Using Syndromic Surveillance to Explore Respiratory Illness in the Community

By

Kirsty Morrison

Thesis
Submitted for the degree of
Doctor of Philosophy
to the University of East Anglia, School of Environmental Sciences

Supervisory committee:
Prof. Iain R. Lake, Dr. Felipe J. Colón-González, Prof. Paul R. Hunter and Dr. Gillian E. Smith

May 2022

Abstract

Syndromic surveillance is the "near real-time collection, analysis and interpretation" of health data. It is based on non-specific, pre-diagnostic signs and symptoms of disease, which can be used as an indicator for specific illnesses. In England, the Real-time Syndromic Surveillance Team (ReSST) operate a suit of national syndromic surveillance systems for early detection of outbreaks, situational awareness and reassurance to the lack of threat to public health.

Using several spatial and temporal statistical methods we highlight how this unique and comprehensive syndromic dataset can be used in observational epidemiological studies. In this thesis we used this data to explore demographic and socioeconomic patterns in healthcare-seeking behaviour for respiratory symptoms; estimated the community burden of healthcare presentations attributable to respiratory syncytial virus (RSV) in children; explored regional differences in the seasonality of RSV, and to explore the relationship between meteorological conditions and acute respiratory infections in children under-5 years. We utilised the frequency at which data is available, and the granularity of the local geography to explore healthcare usage in ways not previously explored. By focusing on healthcare services that provide healthcare in the community we were able to investigate a wider burden of disease than data from acute services, such as hospitalisations.

We successfully used data from syndromic surveillance in a variety of observational spatial and temporal epidemiological studies. These studies highlight that this data can provide similar observations to those from hospitalisation and laboratory data, but also provides a unique insight into healthcare-seeking behaviour in the community, which is often poorly defined. Although this data has been used successfully, this research highlights the limitations of data from syndromic surveillance. Data from syndromic surveillance has enormous potential for a variety of epidemiological research designs; however, data needs to be used within its limitations and the principles of syndromic surveillance.

Contents

List of Tables

List of Figures

Dedicated to Dr. Andrew Morrison, my dad

Acknowledgements

I am truly grateful for the guidance and support from my supervisory team, Iain R. Lake, Felipe J. Colón-González, Paul R. Hunter and, Gillian E. Smith for the expertise they have provided me throughout this project. I am particularly thankful to Iain R. Lake for the patience he has shown me, and the time and expertise he has dedicated to me to improve the quality of my work. I am also truly grateful to Felipe J Colón-González for providing me his expertise in R and R programming, and support in the development of the statistical models used in this project. Both Iain R. Lake and Felipe J Colón-González have provided me with a vast amount encouragement through this journey, for which I am sincerely grateful.

I would also like to show my gratitude to the real-time syndromic team who provided the data for this thesis and technical assistance on the syndromic surveillance systems. A special thanks to Gillian, E. Smith, Alex. J. Elliot and Roger A. Morbey for their guidance on how to effectively use the syndromic data in this project, and their experience in interpreting the outcomes from this project. I am also grateful to Sue Smith and Sally Harcourt for their guidance on the GP in-hours and GP-out-of-hours surveillance systems.

For their assistance on helping me publish my research, I am grateful to Judith Rutter, Gareth Stuttard, Simon de Lusignan, Alex Yeates and Richard Pebody. Their expertise and guidance greatly improved my work and scientific presentation.

Finally, I am thankful to my family for their love and encouragement through this journey. I am particularly grateful to my partner Will for his unconditional love, comfort and support through the ups and downs that we have gone through, and his guidance in life, I could not have done this PhD without him. To my son Lewis, I will forever be grateful for the joy you have brought me.

Chapter 1
Introduction

1.1 Syndromic Surveillance in Public Health Practice

What is Syndromic Surveillance?

Syndromic surveillance is the "near real-time collection, analysis and interpretation" of health data which allows for the timely dissemination of information about potential public health threats (Triple-S Project 2011). It is based on non-specific, pre-diagnostic signs and symptoms of disease, which can be used as an indicator for specific illness. The primary role of syndromic surveillance is to provide the early identification of public health threats, with the aim of; detecting emerging health threats, providing reassurance that there is no public health impact, assessing the impact of a public health event and detecting the start of expected events.

Syndromic surveillance systems monitor data sources, such as school absenteeism, internet searches, telehealth services, primary care consultations and hospital admissions for certain syndromes that may be indicative of an illness or pathogen. Syndromic surveillance focuses on indicators of disease rather than confirmed diagnosis with the aim of detecting public health threats before traditional surveillance systems. It exploits steps in healthcare-seeking behaviour that are undertaken before laboratory or clinical confirmation of disease takes place. A variety of general and specific syndromic indicators are used in these systems, which can be used in the detection of a range of diseases. Public Health England (PHE) have found that syndromes such as *Influenza-Like-Illness* are associated with influenza (Harcourt et al. 2012a), and *Cough* and *Bronchitis* are associated with respiratory syncytial virus (RSV) in children under-5 years (Cooper et al. 2007; Morbey et al. 2017a; 2017b). Although the data syndromic surveillance uses are non-specific, the rapid and sensitive nature of the data can

supplement traditional laboratory surveillance to help estimate the burden and impact of disease in the community.

Data can be obtained continuously in daily transmissions for real-time data collection, analysis and interpretation (Triple-S Project 2011). Traditional forms of surveillance, such as laboratory confirmation, may be affected by closures during holidays or delays due to increased workload during seasonal disease outbreaks (Buckingham-Jeffery et al. 2017). Syndromic surveillance resources have automated electronic transfer of information from its sources, so these delays are eliminated. These aspects of syndromic surveillance allow for the near-real time, continuous collection of pre-diagnostic health data, which can result in the rapid detection of health threats (Triple-S Project 2013).

Timeliness is a key factor in detection of an outbreak to contain and prevent further cases and minimise the health and economic impact of an outbreak. By utilising services that are used when an individual first becomes ill, outbreaks could be identified earlier than through traditional surveillance methods. Zeng and Wagner (2002) developed a model based on people's behaviour when they become ill, which consists of four phases; recognition of symptoms, interpretation of symptoms, cognitive representation of illness, and seeking of treatment. It was proposed that these phases could be targeted for outbreak detection. During the first phase, when a person first recognises their symptoms, they might stay home from work or school suggesting that absenteeism could be used as a proxy indicator for ill health. During the second and third phase an individual may seek to understand their symptoms through web searches or telehealth services, and finally in the fourth stage an individual will seek treatment for their illness (Hulth, Rydevik and Linde 2009; Zeng and Wagner 2002) (Figure 1.1). In reality, these phases are not that distinct, human behaviour is complex and can vary due to multiple factors such as severity of symptoms, onset time, and case demographics (Peppa, Edmunds and Funk 2017). However, this theory of early phases of healthcare-seeking behaviour allows for the basis of using syndromic surveillance as an early detection of public health threats. Ziemann et al. (2016) assessed the timeliness of nineteen of the syndromic surveillance systems in current use in Europe. Out of the nine systems used to detect the onset or peak of an influenza season 75% (n=seven) earlier than traditional

methods of surveillance, with a range of timeliness between 2.5 weeks to two weeks (mean 0.75 weeks). Of the eight systems used for outbreak detection, four detected one or more events within one day, with timeliness defined as detection within three days. The four other systems detected the events within four to fourteen days. A study conducted in the Netherlands assessed the difference in timeliness between syndromic and traditional forms of surveillance and pathogen counts. Absenteeism detected RSV two weeks and influenza four to five weeks ahead of pathogen counts. Pharmacy prescriptions detected RSV one week behind and influenza, zero to two weeks ahead of pathogen counts. Primary care consultations detected RSV one week behind and influenza one to two weeks ahead of pathogen counts. Hospitalisations were in concurrence with pathogen counts for RSV and one to two weeks ahead for influenza (van den Wijngaard et al. 2011).

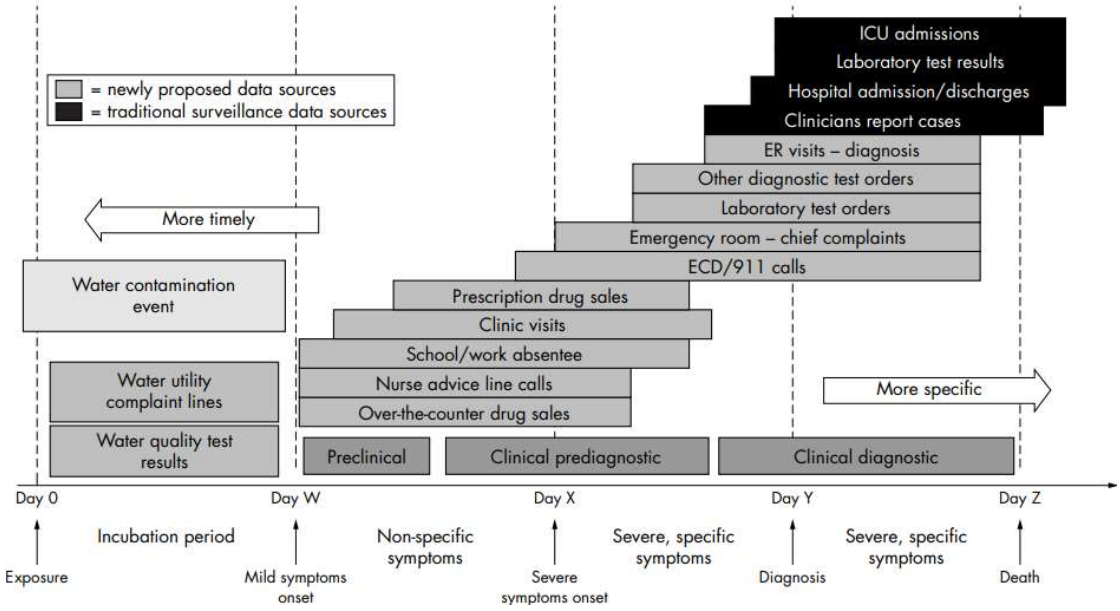

Figure 1.1: Surveillance data sources and healthcare-seeking behaviour (Berger, Shiau and Weintraub 2006).

Reliable and effective surveillance systems are imperative for the detection and monitoring of public health threats and disease outbreaks. They also provide vital information to support public health planning, and the targeting and evaluation of health interventions (Keramarou and Evans 2012). Although traditional forms of surveillance based on clinical diagnosis, such as disease notification and laboratory reports, are seen as the gold standard, they often underestimate the true burden of disease (Wheeler et al. 1999). This is particularly evident in infectious diseases where factors such as asymptomatic carriers and a self-limiting course of infection may affect the number of cases which seek healthcare and are included in disease burden estimates (Gibbons et al. 2014). It is estimated that of all infections 10% are asymptomatic, 40% do not attend healthcare services and 45% are underdiagnosed and under notified (Gibbons et al. 2014) (Figure 1.2). By utilising alternative data sources which are used before a clinical diagnosis is made or if no healthcare is sought, not only is there the possibility of early detection of disease outbreaks but, there is also the opportunity to estimate the true burden of disease. However, syndromic surveillance will not detect asymptomatic cases, which is an important part of disease transmission dynamics (Chisholm et al. 2018).

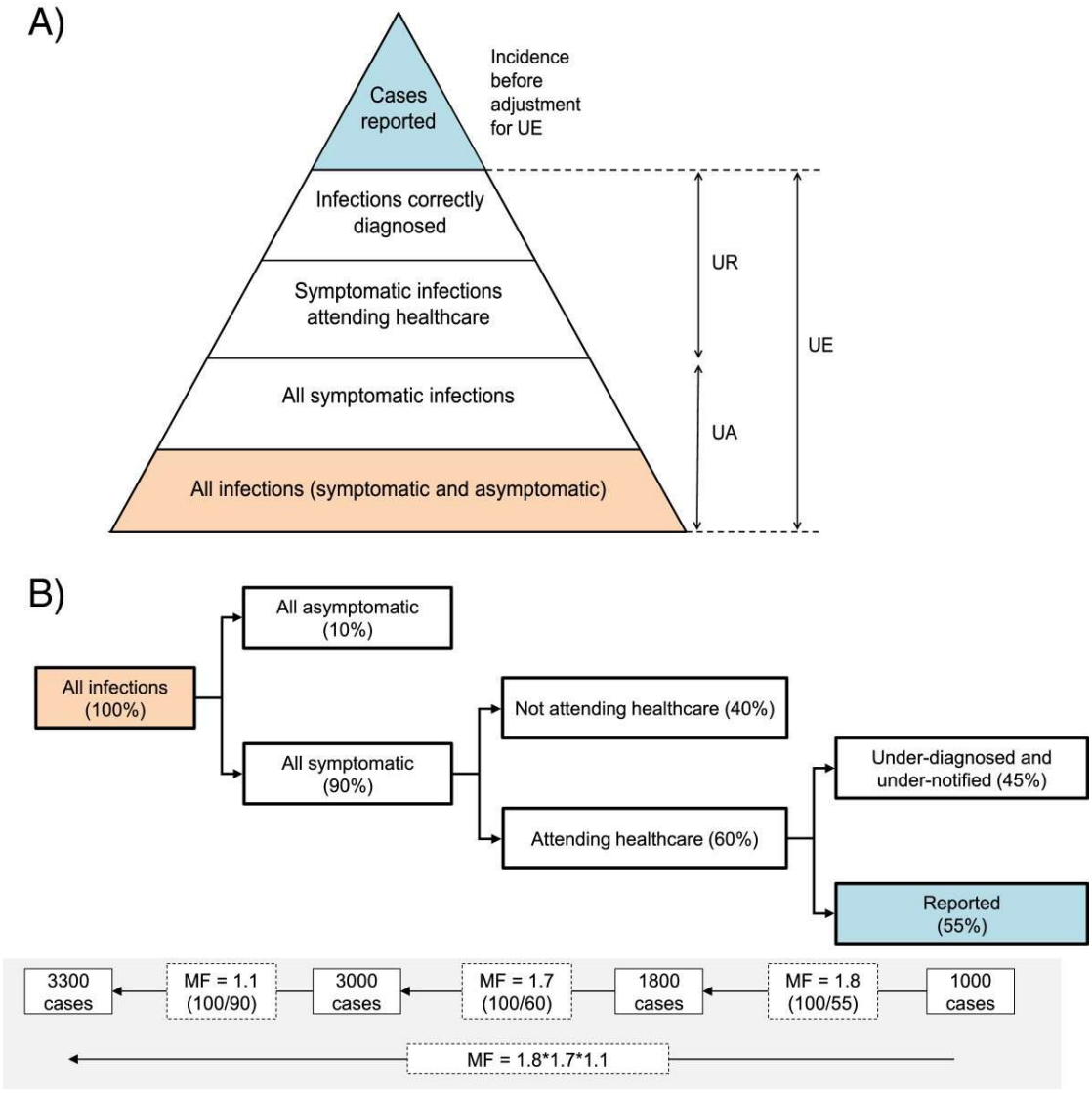

Figure 1.2: Morbidity surveillance pyramid estimating number of infection cases reported. A: the morbidity surveillance pyramid is often used to illustrate the availability of morbidity data at each surveillance level. B: the proportions of infections that are symptomatic, that attend healthcare, and that are reported are represented in this decision tree model (Gibbons et al. 2014).

One of the main limitations of syndromic surveillance is that it does not provide a diagnosis of illness, but instead relies on the detection and identification of signs and symptoms that may be associated with a specific disease. It therefore only provides an early warning of syndromes and cannot identify the responsible pathogen or disease, although these can be indicative the illness or pathogen. I infectious diseases, depending

on the symptom and pathogen, syndromic surveillance has been closely associated with pathogen counts. An evaluation of syndromic surveillance in the Netherlands found that up to 86% of syndromes associated with respiratory diseases could be explained by variations in weekly respiratory pathogen counts. Whereas only up to 62% of neurological syndromes, and up to 40% of gastroenteritis syndromes could be explained by the pathogen counts. In the case of gastroenteritis syndromes, this variation increased to 85% when limiting the analysis to young children (van Asten et al. 2007; van den Wijngaard et al. 2011).

The Evolving Definition of Syndromic Surveillance

One of the earliest examples of syndromic surveillance in the literature was by Robertson et al. (1994), where syndromic surveillance of acute flaccid paralysis was used to detect and identify cases of acute paralytic poliomyelitis in Oman. Since this early example of syndromic surveillance, its utility for detecting and monitoring disease threats has adapted to reflect the evolving requirements of disease surveillance.

Syndromic surveillance first gained traction as novel mechanism for surveillance of emerging threats after the 9/11 terrorist attacks and deliberate anthrax distribution in the United States of America (USA) in 2001. Prior to these attacks, syndromic surveillance systems were not common but the need for novel mechanisms allowing early detection of emerging public health threats, such as bio-terrorism attacks, had been identified (CDC 1998; Khan, Levitt and Sage 2000). In post-9/11 USA, syndromic surveillance was intended for the early detection of chemical or bio-terrorism attacks. The idea was that health data sources, such as absenteeism, over the counter drug use and physician visits could be utilised to detect higher than expected indicators for disease before the pathogen was confirmed, resulting in a quicker response to the threat. Other factors that can influence the scale of a bio-terrorist attack include: geographical distribution, time of release, time of exposure, host behaviours and environmental factors. It was therefore identified that new surveillance mechanisms for the early detection of attacks would need to have wide coverage, be flexible and utilise syndrome-based data sources (Henning 2003). By 2003, over 100 USA health jurisdictions were using syndromic surveillance to enhance their existing surveillance methods (Buehler et al. 2003).

Eventually, as the immediate threat of biological and chemical terrorism diminished, interest in syndromic surveillance shifted to detect emerging infections (Paterson and Durrheim 2013). After the severe acute respiratory syndrome (SARS) outbreak in 2002, public health activities changed to focus on emerging diseases and pandemics. This was further confounded by the emergence of avian influenza in 2008 and H1N1 influenza pandemic in 2009. Emerging diseases often present with specific syndromes such as influenza like illness. Therefore, through surveillance of these syndromes, unexpected increases which might indicate an emerging disease can be investigated in a timely manner. This type of surveillance is particularly important for emerging diseases because laboratories may lack the diagnostic methods to detect novel infections. This shift in focus is highlighted by Paterson and Durrheim (2013) who noted that the number of syndromic surveillance publications relating to influenza and outbreaks increased between 2005 and 2011, reflecting public health concerns at the time. During emerging disease outbreaks, syndromic surveillance can also serve as a tool to monitor and estimate community burden of disease in situations where laboratories lack the capacity to confirm every case (Elliot 2009a).

Over time, syndromic surveillance has shifted its definition from primarily detecting public health threats to also to fill another area, one of "situational awareness". Situational awareness is used to describe the ability to characterise and monitor an outbreak or public health threat regardless of when and how it was detected (Buehler et al. 2008), and to then inform decision makers during events that impact public health (Velsko and Bates 2016). This follows on from outbreak detection to further monitor and characterise the disease outbreak along with traditional forms of surveillance to complete the "epidemiological puzzle" to fully understand the impact on public health (Buehler et al. 2008; CDC 2010; Paterson and Durrheim 2013). Syndromic surveillance can also be used to monitor public health trends of endemic and seasonal communicable diseases and non-communicable disease events. Continued syndromic surveillance of disease symptoms can also allow for an interventions effect on disease and healthcare burden to be evaluated. Bawa et al. (2015) used syndromic surveillance to assess the impact of the rotavirus vaccine in England; it was found that the vaccine resulted in a

26%-33% reduction of incidence among children under-1 year and a statistically significant reduction in GP (general practitioner) in and out-of-hours consultations.

There is a growing amount of literature on the use of syndromic surveillance for situational awareness of non-communicable disease events. It has been used to investigate the impact, or lack thereof, of mass-gathering events, severe weather events and emergencies on public health and health services. During the 2015 Super Bowl in Glendale, Arizona syndromic surveillance found an increase in influenza like activity, a single event of cyanide poisoning and sporadic cases of gastrointestinal and neurological diseases (Ayala et al. 2016; G. E. Smith et al. 2016; Todkill et al. 2016). In 2013, syndromic surveillance was used by PHE to assess the burden of illness associated with a heatwave that occurred in England. GP consultations for heat related illness were found to be double that of non-heatwave years, (Elliot et al. 2016a; Hughes et al. 2014; S. Smith et al. 2016a; Tsai et al. 2016; Vilain et al. 2015). Between 2011 and 2013, syndromic surveillance was used to detect potential health emergencies after an influx of migrants to Italy. During this period of surveillance 20 statistical alarms were investigated, three of which were in relation to a scabies, the only outbreak detected (Napoli et al. 2014; Riccardo et al. 2011). In 2015, syndromic surveillance detected visits to emergency departments (ED) related to stress, post-traumatic stress disorder, anxiety disorder and acute stress reaction, increased significantly after the November terrorist attack in Paris (Vandentorren et al. 2016).

Since the concept of syndromic surveillance became popular in the early 2000's, it has developed from a novel mechanism for early detection of biological or chemical terrorist attacks to a tool which is used alongside other forms of surveillance for the situational awareness of public health. It has adapted to constantly shifting public health concerns and needs; which can be largely contributed to its flexibility, timeliness and cost effectiveness (Ziemann et al. 2015).

1.2 Syndromic Data

1.2.1 Data Collection

Syndromic surveillance, is an electronic data driven approach to public health surveillance. Data for is evaluated on its ability to be available in a timely manner, low cost, flexible manner to changing needs, data completeness and validity (Ziemann and Krafft 2013).

Syndromic surveillance relies on the timely collection of non-specific health indicators, detection of anomalies in the data and accurate interpretation of the threat of these anomalies. For anomalies to be detected, geographic, demographic and temporal coverage must be sufficient and the appropriate data should be available in a timely and consistent manner (Mandl et al. 2004). Syndromic surveillance resources often take advantage of systems that collect data for other purposes, which reduces the need for new systems to be implemented. Another advantage of syndromic surveillance is that data transmission is frequent, electronic and automated which greatly reduces the time and cost associated with disease surveillance (Ding et al. 2015; Elliot 2009a; Lateef 2012). May, Chretien and Pavlin (2009) highlighted, the implementation of syndromic surveillance in developing countries that lack laboratory resources for surveillance offers a feasible approach to effective disease surveillance at low cost.

Although syndromic surveillance can be low cost and requires reduced administrative input, there are a variety of technical, financial, political and ethical considerations that need to be addressed before the systems are set up. Importantly, there needs to be a technical collaboration between data collection services and public health agencies which is long lasting and aims to improve system development and operation when needed. These collaborations can often be time consuming and complex to develop, but once in place are essential for the collection of quality data (Chretien et al. 2008). When new data transfer systems are developed, privacy safeguards need to be developed and risk assessed to eliminate the risk of data breaches and elevate any concerns patients might have about electronic data capture and transmission (Chretien et al. 2008; Medina et al. 2014).

Specific data requirements have been identified for a successful syndromic surveillance system by a Dutch review of syndromic surveillance (van den Wijngaard et al. 2011). One of the most important requirements is data quality. Data artefacts; such as duplicate details and reporting delays can result in false alarms (van den Wijngaard et al. 2011). This can be a major limitation of syndromic surveillance, because, in many cases, it is an automated system and therefore data cannot be amended. Although, if these artefacts are known, false alarms can be identified and appropriate action taken. In systems with a high population coverage sensitivity for outbreak detection in a region is increased. Additional patient characteristics and laboratory trends are also required to identify usual patterns of disease (van den Wijngaard et al. 2011).

