evidence for an agglomeration effect, with urban centres importing larger number of cases from neighbouring districts due to commuter-driven spread. The added weight given to the $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$ parameter results in a smaller estimation of the decay parameter compared to the power-law HHH model formulation (Fig. 7.7), falling to a mean of 2.25 (95% CI: 2.12–2.37) from 3.03 and all other effects remain relatively unchanged (Fig. 7.7). With the inclusion of random effects, $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$ falls by a fraction from 0.46 to $\bar{x} = 0.44$ (95% CI: 0.33–0.55).

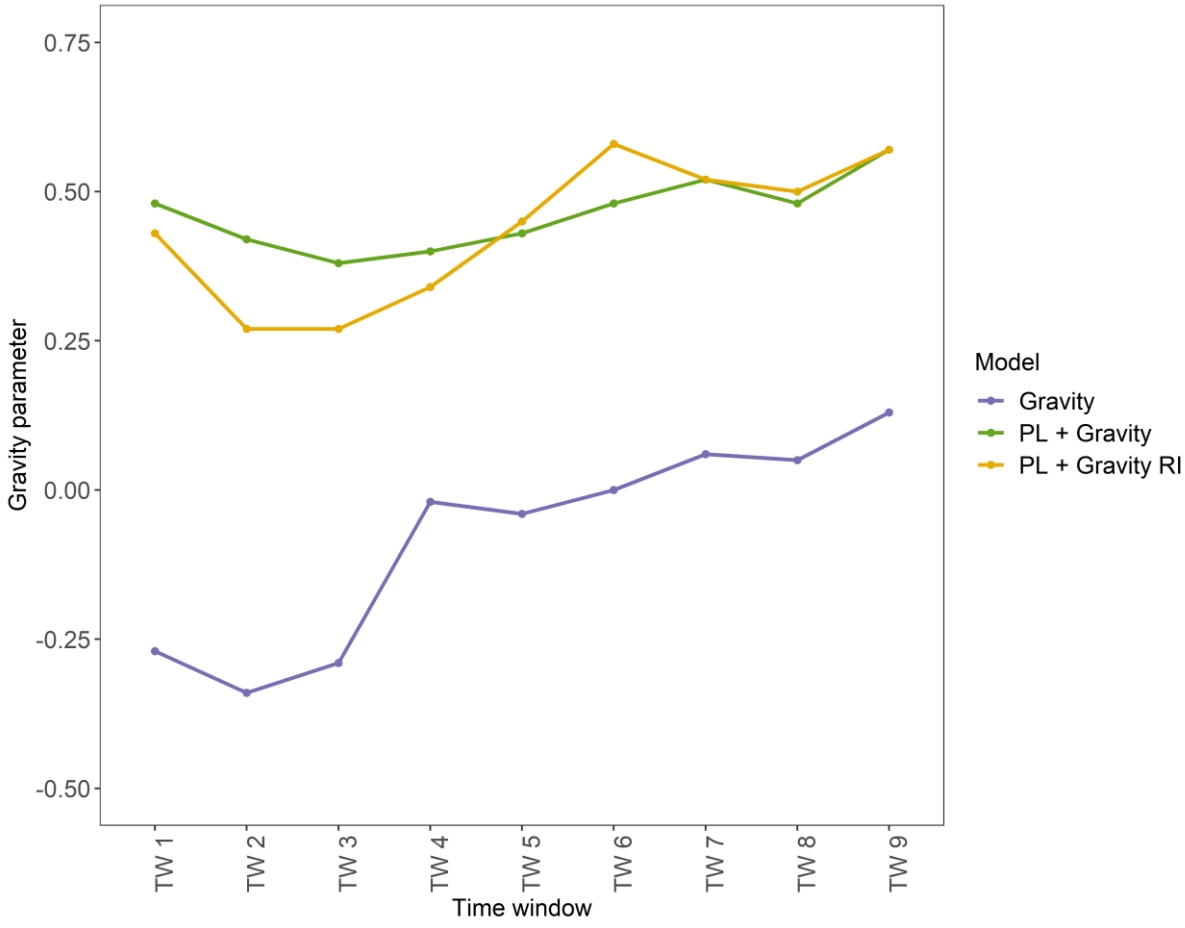

Figure 7.7 *Gravity parameter estimates for three HHH model formulations analysing measles spread in Lancashire, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

maxEigenvalues

Regarding measles in Lancashire, the combined measure is more or less unchanged upon accounting for higher-order neighbours with a power-law, with a mean max eigenvalue of 0.92. This is in line with the first-order HHH model, which assumes a fixed adjacency between neighbours. With the addition of random effects, the epidemic proportion increases disease incidence increases markedly (see Fig. 7.8).

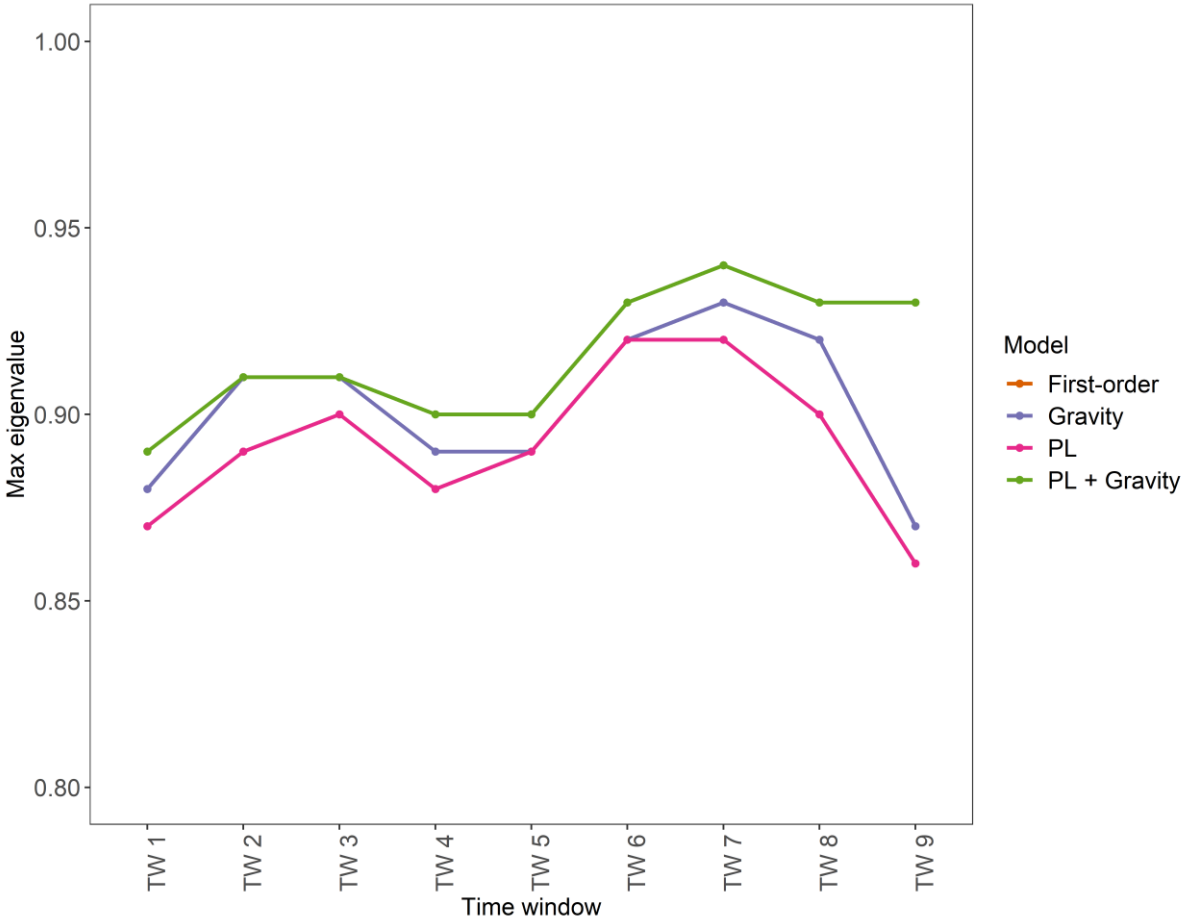

Figure 7.8 *Maximum eigenvalues for four HHH model formulations analysing measles spread in Lancashire, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend. Endemic model excluded due to absence of epidemic component.*

Random Effects

With the inclusion of random effects, the $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$ parameter falls by a fraction from 0.46 to a mean of 0.44 (95% CI: 0.33–0.55). A comparison of HHH models using proper scoring rules reveals only a marginal improvement in predictive performance, with the mean *LogS* score across all time–window models decreasing from -0.207 to -0.259 (p = 0.152). The variance of the random effect of the spatiotemporal component is slightly reduced from 0.65 to 0.43, reflecting a decrease in residual heterogeneity between districts.

Predictive Assessment Scores

Based on the AIC values calculated, of the five models fitted, the power-law + gravity HHH model formulation provides the best fit to the measles incidence data across the nine time–windows with a mean AIC score = 13761.3 (1.3761×10^5). A full breakdown of AIC values and predictive mean scores by time–window and model fit for measles in Lancashire can be found in Appendix V (Tables V.34 and V.40).

7.1.3 Scarlet fever

Overdispersion

Comparison of the strength of the overdispersion parameter of the five fixed-effects HHH model formulations and two formulations incorporating random effects indicates that the endemic HHH model may provide the best fit with observed scarlet fever incidence across the nine time–windows. The endemic HHH model consistently demonstrates significantly less underdispersion in parameter estimates (see Fig. 7.9), which would suggest significant autocorrelation in the data. Residual heterogeneity rises across the time–windows over the course of the study period as scarlet fever incidence gradually declines across the region, rising from a low of ψ = 0.66 (95% CI: 0.64–0.69) in time–window one to ψ = 1.12 (95% CI: 1.07–1.18) in time–window nine. Overdispersion peaks in time–window eight (ψ = 2.09

(95% CI: 1.98–2.19). The overall mean for the endemic HHH model formulation is ψ = 1.12 (95% CI: 1.07–1.18). With the inclusion of epidemic and spatiotemporal components, significantly greater levels of underdispersion are estimated, with the mean parameter value for first-order, gravity, power-law and power-law + gravity HHH formulations ranging between 0.65–0.67. Overdispersion is only found in time–windows eight and nine for these model formulations (Fig. 7.9). Inclusion of random effects with the power-law + gravity HHH formulation only sees the levels of underdispersion detected, and thus presence of autocorrelation, increase.

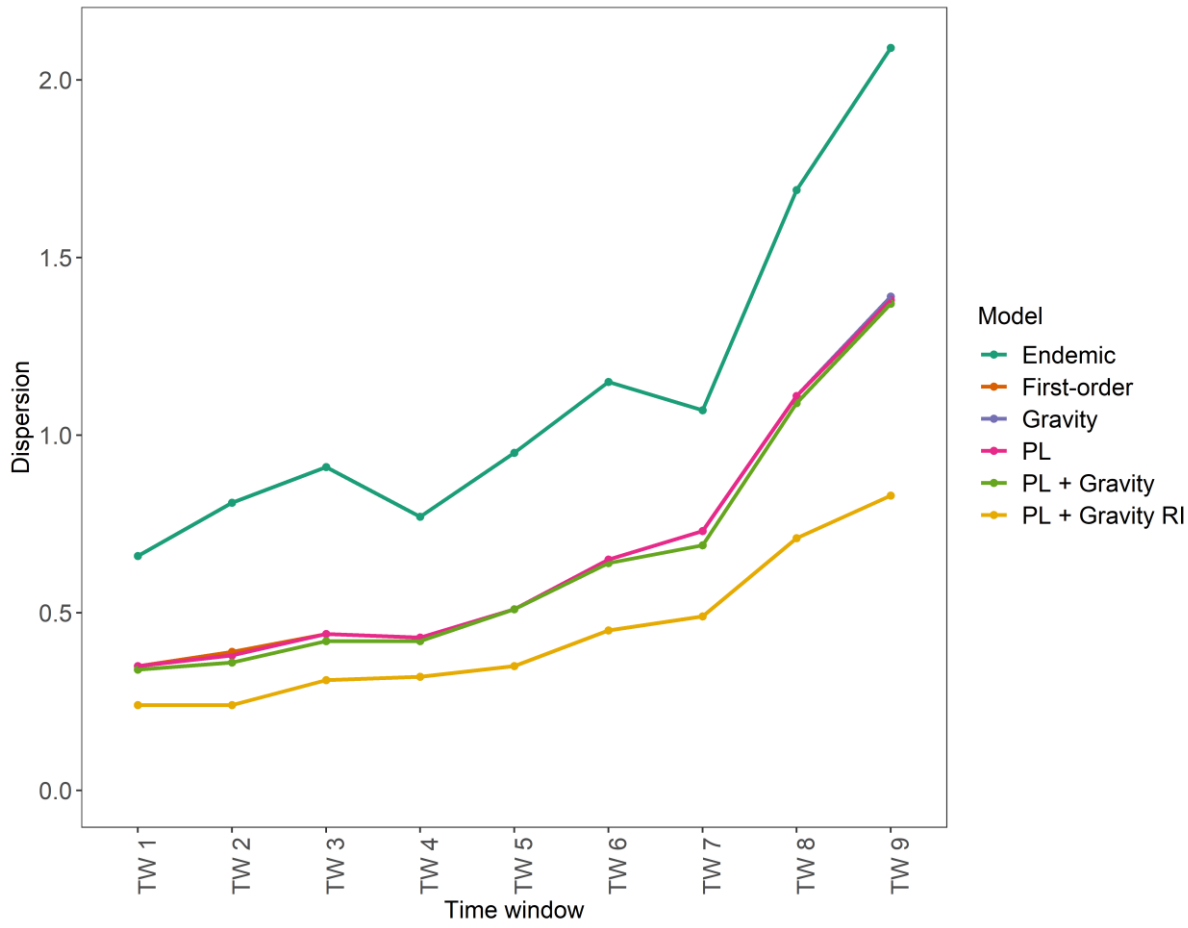

Figure 7.9 *Dispersion parameter estimates for five HHH model formulations analysing scarlet fever spread in Lancashire, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

Distance Decay

Positive distance decay estimates are observed for the power-law extended HHH models across all time–windows (see Fig. 7.10). The overall mean d = 0.96 (95% CI: 0.74–1.19). As observed with measles and pertussis incidence in the Lancashire region, there is a significant fall in the strength of the distance decay parameter after inclusion of a gravity measure. The overall mean d parameter = 0.47 (95% CI: 0.31–0.63). There is no clear downward trend in the distance decay estimates over the course of the nine time–windows, (Fig. 7.10). The distance decay parameter falls from 0.73 in time–windows six (95% CI: 0.59–0.88) to d =0.32 by the final time–window (95% CI: 0.10–0.55).

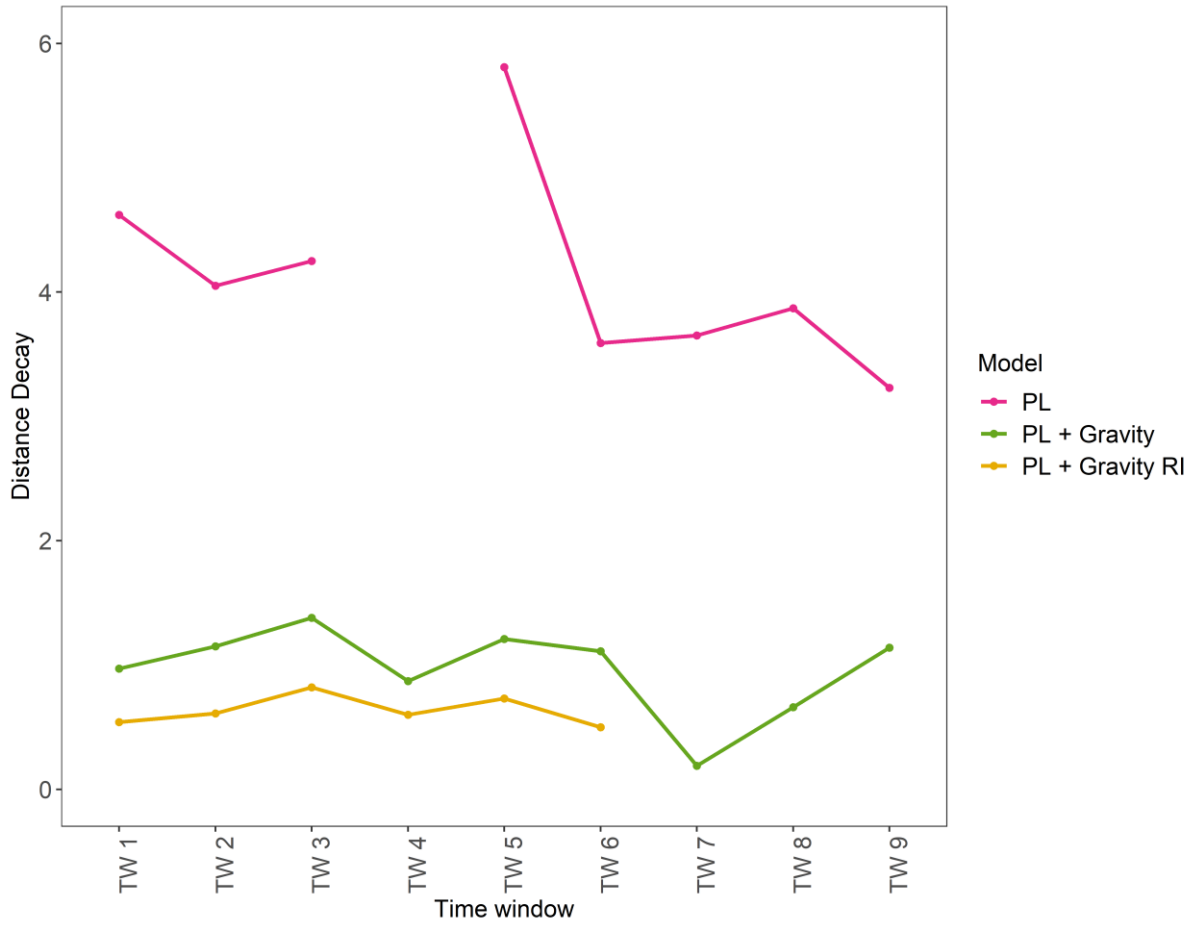

Figure 7.10 *Distance decay parameter estimates for three HHH model formulations analysing scarlet fever spread in Lancashire, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

The inclusion of random effects only serves to further reduce the strength of the distance decay parameter estimate, although this is only marginal (Fig. 7.10). The largest distance decay parameter is observed in time–window three and falls significantly from time–window six onwards. The overall mean d parameter = 0.45 (95% CI: 0.28–0.62). In time–window nine, there is no distance decay parameter estimate due to a failure of convergence in the model.

Gravity Parameter

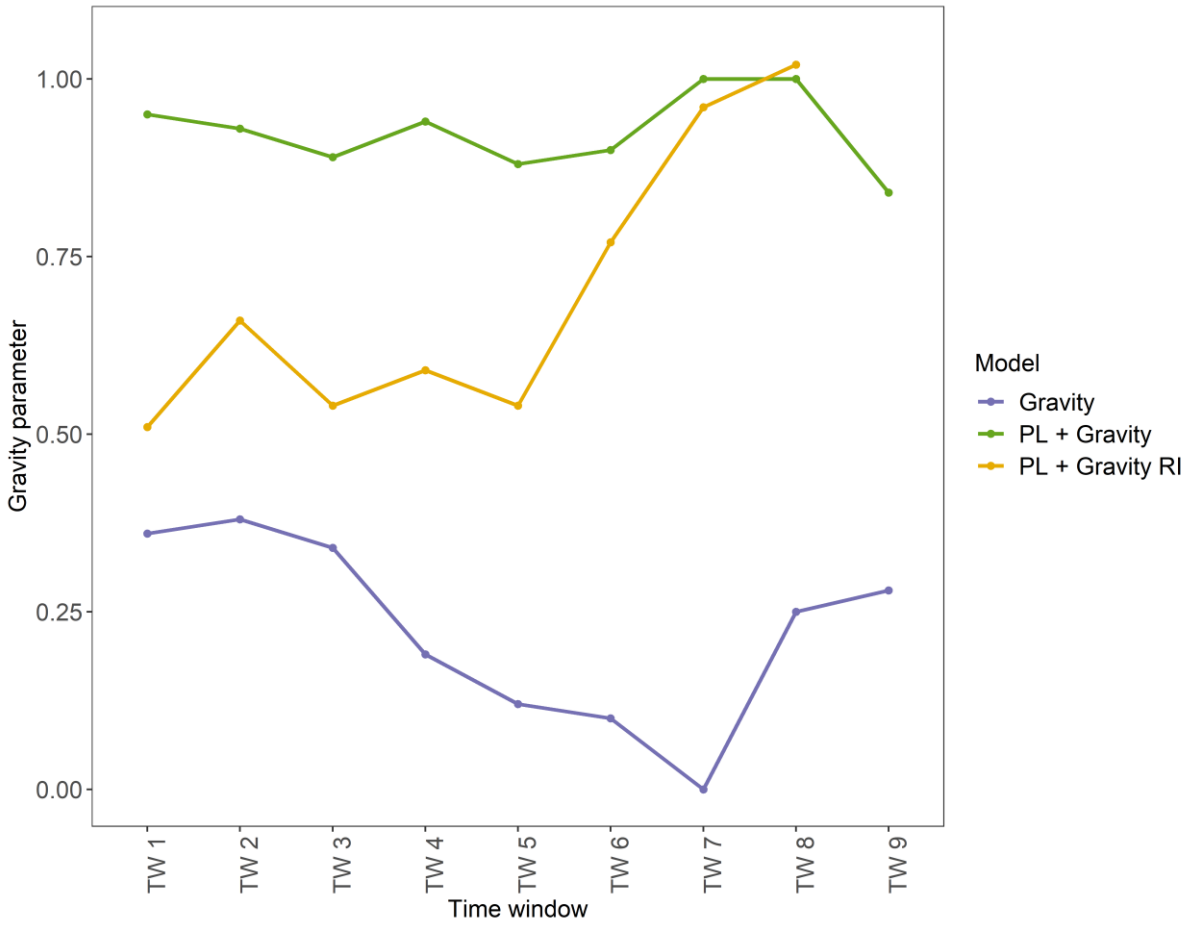

Figure 7.11 *Gravity parameter estimates for three HHH model formulations analysing scarlet fever spread in Lancashire, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

The mean parameter estimate for the gravity HHH models is $\boldsymbol{\beta}^{(\boldsymbol{\phi})}_{\mathbf{log}\,(\boldsymbol{pop})}$ = 0.22 (95% CI: 0.10–0.35). The power-law + gravity HHH model fit sees a significant increase in the strength of the gravity parameter estimates across all nine time–windows (see Fig. 7.11). Estimates range between 0.84 and 1.00, with a mean $\boldsymbol{\beta}^{(\boldsymbol{\phi})}_{\mathbf{log}\,(\boldsymbol{pop})}$ estimate of 0.93 (95% CI: 0.88–0.97). The inclusion of random effects results in slightly weaker estimates of the gravity parameter across the time–windows. The power-law + gravity RI HHH model fails to converge in time–window nine (Fig. 7.11). The mean gravity parameter across all nine time–windows $\boldsymbol{\beta}^{(\boldsymbol{\phi})}_{\mathbf{log}\,(\boldsymbol{pop})}$= 0.70 (95% CI: 0.59–0.81). Noticeably, a pattern of growth emerges in later time–windows, with the gravity parameter increasing from 0.54 (95% CI: 0.38–0.70) in time–window five to a high of 1.02 (95% CI: 0.88–1.16) in time–window eight. Excluding time–window two, parameter estimates in time–windows one to five range between $\boldsymbol{\beta}^{(\boldsymbol{\phi})}_{\mathbf{log}\,(\boldsymbol{pop})}$= 0.50–0.60. Increases in successive windows between time–windows five and eight coincide with the advent of pertussis vaccination and retreat of pertussis infection.

MaxEigenvalues

Assessment of the maximum eigenvalues reveals increasing weight given to the epidemic proportion of scarlet fever cases in each of the nine time–windows, with the stepwise extension of the HHH model formulation (Fig. 7.12). The growth in the mean epidemic eigenvalue is limited with the extension of the first-order model formulation to include, separately, gravity and power-law components, rising from $maxEV$ = 0.61 for the first-order HHH model to $maxEV$ = 0.64 for the gravity HHH model, and $maxEV$ = 0.65 for the power-law HHH model. With the extension of the HHH model formulation to include both gravity and power-law components, the epidemic proportion of scarlet fever incidence rises to a mean $maxEV$ = 0.76 across the nine time–windows. With the power law + gravity HHH model extended to include random effects, major epidemic outbreaks are observed in only two of the nine time–windows: six and eight. Although marginally, the epidemic proportion of scarlet fever incidence in the other seven time–windows are nevertheless extremely high.

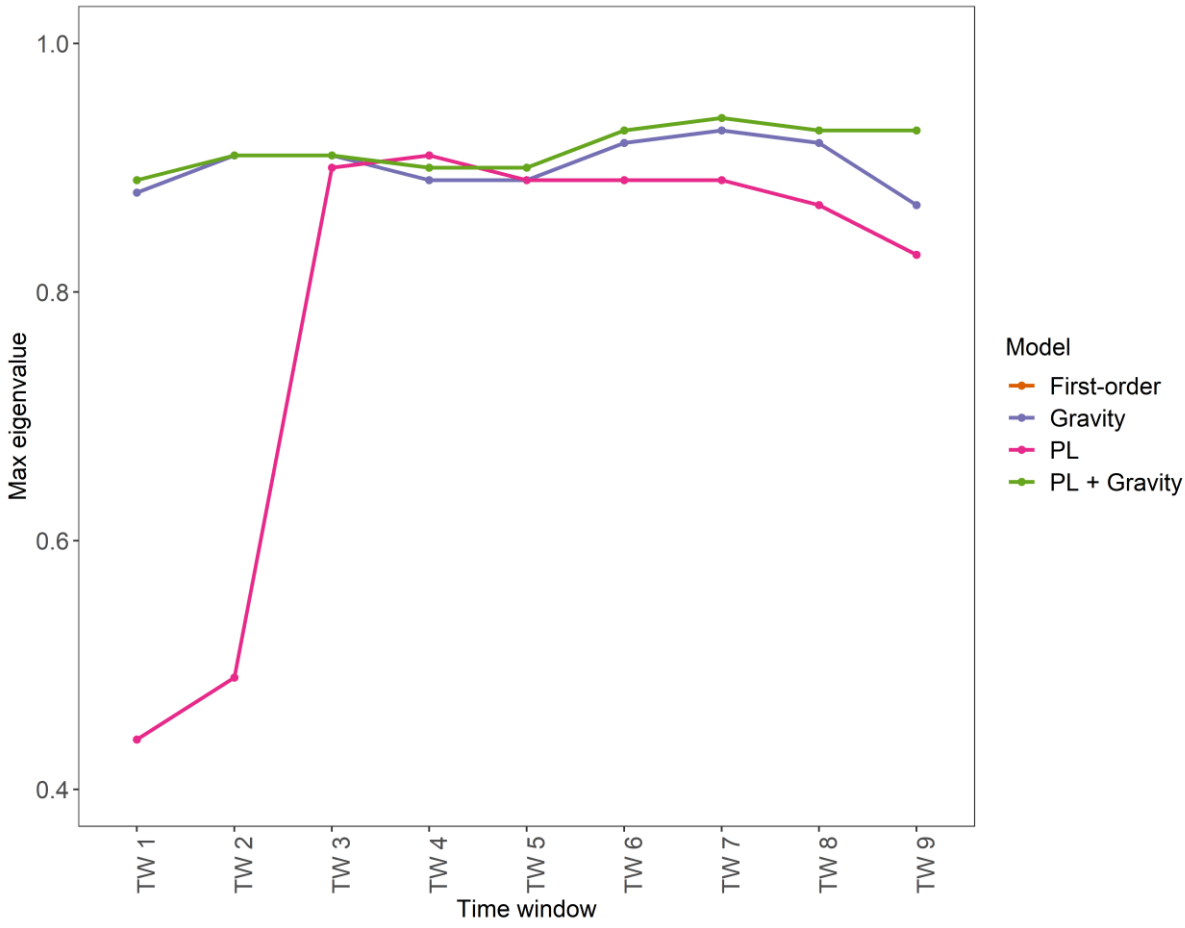

Figure 7.12 *Maximum eigenvalues for four HHH model formulations analysing scarlet fever spread in Lancashire, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

Predictive Assessment Scores

A goodness-of-fit assessment using AIC values for each HHH model formulation reveals the HHH power-law + gravity model provides the best fitting model across the nine time-windows compared to the four other models, with a mean AIC score = 73931.3 (7.3931 x 10^5). Assessment of predictive mean scores using scoring rules indicate the power-law + gravity RI HHH model is the most parsimonious model, suggesting the best fit of the seven models fitted across the nine time–windows in Lancashire (*LogS* = 1.32, *RPS* = 1.19). A breakdown of AIC values and predictive mean scores by time–window and model fit for scarlet fever in Lancashire can be found in Appendix V (Tables V.38 and V.42).

7.2 Results: South Wales Region

7.2.1 Pertussis

Overdispersion

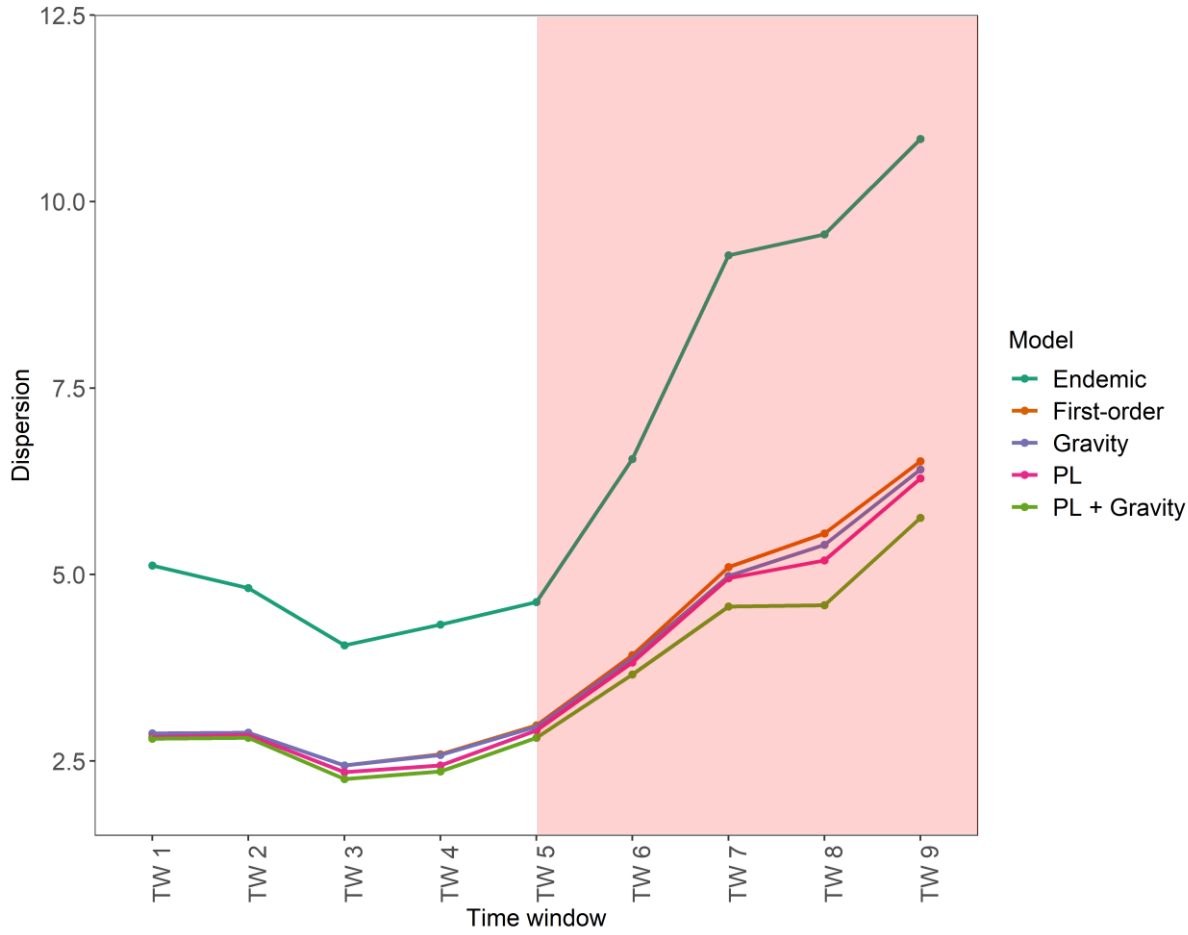

Figure 7.13 *Dispersion parameter estimates for four HHH model formulations analysing pertussis spread in South Wales, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend. Pink shaded area denotes the vaccine era.*

Assessment of the overdispersion parameter for the HHH model formulations reveals significantly greater levels of overdispersion estimated with the endemic HHH model compared to other model formulations (See Fig. 7.13). The overall mean ψ parameter =

6.58 (95% CI: 6.15–7.00). The least amount of residual heterogeneity is detected in time–window three (ψ = 4.05, 95% CI: 3.87–4.22). From time–window three onwards, there is a gradual increase in the overdispersion parameter estimates in the next two time–windows (Fig. 7.13), rising to ψ = 4.63 (95% CI: 4.43–4.84) by time–window five. With the advent of routine mass vaccination by the end of time–window six (1952-57), a significant increase in the level of overdispersion is observed, increasing to ψ = 6.55 (95% CI: 6.19–6.91). The rise in overdispersion is dramatic over the course of the further three time–windows, reaching a high of ψ = 10.84 (95% CI: 9.93–11.75) in the final time–window (1964-69).

With the inclusion of epidemic components and neighborhood adjacency in the first-order HHH model formulation, less residual heterogeneity is observed (Fig. 7.13), although residual heterogeneity is still significant (ψ = 3.87, 95% CI: 3.58–4.16). Echoing the overdispersion trend observed in the endemic HHH models, the least parameter estimate was observed in time–window three (ψ = 2.44, 95% CI: 2.31–2.56). The level of overdispersion gradually increases across successive time–windows rising to ψ = 3.92 (95% CI: 3.67–4.16) in time–window six, before growing substantially in vaccine-era time–windows, reaching a maximum of ψ = 6.52 (95% CI: 5.90–7.14). There is negligible difference in the level of overdispersion with the inclusion of a gravity component in the HHH model (Fig. 7.13), with a slightly lower mean of ψ = 3.82 (95% CI: 3.54–4.10). Once again, the same pattern of growth in the level of residual heterogeneity by individual time–window is visible for the gravity HHH models as observed with previous first-order and endemic formulations (Fig. 7.13). A power-law + gravity HHH model formulation sees further reduction in the overall level of residual heterogeneity across the nine time–windows, with a mean ψ parameter = 3.51 (95% CI: 3.25–3.77).

Distance Decay

Negative distance decay estimates are observed for the power-law HHH models across all time–windows (see Fig. 7.14), with the exception of time–window one (d = 0.95, 95% CI: 0.48–1.42). The overall mean d = -0.73 (95% CI: -1.40–-0.05). The strongest negative distance decay parameter estimated was in time–window five (d = -1.38: 95% CI: -2.20–-0.57), coinciding with the dramatic fall in birth rate, and thus rate of susceptible recruitment, in the immediate years following the baby boom. The distance decay parameter does recover some strength in the final two time–windows, rising from d = -1.38 in time–window five (1952-57) to d = -0.32 in time–window nine (1964-69).

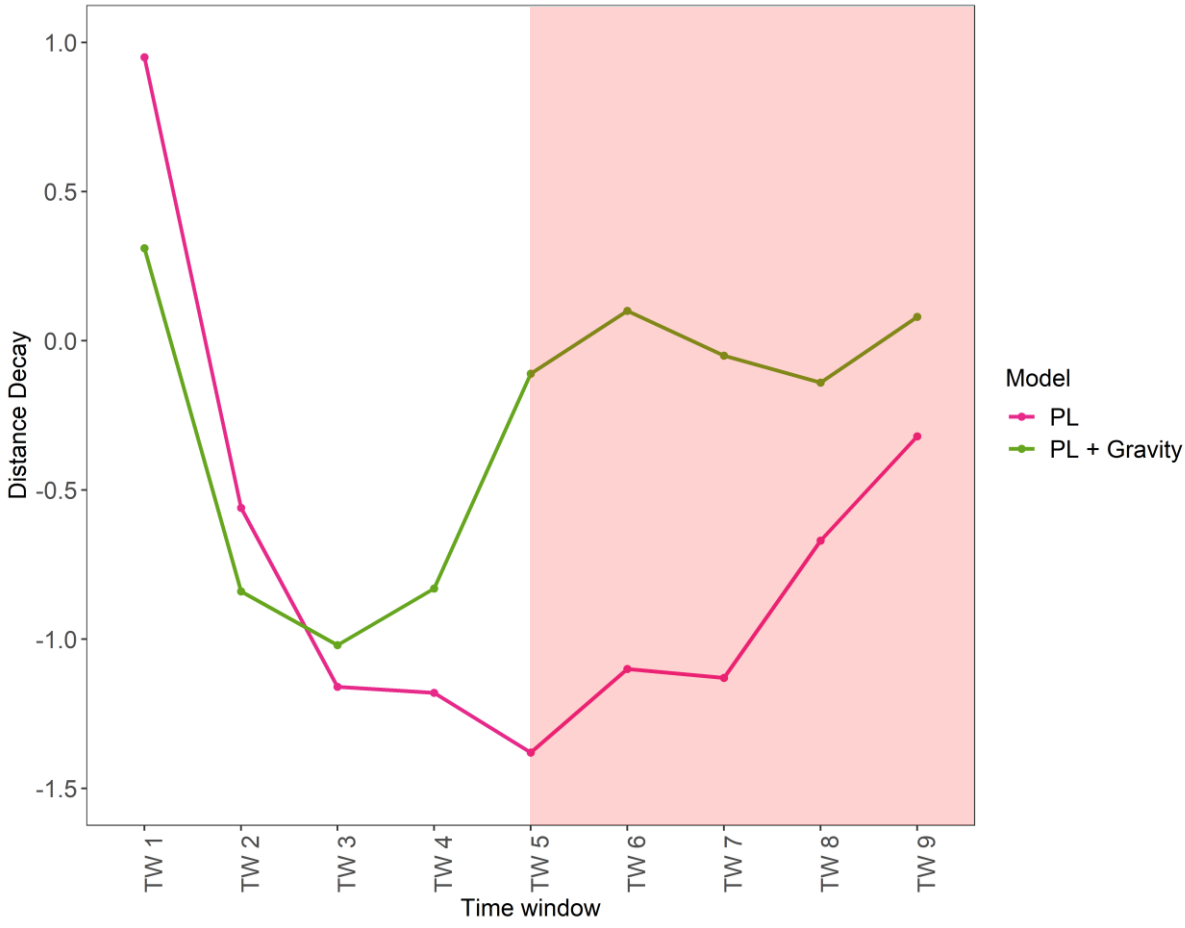

Figure 7.14 *Distance decay parameter estimates for two HHH model formulations analysing pertussis spread in South Wales, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend. Pink shaded area denotes the vaccine era.*

Gravity Parameter

The mean parameter estimate for the gravity-extended HHH models is $\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$ = 0.54 (95% CI: 0.17–0.92), with negative parameter values estimates reported in the first two time–windows, coinciding with the wartime years (see Fig. 7.15). With the addition of a power-law to the gravity HHH model, there is an overall positive increase in the strength of the gravity parameter estimates with a rising trend of growth in the gravity parameter estimates across successive time–windows are notable (Fig. 7.15).

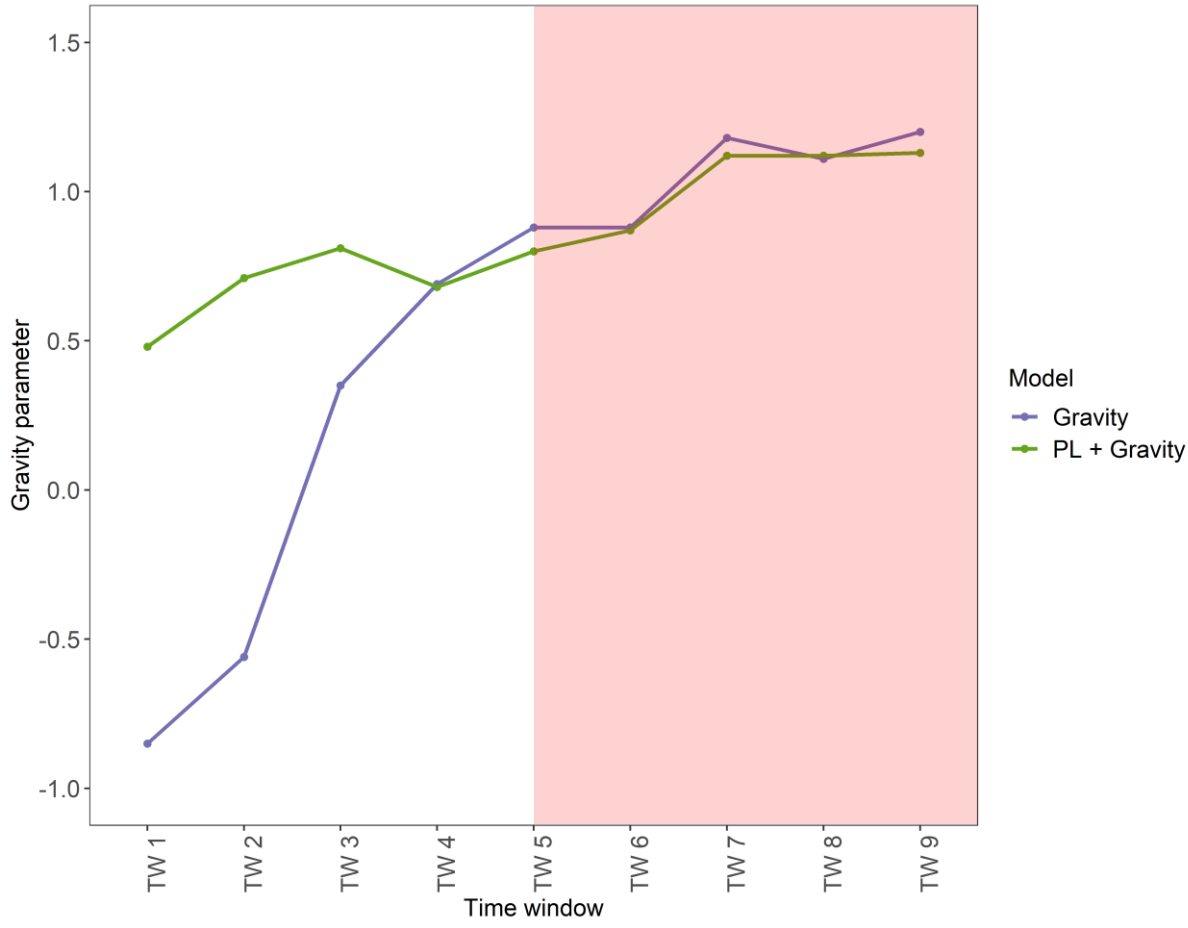

Figure 7.15 *Gravity parameter estimates for two HHH model formulations analysing pertussis spread in South Wales, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend. Pink shaded area denotes the vaccine era.*

maxEigenvalues

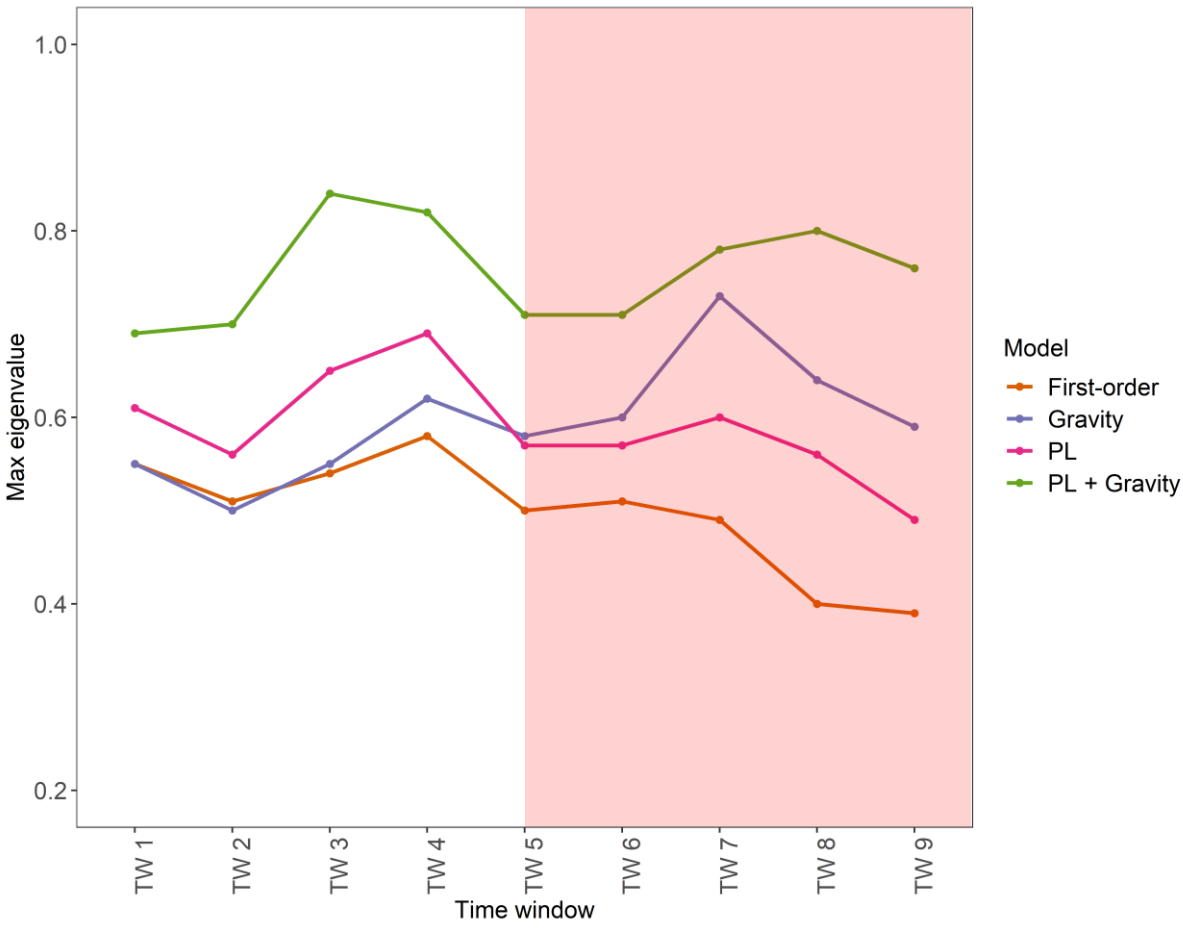

Figure 7.16 *Maximum eigenvalues for four HHH model formulations analysing pertussis spread in South Wales, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend. Endemic model excluded due to absence of epidemic component. Pink shaded area denotes the vaccine era.*

Assessment of the maximum eigenvalues reveals much lower epidemic proportions of pertussis incidence across the nine time–windows compared to measles incidence (see Fig. 7.16). The mean maximum eigenvalue for the first-order HHH model formulation is 0.50, rising to 0.60 when the model is extended to include a gravity model component. When extending the first-order model to include power-law transmission, there is an increase of 0.09 (in other words, nine percent), with $MaxEV$ = 0.59. Extending the first-order HHH

formulation to include both a gravity model component and power-law transmission increases the mean maximum eigenvalue to 0.76, a 26% increase in the epidemic proportion of pertussis incidence compared to the first-order HHH model. According to this model formulation, the most prominent period of epidemic pertussis activity was recorded in time–window three (1946-51; $MaxEV$ = 0.84), closely followed by time–window four (1949-54; $MaxEV$ = 0.82) and time–window eight (1961-66; $MaxEV$ = 0.80), all coinciding with growth periods in birth rates.

Predictive Assessment Scores

An assessment of the goodness-of-fit of the five HHH model formulations (endemic, first–order, power-law, gravity and power-law + gravity), based upon AIC values, reveals the power-law + gravity HHH model provides the most consistent best fit, with a mean AIC value of 30161.1 (3.01611×10^5) across the nine time–windows. A full breakdown of AIC values time–window and model fit for pertussis in South Wales can be found in Appendix V (Tables V.37).

7.2.2 Measles

Overdispersion

Assessment of residual heterogeneity in the HHH model formulations reveals there is significantly greater levels of overdispersion with the endemic HHH model, compared to other model formulations (see Fig. 3.17). The overall mean ψ parameter = 9.63 (95% CI: 9.36–9.90). The exclusion of epidemic components from the HHH endemic models results in a significant rise in residual heterogeneity and an inferior fit to the weekly incidence data modelled. With the inclusion of epidemic components and disease transmission via neighbourhood adjacency in the first-order HHH models, there is a sizable fall in residual heterogeneity in each time–window (Fig. 3.17), although overdispersion is still significant.

The overall mean ψ parameter = 3.06 (95% CI: 2.95–3.18). The overdispersion parameter increases progressively over successive windows from a low of 2.28 (95% CI: 2.10–2.37) in time–window three to a high of 3.86 (95% CI: 3.72–4.00) at the end of the study period, in time–window nine. Incorporating a gravity measure results almost no change in the level of residual heterogeneity compared to the first-order HHH models. The overall mean ψ parameter = 3.04 (95% CI: 2.93–3.16). However, the inclusion of a power law to account for long-distance transmission events in lead to a further reduction in the overall level of overdispersion across the nine time–windows (Fig. 3.17), with an overall mean ψ parameter = 2.51 (95% CI: 2.42–2.61).

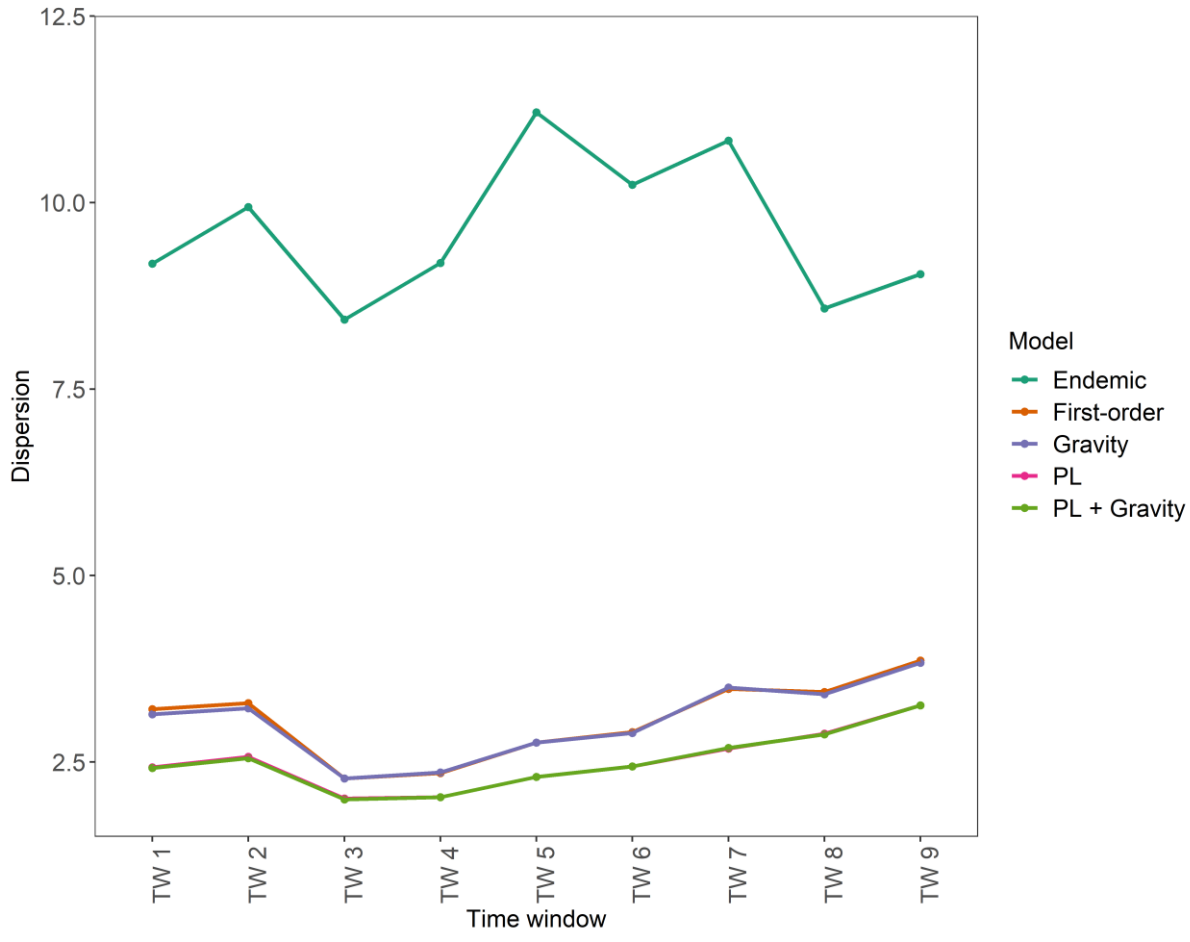

Figure 7.17 *Dispersion parameter estimates for four HHH model formulations analysing measles spread in South Wales, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

The least level of overdispersion was detected in time–windows time–window three (1946-1951), coinciding with the beginning of the baby boom period. Across successive windows, the value of the overdispersion parameter slowly climbed, rising from 2.01 (95% CI: 1.93–2.09) in time–window three to 3.26 (95% CI: 3.14–3.38) in time–window nine. Extending the HHH model formulation to include a power-law and gravity measure reveals almost no difference in the value of the overdispersion parameter estimated for individual time–windows (Fig. 7.17) or the mean ψ parameter = 2.51 (95% CI: 2.41–2.60).

Distance Decay

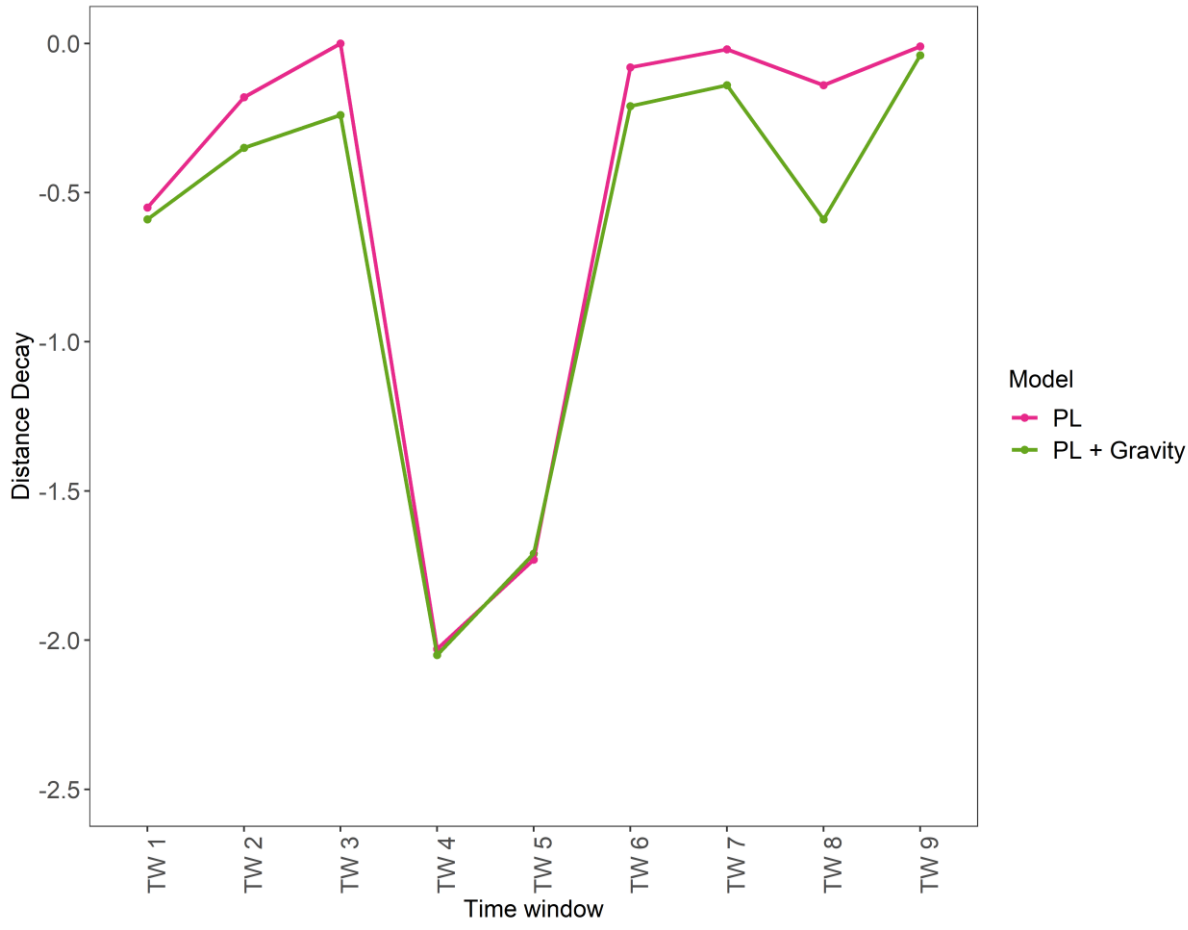

Figure 7.18 *Distance decay parameter estimates for two HHH model formulations analysing measles spread in South Wales, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

Negative distance decay estimates are observed for power-law HHH models across all time–windows with the exception of time–window three (Fig. 7.18). The overall mean d parameter = -0.53 (95% CI: -0.81–-0.24). The most negative distance decay parameter estimated was in time–window four (d = - 2.03: 95% CI: -2.38–-1.68), coinciding with the dramatic fall in fertility rate during this period, and thus rates of susceptible recruitment. In later time–windows, the distance decay parameter estimates are weakly negative, with the weakest parameter value observed in time–window nine (d = -0.01, 95% CI: -0.22–0.19). The power-law + gravity HHH model formulation produces more negative distance decay parameter estimates, with an overall mean d = -0.66 (95% CI: -0.99–-0.33).

Gravity Parameter

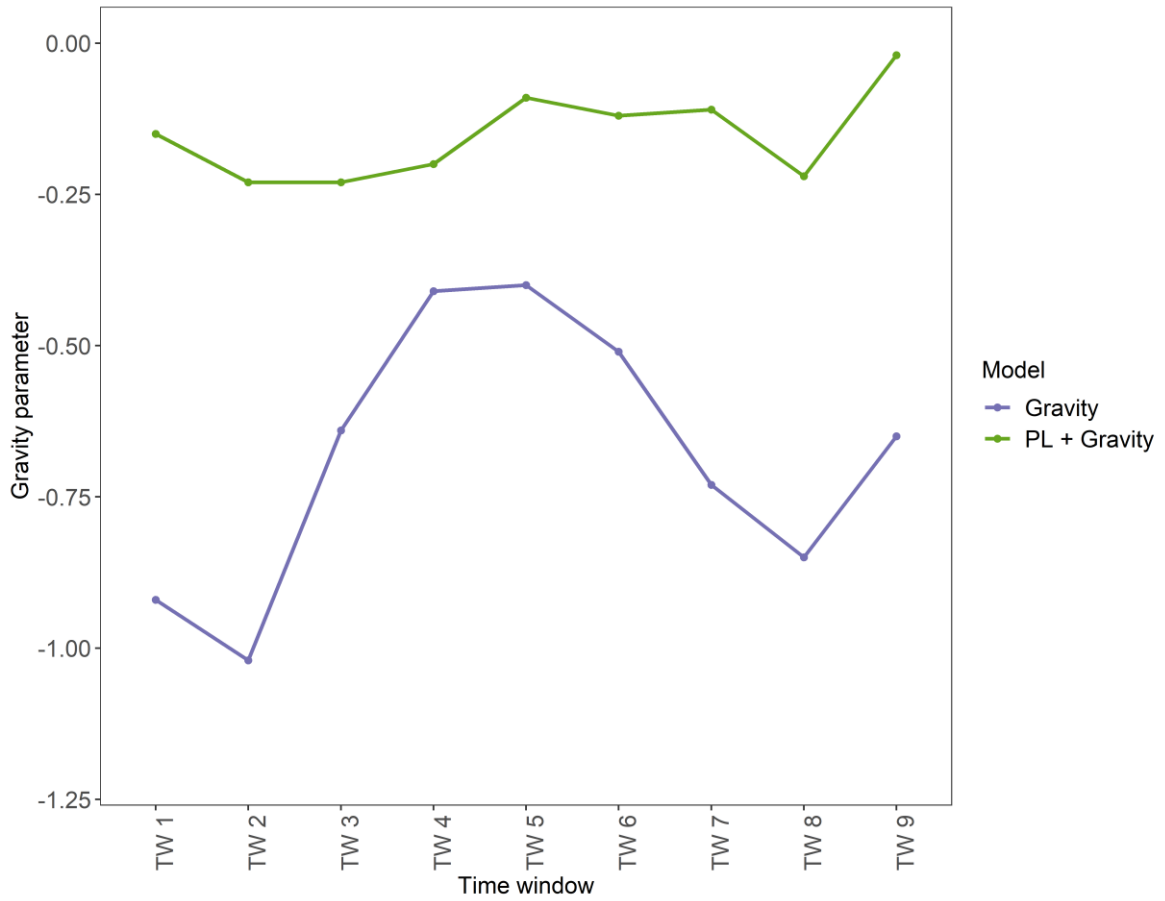

Figure 7.19 *Gravity parameter estimates for two HHH model formulations analysing measles spread in South Wales, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

The mean parameter estimate for the gravity HHH models is $\boldsymbol{\beta}^{(\boldsymbol{\phi})}_{\mathbf{log}\,(\boldsymbol{pop})}$ = -0.68 (95% CI: -0.79–-0.57), with negative parameter estimate values estimated for all time–windows. The largest gravity parameter estimates are found in the first two time–windows (see Fig. 7.19). The weakest negative gravity parameter estimate was found in time–window five ($\boldsymbol{\beta}^{(\boldsymbol{\phi})}_{\mathbf{log}\,(\boldsymbol{pop})}$ = -0.40, 95% CI: -0.50–-0.30), before rising across successive time–windows (Fig. 7.19). A mean $\boldsymbol{\beta}^{(\boldsymbol{\phi})}_{\mathbf{log}\,(\boldsymbol{pop})}$ parameter value of -0.15 (95% CI: -0.21–-0.10) is calculated across the nine time–windows for the power-law + gravity HHH formulation, with significantly weaker negative gravity parameter estimates in each time–window. The value of the gravity parameters estimated are relatively stable and weakly negative across the study period (Fig. 7.19).

maxEigenvalues

Assessment of the maximum eigenvalues reveals significant similarities across the first–order, gravity, power-law and power-law + gravity HHH model formulations (see Fig. 7.20). Since the power-law + gravity HHH models were found to provide the overall best goodness-of fit (Table V.35, Appendix V), the maximum eigenvalues for the gravity + power-law HHH models will be reported here. Epidemics outbreaks were detected in time–windows five, six, seven and nine, with the eigenvalue exceeding unity (value of 1). The scale of epidemic activity is extremely significant in other time–windows in which an outbreak was not detected; the epidemic proportion of measles cases in South Wales is equivalent to 99% (0.96) in time–window one, while in time–window eight this had fallen slightly to 96% (0.96). This fall in time–window eight might reflect an exhaustion of susceptibles due to the outbreak detected in the preceding time–window.

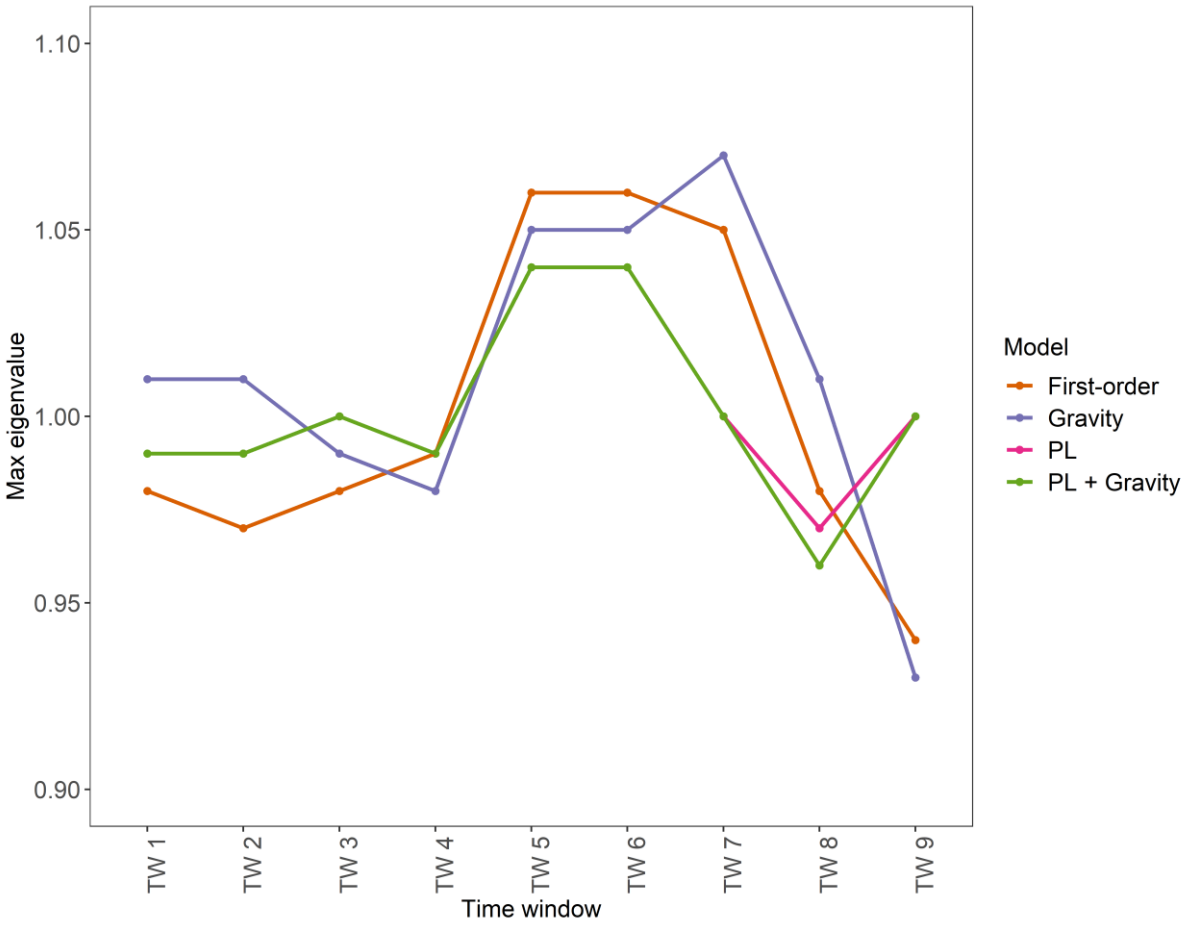

Figure 7.20 *Maximum eigenvalues for four HHH model formulations analysing measles spread in South Wales, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend. Endemic model excluded due to absence of epidemic component.*

Predictive Assessment Scores

An assessment of the goodness-of-fit of the HHH model formulations based upon the AIC values reveals power-law + gravity HHH model provides the most consistent best fit of the five model formulations (endemic, first–order, power-law, gravity and power-law + gravity), with a mean AIC value of 62699.0 (6.2699×10^5) across the nine time–windows. In time–window nine (1964-69), the power-law HHH model formulation was found to provide the best fit. A breakdown of AIC values by time–window and model fit for measles in South Wales can be found in Appendix V (Tables V.35).

7.2.3 Scarlet fever

Overdispersion

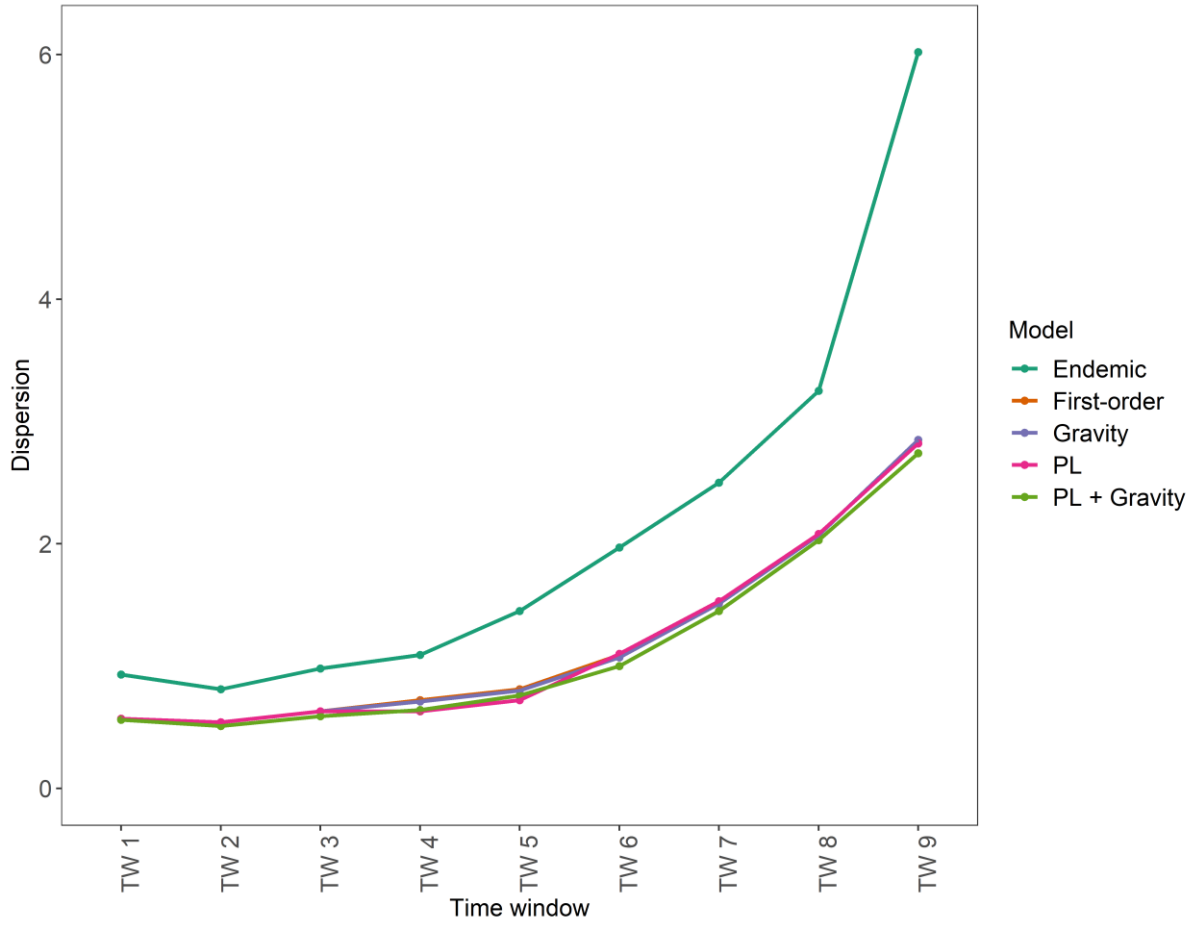

Figure 7.21 *Dispersion parameter estimates for four HHH model formulations analysing scarlet fever spread in South Wales, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

Assessment of residual heterogeneity in HHH model formulations reveals significantly lower levels of overdispersion for scarlet fever, compared to measles and pertussis, using the endemic HHH model. The overall mean ψ parameter = 2.11 (95% CI: 1.94–2.28). Overdispersion parameter values rise consistently across the nine time–windows (see Fig.

7.21), from a low of 0.93 (95% CI: 0.88–0.98) in time–window one, indicating underdispersion, to a high of 6.02 (95% CI: 5.55–6.50) in time–window nine. The velocity of growth in the level of residual heterogeneity increases substantially in time–windows seven to nine. The largest rise came between time–windows eight and nine, jumping from ψ = 3.25 (95% CI: 2.92–3.57) to ψ = 6.02 in the final time–window.

With the inclusion of the epidemic and spatiotemporal components in first-order HHH models, there is a sizable fall in the estimated level of residual heterogeneity in each time–window, with considerable underdispersion detected in time–windows one to five (Fig. 7.21). However, the presence of underdispersion declines progressively over time, with the overdispersion parameter rising across successive time–windows. The overall mean ψ parameter = 1.20 (95% CI: 1.09–1.32). The gravity HHH formulation produces a slight increase in the level of residual heterogeneity observed for each time–window model fit (Fig. 7.21). The overall mean ψ parameter = 1.25 (95% CI: 1.13–1.37). Overdispersion is detected from time–window six onwards, rising to a high of 2.85 (95% CI: 2.56–3.12) by time–window nine. The power-law HHH model sees an increase in the overall level of overdispersion, with mean ψ = 1.26 (95 CI: 1.14–1.38). Mirroring the pattern of previous HHH model formulations, high levels of underdispersion are detected in time–windows one to five (Fig. 7.21).

Applying a power-law + gravity HHH model formulation reveals a fall in the strength of overdispersion parameters estimated for individual time–windows, with a reduced overall mean overdispersion parameter estimate of ψ = 1.20 (95 CI: 1.09–1.32). Underdispersion is detected in time–windows one, three and five. From time–windows six onwards, the level of overdispersion parameter increases consistently across successive time–windows, reaching a high of ψ = 2.74 in time–window nine (95 CI: 2.47–3.02).

Distance Decay

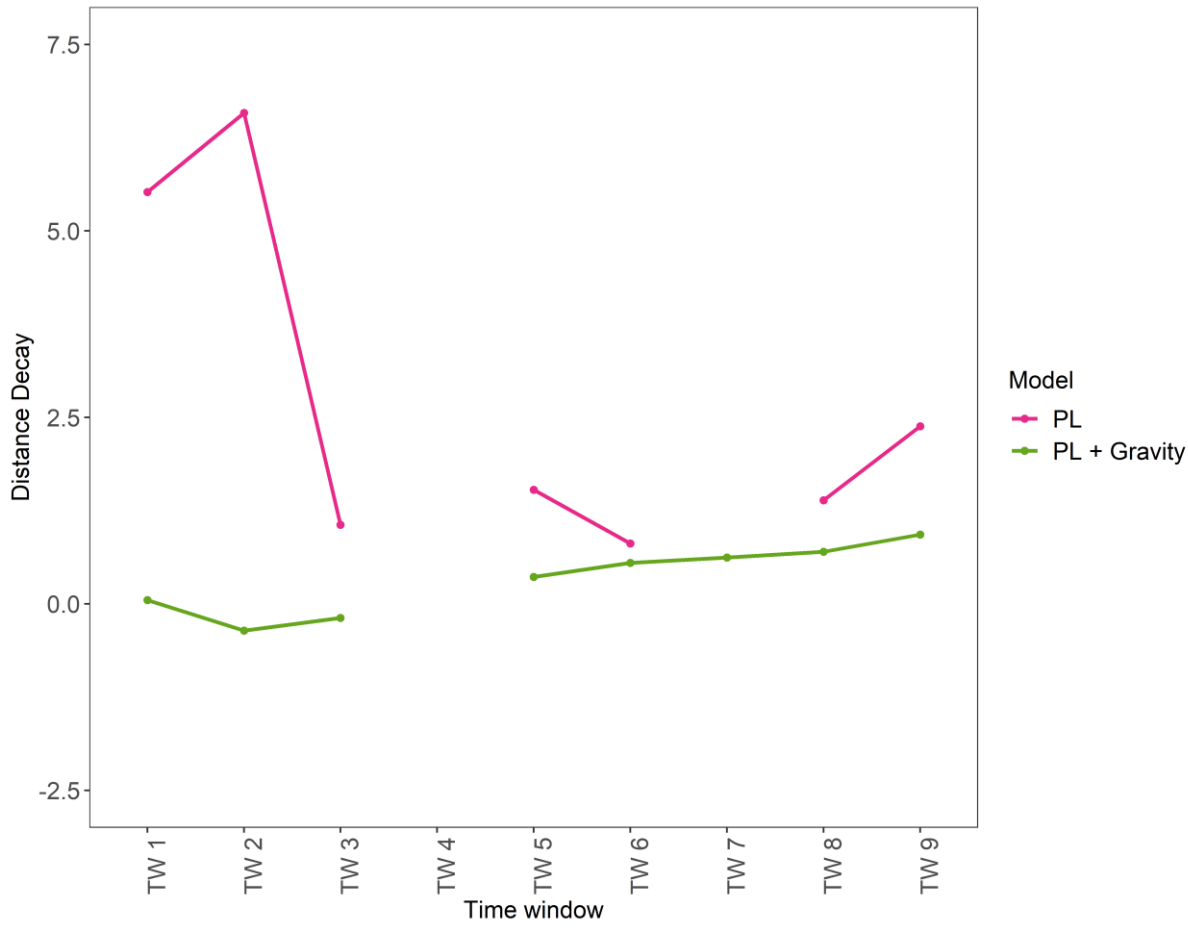

Figure 7.22 *Distance decay parameter estimates for two HHH model formulations analysing scarlet fever spread in South Wales, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

Positive distance decay estimates are observed using power-law HHH models across seven of the nine time–windows for which convergence was achieved in the modelling process and estimates for distance decay were generated (see Fig. 7.22). There was a failure of convergence in time–windows four and seven. The largest distance decay parameter was estimated in time–window two (d = 6.58, 95 CI: -1.74–10.77). The mean distance decay parameter for the power-law HHH models is d = 2.75 (95 CI: -3.81–3.55).

As observed with measles and pertussis incidence, addition of a gravity measure sees a fall in the strength of the distance decay parameter (Fig. 7.22). No distance decay parameter estimate is provided for time–window four due to lack of convergence. There are two negative distance decay parameter estimates: time–window three (d = 0.36, 95 CI: -0.58–-0.13) and time–window five (d = -0.19, 95 CI: -0.38–-0.01). The strength of the distance decay parameter estimates increases across successive windows, from time–window five to time–window nine, rising to a maximum estimate of d = 0.93 (95 CI: 0.61–1.25).

Gravity Parameter

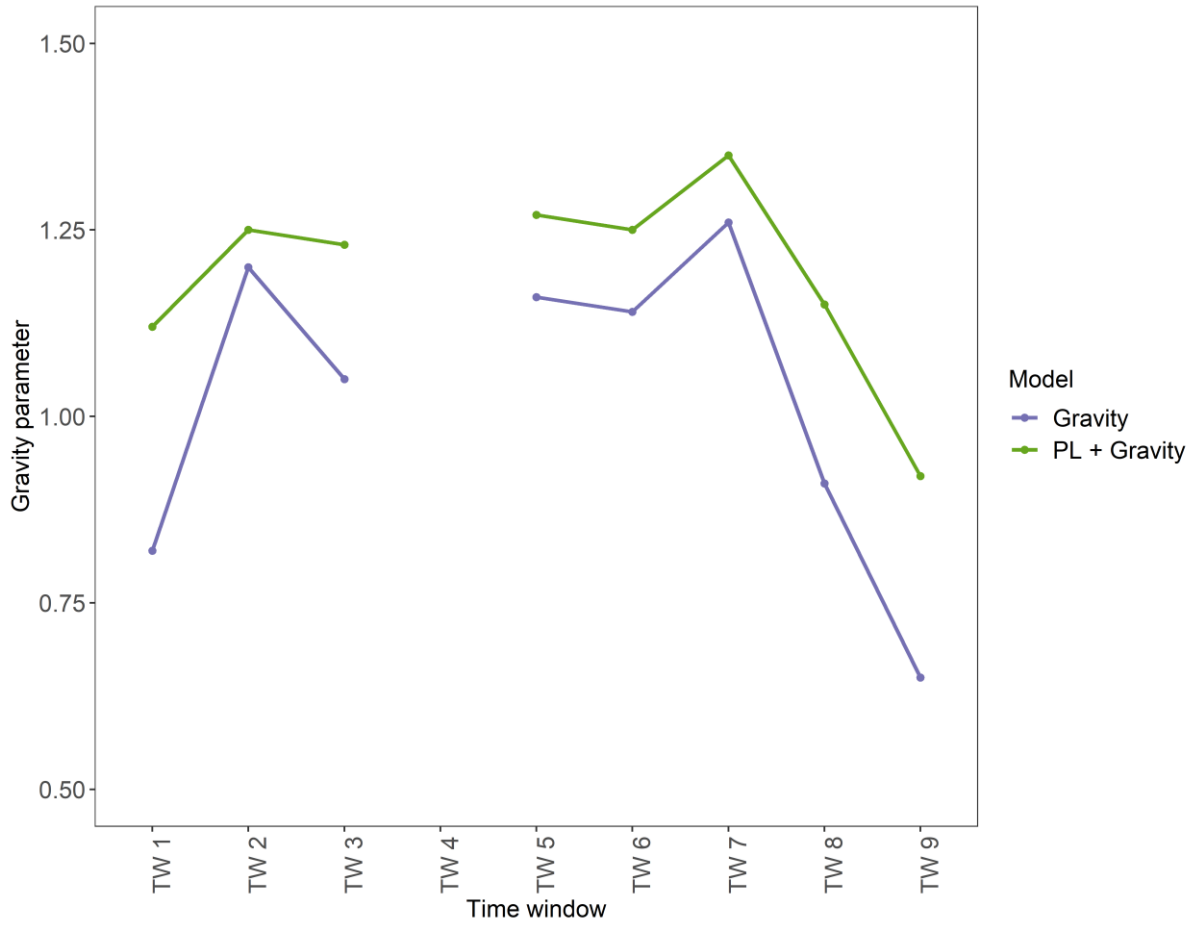

Figure 7.23 *Gravity parameter estimates for two HHH model formulations analysing scarlet fever spread in South Wales, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend.*

The mean parameter estimate for the gravity HHH models is $\boldsymbol{\beta}^{(\boldsymbol{\phi})}_{\mathbf{log}\,(\boldsymbol{pop})}$= 1.02 (95% CI: 0.79–1.26). There is no gravity parameter estimate for time–window four due to lack of model convergence. The largest parameter estimates for the gravity component are visible in the middle of the study period, in time–windows five, six and seven (see Fig. 7.23). The strength of gravity parameter estimates decreases by nearly 50% between time–windows seven and nine, falling to $\beta^{(\phi)}_{\log\,(pop)}$= 0.65 (95 CI: 0.41–0.88). Inclusion of a gravity measure in the power-law HHH model sees a significant increase in the strength of the gravity parameter estimates across the time–windows (excluding time–window four). The overall mean $\boldsymbol{\beta}^{(\boldsymbol{\phi})}_{\mathbf{log}\,(\boldsymbol{pop})}$= 1.19 (95 CI: 1.10–1.29). It is noticeable that there is significantly less variation in the size of the parameter estimates across time–windows (Fig. 7.23). However, there is a significant downward trend in the final two windows in the strength of the gravity parameter, echoing the previous gravity HHH model formulation.

maxEigenvalues

Assessment of maximum eigenvalues reveals the epidemic proportion of scarlet fever cases across individual time–windows is far lower than the epidemic proportions observed for measles and pertussis incidence in the South Wales region (see Fig. 7.24). A mean $MaxEV$ of 0.34 using the first-order HHH model formulation suggests approximately just one-third (30%) of scarlet fever cases in South Wales are the result of epidemic activity. Extending the first-order HHH model to include a power-law only increases the mean epidemic proportion of scarlet fever incidence by 2%, to 0.36. However, extending the first-order model to incorporate a gravity component increases the epidemic proportion of scarlet fever incidence to 0.47, an increase of 17%. Inclusion of both a power-law and gravity measure see the mean epidemic proportion of scarlet fever incidence jumps considerably ($maxEV$ = 0.63).

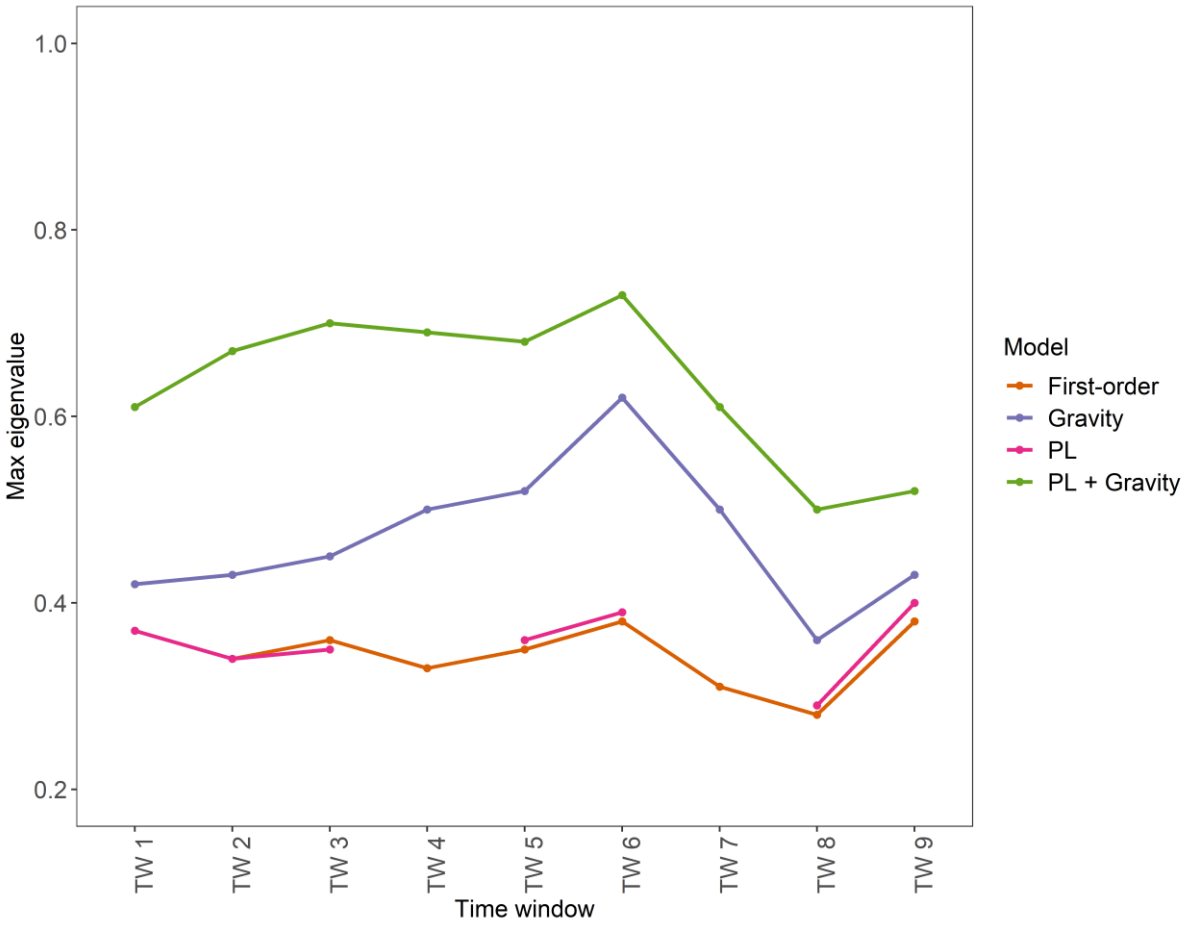

Figure 7.24 *Maximum eigenvalues for four HHH model formulations analysing scarlet fever spread in South Wales, across nine time–windows (1940–1969). Dots represent estimates for each window. Fitted lines have been applied to the points for the time trend. Endemic model excluded due to absence of epidemic component.*

Predictive Assessment Scores

A goodness-of-fit assessment comparing AIC values for each HHH model formulation reveals the gravity + power-law HHH model provides the most consistent, best fitting model across the nine time–windows, with a mean AIC score of 3.0561×10^5 (30561.1). A breakdown of AIC values by time window and model formulation for scarlet fever in South Wales can be found in Appendix V (Tables V.39).