1.2.2 Data Sources

There are two types of data that can be collected for syndromic surveillance; pre-diagnostic clinical data and ill-health proxies. Pre-diagnostic clinical data composes of a suspected clinical diagnosis of an illness or symptom made by a health professional but is not confirmatory. These diagnoses are typically given when a case uses telehealth services, attends their GP or attends out-of-hour services. Ill health proxies are sources, such as internet searches, absenteeism, and drug sales, that are associated with diseases or symptoms but are not confirmatory, disease specific or diagnosed by a health professional. The main advantage of ill health proxies over pre-diagnostic clinical data is that these data sources may be utilised before a clinical diagnosis is given and therefore outbreak may be detected earlier. Conversely, pre-diagnostic clinical data is more disease specific and less likely to be affected by external factors such as drug sales or media interest. Although pre-diagnostic data from health services can provide more specific data, the use of syndromic surveillances systems focused on health proxies or health seeking behaviours before healthcare is sought are becoming more common (Cheng, Channarith and Cowling 2013). These systems can improve the timeliness of outbreak detection and therefore situational awareness and response (Cheng, Channarith and Cowling 2013). Common data sources for syndromic surveillance include; telehealth, hospital and primary care health data, absenteeism, over the counter drug sales and internet searches.

Pre-diagnostic Clinical Data

Telehealth services provide a non-emergency medical helpline for urgent but not life-threatening issues. Typically, they operate 24/7 which means data can be collected and analysed all year providing a valuable source of continuous health data regardless of holidays or weekends (Abat et al. 2016). The health data produced by these hotlines is inherently not as specific as diagnostic data, but additional data such as time of call, demographics and residence of the caller can be obtained (Lombardo and Ross 2006). This additional data is essential for epidemiological investigation and characterisation of a health event. Telehealth services can be the first contact in the healthcare system but unlike ill health proxies, which may be used or displayed earlier than telehealth, they provide more specific and accurate symptomatic data and are therefore more disease dependent (Moore 2004). Data is also electronically stored, and systems from the same country may uniform triage data, therefore national data can easily be analysed and compared on one computerised system (Moore 2004). In the United Kingdom, two telehealth services are used; NHS111 in England, Wales and Northern Ireland (known as NHS Direct between 1998 and 2014), and NHS24 in Scotland. Both run as a non-emergency telephone advice and triage service. The main difference between the two services is that NHS24 is nurse led, but only operates out-of-hours (with online help available 24 hours), and NHS111 operates 24 hours a day and is operated by trained advisors that are supported by healthcare professionals. Morbey et al. (2017a) assessed the use of NHS111 as a potential early warning for respiratory infections. Over 83% of calls for *Cold/Flu*, *Cough* and *Difficulty Breathing* were associated with respiratory pathogens, with the greatest burden of calls associated with RSV and influenza. The best fitting model indicated that calls related to respiratory infections increased a week before respiratory specimen dates, highlighting its ability to be used in a timely manner to enhance traditional surveillance. Among other places, telehealth syndromic surveillance systems have also been used in Ireland, Canada and Sweden (Andersson et al. 2014; Rolland et al. 2006; Ziemann et al. 2015).

Media reporting can affect telehealth services. Elliot et al. (2016b) described how media interest following the detection of Cryptosporidium oocysts in the public water supply

led to a statistically significant increase in gastroenteritis and diarrhoea related enquires to telehealth, which were unrelated to disease burden and primary care services. After the contamination was announced in the media there was a surge in signals related to gastroenteritis and diarrhoea, but investigation revealed there was no laboratory confirmed cases associated with the outbreak suggesting the surge in calls to NHS111 and contact with primary care was driven by the local and national media coverage.

ED and primary care presentation data includes chief complaints (codes which summarise the reason for an accident and emergency admission), and medical record data (health data, often in the form of codes that have been assigned to a patient by a physician to describe their differential diagnosis).

Primary care practices can provide a vital role in syndromic surveillance. Patients with developing illnesses often visit primary care services first, adding to the timeliness of outbreak detection. Coding of the consultation is often done on the same day, and is conducted by a physician allowing for more disease specific codes being given (Sloane et al. 2006). In England, a GP out-of-hours (GPOOH) service covers come primary care services on evenings, weekends and holidays. These services can be utilised for syndromic surveillance, allowing for near continuous surveillance of primary care services, and reducing the impact of closures on surveillance.

ED records are a frequent source for syndromic data (Travers et al. 2006), and has been shown to detect outbreaks one to two weeks earlier than traditional methods. A US study which assessed complaints given at ED for the early detection of outbreaks found chief complaints of pneumonia and influenza detected epidemics one week before deaths, which were used as the gold standard (Tsui et al. 2001). Josseran et al. (2006) found a significant correlation between influenza outbreaks and influenza related syndromic data from ED and, with a one-two week lag, mortality increased significantly. One of the first examples of the use of ED syndromic surveillance was in 2001 in the US where, Heffernan et al. (2004) used chief complaint information to detect increases in respiratory and fever syndromes and signals for diarrhoea. These signals were consistent with seasonal influenza, norovirus and rotavirus outbreaks at

the time, the information was communicated to the medical community so they could prepare for the arrival of these infections.

For syndromic surveillance of chief complaints, appropriate data coding is crucial for the detection of increases in syndromic indicators. Typically, when patients attend these services, their chief complaint is recorded and usually re-coded to fit specific terms used for syndromic surveillance (Travers et al. 2006). This leaves the system open to errors due to miscoding or delays in detection due to lack of timely recording of health information (Travers et al. 2006). These inconsistencies can result in the creation of false signals or true signals being obscured (Yih et al. 2010). It is difficult to determine the extent which miscoding results in hidden or false signals because once these, mostly automatic, systems are set up there is little opportunity to assess the data (Yih et al. 2010).

In many of the coding systems used for these systems, a variety of codes can be applied to the same syndrome or illness, some of which are more specific than others. Although broad syndrome definitions, such as fever, can maximize sensitivity, it may reduce specificity resulting in false signals as well a true outbreak being hidden (Pendarvis et al. 2007; Yih et al. 2010). Having a variety of codes for the same illness creates the opportunity for discrepancies in the way individual clinicians assign a code for what could be the same illness. Again, although the use of multiple and broad syndromic codes allows the system to be more specific, it introduces the chance for more errors or miscoding and therefore reducing specificity (Yih et al. 2010).

Use of chief complaints in ED and primary care facilities data, although more accurate than telehealth services or ill health proxies, could result in a delay in outbreak detection because attendance to ED or primary care often only occurs after someone has been ill for several days, depending on the illness in question.

Ill Health Proxies

Absenteeism can be used as a proxy for ill health. It follows the principle that when someone is ill, they might take time off work or school, therefore higher than expected absenteeism could indicate an outbreak of disease. This data can be used for the early

identification of an outbreak because absenteeism often occurs before seeking medical help (Cheng et al. 2012). Schmidt, Pebody and Mangtani (2010) used school absence records as a tool for influenza surveillance in England, and although they found some peaks of absenteeism which coincided with peaks in laboratory confirmed influenza activity, others did not. Influenza activity was also found to have a much stronger correlation with absenteeism prevalence than incidence (duration of absence vs number absent). The study concluded that absenteeism, as an additional surveillance tool could be useful in the detection of localised outbreaks. The school absenteeism predictor for influenza activity has also been demonstrated, to varying extents in Hong Kong (Cheng et al. 2012), the USA (Egger et al. 2012), Japan (Suzue et al. 2012) and Canada (Kom Mogto et al. 2012).

The advantages of this data source are that the data is pre-existing, easily accessible and requires limited time and resources once the system is implemented (Cheng, Channarith and Cowling 2013). However, there are some major disadvantages to the data source, which greatly affect data completeness and quality. During periods where schools are not open, such as weekends and holidays, there are gaps in the data which can make data analysis and interpretation difficult (Cheng, Channarith and Cowling 2013; Schmidt, Pebody and Mangtani 2010). Absenteeism can be higher before and after holidays or weekends due to "holiday effects", which can further exacerbate the issues with data gaps (Besculides et al. 2005; Cheng, Channarith and Cowling 2013). Finally, absenteeism can be due to a number of reasons, illness related or not which can mask a true signal of an outbreak (Cheng, Channarith and Cowling 2013).

Drug purchase is a form of early healthcare-seeking behaviour that can be targeted for the early detection of increases in disease activity. Much like absenteeism, obtaining medication to treat symptoms is a behaviour people undertake when they first become ill or have mild symptoms (Magruder 2003). Therefore, certain drugs, which are used to treat certain syndromes or illnesses (such as cold or flu medicine) can be monitored and used for the early detection of a disease outbreak. A study by Magruder (2003) found a 90% correlation between flu remedy sales and physician diagnosis of acute respiratory conditions. Pivette et al. (2014a) also found a high correlation between over-

the-counter drug sales for treatment of gastroenteritis and GP consultations for the condition and an epidemic detection, on average, 2.25 weeks earlier than the sentinel network of GPs. A systematic review by Pivette et al. (2014b) found a high correlation between drug sales and the disease surveillance data in 89% (17/19) of the studies reviewed.

Drug sales are particularly useful in the surveillance of diseases that tend to be under-reported or where clinical surveillance is costly (Pivette et al. 2014b). Unlike other forms of syndromic surveillance data sources that are used as proxies for ill-health, such as absenteeism and Google searches, drug sales can be directly linked to a health condition, may be more specific that other disease proxies, and are more likely to reflect actual disease activity (Pivette et al. 2014b). The drugs that are monitored should be disease specific and widely used to maximise signal sensitivity and signal detection (Pivette et al. 2014b). For example, drugs such as paracetamol and ibuprofen can be used to treat a wide range of symptoms, such as fever and muscle ache, with varying severity therefore would not be suitable to target specific diseases or symptoms for surveillance. Infections such as measles or tuberculous, which are not typically treated with a specific drug or treatment is administered in a hospital setting, would be inappropriate targets (Pivette et al. 2014b). This limits the diseases that can be detected through syndromic surveillance of drug sales. External influences, such as cultural, demographic, economic and social factors can confound the association between drug sales and illness (Pivette et al. 2014b; Frosst, Majowicz and Edge 2006). The sale of drugs could be influenced by promotions, the media, or holidays with people buying medication to treat the disease they may get, rather than have currently (CDC 2006a; Frosst, Majowicz and Edge 2006; Pivette et al. 2014b).

Internet based surveillance is novel approach and allows for the near real-time monitoring of diseases and disease indicators over large geographical areas through the aggregation and analysis of internet search terms and the use of social media such as Twitter and Facebook. One of the major benefits of syndromic surveillance systems is its flexibility of being able to utilise different approaches and data sources, giving it the ability to adapt to changing public health needs depending on the situation. This

flexibility also allows it to detect new syndromic indicators when required, take advantage of shifting patterns in human behaviour, such as changes in the way we seek health information, and the development and advancement of new tools which can be used as data sources (Paterson and Durrheim 2013). Over the past 10 years, and as syndromic surveillance has been evolving, internet availability and use has increased significantly (Bahk, Kim and Park 2015). This has changed the way people seek information about their health and has led to the emergence of internet based public health surveillance (Rice 2006). Although this emerging form of syndromic surveillance lacks the capacity to replace traditional surveillance methods, internet-based surveillance can use internet searches and social media posts as proxies for ill health, to detect and monitor disease as an extension to traditional surveillance methods (Milinovich et al. 2014). Internet based surveillance is both economically and logistically beneficial, with large amounts of electronic data available over large geographical areas (Milinovich et al. 2014).

Although there are few examples where internet-based surveillance has been used in health protection practice, its possible utility has been assessed in the literature. A study in South Korea found a significant correlation between internet search terms for foodborne illness and the total number of in-patient stays related to foodborne illness, with searches for food poisoning correlating strongly with chief compliant codes for infectious enteritis and bacterial foodborne illness. The five search terms studied; 'food-poisoning', 'diarrhoea', 'vomiting', 'abdominal pain' and gastroenteritis', correlated strongly with the total number of hospital stays for all surveyed foodborne related illnesses the in the following month of when the internet searches were conducted (Bahk, Kim and Park 2015). Hulth, Rydevik and Linde (2009) compared trends in queries for influenza related terms to sentinel and laboratory surveillance data for influenza, and found that web queries follow the same trend as the other two forms of surveillance and all three have an equal power for the true estimate of influenza burden. Social media has also been proposed as a way to detect seasonal outbreaks, Corley et al. (2010) found a correlation between terms used to describe influenza in blog posts and Centres for Disease Control (CDC) ILINet data. Chew and Eysenbach (2010)

observed an association between Twitter posts containing terms related to the influenza H1N1 outbreak in 2009 and incidence rates in the US.

One of the earliest and most well-known forms of internet-based disease surveillance was Google Flu Trends (Ginsberg et al. 2009). This service, operated by Google, estimated influenza activity by aggregating Google search queries that were related to influenza. It was launched in 2008, and by 2013 it was being used to predict influenza activity in 29 countries and monitor dengue trends in 10 countries. The model, created using Google search queries, was found to be highly correlated to physician visits for influenza like symptoms and predicted regional outbreaks of influenza seven to ten days before conventional CDC surveillance systems which used laboratory and clinical data (Carneiro and Mylonakis 2009; Ginsberg et al. 2009). But between 2011-2013 the predictions were inaccurate and often overestimated, at some points more than double the true prevalence, and in some years, they underestimated prevalence, failing to detected the unseasonal influenza H1N1 pandemic in 2009, highlighting the shortcomings of internet-based surveillance (Lazer et al. 2014). The translation of big, raw data into meaningful and accurate information is challenging, bias can be introduced to the data through unknown search motives (Milinovich et al. 2014). Therefore, peaks in activity could be attributed to actual illness or media driven interest in the disease (Desai et al. 2012). A study by Wilson and Brownstein (2009) highlighted that media can drive online disease interest after a disease outbreak. They found that an increase in searches of the term "Listeria" coincided with media coverage and not the outbreak itself. Between 2005 and 2006 there was in increase in searches for “bird flu” in the USA, despite no avian influenza outbreak there at the time. This increase was attributed media interest about the outbreak in Asia (Carneiro and Mylonakis 2009). Changes in search behaviours can also affect the ability for these systems to detect real outbreaks, for example searches for possible disease related terms that are not in fact related to disease need to be accounted for (Milinovich et al. 2014).

Data is typically aggregated to a large geographical area or at national level and contains no demographic data (Hulth, Rydevik and Linde 2009), therefore these methods lack spatial and demographic resolution to detect local outbreaks or characterise outbreaks

epidemiologically. Theoretically IP (internet protocol) addresses could be used to identify where the user is, but not all search engines log this information (Hulth, Rydevik and Linde 2009). A study conducted in America by Polgreen et al. (2008) used search queries on Yahoo! to investigate the relationship between internet searches and influenza activity. IP addresses associated with the search were used to identify geographic location. Although the models generated from this study successfully predicted influenza one to three weeks before laboratory surveillance, and mortality from pneumonia and influenza up to five weeks ahead, geographical identification proved to be a limitation of the study. Not all searches were associated with geographical data and those that did were not always accurate. Importantly it was also discussed that in some cases very specific geographic information could be obtained and linked to individuals which, for both individuals and health investigators, would be a privacy concern.

1.2.3 Data Analysis

Statistical analysis of syndromic surveillance can be challenging due to many reasons such as; multiple data sources, lack of follow-up, unable to control data duplication and lack of data evaluation (Unkel et al. 2012). Statistical algorithms, based on historic data, are used to detect aberrations that signal a public health threat. The availability of historical baseline data is crucial for the development of the statistical algorithms that are used to detect abnormalities in the data and alert public health officials to a possible threat. These algorithms draw on historical data to account for the effects of natural changes in data, such as seasonal patterns and data artefacts, and estimate a typical level of reporting to detect unusual events (Chretien et al. 2008). Although alert systems may vary across different systems, many rely on the same principles. For example, the real-time syndromic surveillance team (ReSST), which are part of PHE, detect unusual activity when there is a statistically significant increase in a signal, which are syndromes monitored by geography and system. When an alarm is produced public health, officials investigate to determine if it is a true public health threat. If it is deemed a threat or requires communication, an alert is created to inform others of the unusual activity (Morbey et al. 2015b; G. E. Smith et al. 2016).

For the 2012 Olympic games in London, enhanced syndromic surveillance was introduced leading to the development and refinement of new statistical methods by the ReSST (Morbey et al. 2015a). Historical data was available for all four systems implemented, NHS Direct (telehealth), GP out-of-hours, GP in hours and ED, but only the GP in hours systems had data that was comparable. For this system, this data was used to create a baseline by comparing activity with a five-week rolling average from the previous three years. An upper threshold was set for unusual activity using a 99% prediction level. For the three systems and where appropriate historical data was available, a baseline was created using data from the previous two to three weeks. In all four systems, day of the week effects were accounted for. Over the period of the games (73 days) these systems produced 347,754 signals from which 3,946 alarms were generated and investigated, 202 of these alarms were of sufficient interest and after a stringent risk assessment were mentioned in daily reports (Morbey et al. 2015a).

During the 2012 Olympic Games in London, the ReSST developed a public health risk-assessment tool to prioritise statistical alarms and ensure health officials were only alerted to the statistical alarms that were of importance (G. E. Smith et al. 2016). The assessment was divided into two stages. The first stage was carried out by a syndromic surveillance scientist where an alarm was assessed on the following criteria; size of excess observed, if it was unusual for the syndrome, whether it was a repeat alarm, was the alarm different to the national trend, and if a similar alarm has occurred independently on another system. If this assessment met the scoring criteria, using a Likert-type scale, then a consultant epidemiologist would conduct a second final risk assessment using the following criteria; if the syndrome was unusual for the time of year, if there was geographical clustering to suggest an incident, if the increase was observed in a certain age group and if there was evidence of increases severity of a syndrome. If an alarm scored highly on this second assessment, the threat was communicated to other relevant colleagues. This risk assessment allowed for the systems to be implemented into the enhanced surveillance systems used by PHE successfully by streamlining workload and providing consistency in epidemiological interpretation (G. E. Smith et al. 2016).

One of syndromic surveillances greatest advantages is also one of its greatest weaknesses. Although the use of pre-diagnostic health data allows for the early identification of events, the data is not as specific and accurate as confirmatory data (Elliot 2009a). Syndromic surveillance, at the expense of specificity, is sensitive to changes in indicator trends (Elliot 2009a). The lack of specificity increases the chance of a false alarm. Analysing data for longer could improve accuracy and reduce false alarms but this would compromise the timeliness of the system (Stoto, Schonlau and Mariano 2004). Non-specific indicators, such as fever, can result in frequent false alarms and a large number of resources investigating them but inadequate sensitivity could result in the failure to detect health events (Gault et al. 2009). Therefore, there needs to be a trade-off between accurate data and early response (Stoto, Schonlau and Mariano 2004). The accuracy and sensitivity of syndromic data to detect aberrations in data relies on the availability of baseline data (Berger, Shiau and Weintraub 2006). Historic data can be used to create accurate, comparable baseline data which can be used to reduce false positive rates (Lateef 2012). The non-specific "broad bush" approach of using syndromic indicators may capture people who do not fit the case definition or those with mild illness (Elliot 2009a; Lateef 2012).

Although it is highly automated, the signals that indicate an event still need to be assessed by a trained public health practitioner due to the possibility of false positives. Wang et al. (2005) developed a fully automated syndromic surveillance system for the detection of outbreaks and although it only had an overall 84.8% true detection accuracy, it was stated that knowledge of exogenous factors improves detection accuracy.

1.3 Integrated Syndromic Surveillance Systems

For syndromic surveillance to fulfil its potential of early outbreak detection and situational awareness it needs to be integrally linked to public health investigation and response. As previously highlighted, data sources for syndromic surveillance have their limitations, therefore, when an outbreak signal is detected, investigations need to be

undertaken to differentiate between natural variability, pseudo-outbreak (data errors) and a true increase in incidence (Mandl et al. 2004). Validation from other syndromic surveillance data sources can be used to corroborate signals indicating a true increase in incidence. When a true signal has been detected, the signal should be investigated further to identify the cohort affected and determine the public health significance of the signal. If appropriate, the observed information should be communicated to front line medical staff to heighten awareness and the appropriate response should be undertaken (Mandl et al. 2004).

Syndromic surveillance has been integrated into several national surveillance systems to augment traditional surveillance systems. As of 2013, there were 124 syndromic surveillance systems worldwide, 60 of which were in Europe at both local and national level (this includes both veterinary and human health systems) (Ziemann et al. 2015). They utilise a variety of data sources, the majority involve automated electronic reporting of a variety of diseases or disease indicators. Although primarily used for infectious disease outbreak detection, more systems are now being used to monitor non-infectious diseases and major events for situational awareness.

In England, ReSST coordinates 5 national syndromic surveillance systems as part of PHE. The five systems are: NHS111, GP in hours, GP out-of-hours, ED, and ambulance services. These healthcare services are operated by the National Health Service (NHS), which provides consistent and comprehensive healthcare services across England. Most importantly it is free at the point of delivery, and is accessible but anyone in England. This integrated national healthcare system has allowed for the development of a truly integrated and comprehensive syndromic surveillance service in England. Population coverage varies across the systems, with the remote health advice system covering, around, 100%, GP out-of-hours covering 80% and GP in hours covering 55% of the population, the ED system covers 35 departments (PHE 2015). Syndromes associated with specific diseases of interest were identified and disease indicators that capture these syndromes are monitored by the surveillance systems. Examples of syndromic indicators monitored by ReSST are provided in Table 1.1. These systems aggregate data of indicators that are indicative of disease. Weekly, a syndromic surveillance summary

of changes of disease indicator trends observed from the systems is disseminated to appropriate people for their awareness. In England, these systems have been used for conventional surveillance system purposes; to evaluate health interventions (Bawa et al. 2015; Pebody et al. 2015), monitoring disease trends and identify seasonal disease outbreaks (Hughes et al. 2014). However, increasingly these systems are being used for situational awareness of severe weather/natural events and mass gatherings. During both the London Olympic and Paralympic games in 2012, syndromic surveillance was used to assess the impact, or lack of, mass gathering events on health services (Morbey et al. 2014; Morbey et al. 2015a; G. E. Smith et al. 2016; Todkill et al. 2016). It has also been used to assess the health impact following cold weather (Hughes et al. 2014), heatwaves (Elliot et al. 2014; S. Smith et al. 2016a; 2016b) and air pollution events (Elliot et al. 2016a; G. E. Smith et al. 2015).

NHS111	**Emergency Departments**	**GP in Hours**	**GP Out of Hours**
- Cold/flu - Fever - Cough - Difficulty breathing - Diarrhoea - Vomiting - Eye problems - Heat/sunstroke	- All respiratory disease - Acute respiratory infections - Acute bronchitis/ bronchiolitis - Influenza-like illness - Pneumonia - Gastroenteritis	- Influenza-like illness - Upper respiratory tract infections - Lower respiratory tract infections - Pneumonia - ILI with antivirals prescribed - Pneumonia with antibiotics prescribed - Gastroenteritis - Diarrhoea	- Gastroenteritis - Diarrhoea - Vomiting

Table 1.1: Examples of syndromic indicators used in PHE syndromic surveillance systems in England (Bawa et al. 2015; Harcourt et al. 2012b; 2016; Hughes et al. 2016).

National and local, integrated syndromic surveillance systems have also been used to detect possible health emergencies among migrants during the migrant crisis in Italy between 2011-2013 (Napoli et al. 2014; Riccardo et al. 2011), estimate the impact of the 2015 Paris terrorist attack on ED (Vandentorren et al. 2016) and investigate the impact of a major power outage on diarrheal illness in New York (Marx et al. 2006). In response to Hurricanes Katrina and Rita in 2005, the CDC and Louisiana Department of Health and Hospitals implemented syndromic surveillance in hospitals and acute-care facilities to monitor injury and illness (CDC 2005; 2006). These examples highlight how integrated syndromic surveillance systems can be used to determine, evaluate and provide valuable information for an effective public health response for a variety of public health threats and needs to enhance existing surveillance systems.

1.4 Conclusion

Since syndromic surveillance first became widely used in the beginning of the 21st century, it has demonstrated remarkable adaptability to the changing needs of public health surveillance. Its flexibility of utilising a wide range of data sources has allowed it to take advantage of changing healthcare-seeking behaviours. Although syndromic surveillance is intrinsically non-specific, the utilisation of syndromic surveillance data sources allows for the early detection of outbreaks, and broader estimation of disease in the community. When integrally linked to public health investigation and response, syndromic surveillance can be used to augment traditional surveillance methods, and provide vital information to allow informed decision on public health threats the and appropriate response.

1.5 What has Been Described and What is Missing in the Literature?

When syndromic surveillance was first developed its primary aims was to detect bio-terrorism events, since then, its purpose has moved evolved. Currently the primary purpose of the syndromic surveillance service operated by PHE is the early detection of outbreaks, situational awareness and for reassurance to the lack of threat from large events (G. E. Smith et al. 2019), including monitoring seasonal respiratory virus activity, extreme weather events, mass gatherings, air pollution, and norovirus activity. Data from the syndromic surveillance service at PHE is available from multiple healthcare services at daily time points, covers a large proportion of the population, with data available at small geographical units. This rich data source has enormous potential to detect and monitor public health threats outside the routine purpose of ReSST. Two previous studies have successfully utilised the large population and high spatial resolution of this data to estimate the relationship between socioeconomic status and presentations to services monitored by syndromic surveillance. Todkill et al. (2017a) estimated the relationship between socioeconomic status, urbanicity and allergic rhinitis presentations to GP services, it was observed that consultations rates were higher in more urban areas and areas that were more socioeconomically deprived. Adams et al. (2018) estimated the relationship between call related to gastrointestinal illness to NHS111 and socioeconomic status, with those from more deprived backgrounds at higher risk of calling NHS111. The data from ReSST provided access to novel datasets, allowing a better picture of what is going on in the community compared to traditional data sources.

Although previous studies have utilised the geographical resolution of syndromic surveillance, there has been few previous studies exploring both its spatial and temporal utilities. The frequency at which data is available and the granularity of the local geography may allow us to explore healthcare usage in ways not previously explored. In this thesis we will use respiratory presentations to three healthcare services in the community (NHS111, GPIH and GPOOH) to explore how syndromic surveillance can be used in spatial and temporal epidemiological studies to identify and quantify public

health issues. Respiratory issues were chosen as their high burden of illness will result in more presentations compared to other types of disease, giving us more opportunity to explore spatial and temporal relationships.

Although syndromic surveillance data is available from five healthcare services; NHS111, GPIH, GPOOH, ambulance services and ED, for this thesis, only data from NHS111, GPIH, GPOOH will be used. In the literature when observational epidemiological studies are conducted, they primarily use data use data from hospitalisations or laboratory data, leaving information in the community poorly defined. By focusing on NHS111, GPIH and GPOOH we can explore respiratory presentations in the community.

For part of this thesis, we will focus on respiratory presentations in children under-5 years, particularly in relations to RSV. This is because of the high burden of respiratory issues and RSV in children under-5 years. RSV in particular has a high burden on children under-5 years, with previous studies on burden estimates poorly defining cases in the community (Ajayi-Obe et al. 2008; Cromer et al. 2017; Hardelid, Pebody and Andrews 2013; Müller-Pebody et al. 2002; J. Murray 2013; Reeves et al. 2017; Taylor et al. 2016). RSV in children under-5 years also has distinct symptoms with lower respiratory tract involvement (Ogra 2004), which may allow us to detect RSV cases from the respiratory symptoms used syndromic surveillance. The primary aim of this thesis is to explore spatial and temporal associations of respiratory presentations in the community using syndromic surveillance. This will allow us to observe the utility of syndromic surveillance data in identifying and qualifying public health threats in spatial and temporal observational epidemiological studies.

The aims of this thesis are:

1. To identify and quantify the demographics who use these services monitored by syndromic surveillance.
2. To estimate the community burden of RSV in children under-5 years using data from syndromic surveillance use these systems to estimate the burden of RSV.

3. Use RSV burden estimated to explore the spatial and temporal distribution of RSV presentations in young children.
4. Explore the relationship between weather and respiratory presentations in children under-5 years.

1.6 Contribution to the Literature

In this thesis we have successfully demonstrated how data from syndromic surveillance can be used for spatial and temporal observational epidemiological studies, despite its limitations. Deprivation, age and gender have been shown to play an important role in who seeks healthcare in the community. Although these factors and how they contribute to ill health have been widely described in the literature, we demonstrated that, for these healthcare services, there are little differences in who seeks healthcare for only respiratory illness and for all illnesses. This highlights there are similar drivers of all diseases, whether they are aetiological or due to behavioural factors, or both.

Using syndromic surveillance data, we estimated the burden of RSV in children under-5 years in the community using temporal analysis, providing further insight into this illness. Previous estimates of RSV have not managed to fully capture community burden, here we have highlighted the high burden of RSV in this age group, especially in comparison to influenza. By using data from syndromic surveillance, we managed to detect many more presentations attributed to RSV, that would have been detected through laboratory surveillance. Estimating disease burden at the community as well as in acute setting is an important step in further understanding disease transmission dynamics.

In order to explore possible regional differences in RSV activity we estimated the burden of RSV, as well as other metrics for activity. Although we did not observe meaningful differences in regional RSV activity, we highlighted the limitations of syndromic surveillance data. It is empirical to fully understand the strengths and weaknesses of data when undertaking epidemiological research. In this case, although

there is the potential to cover the full geography of England, the passive nature of syndromic surveillance meant that there were long periods where large geographical areas have no surveillance coverage for the GPOOH surveillance system. This limited the ability to conduct analysis that compared geographical regions. We have also highlighted the role of meteorological conditions in presentations for respiratory infections in children under-5 years. We have demonstrated that these factors play a role geographically and all year round. With growing interest in forecasting disease or healthcare presentations, these observations can be used in forecasting models as we have demonstrated that there is a relationship between meteorological conditions and respiratory presentations to healthcare services.

1.7 Thesis Structure

This thesis is divided into five chapters.

Chapter 2 describes the syndromic data used in this thesis. Data was obtained from ReSST, from three healthcare services that are operate in the community; NHS111, general practice in-hours and general practice out-of-hours. In this chapter the healthcare services that provide the data are described as well as the remit for ReSST.

Chapter 3 estimates the demographic and socioeconomic patterns in healthcare-seeking behaviour for respiratory presentations to three healthcare services in England (NHS111, GPIH and GPOOH). Data was obtained annually for 2015 and 2016 and analysis was conducted using generalised linear models (GLM) and generalised linear mixed models (GLMM). Comparisons were made with non-respiratory presentations to identify whether associations are specific to respiratory disease. Young males (<5 years) were observed to have the highest rate respiratory presentations to all three healthcare services. Those from more deprived areas were more likely to seek healthcare due to respiratory symptoms compared to those from the least deprived areas. Comparable results were observed between respiratory and non-respiratory presentations suggesting that demographic and socioeconomic factors may be the strongest influencers of healthcare-seeking behaviours.

A similar version of this chapter has been published as: Morrison, K. E.; Colón-González, F. J.; Morbey, R. A.; Hunter, P. R.; Rutter, J.; Stuttard, G.; de Lusignan, S.; Yeates, A.; Pebody, R.; Smith, G.; Elliot, A. J.; Lake, I. R. (2020). Demographic and socioeconomic patterns in healthcare-seeking behaviour for respiratory symptoms in England: a comparison with non-respiratory symptoms and between three healthcare services. BMJ Open 2020;10:e038356. doi: 10.1136/bmjopen-2020-038356

Author contributions: conception and design: KEM, analysis and interpretation of data: KEM, drafting the manuscript: KEM, significant input in the conception and design: IRL, FJC-G, RAM, GS and AJE, significant input in analysis and interpretation: IRL, FJC-G, RAM, GS and AJE, drafting the manuscript and critically reviewed the final version of the work: IRL, FJC-G, RAM, GS and AJE, critically reviewed the final version of the work: PRH, JR, GS, SdeL, AY, RP.

Chapter 4 estimates the weekly community burden of RSV in children under-five years in England, using respiratory presentations to three healthcare services (NHS111, GPIH and GPOOH), from 11th of November 2013 to 18th of June 2018. Time-series Poisson and negative binomial GLM was used to model each syndromic indicator included in the analysis. Estimates were made while controlling for; other respiratory pathogens, public holidays, seasonality, and temporal autocorrelation. Comparisons were made between age groups (< 1 year and 1-4 years), with influenza (A+B) and with the number of cases detected by laboratory testing. RSV was estimated to account for almost 200,000 presentations to NHS111, GPIH and GPOOH per year. The ratio of influenza presentations to RSV presentations was 1:3.1, the ratio difference was highest in children under-1 year. For every case of RSV reported by laboratory surveillance we estimate there were 69 healthcare presentations.

Chapter 5 estimates differences in RSV activity (burden, start, peak and length) across nine PHE centres in England, in children under-1 year. Activity was measured using weekly respiratory presentations to two healthcare services: NHS111 and GPIH from 30th June 2014 to 25th June 2018. Poisson GLM, using a Bayesian approach, were used to estimate the number of RSV attributable presentations from respiratory syndromic indicators while controlling for; other respiratory pathogens, public holidays,

seasonality, and temporal autocorrelation. The burden of RSV appeared to vary spatially, with burden highest in the southern regions of England. There were no meaningful spatial differences in start, peak and length of RSV activity.

Chapter 6 explores the association between meteorological conditions (temperature, humidity, rainfall and wind speed) on weekly NHS111, GPIH and GPOOH presentations due to acute repository illness in children under-5 years in England from 11th November 2013 to 18th June 2018. Analysis was conducted using Bayesian spatial-temporal Poisson GLM while accounting for gender, public holidays, seasonality, and spatially and temporally structured random effects. We explored both linear and non-linear relationships, as well as the spatial and monthly associations. Most of the respiratory indicators studied had a negative important relationship with temperature and rainfall, and positive important relationship with humidity and wind speed. No meaningful spatial patterns could be observed and when monthly relationships were explored, we observed that meteorological conditions have an influence all year round.

Chapter 2

Public Health England's Real Time Syndromic Surveillance Service

Data for this thesis was obtained from Public Health England's Real Time Syndromic Surveillance Service between 2015 and 2019. Descriptions of the healthcare services, data and surveillance systems are correct as of the time data was received.

2.1 Background

Public Health England (PHE) operates a national near real-time syndromic surveillance service (ReSST) monitoring five healthcare services: telehealth, general practice (in and out of hours), ambulance services, and emergency departments. The healthcare services monitored are operated by the national health services (NHS) and are free at the point of delivery. The presence of the NHS has allowed for the establishment of ReSST as it provides consistent and comprehensive healthcare across England. These systems routinely collect health-related data comprising of symptoms or clinical signs of disease reported by patients or clinicians before a laboratory or confirmatory clinical diagnosis of illnesses is made. The data collected by these systems are analysed through descriptive and statistical modelling methods to generate alerts for public health action (Morbey et al. 2015a; 2015b). Alerts are, typically, generated when activity is high in comparison to historic trends, but alerts can also be generated when there is a recent increase in activity even if low by historical level (Morbey et al. 2015b). When syndromic surveillance was first established in the US in the 2000's its focus was on the detection of bioterrorism events and influenza activity. Now, this remit has evolved to encompass more uses. Currently, the remit of ReSST is for the early detection of outbreaks, both seasonal (such as influenza) and sporadic (such as cryptosporidium),

situational awareness (e.g., describing the impact of weather events), and for reassurance to the lack of threat to public health (such as from mass gatherings) (G. E. Smith et al. 2019). The systems monitored by ReSST have been developed to meet the three main definitions of syndromic surveillance as described by the Triple S Project (Triple S Project 2011):

1. the near-real time collection, analysis, interpretation and dissemination of health-related data, for the identification of impact of public health threats;
2. collection of non-specific health indicators that are based on clinical signs and symptoms of disease, and not laboratory-confirmed diagnosis;
3. data is already collected for purposes other than surveillance and is generated as to not create an additional burden to healthcare providers.

In England, a national unit of syndromic surveillance has been in development since 1998 when it was part of the Health Protection Agency (HPA), which was integrated into PHE in 2011. In preparation for the London 2012 Olympic games, the existing systems (general practitioner in-hours, NHS Direct and emergency departments) were strengthened and new systems (general practitioner out-of-hours) were developed in order to deliver the comprehensive service, which is in use now (Elliot et al. 2013). Since the London Olympic games ReSST has developed two new surveillance systems monitoring the telehealth system NHS111 (Harcourt et al. 2016), an evolution of the NHS Direct service, and the ambulance service (Todkill et al. 2017b). Information on presenting symptoms as well as demographic information is collected daily, from activity the previous day for NHS111, general practitioner in-hours (GPIH), general practitioner out-of-hours (GPOOH) and ambulance surveillance systems. Emergency departments (ED) daily reports are from two days prior, because not all information is available within 24 hours. The data collected from these services is provided in aggregated form and is anonymous with no personal/practice identifiers recorded.

In this chapter, we describe the main sources of data used in this thesis which were obtained from ReSST. As this thesis focuses on data from the community, and not acute healthcare settings, and utilises spatial analysis, we only included data from the NHS111, GPOOH and GPIH systems. At the time of receiving the data for this thesis

there was limited coverage in England for the ED and ambulance surveillance systems. By focusing on these three community healthcare services, we are able to explore healthcare-seeking behaviours at the community level as opposed to the more acute healthcare level.

2.2 NHS111

NHS111 is a medical helpline, run by trained call handlers, for advice on urgent, non-life-threatening health problems, it also acts as a triage system and has the capability to provide access to general practitioners (GP) (both in and out-of-hours) or can trigger an ambulance response. It is free to access through a three-digit (111) telephone number and is available to anyone in England. This service is available 24 hours a day, 365 days a year (NHS 2017). ReSST has monitored NHS111 as part of its national suite of surveillance systems since 2013 (Harcourt et al. 2016). The use of NHS111 does not require pre-registration and anyone can contact this service, such as foreign visitors.

NHS111 uses a symptom-based Clinical Decision Support Software (CDSS), to assess the need of the caller and provide appropriate advice, such as signposting to a pharmacist, or an action (such as providing an ambulance). NHS111 uses the CDSS NHS Pathways, the same tool is also used by the ambulance service, in urgent care assessment units and with patients presenting at emergency departments (NHS Digital 2020). Non-clinical trained call handlers are presented with a series of questions to ask the patient. Questions are asked in a clinical hierarchy with the most serious questions asked first, progressing to questions about less serious symptoms. Through a series of algorithms that link questions and advice a clinical endpoint is reached whereby the call handler can give advice or trigger an appropriate clinical response (NHS Digital 2020). The NHS Pathways systems is split into three symptom-based modules. Module 0, is the start of the pathway, where questions about the most serious symptoms are asked, such those on blood loss, loss of consciousness, and heart attacks. The NHS111 surveillance system does not receive data from the Module 0 pathway. If the answers provided are assessed as serious the handler will trigger the dispatch of ambulance

services without further questioning. Module 1 uses a body map relevant to the patients age and gender, with questions selected based on the patient's main symptom, again these questions are asked in a hierarchical order with questions regarding the most serious symptoms first. Module 2, is used when calls become too complex for the trained call handlers to safely triage and in these scenarios the call is transferred to trained in-house clinicians for assessment (NHS Digital 2020).

The NHS111 surveillance system obtains syndromic indicators based on the chief complaint at the clinical endpoint of the call after the triage process. A number of syndromic indicators are collected by ReSST to enable the monitoring of key infections, such as influenza, or environmental issues, such as gastrointestinal outbreaks, for health protection purposes (Harcourt et al. 2016). Syndromic indicators monitored through this surveillance system include: *Cold/Flu, Fever, Cough, Difficulty Breathing, Diarrhoea, Vomiting,* and *Heat/Sunstroke* (Harcourt et al. 2016). Information on the patients age, gender, geographic location (postcode district) and call outcome (i.e., ambulance call out or self-care advice given) is collected through the surveillance system. The total number of calls to the NHS111 service is also collected for epidemiological analysis and can be used as a denominator. This is to adjust for fluctuations in the daily call numbers and to compensate for the staggered roll out of the service (Harcourt et al. 2016).

NHS111 data has distinct weekly pattern. An analysis of the service conducted by ReSST between November 2013 and November 2014 found that the total number of calls to this service was highest over the weekends and during public holidays. It was also observed that 20-30% of calls did not have syndromic information and <2% did not include a postcode district (the lowest geography of caller) allocated (Harcourt et al. 2016). There are several reasons as to why a call to NHS111 may not results in a syndromic indicator being recorded. Threats to life, where the Pathway detects the most serious symptoms such as heart attacks or heavy blood loss will results in an immediate dispatch of ambulance services and therefore a symptom would not be recorded. Callers with pre-determined management plans or frequent callers (both identified through the NHS111 service) may also not have a symptom recorded (Harcourt et al. 2016).

Analysis of NHS111 data as part of this thesis between the 1st of January 2015 and the 31st of December 2016 found 2.9% of callers did not have geographical information recorded.

Since its implementation in 2013, syndromic surveillance of NHS111 has routinely been used in epidemiological studies to monitor and evaluate outbreaks and public health incidents. In conjunction with surveillance of GPOOH, surveillance of NHS 111 was used to develop thresholds for influenza and respiratory syncytial virus activity to detect and predict significant episodes of activity in these viruses to aid monitoring (Harcourt et al. 2019). Surveillance of NHS111 was used to contribute to the understanding of an atypical winter respiratory illness season in 2014/15; where a higher level of activity of acute respiratory disease (*influenza-like illness, severe asthma*, and *pneumonia*) was noted in older age groups in comparison to previous years (S. Smith et al. 2017). It has also been used as part of an ecological study investigating the social pattering of telehealth calls regarding diarrhoea and vomiting, it was observed that more deprived areas of England had a higher incidence of calls regarding gastrointestinal illness compared to the least deprived areas (Adams et al. 2018). During a period of widespread poor air quality in England in March and April 2014, syndromic surveillance was used to monitor community health; for which respiratory indicators from the surveillance of NHS111, GPIH and GPOOH were used. During this period of poor air quality, statistically significant rises in presentations for a variety of respiratory conditions, including asthma and wheeze, were detected, these rises were observed to be short-lived (G. E. Smith et al. 2015; Elliot et al. 2016a).

2.3 General Practice In-Hours (GPIH)

GP services, also known as primary care or family doctors/medicine, provide care for routine illnesses, and are often a patient's first point of contact for healthcare. They diagnose illness, provide treatment, and can refer patients to more specialised clinics for treatment if necessary. They have a wider role in public health in the community by promoting healthy lifestyles and the prevention of chronic illness. GP services include;

treatment for minor injuries or illness, health advice (smoking cessation or weight loss), vaccinations, health screening, monitoring of chronic illness and prescriptions. Generally, most GP services are run under a nationally negotiated contract, negotiated by the General Medical Services, whereby the GP is paid a sum per patient registered, and for the provision of certain enhanced services, such as extended opening hours, or reaching targets, such as the number of influenza vaccinations given (The King's Fund 2011). GP in hours services can only be accessed by those who are registered with the service, and, typically, require a pre-arranged appointment to attend.

Syndromic surveillance of GPIH services began in 2004 as part of a collaboration between the University of Nottingham Division of Primary and the HPA. The collaborative project with HPA was establish in response to the UK Influenza Pandemic Contingency Plan, which emphasised the importance of increased, up-to-date surveillance from general practice (NHS 2005). Key areas of information were lacking from previous community surveillance systems such as, sub-regional and prescription data, and there was a large time delay in receiving the information. Through this collaboration, key conditions were identified that may indicate areas of health protection concern and could be used for the timely surveillance of incidents. The conditions were based on the Read coding system used in GP settings (Robinson et al. 1997; NHS Digital 2018). Read codes are clinical terminology, primarily used in GP settings, to provide a standardised vocabulary for clinicians to record patient interactions, and have been in use since 1985 (Robinson et al. 1997; NHS Digital 2018). These key conditions included symptomatic data that were related to respiratory, and gastrointestinal related illnesses and information of vaccine uptake with data extracted weekly with additional information on the patients age, sex and the strategic health authority (SHA) of residence (G. E. Smith et al. 2007).

Since its implementation in 2004, the syndromic surveillance of GP services has evolved considerably with data now submitted daily, more clinical codes are obtained and information from another GP IT provider, TPP (The Phoenix Partnership), is received. The surveillance system is now estimated to cover 55% of England's population (PHE 2015), although as data submission is voluntary, coverage can vary

spatially and temporally. Given that two different computer systems submit data, there are some differences between the information available. Data from EMIS Health is available subdivided by age and sex at SHA level only, whereas data from TPP is available subdivided at lower levels of geography, postcode district, but this data is not subdivided by age and gender. Some syndromic indicators, for example *Acute Bronchitis* are only available from the TPP.

Surveillance of GPIH services was used successfully to monitor the influenza A (H1N1) pandemic in 2009, the system obtained information from 3,400 EMIS practices which covered around 38% of the UK population, 23 million people (Harcourt et al. 2012a). During the pandemic the surveillance system was able to provide real-time data relating to the situation and the overall impact and burden on primary-care services at both national and sub-national level (Harcourt et al. 2012a). The GPIH surveillance system was also vital in providing reassurance to planners of the London Olympic games on the impact on mass gatherings (Todkill et al. 2016; Severi et al. 2012). It was used to monitor community health during the 2010 Icelandic volcanic ash cloud that travelled across the UK; the system was able to show there was no immediate impact to community health using real-time data (Elliot et al. 2010). It has also been used to monitor the health impact of heatwaves (S. Smith et al. 2016a; 2016b; Elliot et al. 2014), assess the possible impact of the rotavirus vaccination campaign (Bawa et al. 2015), monitor the trends of asthma presentations in relation to school terms (Bundle et al. 2019), estimate the socioeconomic and geographical variation in *Allergic Rhinitis* consultations (Todkill, et al. 2017a) and monitor primary-care attendances for *Fever* after meningococcal B vaccination in infants (Harcourt et al. 2018).