7.3 Discussion of Findings

In this chapter, several sub-model formulations of the HHH model have been fitted and utilised to analyse the endemic-epidemic dynamics of pertussis, measles and scarlet fever, in a quest to build a more detailed understanding of the spatiotemporal dynamics of measles, pertussis and scarlet fever incidence and transmission within the regional metapopulations of South Wales and Lancashire. Such understanding can help inform interpretation of the rise and fall of endemic threshold populations over time and space, elucidating on the influence of spatial structure, mobility and demographic stochasticity on disease persistence.

The first major extension to the baseline first-order HHH models fitted to better capture the endemic-epidemic dynamics of the three childhood infections under analysis in the Lancashire and South Wales region was the inclusion of a power-law decay of spatial interaction. This was embedded into the spatiotemporal epidemic component and estimated jointly with all other unknown parameters using likelihood inference. The application of power-law transmission kernels to model the spatial dynamics of infectious diseases can be found in models for the 2001 UK foot-and-mouth epidemic (Keeling et al., 2001; Ster and Ferguson, 2007) and modelling human infectious diseases. Geilhufe et al. (2014) used power-law weights in place of traffic data to predicting influenza incidence in Northern Norway. Motivated by the work of Brockmann et al. (2006), who claimed human travel can be well-described by a decreasing power-law of distance thus serving as a ‘starting point for the development of a novel class of models for the spread of human infectious diseases’ (Brockmann et al., 2006: 465), Meyer and Held (2014) incorporated a power-law in the neighborhood component of the HHH model to describe the spread of invasive meningococcal diseases and influence in Germany between 2002 and 2008, using aggregated count data.

The incorporation of power-law distance decay within the networks of geographical districts of the regional metapopulations of Lancashire and South Wales required a distance measure on which the power-law acts require definition. An important characteristic of the power-law is its slow convergence to zero, indicative of a heavy tailed distribution in which the distribution is not exponentially bounded (Cappé et al., 2002; Clauset et al., 2009). Within the context of HHH modelling, this 'heavy tail' enables intermittent long-range transmissions of infection in addition to primary short-distance disease transmissions to be accounted, thus allowing consideration of a mixture of diffusion processes and their role in disease persistence within a regional metapopulation. The analysis here found that power-law formulations performed better than the baseline first-order interaction models, with significant improvements in goodness-of-fit based on assessment of AIC Values. This confirms that the power-law distribution of short-time human travel translates to the modelling of infectious disease spread. The heavy tail of the power-law which enables long-range dependence between cases to be accounted results in greater weight given to the importance and strength of the epidemic component in the power-law HHH and power-law + gravity HHH models. This is reflected in the growth of the epidemic proportion of disease cases with increases observed in maximum eigenvalues in time–windows with the inclusion of a power-law**,** and outbreaks more frequently detected in time–windows.

Allowing for the long-range transmission of cases recognises the hierarchical diffusion of childhood diseases which would correlate with the existence of a well-established urban hierarchy, with the infection waves originating in the largest urban centres before spreading to settlements next in size and so on, through to the smallest, least densely populated, and perhaps least connected, settlements (Bartlett, 1957; Grenfell et al., 2001). This is arguably more prominently the case for the Lancashire region. The spatial structure of the regional metapopulation and network of connectivity which exists between larger towns and the endemic centres of Liverpool CB and Manchester CB is heavily associated with the emergence and development of the textile industry in the late-eighteenth and early-

nineteenth centuries. In South Wales, the industrial revolution arrived later with the exploitation of the South Wales coalfield gathering pace midway through the nineteenth century. But unlike Lancashire, where the various functions and markets associated with the textile industry were widely distributed across the region and thus nurturing the development of urban settlements dozens of kilometres away from the two dominant urban centres in the South of the region, urban development, and thus the population distribution, of South Wales became concentrated in a relatively narrow geographical area, within the mineral-rich valleys of Glamorgan and West Monmouthshire, and the towns located at the valley mouths on the coast, from which resources and goods were exported, and saw these settlements grow exponentially to become the main regional urban and population centres.

The significant fall in the distance decay parameter in the final two time–windows, for both the power-law and power-law + gravity HHH model formulations, reflects the extensive impact of disease intervention in reducing the viability of long–distance transmission of pertussis infection. This mirrors findings of the survival analysis of pertussis hotspots in Lancashire and South Wales in Section 6.3. The survival time of fadeout periods increases significantly after the onset of vaccination, as chains of transmission breakdown and inhibit long-range transmission of infection. The vaccination of susceptibles in districts that are home to satellite settlements or major towns that are located, but do not immediately neighbour, the regional endemic centres of Manchester and Liverpool CB creates a form of productive barrier against commuter–driven spread of infection. High vaccination uptake shields susceptibles from disease transmission fuelled by significant spatial interaction with the two conurbations.

Adopting a power-law approach is very useful in the absence of movement network data, which is certainly difficult to obtain when working with historical epidemiological surveillance data. However, if such data were available, neighbourhood weights in a power-law HHH model could instead be based on the more realistic connectivity between districts described

by commuter traffic data for instance rather than using discrete spatial measures. This possibility has been explored by Schrödle et al. (2012) in their study of the spatiotemporal spread of Coxiellosis in Swiss cows, as well as by Geilhufe et al. (2014) in their study of influenza spread in Northern Norway.

In the past, spatial coupling between district and regions has been assumed to be an inverse function of distance (Okubo and Levin, 2013). However, this assumption has been found to be too simplistic to account for the complex interactions between humans since movement between large communities is often disproportionately more frequent than between smaller settlements (Erlander and Stewart, 1990). This more complex pattern of human movement results in 'gravity transmission' (Murray and Cliff, 1975; Cliff et al., 1993). What is more certain is that the network of spatial spread, in other words epidemiological coupling, can often relate to connectivity network within a host regional metapopulation and urban hierarchy. Influenced by the work of Xia et al. (2004), who proposed a gravity model for modelling the spread of measles in regional metapopulation, a gravity model was incorporated in a sub-model HHH formulation. This is based on the straightforward assumption that larger settlements attract more individuals, resulting in greater likelihood of disease spread; for instance, one would expect more traffic to a regional conurbation such as Manchester CB, which are thus expected to import a higher number of cases from surrounding districts (Bartlett, 1957), as well as function as a centre of endemicity.

The findings described for the sub-model formulations which include a gravity measure provide significant insights into how the interplay between temporal changes in population distribution, epidemiological coupling and regional spatial structure affect the nature of disease spread, and thus influence changes in disease endemicity and growth of endemic threshold populations over time. In both regional metapopulations, there is significant growth in the strength of the gravity parameter over the course of the nine time-windows, further increasing with the inclusion of a power law decay of spatial interaction, and the

inclusion of a gravity measure almost uniformly improves the goodness-of-ft of the HHH models in each time window. This indicates that commuting between subpopulations, with spatial interactions often concentrated around urban centres which function as the major population and employment centres with the region, supports the transmission and persistence of childhood disease. This reflects the evidence of tight spatial coupling between the urban centres of Manchester CB, Liverpool CB and Cardiff CB and surrounding districts generated by the correlation analyses performed in Sections 4.6 and 4.7. Urban centres are found to possess the strongest correlation with the regional pattern of disease activity, followed by neighbouring districts regardless of the significant differences in total, and by extension susceptible, population size. The finding is also corroborated by the conclusions of past research on the spatial dynamics of influenza and measles, which indicate more heavily populated districts draw a greater quantity of infection from neighbours than smaller districts, reflecting commuter-type imports (Viboud et al., 2006; Meyer and Held, 2017).

Manchester CB and Liverpool CB in Lancashire, and Cardiff CB in South Wales, dominate their respective regions. With growing deindustrialisation during the study period crippling free-standing towns in the surrounding metropolitan hinterland of the two regions, whether they be the historic mining districts of the Valleys in Glamorgan or the once thriving cotton mill towns of South Lancashire, a significant proportion of the urbanised populations of South Wales and Lancashire became reliant on a handful of regional urban centres for employment and commerce (Bax and Fairfield, 1978; Hebbert and Deas, 2000; Jenkins, 2014). Many urban districts and municipal boroughs home to small and medium-sized settlements were relegated from operating relatively autonomously as centres of industrial activity to serve a primarily residential function (Hartwell et al., 2004). This transformed the nature of individual movement and spatial interaction between urban centres and surrounding settlements, placing greater and greater importance on commuting over time. For instance, In his analysis of the changing nature of commuting patterns in North West

England, focusing on the counties of Lancashire and Cheshire, Warnes (1972) found the distribution of commuting distances changed significantly between the 1920s and the mid-1960s, from one of dominant regional variations, with long journeys characterising South Lancashire and short journeys being found in north east and the extreme east of Lancashire, to one where the evolving metropolitan situation in postwar Lancashire transformed the nature of travel. By the end of the study period, the large urban, employment centres, Liverpool CB and Manchester CB, had the shortest average commuting distances, with peripheral neighbours supporting the longest mean journey-to-work distances. This evolution in the pattern of travel ensures that higher rates of pertussis transmission, and elevated levels of pertussis by extension, are sustained in districts in the metropolitan hinterlands close to endemic centres of infection, by maintaining tight epidemiological coupling between districts through consistent spatial import and export of infection. However, some caution should be exercised when making inferences on the importance of commuting as a factor in the spread and persistence of pertussis in the Lancashire and South Wales metapopulations.

Firstly, a difficulty in interpreting high levels of spatial coupling due to greater levels of commuting is partly borne from the fact that mobile infectious individuals, whose presence in subpopulations in which they have travelled might be fleeting, do not necessarily pose significant dynamical consequences. Past research has highlighted that epidemic immigration within the context of measles sometimes has negligible dynamical consequences (Bjørnstad et al., 2002); transients infections are irrelevant as soon as the morbidity of the infection exceeds a certain threshold (in the case of measles, approximately 10 cases). This threshold is only crossed in minor populations during the troughs between epidemic outbreaks. However large settlements such as Liverpool CB and Manchester CB have population greater than the endemic threshold size for measles, as well as pertussis and scarlet fever based on the size of endemic threshold populations for each disease estimated in Chapter 5 and thus will almost consistently have number of cases notified

above the minimum mentioned for measles on a weekly basis. The influence of transient infectious individuals on the dynamical behaviour of childhood infection is of greater relevance if the number and density of infected individuals travelling is high and if transient individuals, whether they be susceptible or infected, function in local subpopulations in a similar manner to resident individuals (Xia et al., 2004). In this context, the probability of susceptibles acquiring non-local infection, from an endemic centre or large town for example, from mobile individuals would be much higher.

Secondly, processes of population and economic decentralisation in the postwar period, fundamentally changed commuting patterns among adults in Lancashire (Warnes, 1972; Hall, 1974), whilst the provision of rail and road infrastructure serving the region's subpopulations of Lancashire also transformed in the postwar period, due to the systematic closure of many sub-regional rail lines (as exemplified by the 'Beeching axe') and mass road construction driven by growth in private car ownership (Hartwell et al., 2004; Pooley et al., 2010). However, this does mean that the mobility of children of school age, or infants, who are the primary group susceptible to pertussis infection, also changed in such a way that it influenced the spatial dynamics and persistence of the disease on a local or regional scale. Pooley et al. (2010) found that the everyday mobility and range of travel among school children in Manchester and Lancaster had changed very little since between the 1940s and 2010. For instance, in Manchester, children travelled a mean distance of 1.3km in the 1940s, yet the mean trip distance among school children had only increased by 200 metres by the 2000s. Xia et al. (2004) emphasise that the transmission of childhood diseases, whether it be pertussis, measles or scarlet fever, is not a spatially random diffusion process, and instead follows the fastest route to a target location of individual due to directed movement, often preferring susceptibles or susceptible environments in close proximity rather than over greater distances (Xia et al., 2004).

In the pre-vaccine eras for pertussis and measles in particular, incidence of such infection was almost entirely restricted to children of primary school age or infants of nursey age. In this context, the scaling of diffusive movement of individuals and infection using a gravity model may be less advantageous than modelling with explicit movement data on travel distance. However, to acquire such data given the historical context, in this case mid-twentieth century England and Wales, is far easier said than done. Data on journey-to-school distances are not readily available in archives in the way that journey-to-work distances obtained from the Census of England and Wales utilised by Warnes (1972) might be. Data from national travel surveys provide some general information such as mean distance travelled to school from around the mid-1960s, but data prior to the 1960s is much more limited. In their study, Pooley et al. (2010) turned to a qualitative oral history approach in which cohorts of individuals were questioned about their mobility experience at different stages of their life in Manchester, Salford, Lancaster and Morecambe, to collect qualitative explanatory data of sufficient detail to generate estimates of travel distance in the 1940s. However, such data is drawn from a small sample and is the product of self-reporting, and thus must be handled with caution.

Although not included in the model formulation, one must consider the influence of vaccination coverage on the pertussis spread and persistence. Effective rates of vaccination coverage in later time–windows places greater importance on rescue effects to maintain regional persistence of an infection. There must be sufficient levels of geographical coupling between subpopulations achieved through frequent connections to fuel disease spread. This is demonstrated in the findings of the hotspot and survival analyses discussed in Section 6.4. With declining pertussis incidence in later time–windows in Lancashire and South Wales due to the onset of routine mass vaccination in 1957 (see Sections 5.4.2 and 5.5.2), commuting may take much greater prominence for ensuring disease persistence, as vaccination serves to deplete pools of susceptible. Commuting is a means by which coupling remains sufficiently tight between reservoirs of infection in populations of greater

size than the endemic threshold population, such as Manchester CB, and neighbouring districts, thus limiting growth in threshold populations in highly connected districts and Lancashire as a whole in the vaccine era (as is observed in Section 5.2), but also the greater strength of gravity parameters in later time–windows.

The gravity HHH models for scarlet fever incidence in South Wales reveal a notable decline in the strength of gravity parameters in later time–windows, even when the formulation is extended to include distance decay. This fall coincides with a period of dramatic fall in scarlet fever incidence across Wales, with the pace of decline accelerating in from the 1950s onwards. This is demonstrated by the exploratory analysis of scarlet fever time series for South Wales and a small subset of districts (Cardiff CB, Merthyr Tydfil CB and Fishguard & Goodwick UD) of varying population size in Section 4.3.3. The substantial fall in the strength of the gravity parameter in later time–windows also mirror the unrelenting decline in scarlet fever endemicity, and the significant growth in the size of the regional endemic threshold population for scarlet fever in South Wales (see Sections 5.3 and 5.5.3). One can conclude that the fall in the strength of the gravity parameter estimates reflects the diminishing caseload of scarlet fever in Lancashire and South Wales from the late 1950s onwards. Limited incidence, alongside improvements in hygiene, nutrition and reduced crowding, factors previously cited for explain scarlet fever's declining morbidity in the postwar period (Lamagni et al., 2018), may have resulted in constricting spread of the disease and isolate the infection to within-district, endemic activity. To some extent, this is supported by analysis of the residual heterogeneity detected in the HHH model formulations for scarlet fever in South Wales, which reveals the endemic HHH model results in the least levels of overdispersion and thus variation round the mean. The inclusion of epidemic and spatiotemporal components only serves to increase the presence of overdispersion. This is in stark contrast to HHH formulations of the other two childhood infections measles and pertussis, for which increasing model complexity and inclusion of covariates reduced levels of dispersion significantly compared to the endemic model.

Despite the decline in later windows, the strength of the gravity measure is notably higher in South Wales compared to Lancashire, where gravity parameter estimates are weakly positive across multiple time–windows, declining from time window one to seven before staging some recover in the final time–windows. Similarities between gravity parameters in Lancashire and South Wales are evident once a power-law is incorporated in the gravity HHH model formulations, substantially increasing parameter strength while also reducing the strength of the distance decay considerably. With both short- and long- range disease transmission accounted for (Meyer et al., 2017), the importance of transportation for maintaining persistence of infection via spatial coupling between lightly populated urban districts in the metropolitan hinterland and endemic centres of infection is emphasised. This relationship corresponds crudely to the development of a 'core and satellite' metapopulation (Grenfell and Harwood, 1997), which poses a barrier for disease control, and limits the growth of endemic threshold population size overtime, due to the interplay between intermediate levels of vaccination coverage, frequency and density of contacts between susceptibles and infected individuals, and geographical rescue effects (Bolker and Grenfell, 2016; Keeling and Rohani, 2008; Meyer et al., 2014).

The extension of the HHH model formulation to include random effects alongside fixed effects, a power-law decay of spatial interaction and gravity model measure enables unobserved covariates which influence disease spread to be considered. From a geographical perspective, edge effects on borders of the regional metapopulations are of particular interest. In the context of Lancashire, with the major conurbations of Liverpool CB and Manchester CB situated very close to the regional border, there are undoubtedly missing potential sources on infection from the from the unobserved side of the border, since the metropolitan hinterland of both urban centres stretches into the neighbouring county of Cheshire, with settlements such as Chester and Birkenhead in the Wirral Peninsula, tightly associated economically with Liverpool CB (Pollard et al., 2006), and the town of Stockport, as well as communities in North Derbyshire and South Yorkshire have

historically strong economic ties with Manchester CB (Carter, 1962). One must also take into the account edge effects as a result of commuting between large towns and the main urban centre of the Greater Manchester area with large urban centres in neighbouring Yorkshire, particularly the cities of Leeds, Sheffield and Bradford as well as towns such as Halifax and Huddersfield (Baker and Hebbert, 1995).

By incorporating district-specific random effects in the endemic component, unobserved heterogeneity due to immigration into the region and edge effects was accounted for in the power-law HHH and power-law + gravity HHH model formulations. Additionally, the inclusion of random effects allows one to account for unobserved heterogeneity as a consequence of underreporting of disease incidence, something which affects the surveillance data of all three childhood diseases under analysis in this thesis. The results reveal that the inclusion of random effects increase the proportion of cases attributed to epidemic activity, with rises in maximum eigenvalues and increases of disease decay parameter values/ gravity parameter values. Analysis of predictive performance using proper soring rules also reveal that the inclusion of random effects significantly improves predictive performance compared to previous sub-model formulations, with the power-law + gravity RI HHH model providing the best predictive performance for all three diseases in the Lancashire region across the nine time–windows. A better way of accounting for edge effects would thus be to explicitly incorporate immigration between regions and districts, commuting data such as journey-to-work distances from census statistics (Warnes, 1972) or data on labour flows. For instance, Geilhufe et al. (2014) utilise incoming road or air traffic from outside North Norway as a suitable proxy for the risk of importing cases of influenza. This revealed improved predictive performance when coupled with a gravity model of spatial interaction to account for population in the spatio-temporal component.

7.3.1 Convergence Issues

Due to convergence issues in the modelling process, results are not provided for HHH model formulations that were extended to include random effects for measles or pertussis in time-window in the South Wales region. Convergence issues prevented the modelling of an uninterrupted period of disease activity across all nine time-windows in the study period, handicapping attempts at meaningful inference and interpretation of results.

A major issue encountered in the application of HHH modelling relates to convergence issues, primarily experienced when including random effects to the power-law HHH models and power-law + gravity HHH models. Difficulties with convergence are not unexpected when fitting and running complex models, but convergence issues prevent inferences from being made, thus representing a barrier for productive analysis. The primary issue is a failure of the likelihood maximisation algorithm to converge. Lack of convergence can be an indication that the data does not fit the model well, due to the presence of too many poorly fitting observations (Allison, 2004). More often, this failure of convergence is as a result of data patterns known as complete or quasi-complete separation (Allison, 2009). For these patterns, the maximum likelihood estimates simply cannot be calculated. This is of course highly problematic when in most applied circumstances, every parameter dimension must be converged for the model to converge (Gill, 2008).

Complete separation occurs when a linear combination of the predictors yields a precise prediction of the response variable (Albert and Anderson, 1984). Quasi-complete separation is similar, only predictors yield a perfect prediction of the response variable for most values of the predictors, but not all (Allison, 2004). Separation can occur when the dataset is too small to observe events with low probabilities (Allison, 2009). This may explain why modelling of infectious disease counts in South Wales with the inclusion of random effects resulted in far more numerous convergence failures in contrast to HHH models with random

effects applied to the time-window datasets for Lancashire. Other factors which may result in complete or quasi-compete separation include the incorporation of several fixed-effect or random effect parameters in the model which can increase the likelihood of separation due to individual groups in the data having smaller sample sizes (Allison, 2004). Greater complexity of the model fit, with more predictors included in the model, only makes it more difficult to identify the potential cause of complete of quasi-complete separation. Maximum likelihood estimates may fail to converge and result in significant standard errors, without prior warning. Complete separation can be addressed by regularisation; using penalised regression on the fixed effects has been cited as one such approach (Allison, 2009). However, the HHH model formulation already maximises penalised and marginal log-likelihoods alternately in order to achieve convergence.

The R package `surveillance` does not feature a diagnostic tool which can help aid the identification of potential causes of convergence errors. An oft-cited method for addressing convergence issues is to increase the maximum number of iterations in the optimisation of regression and variance parameters. The default number of inner iterations for the HHH model is 20 and this can be increased to 50 (Höhle, 2007). However, this did not provide any solution to convergence errors encountered in the modelling of measles and pertussis incidence with random effects in time-windows for the South Wales region. Due to the inherent errors within the modelling procedure, it was decided that, since the absence of random effects allows maximum likelihood estimates to converge and parameters to be estimated, the HHH models of measles, pertussis and scarlet fever spread in South Wales were limited to fixed-effects formulations.

7.4 Chapter Summary

In the analysis presented in this chapter, several sub-model formulations of the HHH model were fitted and utilised to analyse the endemic-epidemic dynamics of pertussis, measles

and scarlet fever, in a quest to build a more detailed understanding of the spatiotemporal dynamics of measles, pertussis and scarlet fever incidence and transmission within the regional metapopulations of South Wales and Lancashire. Such understanding can help inform interpretation of the rise and fall of endemic threshold populations over time and space, elucidating on the influence of spatial structure, mobility and demographic stochasticity on disease persistence.

The HHH modelling framework reveals commuter-driven spread as a key mechanism for disease persistence in both regional metapopulations. The extension of the baseline HHH models to include gravity and power-law components to capture the short and long-range transmission of infection significantly improved the model fit to the observed data, for all three childhood diseases 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 pertussis, measles and scarlet fever incidence all concentrated within 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. Regarding Lancashire, districts immediately surrounding the Liverpool and Manchester conurbations were found to be of disproportionate importance in the spatial transmission of pertussis, suggestive of an 'agglomeration effect', as indicated in the hotspot and survival analyses performed in Chapter 6.

In the context of analysing and understanding spatiotemporal changes in the size of endemic threshold populations, these findings indicate that geographical mobility and connectivity play not only a pivotal role in both the spread of disease but also limiting the growth of endemic threshold populations, by ensuring subpopulations remain tightly coupled with regional endemic centres. High levels of connectivity, nurtured by a historically close spatial structure of settlements and subpopulations constructed from a myriad of

economically co-dependent relationships, only serve to foster these epidemiological agglomeration effects, ensuring infections are regularly reintroduced into susceptible populations with high levels of mixing. The net effect of this is significantly limiting the growth of endemic threshold populations even in the presence of effective disease interventions, such as vaccination. This is most evident from the very limited increase in the size of regional endemic threshold estimates for pertussis in Lancashire during in vaccine era time windows (see Section 5.2).

The following and final chapter of thesis shall now provide a broader discussion of the results presented in this and preceding chapters, summarising key research findings, considering the limitations of the research undertaken, identifying areas to improve and avenues for future research before making some final remarks.

Chapter 8: Conclusion

8 Summary of Research Findings

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 study reveals significant regional differences across endemic threshold populations for multiple childhood infections, despite sharing metapopulation dynamics. These findings are notwithstanding the successful uptake of the pertussis vaccine following its introduction in 1957. The empirical work documented in study has provided a comprehensive account of spatiotemporal changes in endemic threshold populations for three childhood infections 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 childhood disease through time and space.

Endemic threshold estimates for the pre-vaccine era reveal stark regional differences in endemic threshold populations for pertussis, demonstrating the influence of geographical variability in population density and spatial connectivity on shaping the size of endemic threshold populations (see Sections 5.1, 5.2 & 5.3). 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. Incidence rates are often low (see Section 4.5), the risk of fadeout, even in the absence of public health interventions such as mass vaccination, is high (see Section 6.3.2), and many

districts located in rural areas in South Wales are weakly correlated with wider regional disease activity (see Section 4.7), resulting in a lack of spatial synchrony. Immediately following World War II, the post-war baby boom resulted in substantial growth in the number and density of susceptibles within districts of all sizes across South Wales, leading to a dramatic fall in the value of endemic threshold estimates. 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, communities often intimately connected economically to the urban centres, 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 twentieth century (Grenfell et al., 2001).

As has been extensively detailed in Chapter Seven, a number of HHH model formulations were implemented across the nine time–windows to analyse the nature of measles, pertussis and scarlet fever spread through time in the regional metapopulations of Lancashire and South Wales, generating insights which help to illuminate the influence of mobility, connectivity and population density in shaped local and regional disease persistence. The HHH models generally reveal commuter-driven spread as a key driver for measles and pertussis persistence in both regional populations, with the most complex HHH model formulation including a power-law decay of spatial interaction, gravity model component and random effects, to account for unobserved heterogeneities such as edge effects and underreporting. found to provide the best performing model fit to observed measles, pertussis and scarlet fever data across successive time windows. In Lancashire

and South Wales, disease persistence in communities with smaller populations was driven primarily by commuter spread to and from endemic reservoirs, with incidence concentrated in districts with high levels of population mobility and significant degree of spatial coupling with largely populated districts. In South Wales, there was significant commuter-driven 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.

A key internal factor for limiting disease persistence in South Wales is the topography of the landscape, playing profound role in shaping the geographical characteristics of the regional population structure. The steep ridges and rugged landscape which hang over the long pre-industrial valleys of Carmarthenshire, Glamorgan and Monmouthshire significantly limit the direction in which disease spread travels from the densely populated urban centres that lie on the coast; Cardiff, Swansea and, to a lesser extent, Newport. Communities in urban and rural districts located in the Valleys tend to be linear in their distribution, hugging the communication links, road and rail lines, found on the valley floor. Reinforced by the intense socio-economic interactions which exist between the major towns and cities on the coast and the once thriving industrial communities in the valleys, internal mobility is constricted and disease spread outside of Glamorgan, in which the majority of Valley communities are located alongside Cardiff and Swansea, is much more restricted. The reduced rate of disease transmission from Glamorgan to neighbouring counties in the South Wales region

aids in sustaining lower levels of disease persistence, which are then exacerbated by the high levels of population dispersion and extremely limited rates of susceptible recruitment in the urban and rural districts of Carmarthenshire, Pembrokeshire and Monmouthshire, the latter of which are predominant in the geographical make-up of the three aforementioned counties.

In Lancashire, there is a consistent pattern of lower endemic threshold populations across the study period, reflecting the greater levels of connectivity, accessibility and density across subpopulations. The regional population of Lancashire has a greater overall population size than South Wales, with a number of large towns and urban centres distributed across the region, despite the two largest settlements, the conurbations of Liverpool and Manchester, being located on the eastern and western edges of the southern portion of the region. These two cities function as lungs of endemicity, regularly breathing life into infectious waves of pertussis, measles and scarlet fever, which spread across to neighbouring districts but also travel down the population hierarchy in large towns further afield which has strong economic ties with Liverpool and Manchester. Both these cities provide a constant source of epidemic activity and a ready pool of susceptibles, recharged by high birth rates.

In summary, the higher levels of mobility and accessibility among Lancashire's districts with greater density and availability of susceptibles in the subpopulations ensure that the endemic threshold population is consistently lower across the nine time windows in Lancashire compared to South Wale. The profound differences in the spatial structure of the regional populations which shape this trend also explain why high density and high connectivity districts in Lancashire have a much lower endemic threshold population size than the high density and highly connected districts in South Wales. In the former, population centres are more evenly distributed across the region yet tightly bound to endemic centres, despite significant distance, while in the latter, highly connected districts

with dense populations go hand-in-hand, located in the Vale of Glamorgan and the valleys immediately adjacent.

8.1 Limitations & Areas for Future Work

An important caveat when working with infectious disease data, and in particular when calculating thresholds using such data is underreporting, which introduces uncertainties. 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 subclinical cases of the disease being a common occurrence. Widespread under-reporting may result in an over-estimate of weeks experiencing fade-out.

Childhood diseases often share symptoms with other viral and bacterial infections that affect young children, and subclinical cases are also common. Although there have been studies which have assessed the quality of historical measles and pertussis disease data in England and Wales (Clarkson and Fine, 1985; Gunning et al., 2014), there has been an absence of work assessing the quality of scarlet fever incidence data, despite being a statutorily notifiable disease consistently since 1899 (Lamagni et al., 2018). According to one study of contemporary scarlet fever surveillance in the two countries, which compares scarlet fever notification data with the Public Health England syndromic surveillance system, a sentinel network of primary care general practitioners, it was estimated that scarlet fever cases were underreported by approximately 50% in the mid-2010s (Lamagni et al., 2018). It is therefore possible that a similar rate of underreporting existed in Lancashire and South Wales during the mid-twentieth century, particularly if the issue of subclinical cases was compounded by misdiagnoses and the failure of parents to seek GPs for whatever reason when their children fell ill to scarlet fever. It is therefore possible that the magnitude of cases was greater than reported, and the disease was more persistent than notification data in the *Weekly Returns* suggests.

Another issue with scarlet fever in terms of endemic threshold estimation and analysing the persistence of the infection is the infectious period of the disease. Without antibiotics, the infectious period of the disease can last up to three weeks; with antibiotic treatment, the infectious period shortens dramatically to around one week (Lamagni et al., 2018). The increasing use of antibiotics to treat illnesses like scarlet fever during the postwar period, may explain the reduced persistence and incidence of the infection as the study period progresses. One would expect that if a significant volume of scarlet fever cases had been regularly left untreated, the longer infectious period would result in lower endemic threshold population estimates and greater levels of endemicity, providing greater opportunity to form new chains of transmission that maintain persistence of the disease (Metcalf et al., 2013).

Moreover, 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 research (Wearing and Rohani, 2009).

The HHH models utilised in this study do not represent an exhaustive list of formulations that could have implemented and there is considerable scope to expand this work in future. Historical-based infectious disease surveillance data possess characteristics which can challenge classical statistical approaches (Lloyd-Smith et al., 2015); these include non-independent observations and the underlying epidemic process being only partially observable when analysing multivariate time series of disease counts if the process is continuous in space. Another problematic feature of surveillance data is the absence of subclinical cases of infection which fuel underreporting, alongside the misclassification of cases. However, the versatility and utility of the multivariate regression modelling framework for endemic-epidemic disease dynamics enables a range of possibilities, too numerous for

one chapter to explore in great detail, to account and control for a wide range of exogenous and unobserved covariates which may aid understanding for disease spread in local and regional populations.

A particularly fruitful avenue to purse with regards to further enriching the HHH model formulation would be to incorporate the impact of disease interventions, namely the implementation of mass vaccination, in the case of pertussis during the study period. Including some measure of vaccination coverage would importantly reflect changing nature of the (remaining) susceptible population, in response to effective disease intervention. Past research has observed that vaccination coverage is be associated with outbreak size, whilst also capturing evidence of considerable heterogeneity in vaccination coverage (Herzog et al., 2011). The work of Herzog et al. (2011) provides the template for the inclusion of vaccination coverage as an explanatory variable, demonstrating the covariate to be strongly associated with the size and occurrence of measles epidemics in Germany at state level. They propose an extension to the two component model for disease counts outlined by Held et al. (2005, 2006). Herzog et al. (2011) found that the inclusion of the susceptible proportion of the population in the autoregressive component of the model to be the most effective approach , according to the mass action principle (Keeling and Rohani, 2008). The mass action principle assumes that the rate of disease propagation is proportional to the density of susceptibles multiplied by the density of infected individuals.

Although patchy in quality, owing to the idiosyncratic and qualitative nature of reporting, annual data on the number of infants and children immunised for pertussis, either by the whole-cell (*wP*) or triple antigen vaccine, can be obtained from the Medical Officer of Health (MOH) reports for several local government districts in Lancashire and South Wales from 1957 to 1969. The Wellcome Library holds printed copies of MOH reports in its archives and beginning in August 2016, digitised and uploaded all existing MOH reports it holds for LGDs in England and Wales from the late nineteenth century onwards, in pdf format. As of

5th February 2020, 68,410 documents have been uploaded, including all annual MOH reports for districts in the Lancashire and South Wales regions for the period 1957–1969. These reports can be accessed from the following source: https://archive.org/details/medicalofficerofhealthreports.

MOH reports can provide information on the total number of children under five years of age immunised by GPs, mobile units and infant welfare and special clinics. MOH reports were produced annually from the late-nineteenth century to 1972 by the Medical Officer of Health for the reporting district, who would describe the work carried out by district public health officers during the year, provide detailed statistical information on birth rates, death rates, infant mortality and incidence of notifiable infectious diseases, as well as publish a general statement on the overall health of the district population.

As alluded to, the format of MOH reports vary considerably over time, by district and by author. As a consequence, immunisation data is not always present in the reports for notifiable diseases; data might be present in a district MOH report one year yet absent for the same district in the next. Overcoming the hurdles posed by the extremely patchy nature of such historical epidemiological data might prove a time-consuming and complex task, but one that may hold tremendous insight into illustrating the impact of vaccination on the spatial dynamics and persistence of pertussis in local subpopulations and region-wide, and how this contributes to growth in endemic threshold populations, which hold great relevance in informing vaccination strategies.

With regards to the incorporation of gravity models, in addition to population, the gravity extension of the HHH model formulation could utilise a second complementary ‘gravity measure’ to model spatiotemporal spread of infection. Past research has proposed an ‘urbanicity’ measure, founded on the size of the largest place within a geographical district, for including urban effects on epidemiological variables (Kafadar and Tukey, 1993). This

would address a key shortcoming of using population-based measures such as population fractions or population density, which are not necessarily based upon large towns or cities and thus are found to be sometimes inconsistent with what one might consider a logical urban hierarchy. For instance, Narberth UD, a small community in South Pembrokeshire with a mean population of just 1,068 in time–window one (1940-45) has a mean population density of 2,125.1 per square kilometre during the same period. This exceeds the population density of Swansea CB, which has a mean population density of 1,863.21 per square kilometre over the corresponding period yet a mean population size which is approximately 133 times greater than that of Narberth UD, standing at 142,273. In this example, population density fails to distinguish between the second largest town and second smallest district by population size in South Wales in time–window one.

Goodall et al. (1998) propose three variations on urbanicity: the population of the largest subunit in each geographical unit, the square root of the sum of the squared population of the top three largest subunits in each geographical unit, and the square root of the sum of the squares of all subunit populations. In the analyses detailed and discussed in this chapter, the geographical units are the 125 local government districts (LGDs) of Lancashire and 75 of the four counties of South Wales respectively, with the most likely candidates for subunits being urban settlements, villages and hamlets. However, since many LGDs contain no smaller subunits, specifically the county boroughs and many of the municipal boroughs found in each region, where the district is essentially the subunit, population of the largest subunit would be the most logical choice as measure of urbanicity. This would allow the one to consider more explicitly the epidemiological effects of shifting urban patterns in regional populations over time and space, on disease spread and persistence. One could analyse, with more confidence perhaps, to what extent significant fluctuations in population densities and local population size as a result of population decentralisation, slum clearance and deindustrialisation experienced in Lancashire and South Wales affected the nature of disease endemicity in the 1950s and 1960s. In addition to gravity

measures including a population effect in the epidemic component as well as the spatiotemporal component may be beneficial to reflect higher contact rates and thus infectivity in districts with more dense populations, which results in the depletion of susceptible pools and more regular occurrence of fadeout events in large and medium-sized towns.

There are patterns in the spatiotemporal spread of infection which can easily go unobserved when utilising geographical distance to explore the impact of human mobility and accessibility on infectious disease persistence. In their work describing describe the geographical spread of 2009 H1N1 influenza pandemic, Brockmann and Helbing (2013) introduce the concept of 'effective distance' to describe the spatiotemporal patterns not easily seen when using geographical distances, collapsing geographical space by visualising and computing only effective-distances based upon mobility or population flow data (Brockmann and Helbing, 2013). This approach relates to what Brown and Horton (1970) terms 'functional distance'; the function of (inter-) regional properties such as commuter or travel flows such that it "reflects the net effect of entity properties upon the propensity of the entities to interact" (Brown and Holmes, 1971: 388).

Another fertile area of future research is the statistical analysis of age-stratified surveillance data. For directly transmitted human diseases, the social tendency for individuals to congregate and gravitate around each other creates contact patterns between subgroups of a population, potentially acting to broaden the pure distance decay of interaction. Since contact patterns vary across age (Mossong et al., 2008; Truscott et al., 2012), there is a need for spatiotemporal models of disease spread which attempt to unify across age groups and geographical districts. The Censuses for England and Wales in 1951, 1961 and 1971 provide a breakdown of local government district populations stratified by age, providing statistical data on the number of children aged between 0-4 years of age, 5-9 years of age and 10-14 years of age. Coupled with spatially aggregated birth rate data reported in the

annual publications of the Registrar–General's *Statistical Review,* it may be possible to construct age-stratified spatiotemporal datasets of areal-level measles, pertussis and scarlet fever counts which could be used to incorporate the inherent contact structure in the regression-oriented, endemic–epidemic HHH model. This three-dimensional approach would offer a more detailed description of disease spread than unstratified or non-spatial models, which characteristically assume homogeneous mixing within each district subpopulation (Meyer and Held, 2017). This provides a much deeper understanding of the nature of disease persistence in susceptible populations on a local scale and within the wider regional metapopulation. Although the historical nature of the epidemiological data analysed here may preclude attempts to enact the following approach, replacing the parametric formulation of distance decay with a social contact matrix, stratified by spatial distance in addition to age group could prove very fruitful for understanding the short-term spread and long term-persistence of childhood infections. Separate movement data for school children and adults could be utilised to quantify the strength of epidemiological coupling between districts (Kucharski et al., 2015).

Although weekday versus weekend differences in contact patterns are not relevant for models which used weekly counts of infectious disease data, there are possibly relevant seasonal effects on larger time scales which need to be considered. Soper (1929) first demonstrated temporal heterogeneity in the transmission rate of disease, in his study of high amplitude outbreaks of measles in Glasgow. Elevated transmission rates were observed in October and fell throughout the academic year, resulting a trough in the summer months, implying higher transmission rates as a result of term-time forcing. Undoubtedly, the contact structure of school children changes considerably between regular and school holiday periods (Hens et al., 2009; Rohani and King, 2010), and the conclusions of Soper (1929) have been supported by a large number of studies analysing the transmission and incidence of childhood infections such as measles, chickenpox, pertussis, and mumps, in England and Wales and the United States (London and Yorke, 1973; Finkenstädt and

Grenfell, 2000). One approach for accounting seasonality in the coefficient of transmission may be to adopt a binary function, with two different values of the coefficient of transmission; one for the school terms and one for the holidays. However, this may necessitate a knowledge of the academic calendar in local populations, which is not always possible, particularly for historical data. A more straightforward approach, in the context of HHH modelling, would be to include a time-varying contact matrix by estimating seasonality in the epidemic component (Held and Paul, 2012), which is supported by the model framework. Relatedly, following an alternative approach adopted by Fanshawe et al. (2008), seasonality parameters could be allowed to change annually according to a random walk model. Implementation would then necessitate the need to use Markov chain Monte Carlo or other more demanding techniques for inference.

Seeking a solution to the multiple convergence issues faced when running the HHH models including random effects discussed in Section 7.3 opens the door to an alternative modelling framework; using Bayesian methods. Complete separation can be addressed by Bayesian regression, with appropriate priors with a prior distribution on the fixed effects (Hsu and Leonard, 1997; Kahn and Raftery, 1996). Although it has often be claimed that the main advantage of utilising a Bayesian modelling framework over frequentists method is the ability to incorporate of prior knowledge by specifying appropriate prior probabilities (Greenland, 2007), Bayesian methods are particularly useful for statistical inference of complex models which present significant difficulties for frequentist methods. Calculating the maximum of very complex likelihood functions can be a difficult task in practice, despite the advances in computer software and hardware in recent years. In such situations, the frequentist approach usually involves numerical tools, such as Newton-Raphson or Quasi-Newton algorithms in the HHH model. However, convergence problems may occur, and solutions may be highly dependent on initial values. Bayesian methods can overcome these issues by utilising MCMC methods (Smith and Roberts, 1993; Brooks, 1998).Adopting this approach, inference is centred on the simulated samples, utilising parameters of interest.

The missing data alongside parameters are treated as random variables, and a MCMC algorithm is employed to use the missing data and parameter values and provide robust inferences and quantify uncertainty in parameter estimates. An advantage of using a Bayesian approach for the analysis of epidemiological data is that it enables the inclusion of unobserved variables such as infectious and incubation period (Cauchemez,et al., 2004; Lekone and Finkenstädt, 2006), in stochastic modelling approaches, which are quantified along with all other model parameters to enrich insights on the mechanisms of disease transmission.

8.2 Final Remarks

To conclude, the South Wales region exhibits consistently higher endemic threshold populations than the Lancashire region over the course the study period charting spatiotemporal changes in endemic and epidemic activity of measles, pertussis and scarlet fever across the mid-twentieth century, from January 1940 to December 1969. A range of internal and external factors explain the regional differences in disease persistence and the respective thresholds of childhood infections. South Wales represents a more sparsely inhabited regional population, in which communities are more widely dispersed and population densities are notably lower across the region's urban and rural districts compared to the Lancashire region. In a more dispersed metapopulation, coupled with lower rates of population growth, rates of susceptible recruitment are more limited due to the greater degree of spacing between susceptibles and infected individuals. Chains of disease transmission are also more prone to fracture and collapse, reducing the duration of epidemic outbreaks and limiting the potential for disease spread. Population mobility exercised through short-range travel patterns that typify commuting play an important role in maintaining high levels of spatial coupling between the metropolitan hinterland and urban centres which function as endemic reservoirs of infection. Such dynamics maintain persistence of disease, which are only reinforced by dense populations and high levels of

susceptible recruitment in large urban districts where vaccination coverage is at intermediate levels. Breaking the strong, intimate ties of spatial coupling between urban centres and their satellite settlements through aggressive diseases intervention is vital to prevent constraints on growth of endemic threshold populations in attempts to eliminate childhood infections such as pertussis and measles.

The methodological approaches adopted and presented in this research provide 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. This has even greater currency today, with vaccine supply issues across the world inhibiting progress to bring the COVID-19 pandemic under control, stressing the practical importance of spatially targeted, geographical informed vaccination 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 metapopulation with core-satellite dynamics. In such a context, district-based or regionally-targeted mass vaccination programmes might represent more effective strategies for disease elimination than nationwide mass vaccination.

Bibliography

Albert, A. and Anderson, J. A. (1984). On the Existence of Maximum Likelihood Estimates in Logistic Regression Models. *Biometrika*, 71, 1-10.

Alkema, L., Raftery, A. E., Gerland, P., Clark, S. J., Pelletier, F., Buettner, T. and Heilig, G. K. (2011). Probabilistic Projections of The Total Fertility Rate for All Countries. *Demography*, 48, 815-839.

Allen, J. C., Schaffer, W. M. and Rosko, D. (1993). Chaos Reduces Species Extinction by Amplifying Local Population Noise. *Nature,* 364, 229-232.

Allison, P. D. (2004). Convergence Problems in Logistic Regression, pp. 238-252 in Altman, M., Gill, J. and McDonald, M. P. (Eds.). *Numerical Issues in Statistical Computing for the Social Scientist*. John Wiley & Sons: Hoboken, NJ.

Allison, P. D. (2009). *Fixed Effects Regression Models*. Sage: London.

Altman, D. G. and Bland, J. M. (1998). Time to Event (Survival) Data. *BMJ*, 317, 468-469.

Amirthalingam, G., Gupta, S. and Campbell, H. (2013). Pertussis Immunisation and Control in England And Wales, 1957 To 2012: A Historical Review. *Eurosurveillance*, 18, 20587.

Anderson, H. and Britton, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis*. Lecture Notes in Statistics, Volume 151. Springer: New York, NY.

Anderson, R. M. (1990). *The Transmission Dynamics of Sexually Transmitted Diseases: The Behavioural Component*. International Union for The Scientific Study of Population: Liege, Belgium.

Anderson, R. M. (2016). The Impact of Vaccination on Epidemiology of Infectious Diseases, pp. 3-31 in Bloom, B. R. and Lambert, P. H. (eds.), *The Vaccine Book.* 2nd ed. Academic Press: London.

Anderson, R. M. and May, R. M. (1979). Population Biology of Infectious Diseases: Part I. *Nature*, 280, 361-367.

Anderson, R. M. and May, R. M. (1982). Directly Transmitted Infectious Diseases: Control by Vaccination. *Science*, 215, 1053-1060.

Anderson, R. M. and May, R. M. (1983). Vaccination Against Rubella and Measles: Quantitative Investigations of Different Policies. *Journal of Hygiene*, 90, 259-325.

Anderson, R. M. and May, R. M. (1985). Age-Related Changes in the Rate of Disease Transmission: Implications for The Design of Vaccination Programmes. *Journal of Hygiene*, 94, 365-436.

Anderson, R. M. and May, R. M. (1990). Immunisation and Herd Immunity. *The Lancet*, 33, 641-645.

Anderson, R. M. and May, R. M. (1992). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press: Oxford.

Anderson, R. M., Grenfell, B. T. and May, R. M. (1984). Oscillatory Fluctuations in the Incidence of Infectious Disease and the Impact of Vaccination: Time Series Analysis. *Journal of Hygiene*, 93, 587-608.

Aron, J. L. and Schwartz, I. B. (1984). Seasonality and Period-Doubling Bifurcations in An Epidemic Model. *Journal of Theoretical Biology*, 110, 665-679.

Bailey, N. T. (1975). *The Mathematical Theory of Infectious Diseases and Its Applications*. Charles Griffin: London.

Baker, M. and Hebbert, M. (1995). *The Regional Economy of the Trans-Pennine Corridor*. Department of Planning and Landscape, University of Manchester: Manchester.

Bansal, S., Grenfell, B. T., and Meyers, L. A. (2007). When Individual Behaviour Matters: Homogeneous and Network Models in Epidemiology. *Journal of The Royal Society Interface*, 4, 879-891.

Barrie, D. S. M. (1982). *The Taff Vale Railway*. 2nd ed. Oakwood Press: Tisbury.

Barrie, D. S. M. (1994). *A Regional History of the Railways of Great Britain. Volume 12: South Wales.* 2nd ed. David & Charles: Exeter.

Barrios, J., Verstraeten, W., Maes, P., Aerts, J. M., Farifteh, J. and Coppin, P. (2012). Using the Gravity Model to Estimate the Spatial Spread of Vector-Borne Diseases. *International Journal of Environmental Research and Public Health*, 9, 4346-4364.

Barron, D. N. (1992). The Analysis of Count Data: Overdispersion and Autocorrelation. *Sociological Methodology*, 22, 179-220.

Barthélemy, M. (2011). Spatial Networks. *Physics Reports*, 499, 1-101.

Bartlett, M. S. (1955). *An Introduction to Stochastic Processes: With Special Reference to Methods and Applications*. Cambridge University Press: Cambridge.

Bartlett, M. S. (1956). Deterministic and Stochastic Models for Recurrent Epidemics. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 4, 81-109.

Bartlett, M. S. (1957). Measles Periodicity and Community Size. *Journal of The Royal Statistical Society. Series A (General)*, 120, 48-70.

Bartlett, M. S. (1960). The Critical Community Size for Measles in the United States. *Journal of The Royal Statistical Society. Series A (General)*, 123, 37-44.

Batty, M. (1970). Models and Projections of The Space Economy: A Sub-Regional Study in North West England. *The Town Planning Review*, 41, 121-147.

Bax, A. and S. Fairfield. (1978). Lancashire, pp. 364-384 in Bax, A. and S. Fairfield. (Eds.). *The Macmillan Guide to The United Kingdom 1978–79*. Palgrave Macmillan: London.

Becker, N. G. and Britton, T. (1999). Statistical Studies of Infectious Disease Incidence. *Journal of The Royal Statistical Society: Series B (Statistical Methodology)*, 61, 287-307.

Becker, N.G. (1989). *Analysis of Infectious Disease Data.* Chapman & Hall: London.

Beckford, A. P., Kaschula, R. O. C. and Stephen, C. 1985). Factors Associated with Fatal Cases of Measles. *South African Medical Journal,* 68, 858-863.

Bell, J. A. (1948). Diphtheria Immunization: Use of An Alum-Precipitated Mixture of Pertussis Vaccine and Diphtheria Toxoid. *Journal of The American Medical Association*, 137, 1009-1016.

Beyer, H. L., Hampson, K., Lembo, T., Cleaveland, S., Kaare, M. and Haydon, D. T. (2011). Metapopulation Dynamics of Rabies and The Efficacy of Vaccination. *Proceedings of The Royal Society B: Biological Sciences*, 278, 2182-2190.

Beyer, H. L., Hampson, K., Lembo, T., Cleaveland, S., Kaare, M. and Haydon, D. T. (2012). The Implications of Metapopulation Dynamics on The Design of Vaccination Campaigns. *Vaccine*, 30, 1014-1022.

Bharti, N., Djibo, A., Ferrari, M. J., Grais, R. F., Tatem, A. J., McCabe, C. A., Bjornstad, O. N. and Grenfell, B. T. (2010). Measles Hotspots and Epidemiological Connectivity. *Epidemiology & Infection*, 138, 1308-1316.

Bisno, A.L. (1995). Streptococcus Pyogenes, pp. 1786-1799 in Mandell, G. L, Bennett, R. G., Dolin, R., (Eds.). *Principles and Practice of Infectious Diseases*. Vol 2. Churchill Livingstone: New York, NY.

Bivand, R. S. and Piras, G. (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, 63, 1-36.

Bivand, R.S., Ledesma, E. and Gomez-Rubio, V. (2013). *Applied Spatial Data Analysis with R*. 2nd ed. Springer-Verlag: Berlin.

Bjørnstad, O. N. and Grenfell, B. T. (2008). Hazards, Spatial Transmission and Timing of Outbreaks in Epidemic Metapopulations. *Environmental and Ecological Statistics*, 15, 265-277.

Bjørnstad, O. N., Finkenstädt, B. F. and Grenfell, B. T. (2002). Dynamics of Measles Epidemics: Estimating Scaling of Transmission Rates Using a Time Series SIR Model. *Ecological Monographs*, 72, 169-184.

Black, F. L. (1966). Measles Endemicity In Insular Populations: Critical Community Size and Its Evolutionary Implication. *Journal of Theoretical Biology*, 11, 207-211.

Black, F. L. (2013). Measles, pp. 451-465 in Evans, A. S. (Ed.). *Viral Infections of Humans: Epidemiology and Control*. 3rd ed. Plenum Medical Book Company: London.

Black, F. L. and Gudnadottir, M. (1963). Measles Vaccination of Adults in Iceland. *The Lancet*, 281, 418-420.

Blackwood, J. C., Cummings, D. A. T., Broutin, H., Iamsirithaworn, S. and Rohani, P. (2012). The Population Ecology of Infectious Diseases: Pertussis in Thailand As A Case Study. *Parasitology*, 139, 1888-1898.

Bland, J. M. and Altman, D. G. (1998). Survival Probabilities (The Kaplan-Meier Method). *BMJ*, 317, 1572-1580.

Bland, J. M. and Altman, D. G. (2004). The Log Rank Test. *BMJ*, 328, 1073.

Blasius, B., Huppert, A. and Stone, L. (1999). Complex Dynamics and Phase Synchronization in Spatially Extended Ecological Systems. *Nature*, 399, 354-359.

Bolker, B. M. and Grenfell, B. T. (1993). Chaos and Biological Complexity in Measles Dynamics. *Proceedings of The Royal Society of London B: Biological Sciences*, 251, 75-81.

Bolker, B. M. and Grenfell, B. T. (1995). Space, Persistence and Dynamics of Measles Epidemics. *Philosophical Transactions of The Royal Society of London B: Biological Sciences*, 348, 309-320.

Bolker, B. M. and Grenfell, B. T. (1996). Impact of Vaccination on the Spatial Correlation and Persistence of Measles Dynamics. *Proceedings of the National Academy of Sciences*, 93, 12648-12653.

Bonanni, P. (1999). Demographic Impact of Vaccination: A Review. *Vaccine*, 17, S120-S125.

Bonds, M. H. and Rohani, P. (2009). Herd Immunity Acquired Indirectly from Interactions Between the Ecology of Infectious Diseases, Demography and Economics*. Journal of the Royal Society Interface*, 7, 541-547.

Bordet, J. (1906). Le Microbe De La Coqueluche. *Ann Inst Pasteur*, 20, 48-68.

Bowen, E. G. (1960). *Wales: A Study in Geography and History*. University of Wales Press: Cardiff.

Breslow, N. E. (1984). Extra-Poisson Variation in Log-Linear Models, *Applied Statistics*, 33, 38-44.

Breslow, N. E. and Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models. *Journal of The American Statistical Association*, 88, 9-25.

Breverton, T. (2012). *Wales: A Historical Companion.* Amberley Publishing Limited: Stroud.

Brockmann, D. and Helbing, D. (2013). The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science*, 342, 1337-1342.

Brockmann, D., Hufnagel, L. and Geisel, T. (2006). The Scaling Laws of Human Travel. *Nature*, 439, 462-465.

Brooks, S. (1998). Markov Chain Monte Carlo Method and its Application. *Journal of The Royal Statistical Society: Series D (The Statistician)*, 47, 69-100.

Broutin, H., Elguero, E., Simondon, F. and Guégan, J. (2004b). Spatial Dynamics of Pertussis in A Small Region of Senegal. *Proceedings of the Royal Society of London B: Biological Sciences*, 271, 2091-2098.

Broutin, H., Mantilla-Beniers, N. and Rohani, P. (2007). Ecology of Infectious Diseases: An Example with Two Vaccine-Preventable Infectious Diseases, pp. 189-198 In Tibayrenc, M. (Ed.). (2007). *Encyclopaedia of Infectious Diseases: Modern Methodologies*. John Wiley & Sons: New York, NY.

Broutin, H., Mantilla-Beniers, N. B., Simondon, F., Aaby, P., Grenfell, B. T., Guégan, J. F. and Rohani, P. (2005). Epidemiological Impact of Vaccination on The Dynamics of Two Childhood Diseases in Rural Senegal. *Microbes and Infection*, 7, 593-599.

Broutin, H., Simondon, F. and Guégan, J. (2004a). Whooping Cough Metapopulation Dynamics in Tropical Conditions: Disease Persistence and Impact of Vaccination. *Proceedings of The Royal Society of London B: Biological Sciences*, 271, S302-S305.

Broutin, H., Viboud, C., Grenfell, B. T., Miller, M. A. and Rohani, P. (2010). 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*, 277, 3239-3245.

Brown, L. A. and Holmes, J. (1971). The Delimitation of Functional Regions, Nodal Regions, and Hierarchies by Functional Distance Approaches. *Ekistics*, 32, 387-391.

Brown, L. A. and Horton, F. E. (1970). Functional Distance: An Operational Approach. *Geographical Analysis*, 2, 76-83.

Burges, S. and Moles, K. (2015). The Heads of The Valleys, pp. 79-94 in Jones, M., Orford, S. and MacFarlane, V. (Eds.). *People, Places and Policy: Knowing Contemporary Wales Through New Localities*. Vol. 88. Routledge: London.

Butt, R. V. J. (1995). *The Directory of Railway Stations: Details Every Public and Private Passenger Station, Halt, Platform and Stopping Place, Past and Present.* 1st ed. Patrick Stephens: Yeovil.

Campbell, H., Amirthalingam, G., Andrews, N., Fry, N. K., George, R. C., Harrison, T. G. and Miller, E. (2012). Accelerating Control of Pertussis in England and Wales. *Emerging Infectious Diseases*, 18, 38-47.

Capaldi, A., Behrend, S., Berman, B., Smith, J., Wright, J. and Lloyd, A. L. (2012). Parameter Estimation And Uncertainty Quantification For An Epidemic Model. *Mathematical Biosciences and Engineering*, 9, 553-576.

Cappé, O., Moulines, E., Pesquet, J. C., Petropulu, A. P. and Yang, X. (2002). Long-Range Dependence and Heavy-Tail Modeling For Teletraffic Data. *IEEE Signal Processing Magazine*, 19, 14-27.

Carter, C. F. (1962). *Manchester and its Region.* Manchester University Press: Manchester.

Cauchemez, S., Carrat, F., Viboud, C., Valleron, A. J. and Boelle, P. Y. (2004). A Bayesian MCMC Approach to Study Transmission of Influenza: Application to Household Longitudinal Data. *Statistics in Medicine*, 23, 3469-3487.

Chaudhuri, A., Mandaviya, K., Badelia, P. and Ghosh, S. K. (2017). Optical character recognition systems, pp. 9-41 in Chaudhuri, A., Mandaviya, K., Badelia, P. and Ghosh, S. K. (Eds.). *Optical Character Recognition Systems for Different Languages with Soft Computing*. Springer Nature: London.

Cherry, J. D. (1998). Pertussis in Adults. *Annals of Internal Medicine*, 128, 1047-1048.

Cherry, J. D. (1999). Pertussis in The Preantibiotic and Prevaccine Era, With Emphasis on Adult Pertussis. *Clinical Infectious Diseases*, 28, S107-S111.

Cherry, J. D. (2012). Epidemic Pertussis in 2012—The Resurgence of a Vaccine-Preventable Disease. *New England Journal of Medicine*, 367, 785-787.

Choisy, M. and Rohani, P. (2012). Changing Spatial Epidemiology of Pertussis in Continental USA. *Proceedings of the Royal Society of London B: Biological Sciences*, 279, 4574-4581.

Christensen, P. E., Schmidt, H., Bang, H. O., Andersen, V., Jordal, B. and Jensen, O. (1953). An Epidemic of Measles in Southern Greenland, 1951; Measles in Virgin Soil. II. The Epidemic Proper. *Acta Medica Scandinavica*, 144, 430.

Clapson, M. (1998). *Invincible Green Suburbs, Brave New Towns: Social Change and Urban Dispersal in Post-war England*. Manchester University Press: Manchester.

Clark, J. S. and Bjørnstad, O. N. (2004). Population Time Series: Process Variability, Observation Errors, Missing Values, Lags, and Hidden States. *Ecology*, 85, 3140-3150.

Clarkson, J. A. and Fine, P. E. (1985). The Efficiency of Measles and Pertussis Notification in England and Wales. *International Journal of Epidemiology*, 14, 153-168.

Clauset, A., Shalizi, C. R. and Newman, M. E. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*, 51, 661-703.

Cliff, A. D. and Haggett, P. (1980). Changes in The Seasonal Incidence of Measles in Iceland, 1896–1974. *Journal of Hygiene*, 85, 451-457.

Cliff, A. D. and Haggett, P. (1989). Spatial Aspects of Epidemic Control. *Progress in Human Geography*, 13, 315-347.

Cliff, A. D. and Haggett, P. (1995). The Epidemiological Significance of Islands. *Health & Place*, 1, 199-209.

Cliff, A. D., Haggett, P. and Ord, J. K. (1986). *Spatial Aspects of Influenza Epidemics*. Pion: London.

Cliff, A. D., Haggett, P. and 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. and Smallman-Raynor, M. R. (2000). *Island Epidemics*. Oxford University Press: Oxford

Cliff, A. D., Haggett, P. and Stroup, D. F. (1992). The Geographic Structure of Measles Epidemics in the North-eastern United States. *American Journal of Epidemiology*, 136, 592-602.

Cliff, A. D., Haggett, P. and Stroup, D. F. (1992b). The Geographic Structure of Measles Epidemics in The North-eastern United States. *American Journal of Epidemiology*, 136, 592-602.

Cliff, A. D., Haggett, P., Stroup, D. F. and Cheney, E. (1992a). The Changing Geographical Coherence of Measles Morbidity in the United States, 1962–88. *Statistics in Medicine*, 11, 1409-1424.

Cliff, A. D., Ord, J. K., Haggett, P. and Versey, G. R. (1981). *Spatial Diffusion: An Historical Geography of Epidemics in An Island Community*. Cambridge University Press: Cambridge.

Cliff, A., Haggett, P. and Smallman-Raynor, M. (1998). *Deciphering Global Epidemics: Analytical Approaches to The Disease Records of World Cities*, 1888-1912. Cambridge University Press: Cambridge.

Cole, B., King, S., Ogutu, B., Palmer, D., Smith, G. and Balzter, H. (2015). *Corine Land Cover 2012 for the UK, Jersey and Guernsey*. NERC Environmental Information Data Centre.

Coleman, S. (2015). The Historical Association Between Measles and Pertussis: A Case of Immune Suppression? *SAGE Open Medicine*, 3, 2050312115621315.

Coleman, S. (2016). The Association Between Varicella (Chickenpox) and Group A Streptococcus Infections in Historical Perspective. *SAGE Open Medicine*, *4*, 2050312116658909.

Conduit, B. (1997). *Brecon Beacons and Glamorgan Walks. Pathfinder Guide*. Jarrold Publishing and Ordnance Survey: Norwich.

Congdon, P. and Shepherd, J. (1986). Modelling Population Changes in Small English Urban Areas. *Environment and Planning A*, 18, 1297-1322.

Conlan, A. J and Grenfell, B. T. (2007). Seasonality and The Persistence and Invasion of Measles. *Proceedings of The Royal Society of London B: Biological Sciences*, 274, 1133-1141.

Conlan, A. J., Rohani, P., Lloyd, A. L., Keeling, M. and Grenfell, B. T. (2009). Resolving the Impact of Waiting Time Distributions on The Persistence of Measles. *Journal of The Royal Society Interface*, Rsif20090284.

Connop-Price, M. R. (2004). *Pembrokeshire: The Forgotten Coalfield*. Landmark Publishing: Ashbourne.

Coombes, M. (2019). Commuting Contrasts in Post-Industrial England: Mobility in the World's First Urban Industrial City Regions. *Built Environment*, 45, 476-492.

Coombs, A. S. and Hinch, L. W. (1969). The Heads of The Valleys Road. *Proceedings of The Institution of Civil Engineers*, 44, 89-118.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of The Royal Statistical Society: Series B (Methodological)*, 34, 187-202.

Crowcroft, N. S. and Pebody, R. G. (2006). Recent Developments in Pertussis. *The Lancet*, 367, 1926-1936.

Cutts, F. T. and Markowitz, L. E. (1994). Successes and Failures in Measles Control. *Journal of Infectious Diseases*, 170, S32-S41.

Czado, C., Gneiting, T. and Held, L. (2009). Predictive Model Assessment for Count Data. *Biometrics*, 65, 1254-1261.

Daley, D. J. and Gani, J. (1999). *Epidemic Modelling: An Introduction*. Cambridge University Press: Cambridge.

Daley, D. J. and Gani, J. (1999). Models for The Spread of Infection Via Pairing at Parties, pp. 95-113 in Feldman, R. M. and Valdez-Flores, C. (Eds) *Applied Probability and Stochastic Processes*: Springer, Boston, MA.

Dalziel, B. D., Bjørnstad, O. N., van Panhuis, W. G., Burke, D. S., Metcalf, C. J. E. and Grenfell, B. T. (2016). Persistent Chaos of Measles Epidemics in the Pre-vaccination United States Caused by A Small Change in Seasonal Transmission Patterns. *PLoS Computational Biology*, 12, e1004655.

David, R., Blewitt, N., Johnston, E. and Grazier, S. (2004). *The Socio-Economic Characteristics of the South Wales Valleys in a Broader Context*. Institute of Welsh Affairs: Cardiff.

Davidkin, I., Valle, M., Peltola, H., Hovi, T., Paunio, M., Roivainen, M., Linnavuori, K., Jokinen, S. and Leinikki, P. (1998). Etiology Of Measles-And Rubella-Like Illnesses in Measles, Mumps, and Rubella—Vaccinated Children. *The Journal of Infectious Diseases*, 178, 1567-1570.

Davies, J. (2007). *A History of Wales*. Penguin: London.

Davies, J., Jenkins, N. and Baines, M. (2008). *The Welsh Academy Encyclopaedia of Wales*. University of Wales Press: Cardiff.

Deeks, S., De Serres, G., Boulianne, N., Duval, B., Rochette, L., Déry, P. and Halperin, S. (1999). Failure of Physicians to Consider the Diagnosis of Pertussis in Children. *Clinical Infectious Diseases*, 28, 840-846.

Deng, D. and Paul, S. R. (2005). Score Tests for Zero-Inflation and Over-Dispersion in Generalized Linear Models. *Statistica Sinica*, 15, 257-276.

Dewhurst, R. (1984). Forecasting in Greater Manchester: A Multi-Regional Approach. *The Town Planning Review*, 453-472.

Dick, G. F. and Dick, G. H. (1924a). The Etiology of Scarlet Fever. *Journal of The American Medical Association*, 82, 301-302.

Dick, G. F. and Dick, G.H. (1924b). A Skin Test for Susceptibility to Scarlet Fever. *Journal of The American Medical Association*, 82, 265-266.

Dietz, K. (1995). Some Problems in The Theory of Infectious Disease Transmission and Control, pp 3-16 in Mollison, D. (ed.). *Epidemic Models: Their Structure and Relation to Data*. Cambridge University Press: Cambridge.

Diggle, P. J., Heagerty, P., Liang, K. Y., Heagerty, P. J. and Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford University Press: Oxford.

Dowell, S. F. and Bresee, J. S. (2008). Pandemic Lessons from Iceland. *Proceedings of The National Academy of Sciences*, 105, 1109-1110.

Duncan, C. J., Duncan, S. R. and Scott, S. (1996). The Dynamics of Scarlet Fever Epidemics in England and Wales in the 19th Century. *Epidemiology & Infection*, 117, 493-499.

Duncan, D. L. (2019). Scarlet Fever: Clinical Overview of a Rising Public Health Concern. *British Journal of School Nursing*, 14, 65-68.

Duncan, S. R., Scott, S. and Duncan, C. J. (2000). Modelling the Dynamics of Scarlet Fever Epidemics in the 19th Century. *European Journal of Epidemiology*, 16, 619-626.

Earn, D. J., Rohani, P. and Grenfell, B. T. (1998). Persistence, Chaos and Synchrony in Ecology and Epidemiology. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265, 7-10.

Earn, D. J., Rohani, P., Bolker, B. M. and Grenfell, B. T. (2000). A Simple Model for Complex Dynamical Transitions in Epidemics. *Science*, 287, 667-670.

Edwards, J. A. (1980). The Swansea City-Region: A Case Study of a Gateway System. *Geography*, 65, 81-94.

Edwards, K. M. (2005). Overview of Pertussis: Focus on Epidemiology, Sources of Infection, and Long-Term Protection After Infant Vaccination. *The Pediatric Infectious Disease Journal*, 24, S104-S108.

Eggo, R. M., Cauchemez, S. and Ferguson, N. M. (2010). Spatial Dynamics of the 1918 Influenza Pandemic in England, Wales and The United States. *Journal of the Royal Society Interface*, 8, 233-243.

Eichner, M., Hadeler, K. P. and Dietz, K. (1996). Stochastic Models for The Eradication of Poliomyelitis: Minimum Population Size for Polio Virus Persistence, pp. 315-327 in Isham, V. and Medley, G. (Eds). *Models for Infectious Human Diseases: Their Structure and Relation to Data*. Cambridge University Press: Cambridge.

Eidlitz-Markus, T., Mimouni, M. and Zeharia, A. (2007). Pertussis Symptoms in Adolescents and Children Versus Infants: The Influence of Vaccination and Age. *Clinical Pediatrics*, 46, 718-723.

Epstein, E. S. (1969). A Scoring System for Probability Forecasts of Ranked Categories. *Journal of Applied Meteorology*, 8, 985-987.

Erlander, S. and Stewart, N. F. (1990). *The Gravity Model in Transportation Analysis: Theory and Extensions*. VSP: Utrecht, The Netherlands.

Eubank, S., Guclu, H., Kumar, V. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z. and Wang, N. (2004). Modelling Disease Outbreaks in Realistic Urban Social Networks. *Nature*, 429, 180-184.

Evans, C. J. O. (1948). *Glamorgan, its History and Topography*. William Lewis: Cardiff.

Fair, T. J. (1974). Decentralisation, Dispersal and Dispersion. *South African Geographical Journal,* 56, 94-96.

Fanshawe, T. R., Diggle, P. J., Rushton, S., Sanderson, R., Lurz, P. W. W., Glinianaia, S. V., Pearce, M. S., Parker, L., Charlton, M. and Pless-Mulloli, T. T. (2008). Modelling Spatio-Temporal Variation in Exposure to Particulate Matter: A Two-Stage Approach. *Environmetrics: The Official Journal of the International Environmetrics Society*, 19, 549-566.

Farmer, E. and Smith, R. (1975). Overspill Theory: A Metropolitan Case Study. *Urban Studies*, 12, 151-168.

Farrington, C. P., Andrews, N. J., Beale, A. D. and Catchpole, M. A. (1996). A Statistical Algorithm for The Early Detection of Outbreaks of Infectious Disease. *Journal of The Royal Statistical Society: Series A (Statistics in Society)*, 159, 547-563.

Farrington, C. P., Kanaan, M. N. and Gay, N. J. (2003). Branching Process Models for Surveillance of Infectious Diseases Controlled by Mass Vaccination. *Biostatistics*, 4, 279-295.

Felton, H. M. and Willard, C. Y. (1944). Current Status of Prophylaxis by Hemophilus Pertussis Vaccine. *Journal of The American Medical Association*, 126, 294-299.

Ferrari, M. J., Grais, R. F., Bharti, N., Conlan, A. J., Bjørnstad, O. N., Wolfson, L. J., Guerin, P. J., Djibo, A. and Grenfell, B. T. (2008). The Dynamics of Measles in Sub-Saharan Africa. *Nature*, 451, 679-684.

Ferrari, M. J., Grais, R. F., Bharti, N., Conlan, A. J., Bjørnstad, O. N., Wolfson, L. J., Guerin, P.J., Djibo, A. and Grenfell, B. T. (2008). The Dynamics of Measles in Sub-Saharan Africa. *Nature*, 451, 679-684.

Ferretti, J. and Köhler, W. (2016). History of Streptococcal Research, pp. 1-26 In Ferretti, J. J., Stevens, D. L. and Fischetti, V. A. (Eds.) *Streptococcus Pyogenes: Basic Biology to Clinical Manifestations* [Internet]. University of Oklahoma Health Sciences Center: Oklahoma City, OK.

Field, D. (1968). New Town and Town Expansion Schemes. Part III. *Town Planning Review*, 39, 196.

Fine, P. E. (1993). Herd Immunity: History, Theory, Practice. *Epidemiologic Reviews*, 15, 265-302.

Fine, P. E. and Clarkson, J. (1982). The Recurrence of Whooping Cough: Possible Implications for Assessment of Vaccine Efficacy. *The Lancet*, 319, 666-669.

Fine, P. E. and Clarkson, J. A. (1982). Measles in England and Wales – II: The Impact of The Measles Vaccination Programme on The Distribution of Immunity in The Population. *International Journal of Epidemiology*, 11, 15-25.

Fine, P. E. and Clarkson, J. A. (1983). Measles in England and Wales – III: Assessing Published Predictions of The Impact of Vaccination on Incidence. *International Journal of Epidemiology*, 12, 332-339.

Fine, P. E. and Clarkson, J. A. (1987). Reflections on The Efficacy of Pertussis Vaccines. *Reviews of Infectious Diseases*, 9, 866-883.

Fine, P., Eames, K. and Heymann, D. L. (2011). “Herd Immunity”: A Rough Guide. *Clinical Infectious Diseases*, 52, 911-916.

Finkenstädt, B. and Grenfell, B. (1998). Empirical Determinants of Measles Metapopulation Dynamics in England and Wales. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265, 211-220.

Finkenstädt, B. F. and Grenfell, B. T. (2000). Time Series Modelling of Childhood Diseases: A Dynamical Systems Approach. *Journal of The Royal Statistical Society: Series C (Applied Statistics)*, 49, 187-205.

Finkenstädt, B. F., Bjørnstad, O. N. and Grenfell, B. T. (2002). A Stochastic Model for Extinction and Recurrence of Epidemics: Estimation and Inference for Measles Outbreaks. *Biostatistics*, *3*, 493-510.

Frangopulo, N. J. (1977). *Tradition in Action: The Historical Evolution of the Greater Manchester County*. EP Publishing: New York.

Freeman, T. W. and Snodgrass, C. P. (1959). *The Conurbations of Great Britain.* Manchester University Press: Manchester.

Gale, A. H. (1945). A Century of Changes in The Mortality and Incidence of The Principal Infections of Childhood*. Archives of Disease in Childhood*, 20, 2-21.

Gareth Evans, D. (1989). *A History of Wales 1815–1906*. University of Wales Press: Cardiff.

Gatrell, A. C. and Bailey, T. C. (1996). Interactive Spatial Data Analysis in Medical Geography. *Social Science & Medicine*, 42*,* 843-855.

Gehlenborg, N. and Wong, B. (2012). Points of View: Heat Maps. *Nature Methods*, 9, 213.

Geilhufe, M., Held, L., Skrøvseth, S. O., Simonsen, G. S. and Godtliebsen, F. (2014). Power Law Approximations of Movement Network Data for Modeling Infectious Disease Spread. *Biometrical Journal*, 56, 363-382.

Gelfand, A. E. and Ghosh, S. K. (1998). Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika,* 85, 1-11.

General Register Office (1964a). *Census 1961: England and Wales: County Report: Brecknockshire and Carmarthenshire*: H.M.S.O: London.

General Register Office (1964b). *Census 1961: England and Wales: County Report: Cardiganshire and Pembrokeshire*: H.M.S.O: London.

Gibb, R. (2005). *Greater Manchester: A Panorama of People and Places in Manchester and Its Surrounding Towns*. Myriad Books: London.

Gibbons, C. L., Mangen, M. J. J., Plass, D., Havelaar, A. H., Brooke, R. J., Kramarz, P., Peterson, K. L., Stuurman, A. L., Cassini, A., Fèvre, E. M. and Kretzschmar, M. E. (2014).

Measuring Underreporting and Under-Ascertainment in Infectious Disease Datasets: A Comparison of Methods. *BMC Public Health*, 14, 147.

Gill, J. (2008). Is Partial-Dimension Convergence A Problem for Inferences From MCMC Algorithms? *Political Analysis*, 16, 153-178.

Glass, K., Kappey, J. and Grenfell, B. T. (2004). The Effect of Heterogeneity in Measles Vaccination on Population Immunity. *Epidemiology & Infection*, 132, 675-683.

Gneiting, T. and Raftery, A. E. (2007). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of The American Statistical Association*, 102, 359-378.

Good, I. J. (1952). Rational Decisions. *Journal of The Royal Statistical Society, Series B,* 14, 107-114.

Goodall, C. R., Kafadar, K. and Tukey, J. W. (1998). Competing and Using Moral Versus Urban Measures in Statistical Applications. *The American Statistician*, 52, 101-111.

Gopal, D. P., Barber, J. and Toeg, D. (2019). Pertussis (Whooping Cough). *BMJ*, 364, L401.

Gottfredsson, M., Halldórsson, B. V., Jónsson, S., Kristjánsson, M., Kristjánsson, K., Kristinsson, K. G., Löve, A., Blöndal, T., Viboud, C., Thorvaldsson, S. and Jónsdóttir, I. (2008). Lessons from The Past: Familial Aggregation Analysis of Fatal Pandemic Influenza (Spanish flu) in Iceland in 1918. *Proceedings of the National Academy of Sciences*, 105, 1303-1308.

Greenberg, B. L., Sack, R. B., Salazar-Lindo, E., Budge, E., Gutierrez, M., Campos, M., Visberg, A., Leon-Barna, R., Vi, A., Maurutia, D. and Gomez, M. (1991). Measles-Associated Diarrhea in Hospitalized Children in Lima, Peru: Pathogenic Agents and Impact on Growth. *Journal of Infectious Diseases*, 163, 495-502.

Greenland, S. (2007). Bayesian Perspectives for Epidemiological Research. II. Regression Analysis. *International Journal of Epidemiology*, 36, 195-202.

Gregory, I. N., Bennett, C., Gilham, V. L. and Southall, H. R. (2002). The Great Britain Historical GIS Project: From Maps to Changing Human Geography. *The Cartographic Journal*, 39, 37-49.

Grenfell, B. and Harwood, J. (1997). (Meta) Population Dynamics of Infectious Diseases. *Trends in Ecology & Evolution*, 12, 395-399.

Grenfell, B. T. (1998). Cities and Villages: Infection Hierarchies in A Measles Metapopulation. *Ecology Letters*, 1, 68-70.

Grenfell, B. T. and Anderson, R. M. (1989). Pertussis in England and Wales: An Investigation of Transmission Dynamics and Control by Mass Vaccination. *Proceedings of The Royal Society of London. B. Biological Sciences*, 236, 213-252.

Grenfell, B. T. and Bolker, B. M. (1998). Cities and Villages: Infection Hierarchies in A Measles Metapopulation. *Ecology Letters*, 1, 63-70.

Grenfell, B. T. and Harwood, J. (1997). (Meta) Population Dynamics of Infectious Diseases. *Trends in Ecology & Evolution*, 12, 395-399.

Grenfell, B. T., Bjørnstad, O. N. and Kappey, J. (2001). Travelling Waves and Spatial Hierarchies in Measles Epidemics. *Nature*, 414, 716-723.

Grenfell, B. T., Bolker, B. M. and Kleczkowski, A. (1995). Seasonality and Extinction in Chaotic Metapopulations. *Proceedings of The Royal Society of London B: Biological Sciences*, 259, 97-103.

Griffith, A. H. (1978). Reactions After Pertussis Vaccine: A Manufacturer's Experiences and Difficulties Since 1964. *BMJ*, 1, 809-815.

Griffiths, D. A. (1973). The Effect of Measles Vaccination on The Incidence of Measles in The Community. *Journal of The Royal Statistical Society. Series A (General)*, 136, 441-449.

Gschlößl, S. and Czado, C. (2008). Modelling Count Data with Overdispersion and Spatial Effects. *Statistical Papers*, 49, 531-556.

Guerra, F. M., Bolotin, S., Lim, G., Heffernan, J., Deeks, S. L., Li, Y. and Crowcroft, N. S. (2017). The Basic Reproduction Number (R_0) Of Measles: A Systematic Review. *The Lancet Infectious Diseases*, 17, E420-E428.

Guerrant, R. L., Walker, D. H. and Weller, P. F. (1999), pp. 203-211 in (Eds). *Tropical Infectious Diseases: Principles, Pathogens and Practice*. Vol. 1 & 2. Churchill Livingstone: New York, NY.

Gunning, C. E. and Wearing, H. J. (2013). Probabilistic Measures of Persistence and Extinction in Measles (Meta) Populations. *Ecology Letters*, 16, 985-994.

Gunning, C. E., Erhardt, E. and Wearing, H. J. (2014). Conserved Patterns of Incomplete Reporting in Pre-Vaccine Era Childhood Diseases. *Proceedings of the Royal Society B: Biological Sciences*, 281, 1-9.

Guttorp, P. and Minin, V. N. (1995). *Stochastic Modelling of Scientific Data*. CRC Press: London.

Hagenaars, T. J., Donnelly, C. A. and Ferguson, N. M. (2004). Spatial Heterogeneity and The Persistence of Infectious Diseases. *Journal of Theoretical Biology*, 229, 349-359.

Haining, R., Law, J. and Griffith, D. (2009). Modelling Small Area Counts in The Presence of Overdispersion and Spatial Autocorrelation. *Computational Statistics & Data Analysis*, 53, 2923-2937.

Hall, P. (1974). The Containment of Urban England. *Geographical Journal*, 140, 386-408.

Hamer, W. H. (1906). *The Milroy Lectures on Epidemic Disease in England: The Evidence of Variability and of Persistency of Type*. Bedford Press: London.

Hamnett, C. and Randolph, W. G. (1982). The Changing Population Distribution of England and Wales, 1961–81: Clean Break or Consistent Progression? *Built Environment (1978-)*, 8, 272-280.

Hanski, I. and Gaggiotti, O. E. (2004*). Ecology, Genetics, and Evolution of Metapopulations.* Academic Press: London.

Hanski, I. and Gyllenberg, M. (1993). Two General Metapopulation Models and The Core-Satellite Species Hypothesis. *The American Naturalist*, 142, 17-41.

Harrison, J. (2012). Life After Regions? The Evolution of City-Regionalism in England. *Regional Studies*, 46, 1243-1259.

Harrison, S. (1991). Local Extinction in A Metapopulation Context: An Empirical Evaluation. *Biological Journal of The Linnean Society*, 42, 73-88.

Harrison, S. and Taylor, A. D. (1997). Empirical Evidence for Metapopulation Dynamics, pp. 27-42 in Hanski, I. and Gulping, M. E. (eds), *Metapopulation Dynamics: Ecology, Genetics and Evolution*. Academic Press: London.

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E., Robinson, B. S., Hodgson, D. J. and Inger, R. (2018). A Brief Introduction to Mixed Effects Modelling and Multi-Model Inference in Ecology. *PeerJ*, 6, e4794.

Harvey, A. M. (1997). Fluvial Geomorphology of North-West England, pp. 173–200 in Gregory, K. J. (1997) (Eds.). *Fluvial Geomorphology of Great Britain.* 1st ed. Springer: Dordrecht.

Harville, D. A. (1998). *Matrix Algebra from A Statistician's Perspective*. Springer: New York, 1997.

Hartwell, C., Hyde, M. and Pevsner, N. (2004). *Lancashire: Manchester and the South-east*. Yale University Press: Yale.

Hasluck, E. L. (1948). *Local Government in England*. Cambridge University Press: Cambridge.

Heathcote, C. R. (1965). A Branching Process Allowing Immigration. *Journal of The Royal Statistical Society: Series B (Methodological)*, 27, 138-143.

Hebbert, M. and Deas, I. (2000). Greater Manchester-'Up and Going'? *Policy & Politics*, 28, 79-92.

Held, L. and Paul, M. (2012). Modelling Seasonality in Space-Time Infectious Disease Surveillance Data. *Biometrical Journal*, 54, 824-843.

Held, L., Hofmann, M., Höhle, M. and Schmid, V. (2006). A Two-Component Model for Counts of Infectious Diseases. *Biostatistics*, 7, 422-437.

Held, L., Höhle, M. and Hofmann, M. (2005). A Statistical Framework for The Analysis of Multivariate Infectious Disease Surveillance Counts. *Statistical Modelling*, 5, 187-199.

Hens, N., Ayele, G. M., Goeyvaerts, N., Aerts, M., Mossong, J., Edmunds, J. W. and Beutels, P. (2009). Estimating the Impact of School Closure on Social Mixing Behaviour and The Transmission of Close Contact Infections in Eight European Countries. *BMC Infectious Diseases*, *9*, 1-12.

Herzog, S. A., Paul, M. and Held, L. (2011). Heterogeneity in Vaccination Coverage Explains the Size and Occurrence of Measles Epidemics in German Surveillance Data. *Epidemiology and Infection*, 139, 505–515.

Hobcraft, J. (1996). Fertility in England and Wales: A Fifty-Year Perspective. *Population Studies*, 50, 485-524.

Höhle, M. (2007). Surveillance: An R Package for The Monitoring of Infectious Diseases. *Computational Statistics*, 22, 571-582.

Höhle, M. and Mazick, A. (2010). Aberration Detection in R Illustrated by Danish Mortality Monitoring. *Biosurveillance: Methods and Case Studies*, 3, 215-237.

Hsu, J. S. and Leonard, T. (1997). Hierarchical Bayesian Semiparametric Procedures for Logistic Regression. *Biometrika*, 84, 85-93.

Humphrys, G. (1986). Dealing with Dilemmas: Industrial Landscapes in South Wales. *Landscape Research*, 11, 11-14.

Hunt, M. (2011). *Cardiff and South East Wales: Social, Economic and Sustainability Context. Retrofit 2050 Working Paper*. Cardiff University Press: Cardiff.

Huppert, A. and Katriel, G. (2013). Mathematical Modelling and Prediction in Infectious Disease Epidemiology. *Clinical Microbiology and Infection*, 19, 999-1005.

Jacob, C. (2010). Branching Processes - Their Role in Epidemiology. *International Journal of Environmental Research and Public Health*, 7, 1186-1204.

Jandarov, R., Haran, M., Bjørnstad, O. and Grenfell, B. T. (2014). Emulating A Gravity Model to Infer the Spatiotemporal Dynamics of An Infectious Disease. *Journal of The Royal Statistical Society: Series C (Applied Statistics)*, 63, 423-444.

Jansen, V. A., Stollenwerk, N., Jensen, H. J., Ramsay, M. E., Edmunds, W. J. and Rhodes, C. J. (2003). Measles Outbreaks in A Population with Declining Vaccine Uptake. *Science*, 301, 804-804.

Jenkins, P. (2014). *A History of Modern Wales, 1536–1990*. Longman: London.

Jenkinson, D. (1983). Whooping Cough: What Proportion of Cases is Notified in an Epidemic? *BMJ*, 287, 185.

John, A. H. (1980) Introduction: Glamorgan 1700-1750, pp. 1-43 in John, A.H. and Williams, G. (Eds.), *Glamorgan County History, Vol. V – Industrial Glamorgan from 1700 to 1970*. University of Wales Press: Cardiff.

Joint Committee on Vaccination and Immunisation (JCVI). (1977). *Report on The Review of The Evidence on Whooping Cough Vaccination, 1977*. The National Archives: London.