2.4 General Practice Out Of-Hours (GPOOH)

General practice out-of-hours or unscheduled care services provide access to healthcare when GP surgeries are closed, typically between 6:30 pm and 8:00 am, at weekends, and during public holidays. This service is for patients who have urgent healthcare needs that cannot wait until their GP practice is open, but where the need is not serious

enough for them to be referred to emergency departments. Prior to 2004, GPs provided out of-hours care either as a practice or as part of a co-operative. From 2004, a change of contract allowed GP practices to opt out of out of-hours care and transfer responsibility to primary care trusts. Through primary care trusts out-of-hours (OOH) services were provided through either GP co-operatives or commercial providers. By 2012, only 10% of GP practices were still responsible for OOH services (Care Quality Commission 2014). The GP's that work in OOH services can be local GP's, dedicated OOH GPs, or locum GP's. With the introduction of NHS111, the provision of GPOOH services changed, with NHS111 able to better sign-post patients to the appropriate level of care, and provide the access to OOH for many of the GPOOH providers, this resulted in a reduction in the number of cases being handled by GPOOH with 8.6 million cases in 2007/08 to 5.8 million in 2013/14 (National Audit Office 2014). In some areas, NHS111 and GPOOH services are provided by the same organisations (Care Quality Commission 2014; National Audit Office 2014). GPOOH services can be accessed through NHS111, but also include walk in centres. GPOOH services do not require pre-registration and can be used by anyone, including foreign visitors.

Syndromic surveillance of the GPOOH service was also introduced as part of the expansion of the syndromic surveillance programme by the HPA in response to the London 2012 Olympic games. Its introduction aimed to fill the gaps identified in the existing monitoring of GPIH services (Harcourt et al. 2012a). The provision of daily data from the GPOOH services to ReSST is voluntary. Each provider submitted automated daily extracts of anonymised consultation activity. Data from the providers included age, gender, postcode district, clinical diagnostic codes, prescribing information and informational outcomes (Harcourt et al. 2012a).

Clinical diagnostic codes used in GPOOH services are based on Read codes. The syndromic indicators used in this surveillance system were developed by aggregating similar Read codes into clinical syndromic indicators. The indicators that were chosen to be monitored by the surveillance system were developed in response to the requirements of enhanced surveillance of mass gatherings, and to the needs in responding to public health incidents (Harcourt et al. 2012b). The syndromic indicators

developed for this surveillance system from Read codes include; *Acute Respiratory Illness, Influenza-Like Illness*, *Difficulty Breathing/Wheeze/Asthma, Gastroenteritis, Vomiting, Diarrhoea,* and *Heatstroke* (Harcourt et al. 2012b). Some data quality issues have been noted with the surveillance of GPOOH, as providing a Read code for consultations is not mandatory for all providers, therefore the level of coding between providers can range from 10% to over 90%. When a code is not present, a syndromic indicator cannot be assigned, making the record unusable for surveillance purposes. This can have a severe impact of monitoring data at the local level, as a provider submits data covering a certain geographical area (Harcourt et al. 2012b). This surveillance system is estimated to cover 80% of the population of England (PHE 2015).

The introduction of the GPOOH surveillance system, along with surveillance of emergency departments, was successfully able to provide reassurance to planners of the London Olympic games on the impact on mass gatherings (Todkill et al. 2016). Since its implementation for the 2012 London Olympic Games, it has successfully been used as part of the programme of syndromic surveillance to monitor the impact of public health events. During the 2013 summer heatwave in England surveillance of GPOOH services, along with surveillance of GPIH, NHS Direct and emergency departments, was used to monitor the impact to public health. The syndromic indicators *Impact of Heat* and *Heat/Sun Stroke* were monitored during the period of interest. Significant increases in these indicators were observed during the heatwave period, and ReSST were able to provide near-real time information to healthcare planners and clinicians (S. Smith et al. 2016a; Elliot et al. 2014). It has also been used to estimate the impact of arthropod bites requiring healthcare (Newitt et al. 2016), assess the likely impact on the Rotavirus vaccination program in England (Bawa et al. 2015), and assess the burden of seasonal respiratory disease on GP services (Morbey et al. 2018).

2.5 Developments in the Surveillance Systems Since 2019

All the data for this thesis was collected from ReSST by the end of 2019. For information we consider developments in the surveillance systems since 2019. Due to

the COVID-19 Pandemic the way healthcare was accessed changed to protect patents and healthcare workers. Those who required advice or support from a GP were required to call their GP surgery for advice on their needs, where a telephone or video call would be booked if required. Only in cases where it was absolutely necessary were patents asked to visit the surgery for in-person appointments. Those who required urgent help through NHS111 services were asked to use the NHS111 online service. This online service is similar to the telephone service where questions are asked to assess your needs. Through this online service people can check if symptoms require further help, see if their symptoms could be due to COVID-19, get emergency prescribed medication, book a COVID-19 vaccination and find out about testing for coronavirus. People were requested not to call NHS111 unless they needed urgent help for a child under-5 years or could not get help online. Those with life threatening emergencies were requested to call 999 and avoid going straight to emergency departments (NHS 2021). These had large impacts upon the data received by ReSST. In response to the changes in healthcare usage ReSST introduced a 6th surveillance system, NHS online which monitored potential COVID-19 assessments from March 2020 and went live in May 2020. A new "potential COVID-19" NHS111 indicator was also developed to monitor COVID-19 cases, this indicator included NHS111 calls triaged through the COVID-19 Pathway and with a COVID-19 disposition (outcome). A "COVID-19-like" indicator was also introduced to the GPIH surveillance system. This new indicator is based on new diagnostic codes released in March 2020 to record patients presenting to primary care services with potential COVID-19 symptoms. It has also been noted by ReSST that due to changes in the way patients have accessed healthcare during the COVID-19 pandemic, syndromic data has to be interpreted with context and caution during this time.

The clinical data received from GP services used in this thesis was based on Read codes. These were used widely up until 2018 when NHS England switched to SNOMED Clinical Terms (CT). By 2020 all services which used Read codes had switched to SNOMED CT. This change allowed clinical information to be recorded consistently and accurately, as well as allowing the simplified exchange of information between

services (for example hospital discharge information no-longer needs to be re-entered into GP systems) (NHS Digital 2021a).

In this thesis the data from the GPIH service was received from GPs who used two software providers, EMIS and TPP. From April 2021 the contract between ReSST and EMIS ended therefore data is no longer used from this provider for syndromic surveillance.

2.6 Considerations for Data Used in this Thesis

Data from the three syndromic surveillance systems described (NHS111, GPIH, GPOOH) are used throughout this thesis. The focus of this thesis is on how syndromic data can be used to monitor respiratory illness. Therefore, we only use indicators that may be indicative of respiratory infections or disease. These indicators are described in Table 2.1. Data obtained for this thesis was aggregated counts (number of presentations for each indicator) per unit time and by geographical location, subdivided by age and gender. The data used in each chapter is described in Table 2.2.

There were some general considerations to be made when obtaining the data. Buckingham-Jeffery et al. (2017) outlined that there are strong influences of weekday, weekend and public holidays on the daily syndromic data. GPIH service are closed during evening, weekends and public holidays, with GPOOH services taking over during this time therefore both these services will have very different usage patterns during these times. NHS111 operates 24/7, 365 days a year, but also observes different usage patterns when other services are not in operation. The same study identified that failing to account for these biasing effects can lead to misinterpretations (Buckingham-Jeffery et al. 2017). Given this situation, we will analyse data at weekly or annual temporal resolutions, while accounting for the number of public holidays during the time period studied.

Syndromic data collects data on healthcare-seeking behaviour not disease incidence. Many factors may influence this including; age, gender, socioeconomic status and

media reporting. One good example of this is Elliot et al. (2016b), which identified the impact of media reporting on public health events. Between 31st of July and 4th August 2015 *Cryptosporidium spp.* oocysts were identified in water treatment works that supplied the public drinking supply in the North West of England. This resulted in a boil water notice being issued, which instructed that water from the public mains was not safe to drink unless boiled first. During this incident there was significant media coverage. Syndromic surveillance was used to monitor increases in gastrointestinal presentations possibly related to this incident. In the areas affected, GPIH consultations for gastroenteritis increased by 24.8% and diarrhoea consultations increased by 28.5%; however, no laboratory confirmed cases associated with the incident were reported. The authors suggested that these observations were caused by changes in healthcare-seeking behaviour driven by the intense media coverage (Elliot et al. 2016b). This highlights that healthcare-seeking behaviours need to be considered when interpreting the findings of any research using this data.

Routinely, ReSST use the total number of presentations to account for daily fluctuations of coverage, the English population that is monitored by the surveillance systems, for the NHS111 and GPOOH systems. For the purposes of this thesis the denominator used was the population of England, and periods of poor coverage were estimated using statistical methods. This change in approach was undertaken because it was noted that there were instances where data on the total number of contacts was collected, but not information on the specific syndromic indicators. By estimating population coverage, we will also be able to provide incidence rates based on the population rather than based on the proportion of presentations, which can be influenced by the number of presentations due to other illnesses. The total number of registered patients monitored by the GPIH surveillance system is collected daily, therefore this was used as the population denominator for analysis of GPIH data.

Another consideration when interpreting findings when using this data is that in both between the surveillance systems and within the surveillance systems there is a hierarchy. NHS111 can be used to access GPOOH services, and can also be used to get appointments for GPIH services. Therefore, the services do not operate in isolation from

each other, and one person can be presented to all three services for the same illness. Therefore, when interpreting findings from these services, it is not possible to equate one presentation to one case if illness, as this could result in double counting. Within the GPIH and GPOOH systems some of the indicators are subsets of other, broader, indicators. For example, the GPOOH indicator *All Respiratory Disease* consists of two other indicators: *Difficulty Breathing/Asthma* and *Acute Respiratory Infection*. This needs to be accounted for in the analysis and interpretation of the data. The hierarchy of the systems and indicators in described in Figure 2.1.

2.7 Conclusion

Here we describe the three community healthcare services monitored by ReSST that will be used in this thesis: NHS111, GPIH and GPOOH, how clinical data is recorded by ReSST. Data from these services is collected in near real-time and comprises of symptoms and clinical diagnosis, before laboratory confirmation can take place. This allows it to detect and monitor public health threats in real-time and in the community. Data has primarily been used to monitor and detect outbreaks and for situational awareness. This thesis will focus on this syndromic data to explore its utility in epidemiological studies.

Service	Indicator	Definition	Chapter Used
NHS 111	Difficulty Breathing	Patients calling NHS 111 and reporting symptoms of breathing problems, breathlessness or wheeze. This syndromic indicator excludes callers reporting symptoms of difficulty breathing which the handler assesses as being of immediate threat to life for which an emergency ambulance is required. These patients would not routinely continue with telephone triage and therefore would not have a Pathway selected.	3, 4, 5, 6
	Cough	Patients calling NHS 111 and reporting symptoms of cough	3, 4, 5, 6
	Cold/Flu	Patients calling NHS 111 and reporting symptoms of colds and/or influenza	3, 4, 6
	Fever	Patients calling NHS 111 and reporting symptoms of fever	4
	Sore Throat	Patients calling NHS 111 and reporting symptoms of sore throat	4
GPOOH	All Respiratory Disease (ARD)	All indicators related to respiratory diseases	3, 4, 6
	Difficulty Breathing/ Wheeze/Asthma (DBWA)	Difficulty breathing/wheeze/asthma - includes codes indicative of asthma, wheeze and difficulty breathing e.g., dyspnoea, bronchial breathing, expiratory wheeze and stridor.	3, 4, 6
	Asthma	Diagnoses of acutely presenting or severe asthma including asthma attack, not including routine consultations (where possible to distinguish and not including prescription links)	3, 4
	Acute Respiratory Infection (ARI)	Acute respiratory infection - includes all codes indicative of an acute respiratory infection e.g., acute sinusitis, viral pneumonia, influenza and pleurisy	3, 4
	Influenza Like Illness (ILI)	Codes indicative of influenza	3, 4, 6
	Bronchitis/Bronchiolitis	Codes indicative of bronchitis and bronchiolitis	3, 4, 6
	Pharyngitis/Scarlet Fever	Codes indicative of pharyngitis and Scarlet Fever	3, 4, 6
GPIH	Upper Respiratory Tract Infection (URTI)	Codes suggestive of an acute bacterial or viral infection mainly affecting the upper respiratory tract (includes colds/flu, sinusitis, throat infections, ear infections).	3, 4, 5, 6
	Influenza Like Illness (ILI)	Codes suggestive of influenza	3, 4, 6
	Pharyngitis or Scarlet Fever	Codes indicative of pharyngitis and Scarlet Fever	4
	Scarlet Fever	Codes indicative of Scarlet Fever	4
	Lower Respiratory Tract Infection (LRTI)	Codes suggestive of an acute bacterial or viral infection mainly affecting the lower respiratory tract including pneumonia, bronchitis, bronchiolitis, pleurisy and complications. Includes codes for the pneumonia indicator.	3, 4, 5, 6
	Pneumonia	Codes for pneumonia	3, 4
	Severe Asthma	Codes for asthma	3, 4
	Bronchitis/Bronchiolitis	Codes for bronchitis and bronchiolitis	3, 4, 6

Table 2.1: Definitions of the syndromic indicators used in this thesis.

Chapter Title	Syndromic Systems Used	Respiratory Indicators Used	Age Group	Spatial Resolution	Temporal Resolution
Demographic and socioeconomic patterns in healthcare-seeking behaviour for respiratory symptoms in England; A comparison with non-respiratory symptoms and between three healthcare services	NHS 111; GPOOH; GPIH	-NHS 111: Difficulty Breathing + Cough + Cold/Flu as a combined respiratory indicator. -GPOOH: Acute Respiratory Infection + Difficulty Breathing/Wheeze/Asthma + Asthma as a combined respiratory indicator. -GPIH: URTI + LRTI as a combined respiratory indicator.	All ages in age bands of: <1, 1-4, 5-14, 15-44, 45-64, 65-74 and >75 years	-NHS 111 and GPOOH: Postcode district. -GPIH: PHE Centre	Annual
Estimating the Community Burden of Respiratory Syncytial Virus (RSV) in Children Under-Five Years, England.	NHS 111; GPOOH; GPIH	-NHS 111: Difficulty Breathing, Cough, Cold/Flu, Fever, Sore Throat. -GPOOH: ARD, DBWA, Asthma, ARI, ILI, Bronchitis/Bronchiolitis, Pharyngitis/Scarlet Fever. -GPIH: URTI, ILI, Pharyngitis or Scarlet Fever, Scarlet Fever, LRTI, Pneumonia, Severe Asthma, Bronchitis/Bronchiolitis	Children under-5 years in age bands: under-1 year and 1-4 years	National	Weekly
Identifying Differences in Regional RSV Community Attributable Burden and Seasonality in Children Under-One Year, England	NHS 111; GPIH	NHS 111: Difficulty Breathing + Cough as a combined respiratory indicator. -GPIH: LRTI	Children under-1 year	-NHS 111: UTLA. -GPIH: PHE Centre	Weekly
Exploring the Effect of Meteorological Conditions on Respiratory Syndromic Indicators in Children Under-Five Years, England.	NHS 111; GPOOH; GPIH	-NHS 111: Difficulty Breathing, Cough, Cold/Flu. -GPOOH: ARD, DBWA, ARI, Bronchitis/Bronchiolitis. -GPIH: URTI, ILI, LRTI, Bronchitis/Bronchiolitis	Children under-5 years	-NHS 111 and GPOOH: UTLA. -GPIH: PHE Centre	Weekly

Table 2.2: Summary of the data used in each chapter of this thesis.

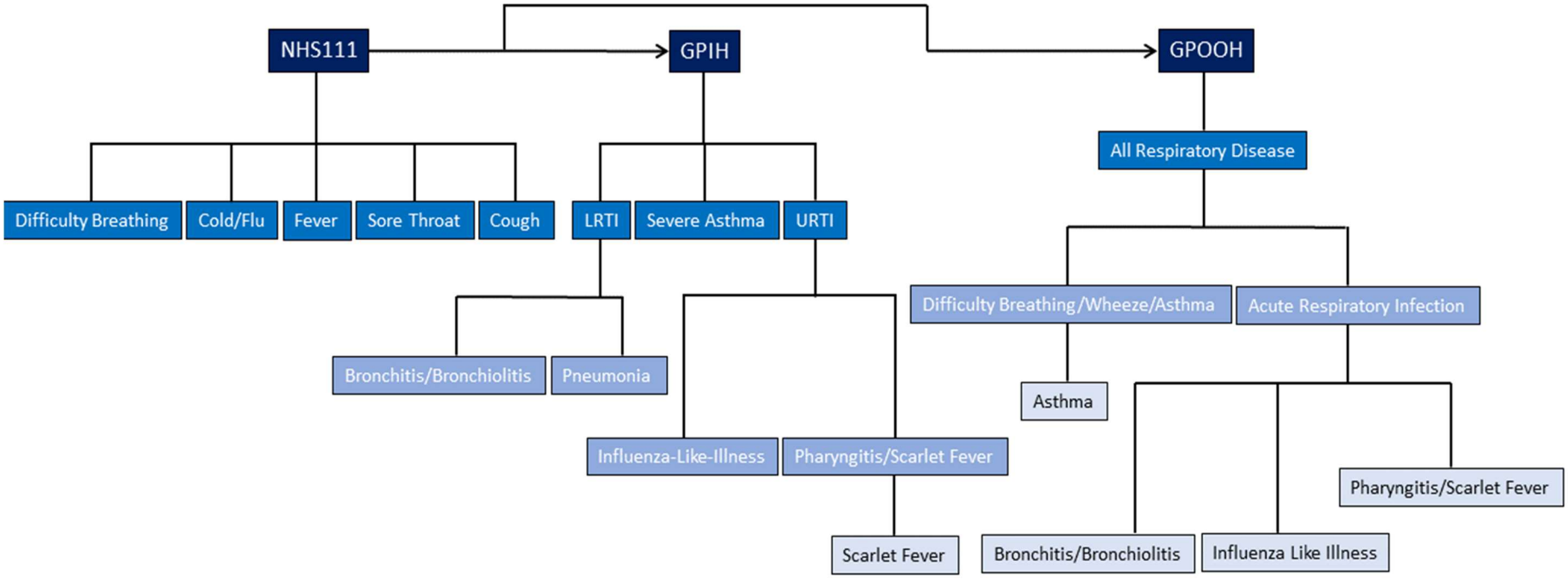

Figure 2.1: Schematic hierarchy of the syndromic surveillance indicators used in the study and the services they relate to, using the example of respiratory conditions.

Chapter 3

Demographic and Socioeconomic Patterns in Healthcare-Seeking Behaviour for Respiratory Symptoms in England; A Comparison with Non-Respiratory Symptoms and Between Three Healthcare Services

3.1 What is Already Known on this Subject?

- Existing studies have demonstrated an association between respiratory illness and age, gender and deprivation, with more healthcare presentations for females, the young and old and those from more deprived backgrounds. These studies have routinely focused on acute measures such as hospitalisations and previous research tends to focus upon respiratory presentations only, and does not compare results to non-respiratory presentations making it difficult to know whether the results are due to underlying aetiologies or demographic of social drivers of healthcare-seeking behaviour.

3.2 What this Study Adds?

- Similar trends were observed between respiratory presentations to three community healthcare services and age, gender and deprivation, these were comparable to associations seen in acute healthcare settings. This highlights the similar demographic patterns of healthcare usage for respiratory diseases at different points of access.

- Patterns for respiratory and non-respiratory presentations, were comparable suggesting similar drivers of healthcare-seeking behaviours for both respiratory and non-respiratory illnesses.
- It was observed that presentation rates were highest in male children under-5 years in comparison to females of the same age, for both respiratory and non-respiratory presentations in the services studied. This observation requires further research as to the social and biological factors drive this relationship.

3.3 Abstract

Acute respiratory diseases are a global issue with 4.7 million disability-adjusted life years lost due to upper respiratory tract infections, over 2.3 million deaths in 2016 due to lower respiratory tract infections, and 358 million people living with asthma, globally. It is crucial to investigate the demographic and socioeconomic patterns in healthcare-seeking behaviour, enabling targeted public health interventions. Previous studies focus upon acute measures of respiratory disease, which can underestimate disease burden. Here we identify the demographic and socioeconomic patterns of over 13 million respiratory healthcare-seeking presentations to three healthcare services which capture more of the disease burden

Generalised Linear Mixed Models were used to estimate the relationship between respiratory presentations to three healthcare services in England (NHS111; telehealth helpline, GPIH (general practitioner in-hours); family doctor services in-hours and GPOOH (general practitioner out-of-hours); unscheduled care) and patient's age, gender, and deprivation. Results were compared between healthcare services, and with non-respiratory presentations.

All services showed similar healthcare-seeking behaviours and similar patterns between respiratory and non-respiratory presentations. More respiratory presentations were observed for females, with 1.59, 1.73, and 1.95 times the rate of presentations to NHS111, GPOOH and GPIH, respectively. When compared to 15-44 year olds, there were 37.32, 18.66, and 6.21 times the rate of respiratory presentations to NHS111, GPOOH and GPIH in children under-1 year. There were 1.75 and 2.70 times the rate of respiratory presentations in the most deprived areas compared to the least deprived to NHS111 and GPOOH. Elevated respiratory presentations were observed for males under-5 years. Healthcare-seeking behaviours between respiratory and non-respiratory presentations were observed to be similar.

When presentation rates in services that capture a more of the disease burden are explored, the demographic patterns are similar to those observed in acute settings.

Comparable results were observed between respiratory and non-respiratory presentations suggesting that when a wider spectrum of disease is explored, demographic and socioeconomic factors may be the strongest influencers of healthcare-seeking behaviour. Higher presentation rates in male's under-5 years for both respiratory and non-respiratory presentations require further research.

3.4 Introduction

Acutely presenting respiratory diseases, henceforth referred to as respiratory disease, including upper (URTI) and lower respiratory tract infections (LRTI) and asthma, have a substantial impact on individual health and healthcare systems. Globally, respiratory diseases are the second largest cause of lost disability-adjusted life years (DALYs) with an estimated one billion people suffering from acute or chronic respiratory diseases, and they result in four million premature deaths annually (WHO 2014; Nair et al. 2010). Globally, LRTI were the sixth leading cause of death, causing over 2.3 million deaths in 2016 (Troeger et al. 2018). URTI have a substantial impact on health burden with over 4.7 million DALYs lost globally in 2016 (Hay et al. 2017). Although asthma is a chronic condition, it often presents with acute exacerbations. Worldwide, as of 2015, around 358 million people were living with asthma (Soriano et al. 2017).

Respiratory disease is a particular issue in the United Kingdom (UK). In the Global Burden of Disease Study 2010, the UK had the second highest number of age-standardised DALYs due to LRTI and asthma out of 19 other high-income countries (C. Murray et al. 2013; GBDCN 2012). In comparison to 15 other European countries, the UK had one of the highest mortality rates due to respiratory infections (Mukherjee et al. 2016). In the UK, 15.8% of the population are predicted to develop asthma in their lifetime (Mukherjee et al. 2016). LRTI and URTI are estimated to cost the UK over £1.7 billion, and asthma over £3 billion annually (Trueman, Woodcock and Hancock 2017).

To help develop and target interventions it is necessary to understand the sections of society most at risk from respiratory disease. In the UK, deprivation has repeatedly been linked to both respiratory morbidity and mortality (Al Sallakh et al. 2017; Jordan et al. 2006; Riaz et al. 2011; Roberts et al. 2012; Snell et al. 2016). Factors attributed to higher rates of respiratory disease among more deprived areas include increased rates of smoking (Newton et al. 2015), higher levels of outdoor air pollution (Sofianopoulou et al. 2013), poorer quality housing (Gibson et al. 2011) and occupational hazards due to manual labour (Pleasants, Riley and Mannino 2016). Reduced prevalence of asthma (Iversen et al. 2005) and fewer deaths due to respiratory disease have been observed in more rural areas (Gartner et al. 2008).

There is a complex relationship between respiratory diseases and gender, with many respiratory diseases affecting men and women differently in both prevalence and severity. There is growing evidence to suggest that there is a greater female incidence of chronic obstructive pulmonary disease (COPD), asthma, cystic fibrosis (CF) and non-CF bronchiectasis (Raghavan and Jain 2016). Respiratory tract infections, on the other hand, occur at higher rates in males and tend to be more severe leading to higher mortality rates (Falagas, Mourtzoukou and Vardakas 2007). Although, females have higher rates of presentation with URTI compared to males (Falagas, Mourtzoukou and Vardakas 2007), males succumb more to LRTI, which are more severe than URTI and can lead to higher mortality rates (Falagas, Mourtzoukou and Vardakas 2007). Gender differences in respiratory diseases are affected by age, with male children more prone to illness, hospitalisation, and death due to respiratory diseases (Liptzin, Landau and Taussig 2015). This interaction is particularly prevalent in asthma, with male children twice as likely to develop the condition, but by adolescence the prevalence equalises between the genders (Vink et al. 2010). By adulthood there is a female predominance with women over 35 years having a 20% higher risk of asthma (Leynaert et al. 2012).

Most of the studies discussed use laboratory confirmed, or diagnostic health data as measures of respiratory health from laboratory or hospital-based surveillance. One prominent issue is that these data are known to represent a fraction of the overall healthcare burden, and often do not capture cases that occur in the community and non-

acute healthcare settings. This under-reporting of disease burden has been shown to be influenced by factors such as severity of symptoms and health literacy (MacDougall et al. 2008). Under-ascertainment occurs when an individual does not seek healthcare, and therefore cannot be captured by any surveillance system. Melbye et al. (2012) reported that only 5.1% of individuals with respiratory symptoms reported seeking medical consultations, highlighting the high degree of under-ascertainment. In the case of chronic respiratory diseases such as COPD and asthma, cases often go undiagnosed until disease is apparent and moderately advanced (Okoromah and Oviawe 2002; Ford et al. 2013). A survey conducted in the UK found that of those with recent respiratory tract infection symptoms (n = 1000), only 19.7% visited their general practitioner (GP) (McNulty et al. 2013). Figure 3.1 illustrates the pyramid of infection in disease surveillance, and how the healthcare services used in this study can capture a wider burden of disease, compared to other forms of surveillance.

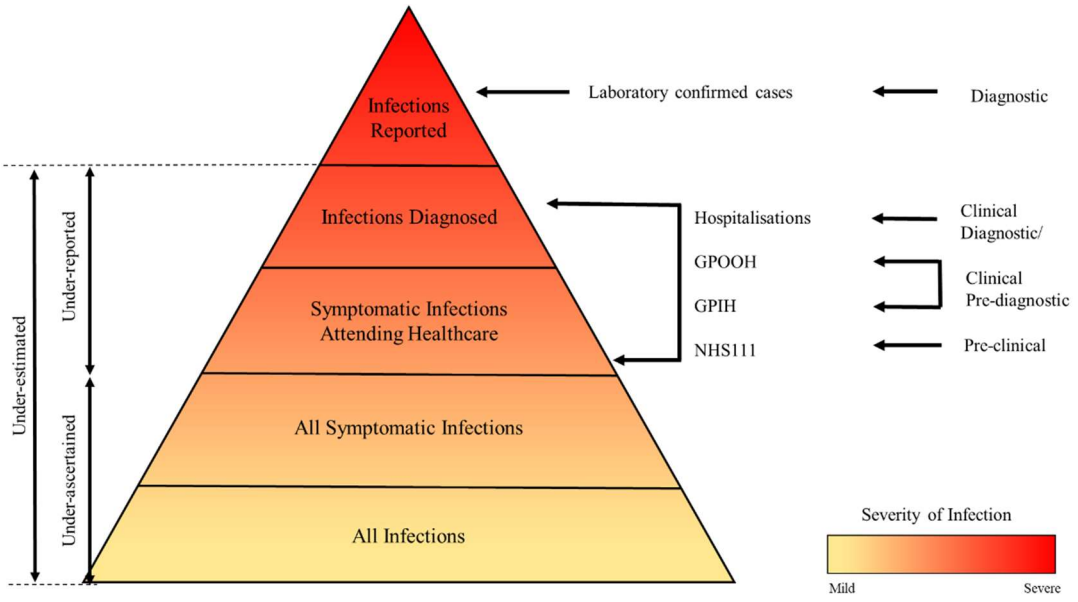

Figure 3.1: Morbidity surveillance pyramid estimating number of infection cases reported adapted from Gibbons et al. (2014) and Berger, Shiau and Weintraub (2006).

Here, we focus on the sociodemographic patterning of respiratory disease in England. Respiratory diseases can often be self-limiting, and therefore may be underreported in national surveillance from traditional data sources such as laboratory reports and hospitalisations. This can lead to bias in the reported relationship between healthcare-seeking behaviour for respiratory diseases, and sociodemographic factors. Unlike previous studies, which often focus on traditional data sources we utilise data sources from telehealth, family doctors and unscheduled care that may provide a more complete reflection of community burden. The data used in this study is defined as non-specific, pre-diagnostic-syndromic data which we use as a proxy for disease (Triple S Project 2011).

Syndromic surveillance has previously been used to investigate the links between age, gender and deprivation with disease. Todkill et al. (2017a) used syndromic surveillance records of family doctor consultations, which comprised of approximately 35 million registered patients in England, for allergic rhinitis to investigate socioeconomic and geographical variation. This study was conducted at upper tier local authority, a coarse geographic level of 157 authorities in England, with an average of 364,664 residents per authority. Higher rates were found in females, children and those from more deprived areas. Adams et al. (2018) utilised 24 million gastrointestinal related calls to a telehealth system (NHS111) in England to investigate the relationship between deprivation and gastrointestinal infections, a higher risk of gastrointestinal infection in more deprived areas was observed. Baker et al. (2012) conducted a large-scale study monitoring infectious diseases in New Zealand, using hospitalisation codes (ICD-9 and ICD-10). All hospitalisations for infectious and non-infectious diseased from 1989 to 2008 were included. It was observed that there were higher rates of hospitalisation due to infectious diseases in the most socioeconomically deprived quintile compared with the least deprived quintile. de Lusignan et al. (2016), used GP coded diagnosis and symptoms to conduct a large-scale investigation on the impact of age, gender, ethnicity and deprivation in England on six common conditions including three respiratory illnesses; the common cold, pneumonia and influenza-like illness. People in the most deprived quintile had a higher probability of presenting with the common cold, whereas

there was no clear deprivation trend in people presenting with pneumonia or influenza-like illness. Charland et al. (2011) used medical billing claims for outpatient clinic and emergency department visits to investigate the relationship between deprivation with burden of influenza infection in Quebec, Canada. There was no evidence of a relationship between material deprivation and influenza burden, however, rates of healthcare utilisation due to influenza decreased as social deprivation increased. Nilsson and Laurell (2005), investigated associations between antibiotic prescriptions for penicillin-non-susceptible Streptococcus pneumonia (PNSP) and deprivation in Sweden. Although deprivation was not linked with higher rates of PNSP, higher deprivation was associated with increased rates of antibiotic prescribing.

One issue with the use of syndromic data in such studies is the difficulty in ascertaining whether associations arise from healthcare-seeking behaviours or disease incidence. This situation could be explored by comparison of results to all healthcare-seeking behaviours, but this appears absent in the literature. Furthermore, most studies utilise a single source of syndromic data making it difficult to know the generability of results.

This study aims to:

1. Explore association between respiratory-related presentations to three community healthcare services; telehealth, family doctors and unscheduled care, and age, gender and ecological measures of deprivation, in England.
2. Compare these results to all non-respiratory presentations to identify whether associations are specific to respiratory disease.

3.5 Methods

3.5.1 Data Collection

Epidemiological Data

The Real-Time Syndromic Surveillance Team (ReSST), part of Public Health England (PHE), coordinates a national programme of syndromic surveillance of multiple healthcare services (PHE 2015). Our study utilises routinely available syndromic data from a telehealth service (NHS111) which operates continuously all year; a family doctor service (general practitioner in-hours; GPIH) that operates during weekday working hours, and an out-of-hours family doctor service (general practitioner out-of-hours; GPOOH) (Table 3.1). These services run as part of the National Health Service (NHS) which is universal and free at the point of delivery.

Syndromic data, obtained from the three surveillance systems coordinated by PHE, comprised of annual counts of respiratory and non-respiratory presentations received between 1 January 2015 and 31 December 2016. Syndromic indicators classified as respiratory for this study are presented in Table 3.1 and are described in Table 3.2. Previous studies have demonstrated the association between acute respiratory diseases and these syndromic indicators (Morbey et al. 2017a; 2017b). Based on expert knowledge these indicators were chosen to characterise acutely presenting respiratory illnesses such as asthma, and respiratory infections, and to be as comparable as possible between the three healthcare services. Respiratory presentations for each service comprised of the sum of the respiratory indicators listed in Table 3.1. Non-respiratory counts comprised of the difference between the total number of presentations and the number of respiratory presentations. The total number of presentations was not available through the GPIH surveillance system, and therefore non-respiratory counts were not available.

Data were obtained at the finest geographical level available; postcode district (PD) (e.g., SW1) for NHS111 and GPOOH, and PHE Centre (e.g., London) for GPIH. In

England, there are 2,234 PD with, on average, 25,660 (Range:142-162,266) residents. There are nine PHE Centres with an average of 6,181,375 (Range:2,644,727-9,080,825) residents in England. Count data were subdivided by age group: under-1 year, 1-4, 5-14, 15-44, 45-64, 65-74 and over-75 years; by gender (male/female), year (2015/2016); and by geographic location.

Healthcare Service	Healthcare Service Provided	Coding System for Healthcare Service	Presentation Type	Routine syndromic indicator included for this study	Number of Presentations used in study
NHS111	NHS111 is a free non-emergency medical helpline. It operates 24/7 and is staffed by trained call handlers. A clinical decision support system is used to structure the response to the call, with the call disposal ranging from advice about self-care to dispatch of an emergency ambulance.	NHS Pathways (NHS Digital 2020)	Acute respiratory presentations	"Cold/flu", "cough" and "difficulty breathing"	1,721,034
			Total number of presentations	All presentations	21,242,154
			Non-respiratory presentations	All non-acute respiratory presentations (All presentations – acute respiratory presentations)	19,521,120
GPIH	GPIH are primary care services that provide free scheduled day-to-day healthcare in England. General practitioners (GPs) treat all common medical conditions and depending on the condition will refer patients to hospitals and other medical services for urgent and specialist treatment.	Read codes v2 (hierarchical) and v3 (non-hierarchical). Full description in Robinson et al. (Robinson et al. 1997)	Acute respiratory presentations	"Upper and lower respiratory tract infections" and "asthma"	10,310,626
			Total number of presentations	Not available	Not available
			Non-respiratory presentations	Not available	Not available
GPOOH	GPOOH services provide free access to primary healthcare when GPIH services are closed, which is typically weekdays 6:30pm – 8:00am, weekends and bank holidays.	Read codes (Robinson et al. 1997)	Acute respiratory presentations	"Acute respiratory infection" "difficulty breathing/wheeze/asthma"	1,562,883
			Total number of presentations	All presentations	8,500,540
			Non-acute respiratory presentations	All non-respiratory presentations (All presentations – acute respiratory presentations)	6,937,657

Table 3.1: Syndromic indicators indicative of respiratory diseases for each syndromic surveillance system and the number of presentations received from each healthcare service, 1st January 2015 to 31st December 2016.

Healthcare Service	Syndromic Indicator	Description
NHS111	Difficulty breathing	Patients calling NHS111 and reporting symptoms of breathing problems, breathlessness or wheeze This syndromic indicator excludes those callers' reporting symptoms of difficulty breathing which the call handler assesses as being of immediate threat to life for which an emergency ambulance is required. These patients would not routinely continue with telephone triage and therefore would not have a Pathway selected.
	Cough	Patients calling NHS111 and reporting symptoms of cough
	Cold/flu	Patients calling NHS111 and reporting symptoms of colds and/or influenza
GPIH	Upper respiratory tract infection (URTI)	Codes suggestive of an acute bacterial or viral infection mainly affecting the upper respiratory tract (includes colds/flu, sinusitis, throat infections, ear infections).
	Lower respiratory tract infection (LRTI)	Codes suggestive of an acute bacterial or viral infection mainly affecting the lower respiratory tract including pneumonia, bronchitis, bronchiolitis, pleurisy and complications. Includes codes for the pneumonia indicator.
GPOOH	Asthma	Diagnoses of acutely presenting or severe asthma including asthma attack, not including routine consultations (where possible to distinguish and not including prescription links)
	Acute respiratory infection (ARI)	Acute respiratory infection - includes all codes indicative of an acute respiratory infection eg. acute sinusitis, viral pneumonia, influenza and pleurisy
	Difficulty breathing/wheeze/asthma	Difficulty breathing/wheeze/asthma - includes codes indicative of asthma, wheeze and difficulty breathing. Includes eg. dyspnoea, bronchial breathing, expiratory wheeze and stridor.

Table 3.2: Descriptions of syndromic indicators used in study.

Demographic Data

Total population of PD, subdivided by age, gender and postcode district, was used as a denominator in the analysis of NHS111 and GPOOH presentations (UK Data Service 2017). These population data were based upon 2011 data and so was adjusted by local authority level population change to account for the change in population between 2011 and 2015. GPIH populations (subdivided by age and sex) were derived from the annual sum of daily number of registered populations at each GP, this demographic breakdown was only possible at PHE centre from the GPIH surveillance system.

Independent Variables

The Index of Multiple Deprivation (IMD) was used as an area level measure of deprivation. IMD scores were obtained at Lower Layer Super Output Area (LSOA), which have been generated to have an average population of 1,500 (DCLG 2015). This index in calculated from seven domains; income, employment, education, health (premature death and poor physical/mental health), crime, barriers to housing and services, and living environment. This index was used, rather than more specific variables (e.g., smoking behaviour) to avoid issues with collinearity. A weighted mean for each PD was calculated using the portion of LSOA in each PD. The weighted mean scores for each PD were then equally divided into quintiles of most (1) and least (5) deprived.

The Office for National Statistics Rural Urban Classification were used to obtain the percentage of PD classified as urban (DEFRA, 2017).

3.5.2 Data Cleaning and Exploration

NHS111 respiratory presentations from Essex in 2015 & 2016 and Norfolk in 2016 were excluded because syndromic data was unavailable. Syndromic surveillance coverage maps of GPOOH were obtained at the upper tier local authority (UTLA) geographical level, and data were excluded from any UTLA where the PHE surveillance programme received little or no syndromic data.

For all systems, data were excluded if location, age or gender of healthcare seeker was unknown (Table 3.3). PDs that were demarcated for large organisations (e.g., Heathrow Airport), or had less than 200 residents were excluded from analysis as their small populations or unique nature are unlikely to produce reliable estimates. PDs that overlap borders with Scotland and Wales were excluded. For each system, presentation rates were mapped for both study years to visualise spatial variation of the data (Figure 3.2).

For each system, outliers were investigated in the dependent and independent variables using Cleveland plots. Multicollinearity within the independent variables was investigated using Variance Inflation Factor (VIF), with a threshold of 10 being used (Hair et al. 2010).

	NHS111 (% of total)	**GPOOH (% of total)**
Total number provided	21,905,099	9,623,939
Reason for exclusion		
No valid postcode provided/Not in England	613,495 (2.9)	92,815 (9.6)
No gender given	9,536 (0.04)	12,312 (0.13)
No age given	-	1,601 (0.017)
Postcode District with < 200 population	44 (0.0002)	12 (0.0001)
Overlapping borders with Scotland or Wales	17,072 (0.08)	3,795 (0.04)
Large area users/City centres	1,517 (0.007)	2,379 (0.03)
Poor coverage/Data Issues	21,281 (0.1)	1,010,485 (10.50)
Total excluded	**662,945 (3.02)**	**1,123,399 (11.67)**

Table 3.3: Number of presentations excluded from study by reason for exclusion. [+]GPIH had no exclusions.

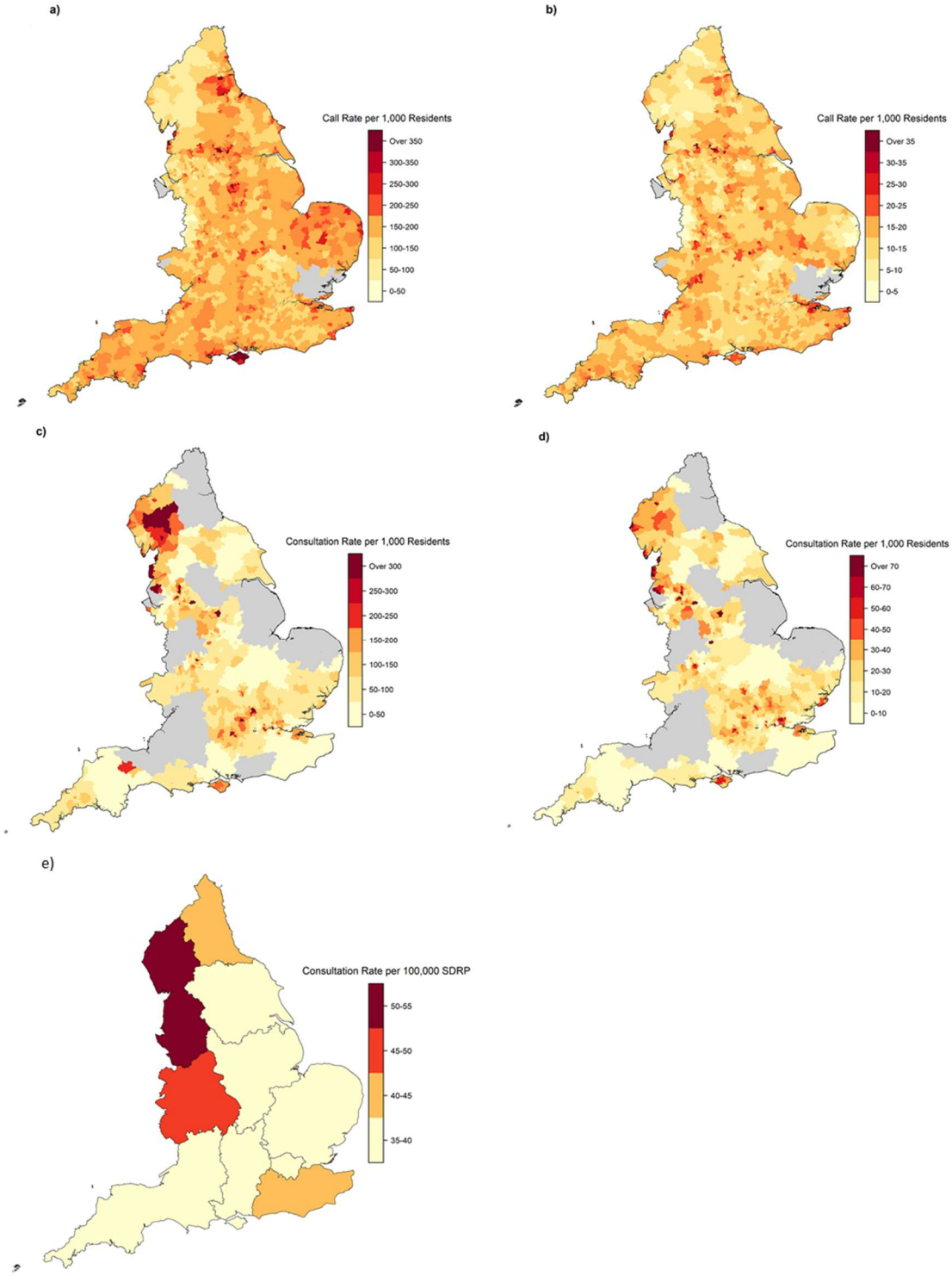

Figure 3.2: Spatial variation of presentations to three health services in England between 1st January 2015 & 31st December 2016: a) All non-respiratory calls to NHS111 at PD, b) Respiratory calls to NHS111 at PD, c) All non-respiratory consultations to GPOOH services at PD, d) Respiratory consultations to GPOOH services at PD, e) Respiratory calls to GPIH services PHE centre. SDPR = sum of daily registered population. The grey areas indicate the PD that were excluded for the duration of the study period.

3.5.3 Statistical Analysis

To measure the relationship between the dependent and independent variables generalised linear mixed models (GLMMs) were used for the NHS111 and GPOOH data, while a generalised linear model (GLM) was used for the GPIH data. GPIH data was modelled using both GLM and GLMM methods, however the GLM provided a slightly better model fit (Table 3.4).

	GLM		GLMM	
Variable	Rate Ratio	95% CI	Rate Ratio	95% CI
Intercept	142.693	0.000 –1.455e+16	142.264	0.000-2.619e+16
Main Effects				
Age		***		***
Under 1 year	6.213	5.955 – 6.482***	6.213	5.950-6.487***
1-4 years	3.492	3.348 – 3.643***	3.492	3.450-3.645***
5-14 years	1.172	1.123 – 1.223***	1.172	1.123 - 1.223***
15-44 years	ref	ref	ref	ref
45-64 years	1.032	0.989 – 1.076	1.032	0.988 - 1.077
65-74 years	1.170	1.122 – 1.221***	1.170	1.121 - 1.222***
75 years and above	1.500	1.438 – 1.564***	1.500	1.436 - 1.566***
Gender				***
Female	ref	ref	ref	ref
Male	0.513	0.492 – 0.535***	0.513	0.492-0.536***
PHE Centre	-	***	-	-
Year	-	ns	-	ns
Interaction Effects				
Age:Gender	-	***	-	***
AIC		5083.492		5117.537
Deviance Explained		0.993		0.992
Dispersion Statistic		1.100		1.0332

Table 3.4: Comparison between GLM and GLMM results of respiratory calls to GPIH (ns = not significant (overall effect only), * = ≤0.05, ** = ≤0.01, *** = ≤0.001).

The variables of interest for analysis of NHS111 and GPOOH data were age, gender, and deprivation. Two-way interactions between age and gender, and age and deprivation were also investigated. A categorical variable for year was included in the model to account for inter-annual variation. Percentage urban area was included to account for differences in healthcare-seeking behaviour or disease risk related to urbanicity. Due to the highly aggregated deprivation and percentage urban data at PHE centre level these variables were not included in the analysis of GPIH dataset, as non-

significant estimates would likely be due to the lack of variation within this aggregated data. Therefore, the variables of interest age and gender and their interactions were investigated in the analysis of GPIH data. To account for population differences at the geographical level, the logarithm of the population plus one was included as a model offset.