Jones, J.G. (2014). The History of Wales. University of Wales Press: Cardiff.

Jones, P. N. (1969). *Colliery Settlement in the South Wales Coalfield 1850-1926*. University of Hull Publication: Hull.

Kafadar, K. and Tukey, J. W. (1993). US Cancer Death Rates: A Simple Adjustment for Urbanization. *International Statistical Review/Revue Internationale De Statistique*, 61, 257-281.

Kahn, M. J. and Raftery, A. E. (1996). Discharge Rates of Medicare Stroke Patients to Skilled Nursing Facilities: Bayesian Logistic Regression with Unobserved Heterogeneity. *Journal of The American Statistical Association*, 91, 29-41.

Kaplan, E. L. and Meier, P. (1958). Nonparametric Estimation from Incomplete Observations. *Journal of The American Statistical Association*, 53, 457-48

Katz, A.R. and Morens, D.M. (1992). Severe Streptococcal Infections in Historical Perspective. *Clinical Infectious Diseases*, 14, 298-307.

Katzmann, W. and Dietz, K. (1984). Evaluation of Age-Specific Vaccination Strategies. *Theoretical Population Biology*, 25, 125-137.

Keeble, D. E. (1972). Industrial Movement and Regional Development in The United Kingdom. *Town Planning Review*, 43, 3.

Keeling, M. J. (1997). Modelling the Persistence of Measles. *Trends in Microbiology*, 5, 513-518.

Keeling, M. J. (1999). Correlation Equations for Endemic Diseases: Externally Imposed and Internally Generated Heterogeneity. *Proceedings of the Royal Society of London B: Biological Sciences*, 266, 953-960.

Keeling, M. J. (1999). The Effects of Local Spatial Structure on Epidemiological Invasions. *Proceedings of The Royal Society of London. Series B: Biological Sciences*, 266, 859-867.

Keeling, M. J. (2000). Metapopulation Moments: Coupling, Stochasticity and Persistence. *Journal of Animal Ecology*, 69, 725-736.

Keeling, M. J. and Grenfell, B. T. (1997). Disease Extinction and Community Size: Modelling the Persistence of Measles. *Science*, 275, 65-67.

Keeling, M. J. and Grenfell, B. T. (2002). Understanding the Persistence of Measles: Reconciling Theory, Simulation and Observation. *Proceedings of The Royal Society of London. Series B: Biological Sciences*, 269, 335-343.

Keeling, M. J. and Rohani, P. (2002). Estimating Spatial Coupling in Epidemiological Systems: A Mechanistic Approach. *Ecology Letters*, 5, 20-29.

Keeling, M. J. and Rohani, P. (2008). Introduction to Simple Epidemic Models, pp. 15-53 in Keeling, M. J. and Rohani, P. (eds.). *Modelling Infectious Diseases in Humans and Animals*. Princeton University Press: Princeton.

Keeling, M. J. and Ross, J. V. (2008). On Methods for Studying Stochastic Disease Dynamics. *Journal of The Royal Society Interface*, 5, 171-181.

Keeling, M. J., and Eames, K. T. (2005). Networks and Epidemic Models. *Journal of The Royal Society Interface*, 2, 295-307.

Keeling, M. J., Bjørnstad, O. N. and Grenfell, B. T. (2004). Metapopulation Dynamics of Infectious Diseases, pp. 415-445 in Hanski, I. A. and Gaggiotti, O. E. (Eds.). *Ecology, Genetics and Evolution of Metapopulations*. Academic Press: London.

Keeling, M. J., Woolhouse, M. E. J., Shaw, D. J., Matthews, L., Chase-Topping, M., Haydon, D. T., Cornell, S. J., Kappey, J., Wilesmith, J. and Grenfell, B. T. (2001). Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in A Heterogeneous Landscape. *Science*, 294, 813-817.

Kermack, W. O. and McKendrick, A. G. (1927). A Contribution to The Mathematical Theory of Epidemics. *Proceedings of The Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 115, 700-721.

Kidner, R. W. (1995). *The Rhymney Railway.* Oakwood Press: Headington.

Kleinman, K., Lazarus, R. and Platt, R. (2004). A Generalized Linear Mixed Models Approach for Detecting Incident Clusters of Disease in Small Areas, With an Application to Biological Terrorism. *American Journal of Epidemiology*, 159, 217-224.

Kneib, T. and Fahrmeir, L. (2007). A Mixed Model Approach for Geoadditive Hazard Regression. *Scandinavian Journal of Statistics*, 34, 207-228.

Kucharski, A. J., Conlan, A. J. and Eames, K. T. (2015). School's Out: Seasonal Variation in The Movement Patterns of School Children. *Plos One*, 10, E0128070.

Lamagni, T., Guy, R., Chand, M., Henderson, K. L., Chalker, V., Lewis, J., Saliba, V., Elliot, A.J., Smith, G.E., Rushton, S. and Johnson, A. P. (2018). Resurgence of Scarlet Fever in England, 2014–16: A Population-Based Surveillance Study. *The Lancet Infectious Diseases*, 18, 180-187.

Langlois, D. M. and Andreae, M. (2011). Group A Streptococcal Infections. *Pediatrics In Review-Elk Grove*, 32, 423-433.

Langmuir, A. D. (1962). Medical Importance of Measles. *American Journal of Diseases of Children*, 103, 224-226.

Lavine, J. S. and Rohani, P. (2012). Resolving Pertussis Immunity and Vaccine Effectiveness Using Incidence Time Series. *Expert Review of Vaccines*, 11, 1319-1329.

Lee, C. F., Cowling, B. J. and Lau, E. H. (2017). Epidemiology of Re-emerging Scarlet Fever, Hong Kong, 2005–2015. *Emerging Infectious Diseases*, *23*, 1707.

Lee, Y., Nelder, J. A. and Pawitan, Y. (2018). *Generalized Linear Models with Random Effects: Unified Analysis Via H-Likelihood.* Chapman and Hall: London.

Lekone, P. E. and Finkenstädt, B. F. (2006). Statistical Inference in A Stochastic Epidemic SEIR Model with Control Intervention: Ebola As A Case Study. *Biometrics*, 62, 1170-1177.

Leroux, B. G., Lei, X. and Breslow, N. (2000). Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence, pp. 179-191 in Halloran, M. E. and Berry, D.

(Eds.) *Statistical Models in Epidemiology, The Environment, and Clinical Trials*. Springer: New York.

Lessler, J. and Cummings, D. A. (2016). Mechanistic Models of Infectious Disease and Their Impact on Public Health. *American Journal of Epidemiology*, 183, 415-422.

Levins, R. (1969). Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bulletin of The Entomological Society of America*, 15, 237-240.

Lewis, E. D. (1959). *The Rhondda Valleys*. Phoenix House: London.

Li, N., Qian, G. and Huggins, R. (2003). A Random Effects Model for Diseases with Heterogeneous Rates of Infection. *Journal of Statistical Planning and Inference*, 116, 317-332.

Lloyd, A. L. and May, R. M. (1996). Spatial Heterogeneity in Epidemic Models. *Journal of Theoretical Biology*, 179, 1-11.

Lloyd, A. L. and Sattenspiel, L. (2009). Spatiotemporal Dynamics of Measles: Synchrony and Persistence in a Disease Metapopulation, pp. 251-272 in Cantrell, S., Cosner, C. and Ruan, S. (eds). *Spatial Ecology*. Chapman & Hall/CRC Press: London.

Lloyd, T., Orbach, J. and Scourfield, R. (2006). *Carmarthenshire and Ceredigion*. Yale University Press: Yale.

London, W. P. and Yorke, J. A. (1973). Recurrent Outbreaks of Measles, Chickenpox and Mumps I. Seasonal Variation in Contact Rates. *American Journal of Epidemiology*, 98*,* 453-468.

Long, S. S. (2004). Pertussis (Bordetella Pertussis and B. Parapertussis), pp. 908-912 In Behrman, R. E., Kliegman, R. M., Jenson, H.B, (Eds.). *Nelson's Textbook of Pediatrics*. 17th Ed. Saunders: Philadelphia.

Long, S. S., Welkon, C. J. and Clark, J. L. (1990). Widespread Silent Transmission of Pertussis in Families: Antibody Correlates of Infection and Symptomatology. *Journal of Infectious Diseases*, 161, 480-486.

Lu, Q., Wu, H., Ding, Z., Wu, C. and Lin, J. (2019). Analysis of Epidemiological Characteristics of Scarlet Fever in Zhejiang Province, China, 2004–2018. *International Journal of Environmental Research and Public Health*, 16, 3454.

Mackie, P. J., Richardson, T. and Tweddle, G. (1995). *Transport in the Trans-Pennine Corridor: Present Conditions and Future Options.* Interregional Study Working Paper 3.

Madge, J. (1962). The New Towns Program in Britain. *Journal of The American Institute of Planners*, 28, 208-219.

Magpantay, F. M. G. and Rohani, P. (2015). Dynamics of Pertussis Transmission in The United States. *American Journal of Epidemiology*, Kwv024.

Mantel, N. (1966). Evaluation of Survival Data and Two New Rank Order Statistics Arising in its Consideration. *Cancer Chemother Rep*, *50*(3), 163-170.

Marks, J. S., Halpin, T. J. and Orenstein, W. A. (1978). Measles Vaccine Efficacy in Children Previously Vaccinated at 12 Months of Age. *Pediatrics*, 62, 955-960.

Martinón-Torres, F., Heininger, U., Thomson, A. and Wirsing von König, C. H. (2018). Controlling Pertussis: How Can We Do It? A Focus on Immunization. *Expert Review of Vaccines*, 17, 289-297.

May, R. M. and Anderson, R. M. (1984). Spatial Heterogeneity and The Design of Immunization Programs. *Mathematical Biosciences*, 72, 83-111.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. CRC Press: London.

McKinnon, N. E. (1946). Mortality Reductions in Ontario, 1900-42 Vi. Scarlet Fever. *Canadian Journal of Public Health/Revue Canadienne De Sante'e Publique*, 37, 407-410.

Metcalf, C. J. E., Bjørnstad, O. N., Grenfell, B. T. and Andreasen, V. (2009). Seasonality and Comparative Dynamics of Six Childhood Infections in Pre-Vaccination Copenhagen. *Proceedings of The Royal Society B: Biological Sciences*, 276, 4111-4118.

Metcalf, C. J. E., Cohen, C., Lessler, J., McAnerney, J. M., Ntshoe, G. M., Puren, A., Klepac, P., Tatem, A., Grenfell, B.T. and Bjørnstad, O. N. (2013a). Implications of Spatially Heterogeneous Vaccination Coverage for The Risk of Congenital Rubella Syndrome in South Africa. *Journal of The Royal Society Interface*, 10, 20120756.

Metcalf, C. J. E., Hampson, K., Tatem, A. J., Grenfell, B. T. and Bjørnstad, O. N. (2013b). Persistence in Epidemic Metapopulations: Quantifying the Rescue Effects for Measles, Mumps, Rubella and Whooping Cough. *PLoS One*, 8, E74696.

Meyer, H. M., Brooks, B. E., Douglas, R. D. and Rogers, N. G. (1962). Ecology of Measles in Monkeys. *American Journal of Diseases of Children*, 103, 307-313.

Meyer, S. and Held, L. (2014). Power-Law Models for Infectious Disease Spread. *The Annals of Applied Statistics*, *8*, 1612-1639.

Meyer, S., Held, L. and Höhle, M. (2014). Spatio-Temporal Analysis of Epidemic Phenomena Using the R Package Surveillance. *Journal of Statistical Software*, 77, 1-55.

Miller, E., Vurdien, J. E. and White, J. M. (1992). The Epidemiology of Pertussis in England and Wales: Communicable Disease Report. *CDR Review*, 2, 152-154.

Milligan, I. (2013). Illusionary Order: Online Databases, Optical Character Recognition, and Canadian History, 1997–2010. *Canadian Historical Review*, 94, 540-569.

Mina, M. J., Kula, T., Leng, Y., Li, M., De Vries, R. D., Knip, M., Siljander, H., Rewers, M., Choy, D. F., Wilson, M. S. and Larman, H. B. (2019). Measles Virus Infection Diminishes Pre-existing Antibodies That Offer Protection from Other Pathogens. *Science*, 366, 599-606.

Mollison, D., Isham, V. and Grenfell, B. (1994). Epidemics: Models and Data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 157, 115-149.

Mooney, G. (2015). *Intrusive Interventions: Public Health, Domestic Space, and Infectious Disease Surveillance in England, 1840-1914.* University of Rochester Press: Rochester, NY.

Moran, P. A. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, 37, 17-23.

Morgan, K. O. (1981). *Rebirth of A Nation: Wales, 1880-1980*. Vol. 6. Clarendon Press: Oxford.

Morris, J. H. and Williams, L. J. (1958). *The South Wales Coal Industry, 1841-1875*. University of Wales Press: Cardiff.

Moss, W. J. and Griffin, D. E. (2006). Global Measles Elimination. *Nature Reviews Microbiology*, 4, 900-908.

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G.S., Wallinga, J. and Heijne, J. and Edmunds, W. J. (2008). Social Contacts and Mixing Patterns Relevant to The Spread of Infectious Diseases. *Plos Med*, 5, E74.

Muloiwa, R., Wolter, N., Mupere, E., Tan, T., Chitkara, A. J., Forsyth, K. D., von König, C. H. W. and Hussey, G. (2018). Pertussis in Africa: Findings and Recommendations of The Global Pertussis Initiative (GPI). *Vaccine*, 36, 2385-2393.

Munro, A.D., Smallman-Raynor, M. and Algar, A. C. (2021). Long-Term Changes in Endemic Threshold Populations for Pertussis in England and Wales: A Spatiotemporal Analysis of Lancashire and South Wales, 1940-69. *Social Science and Medicine*, 288, 113295.

Murphy, M. (2011). Long-Term Effects of The Demographic Transition on Family and Kinship Networks in Britain. *Population and Development Review*, 37, 55-80.

Murray, G. D. and Cliff, A. D. (1977). A Stochastic Model for Measles Epidemics in A Multi-Region Setting. *Transactions of The Institute of British Geographers*, 2, 158-174.

Nåsell, I. (2005). A New Look at The Critical Community Size for Childhood Infections. *Theoretical Population Biology*, 67, 203-216.

Nathanson, N. (1982). Eradication of Poliomyelitis in the United States. *Review of Infectious Diseases,* 4, 940-950.

Nathanson, N. and Kew, O. M. (2010). From Emergence to Eradication: The Epidemiology of Poliomyelitis Deconstructed. *American Journal of Epidemiology*, Kwq320.

Nelder, J. A. and Wedderburn, R. W. (1972). Generalized Linear Models. *Journal of The Royal Statistical Society: Series A (General)*, 135, 370-384.

Newman, J., Hughes, S. R. and Ward, A. (1995). *Glamorgan (Mid Glamorgan, South Glamorgan and West Glamorgan)*. Yale University Press: Yale.

Noah, N (2006). *Controlling Communicable Disease*. Open University Press: Maidenhead.

Nokes, D. J. and Anderson, R. M. (1988). The Use of Mathematical Models in The Epidemiological Study of Infectious Diseases and In the Design of Mass Immunization Programmes. *Epidemiology & Infection*, 101, 1-20.

Okubo, A. and Levin, S. A. (2013). *Diffusion and Ecological Problems: Modern Perspectives*. 2nd ed. Springer Science & Business Media: New York.

Orenstein, W. A., Strebel, P. M., Papania, M., Sutter, R. W., Bellini, W. J. and Cochi, S. L. (2000). Measles Eradication: Is It in Our Future? *American Journal of Public Health*, 90, 1521.

Page, J. (1988), *Forgotten Railways: Volume 8 – South Wales*. 2nd ed. David & Charles Publishers: Newton Abbott.

Pal, U. and Dash, N. S. (2014). Language, Script, and Font Recognition, pp. 291-330 in Doermann, D. and Tombre, K. (Eds.). *Handbook of Document Image Processing and Recognition*. Springer: London

Panum, P. L. and Petersen, J. J. (1940). *Observations Made During the Epidemic of Measles on The Faroe Islands in The Year 1846*. Delta Omega Society: New York.

Parker, R, (2008). *The Railways of Pembrokeshire.* Crecy Publishing: Manchester.

Patmore, J. A. (1961). The Railway Network of Merseyside. *Transactions and Papers (Institute of British Geographers)*, 29, 231-244.

Patmore, J. A. (1964). The Railway Network of the Manchester Conurbation. *Transactions and Papers (Institute of British Geographers)*, 34, 159-173.

Paul, M. and Held, L. (2011). Predictive Assessment of a Non-Linear Random Effects Model for Multivariate Time Series of Infectious Disease Counts. *Statistics in Medicine*, 30, 1118-1136.

Paul, M., Held, L. and Toschke, A. M. (2008). Multivariate Modelling of Infectious Disease Surveillance Data. *Statistics in Medicine*, 27, 6250-6267.

Pawitan, Y. (2001). *In All Likelihood: Statistical Modelling and Inference Using Likelihood*. Oxford University Press: Oxford.

Petrova, V.N., Sawatsky, B., Han, A.X., Laksono, B.M., Walz, L., Parker, E., Pieper, K., Anderson, C.A., De Vries, R.D., Lanzavecchia, A. and Kellam, P. (2019). Incomplete Genetic Reconstitution of B Cell Pools Contributes to Prolonged Immunosuppression After Measles. *Science Immunology*, 4(41), eaay6125.

Philip's (1994). *Philip's Atlas of the World*. George Philip & Son: London.

Plotkin, S. A. (2005). Vaccines: Past, Present and Future. *Nature Medicine*, 11, S5-S11.

Plotkin, S. A. (2014). History of Vaccination. *Proceedings of The National Academy of Sciences*, 111, 12283-12287

PMC, J. E. and API, G. R. (1969). Efficacy of Whooping-Cough Vaccines Used in The United Kingdom Before 1968. A Preliminary Report to The Director of the Public Health Laboratory Service by The Public Health Laboratory Service Whooping-Cough Committee and Working Party. *BMJ*, 4, 329-33.

Pollard, R., Pevsner, N. and Sharples, J. (2006). *Lancashire: Liverpool and the Southwest*. Yale University Press: Yale.

Pooley, C. G. (2010). Landscapes Without the Car: A Counterfactual Historical Geography of Twentieth-Century Britain. *Journal of Historical Geography*, 36, 266-275.

Pratt, I. S. (1977)*. "Workedslegh": A History of Worsley.* Manchester Pratt: Manchester.

Public Health England [PHE] (2017). Measles notifications and deaths in England and Wales 1940 to 2013. Available from: https://www.gov.uk/government/publications/measles-deaths-by-age-group-from-1980-to-2013-ons-data/measles-notifications-and-deaths-in-england-and-wales-1940-to-2013. [Accessed 19th October 2019].

Quadros, C. A. D. (2004). Can Measles Be Eradicated Globally? *Bulletin of The World Health Organization*, 82, 134-138.

Ravetz, J. and Warhurst, P. (2013). Manchester: Re-Inventing the Local–Global in The Peri-Urban City-Region, pp. 169-207 in Nilsson, K., Pauleit, S., Bell, S., Aalbers, C. and Nielsen, T. A. S. (Eds.). *Peri-Urban Futures: Scenarios and Models for Land Use Change in Europe*. Springer: Berlin.

Read, J. M., Eames, K. T., and Edmunds, W. J. (2008). Dynamic Social Networks and The Implications for The Spread of Infectious Disease. *Journal of The Royal Society Interface*, 5, 1001-1007.

Real, L. A. and Biek, R. (2007). Spatial Dynamics and Genetics of Infectious Diseases on Heterogeneous Landscapes. *Journal of The Royal Society Interface*, 4, 935-948.

Redcliffe-Maud, J. M. B. (1969). *Report of the Royal Commission on Local Government in England, 1966-1969.* H.S.M.O: London.

Redcliffe-Maud, Lord. and Wood, B. (1974). *English Local Government Reformed.* Oxford University Press: Oxford.

Rees, W. (1967). *An Historical Atlas of Wales from Early to Modern Times*. Faber & Faber: London.

Riley, S. (2007). Large-Scale Spatial-Transmission Models of Infectious Disease. *Science*, 316, 1298-1301.

Robert, S. and Randolph, W. G. (1983). Beyond Decentralization: The Evolution of Population Distribution in England and Wales, 1961–1981. *Geoforum*, 14, 75-102.

Rohani, P. and Drake, J. M. (2011). The Decline and Resurgence of Pertussis in The Us. *Epidemics*, 3, 183-188.

Rohani, P., Earn, D. J. and Grenfell, B. T. (1999). Opposite Patterns of Synchrony in Sympatric Disease Metapopulations. *Science*, 286, 968-971.

Rohani, P., Earn, D. J. and Grenfell, B. T. (2000). Impact of Immunisation on Pertussis Transmission in England and Wales. *The Lancet*, *355*, 285-286.

Rohani, P., Keeling, M. J. and Grenfell, B. T. (2002). The Interplay Between Determinism and Stochasticity in Childhood Diseases. *The American Naturalist*, 159, 469-481.

Rothstein, E. and Edwards, K. (2005). Health Burden of Pertussis in Adolescents and Adults. *The Pediatric Infectious Disease Journal*, *2*4, S44-S47.

Roush, S. W., Murphy, T. V. and Vaccine-Preventable Disease Table Working Group. (2007). Historical Comparisons of Morbidity and Mortality for Vaccine-Preventable Diseases in The United States. *JAMA*, 298, 2155-2163.

Row, B. W. and Squire, F. G. (1974). Cardiff 1889–1974: The Story of the County Borough. The Corporation of Cardiff: Cardiff.

Sartwell, P. E. (1966). The Incubation Period and The Dynamics of Infectious Disease. *American Journal of Epidemiology*, 83, 204-216.

Schenzle, D. (1984). An Age-Structured Model of Pre-and Post-Vaccination Measles Transmission. *Mathematical Medicine and Biology: A Journal of the IMA*, *1*, 169-191.

Schrödle, B., Held, L. and Rue, H. (2012). Assessing the Impact of a Movement Network on The Spatiotemporal Spread of Infectious Diseases. *Biometrics*, 68, 736-744.

Sedgwick, P. (2012). Hazards and Hazard Ratios. *BMJ*, 345.

Sedgwick, P. (2014). How to Read a Kaplan-Meier Survival Plot. *BMJ*, 349, G5608.

Sencer, D. J., Dull, H. B. and Langmuir, A. D. (1967). Epidemiologic Basis for Eradication of Measles in 1967. *Public Health Reports*, 82, 253-256.

Shelton, J. D., Jacobson, J. E., Orenstein, W. A., Schulz, K. F. and Donnell, H. D. (1978). Measles Vaccine Efficacy: Influence of Age at Vaccination Vs. Duration of Time Since Vaccination. *Pediatrics*, 62, 961-964.

Smallman-Raynor, M. (2006). *Poliomyelitis: Emergence to Eradication.* Oxford University Press on Demand.

Smallman-Raynor, M. R. and Cliff, A. D. (2015). Operation Pied Piper: A Geographical Reappraisal of the Impact of Wartime Evacuation on Scarlet Fever and Diphtheria Rates in England and Wales, 1939–1945. *Epidemiology & Infection*, 143, 2923-2938.

Smallman-Raynor, M., Nettleton, C. and Cliff, A. (2003). Wartime Evacuation and the Spread of Infectious Diseases: Epidemiological Consequences of the Dispersal of Children from London During World War II. *Journal of Historical Geography*, 29, 396-421.

Smith, A. F. and Roberts, G. O. (1993). Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of The Royal Statistical Society: Series B (Methodological)*, 55, 3-23.

Soper, H. E. (1929). The Interpretation of Periodicity in Disease Prevalence. *Journal of The Royal Statistical Society*, 92, 34-73.

Spruance, S. L., Reid, J. E., Grace, M. and Samore, M. (2004). Hazard Ratio in Clinical Trials. *Antimicrobial Agents and Chemotherapy*, 48, 2787-2792.

Ster, I. C. and Ferguson, N. M. (2007). Transmission Parameters of the 2001 Foot and Mouth Epidemic in Great Britain. *PloS One*, 2, e502.

Stobart, J. (1996). An Eighteenth-Century Revolution? Investigating Urban Growth in North-West England, 1664-1801. *Urban History*, 23, 26-47.

Stobart, J. (1996). Geography and Industrialization: The Space Economy of Northwest England, 1701-1760. *Transactions of The Institute of British Geographers*, 21, 681-696.

Stobart, J. (1998). Textile Industries in North-West England in the Early Eighteenth Century: A Geographical Approach. *Textile History*, 29, 3-18.

Stocks, P. (1941). Diphteria and Scarlet Fever Incidence During the Dispersal of 1939-40. *Journal of The Royal Statistical Society*, 104, 311-345.

Stocks, P. (1942). Measles and Whooping-Cough Incidence Before and During the Dispersal of 1939-41. *Journal of The Royal Statistical Society*, 105, 259-291.

Stocks, P. (1949). *Sickness in the population of England and Wales: in 1944-1947* (No. 2). H.M.S.O: London.

Stojanović, O., Leugering, J., Pipa, G., Ghozzi, S. and Ullrich, A. (2019). A Bayesian Monte Carlo Approach for Predicting the Spread of Infectious Diseases. *Plos One*, 14, E0225838.

Strebel, P. M., Papania, M. J., Fiebelkorn, A. P. and Halsey, N. A. (2012). Measles Vaccine. *Vaccines*, 6, 352-387.

Swedlund, A. C. and Donta, A. K. (2003). Scarlet Fever Epidemics of The Nineteenth Century: A Case of Evolved Pathogenic Virulence? Pp. 159-177 in Herring, D. A., & Swedlund, A. C. (Eds.). (2003). *Human Biologists in The Archives: Demography, Health, Nutrition and Genetics in Historical Populations*. Cambridge University Press: Cambridge.

Tanaka, M., Vitek, C. R., Pascual, F. B., Bisgard, K. M., Tate, J. E. and Murphy, T. V. (2003). Trends in Pertussis Among Infants in The United States, 1980-1999. *JAMA*, 290, 2968-2975.

Thompson, R. E. (1933). The Fylde. *Geography*, 18, 307-320.

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in The Detroit Region. *Economic Geography*, 46, 234-240.

Tolles, J. and Luong, T. (2020). Modeling Epidemics with Compartmental Models. *JAMA*. 23, 2515-2516.

Trevelyan, B., Smallman-Raynor, M. and Cliff, A. D. (2005). The Spatial Structure of Epidemic Emergence: Geographical Aspects of Poliomyelitis in North-Eastern USA, July–October 1916. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168, 701-722.

Trottier, H. and Philippe, P. (2001). Deterministic Modeling of Infectious Diseases: Theory and Methods. *The Internet Journal of Infectious Diseases*, *1*, 1528-8366.

Truscott, J., Fraser, C., Cauchemez, S., Meeyai, A., Hinsley, W., Donnelly, C.A., Ghani, A. and Ferguson, N. (2012). Essential Epidemiological Mechanisms Underpinning the Transmission Dynamics of Seasonal Influenza. *Journal of The Royal Society Interface*, 9, 304-312.

Turner, C.E., Pyzio, M., Song, B., Lamagni, T., Meltzer, M., Chow, J.Y., Efstratiou, A., Curtis, S. and Sriskandan, S. (2016). Scarlet Fever Upsurge in England and Molecular-Genetic Analysis in North-West London, 2014. *Emerging Infectious Diseases*, 22, 1075.

Van Bavel, J. and Reher, D. S. (2013). The Baby Boom and Its Causes: What We Know and What We Need to Know. *Population and Development Review*, 39, 257-288.

Viboud, C., Bjørnstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A. and Grenfell, B. T. (2006). Synchrony, Waves, and Spatial Hierarchies in The Spread of Influenza. *Science*, 312, 447-451.

Vining Jr, D. R. and Kontuly, T. (1978). Population Dispersal from Major Metropolitan Regions: An International Comparison. *International Regional Science Review*, 3, 49-73.

Von König, C. W., Halperin, S., Riffelmann, M. and Guiso, N. (2002). Pertussis of Adults and Infants. *The Lancet Infectious Diseases*, 2, 744-750.

Vyse, A. J., Gay, N. J., White, J. M., Ramsay, M. E., Brown, D. W. G., Cohen, B. J., Hesketh, L. M., Morgan-Capner, P. and Miller, E. (2002). Evolution of Surveillance of Measles, Mumps, and Rubella in England and Wales: Providing the Platform for Evidence-Based Vaccination Policy. *Epidemiologic Reviews*, 24, 125-136.

Walker, M. J. and Brouwer, S. (2018). Scarlet Fever Makes a Comeback. *The Lancet Infectious Diseases,* 18, 128-129.

Wannop, U. (1999). New towns, pp. 213-31 in Cullingworth, J. B. (Ed.). (1999). *British Planning: 50 Years of Urban and Regional Policy*. A&C Black: London.

Warnes, A. M. (1972). Estimates of Journey-To-Work Distances from Census Statistics. *Regional Studies*, 6, 315-326.

Wearing, H. J. and Rohani, P. (2009). Estimating the Duration of Pertussis Immunity Using Epidemiological Signatures. *Plos Pathog*, *5*, E1000647.

Webb, J. L. (2015). The Historical Epidemiology of Global Disease Challenges. *The Lancet*, 385, 322-323.

Wei, W. and Held, L. (2014). Calibration Tests for Count Data. *Test*, 23, 787-805.

Welsh Assembly Government (2008). *People, Places, Futures – The Wales Spatial Plan 2008 Update*. Welsh Assembly Government: Cardiff. Available from: http://wales.gov.uk/dpsp/wspatialplan/documents/wsp2008update/wsp2008updatee.pdf?lang=en. Accessed: 22nd September 2019.

Welsh Office (1967). *Wales: The Way Ahead*. H.M.S.O: London

Wendelboe, A. M. and Van Rie, A. (2006). Diagnosis of Pertussis: A Historical Review and Recent Developments. *Expert Review of Molecular Diagnostics*, 6, 857-864.

Wendelboe, A. M., Van Rie, A., Salmaso, S. and Englund, J. A. (2005). Duration of Immunity Against Pertussis After Natural Infection or Vaccination. *The Pediatric Infectious Disease Journal*, 24, S58-S61.

Williams, C. (2011). The Question of Monmouthshire, pp. 348–359 in Williams, C. and Rhiannon, S. (eds.), *The Gwent County History.* University of Wales Press: Cardiff.

Wong, S. S. and Yuen, K. Y. (2012). Streptococcus Pyogenes and Re-Emergence of Scarlet Fever as A Public Health Problem. *Emerging Microbes & Infections*, 1, e2.

Xia, Y., Bjørnstad, O. N. and Grenfell, B. T. (2004). Measles Metapopulation Dynamics: A Gravity Model for Epidemiological Coupling and Dynamics. *The American Naturalist*, 164, 267-281.

Yelling, J. (2000). The Incidence of Slum Clearance in England and Wales, 1955–85. *Urban History*, 27, 234-254.

Youngs, F. A. (1991) *Guide to the Local Administrative Units of England, Vol. II: Northern England*. Offices of the Royal Historical Society: London.

Zeger, S. L. and Karim, M. R. (1991). Generalized Linear Models with Random Effects: A Gibbs Sampling Approach. *Journal of The American Statistical Association*, 86, 79-86.

Zeger, S. L. and Qaqish, B. (1988). Markov Regression Models for Time Series: A Quasi-Likelihood Approach. *Biometrics*, 44, 1019-1031.

APPENDIX I

I. Additional Maps: Regional Administrative Geographies

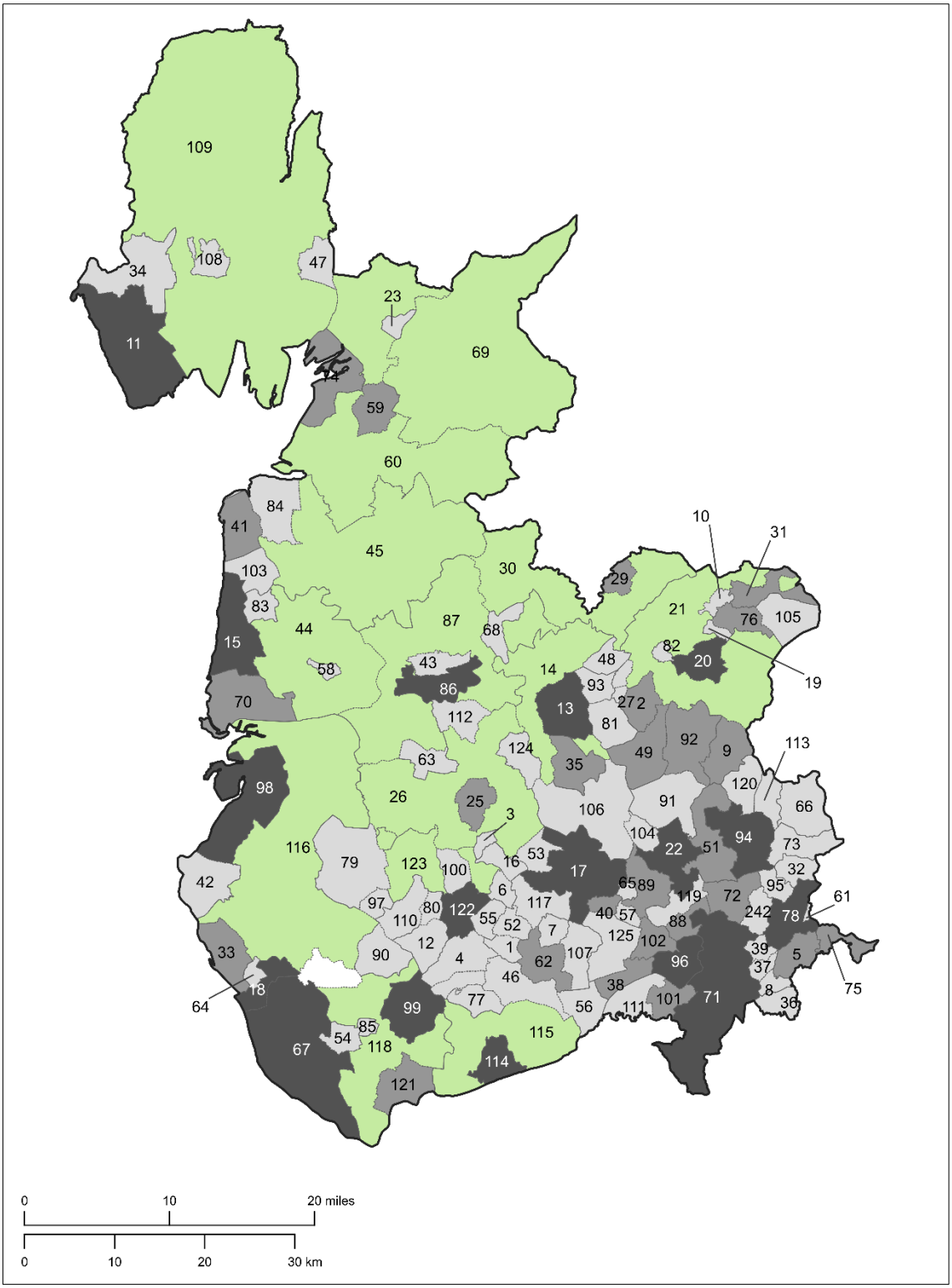

Figure I.1 *Administrative map of the Lancashire region, with local government districts (LGDs) numbered for identification purposes (See following page).*

List of Lancashire LGD numbers with corresponding name:

1. Abram UD
2. Accrington MB
3. Adlington UD
4. Ashton-in-Makerfield UD
5. Ashton-under-Lyne MB
6. Aspull UD
7. Atherton UD
8. Audenshaw UD
9. Bacup MB
10. Barrowford UD
11. Barrow-in-Furness CB
12. Billinge & Winstanley UD
13. Blackburn CB
14. Blackburn RD
15. Blackpool CB
16. Blackrod UD
17. Bolton CB
18. Bootle CB
19. Brierfield UD
20. Burnley CB
21. Burnley RD
22. Bury CB
23. Carnforth UD
24. Chadderton UD
25. Chorley MB
26. Chorley RD
27. Church UD
28. Clayton-le-Moors UD
29. Clitheroe MB
30. Clitheroe RD
31. Colne MB
32. Crompton UD
33. Crosby MB
34. Dalton-in-Furness UD
35. Darwen MB
36. Denton UD
37. Droylsden UD
38. Eccles MB
39. Failsworth UD
40. Farnworth MB
41. Fleetwood MB
42. Formby UD
43. Fulwood UD
44. Fylde RD
45. Garstang RD
46. Golborne UD
47. Grange UD
48. Great Harwood UD
49. Haslingden MB
50. Haydock UD
51. Heywood MB
52. Hindley UD
53. Horwich UD
54. Huyton-with-Roby UD
55. Ince-in-Makerfield UD
56. Irlam UD
57. Kearsley UD
58. Kirkham UD
59. Lancaster MB
60. Lancaster RD
61. Lees UD
62. Leigh MB
63. Leyland UD
64. Litherland UD
65. Little Lever UD
66. Littleborough UD
67. Liverpool CB
68. Longridge UD
69. Lunesdale RD
70. Lytham St Anne's MB
71. Manchester CB
72. Middleton MB
73. Milnrow UD
74. Morecambe & Heysham MB
75. Mossley MB
76. Nelson MB
77. Newton-le-Willows UD
78. Oldham CB
79. Ormskirk UD
80. Orrell UD
81. Oswaldtwistle UD
82. Padiham UD
83. Poulton-le-Fylde UD
84. Preesall UD
85. Prescot UD
86. Preston CB
87. Preston RD
88. Prestwich MB
89. Radcliffe MB
90. Rainford UD
91. Ramsbottom UD
92. Rawtenstall MB
93. Rishton UD
94. Rochdale CB
95. Royton UD
96. Salford CB
97. Skelmersdale UD
98. Southport CB
99. St. Helens CB
100. Standish-with-Langtree UD
101. Stretford MB
102. Swinton & Pendlebury MB
103. Thornton Cleveleys UD
104. Tottington UD
105. Trawden UD
106. Turton UD
107. Tyldesley UD
108. Ulverston UD
109. Ulverston/North Lonsdale RD
110. Upholland UD
111. Urmston UD
112. Walton-le-Dale UD
113. Wardle UD
114. Warrington CB
115. Warrington RD
116. West Lancashire RD
117. Westhoughton UD
118. Whiston RD
119. Whitefield UD
120. Whitworth UD
121. Widnes MB
122. Wigan CB
123. Wigan RD
124. Withnell UD
125. Worsley UD

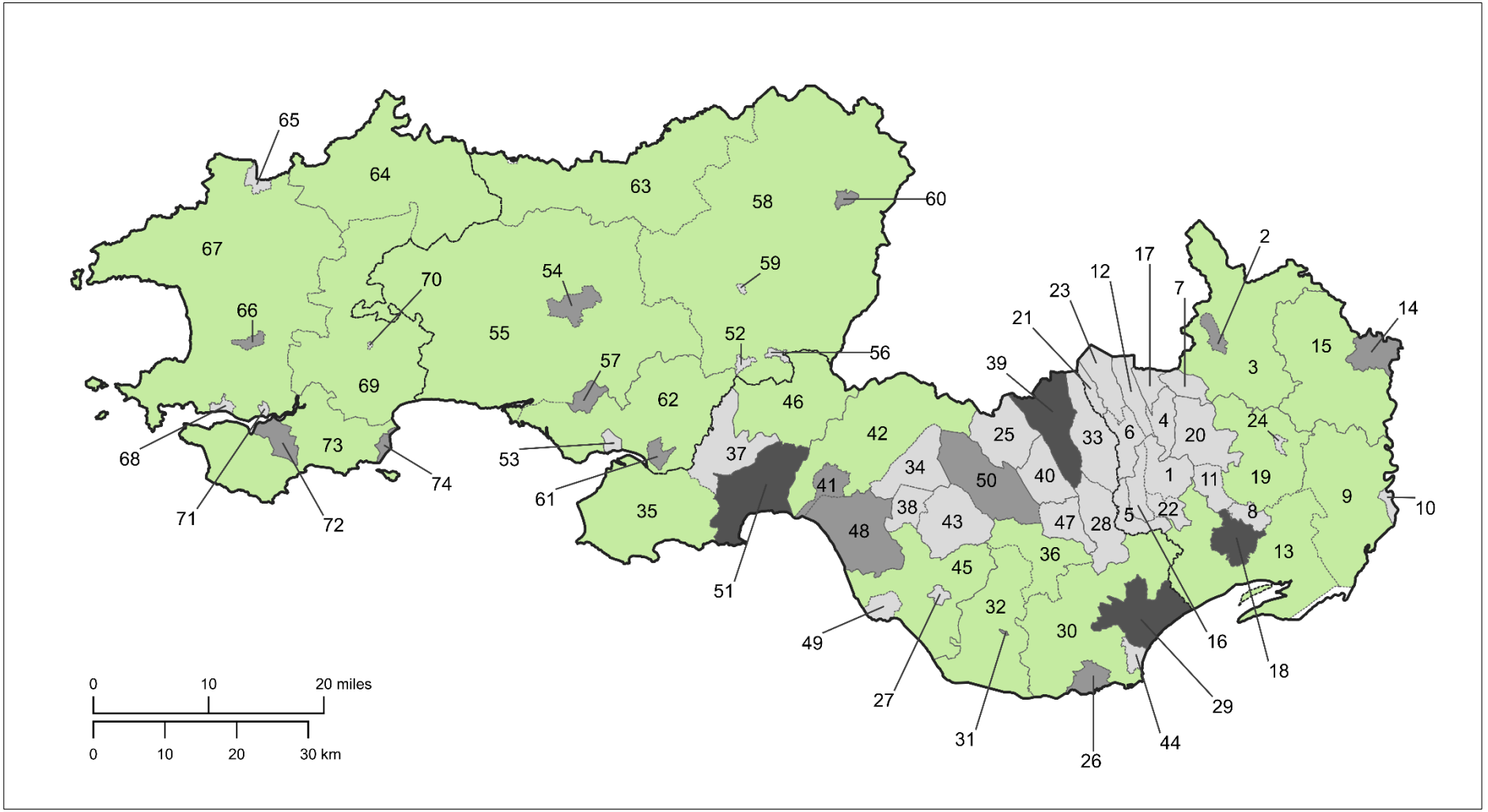

Figure I.2 *Administrative map of the South Wales region, with LGDs numbered for identification purposes (See following page).*

List of South Wales LGD numbers with corresponding name:

1. Abercarn UD
2. Abergavenny MB
3. Abergavenny RD
4. Abertillery UD
5. Bedwas & Machen UD
6. Bedwellty UD
7. Blaenavon UD
8. Caerleon UD
9. Chepstow RD
10. Chepstow UD
11. Ebbw Vale UD
12. Magor & St. Mellons RD
13. Monmouth MB
14. Monmouth RD
15. Mynyddislwyn UD
16. Nantyglo & Blaina UD
17. Newport CB
18. Pontypool RD
19. Pontypool UD
20. Rhymney UD
21. Risca UD
22. Tredegar UD
23. Usk UD
24. Aberdare UD
25. Barry MB
26. Bridgend UD
27. Caerphilly UD
28. Cardiff CB
29. Cardiff RD
30. Cowbridge MB
31. Cowbridge RD
32. Gelligaer UD
33. Glyncorrwg UD
34. Gower RD
35. Llantrisant & Llantwitfardre RD
36. Llwchwr UD
37. Maesteg UD
38. Merthyr Tydfil CB
39. Mountain Ash UD
40. Neath MB
41. Neath RD
42. Ogmore & Garw UD
43. Penarth UD
44. Penybont RD
45. Pontardawe RD
46. Pontypridd UD
47. Port Talbot MB
48. Porthcawl UD
49. Rhondda MB
50. Swansea CB
51. Ammanford UD
52. Burry Port UD
53. Carmarthen MB
54. Carmarthen RD
55. Cwmamman UD
56. Cwmbran UD
57. Kidwelly MB
58. Llandilo RD
59. Llandilo UD
60. Llandovery MB
61. Llanelli MB
62. Llanelli RD
63. Newcastle Emlyn RD
64. Cemaes RD
65. Fishguard & Goodwick UD
66. Haverfordwest MB
67. Haverfordwest RD
68. Milford Haven UD
69. Narberth RD
70. Narberth UD
71. Neyland UD
72. Pembroke MB
73. Pembroke RD
74. Tenby MB