The study design accounted for the hierarchical structure of the data for each system by including PD and UTLA as random effects in the NHS111 and GPOOH models. UTLA was included as a random effect to account for similar characteristics of neighbouring PD to reduce the effect of spatial autocorrelation (map of UTLA overlaid PD in Figure 3.3). Where a PD was located in more than one UTLA, the largest PD area was allocated. PHE Centre was included as a fixed effect in the GPIH model. These geographical areas were also included in the analysis to account for area specific differences that may influence healthcare-seeking behaviour.

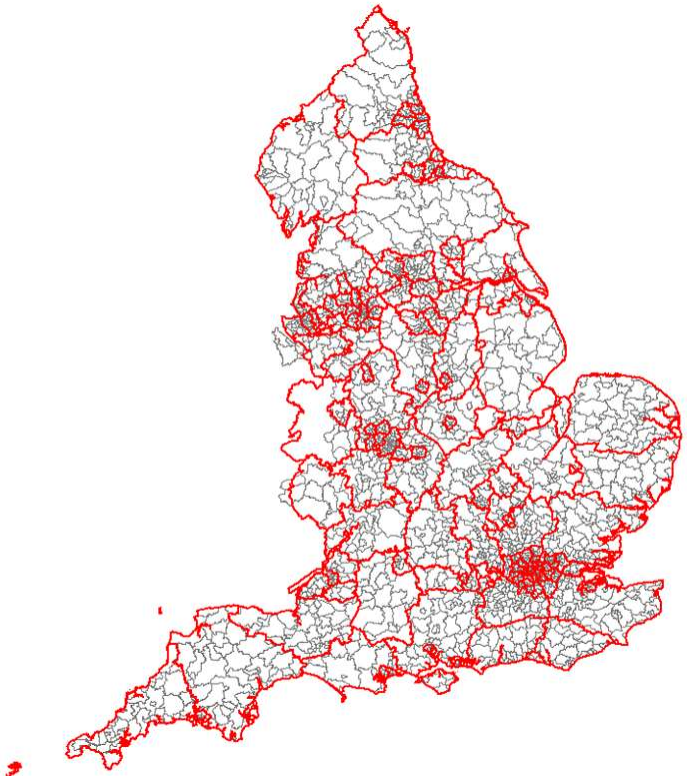

Figure 3.3: Map of upper teir local authorties over-layed postcode dictricts. Upper-tier local authorities (red) overlapped over PD (grey).

We explored Poisson and negative binomial model specifications to account for potential over-dispersion in the data. Over-dispersion was tested by comparing the sum of squared Pearson residuals to the residual degrees of freedom. Models with over-dispersion statistics <1.5 were deemed acceptable (Payne et al. 2018). Wald tests were used to determine the overall significance of variables.

The general algebraic definition of the models is given by:

$$Ya,g,i,y \mid \mu a,g,i,y,\varphi \sim NegBin(\mu a,g,i,y,\varphi),$$

Where $\mu_{a,g,i,y}$ is the is the number of age a and gender g specific presentations at each geographical location (PHE centre or postcode district) i at year y, and $\varphi > 0$ is the negative binomial dispersion parameter.

The expected number of cases for NHS111 and GPOOH services is modelled as:

$$\log(\mu a,g,i,y) = \alpha + \log(Pa,g,i,y) + \beta a + \gamma g + \delta y + \zeta d + \eta p + \beta a \cdot \gamma g + \beta a \cdot \zeta d + ui + vl$$

Where α corresponds to the intercept; $\log(P_{a,g,i,y})$ denotes the logarithm of the population P at risk for postcode district i, age a, gender g and year y included as an offset to adjust counts by population at risk; a denotes a categorical variable age with coefficient β; g denotes a categorical variable gender with coefficient γ; y denotes a categorical variable for year with coefficient δ; d denotes a categorical variable for deprivation with coefficient ζ; p denotes a continuous variable to represent the percentage of postcode district that is classified as urban with coefficient η. $\beta a \cdot \gamma_g$ denotes the interaction between age and gender, and $\beta a \cdot \zeta_d$ denotes the interaction between age and deprivation. Unstructured random effects of postcode district (u_i) and upper tier local authority (v_l) were included to account for unknown confounding factors at the geographical level, the hierarchical structure of the data and spatial dependency.

The expected number of cases for GPIH services is modelled as:

$$\log(\mu_{a,g,i,y}) = \alpha + \log(P_{a,g,i,y}) + \beta a + \gamma_g + \delta_y + \epsilon_i + \beta a \cdot \gamma_g$$

Where α corresponds to the intercept; $\log(P_{a,g,i,y})$ denotes the logarithm of the population P at risk for PHE centre i, age a, gender g and year y included as an offset to adjust counts by population; a denotes a categorical variable age with coefficient β; g denotes a categorical variable gender with coefficient γ; y denotes a categorical variable for each year with coefficient δ; i denotes a categorical variable PHE centre with coefficient ϵ. The interaction between age and gender is denoted by $\beta a \cdot \gamma_g$.

Model overfitting and the predictive ability of the model was assessed using k-fold cross-validation where the data was split into 10 equal groups (k). Each group was used to train the model k-1 times and test the model once. To assess the model, mean absolute error was used and the results are presented in Table 3.5. Rate ratios (RR) with 95% confidence intervals (CI) were estimated for the main effects: age, gender and deprivation. To allow the visualisation of the main and interaction effects, and comparisons of trends between the presentation types and services, the number of presentations to each service were predicted using the models, then standardised, and plotted. The predictions were standardised to a zero mean and unit variance by subtracting the mean of the predictions from each predicted value and then dividing by the standard deviation.

All analyses were conducted in R version 3.5.2 (RStudio Team 2015; R Core Team 2017) and models were specified using the glmmTMB (Brooks et al. 2017) and MASS (Venables and Ripley 2002) packages.

	NSH111		GPOOH		GPIH
	Respiratory	Non-Respiratory	Respiratory	Non-Respiratory	Respiratory
Max presentations per location	391	17,842	2,108	10,175	200,304
Min presentations per location	0	0	0	0	4,274
Mean presentations per location	30.9	350.3	39.2	173.9	36,823.7
Mean of the mean absolute error	7.02	54.6	9.5	40.5	2,533.7

Table 3.5: Cross validation results of final model to assess model overfitting.

3.6 Results

3.6.1 Model Selection

From Table 3.3 it can be observed that relatively few presentations were excluded from analysis due to unknown location, age or gender from NHS111, with only 3.0% presentations excluded. More presentations were excluded from GPOOH with, 11.7% of data excluded. GPIH data had no exclusions because patients that used this service have to pre-register and therefore location, age and gender are known. Any data issues observed in the datasets were because of the passive reporting to the surveillance systems and not due to disruption of healthcare services. Table 3.3 demonstrates that in total 21,242,154 presentations to NHS111 were included in the analysis, of which 8.10% (n = 1,721,034) were respiratory presentations; 6,937,657 GPOOH presentations were included in the analysis of which 22.53% (n = 1,562,883) were respiratory presentations. The different proportions of respiratory presentations between the NHS111 and GPOOH services likely reflect to the different functions of the services, and the severity of illness for which each service would be contacted by patients. Total number of presentations were not available for GPIH, but 10,310,626 respiratory presentations were included in the analysis of GPIH data.

Two model distributions were considered for analysis: Poisson, and negative binomial, with models selected by considering over-dispersion of the data. Negative binomial models handled the over-dispersed data best in all five models (descriptions of model

fit in Table 3.6). Overall, the models performed well with low mean of mean absolute error values (Table 3.5). The fixed effects explained a high amount of variation in both NHS111 and GPIH respiratory models, with a marginal R^2 of 0.86 and an R^2 of 0.99, respectively (Table 3.4). The fixed effects explained less of the variation in the GPOOH respiratory model, with a marginal R^2 of 0.30. When the spatial random effects were considered in the NHS111 and GPOOH respiratory models, both models had a conditional R^2 of over 0.94. In the GPOOH models the difference observed between marginal and conditional R^2 values compared to the NHS1111 models is because less of the variation in the data can be explained by the fixed effects. The difference between the marginal and conditional R-squared in the GPOOH models is likely due to the uncertainty of the underlying study population, with much of this uncertainty explained by including the geographical levels (PD and UTLA) as random effects.

3.6.2 Multivariable Analysis

Standardised predictions, the number of standard deviations from the mean, of the multivariable analysis for each model are visualised in Figures 3.4 to 3.6, alongside the overall significance of each model. The table of main effect RR and 95% CI are presented in Table 3.6.

Respiratory presentations in the under-1 year age group are significantly higher compared to the reference group (15-44 years) in all three services (Table 3.6), with 37.32 (95% CI:36.10-38.85), 18.66 (95% CI:17.78-19.58), and 6.21 (95% CI:5.96-6.48) times the rate of presentations to NHS111, GPOOH and GPIH, respectively. The comparative differences between respiratory presentations in the under-1-year age group and the reference group was highest for NHS111 compared to GPOOH and GPIH. Presentations are highest in the under-1 year age group compared to the reference group in both the non-respiratory models; similar RR were observed between the services; with 7.03 (95% CI:6.87-7.20), and 7.64 (95% CI:7.28-8.02) times the rate of non-respiratory presentations to NHS111, and GPOOH in the under-1 year age group compared to the reference group. This relationship is visualised in Figure 3.4a.

Gender has a significant influence on presentations to each service (Table 3.6); visually the trend appears similar between respiratory and non-respiratory presentations and across all services (Figure 3.4b). There are 1.59 (95% CI:1.56–1.62), 1.73 (95% CI:1.70-1.77), and 1.95 (95% CI:1.87-2.03) times the rate of respiratory presentations regarding females to NHS11, GPOOH and GPIH, respectively.

Deprivation is significant in both NHS111 models (Table 3.6), which found there are 1.75 (95% CI: 1.56-1.95) times the rate of respiratory presentations in the most-deprived areas compared to the least-deprived (IMD quintile 1 vs 5), and 1.81 (95% CI:1.66-1.99) times the rate of non-respiratory presentations. Deprivation was significant in both the respiratory and non-respiratory GPOOH models, with 2.70 (95% CI:1.79-4.08) times the rate of respiratory presentations and 2.70 (95% CI:1.89-3.85) times the rate of non-respiratory presentations in most-deprived areas compared to least-deprived. Similar RR estimates for deprivation were observed in respiratory and non-respiratory NHS111 and GPOOH presentations. This relationship is visualised in Figure 3.4c, where the similarities between NHS111 and GPOOH, and respiratory and non-respiratory presentations can be observed.

Age-gender interactions show similar trends across system types and when compared to non-respiratory presentations. Overall, there are more female presentations; but in all models (Figure 3.5), in the under-1 and 1-4 year age groups there were more presentations regarding males.

Age-deprivation interactions were investigated for NHS111 and GPOOH. In both respiratory and non-respiratory presentations to NHS111 (Figure 3.6a), the trends suggest there are more presentations in the most deprived quintiles across all ages. For respiratory presentations, there is a stronger trend with deprivation in the under-1, 1-4 and over-75-year age groups, and a weaker trend in the remaining age groups. This NHS111 trend is strong across all age groups for non-respiratory presentations, particularly in the age group over-75 years. There was a similar linear trend with deprivation and age for respiratory and non-respiratory presentations to GPOOH (Figure 3.6b), with more presentations in the most-deprived areas in all age groups.

	NHS111				GPOOH				GPIH	
	Respiratory Calls		Non-respiratory Calls		Respiratory Calls		Non-respiratory Calls		Respiratory Calls (PHE Centre	
Variable	**Rate Ratio**	**95% CI**	**Rate Ratio**	**95% CI**	**Rate Ratio**	**95% CI**	**Rate Ratio**	**95% CI**	**Rate Ratio**	**95% CI**
Intercept	0.005	0.004-0.005	0.106	0.099-0.115	0.002	0.001-0.003***	0.015	0.009-0.023***	142.693	0.000 – 1.455e+16
Main Effects										
%Urban	1.002	1.002-1.003***	1.001	1.001-1.002***	1.004	1.003-1.006***	1.004	1.003-1.006***	x	x
Age	–	***	–	***	–	***	–	***	–	***
Under 1 year	37.324	36.104-38.850***	7.031	6.869-7.197***	18.661	17.783-19.581***	7.640	7.281-8.018***	6.213	5.955 – 6.482***
1 – 4 years	16.683	16.177-17.206***	3.220	3.149-3.292***	9.822	9.400-10.263***	2.782	2.656-2.913***	3.492	3.348 – 3.643***
5 – 14 years	1.983	1.915-2.054***	0.871	0.851-0.891***	1.879	1.794-1.966***	0.791	0.755-0.829***	1.172	1.123 – 1.223***
15-44 years	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
45-64 years	0.854	0.826-0.883***	0.670	0.656-0.685***	0.515	0.492-0.540***	0.795	0.761-0.831***	1.032	0.989 – 1.076
65-74 years	1.372	1.324-1.423***	0.912	0.892-0.932***	0.674	0.640-0.709***	1.247	1.192-1.304***	1.170	1.122 – 1.221***
75 years and over	3.999	3.872-4.130***	2.673	2.616-2.732***	2.500	2.388-2.617***	4.029	3.855-4.211***	1.500	1.438 – 1.564***
Deprivation Quintile	–	***	–	***	–	***	–	***	–	–
(Most Deprived) 1	1.745	1.561-1.949***	1.814	1.658-1.985***	2.704	1.792-4.081***	2.700	1.892-3.853***	x	x
2	1.382	1.291-1.480***	1.387	1.313-1.466***	1.654	1.326-2.062***	1.659	1.374-2.003***	x	x
3	1.172	1.104-1.245***	1.179	1.124-1.236***	1.318	1.089-1.596**	1.231	1.045-1.451*	x	x
4	1.075	1.019-1.133**	1.085	1.040-1.132***	1.153	0.986-1.153	1.200	1.050-1.372**	x	x
(Least Deprived) 5	ref	ref	ref	ref	ref	ref	ref	ref	x	x
Gender	–	***		***	–	***	–	***	–	–
Male	ref	ref	ref	ref	ref	ref	ref	ref	ref	ref
Female	1.588	1.558-1.618***	1.748	1.722-1.774***	1.733	1.697-1.770***	1.711	1.665-1.758	1.949	1.868-2.033***
PHE Centre	–	***	–	***	–	***	–	***	–	***

Table 3.6: Final model multivariable regression analysis of respiratory and non-respiratory presentations to three healthcare services. Full results and overall significance presented for main effects, and overall significance for interaction terms (ns = not significant, * = ≤0.05, ** = ≤0.01, *** = ≤0.001, x = variable not modelled, - = reference level).

Year	–	***	–	***	–	***	–	***	–	ns
Interaction Effects										
Age:Gender	–	***	–	**	–	***	–	***	–	***
Age:Deprivation	–	***	–	***	–	***	–	***	x	x
R-Squared										
Marginal		0.864		0.813		0.303		0.227		0.993
Conditional		0.941		0.915		0.959		0.945		-
Dispersion statistics		1.1572		1.262		1.088		1.025		1.100

Table 3.6: Final model multivariable regression analysis of respiratory and non-respiratory presentations to three healthcare services. Full results and overall significance presented for main effects, and overall significance for interaction terms (ns = not significant, * = ≤0.05, ** = ≤0.01, *** = ≤0.001, x = variable not modelled, - = reference level).

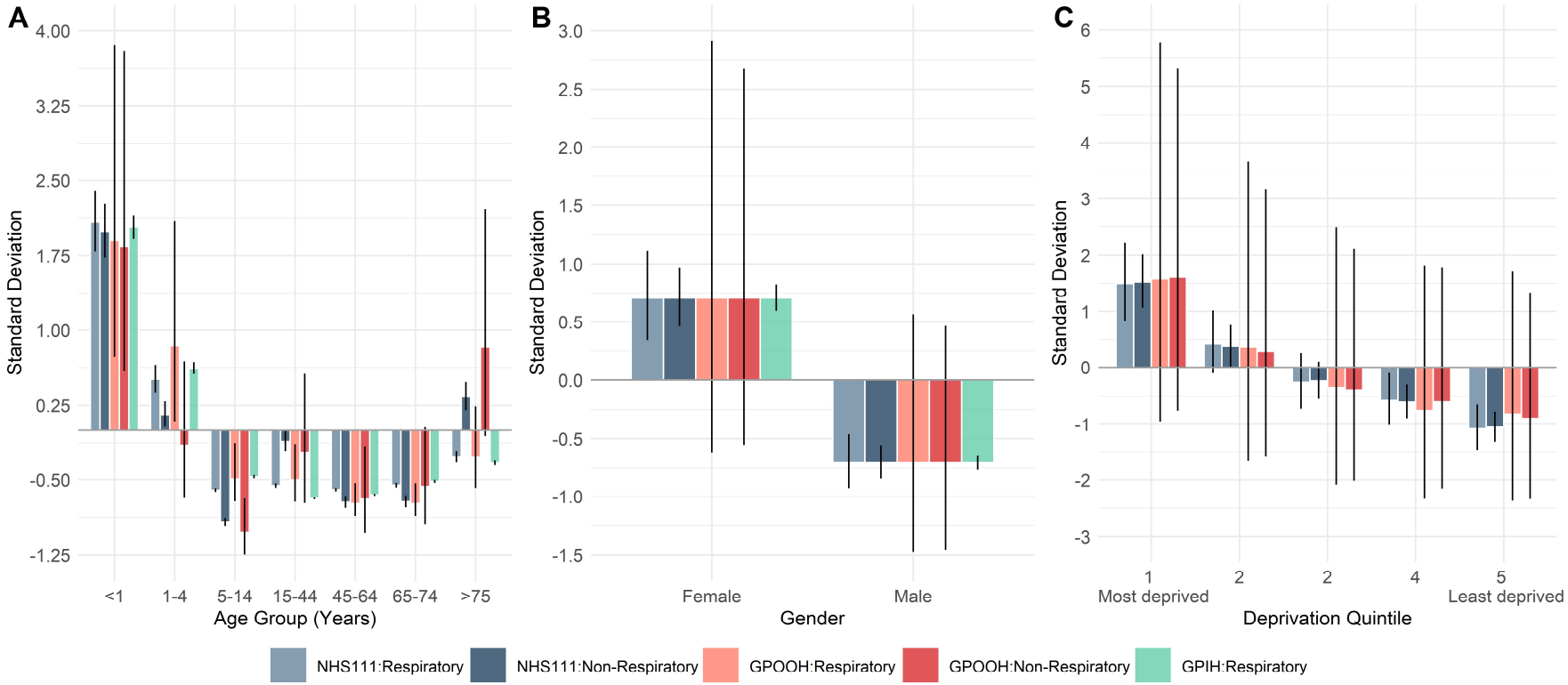

Figure 3.4: Standard deviations from the mean of main effects predictions from the multivariable analysis of respiratory presentations to NHS111, general practitioner-out-of-hours (GPOOH) and general practitioner-in-hours (GPIH): Each plot describes the standard deviations of A) age group, B) gender and C) deprivation. All models had an overall significance ≤0.001.

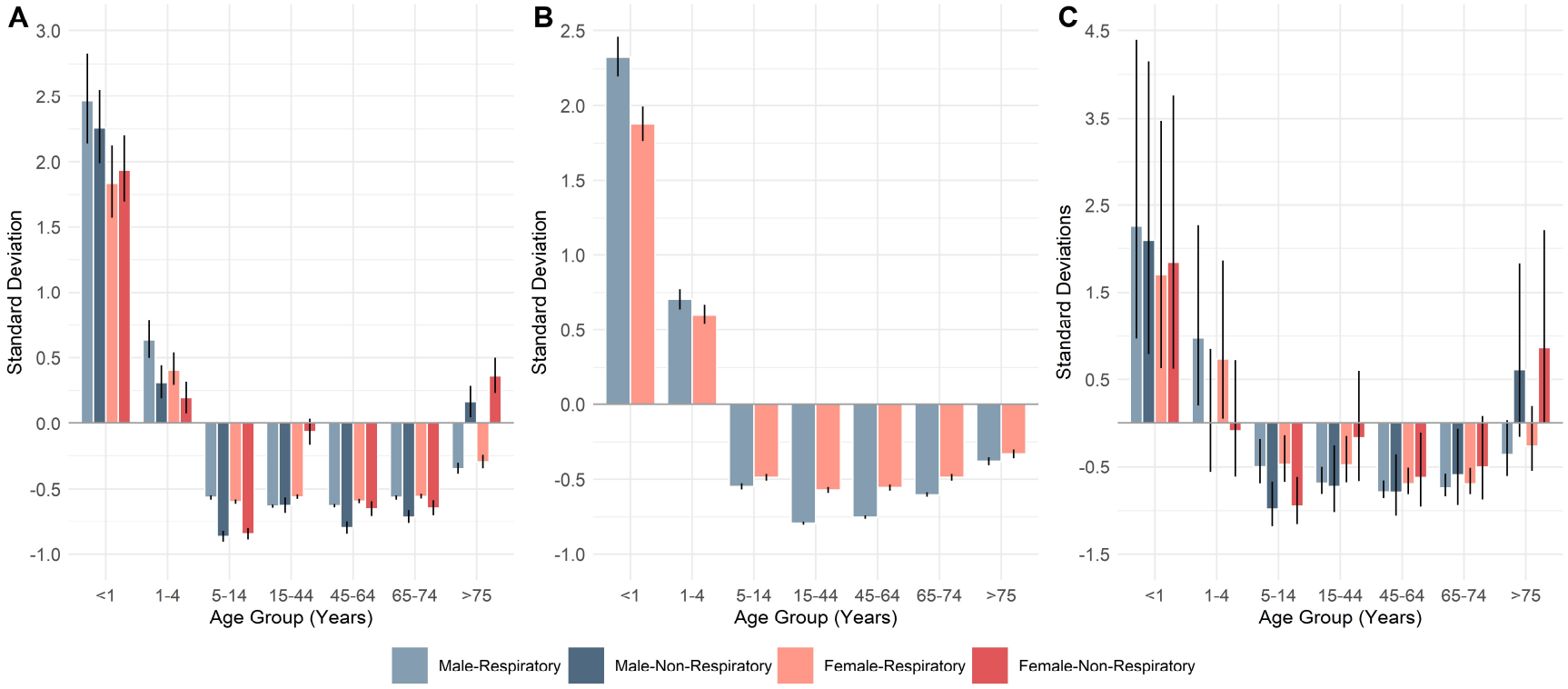

Figure 3.5: Standard deviations from the mean in rates of presentations across subgroups defined by age and gender for respiratory and non-respiratory presentations by each system: A) NHS111, B) general practitioner-in-hours (GPIH) and C) general practitioner-out-of-hours (GPOOH). All models had an overall significance ≤0.001.