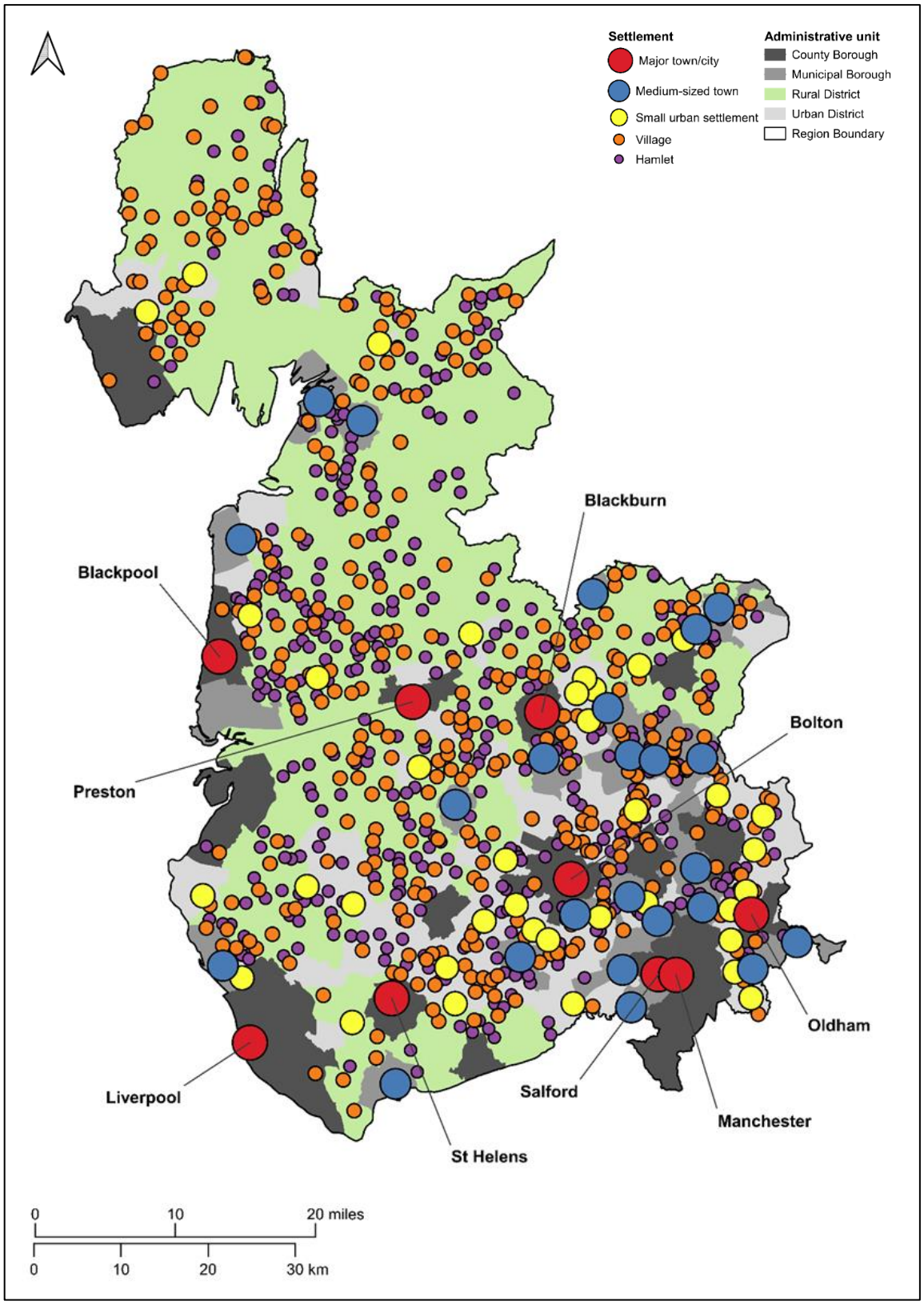

Figure I.3 *Population distribution by form of urban and rural settlement in the Lancashire region, 1940-1969. Settlements with mean population size greater than 100,000 inhabitants are labelled.*

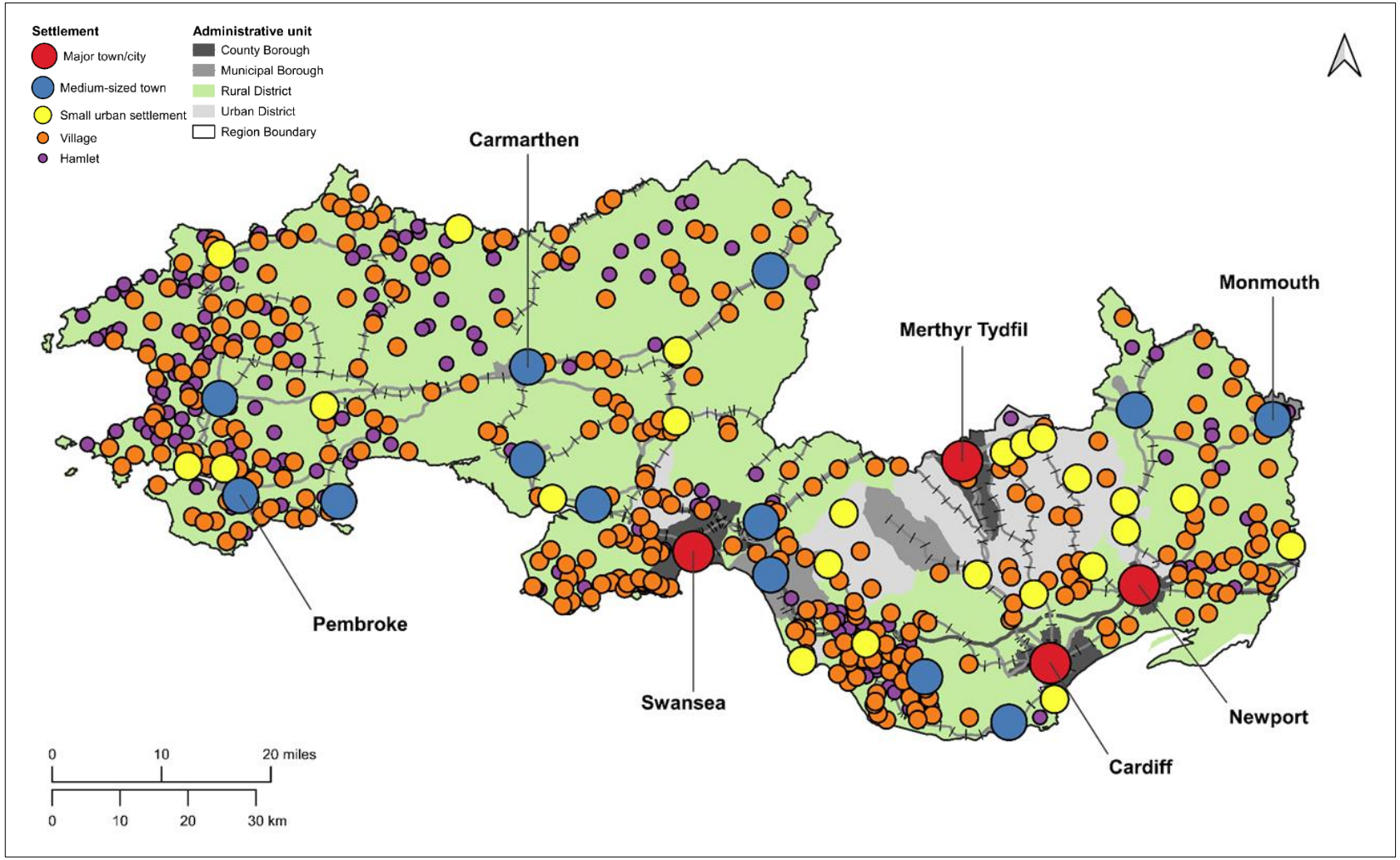

Figure I.4 *Population distribution by form of urban and rural settlement in the South Wales region, 1940-1969. Settlements with mean population size greater than 100,000 inhabitants are labelled.*

APPENDIX II

I. Problem of Using Percentage Data In Regression

There are some inherent issues using percentage data in regression. For linear models, the dependent variable does not need to be normally distributed, but it must possess the following properties; be continuous, unbounded, and measured on an interval or ratio scale. Percentage data does not fit these criteria, despite being continuous and measured on a ratio scale. The issue is the boundaries capped at 0 and 100. Proportional data is, by its nature, bound at 0 and 1, and is therefore often not normally distributed or homoscedastic. When one has a bounded dependent variable but the model being fitted is unbounded, such as an OLS model, this may produce significant problems when describing the mean. Some fitted and predicted values may be impossible, with values below 0 or above 1. The true relationship must eventually become flatter than it is at the middle as it approaches the bounds, resulting in a bend of some form. There may also be issues when describing the variance; as the mean approaches the bound, the variance will tend to decrease as well. With less room between the mean and bounds, the overall variability tends to fall.

Percentage endemicity is derived from discrete counts of “successes” and “failures”, or in this instance “absent” and “present”, with the number of months in which disease is “present” divided by the total number of months. Data of this kind is perhaps better suited to analysis by logistic or beta regression. If the proportion data is for a count variable, a common model for the distribution of the proportion is a binomial GLM. There are several options for the form of the relationship of the mean proportion and the predictors, but the most common one would be a logistic GLM. Beta regression can be conducted with the *betareg* function in the R package `betareg` (Cribari-Neto and Zeileis, 2010). With this function, the dependent variable varies between 0 and 1, but no observation can equal exactly 0 or 1. The model assumes that the data follows a beta distribution.

II. Two-Component Model for Disease Counts

Let $Y = (Y_1, \ldots, Y_n)$ denote the time series of weekly counts of a specific infectious disease. In its most basic form, without temporal or seasonal trends, the model is specified to let previous counts act directly on the conditional mean μ_t of $Y_t|Y_{t-1}$, not the log mean). An identity link is used, rather than a log link

$$\mu_t = \nu + \lambda Y_{t-1}, \tag{1.1}$$

where ν is the endemic parameter and λ is the autoregressive parameter. Model 1.1 would not be appropriate in most cases when applied to infectious disease data since it does not allow for seasonal patterns and temporal trends. Thus, ν is replaced with a time-varying ν_t. The autoregressive parameter λ is also allowed to vary over time, to reflect situations where infectiousness of a disease changes over time for instance, perhaps due to the implementation of immunisation programmes or other public health interventions, or through external factors which influence the spread of infection. Other scenarios include a declining number of susceptibles over time, which would effectively decrease λ, and sudden outbreaks where $\lambda_t > 1$ for a limited time period to be estimated from infectious disease data.

Assume that Y_t follows a Bienaymé–Galton–Watson (BGW) Poisson branching process with immigration and time-varying parameters ν_t and λ_t:

$$Y_t = X_t + Z_t, t = 1, \ldots, n \text{ with}$$

$$X_t \sim Po(\nu_t), \text{and}$$

$$Z_t | Y_t - 1 \sim Po(\lambda t\, Y_{t-1})$$

The observed number of counts Y_t is now separated into two (unknown) components: X_t and Z_t, with conditional rate $\lambda Y_t - 1$, and it is assumed that they are independent. These

two components are termed the ‘endemic’ and ‘epidemic’ components, respectively (Held et al., 2005, 2006). The endemic component X_t parametrically models seasonal variation and trends in endemic incidence and should exhibit a persistent, stable temporal pattern. With regards to the epidemic component Z_t, occasional outbreaks in epidemic incidence should be captured and eventually fade out (provided $\lambda < 1$).

For constant ν and λ, Y_t is a simple branching process with immigration. Here, $\nu > 0$ and $0 < \lambda < 1$, the number of counts (Z_t) is stationary with mean and variance

$$\mu = \frac{\nu}{1-\lambda}, \qquad \sigma^2 = \frac{\nu}{(1-\lambda)(1-\lambda^2)}. \tag{1.2}$$

The stationary mean $\mu = \nu/(1 - \lambda)$ enables an effective interpretation of λ. For endemic incidence, the stationary mean is ν, whilst epidemic incidence has a stationary mean $\nu/(1 - \lambda) - \nu = \lambda\nu/(1 - \lambda)$, therefore the ratio of epidemic to total stationary mean rate is equal to the autoregressive parameter λ.

Held et al. (2006) extend the model to allow seasonal terms in the endemic rate, and to enable the inclusion of an additional influx of endemic cases with rate ν to ensure that whenever $\lambda \geq 1$ an outbreak will occur, while for $\lambda < 1$ the process will be stable and will not fade-out completely with probability 1 (depending on the actual value of λ). This would not be the case according to an ordinary branching process. This is also advantageous as infectious disease surveillance data tends to exhibit a mixture of endemic and epidemic behaviours given the dynamical nature of disease systems. When λ is close to 1, simulations show the two component model displays occasional epidemic outbreaks, indicating the formulation provides a more realistic fit than a solely parameter-driven model (Held et al., 2005, 2006)

The model (1.1) does not fit a generalized linear model (GLM) framework but numerical techniques for likelihood inference can be used. Maximum likelihood estimates can be simply obtained using generic optimisation routines so more complicated simulation-based inferences such as Markov chain Monte-Carlo (MCMC) algorithms are not required.

In most applications of the model outlined, it will be necessary to replace the Poisson distribution with a more flexible observation model to adjust for overdispersion resultant from a variety of factors: underreporting or reporting delays of infectious disease counts, the influence of unobserved covariates and mechanisms which affect disease incidence, spatial autocorrelation, amongst others. Overdispersion can be allowed for by replacing the Poisson distribution with a negative binomial model, where the conditional mean μ_t remains the same but the conditional variance σ_t^2 increases to

$$\sigma_t^2 = \mu_t + \frac{\mu_t^2}{\psi} = \mu_t \left(1 + \frac{\mu_t}{\psi}\right), \tag{1.3}$$

where ψ is the overdispersion parameter. $\psi > 0$ is estimated from the available data.

Table II.A *Glossary of notation for parameters in HHH model formulations.*

Notation	Parameter
$\boldsymbol{\beta}^{(\boldsymbol{\phi})}_{\log(\boldsymbol{pop})}$	Population in the spatio-temporal component ('Gravity' parameter)
$\boldsymbol{d}$	Distance decay
$\boldsymbol{\alpha}^{(\boldsymbol{\lambda})}$	Fixed effect, epidemic component
$\boldsymbol{\alpha}^{(\boldsymbol{\phi})}$	Fixed effect, spatio-temporal component
$\boldsymbol{\alpha}^{(\boldsymbol{v})}$	Fixed effect, endemic component
$\boldsymbol{\psi}$	Dispersion
$\boldsymbol{\sigma}^{\mathbf{2}}_{\boldsymbol{\lambda}}$	Variance, epidemic random effect
$\boldsymbol{\sigma}^{\mathbf{2}}_{\boldsymbol{\phi}}$	Variance, spatio-temporal random effect
$\boldsymbol{\sigma}^{\mathbf{2}}_{\mathbf{v}}$	Variance, endemic random effect
$\boldsymbol{maxEV}$	Maximum Eigenvalues
$\boldsymbol{\ell}_{\boldsymbol{pen}}$	Penalised log-likelihood
$\boldsymbol{\ell}_{\boldsymbol{marg}}$	Marginal log-likelihood
$\boldsymbol{\ell}$	Log-likelihood

APPENDIX III

I. Endemic Threshold Population Estimates: Measles, 1940–1969

Table III.1 Parameter estimates, measures of dispersion and endemic threshold estimates for measles with 95% CI for nine Lancashire and South Wales time windows (all districts).

Table III.2 Parameter estimates, measures of dispersion and endemic threshold estimates for measles with 95% CI for low density and high density districts in Lancashire, for nine time windows.

Table III.3 Parameter estimates, measures of dispersion and endemic threshold estimates for measles with 95% CI for low density and high density districts in South Wales, for nine time windows.

Table III.4 Parameter estimates, measures of dispersion and endemic threshold estimates for measles with 95% CI for low connectivity and high connectivity districts in Lancashire, for nine time windows.

Table III.5 Parameter estimates, measures of dispersion and endemic threshold estimates for measles with 95% CI for low connectivity and high connectivity districts in South Wales, for nine time windows.

Table III.1.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Lancashire							
1940-45	-140.251	44.335	-1.981	0.80	**263**	215	322
1943-48	-134.642	42.075	-2.319	0.73	**378**	295	484
1946-51	-140.398	43.788	-2.246	0.76	**301**	246	390
1949-54	-148.471	45.793	-1.911	0.83	**267**	221	322
1952-57	-139.263	43.312	-1.689	0.84	**335**	281	399
1955-60	-140.645	43.237	-1.611	0.86	**368**	311	436
1958-63	-149.694	45.788	-1.750	0.85	**284**	239	338
1961-66	-151.636	47.308	-1.949	0.83	**209**	174	252
1964-69	-165.784	49.592	-2.018	0.83	**229**	191	275
South Wales							
1940-45	-110.126	34.622	-2.617	0.70	**1,180**	834	1,650
1943-48	-104.077	32.632	-2.463	0.71	**1,795**	1,277	2,523
1946-51	-118.435	36.918	-2.253	0.79	**826**	627	1,088
1949-54	-117.257	35.826	-2.136	0.79	**1,160**	886	1,518
1952-57	-110.099	33.682	-2.377	0.73	**1,730**	1,258	2,378
1955-60	-108.942	33.678	-2.36	0.74	**1,600**	1,167	2,196
1958-63	-102.857	31.824	-2.315	0.72	**2,370**	1,706	3,288
1961-66	-120.000	37.090	-2.733	0.72	**855**	613	1,192
1964-69	-126.170	37.632	-2.868	0.70	**1,025**	726	1,444

Table III.2.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Low density							
1940-45	-113.358	37.314	-3.650	0.63	523	336	813
1943-48	-89.737	30.347	-4.269	0.45	1,788	948	3,375
1946-51	-107.766	35.162	-4.136	0.54	811	477	1,379
1949-54	-132.134	41.473	-3.668	0.68	396	266	590
1952-57	-129.722	40.887	-3.220	0.72	416	292	593
1955-60	-117.174	37.206	-2.797	0.74	688	490	965
1958-63	-138.840	42.903	-3.107	0.76	369	267	512
1961-66	-150.168	46.795	-3.901	0.70	222	153	324
1964-69	-153.220	46.164	-3.570	0.73	306	216	434
High density							
1940-45	-143.505	45.335	-2.968	0.80	236	175	316
1943-48	-142.421	44.301	-3.158	0.77	297	215	410
1946-51	-138.999	43.953	-3.159	0.76	275	199	379
1949-54	-146.218	45.534	-2.795	0.82	256	194	338
1952-57	-143.556	44.315	-2.619	0.83	314	240	410
1955-60	-145.748	44.538	-2.716	0.82	330	251	434
1958-63	-146.650	45.294	-2.829	0.81	279	211	370
1961-66	-144.160	45.789	-2.708	0.83	215	165	281
1964-69	-156.152	47.789	-3.055	0.80	230	172	306

Table III.3.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Low connectivity							
1940-45	-132.681	41.729	-2.725	0.81	**377**	281	506
1943-48	-117.772	37.086	-3.078	0.77	**745**	513	1,084
1946-51	-120.705	38.160	-3.413	0.76	**608**	406	910
1949-54	-134.140	41.575	-3.176	0.79	**430**	304	605
1952-57	-132.877	41.478	-2.708	0.82	**412**	307	553
1955-60	-127.646	39.801	-2.654	0.83	**525**	389	709
1958-63	-133.052	41.350	-3.006	0.82	**433**	312	601
1961-66	-137.316	43.401	-2.979	0.82	**294**	216	401
1964-69	-145.826	44.114	-3.175	0.82	**374**	271	518
High connectivity							
1940-45	-134.239	43.717	-2.677	0.80	**229**	174	301
1943-48	-137.810	43.708	-3.199	0.68	**276**	199	385
1946-51	-146.873	46.150	-2.734	0.76	**224**	172	293
1949-54	-149.670	46.838	-2.015	0.87	**215**	177	260
1952-57	-140.575	43.911	-2.214	0.85	**302**	240	379
1955-60	-149.548	45.581	-1.984	0.87	**299**	246	364
1958-63	-159.838	48.547	-1.926	0.88	**226**	189	270
1961-66	-157.508	49.124	-2.591	0.82	**175**	138	222
1964-69	-171.617	51.628	-2.403	0.84	**183**	148	226

Table III.4.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Low density							
1940-45	-108.748	33.488	-3.906	0.67	**1,713**	1,012	2,899
1943-48	-100.528	31.055	-3.952	0.63	**2,866**	1,614	5,090
1946-51	-124.765	38.117	-3.619	0.75	**789**	514	1,211
1949-54	-99.394	31.414	-3.502	0.69	**2,225**	1,346	3,680
1952-57	-96.978	30.224	-3.795	0.63	**3,291**	1,868	5,800
1955-60	-99.005	31.019	-3.742	0.65	**2,605**	1,511	4,488
1958-63	-95.118	29.922	-3.806	0.63	**3,320**	1,869	5,892
1961-66	-118.245	36.817	-4.788	0.62	**847**	471	1,524
1964-69	-103.895	31.820	-4.838	0.54	**2,560**	1,288	5,090
High density							
1940-45	-104.309	34.006	-3.466	0.73	**1,020**	644	1,614
1943-48	-99.800	32.294	-3.172	0.74	**1,538**	988	2,396
1946-51	-113.070	35.981	-3.036	0.80	**836**	571	1,223
1949-54	-129.071	38.624	-2.744	0.85	**853**	619	1,176
1952-57	-116.179	35.402	-3.098	0.78	**1,278**	861	1,897
1955-60	-113.133	34.926	-3.129	0.77	**1,266**	845	1,897
1958-63	-107.109	32.868	-3.007	0.77	**2,010**	1,325	3,024
1961-66	-121.692	37.346	-3.323	0.78	**864**	578	1,290
1964-69	-136.297	40.473	-3.472	0.79	**690**	468	1,016

Table III.5.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Low connectivity							
1940-45	-90.810	29.416	-3.890	0.73	**3,070**	1,689	5,580
1943-48	-81.686	26.537	-3.560	0.74	**7,030**	3,835	12,870
1946-51	-101.268	32.166	-3.340	0.79	**1,810**	1,132	2,889
1949-54	-104.350	32.185	-3.100	0.80	**2,240**	1,447	3,452
1952-57	-89.017	27.570	-3.400	0.75	**7,180**	4,116	12,515
1955-60	-91.728	28.769	-3.840	0.70	**4,620**	2,530	8,430
1958-63	-95.335	29.840	-3.800	0.68	**3,520**	1,979	6,250
1961-66	-108.260	33.888	-4.270	0.70	**1,400**	792	2,470
1964-69	-101.850	31.047	-4.520	0.71	**3,180**	1,646	6,117
High connectivity							
1940-45	-124.238	38.228	-3.849	0.62	**735**	467	1,157
1943-48	-119.429	36.617	-3.624	0.59	**983**	629	1,537
1946-51	-127.926	39.516	-3.296	0.72	**587**	403	854
1949-54	-121.318	37.146	-3.200	0.74	**908**	616	1,340
1952-57	-112.353	34.982	-3.205	0.67	**1,176**	778	1,778
1955-60	-114.704	35.561	-3.036	0.71	**1,091**	742	1,604
1958-63	-110.115	33.592	-3.269	0.70	**1,800**	1,160	2,792
1961-66	-128.272	39.236	-4.058	0.66	**658**	413	1,049
1964-69	-144.717	42.361	-4.001	0.62	**599**	391	917

II. Endemic Threshold Population Estimates: Pertussis, 1940–1969

Table III.6 Parameter estimates, measures of dispersion and endemic threshold estimates for pertussis with 95% CI for nine Lancashire and South Wales time windows (all districts).

Table III.7 Parameter estimates, measures of dispersion and endemic threshold estimates for pertussis with 95% CI for low density and high density districts in Lancashire, for nine time windows.

Table III.8 Parameter estimates, measures of dispersion and endemic threshold estimates for pertussis with 95% CI for low density and high density districts in South Wales, for nine time windows.

Table III.9 Parameter estimates, measures of dispersion and endemic threshold estimates for pertussis with 95% CI for low connectivity and high connectivity districts in Lancashire, for nine time windows.

Table III.10 Parameter estimates, measures of dispersion and endemic threshold estimates for pertussis with 95% CI for low connectivity and high connectivity districts in South Wales, for nine time windows.

Table III.6.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Lancashire							
1940-45	-1.515	0.690	0.046	0.64	125	92	169
1943-48	-1.379	0.662	0.047	0.62	130	94	175
1946-51	-1.208	0.634	0.037	0.70	115	89	150
1949-54	-1.136	0.623	0.031	0.76	109	87	136
1952-57	-1.328	0.651	0.031	0.78	130	105	161
1955-60	-1.915	0.750	0.037	0.77	166	133	208
1958-63	-2.448	0.816	0.045	0.73	283	221	363
1961-66	-2.688	0.834	0.056	0.64	418	308	568
1964-69	-2.525	0.799	0.058	0.60	461	332	641
South Wales							
1940-45	-1.564	0.652	0.076	0.50	293	173	498
1943-48	-1.598	0.662	0.068	0.56	273	171	435
1946-51	-1.708	0.708	0.055	0.69	173	122	245
1949-54	-1.741	0.722	0.057	0.69	152	107	217
1952-57	-2.099	0.787	0.056	0.73	162	118	223
1955-60	-2.196	0.763	0.060	0.69	316	222	450
1958-63	-2.109	0.681	0.059	0.64	1,082	732	1,603
1961-66	-2.060	0.655	0.063	0.59	1,580	1,024	2,437
1964-69	-2.078	0.654	0.073	0.52	1,720	1,114	2,843

Table III.7.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Low density							
1940-45	-1.275	0.623	0.089	0.44	157	80	307
1943-48	-1.030	0.564	0.092	0.38	98	53	181
1946-51	-1.080	0.593	0.077	0.49	128	76	215
1949-54	-1.378	0.677	0.063	0.65	227	151	339
1952-57	-1.426	0.671	0.065	0.64	579	378	880
1955-60	-1.748	0.700	0.086	0.52	1,025	577	1,806
1958-63	-1.975	0.690	0.100	0.44	1,593	766	3,313
1961-66	-2.092	0.681	0.120	0.34	1,021	462	2,255
1964-69	-1.808	0.614	0.117	0.30	1,593	672	3,774
High density							
1940-45	-1.291	0.647	0.068	0.60	123	76	197
1943-48	-1.104	0.610	0.061	0.62	123	78	193
1946-51	-0.856	0.564	0.044	0.73	116	82	166
1949-54	-0.690	0.528	0.042	0.73	125	88	178
1952-57	-0.947	0.571	0.041	0.76	146	105	202
1955-60	-1.516	0.669	0.042	0.81	181	137	239
1958-63	-2.317	0.795	0.051	0.80	270	202	360
1961-66	-2.809	0.866	0.069	0.73	358	250	512
1964-69	-2.638	0.832	0.071	0.70	376	256	552

Table III.8.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Low density							
1940-45	-1.433	0.608	0.133	0.36	443	166	1188
1943-48	-1.838	0.707	0.129	0.45	269	118	611
1946-51	-1.867	0.747	0.110	0.56	151	78	292
1949-54	-1.55	0.684	0.091	0.61	155	86	282
1952-57	-2.14	0.801	0.086	0.70	148	91	240
1955-60	-2.299	0.789	0.085	0.70	281	173	458
1958-63	-1.937	0.639	0.080	0.64	1,450	826	2,546
1961-66	-1.932	0.626	0.091	0.56	1,911	993	3,678
1964-69	-2.236	0.679	0.119	0.47	1,733	784	3,828
High density							
1940-45	-1.519	0.653	0.097	0.55	245	127	480
1943-48	-1.385	0.625	0.080	0.62	261	147	471
1946-51	-1.648	0.693	0.063	0.77	184	122	278
1949-54	-1.981	0.77	0.075	0.74	148	96	230
1952-57	-2.097	0.783	0.076	0.75	171	111	265
1955-60	-2.143	0.75	0.086	0.68	335	200	561
1958-63	-2.204	0.704	0.088	0.64	937	533	1,648
1961-66	-2.142	0.673	0.090	0.60	1,428	781	2,610
1964-69	-1.923	0.629	0.093	0.56	1,726	888	3,356

Table III.9.

Time Window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						lower	upper
Low connectivity							
1940-45	-1.64	0.708	0.057	0.65	139	97	199
1943-48	-1.441	0.662	0.062	0.59	158	104	241
1946-51	-1.302	0.644	0.046	0.70	135	98	186
1949-54	-1.232	0.635	0.043	0.73	123	91	167
1952-57	-1.297	0.634	0.045	0.76	159	115	219
1955-60	-1.732	0.696	0.049	0.80	231	168	317
1958-63	-1.927	0.681	0.056	0.82	585	404	845
1961-66	-1.974	0.654	0.074	0.75	1,193	714	1,993
1964-69	-1.754	0.603	0.076	0.73	1,681	950	2,976
High connectivity							
1940-45	-1.062	0.599	0.071	0.62	130	77	221
1943-48	-0.953	0.578	0.067	0.61	129	76	218
1946-51	-0.761	0.545	0.055	0.69	117	75	184
1949-54	-0.75	0.546	0.043	0.78	109	77	156
1952-57	-1.126	0.615	0.042	0.78	122	89	166
1955-60	-1.885	0.754	0.055	0.72	143	103	198
1958-63	-2.853	0.919	0.068	0.61	191	137	267
1961-66	-3.376	0.999	0.083	0.50	241	166	350
1964-69	-3.246	0.973	0.085	0.44	247	167	366

Table III.10.

Time Window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						lower	upper
Low connectivity							
1940-45	-1.059	0.518	0.106	0.58	805	321	2,020
1943-48	-1.094	0.53	0.094	0.65	689	317	1,532
1946-51	-1.68	0.694	0.091	0.60	201	111	364
1949-54	-1.872	0.753	0.101	0.55	139	76	254
1952-57	-2.446	0.872	0.084	0.64	126	82	195
1955-60	-2.438	0.821	0.079	0.67	255	165	394
1958-63	-2.104	0.676	0.099	0.55	1,178	609	2,280
1961-66	-1.852	0.594	0.105	0.52	3,054	1,372	6,802
1964-69	-2.059	0.638	0.120	0.39	2,305	987	5,373
High connectivity							
1940-45	-1.97	0.754	0.120	0.33	185	90	379
1943-48	-2.078	0.779	0.108	0.39	172	92	323
1946-51	-1.561	0.68	0.075	0.70	173	105	284
1949-54	-1.521	0.674	0.069	0.77	168	106	267
1952-57	-1.639	0.683	0.081	0.76	213	125	366
1955-60	-1.808	0.676	0.098	0.66	430	224	828
1958-63	-1.991	0.658	0.083	0.64	1,163	658	2,055
1961-66	-2.089	0.669	0.087	0.56	1,295	719	2,332
1964-69	-1.735	0.585	0.103	0.51	2,425	1,097	5,359

III. Endemic Threshold Population Estimates: Scarlet fever, 1940–1969

Table III.11 Parameter estimates, measures of dispersion and endemic threshold estimates for scarlet fever with 95% CI for nine Lancashire and South Wales time windows (all districts).

Table III.12 Parameter estimates, measures of dispersion and endemic threshold estimates for scarlet fever with 95% CI for low density and high density districts in Lancashire, for nine time windows.

Table III.13 Parameter estimates, measures of dispersion and endemic threshold estimates for scarlet fever with 95% CI for low density and high density districts in South Wales, for nine time windows.

Table III.14 Parameter estimates, measures of dispersion and endemic threshold estimates for scarlet fever with 95% CI for low connectivity and high connectivity districts in Lancashire, for nine time windows.

Table III.15 Parameter estimates, measures of dispersion and endemic threshold estimates for scarlet fever with 95% CI for low connectivity and high connectivity districts in South Wales, for nine time windows.

Table III.11.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Lancashire							
1940-45	-1.288	0.678	0.034	0.77	71	57	89
1943-48	-1.365	0.690	-0.033	0.79	76	61	94
1946-51	-1.657	0.739	-0.036	0.78	89	72	111
1949-54	-1.797	0.768	-0.036	0.79	88	72	109
1952-57	-2.102	0.811	-0.041	0.76	115	91	144
1955-60	-2.351	0.831	-0.048	0.71	173	133	224
1958-63	-2.550	0.869	-0.056	0.67	173	129	231
1961-66	-2.964	0.935	-0.066	0.62	204	149	281
1964-69	-2.863	0.913	-0.066	0.61	213	154	294
South Wales							
1940-45	-1.580	0.720	-0.045	0.78	94	71	125
1943-48	-1.809	0.774	-0.054	0.74	84	61	115
1946-51	-2.134	0.835	-0.054	0.77	90	67	120
1949-54	-2.118	0.806	-0.060	0.71	130	92	180
1952-57	-2.425	0.852	-0.060	0.74	157	114	215
1955-60	-2.590	0.869	-0.056	0.77	192	144	256
1958-63	-2.335	0.766	-0.069	0.62	457	304	686
1961-66	-2.393	0.751	-0.079	0.55	708	440	1,137
1964-69	-2.533	0.781	-0.078	0.58	637	406	1,000

Table III.12.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Low density							
1940-45	-0.435	0.492	-0.034	0.78	89	66	122
1943-48	-0.776	0.561	-0.043	0.73	89	63	126
1946-51	-1.154	0.630	-0.045	0.76	102	74	141
1949-54	-1.051	0.605	-0.040	0.79	111	82	149
1952-57	-1.322	0.644	-0.055	0.69	145	98	212
1955-60	-1.804	0.716	-0.063	0.68	206	139	306
1958-63	-2.007	0.757	-0.068	0.67	197	131	295
1961-66	-2.311	0.801	-0.089	0.57	242	146	398
1964-69	-1.979	0.728	-0.092	0.51	293	166	517
High density							
1940-45	-2.136	0.884	0.068	0.73	48	34	68
1943-48	-1.886	0.815	-0.061	0.75	59	42	83
1946-51	-2.042	0.830	-0.072	0.68	75	51	110
1949-54	-2.530	0.945	-0.075	0.72	63	44	89
1952-57	-2.531	0.910	-0.080	0.68	96	65	142
1955-60	-2.550	0.874	-0.106	0.52	161	93	278
1958-63	-2.611	0.876	-0.120	0.46	184	99	341
1961-66	-2.950	0.920	-0.130	0.45	241	127	455
1964-69	-3.085	0.956	0.123	0.49	209	117	373

Table III.13.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Low density							
1940-45	-2.272	0.884	-0.082	0.76	69	45	104
1943-48	-2.800	1.011	-0.093	0.76	56	37	85
1946-51	-2.958	1.031	-0.086	0.80	65	45	94
1949-54	-1.887	0.749	-0.102	0.60	155	84	287
1952-57	-2.065	0.766	-0.097	0.63	203	115	360
1955-60	-2.503	0.848	-0.085	0.73	205	130	322
1958-63	-2.124	0.715	-0.101	0.58	587	310	1,110
1961-66	-1.999	0.654	-0.123	0.43	1,303	558	3,050
1964-69	-2.381	0.727	-0.139	0.42	1,070	449	2,517
High density							
1940-45	-1.244	0.646	-0.052	0.81	106	74	152
1943-48	-1.353	0.671	-0.062	0.76	100	66	152
1946-51	-1.756	0.751	-0.069	0.77	101	67	152
1949-54	-2.270	0.842	-0.076	0.77	118	79	178
1952-57	-2.644	0.903	-0.077	0.79	139	95	205
1955-60	-2.641	0.881	-0.077	0.78	186	125	275
1958-63	-2.455	0.795	-0.099	0.64	403	229	705
1961-66	-2.615	0.805	-0.107	0.61	541	297	985
1964-69	-2.528	0.796	-0.089	0.69	489	295	809

Table III.14.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Low connectivity							
1940-45	-1.455	0.710	-0.04	0.77	74	57	97
1943-48	-1.532	0.720	-0.041	0.78	81	63	105
1946-51	-1.771	0.756	-0.043	0.80	98	76	126
1949-54	-1.834	0.769	-0.053	0.76	97	71	133
1952-57	-2.027	0.783	-0.052	0.80	140	104	188
1955-60	-2.245	0.796	-0.059	0.76	216	156	301
1958-63	-2.475	0.846	-0.076	0.69	195	131	293
1961-66	-2.895	0.905	-0.080	0.67	256	173	383
1964-69	-2.862	0.899	-0.090	0.59	257	163	403
High connectivity							
1940-45	-0.880	0.592	-0.051	0.76	74	50	109
1943-48	-0.897	0.592	-0.048	0.79	79	55	114
1946-51	-1.238	0.653	-0.055	0.75	91	63	133
1949-54	-1.584	0.726	-0.052	0.78	87	63	120
1952-57	-1.978	0.792	-0.064	0.71	106	74	152
1955-60	-2.227	0.813	-0.077	0.63	159	104	243
1958-63	-2.498	0.863	-0.086	0.61	163	105	256
1961-66	-2.676	0.884	-0.105	0.54	195	115	333
1964-69	-2.457	0.835	-0.100	0.56	218	127	374

Table III.15.