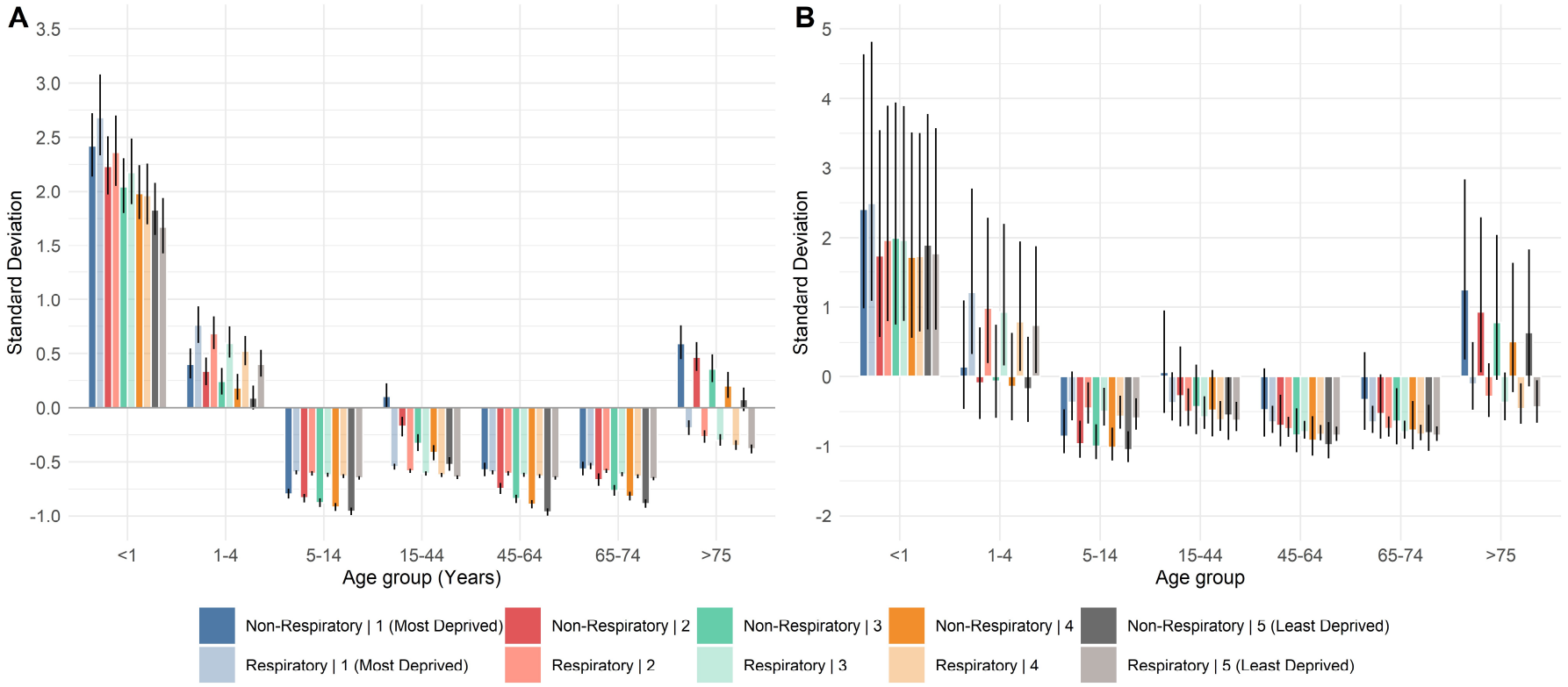

Figure 3.6: Standard deviation from the mean in rates of presentations across subgroups defined by age and deprivation quintile for respiratory and non-respiratory presentations by each system: A) NHS111 and B) general practitioner-out-of-hours (GPOOH). NHS111 respiratory, GPOOH respiratory and GPOOH non-respiratory models had an overall significance ≤0.001. NHS111 non-respiratory model had an overall significance of ≤0.01.

3.7 Discussion

3.7.1 Impact

Here have we presented a large-scale analysis of over 13 million respiratory-related presentations to three healthcare services in England. These services deliver healthcare to patients at the community level and may be better indicators of overall disease patterns as opposed to the relatively small proportion that appear in hospitalisation or laboratory-based surveillance. Access to these large syndromic datasets has allowed a comprehensive analysis of the demographics using of each service and the impact deprivation has on healthcare presentations. It has also allowed the interactions between these factors to be explored. By analysing respiratory and non-respiratory presentations, we were able to identify differences and similarities in usage patterns. Analysing NHS111 and GPOOH presentations at PD, a small level of geography, we have been able to explore deprivation patterns to give a thorough analysis of social patterning of factors associated with respiratory presentations.

3.7.2 Main Findings and Comparison to the Literature

In the three services included in this study, there were higher rates of respiratory and non-respiratory presentations in females. Previous research indicates that women are more likely to seek healthcare than men, even when female-specific concerns are accounted for (Manierre 2015; Thompson et al. 2016). Although we observed higher respiratory presentation rates for in females, males have been observed to have higher death and burden rates due to respiratory disease (Jordan et al. 2006), highlighting possible gender differences in healthcare-seeking behaviour. In terms of age, presentations were highest in those under-1 year, followed by the 1-4 and the over-75 age groups. These trends were observed across all three services, and these observations agree with previous research (Liptzin, Landau and Taussig 2015; Salisbury, Trivella and Bruster 2000). Although the three services displayed similarities in terms of the age groups more likely to use the service, there were differences in the magnitude of the rate ratios. In comparison to the reference group (15-44 years) children under-1 year

were 37.3, 18.7 and 6.2 times more likely to present to NHS111, GPOOH and GPIH, respectively, due to respiratory disease. These differences may reflect the routes to access to the different services, and the severe and sudden nature of the respiratory illness in very young children requiring urgent health advice.

When interactions between age and gender were explored, we found that more presentations were regarding males in the two youngest age groups for respiratory diseases; this effect was observed for all services. Previous studies have reported that male children have higher rates of respiratory illnesses (Liptzin, Landau and Taussig 2015; Vink et al. 2010; de Lusignan et al. 2018). de Lusignan et al. (2018), observed a higher incidence of family doctor presentations in male's under-15 years due to LRTI and asthma compared to females of the same age, with the largest gender difference in the youngest age group (under-1 year). The reasons for which are unclear, but immunologic, genetic and biological differences are thought to increase risk, suggesting that the excess in male children is due to genuine predisposition to respiratory illness rather than sociological factors (Liptzin, Landau and Taussig 2015). This excess in male children was also observed in our data for non-respiratory presentations. The reasons for this excess of non-respiratory presentations in male children is unclear. These findings suggest that there are similar drivers to presentations in male children presenting for all illness types rather than with just respiratory diseases, whether this be due to differences in healthcare-seeking behaviours or disease aetiology. Earp et al. (2019) observed that adults perceived male children (aged 5-years) experienced more pain compared to females of the same age, despite the same clinical circumstances and identical pain behaviour. The Infectious Intestinal Disease (IID) Study observed a higher rate of IID in the community in females under-1-year compared to males, but a higher rate of presentations to family doctors of males compared to females of the same age (FSA 2000). This observation suggests that either males were more likely to have cases of IID, which required medical intervention, or healthcare was more likely to be sought for males compared to females of the same age. There could be gender differences in the perception of ill health in young children from parental caregivers,

which could influence healthcare-seeking behaviour. The factors which influence healthcare-seeking behaviour in young children requires more research.

Deprivation was significant in both the NHS111 and GPOOH models, with a greater risk of respiratory and non-respiratory presentations in those from more deprived areas. Similar results were found when looking at gastrointestinal illness presentations to NHS111 and deprivation, with more presentations in the most deprived areas (Adams et al. 2018). Kelly et al. (2018) observed higher attendances from those from the most deprived areas to GPOOH services. The findings observed in this study are similar to those described in the literature from more acute healthcare settings, with previous studies linking increased deprivation with increased rates of hospitalisations and death due to respiratory infection (Jordan et al. 2006; Hawker et al. 2003; Geyer, Peter and Siegrist 2002).

One of the relatively unique elements of our work was the exploration of the interactions between deprivation and age. For NHS111, when stratified by age, deprivation appears to have a greater impact on non-respiratory presentations than on respiratory presentations. The exception is in the youngest age bracket, where deprivation has a similarly high impact on both respiratory and non-respiratory presentations, with more presentations in most-deprived areas. A similar pattern emerged for GPOOH where there were more presentations in the most derived areas; however, this trend was similar across age groups and in respiratory and non-respiratory presentations.

It is important to consider how these results can be generalised outside of England. It is difficult to make direct comparisons to other countries due to differences in healthcare systems, public health surveillance infrastructure, population demographics and deprivation levels. With the rise in the use of syndromic forms of surveillance (such as those used in this study), we feel it is important for other counties who use these forms of surveillance to undertake similar research.

3.7.3 Limitations

Although the surveillance of these three services allows us to observe a large number of healthcare presentations, these are working surveillance datasets from real-time surveillance system, with periods where data is not received. The large numbers included in the study meant that socioeconomic and demographic patterns could still be observed, and omissions are only likely to bias estimates if correlated with independent variables. Coverage issues will increase the uncertainty of our estimates, which is particularly evident in the wide confidence intervals observed in the GPOOH results.

The analysis presented here is exploring indicators of disease as opposed to actual disease. Therefore, the analyses are reliant on accurate classifications of disease indicators. NHS111, GPIH and GPOOH services have different symptom coding systems, and although we tried to choose indicators which would allow comparisons, the definition of respiratory and non-respiratory may differ between services. The deprivation measures used were composites of multiple factors that contribute to deprivation and were ecological in nature, are subject to the ecological fallacy whereby the associations found at the area level may not hold at the individual level. Data obtained from the surveillance systems did not contain information on co-morbidities or ethnicity, therefore these factors could not be accounted for in the analysis. Data from each surveillance system was available in an aggregated format, and hence it was not possible to identify multiple presentations by the same patient to the services during the study period.

3.7.4 Conclusion

This large-scale study highlights the effects of age, gender and deprivation on healthcare-seeking behaviours for respiratory diseases in the community, with more healthcare presentations from females, the young and old and those from more deprived areas. Similar patterns were observed across the three services, which were in agreement with the literature. Comparable results were also observed between respiratory and non-respiratory presentations suggesting that even when a wider

spectrum of disease is explored, demographic and socioeconomic factors may be the strongest influencers of healthcare-seeking behaviours. When broken down by age there were more presentations regarding male children under-5 years compared to females. This trend was observed in both respiratory and non-respiratory presentations indicating there are more healthcare-seeking behaviours in male children across a range of disease types. These findings could be influenced by sociological factors as well as disease aetiology and requires further research. Further research is required to understand the role that age, gender and deprivation has on healthcare-seeking behaviours.

Chapter 4

Estimating the Community Burden of Respiratory Syncytial Virus in Children Under-Five Years, England

4.1 What is Known About the Subject?

- Most burden estimates of respiratory syncytial virus (RSV) in young children have focused on hospitalisations, while the community burden remains poorly defined.
- High RSV burden has previously been described in children, with the highest burden observed in those under-1 year.

4.2 What this Study Adds?

- We estimate baseline data on the burden of RSV in the community and associated direct costs to healthcare services. These are essential for assessing the impacts of a future RSV vaccine.
- Highlights the high burden of RSV in the community, especially in children under-1 year.
- The highest community healthcare service burden was observed in in children under-1 year referred to unscheduled care, highlighting the acute nature of RSV symptoms in this age group.
- RSV had a higher burden on community healthcare services compared to influenza; this was particularly evident in children under-1 year.

4.3 Abstract

Respiratory syncytial virus (RSV) is a highly contagious viral infection which causes acute respiratory tract infections. RSV is a major cause of disease in children worldwide, infecting over 80% of children before two years of age. Most burden estimates of RSV in young children focus on hospitalisations; with the community burden remains poorly defined. We estimate the attributable RSV burden in the community for children under-5 years in England, using syndromic data from three community healthcare services. With hope for an RSV vaccine in the future, baseline RSV burden estimates are essential for assessing vaccine impact.

The attributable burden of RSV was estimated using syndromic data obtained from presentations to three healthcare services, telehealth (NHS111), family doctors (GPIH) and unscheduled care (GPOOH) services between 11th of November 2013 and 18th of June 2018. Time-series Poisson and negative binomial regression were used to model each syndromic indicator of interest against laboratory confirmed cases of RSV, while controlling for other respiratory pathogens, unknown seasonal drivers, long-term trends and public holidays. Direct healthcare costs were also estimated.

We estimate almost 200,000 RSV attributable presentations to NHS111, GPIH and GPOOH per year in children under-5 years over the study period. The average annual presentation rates attributable to RSV in children under-5 years was 1,240, and 2,051 to NHS111 and GPOOH per 100,000 population and 2,544 per 100,000 registered patients to GPIH. Higher presentation rates were observed in children under-1 year compared to those of 1-4 years across all services. Compared to influenza, more presentations were attributed to RSV among children under-5 years. For every case of RSV reported in laboratory surveillance we estimate between 26 and 69 healthcare presentations in under 5s in the community. The direct costs attributable to RSV related presentations to the three healthcare services were calculated to be almost £10 million per year.

This study highlights the substantial burden of RSV in children under-5 years in the community, and the high direct costs to healthcare services. The burden of disease

exceeds that of influenza. This study adds to evidence of the burden of RSV, which is important in the context of a future vaccine.

4.4 Introduction

Respiratory syncytial virus (RSV) is a highly contagious viral infection, which can cause acute respiratory tract infections (Parrott et al. 1973; Mufson et al. 1973). Clinical manifestations can vary in severity according to age, health status, and environmental exposure. Common and early signs of infection manifest in the upper respiratory tract with cold-like symptoms such as fever and cough. In more severe cases, infection can lead to lower respiratory tract involvement with pneumonia and bronchiolitis. RSV is the most common cause of lower respiratory tract infections (LRTI) in children under-2 years (Anderson et al. 1985).

RSV is a major cause of disease in children worldwide, infecting over 80% of children by two years of age (Glezen et al. 1986). In children RSV is estimated to cause 50-90% of hospitalisations due to bronchiolitis and 5-40% of hospitalisations due to pneumonia (Hall and McCarthy 2000). Worldwide, children under-5 years are most at risk of developing severe infection, and there were ~34 million episodes of acute lower respiratory infection (ALRI) occurring in children under-5 years in 2015, ~4 million of which required hospitalisation (Shi et al. 2017). It is estimated up to 40% of infants under-1 year with RSV develop lower respiratory tract involvement and 0.5-2% of under-1's required hospital admission (Cherry et al. 1998).

RSV in children places a significant burden on healthcare. In the USA, RSV-related bronchiolitis resulted in 77,700 hospital admissions annually between 1997 and 2002 and was the leading cause of hospitalisations in children under-1 year (Leader and Kohlhase 2003). In Canada, RSV was also found to be a leading cause of hospitalisations in children, exceeding those related to influenza (Schanzer, Langley and Tam 2006). In the United Kingdom (UK) over 450 thousand family doctor (general practitioner; GP) consultations, over 29 thousand hospitalisations, and 83 deaths in children and adolescents were estimated to be caused by RSV annually between 1995

and 2009 (Taylor et al. 2016). RSV was also found to contribute to more GP visits, hospitalisations and deaths than influenza in children under-5 years in the UK (Taylor et al. 2016).

For any disease, estimating its total burden is essential to understanding the public health impact. This estimate allows policy-makers and health professionals to make informed decisions for targeted research, to identify and reduce health inequalities, control disease spread, plan and implement effective interventions and, ultimately, improve health outcomes (Lajoie 2013). With a RSV vaccine imminent (Higgins, Trujillo and Keech 2016), retrospective baseline data on total RSV burden is essential for assessing the impact of the vaccine.

Past studies estimating the health and clinical burden of RSV in the UK have primarily focused on laboratory (Ajayi-Obe et al. 2008; Hardelid, Pebody and Andrews 2013) and diagnostic surveillance data (Deshpande and Northern 2003; Fleming, Pannell and Cross 2005a; Fleming et al. 2015; J. Murray 2013; Reeves et al. 2017) taken during a period of hospitalisation. By focusing on diagnostic or laboratory diagnosis given during hospitalisation of RSV associated disease, there is an underestimate of the true burden of disease, as it will not account for less severe cases in the community who have sought healthcare but were not subject to laboratory analyses.

An alternative method to estimate burden of RSV in the community is through an analysis of non-laboratory-based surveillance data such as presentations to telehealth services and family doctors. Where data based on clinical symptoms of the patients are noted during presentations to healthcare services. By utilising alternative surveillance data from healthcare contact that is made before a clinical or laboratory diagnosis is given, it may be possible to capture a greater estimate of disease burden. Syndromic surveillance data allows for more of the burden of disease to be captured compared to traditional data sources such as laboratory or hospitalisation surveillance. Syndromic data is defined as non-specific, pre-diagnostic-syndromic health data which can be used as an indicator for disease (Triple S Project 2011). A limitation of syndromic surveillance is that by focussing upon symptoms, there is an uncertainty as to whether individuals have RSV or another respiratory illness with similar symptoms.

Here, we estimate the relationship between syndromic data potentially indicative of RSV (e.g., calls to a telehealth system for *Difficulty Breathing*), and laboratory data of confirmed RSV cases, to estimate the RSV-attributable burden in three community healthcare services. The attributable burden of RSV is estimated for children under-5 years who sought help from three healthcare services: telehealth, family doctors and unscheduled care. We estimate the attributable RSV burden in children under-5 years in England. Children under-5 years are the focus of this research due to the substantial burden of RSV in this age group (Shi et al. 2017). To help put RSV into context, comparisons are made to influenza, another common respiratory infection with a high healthcare burden. This research can provide better understanding of the community RSV burden and allow evaluation of the potential benefits of a vaccine post-introduction. These estimates can also allow for the further generation of hypotheses regarding factors that drive RSV transmission.

The aims of this study are to:

1. Estimate the community burden of RSV in England, using syndromic indicators of RSV activity.
2. Investigate the community RSV burden by age group; under-1 year and 1-4 years.
3. Compare the community RSV burden to that of influenza.
4. Estimate the direct healthcare cost of the community burden of RSV.

4.5 Methods

4.5.1 Data Collection

Respiratory syndromic data (indicators), potentially linked to RSV, were obtained from three healthcare services monitored by Public Health England's (PHE) national real-time syndromic surveillance team (ReSST); NHS111 (telehealth service), GPIH (general practitioner consultations in hours), and GPOOH (general practitioner consultations out-of-hours) (Table 4.1). Data were obtained at weekly intervals for the

period 11th of November 2013 and 18th of June 2018. Count data was stratified by age group (<1 year and 1-4 years), week, and location. Demographic and location data were obtained from information provided when the patient contacted the healthcare services. Location was provided at upper tier local authority (UTLA) for NHS111 and GPOOH data, and at PHE centre for GPIH data. The demographic detail of these spatial levels varies considerably; with 149 UTLA each with an average of 23,050 (5,530-91,830) residents under-5 years. PHE Centres (n=9) are much larger in terms of geography and population with an average of 378,960 (148,180-628,390) residents under-5 years. For the analysis all data was aggregated to the national level (England) as the focus was on the overall burden in England. Syndromic indicators were chosen based on expert knowledge and previous research (Morbey et al. 2017a; 2017b; 2018). Some syndromic indicators from the GPIH and GPOOH surveillance systems were subsets of a broader indicator (for example, *All Respiratory Disease* includes *Acute Respiratory Infection*).

Weekly counts of laboratory confirmed cases of RSV and a further nine seasonal respiratory pathogens (Table 4.2) were obtained from the PHE Second Generation Surveillance System (SGSS) for the same period as the syndromic data. Data on these nine other pathogens was collected because the syndromic indicators used in this study may be associated with RSV as well as other seasonal respiratory pathogens. Data was stratified by age group and week. Pathogen data was based on specimen collection date. Time series of laboratory confirmed cases of RSV and influenza per 100,000 resident's under-5 years in England are described in Figure 4.1.

Mid-year estimates of English population, stratified by age group, for each year included in the analysis were obtained at the UTLA level for analysis with the NHS111 and GPOOH data (ONS 2014). These populations were aggregated to the national level for the analysis. Populations for GPIH data were obtained from the average weekly number of patients registered at the GP practices that provide data for syndromic surveillance, in England, this figure was stratified by age group and year.

Service/Syndromic Indicator	Number of Episodes		Weekly Mean		Maximum in a Single Week	
Age	Under 1 Year	1-4 Years	Under 1 Year	1-4 Years	Under 1 Year	1-4 Years
NHS111						
Difficulty Breathing	221,396	351,801	907.4	1,441.8	2,406	3,589
Fever	163,578	344,427	670.4	1,411.6	921	2,651
Cough	278,891	534,077	1,143.0	2,188.8	3,644	6,353
Sore Throat	196	176,399	0.8	722.9	8	1,667
Cold/Flu	126	90,314	0.5	370.1	9	1,057
GPOOH						
All Respiratory Disease	322,939	870,191	1,323.5	35,66.4	4,389	8,703
Difficulty Breathing/Wheeze/Asthma	291,50	69,304	119.5	284.0	449	769
Asthma	337	12,321	1.4	50.5	8	162
Acute Respiratory Infection	253,339	705,842	1,038.3	2,892.8	3,509	7,189
Influenza Like Illness	1,907	3,971	7.8	16.3	33	78
Bronchitis/Bronchiolitis	19,937	8,674	81.7	35.5	434	154
Pharyngitis/Scarlet Fever	483	2,649	2.0	10.9	8	34
GPIH						
Upper Respiratory Tract Infection	825,666	2,477,967	3,383.9	10,155.6	7,369	23,920
Influenza Like Illness	2,724	14,139	11.2	58.0	36	237
Pharyngitis or Scarlet Fever	65,793	311,115	270.0	1,275.1	542	2,845
Scarlet Fever	300	5,042	1.2	20.7	10	66
Lower Respiratory Tract Infection	189,201	414,071	775.4	1,697.0	2,707	5,066
Pneumonia	712	4,016	2.9	16.5	13	52
Severe Asthma	136	18,595	0.6	76.2	6	190
Bronchitis/Bronchiolitis	8,052	6,215	33	25.5	75	61

Table 4.1: Number of syndromic indicator presentations received by ReSST between 11th November 2013 and 18th June 2018.

Pathogen	Total		Weekly Mean		Maximum in a Single Week	
Age	Under 1 Year	1-4 Years	Under 1 Year	1-4 Years	Under 1 Year	1-4 Years
Respiratory Syncytial Virus	13,966	3,485	57.2	14.3	474	123
Rhinovirus	6,677	3,979	27.4	16.3	63	40
Parainfluenza	1,896	1,511	7.8	6.2	34	23
Influenza A	844	1,638	3.5	6.7	21	69
Human Metapneumovirus	1,118	830	4.6	3.4	30	22
Influenza B	314	651	1.3	2.7	19	35
Coronavirus	416	337	1.9	1.4	15	13
Streptococcus Pneumonia	262	367	1.1	1.5	5	9
Haemophilus Influenzae	101	71	0.4	0.3	4	3
Mycoplasma Pneumonia	42	98	0.2	0.4	4	6

Table 4.2: Laboratory confirmed pathogens samples between 11th November 2013 and 18th June 2018.