Time window	α	*b*	S.E.	R^2	Threshold estimate (in 000s)	95% CI	
						Lower	Upper
Low connectivity							
1940-45	-1.248	0.628	-0.054	0.85	149	101	220
1943-48	-1.459	0.675	-0.061	0.83	134	89	201
1946-51	-1.742	0.722	-0.062	0.84	153	104	225
1949-54	-1.857	0.724	-0.066	0.80	213	141	321
1952-57	-2.199	0.784	-0.088	0.72	227	137	377
1955-60	-2.488	0.829	-0.092	0.68	260	158	428
1958-63	-2.329	0.757	-0.115	0.52	524	265	1,040
1961-66	-2.169	0.686	-0.138	0.46	1,195	485	2,965
1964-69	-2.423	0.738	-0.138	0.42	985	424	2,290
High connectivity							
1940-45	-1.757	0.769	-0.076	0.65	77	50	121
1943-48	-1.944	0.815	-0.092	0.59	70	42	115
1946-51	-2.223	0.869	-0.090	0.64	73	46	116
1949-54	-1.974	0.788	-0.102	0.58	111	62	199
1952-57	-2.376	0.852	-0.089	0.68	137	86	220
1955-60	-2.329	0.822	-0.072	0.79	185	125	275
1958-63	-2.115	0.722	-0.097	0.63	501	273	919
1961-66	-2.438	0.769	-0.106	0.53	591	317	1,100
1964-69	-2.187	0.715	-0.100	0.60	718	382	1,350

IV. Additional Endemic Threshold Size Plots

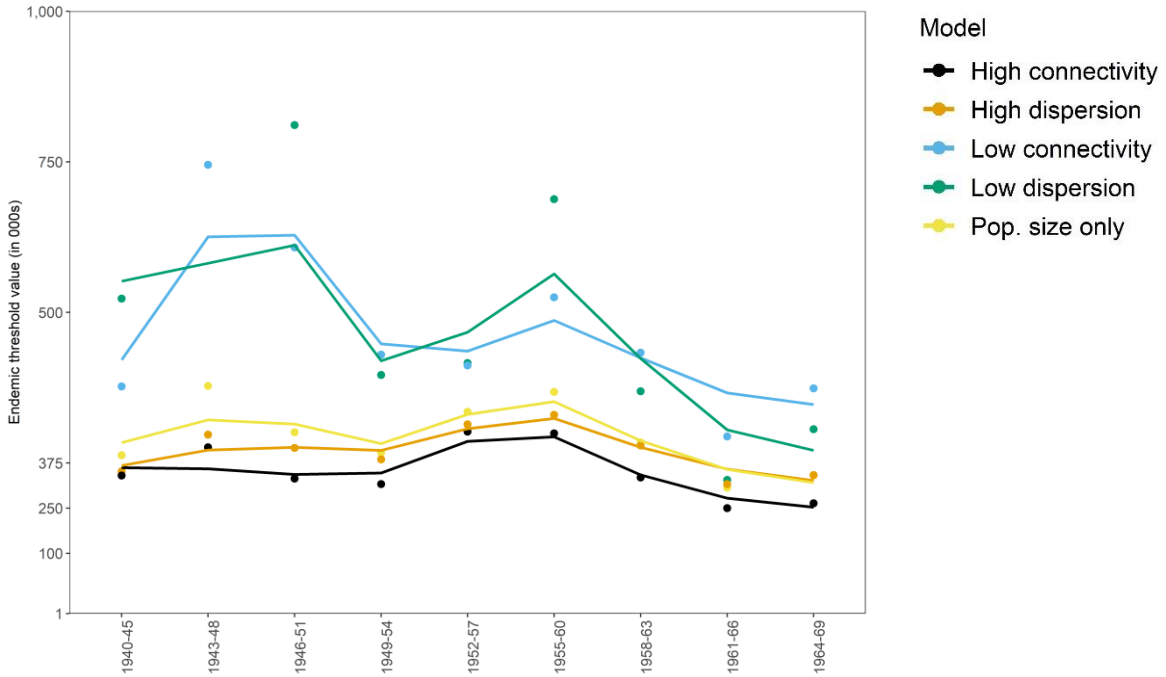

Figure III.1 *Endemic threshold population size estimates for measles in Lancashire for time–windows, accounting for population size, low/high geographical connectivity and low/high population density.*

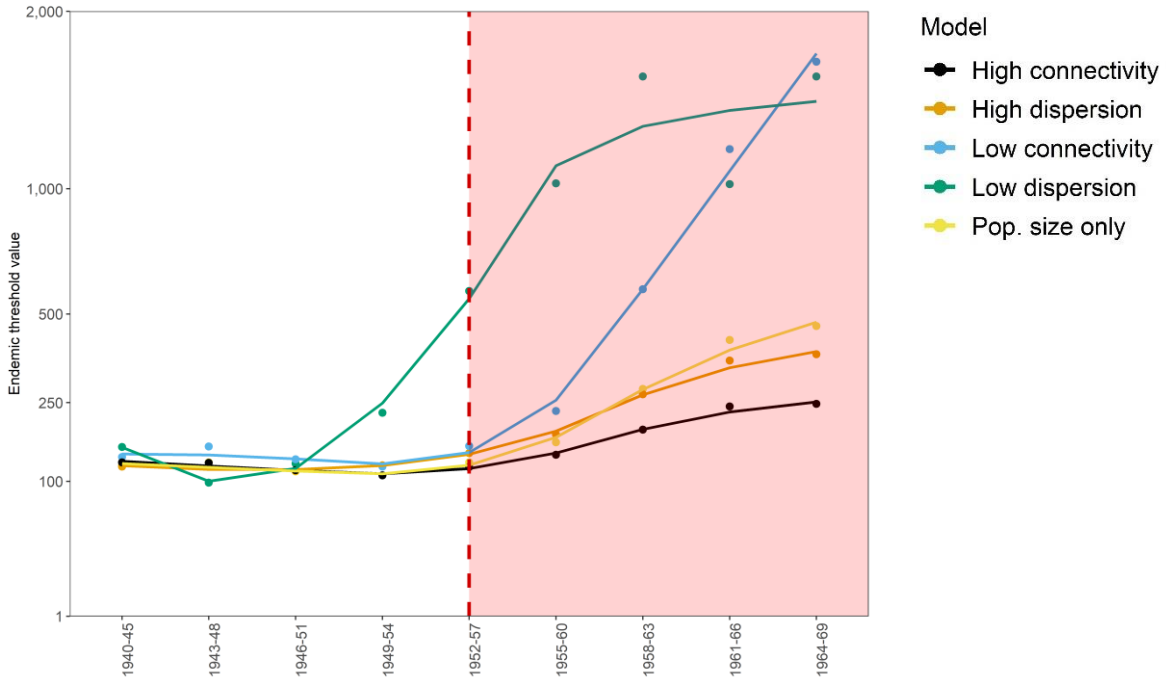

Figure III.2 *Endemic threshold population size estimates for pertussis in Lancashire for nine time–windows, accounting for population size, low/high geographical connectivity and low/high population density. The red shaded area denotes the vaccine era.*

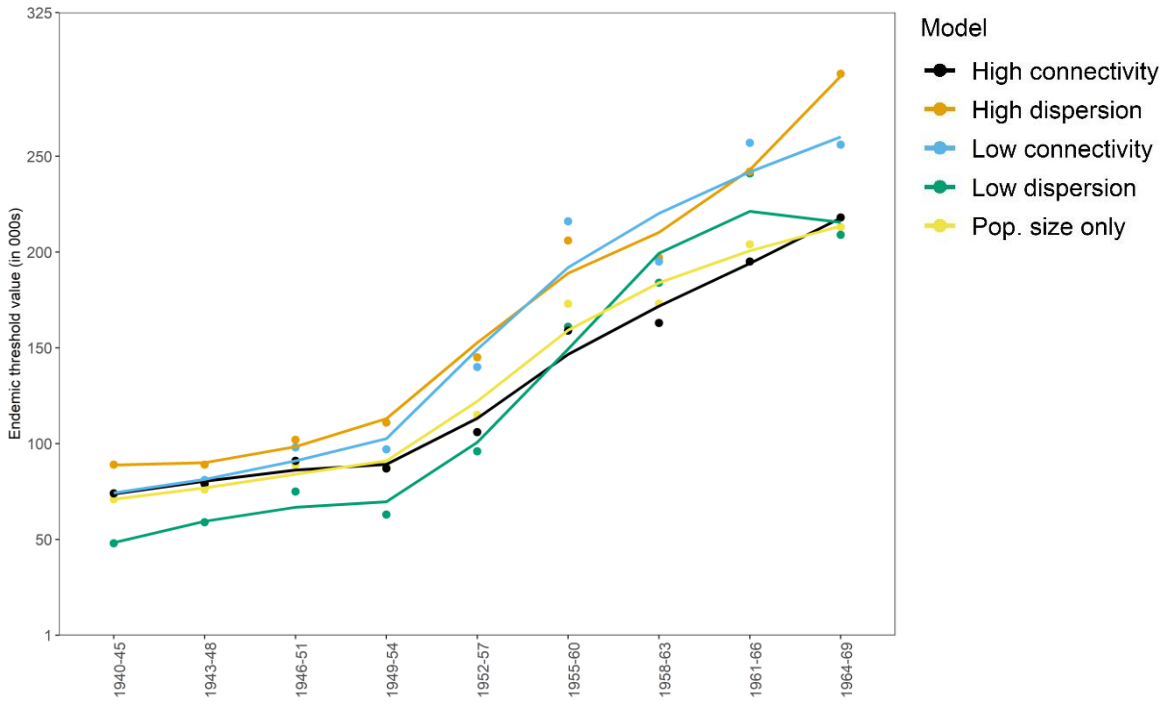

Figure III.3 *Endemic threshold population size estimates for scarlet fever in Lancashire for time–windows, accounting for population size, low/high geographical connectivity and low/high population density.*

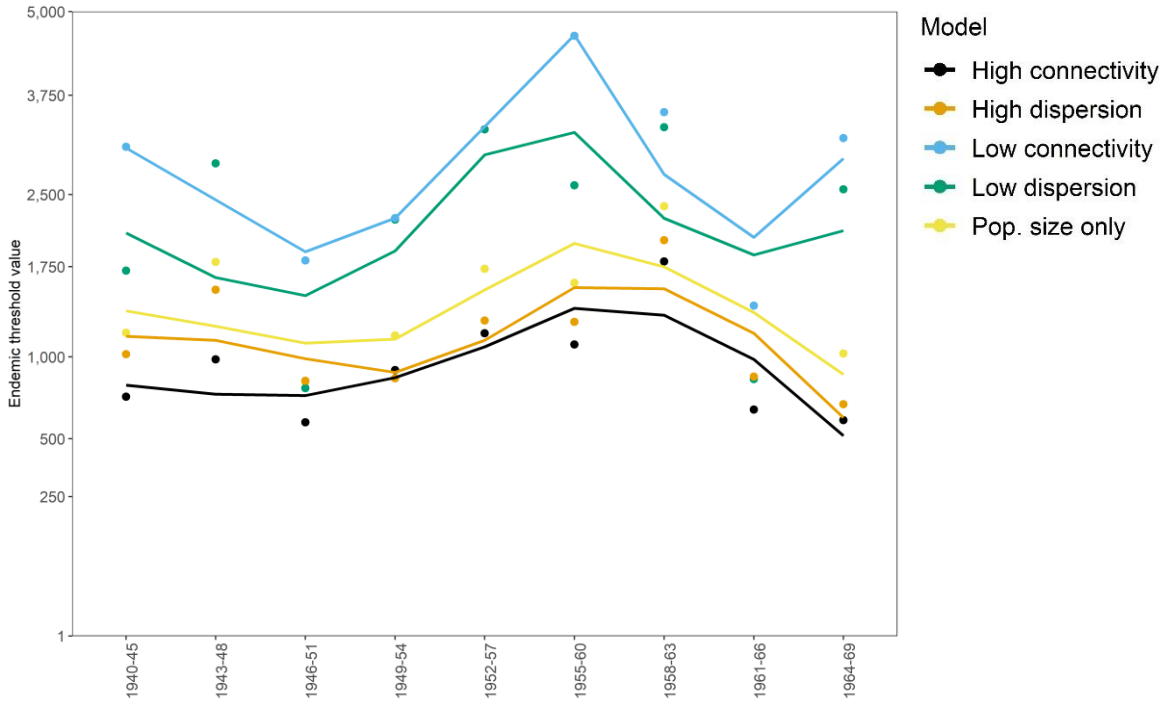

Figure III.4 *Endemic threshold population size estimates for measles in South Wales for nine time–windows, accounting for population size, low/high geographical connectivity and low/high population density.*

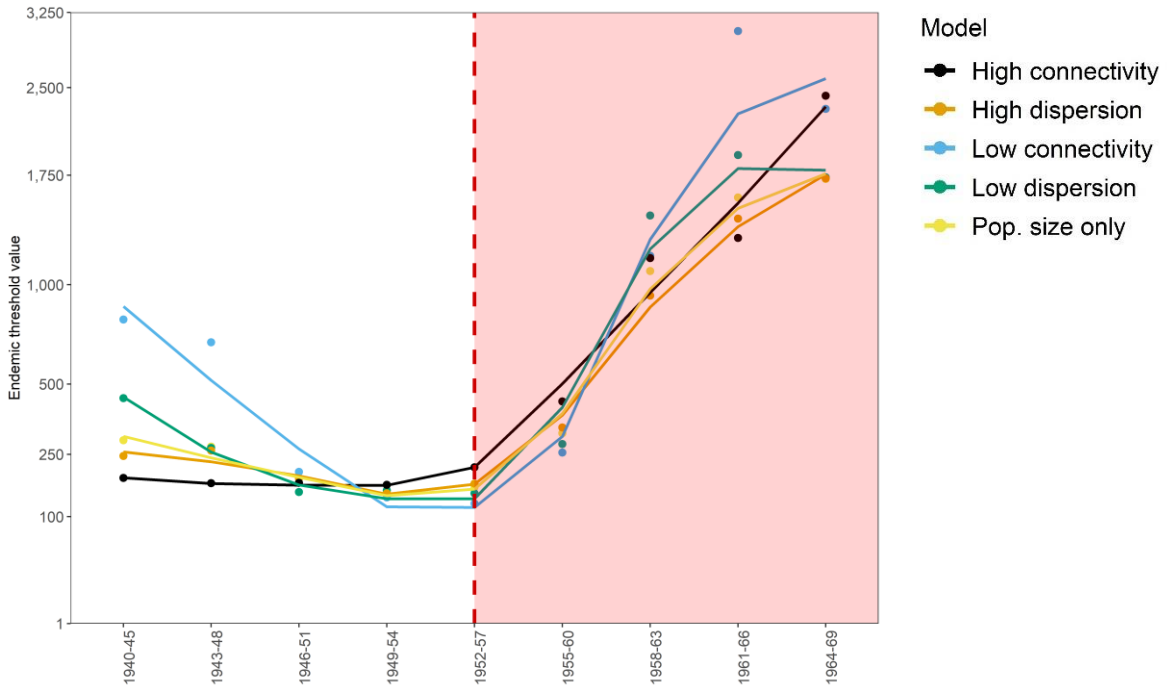

Figure III.5 *Endemic threshold population size estimates for pertussis in South Wales for time–windows, accounting for population size, low/high geographical connectivity and low/high population density. The red shaded area denotes the vaccine era.*

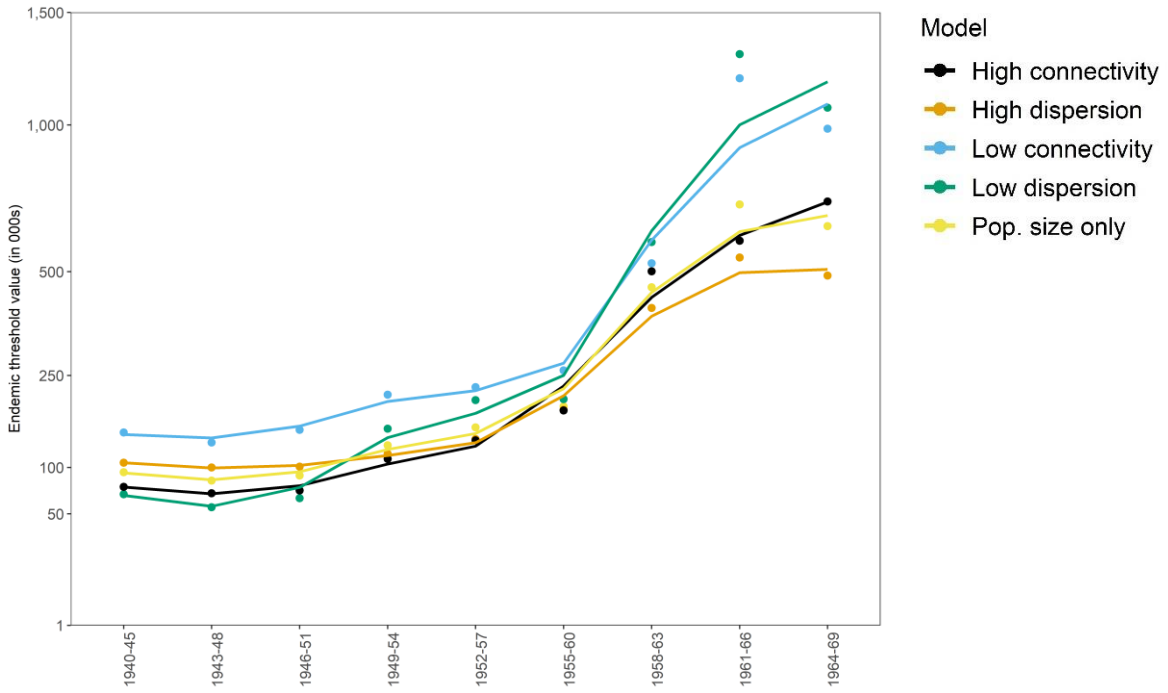

Figure III.5 *Endemic threshold population size estimates for scarlet fever in South Wales for nine time–windows, accounting for population size, low/high geographical connectivity and low/high population density.*

APPENDIX IV

V. Hotspot lists (Findings from *Chapter 7: Hotspot & Survival Analysis*)

Table IV.1 Lists of pertussis hotspots in Lancashire, by era of disease intervention.

Hotspots	N	District
Pre-vaccine era	**55**	Blackburn CB, Blackpool CB, Bolton CB, Bootle CB, Burnley CB, Bury CB, Oldham CB, Preston CB, Rochdale CB, St. Helens CB, Salford CB, Southport CB, Warrington CB, Wigan CB, Ashton-under-Lyne MB, Chorley MB, Crosby MB, Darwen MB, Eccles MB, Farnworth MB, Fleetwood MB, Lancaster MB, Leigh MB, Middleton MB, Morecambe & Heysham MB, Nelson MB, Radcliffe MB, Rawtenstall MB, Stretford MB, Swinton & Pendlebury MB, Widnes MB, Abram UD, Chadderton UD, Crompton UD, Denton UD, Droylsden UD, Golborne UD, Haydock UD, Hindley UD, Huyton-with-Roby UD, Ince-in-Makerfield UD, Kearsley UD, Litherland UD, Newton-le-Willows UD, Royton UD, Skelmersdale UD, Trawden UD, Tyldesley UD, Urmston UD, Walton-le-Dale UD, Withnell UD, Preston RD, Warrington RD, West Lancashire RD, Whiston RD
Vaccine era	**38**	Blackburn CB, Blackpool CB, Bolton CB, Bootle CB, Burnley CB, Bury CB, Liverpool CB, Manchester CB, Oldham CB, Preston CB, Rochdale CB, St. Helens CB, Salford CB, Southport CB, Warrington CB, Wigan CB, Accrington MB, Ashton-under-Lyne MB, Chorley MB, Crosby MB, Darwen MB, Eccles MB, Farnworth MB, Lancaster MB, Middleton MB, Radcliffe MB, Stretford MB, Swinton & Pendlebury MB, Widnes MB, Denton UD, Golborne UD, Huyton-with-Roby UD, Irlam UD, Litherland UD, Urmston UD, Worsley UD, West Lancashire RD, Whiston RD
Both eras	**35**	Blackburn CB, Blackpool CB, Bolton CB, Bootle CB, Burnley CB, Bury CB, Oldham CB, Preston CB, Rochdale CB, St. Helens CB, Salford CB, Southport CB, Warrington CB, Wigan CB, Liverpool CB, Manchester CB, Ashton-under-Lyne MB, Chorley MB, Crosby MB, Darwen MB, Eccles MB, Farnworth MB, Lancaster MB, Middleton MB, Radcliffe MB, Stretford MB, Swinton & Pendlebury MB, Widnes MB, Denton UD, Golborne UD, Huyton-with-Roby UD, Litherland UD, Urmston UD, West Lancashire RD, Whiston RD

Table IV.2 Lists of pertussis hotspots in South Wales, by era of disease intervention.

Hotspots	N	District
Pre-vaccine era	**31**	Abergavenny MB, Barry MB, Bedwas & Machen UD, Blaenavon UD, Burry Port UD, Cardiff CB, Chepstow UD, Cowbridge MB, Cwmamman UD, Cwmbran UD, Ebbw Vale UD, Fishguard & Goodwick UD, Haverfordwest MB, Kidwelly MB, Llandilo UD, Llandovery MB, Llanelli RD, Narberth RD, Narberth UD, Newcastle Emlyn RD, Newport CB, Neyland UD, Ogmore & Garw UD, Pembroke RD, Port Talbot MB, Rhondda MB, Rhymney UD, Swansea CB, Tenby MB, Tredegar UD, Usk UD
Vaccine era	**17**	Barry MB, Cardiff CB, Cardiff RD, Ebbw Vale UD, Gelligaer UD, Llanelli MB, Llanelli RD, Llantrisant & Llantwitfardre RD, Merthyr Tydfil CB, Neath RD, Newport CB, Ogmore & Garw UD, Penybont RD, Pontypool UD, Port Talbot MB, Rhondda MB, Swansea CB
Both eras	**9**	Barry MB, Cardiff CB, Ebbw Vale UD, Llanelli RD, Newport CB, Ogmore & Garw UD, Port Talbot MB, Rhondda MB, Swansea CB

APPENDIX V

I. HHH Model Summaries: Measles

Table V.1 Estimated model parameters (with standard errors) for power-law + gravity HHH model formulation for analysis of measles spread in Lancashire time–windows.

Table V.2 Estimated model parameters (with standard errors) for power-law + gravity HHH model formulation for analysis of measles spread in South Wales time–windows.

Table V.3 Estimated model parameters (with standard errors) for power-law HHH model formulation for analysis of measles spread in Lancashire time–windows.

Table V.4 Estimated model parameters (with standard errors) for power-law HHH model formulation for analysis of measles spread in South Wales time–windows.

Table V.5 Estimated model parameters (with standard errors) for gravity HHH model formulation for analysis of measles spread in Lancashire time–windows.

Table V.6 Estimated model parameters (with standard errors) for gravity HHH model formulation for analysis of measles spread in South Wales time–windows.

Table V.7 Estimated model parameters (with standard errors) for first-order HHH model formulation for analysis of measles spread in Lancashire time–windows.

Table V.8 Estimated model parameters (with standard errors) for first-order HHH model formulation for analysis of measles spread in South Wales time–windows.

Table V.9 Estimated model parameters (with standard errors) for endemic HHH model formulation for analysis of measles spread in Lancashire time–windows.

Table V.10 Estimated model parameters (with standard errors) for endemic HHH model formulation for analysis of measles spread in South Wales time–windows.

Table V.1

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$	0.484 (0.022)	0.417 (0.022)	0.380 (0.022)	0.400 (0.023)	0.430 (0.021)	0.476 (0.021)	0.518 (0.021)	0.482 (0.022)	0.567 (0.019)
$\boldsymbol{d}$ **(se)**	9.065 (0.540)	9.185 (0.590)	12.020 (0.840)	12.810 (0.97	8.940 (0.520)	8.629 (0.500)	8.278 (0.550)	9.104 (0.630)	8.012 (0.440)
$\boldsymbol{\alpha}^{(\lambda)}$ **(se)**	0.741 (0.0090)	0.766 (0.0092)	0.789 (0.0087)	0.777 (0.0087)	0.78 (0.0092)	0.804 (0.0097)	0.797 (0.0094)	0.757 (0.0088)	0.688 (0.0092)
$\boldsymbol{\alpha}^{(\phi)}$ **(se)**	0.001 (0.0002)	0.002 (0.0004)	0.002 (0.0005)	0.002 (0.0004)	0.001 (0.0003)	0.001 (0.0002)	0.001 (0.0001)	0.001 (0.0002)	0.001 (0.0001)
$\boldsymbol{\alpha}^{(v)}_{t}$ **(se)**	0.997 (0.0005)	1.000 (0.0006)	0.998 (0.0005)	1.001 (0.0004)	0.998 (0.0005)	0.999 (0.0006)	1.002 (0.0006)	1.005 (0.0006)	1.002 (0.0008)
$\boldsymbol{\psi}$ **(se)**	1.046 (0.015)	1.005 (0.015)	0.911 (0.013)	0.953 (0.014)	1.041 (0.015)	1.095 (0.016)	1.048 (0.015)	0.996 (0.013)	1.124 (0.016)
$\boldsymbol{maxEV}$	0.89	0.91	0.91	0.90	0.90	0.93	0.94	0.93	0.93
$\boldsymbol{l}_{\log}$	-68996.23	-63795.54	-67590.91	-68443.69	-68422.11	-67275.77	-69467.41	-75177.95	-69997.90

Table V.2

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$	-0.152 (0.026)	-0.229 (0.030)	-0.232 (0.030)	-0.196 (0.029)	-0.090 (0.027)	-0.119 (0.027)	-0.114 (0.026)	-0.221 (0.028)	-0.022 (0.028)
$\boldsymbol{d}$ **(se)**	0.556 (0.078)	0.706 (0.120)	0.784 (0.160)	0.129 (0.024)	0.181 (0.037)	0.807 (0.100)	0.866 (0.100)	0.556 (0.130)	0.963 (0.110)
$\boldsymbol{\alpha}^{(\lambda)}$ **(se)**	0.871 (0.021)	0.884 (0.022)	0.908 (0.019)	0.903 (0.020)	0.9382 (0.022)	0.9503 (0.022)	0.9063 (0.023)	0.8546 (0.022)	0.836 (0.024)
$\boldsymbol{\alpha}^{(\phi)}$ **(se)**	0.508 (0.120)	0.917 (0.240)	0.779 (0.200)	0.561 (0.140)	0.233 (0.056)	0.276 (0.067)	0.289 (0.068)	0.883 (0.220)	0.203 (0.051)
$\boldsymbol{\alpha}^{(v)}_{t}$ **(se)**	1.005 (0.0007)	0.9987 (0.0006)	1.003 (0.0004)	0.9955 (0.0004)	1.001 (0.0007)	0.9959 (0.0005)	1.004 (0.0007)	1.005 (0.0005)	0.9987 (0.0007)
$\boldsymbol{\psi}$ **(se)**	2.422 (0.047)	2.546 (0.052)	2.000 (0.040)	2.031 (0.042)	2.300 (0.048)	2.444 (0.049)	2.685 (0.054)	2.871 (0.052)	3.26 (0.061)
$\boldsymbol{maxEV}$	0.99	0.99	1.00	0.99	1.04	1.04	1.00	0.96	1.00
$\boldsymbol{l}_{\log}$	-32448.37	-30035.44	-32259.99	-30264.71	-29793.09	-30991.59	-30339.91	-35044.89	-30890.48

Table V.3

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
d **(se)**	1.876 (0.123)	1.932 (0.131)	2.632 (0.190)	2.95 (0.226)	1.627 (0.102)	1.669 (0.108)	2.073 (0.149)	2.209 (0.151)	2.049 (0.134)
$\alpha^{(\lambda)}$ **(se)**	0.774 (0.009)	0.792 (0.009)	0.808 (0.009)	0.796 (0.009)	0.807 (0.009)	0.835 (0.010)	0.832 (0.009)	0.794 (0.009)	0.733 (0.009)
$\alpha^{(\phi)}$ **(se)**	0.091 (0.002)	0.099 (0.002)	0.090 (0.002)	0.084 (0.002)	0.080 (0.002)	0.081 (0.002)	0.087 (0.002)	0.107 (0.002)	0.130 (0.003)
$\alpha_t^{(\nu)}$ **(se)**	0.997 (0.0003)	1.001 (0.0004)	0.999 (0.0003)	1.001 (0.0003)	0.998 (0.0003)	1.000 (0.0003)	1.000 (0.0003)	1.003 (0.0003)	1.000 (0.0003)
ψ **(se)**	1.07 (0.015)	1.02 (0.015)	0.92 (0.014)	0.97 (0.014)	1.06 (0.015)	1.12 (0.016)	1.07 (0.015)	1.02 (0.014)	1.16 (0.016)
$maxEV$	0.87	0.89	0.90	0.88	0.89	0.92	0.92	0.90	0.86
$l_{\log}$	-69170.28	-63930.86	-67693.37	-68554.10	-68582.78	-67469.08	-69667.70	-75352.51	-70273.21

Table V.4

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
d **(se)**	0.579 (0.083)	0.837 (0.126)	1.005 (0.143)	0.131 (0.024)	0.177 (0.036)	0.927 (0.104)	0.980 (0.104)	0.865 (0.137)	0.988 (0.103)
$\alpha^{(\lambda)}$ **(se)**	0.860 (0.020)	0.868 (0.022)	0.898 (0.019)	0.893 (0.019)	0.934 (0.022)	0.947 (0.023)	0.900 (0.023)	0.842 (0.022)	0.835 (0.024)
$\alpha^{(\phi)}$ **(se)**	0.130 (0.004)	0.125 (0.004)	0.105 (0.003)	0.100 (0.003)	0.103 (0.003)	0.095 (0.003)	0.103 (0.003)	0.128 (0.004)	0.167 (0.005)
$\alpha_t^{(v)}$ **(se)**	1.006 (0.0008)	0.998 (0.0007)	1.003 (0.0005)	0.996 (0.0004)	1.001 (0.0007)	0.996 (0.0005)	1.004 (0.0007)	1.004 (0.0006)	0.999 (0.0007)
ψ **(se)**	2.43 (0.047	2.57 (0.052)	2.01 (0.040)	2.03 (0.043)	2.30 (0.048)	2.44 (0.049)	2.68 (0.054)	2.88 (0.052)	3.26 (0.061)
$maxEV$	0.99	0.99	1	0.99	1.04	1.04	1	0.97	1
$l_{\log}$	-32467.05	-30067.09	-32293.58	-30289.19	-29798.91	-31001.68	-30349.85	-35077.84	-30890.79

Table V.5

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$	-0.091 (0.031)	-0.114 (0.029	-0.098 (0.028)	-0.022 (0.029)	-0.037 (0.031)	0.004 (0.031)	0.059 (0.029)	0.051 (0.026)	0.128 (0.027)
$\boldsymbol{\alpha}^{(\lambda)}$ **(se)**	0.769 (0.009	0.785 (0.009)	0.800 (0.009)	0.787 (0.009)	0.792 (0.009)	0.818 (0.010)	0.818 (0.010)	0.785 (0.009)	0.717 (0.009)
$\boldsymbol{\alpha}^{(\phi)}$ **(se)**	0.038 (0.011)	0.053 (0.015)	0.042 (0.011)	0.019 (0.005)	0.021 (0.006)	0.014 (0.004)	0.009 (0.003)	0.012 (0.003)	0.006 (0.002)
$\boldsymbol{\alpha}^{(\nu)}_{t}$ **(se)**	0.997 (0.0002	1.001 (0.0003)	1.000 (0.0002)	1.001 (0.0002)	0.998 (0.0002)	1.000 (0.0002)	1.000 (0.0002)	1.002 (0.0002)	0.999 (0.0002)
$\boldsymbol{\psi}$ **(se)**	1.118 (0.015)	1.077 (0.016)	0.965 (0.014)	1.004 (0.014)	1.11 (0.016)	1.163 (0.017)	1.108 (0.015)	1.052 (0.014)	1.194 (0.016)
$\boldsymbol{maxEV}$	0.88	0.91	0.91	0.89	0.89	0.92	0.93	0.92	0.87
$\boldsymbol{l}_{\log}$	-69764.81	-64443.91	-68102.4	-68927.92	-69158.36	-67935.87	-70101.46	-75898.71	-70629.97

Table V.6

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$	-0.92 (0.065)	-1.019 (0.069)	-0.642 (0.053)	-0.409 (0.049)	-0.401 (0.053)	-0.510 (0.061)	-0.730 (0.058)	-0.848 (0.053)	-0.648 (0.056)
$\boldsymbol{\alpha}^{(\lambda)}$ **(se)**	0.890 (0.024)	0.905 (0.025)	0.912 (0.020)	0.907 (0.021)	0.991 (0.024)	0.999 (0.025)	0.965 (0.027)	0.876 (0.024)	0.827 (0.026)
$\boldsymbol{\alpha}^{(\phi)}$ **(se)**	49.22 (27.000)	87.03 (49.000)	3.414 (1.500)	0.546 (0.230)	0.468 (0.210)	0.863 (0.440)	10.06 (4.800)	31.44 (14.000)	5.278 (2.400)
$\boldsymbol{\alpha}^{(\nu)}_{t}$ **(se)**	1.000 (0.0003)	0.999 (0.0003)	1.004 (0.0003)	0.995 (0.0003)	1.001 (0.0003)	0.996 (0.0003)	1.000 (0.0003)	1.004 (0.0003)	0.997 (0.0003)
$\boldsymbol{\psi}$ **(se)**	3.14 (0.059)	3.219 (0.065)	2.277 (0.046)	2.356 (0.049)	2.757 (0.057)	2.891 (0.058)	3.498 (0.069)	3.411 (0.060)	3.831 (0.071)
$\boldsymbol{maxEV}$	1.01	1.01	0.99	0.98	1.05	1.05	1.07	1.01	0.93
$\boldsymbol{l}_{\log}$	-33875.49	-31311.92	-33235.62	-31240.28	-30801.41	-31932.21	-31612.5	-35959.68	-31752.26

Table V.7

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\alpha^{(\lambda)}$ **(se)**	0.765 (0.009)	0.779 (0.009)	0.795 (0.009)	0.786 (0.009)	0.79 (0.009)	0.819 (0.010)	0.821 (0.009)	0.788 (0.009)	0.725 (0.009)
$\alpha^{(\phi)}$ **(se)**	0.016 (0.0004)	0.018 (0.0004)	0.017 (0.0004)	0.016 (0.0004)	0.015 (0.0004)	0.015 (0.0004)	0.016 (0.0004)	0.019 (0.0005)	0.021 (0.0006)
$\alpha_t^{(\nu)}$ **(se)**	0.997 (0.0002)	1.001 (0.0002)	1.000 (0.0002)	1.001 (0.0002)	0.998 (0.0002)	1.000 (0.0002)	1.000 (0.0002)	1.002 (0.0002)	0.999 (0.0002)
ψ **(se)**	1.118 (0.015)	1.078 (0.016)	0.966 (0.014)	1.004 (0.014)	1.11 (0.016)	1.163 (0.017)	1.108 (0.015)	1.053 (0.014)	1.196 (0.016)
$maxEV$	0.88	0.91	0.91	0.89	0.89	0.92	0.93	0.92	0.87
$l_{\log}$	-69769.23	-64451.67	-68108.85	-68928.22	-69159.09	-67935.88	-70103.51	-75900.55	-70640.84

Table V.8

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\alpha^{(\lambda)}$ **(se)**	0.881 (0.024)	0.892 (0.025)	0.905 (0.02)	0.901 (0.021)	0.985 (0.024)	0.995 (0.025)	0.949 (0.027)	0.850 (0.023)	0.815 (0.026)
$\alpha^{(\phi)}$ **(se)**	0.0176 (0.0011)	0.0142 (0.0009)	0.013 (0.0007)	0.015 (0.0008)	0.014 (0.0008)	0.011 (0.0007)	0.018 (0.0011)	0.023 (0.0012)	0.022 (0.0013)
$\alpha_t^{(v)}$ **(se)**	1.001 (0.00003)	0.999 (0.00003)	1.004 (0.0003)	0.995 (0.0003)	1.002 (0.0003)	0.996 (0.0003)	1.000 (0.0003)	1.004 (0.0003)	0.997 (0.0003)
ψ **(se)**	3.205 (0.061)	3.294 (0.066)	2.284 (0.046)	2.349 (0.049)	2.757 (0.057)	2.897 (0.058)	3.482 (0.07)	3.436 (0.061)	3.864 (0.072)
$maxEV$	0.98	0.97	0.98	0.99	1.06	1.06	1.05	0.98	0.94
$l_{\log}$	-34002.36	-31456.21	-33326.66	-31278.07	-30833.46	-31974.09	-31715.81	-36135.75	-31837.27

Table V.9

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\alpha^{(v)}t}$ (se)	0.9972	1.002	1.001	1.001	1.001	1.000	1.001	1.000	0.996
	(0.00011)	(0.00012)	(0.00013)	(0.00012)	(0.00012)	(0.00012)	(0.00012)	(0.00011)	(0.00012)
$\boldsymbol{\alpha^{(v)}}A$ (se)	0.385	0.806	0.736	0.676	0.611	0.559	0.673	0.576	0.317
	(0.015)	(0.016)	(0.015)	(0.015)	(0.016)	(0.017)	(0.015)	(0.014)	(0.014)
$\boldsymbol{\alpha^{(v)}}S$ (se)	0.509	0.767	0.893	0.637	0.418	0.475	0.590	1.061	1.683
	(0.023)	(0.014)	(0.013)	(0.016)	(0.020)	(0.021)	(0.016)	(0.013)	(0.008)
$\boldsymbol{\psi}$ (se)	3.617	3.978	3.793	3.635	4.061	4.302	3.808	3.319	3.247
	(0.036)	(0.041)	(0.038)	(0.036)	(0.040)	(0.043)	(0.038)	(0.032)	(0.033)
$\boldsymbol{l_{\log}}$	-81042.9	-76064.0	-80720.8	-80885.4	-80993.1	-79735.3	-81791.6	-87583.01	-80085.0

Table V.10

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\alpha^{(\nu)}t}$ (se)	0.997	1.001	1.003	0.995	1.001	0.995	1.001	0.998	0.998
	(0.00022)	(0.00025)	(0.00025)	(0.00032)	(0.00026)	(0.00026)	(0.00025)	(0.00022)	(0.00025)
$\boldsymbol{\alpha^{(\nu)}A}$ (se)	1.291	1.179	0.731	0.788	0.241	0.255	0.834	0.737	0.505
	(0.031)	(0.035)	(0.030)	(0.031)	(0.031)	(0.031)	(0.036)	(0.031)	(0.034)
$\boldsymbol{\alpha^{(\nu)}S}$ (se)	-0.287	-0.460	-0.375	0.293	-1.019	-0.508	-0.039	0.039	0.107
	(0.023)	(0.025)	(0.034)	(0.036)	(0.034)	(0.060)	(0.037)	(0.038)	(0.056)
$\boldsymbol{\psi}$ (se)	9.177	9.940	8.426	9.189	11.210	10.240	10.830	8.580	9.041
	(0.128)	(0.145)	(0.117)	(0.132)	(0.161)	(0.147)	(0.158)	(0.118)	(0.134)
$\boldsymbol{l_{\log}}$	-38486.7	-35636.0	-38772.8	-36684.3	-36349.7	-37106.3	-36046.9	-40273.7	-35342.8

II. HHH Model Summaries: Pertussis

Table V.11 Estimated model parameters (with standard errors) for power-law + gravity HHH model formulation for pertussis spread in Lancashire time–windows.

Table V.12 Estimated model parameters (with standard errors) for power-law + gravity HHH model formulation for pertussis spread in South Wales time–windows.

Table V.13 Estimated model parameters (with standard errors) for power-law HHH model formulation for pertussis spread in Lancashire time–windows.

Table V.14 Estimated model parameters (with standard errors) for power-law HHH model formulation for pertussis spread in South Wales time–windows.

Table V.15 Estimated model parameters (with standard errors) for gravity HHH model formulation for pertussis spread in Lancashire time–windows.

Table V.16 Estimated model parameters (with standard errors) for gravity HHH model formulation for pertussis spread in South Wales time–windows.

Table V.17 Estimated model parameters (with standard errors) for first-order HHH model formulation for pertussis spread in Lancashire time–windows.

Table V.18 Estimated model parameters (with standard errors) for first-order HHH model formulation for pertussis spread in South Wales time–windows.

Table V.19 Estimated model parameters (with standard errors) for endemic HHH model formulation for pertussis spread in Lancashire time–windows.

Table V.20 Estimated model parameters (with standard errors) for endemic HHH model formulation for pertussis spread in South Wales time–windows.

Table V.11

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$	0.691 (0.030)	0.299 (0.050)	0.431 (0.030)	0.449 (0.030)	0.557 (0.030)	0.760 (0.020)	0.875 (0.030)	0.926 (0.030)	0.842 (0.040)
$\boldsymbol{d}$ **(se)**	2.362 (0.200)	7.509 (0.930)	6.189 (0.540)	6.478 (0.550)	5.668 (0.540)	3.504 (0.240)	3.361 (0.240)	3.923 (0.300)	3.669 (0.310)
$\boldsymbol{\alpha}^{(\lambda)}$ **(se)**	0.597 (0.010)	0.502 (0.010)	0.596 (0.010)	0.635 (0.010)	0.612 (0.010)	0.564 (0.010)	0.397 (0.020)	0.326 (0.020)	0.300 (0.020)
$\boldsymbol{\alpha}^{(\phi)}$ **(se)**	0.0001 (0.000)	0.006 (0.000)	0.002 (0.000)	0.001 (0.000)	0.001 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\boldsymbol{\alpha}^{(v)}_{t}$ **(se)**	1.000 (0.000)	1.000 (0.000)	1.001 (0.000)	1.000 (0.000)	0.999 (0.000)	0.999 (0.000)	0.995 (0.000)	1.000 (0.000)	0.999 (0.000)
$\boldsymbol{\psi}$ **(se)**	1.383 (0.020)	1.302 (0.020)	0.898 (0.020)	0.823 (0.010)	1.091 (0.020)	1.392 (0.030)	2.371 (0.060)	2.663 (0.080)	3.068 (0.090)
$\boldsymbol{maxEV}$	0.82	0.62	0.74	0.79	0.79	0.86	0.80	0.76	0.67
$\boldsymbol{l}_{\log}$	-47516.79	-47777.18	-55390.75	-57780.20	-50760.32	-39862.60	-24724.57	-19332.27	-20571.67

Table V.12

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$	0.476 (0.070)	0.707 (0.060)	0.809 (0.030)	0.685 (0.030)	0.799 (0.050)	0.869 (0.050)	1.102 (0.060)	1.117 (0.050)	1.124 (0.060)
$\boldsymbol{d}$ **(se)**	1.363 (0.020)	0.432 (0.010)	0.360 (0.050)	0.437 (0.060)	0.896 (0.180)	1.107 (0.200)	0.906 (0.180)	0.874 (0.140)	1.081 (0.170)
$\boldsymbol{\alpha}^{(\lambda)}$ **(se)**	0.532 (0.020)	0.494 (0.020)	0.538 (0.020)	0.551 (0.020)	0.474 (0.020)	0.463 (0.030)	0.409 (0.030)	0.309 (0.030)	0.308 (0.030)
$\boldsymbol{\alpha}^{(\phi)}$ **(se)**	0.002 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\boldsymbol{\alpha}^{(v)}_{t}$ **(se)**	1.000 (0.000)	1.000 (0.000)	0.999 (0.000)	1.002 (0.000)	0.999 (0.000)	0.995 (0.000)	0.994 (0.000)	1.001 (0.000)	0.995 (0.000)
$\boldsymbol{\psi}$ **(se)**	2.795 (0.080)	2.810 (0.080)	2.263 (0.060)	2.358 (0.060)	2.812 (0.070)	3.660 (0.120)	4.571 (0.210)	4.594 (0.240)	5.759 (0.290)
$\boldsymbol{maxEV}$	0.69	0.70	0.84	0.82	0.71	0.71	0.78	0.80	0.76
$\boldsymbol{l}_{\log}$	-18499.51	-18292.76	-21070.92	-22596.68	-19971.09	-13793.08	-7856.72	-6651.05	-6909.03

Table V.13

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
d **(se)**	0.590 (0.055)	1.209 (0.129)	1.213 (0.113)	1.454 (0.141)	1.801 (0.206)	1.387 (0.146)	1.032 (0.092)	1.209 (0.115)	1.027 (0.115)
$\alpha^{(\lambda)}$ **(se)**	0.630 (0.011)	0.505 (0.010)	0.602 (0.008)	0.638 (0.008)	0.616 (0.009)	0.586 (0.011)	0.480 (0.016)	0.413 (0.017)	0.353 (0.016)
$\alpha^{(\phi)}$ **(se)**	0.079 (0.003)	0.084 (0.004)	0.087 (0.003)	0.087 (0.003)	0.084 (0.003)	0.098 (0.004)	0.153 (0.006)	0.163 (0.007)	0.151 (0.008)
$\alpha_t^{(v)}$ **(se)**	1.000 (0.0002)	0.999 (0.0002)	1.001 (0.0002)	0.999 (0.0002)	0.998 (0.0002)	0.999 (0.0002)	0.996 (0.0003)	1.001 (0.0003)	0.998 (0.0003)
ψ **(se)**	1.42 (0.025)	1.31 (0.024)	0.91 (0.016)	0.83 (0.014)	1.10 (0.020)	1.44 (0.030)	2.62 (0.066)	2.97 (0.089)	3.25 (0.093)
$maxEV$	0.71	0.59	0.69	0.73	0.70	0.68	0.63	0.58	0.50
$l_{\log}$	-47646.90	-47789.71	-55444.77	-57847.07	-50826.19	-40037.88	-25022.23	-19545.48	-20703.05

Table V.14

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
d **(se)**	2.596 (0.622)	0.574 (0.334)	0.313 (0.071)	0.307 (0.060)	0.251 (0.105)	0.334 (0.184)	0.322 (0.126)	0.512 (0.117)	0.724 (0.191)
$\alpha^{(\lambda)}$ **(se)**	0.538 (0.021)	0.501 (0.020)	0.580 (0.019)	0.594 (0.019)	0.503 (0.020)	0.501 (0.026)	0.492 (0.036)	0.396 (0.033)	0.387 (0.036)
$\alpha^{(\phi)}$ **(se)**	0.068 (0.006)	0.054 (0.006)	0.071 (0.005)	0.098 (0.006)	0.062 (0.006)	0.072 (0.008)	0.113 (0.011)	0.169 (0.015)	0.106 (0.013)
$\alpha_t^{(\nu)}$ **(se)**	0.9997 (0.0003)	1.001 (0.0003)	1.001 (0.0003)	1.001 (0.0003)	0.9994 (0.0003)	0.9955 (0.0003)	0.9954 (0.0004)	1 (0.0005)	0.9965 (0.0005)
ψ **(se)**	2.82 (0.078)	2.85 (0.079)	2.35 (0.060)	2.44 (0.060)	2.91 (0.075)	3.82 (0.123)	4.95 (0.225)	5.19 (0.265)	6.29 (0.305)
$maxEV$	0.61	0.56	0.65	0.69	0.57	0.57	0.6	0.56	0.49
$l_{\log}$	-18512.33	-18319.57	-21190.47	-22723.08	-20053.69	-13875.44	-7965.11	-6787.95	-7004.3

Table V.15

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$	-0.402 (0.061)	-0.343 (0.067)	-0.162 (0.052)	-0.114 (0.046)	-0.1 (0.056)	0.08 (0.063)	0.485 (0.069)	0.627 (0.079)	0.738 (0.107)
$\boldsymbol{\alpha}^{(\lambda)}$ **(se)**	0.601 (0.011)	0.488 (0.009)	0.593 (0.008)	0.630 (0.008)	0.603 (0.009)	0.568 (0.011)	0.440 (0.016)	0.369 (0.017)	0.309 (0.016)
$\boldsymbol{\alpha}^{(\phi)}$ **(se)**	0.468 (0.261)	0.244 (0.151)	0.051 (0.025)	0.035 (0.016)	0.031 (0.017)	0.006 (0.004)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\boldsymbol{\alpha}^{(\nu)}_{t}$ **(se)**	1.000 (0.0002)	1.000 (0.0002)	1.001 (0.0002)	1.000 (0.0002)	0.999 (0.0002)	0.998 (0.0002	0.997 (0.0002)	1.001 (0.0003)	0.998 (0.0002)
$\boldsymbol{\psi}$ **(se)**	1.445 (0.025)	1.33 (0.024)	0.931 (0.016)	0.85 (0.0146)	1.118 (0.020)	1.466 (0.030)	2.687 (0.068)	2.982 (0.090)	3.282 (0.095)
$\boldsymbol{maxEV}$	0.67	0.55	0.67	0.71	0.68	0.66	0.64	0.63	0.54
$\boldsymbol{l}_{\log}$	-47841.09	-47917.9	-55662.95	-58026.17	-50926.19	-40174.93	-25247.69	-19686.51	-20803.43

Table V.16

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$	-0.848 (0.226)	-0.565 (0.269)	0.353 (0.446)	0.693 (0.114)	0.876 (0.139)	0.884 (0.132)	1.175 (0.111)	1.106 (0.132)	1.195 (0.162)
$\boldsymbol{\alpha}^{(\lambda)}$ **(se)**	0.519 (0.021)	0.486 (0.020)	0.541 (0.019)	0.548 (0.019)	0.470 (0.020)	0.469 (0.026)	0.429 (0.034)	0.350 (0.032)	0.355 (0.035)
$\boldsymbol{\alpha}^{(\phi)}$ **(se)**	5.932 (10.55)	0.361 (0.766)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
$\boldsymbol{\alpha}^{(\nu)}_{t}$ **(se)**	1.000 (0.0002)	1.001 (0.0002)	1.003 (0.0002)	1.001 (0.0002)	0.999 (0.0002)	0.996 (0.0003)	0.995 (0.0004)	0.999 (0.0004)	0.996 (0.0004)
$\boldsymbol{\psi}$ **(se)**	2.866 (0.080)	2.876 (0.080)	2.436 (0.062)	2.578 (0.063)	2.958 (0.076)	3.878 (0.124)	4.983 (0.229)	5.403 (0.279)	6.409 (0.311)
$\boldsymbol{maxEV}$	0.55	0.50	0.55	0.62	0.58	0.60	0.73	0.64	0.59
$\boldsymbol{l}_{\log}$	-18595.14	-18366.12	-21358.21	-22992.07	-20123.84	-13922.72	-8009.94	-6871.31	-7045.75

Table V.17

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\alpha^{(\lambda)}$ **(se)**	0.594 (0.011)	0.482 (0.010)	0.591 (0.008)	0.629 (0.008)	0.602 (0.009)	0.57 (0.011)	0.454 (0.016)	0.396 (0.017)	0.328 (0.016)
$\alpha^{(\phi)}$ **(se)**	0.011 (0.0005)	0.01 (0.0006)	0.011 (0.0005)	0.012 (0.0005)	0.012 (0.0006)	0.013 (0.0007)	0.018 (0.001)	0.02 (0.0012)	0.015 (0.0011)
$\alpha_t^{(\nu)}$ **(se)**	1.000 (0.0002)	1.000 (0.0002	1.001 (0.0002)	1.000 (0.0002)	0.999 (0.0002	0.998 (0.0002)	0.997 (0.0002)	1.001 (0.0003)	0.998 (0.0002)
ψ **(se)**	1.446 (0.025)	1.330 (0.024)	0.930 (0.016)	0.850 (0.015)	1.118 (0.020)	1.467 (0.0298)	2.717 (0.068)	3.061 (0.092)	3.342 (0.096)
$maxEV$	0.67	0.55	0.67	0.71	0.68	0.66	0.58	0.53	0.43
$l_{\log}$	-47864.17	-47931.55	-55667.92	-58029.21	-50927.79	-40175.74	-25274.05	-19719.31	-20831.95

Table V.18

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\alpha^{(\lambda)}$ **(se)**	0.519 (0.021)	0.486 (0.020)	0.541 (0.019)	0.556 (0.019)	0.481 (0.020)	0.483 (0.026)	0.466 (0.036)	0.374 (0.033)	0.370 (0.036)
$\alpha^{(\phi)}$ **(se)**	0.005 (0.0008)	0.003 (0.0008)	0.0004 (0.0005)	0.004 (0.0008)	0.004 (0.0009)	0.004 (0.001)	0.004 (0.0014)	0.006 (0.0018)	0.003 (0.0016)
$\alpha_t^{(v)}$ **(se)**	1.000 (0.0002)	1.001 (0.0002)	1.003 (0.0002	1.001 (0.0002)	0.999 (0.0002)	0.996 (0.0003)	0.996 (0.0004)	0.999 (0.0004)	0.996 (0.0004)
ψ **(se)**	2.869 (0.080)	2.878 (0.080)	2.437 (0.062)	2.591 (0.063)	2.976 (0.077)	3.915 (0.125)	5.1 (0.234)	5.554 (0.286)	6.518 (0.317)
$maxEV$	0.55	0.51	0.54	0.58	0.50	0.51	0.49	0.40	0.39
$l_{\log}$	-18605.12	-18368.89	-21358.45	-23004.69	-20134.17	-13937.23	-8045.13	-6893.59	-7064.25

Table V.19

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\alpha^{(v)}t}$ (se)	1.000	1.000	1.001	0.999	0.998	0.997	0.998	1.000	0.998
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0002)
$\boldsymbol{\alpha^{(v)}}A$ (se)	0.150	0.071	0.099	0.149	0.117	0.096	0.199	0.137	0.240
	(0.015)	(0.014)	(0.013)	(0.013)	(0.014)	(0.017)	(0.021)	(0.024)	(0.024)
$\boldsymbol{\alpha^{(v)}}S$ (se)	0.382	-0.618	-0.335	0.003	-0.323	-1.255	-2.793	2.579	2.415
	(0.036)	(0.023)	(0.038)	(0.085)	(0.041)	(0.014)	(0.008)	(0.009)	(0.010)
$\boldsymbol{\psi}$ (se)	2.786	2.305	2.133	2.067	2.479	3.301	4.920	5.629	5.197
	(0.040)	(0.035)	(0.028)	(0.026)	(0.035)	(0.053)	(0.104)	(0.142)	(0.130)
$\boldsymbol{l_{\log}}$	-51823	-51086.5	-61844.2	-65008.9	-56543.6	-44775.6	-27108.1	-21114.5	-21811.2

Table V.20

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\alpha^{(v)}t}$ (se)	0.999	1.000	1.004	0.999	0.999	0.996	0.996	0.998	0.996
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0003	(0.0003)	(0.0004)	(0.0004)
$\boldsymbol{\alpha^{(v)}}A$ (se)	0.040	0.074	0.127	0.179	0.274	0.135	0.239	0.565	0.273
	(0.026)	(0.026)	(0.024)	(0.024)	(0.025)	(0.030)	(0.040)	(0.044)	(0.044)
$\boldsymbol{\alpha^{(v)}}S$ (se)	1.865	-3.059	-0.05067	0.01432	0.4496	1.088	1.955	0.7809	1.906
	(0.015)	(0.009)	(0.170)	(0.132)	(0.049)	(0.030)	(0.021)	(0.047)	(0.024)
$\boldsymbol{\psi}$ (se)	5.118	4.821	4.047	4.329	4.632	6.550	9.281	9.557	10.840
	(0.119)	(0.115)	(0.090)	(0.089)	(0.103)	(0.182)	(0.367)	(0.429)	(0.463)
$\boldsymbol{l_{\log}}$	-19915.8	-19549.7	-22691.9	-24420.4	-21233.5	-14816.8	-8588.91	-7298.25	-7457.42

III. HHH Model Summaries: Scarlet fever

Table V.21 Estimated model parameters (with standard errors) for power-law + gravity HHH model formulation for analysis of scarlet fever spread in Lancashire time–windows.

Table V.22 Estimated model parameters (with standard errors) for power-law + gravity HHH model formulation for analysis of scarlet fever spread in South Wales time–windows.

Table V.23 Estimated model parameters (with standard errors) for power-law HHH model formulation for analysis of scarlet fever spread in Lancashire time–windows.

Table V.24 Estimated model parameters (with standard errors) for power-law HHH model formulation for analysis of scarlet fever spread in South Wales time–windows.

Table V.25 Estimated model parameters (with standard errors) for gravity HHH model formulation for analysis of scarlet fever spread in Lancashire time–windows.

Table V.26 Estimated model parameters (with standard errors) for gravity HHH model formulation for analysis of scarlet fever spread in South Wales time–windows.

Table V.27 Estimated model parameters (with standard errors) for first-order HHH model formulation for analysis of scarlet fever spread in Lancashire time–windows.

Table V.28 Estimated model parameters (with standard errors) for first-order HHH model formulation for analysis of scarlet fever spread in South Wales time–windows.

Table V.29 Estimated model parameters (with standard errors) for endemic HHH model formulation for analysis of scarlet fever spread in Lancashire time–windows.

Table V.30 Estimated model parameters (with standard errors) for endemic HHH model formulation for analysis of scarlet fever spread in South Wales time–windows.

Table V.21

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$	-0.064 (0.015)	-0.05 (0.016)	-0.124 (0.016	-0.152 (0.017	-0.177 (0.021)	-0.247 (0.022)	-0.268 (0.023	-0.375 (0.030)	-0.517 (0.033)
$\boldsymbol{d}$ **(se)**	1.876 (0.081)	2.072 (0.089)	2.136 (0.095)	2.306 (0.110)	2.063 (0.100)	1.794 (0.096)	1.886 (0.110)	1.856 (0.120)	1.876 (0.120)
$\boldsymbol{\alpha}^{(\lambda)}$ **(se)**	0.672 (0.007)	0.674 (0.007)	0.646 (0.008)	0.633 (0.008)	0.602 (0.009)	0.566 (0.010)	0.528 (0.011)	0.513 (0.013)	0.486 (0.013)
$\boldsymbol{\alpha}^{(\phi)}$ **(se)**	0.452 (0.060)	0.380 (0.060)	0.854 (0.130)	1.211 (0.200)	1.632 (0.310)	3.606 (0.740)	5.036 (1.100)	13.87 (3.800)	54.29 (16.000)
$\boldsymbol{\alpha}^{(v)}_{t}$ **(se)**	0.999 (0.002)	0.998 (0.002)	0.997 (0.0014)	1.002 (0.001)	1.001 (0.001)	1.001 (0.0009)	0.999 (0.0009)	1.001 (0.0008)	0.999 (0.0005)
$\boldsymbol{\psi}$ **(se)**	0.61 (0.013)	0.63 (0.014)	0.75 (0.017)	0.78 (0.018)	0.93 (0.024)	1.18 (0.031)	1.39 (0.039)	1.92 (0.054)	2.16 (0.057)
$\boldsymbol{maxEV}$	0.91	0.91	0.90	0.91	0.89	0.89	0.89	0.87	0.83
$\boldsymbol{l}_{\log}$	-53769.96	-52646.02	-47681.3	-46599.97	-39219.61	-34828.87	-29765.73	-26511.87	-27031.37

Table V.22

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$	1.124 (0.035)	1.254 (0.028)	1.234 (0.026)	–	1.268 (0.040)	1.252 (0.039)	1.348 (0.065)	1.155 (0.088)	0.923 (0.076)
$\boldsymbol{d}$ **(se)**	1.046 (0.160)	0.701 (0.081)	0.823 (0.079)	–	1.431 (0.160)	1.738 (0.190)	1.862 (0.270)	2.008 (0.370)	2.533 (0.410)
$\boldsymbol{\alpha}^{(\lambda)}$ **(se)**	0.341 (0.010)	0.292 (0.010)	0.302 (0.010)	–	0.316 (0.012)	0.327 (0.014)	0.283 (0.015)	0.259 (0.017)	0.301 (0.020)
$\boldsymbol{\alpha}^{(\phi)}$ **(se)**	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	–	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.00002 (0.00001)
$\boldsymbol{\alpha}^{(v)}_{t}$ **(se)**	1.002 (0.00026	0.999 (0.0003	0.998 (0.00035	–	0.999 (0.0004	0.998 (0.00044	0.997 (0.00043	1.000 (0.00043	0.999 (0.00042
$\boldsymbol{\psi}$ **(se)**	0.56 (0.020)	0.51 (0.019)	0.59 (0.022)	–	0.76 (0.036)	1.00 (0.047)	1.45 (0.077)	2.03 (0.120)	2.74 (0.140)
$\boldsymbol{maxEV}$	0.61	0.67	0.70	–	0.68	0.73	0.61	0.50	0.52
$\boldsymbol{l}_{\log}$	-22799.22	-23399.18	-21459.39	–	-15081.97	-13244.19	-9889.25	-8133.12	-8166.20

Table V.23

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
d **(se)**	10.14 (4.998)	5.757 (1.309)	6.989 (1.850)	4.337 (1.808)	3.346 (0.852)	3.618 (0.935)	3.841 (1.328)	4.796 (1.715)	2.525 (0.583)
$\alpha^{(\lambda)}$ **(se)**	0.386 (0.006)	0.420 (0.006)	0.402 (0.007)	0.363 (0.007)	0.352 (0.007)	0.334 (0.008)	0.286 (0.009)	0.308 (0.010)	0.318 (0.010)
$\alpha^{(\phi)}$ **(se)**	0.051 (0.003)	0.072 (0.003)	0.082 (0.003)	0.055 (0.004)	0.053 (0.004)	0.073 (0.005)	0.062 (0.005)	0.070 (0.006)	0.083 (0.007)
$\alpha_t^{(\nu)}$ **(se)**	1.002 (0.0001)	0.997 (0.0001)	1.000 (0.0001)	0.999 (0.00012)	0.996 (0.00013)	1.001 (0.0002)	0.996 (0.0002)	1.002 (0.00012)	0.999 (0.0002)
ψ **(se)**	0.35 (0.009)	0.38 (0.010)	0.75 (0.017)	0.78 (0.018)	0.93 (0.024)	1.18 (0.031)	1.39 (0.039)	1.92 (0.054)	2.16 (0.057)
$maxEV$	0.44	0.49	0.90	0.91	0.89	0.89	0.89	0.87	0.83
$l_{\log}$	-50370.51	-49370.84	-47681.3	-46599.97	-39219.61	-34828.87	-29765.73	-26511.87	-27031.37

Table V.24

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
d **(se)**	249.5 (891.3)	6.58 (3612)	1.063 (266.1)	–	1.531 (1.680)	0.811 (1.479)	–	1.388 (1.079)	2.383 (0.634)
$\alpha^{(\lambda)}$ **(se)**	0.360 (0.010)	0.338 (0.010)	0.354 (0.010)	–	0.357 (0.012)	0.377 (0.014)	–	0.271 (0.017)	0.336 (0.021)
$\alpha^{(\phi)}$ **(se)**	0.0060 (0.003)	0.00000003 (0.00002)	0.00000009 (0.0001)	–	0.0064 (0.004)	0.0122 (0.005)	–	0.0145 (0.009)	0.0682 (0.009)
$\alpha_t^{(v)}$ **(se)**	1.002 (0.0002	0.999 (0.0001	0.999 (0.0002)	–	0.999 (0.0002)	0.998 (0.0002)	–	0.999 (0.0003)	0.9995 (0.0003)
ψ **(se)**	0.57 (0.020)	0.54 (0.019)	0.63 (0.023)	–	0.81 (0.037)	1.10 (0.050)	–	2.08 (0.119)	2.82 (0.144)
$maxEV$	0.37	0.34	0.35	–	0.36	0.39	–	0.29	0.40
$l_{\log}$	-22845.79	-23535.89	-21624.82	–	-15186.49	-13374.72	–	-8159.52	-8209.64

Table V.25

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\beta^{(\phi)}_{\log(pop)}$	0.362 (0.058)	0.377 (0.045)	0.342 (0.045)	0.193 (0.056)	0.118 (0.064)	0.099 (0.064)	-0.004 (0.074)	0.247 (0.078)	0.278 (0.081)
$\alpha^{(\lambda)}$ **(se)**	0.379 (0.006)	0.406 (0.006)	0.389 (0.007)	0.360 (0.007)	0.351 (0.007)	0.329 (0.008)	0.283 (0.009)	0.303 (0.010)	0.310 (0.010)
$\alpha^{(\phi)}$ **(se)**	0.0003 (0.00017)	0.0003 (0.00015)	0.0005 (0.00023)	0.0017 (0.00091)	0.0033 (0.00204)	0.0047 (0.00292)	0.0123 (0.00878)	0.0011 (0.00087)	0.0008 (0.00065)
$\alpha^{(\nu)}_t$ **(se)**	1.002 (0.00011)	0.997 (0.00013)	1.000 (0.00013)	0.999 (0.00012)	0.996 (0.00014)	1.001 (0.00015)	0.996 (0.00016)	1.002 (0.00019)	0.990 (0.00018)
ψ **(se)**	0.35 (0.009)	0.38 (0.010)	0.44 (0.012)	0.43 (0.012)	0.51 (0.016)	0.65 (0.021)	0.73 (0.026)	1.11 (0.037)	1.39 (0.041)
$maxEV$	0.47	0.53	0.51	0.44	0.42	0.42	0.36	0.4	0.41
$l_{\log}$	-50351.74	-49355.03	-44496.31	-43338.37	-36352.93	-32247.92	-27349.2	-24532.22	-25357.24

Table V.26

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\beta}^{(\phi)}_{\log(pop)}$	0.82 (0.150)	1.198 (0.130)	1.054 (0.098)	–	1.157 (0.091)	1.139 (0.069)	1.259 (0.120)	0.909 (0.190)	0.646 (0.120)
$\boldsymbol{\alpha}^{(\lambda)}$ **(se)**	0.359 (0.010)	0.333 (0.010)	0.348 (0.010)	–	0.345 (0.012)	0.360 (0.014)	0.303 (0.016)	0.266 (0.017)	0.328 (0.021)
$\boldsymbol{\alpha}^{(\phi)}$ **(se)**	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	–	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.00002 (0.00003)
$\boldsymbol{\alpha}^{(v)}_{t}$ **(se)**	1.002 (0.0002)	0.999 (0.0002)	0.999 (0.0002)	–	0.999 (0.0002)	0.998 (0.0003)	0.997 (0.0003)	1.000 (0.0003)	1.000 (0.0003)
$\boldsymbol{\psi}$ **(se)**	0.57 (0.020)	0.54 (0.019)	0.63 (0.023)	–	0.80 (0.037)	1.07 (0.049)	1.51 (0.079)	2.07 (0.120)	2.84 (0.140)
$\boldsymbol{maxEV}$	0.42	0.43	0.45	–	0.52	0.62	0.50	0.36	0.43
$\boldsymbol{l}_{\log}$	-22838.74	-23524.01	-21609.57	–	-15160.74	-13320.28	-9928.72	-8152.98	-8220.14

Table V.27

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\alpha^{(\lambda)}$ **(se)**	0.689 (0.010)	0.684 (0.010)	0.664 (0.009)	0.659 (0.009)	0.619 (0.010)	0.592 (0.011)	0.562 (0.013)	0.554 (0.015)	0.546 (0.016)
$\alpha^{(\phi)}$ **(se)**	0.034 (0.001)	0.033 (0.001)	0.036 (0.001)	0.040 (0.001)	0.034 (0.001)	0.036 (0.001)	0.037 (0.002)	0.029 (0.002)	0.028 (0.002)
$\alpha_t^{(\nu)}$ **(se)**	1.001 (0.000)	0.998 (0.000)	1.000 (0.000)	0.999 (0.000)	0.998 (0.000)	1.000 (0.000)	0.997 (0.000)	1.002 (0.000)	0.999 (0.000)
ψ **(se)**	0.885 (0.017)	0.896 (0.017)	1.081 (0.022)	1.115 (0.023)	1.404 (0.031)	1.908 (0.044)	2.283 (0.056)	3.117 (0.08)	3.437 (0.084)
$maxEV$	0.92	0.91	0.91	0.93	0.86	0.84	0.82	0.75	0.74
$l_{\log}$	-56785.72	-55467.43	-50532.84	-49398.59	-41915.59	-37613.81	-32238.02	-28764.63	-29287.5

Table V.28

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\alpha^{(\lambda)}$ **(se)**	0.361 (0.010)	0.34 (0.010)	0.355 (0.010)	0.328 (0.011)	0.355 (0.012)	0.374 (0.014)	0.313 (0.016)	0.27 (0.017)	0.334 (0.021)
$\alpha^{(\phi)}$ **(se)**	0.0021 (0.00)	0.0008 (0.000)	0.0003 (0.000)	0.000 (0.000)	0.000 (0.000)	0.0005 (0.001)	0.000 (0.000)	0.0018 (0.002)	0.0087 (0.002)
$\alpha_t^{(\nu)}$ **(se)**	1.002 (0.0002)	1.00 (0.0002)	0.999 (0.0002)	0.997 (0.0002)	0.999 (0.0002)	0.998 (0.0002)	0.996 (0.000)	1.00 (0.000)	1.00 (0.0003)
ψ **(se)**	0.57 (0.020)	0.54 (0.019)	0.63 (0.023)	0.72 (0.029)	0.81 (0.0370)	1.09 (0.05)	1.53 (0.080)	2.07 (0.120)	2.85 (0.150)
$maxEV$	0.37	0.34	0.36	0.33	0.35	0.38	0.31	0.28	0.38
$l_{\log}$	-22844.73	-23534.95	-21624.73	-18025.31	-15188	-13378.7	-9949.99	-8160.12	-8232.82

Table V.29

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\alpha^{(v)}t}$ (se)	1.002 (0.00008)	0.998 (0.00008)	1.000 (0.00009)	0.998 (0.00009)	0.9964 (0.00010)	1.000 (0.00011)	0.995 (0.00012)	1.002 (0.00013)	0.9983 (0.00014)
$\boldsymbol{\alpha^{(v)}A}$ (se)	0.233 (0.009)	0.268 (0.010)	0.380 (0.011)	0.408 (0.011)	0.385 (0.013)	0.455 (0.014)	0.413 (0.015)	0.470 (0.018)	0.512 (0.018)
$\boldsymbol{\alpha^{(v)}S}$ (se)	2.197 (0.004)	1.946 (0.005)	1.454 (0.007)	1.327 (0.008)	1.22 (0.010)	1.182 (0.011)	1.241 (0.011)	0.901 (0.017)	0.862 (0.017)
$\boldsymbol{\psi}$ (se)	0.66 (0.013)	0.81 (0.015)	0.91 (0.017)	0.77 (0.016)	0.95 (0.022)	1.15 (0.028)	1.07 (0.032)	1.69 (0.047)	2.09 (0.053)
$\boldsymbol{l_{\log}}$	-53170.2	-52828.3	-47577.9	-45679.9	-38460.1	-34061.5	-28453.8	-25652.1	-26474.3

Table V.30

Parameter	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\boldsymbol{\alpha^{(v)}t}$ (se)	1.001 (0.00012)	0.9996 (0.00011)	0.999 (0.00013)	0.9963 (0.00014)	0.9993 (0.00016)	0.9971 (0.00019)	0.9959 (0.00023)	0.9993 (0.00025)	1.002 (0.00030)
$\boldsymbol{\alpha^{(v)}A}$ (se)	0.239 (0.015)	0.280 (0.015)	0.335 (0.016)	0.312 (0.018)	0.365 (0.021)	0.368 (0.023)	0.273 (0.028)	0.322 (0.033)	0.478 (0.038)
$\boldsymbol{\alpha^{(v)}S}$ (se)	1.898 (0.008)	1.392 (0.010)	1.186 (0.013)	1.473 (0.012)	1.162 (0.017)	1.080 (0.020)	1.051 (0.026)	0.812 (0.038)	1.555 (0.023)
$\boldsymbol{\psi}$ (se)	0.93 (0.026)	0.81 (0.024)	0.98 (0.029)	1.09 (0.037)	1.45 (0.052)	1.97 (0.071)	2.50 (0.112)	3.25 (0.167)	6.02 (0.243)
$\boldsymbol{l_{\log}}$	-23945.0	-24450.4	-22563.0	-18764.1	-16088.6	-14205.1	-10552.5	-8611.7	-8885.4

IV. HHH Model Summaries (With Random Effects)

Table V.31 Estimated model parameters (with standard errors) and random effects for power-law + gravity RI HHH model formulation for measles spread in Lancashire time–windows.

Table V.32 Estimated model parameters (with standard errors) and random effects for power-law + gravity RI HHH model formulation for pertussis spread in Lancashire time–windows.

Table V.33 Estimated model parameters (with standard errors) and random effects for power-law + gravity RI HHH model formulation for scarlet fever spread in Lancashire time–windows.

Table V.31

	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\beta^{(\phi)}_{\log(pop)}$	0.432	0.273	0.269	0.341	0.450	0.576	0.524	0.505	0.572
	(0.053)	(0.059)	(0.059)	(0.062)	(0.058)	(0.059)	(0.057)	(0.056)	(0.053)
d **(se)**	3.18	3.56	4.30	4.56	3.46	3.43	3.22	3.16	3.17
	(0.086)	(0.100)	(0.123)	(0.129)	(0.082)	(0.085)	(0.100)	(0.083)	(0.089)
ψ **(se)**	0.93	0.89	0.81	0.85	0.93	0.99	0.96	0.90	1.01
	(0.013)	(0.014)	(0.012)	(0.013)	(0.014)	(0.015)	(0.014)	(0.012)	(0.014)
$maxEV$	0.86 –	0.87 –	0.83 –	0.83 –	0.84 –	0.88 –	0.86 –	0.99 –	0.86 –
	1.03	1.20	1.11	1.10	1.11	1.06	1.09	1.26	1.05
Fixed									
$\alpha^{(\lambda)}$ **(se)**	0.658	0.662	0.676	0.677(0.	0.708	0.749	0.749	0.691	0.623
	(0.016)	(0.018)	(0.018)	015)	(0.014)	(0.012)	(0.014)	(0.015)	(0.016)
$\alpha^{(\phi)}$**(se)**	0.0020	0.0090	0.0094	0.0048	0.0015	0.0004	0.0006	0.0010	0.0006
	(0.001)	(0.005)	(0.005)	(0.003)	(0.001)	(0.0002)	(0.0003)	(0.0005)	(0.0003)
Random									
σ^2_λ	0.046	0.064	0.061	0.041	0.028	0.012	0.020	0.035	0.053
σ^2_ϕ	0.543	0.437	0.459	0.498	0.421	0.402	0.344	0.361	0.353
σ^2_v	1.375	1.207	1.015	0.777	0.999	1.33	0.495	0.742	1.085
ρ^2_λ	-0.008	-0.112	-0.117	0.043	-0.110	-0.174	-0.225	-0.110	-0.188
ρ^2_ϕ	-0.024	0.169	0.251	0.182	-0.009	-0.063	-0.240	-0.225	-0.118
ρ^2_v	-0.407	-0.229	-0.271	-0.277	-0.114	0.006	-0.054	-0.199	-0.352
l_{pen}	-	-	-	-	-	-	-	-	-
	67928.2	62648.8	66430.6	67333.6	67302.2	66158.7	68547.3	74093.9	69008.4
l_{mar}	-549.91	-551.57	-557.98	-528.13	-503.57	-474.65	-455.34	-505.77	-502.08

Table V.32

	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\beta^{(\phi)}_{\log(pop)}$	0.522	0.479	0.039	0.093	0.224	0.859	0.813	1.013	0.897
	(0.063)	(0.070)	(0.080)	(0.070)	(0.076)	(0.058)	(0.073)	(0.071)	(0.066)
d **(se)**	1.99	2.38	2.80	3.46	2.78	2.01	2.09	1.09	0.89
	(0.098)	(0.115)	(0.121)	(0.150)	(0.129)	(0.103)	(0.094)	(0.088)	(0.094)
ψ **(se)**	1.05	1.05	0.70	0.66	0.88	1.11	1.93	2.10	2.19
	(0.021)	(0.021)	(0.014)	(0.012)	(0.017)	(0.025)	(0.053)	(0.069)	(0.069)
$maxEV$	0.88 –	0.79 –	0.90 –	0.82 –	0.90 –	0.87 –	0.80 –	0.94 –	0.96 –
	1.04	0.97	0.97	0.99	1.00	1.12	1.01	1.06	1.00
Fixed									
$\alpha^{(\lambda)}$ **(se)**	0.356	0.335	0.368	0.404	0.402	0.334	0.264	0.193	0.153
	(0.018)	(0.015)	(0.017)	(0.016)	(0.019)	(0.019)	(0.017)	(0.015)	(0.014)
$\alpha^{(\phi)}$**(se)**	0.0010	0.0014	0.08600	0.0544	0.01619	0.00003	0.00006	0.00001	0.00003
	(0.0006)	(0.0010)	(0.0671)	(0.03735)	(0.0121)	(0.00002)	(0.00005)	(0.00001)	(0.00002)
Random									
σ^2_λ	0.221	0.169	0.205	0.154	0.219	0.280	0.253	0.221	0.169
σ^2_ϕ	0.735	0.892	1.228	0.865	0.952	0.413	0.682	0.735	0.892
σ^2_v	1.308	0.957	1.603	1.826	0.787	2.122	1.217	1.308	0.957
ρ^2_λ	0.369	0.316	0.612	0.467	0.392	-0.042	0.225	0.369	0.316
ρ^2_ϕ	-0.274	0.031	-0.353	-0.206	-0.120	0.135	0.128	-0.274	0.031
ρ^2_v	0.051	-0.350	-0.559	-0.481	-0.353	-0.427	-0.295	0.051	-0.350
l_{pen}	-	-		-	-	-	-	-	-
	46025.0	46516.2	-538407	56291.7	49501.2	38799.3	23994.2	46025.0	46516.2
l_{mar}	-586.88	-537.85	-580.49	-593.85	-581.54	-501.62	-424.73	-586.88	-537.85

Table V.33

	Model estimates								
	40–45	**43–48**	**46–51**	**49–54**	**52–57**	**55–60**	**58–63**	**61–66**	**64–69**
$\beta^{(\phi)}_{\log(pop)}$	0.510	0.658	0.544	0.591	0.537	0.766	0.963	1.018	0.94
	(0.037)	(0.035)	(0.044)	(0.053)	(0.082)	(0.066)	(0.056)	(0.071)	(0.064)
d **(se)**	1.72	1.84	2.27	1.81	2.08	1.65	0.84	0.98	0.91
	(0.073)	(0.068)	(0.108)	(0.113)	(0.181)	(0.138)	(0.151)	(0.142)	(0.157)
ψ **(se)**	0.24	0.24	0.31	0.32	0.35	0.45	0.49	0.71	0.91
	(0.008)	(0.008)	(0.010)	(0.010)	(0.013)	(0.018)	(0.021)	(0.028)	(0.031)
$maxEV$	0.70 –	0.86 –	0.70 –	0.53 –	0.55 –	0.48 –	0.43 –	0.47 –	0.40 –
	1.07	1.01	0.88	0.87	0.85	0.86	0.94	0.93	1.02
Fixed									
$\alpha^{(\lambda)}$ **(se)**	0.229	0.227	0.231	0.228	0.205	0.195	0.165	0.129	0.134
	(0.012)	(0.012)	(0.013)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)	(0.011)
$\alpha^{(\phi)}$**(se)**	0.0019	0.0004	0.0011	0.0006	0.0007	0.0001	0.00001	0.00001	0.00001
	(0.0007)	(0.0001)	(0.0005)	(0.0003)	(0.0006)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Random									
σ^2_λ	0.265	0.268	0.284	0.222	0.297	0.250	0.250	0.448	0.379
σ^2_ϕ	0.385	0.253	0.263	0.349	0.521	0.431	0.464	0.701	0.727
σ^2_v	2.222	1.391	0.415	0.274	0.324	0.498	0.582	0.794	0.748
ρ^2_λ	0.742	0.617	0.529	0.479	0.377	0.406	0.063	0.314	0.336
ρ^2_ϕ	-0.277	-0.164	0.250	0.165	0.161	0.199	0.252	0.390	0.426
ρ^2_v	-0.310	-0.293	0.097	0.129	0.205	0.305	0.753	0.693	0.799
l_{pen}	-	-	-	-	-	-	-	-	-
	48922.1	47700.6	43206.7	42230.3	35209.8	31139.4	26252.9	23212.7	23956.3
l_{mar}	-535.59	-543.32	-475.93	-483.76	-494.31	-463.76	-432.14	-444.13	-436.78

Table V.34 *Goodness-of -fit assessments for HHH model formulations using AIC values, for measles in Lancashire. Values highlighted in yellow indicate best fitting model.*

Time Window	Model				
	Endemic	**First-order**	**Gravity**	**Power Law**	**PL + Gravity**
40–45	162095.9	139552.5	139545.6	138356.6	138010.5
43–48	152138.0	128917.3	128903.8	127877.7	127609.1
46–51	161451.6	136231.7	136220.8	135402.7	135199.8
49–54	161780.7	137870.4	137871.8	137124.2	136905.4
52–57	161996.3	138332.2	138332.7	137181.6	136862.2
55–60	159480.5	135885.8	135887.7	134954.2	134569.5
58–63	163593.2	140221.0	140218.9	139351.4	138952.8
61–66	175176.0	151815.1	151813.4	150721.0	150373.9
64–69	160179.9	141295.7	141275.9	140562.4	140013.8
Mean	**161974.5**	**138821.2**	**138815.6**	**137896.9**	**137560.8**

Table V.35 *Goodness-of -fit assessments for HHH model formulations using AIC values, for measles in South Wales. Values highlighted in yellow indicate best fitting model.*

Time Window	Model				
	Endemic	**First-order**	**Gravity**	**Power Law**	**Gravity + PL**
40–45	76983.5	68018.7	67767.0	64950.1	64914.7
43–48	71281.9	62926.4	62639.8	60150.2	60088.9
46–51	77555.7	66667.3	66487.2	64603.2	64537.9
49–54	73378.7	62570.2	62496.6	60594.4	60547.4
52–57	72709.5	61680.9	61618.8	59613.8	59604.2
55–60	74222.6	63962.2	63880.4	62019.4	62001.2
58–63	72103.9	63445.6	63241.0	60715.7	60697.8
61–66	80557.4	72285.5	71935.4	70171.7	70107.8
64–69	70695.7	63688.5	63520.5	61797.6	61799.0
Mean	**74387.7**	**65027.3**	**64843.0**	**62735.1**	**62699.9**

Table V.36 *Goodness-of -fit assessments for HHH model formulations using AIC values, for pertussis in Lancashire. Values highlighted in yellow indicate best fitting model.*

Time Window	Model				
	Endemic	**First-order**	**Gravity**	**Power Law**	**PL + Gravity**
40–45	103656.0	95742.3	95698.2	95309.8	95051.6
43–48	102183.0	95877.1	95851.8	95595.4	95572.4
46–51	123698.5	111349.8	111341.9	110905.5	110799.5
49–54	130027.8	116072.4	116068.3	115710.1	115578.4
52–57	113097.1	101869.6	101868.4	101668.4	101538.6
55–60	89561.2	80365.5	80365.9	80091.8	79743.2
58–63	54226.1	50562.1	50524.5	50060.5	49473.1
61–66	42238.9	39452.6	39391.3	39107.0	38683.1
64–69	43632.4	41677. 9	41627.7	41422.1	41161.4
Mean	**89146.9**	**81441.0**	**81415.3**	**81096.7**	**80844.6**

Table V.37 *Goodness-of -fit assessments for HHH model formulations using AIC values, for pertussis in South Wales. Values highlighted in yellow indicate best fitting model.*

Time Window	Model				
	Endemic	**First-order**	**Gravity**	**Power Law**	**Gravity + PL**
40–45	39841.6	37224.2	37206.3	37040.7	37017.0
43–48	39109.4	36751.8	36748.2	36655.1	36603.5
46–51	45393.7	42730.9	42732.4	42397.0	42159.8
49–54	48850.8	46023.4	46000.1	45462.2	45211.4
52–57	42477.0	40282.3	40263.7	40123.4	39960.2
55–60	29643.6	27888.5	27861.4	27766.9	27604.2
58–63	17187.8	16104.3	16036.7	15946.2	15738.1
61–66	14606.5	13801.2	13758.8	13591.9	13321.2
64–69	14924.9	14142.5	14108.2	14024.6	13839.0
Mean	**32448.4**	**30549.9**	**30524.0**	**30334.2**	**30161.6**

Table V.38 *Goodness-of -fit assessments for HHH model formulations using AIC values, for scarlet fever in Lancashire. Values highlighted in yellow indicate best fitting model.*

Time Window	Model				
	Endemic	**First-order**	**Gravity**	**Power Law**	**PL + Gravity**
40–45	106350.4	100756.0	100719.5	100757.0	100446.6
43–48	105666.6	98789.4	98726.1	98757.7	98172.1
46–51	95165.7	89060.3	89008.6	89027.1	88702.2
49–54	91369.9	86702.3	86692.7	86753.4	86577.2
52–57	76930.2	72723.1	72721.9	72762.5	72714.6
55–60	68133.0	64512.2	64511.8	64547.8	64423.2
58–63	56917.6	54712.4	54714.4	54782.3	54669.2
61–66	51314.2	49087.9	49080.5	49116.2	49041.1
64–69	52958.5	50739.6	50730.5	50719.3	50635.0
Mean	**78311.8**	**74120.4**	**74100.7**	**74135.9**	**73931.3**

Table V.39 *Goodness-of -fit assessments for HHH model formulations using AIC values, for scarlet fever in South Wales. Values highlighted in yellow indicate best fitting model.*

Time Window	Model				
	Endemic	**First-order**	**Gravity**	**Power Law**	**Gravity + PL**
40–45	47900.1	45703.5	45693.5	45707.6	*45616.4*
43–48	48910.8	47083.9	47064.0	47087.8	*46816.4*
46–51	45135.9	43263.5	43235.2	43265.6	*42936.8*
49–54	37538.1	36064.6	–	–	–
52–57	32187.1	30390.0	30337.5	30389.0	*30182.0*
55–60	28420.2	26771.4	26656.6	26765.4	*26506.4*
58–63	21115.1	19914.0	19873.5	–	*19796.5*
61–66	17233.3	16334.3	16321.9	16335.0	*16284.2*
64–69	17780.7	16479.6	16456.3	16435.3	*16350.4*
Mean	**32913.5**	**31333.9**	**30704.8**	**32283.7**	***30561.1***

Table V.40 *Assessment of predictive performance of HHH model formulations for measles in Lancashire, using mean scores. Values highlighted in yellow indicate best performing model fits according to proper scoring rules.*

	Predictive Score Assessment											
	Endemic		First-order		Gravity		Power Law		PL + Gravity		PL + Gravity RI	
TW	LogS	RPS	LogS	RPS	LogS	RPS	LogS	RPS	LogS	RPS	LogS	RPS
40–45	1.587	4.087***	2.245*	2.082*	2.332**	2.168*	2.443*	1.980*	**0.900**	**0.415**	1.803	1.285
43–48	2.147*	2.874**	1.656	1.140	1.699	1.186	0.824	**0.178**	**0.117**	0.814	0.685	0.413
46–51	10.439***	6.962***	3.053**	4.810***	2.886**	4.639***	1.337	**3.345*** **	**1.540**	3.666***	2.623**	6.178***
49–54	6.794***	5.303***	2.461**	1.519	2.463**	1.525	3.536***	2.474*	3.398***	2.065*	**1.884**	**1.271**
52–57	13.774***	4.945***	4.002***	4.217***	3.941***	4.136***	1.514	1.819*	1.457	1.960*	**1.221**	**1.471**
55–60	21.293***	12.433***	9.163***	6.828***	9.162***	6.826***	6.332***	4.620***	3.761***	2.153*	**2.729****	**1.673**
58–63	19.183***	12.086***	3.891***	5.081***	3.988***	5.194***	2.238*	3.180***	**2.013***	**2.994****	3.840***	4.781***
61–66	18.331***	11.584***	6.170***	3.152**	6.174***	3.169**	4.030***	1.164	**2.055***	**0.581**	3.581***	1.464
64–69	1.614	2.563**	0.488	**0.005**	0.416*	0.107	0.997	0.512	1.075	0.380	0.663	1.001
Mean	10.574	6.982	3.681	3.204	3.673	3.217	2.583	2.141	**1.813**	**1.670**	2.114	2.171

*** p < 0.05, ** p< 0.01, *** p < 0.001**

Table V.41 *Assessment of predictive performance of HHH model formulations for pertussis in Lancashire, using mean scores. Values highlighted in yellow indicate best performing model fits according to proper scoring rules.*

	Predictive Score Assessment											
	Endemic		First-order		Gravity		Power Law		PL + Gravity		PL + Gravity RI	
TW	LogS	RPS	LogS	RPS	LogS	RPS	LogS	RPS	LogS	RPS	LogS	RPS
40–45	20.086***	19.045***	13.401***	11.423***	12.668***	11.164***	6.989***	5.632***	2.688**	1.035	**0.905**	**1.147**
43–48	0.174	0.782	1.753	2.644	1.580	2.466*	0.504	1.369	0.554	1.442	1.720	2.523**
46–51	4.611***	**3.015****	4.801***	4.147***	4.847***	4.198***	5.398***	4.842***	5.093***	4.528***	**3.742*****	3.267***
49–54	3.750***	3.234***	1.177	1.548	1.089	1.451	1.120	**1.551**	1.678	2.155*	2.919**	3.610***
52–57	4.087***	2.008*	0.942	0.276	0.882	**0.221**	**0.390**	0.940	1.555	1.916	0.710	0.572
55–60	6.092***	5.086***	3.971***	3.578***	4.017***	3.625***	3.001**	2.800**	**1.599**	**1.325**	2.208*	2.484**
58–63	1.721	0.615	0.826***	0.554	0.721	**0.415**	**0.353**	0.521	1.612	1.741	1.433	1.721
61–66	1.625	1.518	1.523***	1.495	1.496	1.474	–	–	1.538	1.511	**0.896**	**0.780**
64–69	6.304***	5.708***	4.575***	4.038***	4.424***	3.859***	3.903***	3.433***	1.455	0.992	0.465	0.913
Mean	**2.558**	**2.281**	**1.241**	**1.149**	**1.132**	**1.090**	**0.196**	**0.202**	**0.526**	**0.526**	**0.159**	**0.482**

*** p < 0.05, ** p< 0.01, *** p < 0.001**

Table V.42 *Assessment of predictive performance of HHH model formulations for scarlet fever in Lancashire, using mean scores. Values highlighted in yellow indicate best performing model fits according to proper scoring rules.*

	Predictive Score Assessment											
	Endemic		First-order		Gravity		Power Law		PL + Gravity		PL + Gravity RI	
TW	LogS	RPS	LogS	RPS	LogS	RPS	LogS	RPS	LogS	RPS	LogS	RPS
40–45	0.129	0.477	2.013*	1.772	0.343	0.712	0.028	0.324	0.649	0.052	0.327	0.665
43–48	0.552	0.562	2.552*	0.788	0.203	0.182	0.842	0.874	1.243	0.929	1.311	1.061
46–51	3.140**	2.560***	4.379***	3.734***	2.943**	2.372*	3.069**	2.525*	1.634	1.248	1.354	0.89
49–54	3.219***	2.785***	5.269***	5.409***	3.132**	2.683**	3.244***	2.828**	2.471*	2.083*	2.975**	2.424*
52–57	2.652**	3.061***	5.345***	5.195***	2.732**	3.159**	-	-	2.886**	3.292***	-	-
55–60	1.145	1.188	1.404	1.155	1.203	1.248	1.013	1.045	1.479	1.523	0.697	0.927
58–63	4.413***	4.002***	6.410***	5.728***	4.411***	4.000***	4.297***	3.922***	2.336*	2.121*	2.140*	2.115*
61–66	3.580***	3.258***	4.865***	4.428***	3.468**	3.144**	3.711***	3.390***	1.276	1.039	1.084	0.896
64–69	2.918**	2.757***	3.485***	2.764**	3.012**	2.865**	2.462*	2.305*	1.668	1.566	0.664	0.561
Mean	**2.416**	**2.294**	**3.969**	**3.441**	**2.383**	**2.263**	**2.333**	**2.152**	**1.738**	**1.539**	**1.319**	**1.192**

*** p < 0.05, ** p< 0.01, *** p < 0.001**