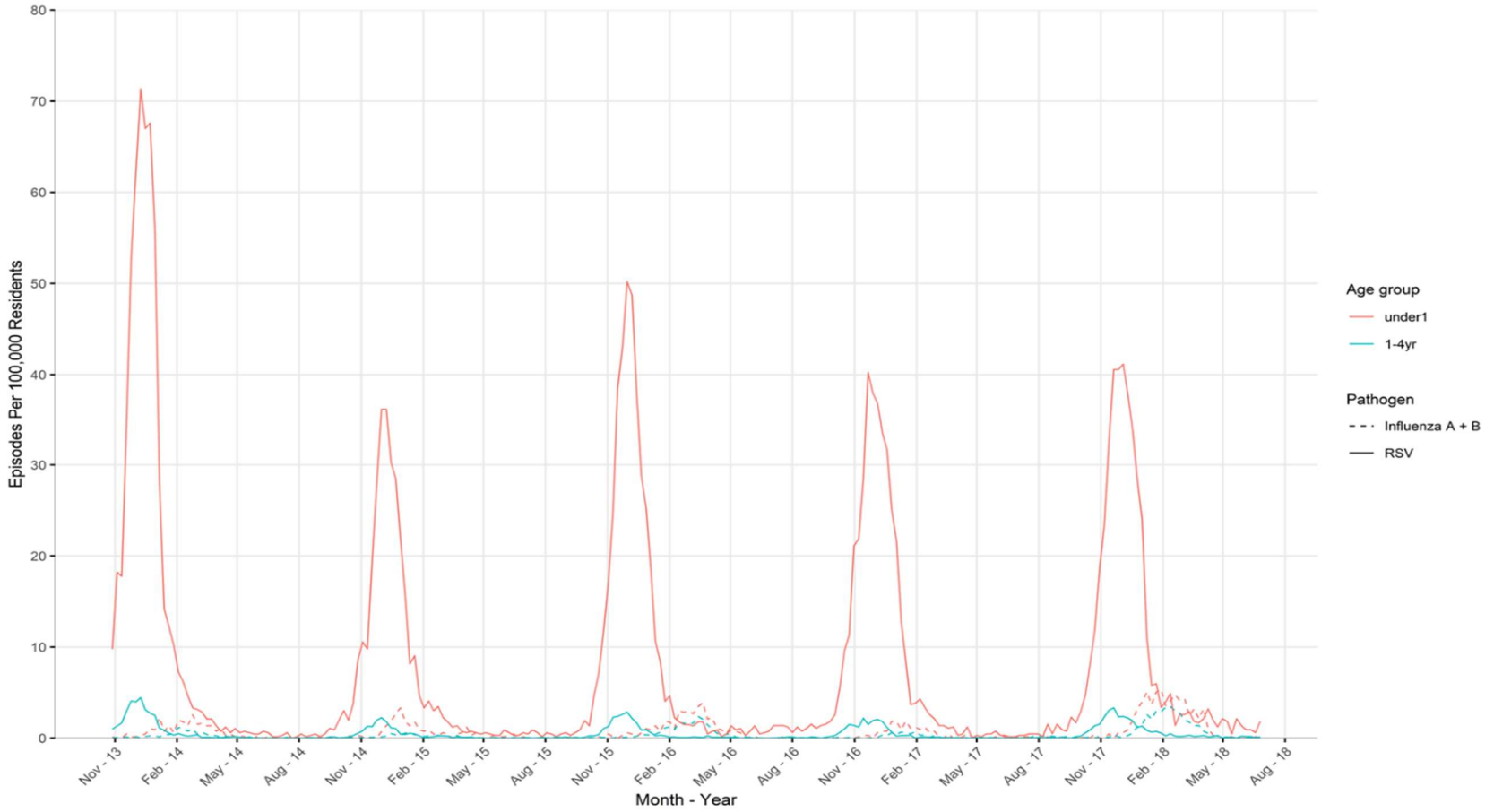

Figure 4.1: Time series of laboratory confirmed cases of RSV (solid lines) and Influenza A+B (dashed lines) per 100,000 residents in England for the period 11th November 2013 and 18th June 2018.

4.5.2 Estimating Areas with Poor Surveillance Coverage

There were multiple periods of incomplete coverage in the NHS111 and GPOOH surveillance systems; this meant that the study population was unknown. Our experience of running real time surveillance systems over years tells us that these periods are due to issues with data transfer to the surveillance systems, as opposed to issues with the provision of healthcare services. In order to estimate the study population and therefore the RSV burden, it is important to identify these periods when data may be incomplete. The periods of poor coverage were detected using the pruned exact linear time method (PELT) of change-point analysis (Killick, Fearnhead and Eckley 2012). PELT was used to estimate changes in mean and variance of data from NHS111 and GPOOH in each UTLA. Change point detection estimates the point at which the statistical properties, in this case mean and variance, of a time series change. PELT allows the detection of multiple change points over the period of observation (Killick, Fearnhead and Eckley 2012). Change point analysis was conducted using the R package *changepoint* (Killick and Eckley 2014). Any change below the mean number of presentations over the time-period was identified as a period of poor coverage and was excluded from the analysis (example Figure 4.2). In total, 38 of 149 UTLAs had periods of poor coverage for NHS111; this constituted 10.5% (3,827 of 36,499) of the total number of weeks included the analysis. In total, 54 of 149 UTLAs had periods of poor coverage for GPOOH; this constituted 53.1% (19,379 of 36,505) of the total number of weeks included the analysis. Once the periods of poor coverage were identified and excluded at the UTLA level, the data was aggregated at the national level. Population estimates used in the statistical models of NHS111 and GPOOH data, were adjusted to remove the periods where poor coverage was identified. Detection of poor coverage was not required for the GPIH data because the number of registered patients was available; therefore, the population of study was known. The total number of residents registered per day were available for the GPIH data, a weekly average, which accounted for public holidays, was taken to estimate the mean number of residents per week. Population estimates for NHS111, GPOOH and GPIH services changed weekly to reflect the number of institutions reporting to the surveillance service. The outcome

measures of interest were presentations to NHS111, GPIH and GPOOH coded as the syndromic indicators listed in Table 4.1, which are associated with RSV in children under-5 years for the time period 11th of November 2013 and 18th of June 2018.

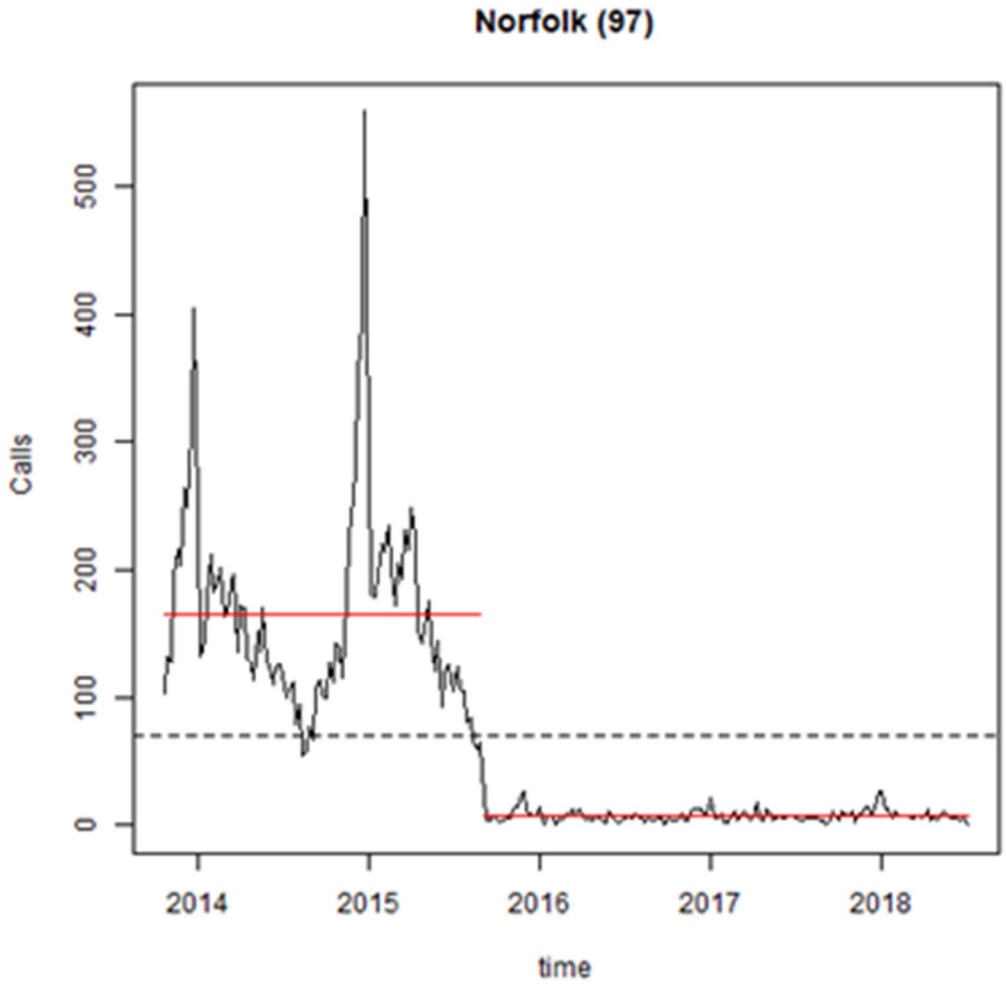

Figure 4.2: Results from pruned exact linear time method (PELT) change point analysis for NHS111 syndromic data of Norfolk for the period 11th November 2013 and 18th June 2018. Black solid lines indicate the number of calls. Red lines indicated periods of similar mean and variance. Dotted line indicates mean.

4.5.3 Statistical Analysis

Poisson and negative binomial generalised linear models (GLM) were used to model each syndromic indicator of interest separately (Equation 1). The models associate the temporal trend of laboratory confirmed cases of RSV to the syndromic indicators of interest, while controlling for other laboratory confirmed seasonal respiratory pathogens, unknown seasonal drivers of the outcome of interest, long-term time trends and public holiday effects. Laboratory data of nine seasonal respiratory pathogens, other than RSV, (Table 4.2) were included in the model as independent variables to account for their potential confounding effects with the outcome of interest, where circulation of these other seasonal respiratory pathogens could result in respiratory presentations to the healthcare services monitored by syndromic surveillance. All models included population as an offset. A directed acyclic graph (Figure 4.3) was created using the R package *dagitty* (Textor et al. 2016) to identify possible causal mechanisms for the syndromic data, and to identify independent variables for inclusion in the study. Factors identified as having a relationship on the outcome of interest (the syndromic indicators) were included in the models, where possible, to reduce bias and to account for potential confounding effects. We also estimated the burden of influenza using the same methods described, to allow for comparisons with RSV. Public holidays were included in the model to account for different healthcare usage during holidays with a lag of one week included to account for delayed healthcare-seeking behaviours. To account for unknown seasonal factors three sine/cosine pairs were included in the model; the number of pairs included in the model were based on goodness of fit metrics, and temporal autocorrelation. A cubic polynomial of time (each week included in the study) was included in the models to account for long-term time trends. To account for residual temporal autocorrelation an auto-regressive term of the residuals was included in the model (Cameron and Trivedi 2013; Imai et al. 2015).

Seasonality needs to be controlled for as both the dependant (syndromic indicators) and independent variables of interest (RSV and influenza) have a seasonal component, that cannot be solely explained by the variables included in the models (Christiansen et al.,

2012). For example, other seasonal factors such as temperature, humidity or changes in behaviour might drive trends in the data (D'Amato et al., 2018; Peppa et al., 2017) and cannot be accounted for by the observed independent variables (other repertory pathogens and public holidays). If seasonality is not adequately controlled for, the relationship between the relationship between the syndromic indicators and RSV and influenza could be overestimated, however over-accounting for seasonality might underestimate the relationship with the parameters of interest, especially when they have a highly cyclic pattern such as RSV. Here we included Fourier terms and a cubic polynomial of time to account for seasonality as described by Bhaskaran et al., (2013). Other commonly used methods to control for seasonality and long terms trends include time stratified models and flexible spline functions, however these methods tend to be a more stringent methods to account for seasonality, attribution a higher degree of the variability on the data to seasonality and long terms trends (Bhaskaran et al., 2013).

During model development the inclusion of up to three Fourier terms was included in the model to establish the impact of different parametrisations of seasonality on model fit and the variables of interest (Table 4.3). In general, AIC was lowest when three Fourier terms were used in the models. There were very modest differences in the deviance explained and dispersion statistics observed with different seasonality parameterisations. When looking at the results of a cross- validation analysis, when one Fourier term was used there tended to be a higher degree of overfitting (higher RME and MAE values) compared to when two or three Fourier terms were included. There were substantial differences in RSV and influenza predictions when different parameters of seasonality were included. Although the highest number of RSV cases were observed when only one Fourier term was included in the models, however these models were not flexible enough to detect influenza trends because there were not enough inflexions within the sine wave masking the trends of influenza (this mainly occurred in ARD, URTI, LRTI, cough and difficulty breathing indicators). When two Fourier terms were included in the model, RSV estimates tended to be much lower than when one or three terms were included in the models, suggesting an underestimation of the relationship. Although the number of RSV cases, tended to be lower when three

Fourier terms was included compared to one, this parametrisation also captured influenza trends. In summary, there was clear differences in terms of model metrics, cross validation, and predictions when using different seasonality parameters. However, the inclusion of three Fourier terms produced more reliable predictions for both RSV and influenza compared to other parametrisations of seasonality as both RSV and influenzas trends could be detected, as well as lower AIC values and lower levels of overfitting compared to the other seasonality parameters, especially for the broad indicators that will capture a higher degree of burden.

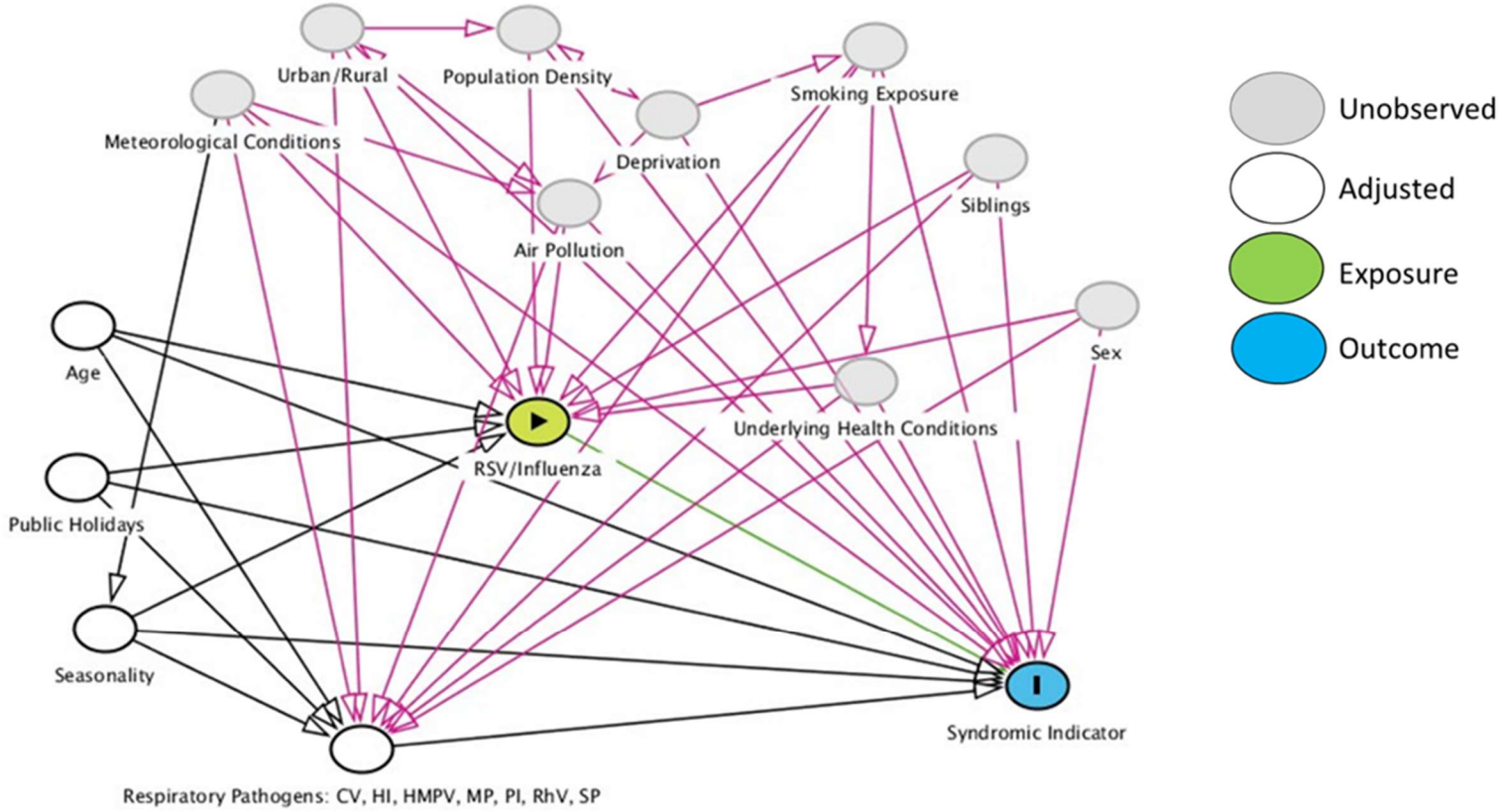

Figure 4.3: Directed acyclic graph to estimate the causal relationship between laboratory cases of RSV and influenza (green), and syndromic indicators (blue). Grey circles represent variables that could not be accounted for, and white circles represent the variables that

were not adjusted for in the models. Abbreviations: Respiratory Syncytial Virus (RSV), Coronavirus (CV), Rhinovirus (RhV), Parainfluenza (PI), Human Metapneumovirus (HMPV), Streptococcus Pneumonia (SP), Haemophilus Influenzae (HI), and Mycoplasma Pneumonia (MP). As analysis was conducted on national data, we could not adjust for variables that were spatially defined (meteorological conditions, air pollution, deprivation, population density, rurality). Due to data aggregation, we could not account for individual characteristics (siblings, underlying health conditions, smoking exposure). Although these variables could not be accounted for in the model, these unobserved variables are likely to have been static over time therefore their omission is unlikely to bias our results. Data was available by sex, however data was national and there was no evidence the distribution of sex changed over time, therefore we did not account for it in the analysis (Bhaskaran et al., 2013).

System	Syndrome	Age	Fourier terms	AIC	Deviance Explained	Dispersion	MAE	RMSE	No. RSV Predictions	No. Flu Predictions	No. Presentation Predictions	% of Total Presentation's attributable to RSV	% of Total Presentation's attributable Influenza
NHS111	Difficulty Breathing	Under 1	1	2834	0.96	1.10	80.58	110.30	48584	-2927	255732	19.00	-1.14
			2	2815	0.96	1.10	73.64	101.02	42034	3988	253614	16.57	1.57
			3	2813	0.96	1.10	77.20	105.00	42288	2845	253338	16.69	1.12
		Over 1	1	3244	0.88	1.11	179.84	230.81	51863	-15681	400002	12.97	-3.92
			2	3209	0.90	1.12	163.74	210.07	28286	-1704	395006	7.16	-0.43
			3	3200	0.91	1.13	171.32	214.93	34303	1181	394983	8.68	0.30
	Fever	Under 1	1	2643	0.66	1.10	53.20	65.95	-606	3793	181974	-0.33	2.08
			2	2643	0.66	1.11	54.24	67.23	-3180	3314	181811	-1.75	1.82
			3	2645	0.67	1.12	55.89	69.00	-3725	2918	181709	-2.05	1.61
		Over 1	1	3073	0.85	1.09	137.00	177.45	12690	13629	384189	3.30	3.55
			2	3068	0.85	1.09	143.54	181.71	2447	18779	383292	0.64	4.90
			3	3070	0.85	1.09	153.65	193.76	8505	19486	383510	2.22	5.08
	Cough	Under 1	1	2992	0.96	1.08	132.56	185.32	74965	-1064	328263	22.84	-0.32
			2	2956	0.97	1.09	118.58	161.75	54100	9161	322369	16.78	2.84
			3	2958	0.97	1.10	122.48	164.20	54429	7089	322060	16.90	2.20
		Over 1	1	3471	0.92	1.08	379.25	521.55	109912	-29689	622349	17.66	-4.77
			2	3427	0.94	1.09	359.79	483.85	58018	1069	609369	9.52	0.18
			3	3424	0.94	1.10	362.04	476.40	65634	4729	609590	10.77	0.78
	Sore Throat		1	2826	0.83	1.07	83.35	107.42	8105	4991	199001	4.07	2.51

Table 4.3: Impact of using three seasonality parameterisations on model fit, cross validation, and model predictions. One, two or three Fourier terms were introduced into the model, while all other model parameters were kept constant.

		Over 1	2	2821	0.83	1.08	84.63	109.33	3470	7771	198386	1.75	3.92
			3	2811	0.84	1.07	86.59	109.31	8011	8343	198547	4.03	4.20
	Cold/Flu	Over 1	1	2652	0.92	1.10	61.16	85.33	8123	6078	103301	7.86	5.88
			2	2626	0.93	1.11	60.18	82.72	1028	10972	101757	1.01	10.78
			3	2624	0.93	1.11	61.58	82.18	2809	11389	101775	2.76	11.19
GPOOH	ARD	Under 1	1	3071	0.93	1.07	210.76	312.52	300798	-1637	1045563	28.77	-0.16
			2	2998	0.95	1.09	170.47	250.93	183680	45003	911724	20.15	4.94
			3	2999	0.95	1.10	185.68	275.12	172224	39863	886592	19.43	4.50
		Over 1	1	3639	0.88	1.07	444.83	572.42	377290	-38727	2343376	16.10	-1.65
			2	3566	0.91	1.09	353.69	450.08	192655	186035	2291762	8.41	8.12
			3	3559	0.92	1.10	357.68	451.06	183842	205278	2282076	8.06	9.00
	ILI	Under 1	1	1233	0.53	1.11	3.49	4.66	1565	782	7061	22.16	11.07
			2	1234	0.53	1.12	3.62	4.80	1634	938	6918	23.62	13.56
			3	1234	0.54	1.11	4.11	5.56	1782	960	7015	25.40	13.68
		Over 1	1	1477	0.74	1.13	5.55	7.72	1122	3653	13184	8.51	27.71
			2	1476	0.75	1.14	5.95	8.24	251	5119	13973	1.80	36.63
			3	1474	0.75	1.15	5.70	7.83	748	5773	14650	5.11	39.40
	DBWA	Under 1	1	2085	0.89	1.09	31.03	45.69	42707	-3143	91933	46.45	-3.42
			2	2036	0.92	1.11	24.46	34.49	28060	705	76798	36.54	0.92
			3	2036	0.92	1.11	27.20	39.21	26598	-39	74941	35.49	-0.05
		Over 1	1	2620	0.76	1.11	47.12	58.55	31978	-13421	168626	18.96	-7.96
			2	2558	0.82	1.19	38.71	46.74	17398	1359	157513	11.05	0.86
			3	2545	0.83	1.19	39.41	48.24	15064	2969	155550	9.68	1.91
	AR I		1	2950	0.93	1.07	157. 51	234.58	226438	6352	813427	27.84	0.78

Table 4.3: Impact of using three seasonality parameterisations on model fit, cross validation, and model predictions. One, two or three Fourier terms were introduced into the model, while all other model parameters were kept constant.

		Under 1	2	2882	0.95	1.10	128.78	189.13	137033	40677	712025	19.25	5.71
			3	2883	0.95	1.10	139.43	206.02	128213	36475	692502	18.51	5.27
		Over 1	1	3523	0.89	1.08	356.67	460.10	308218	-18385	1900177	16.22	-0.97
			2	3456	0.92	1.09	281.59	364.72	156487	167341	1868103	8.38	8.96
			3	3451	0.92	1.10	283.14	365.12	153573	181854	1862522	8.25	9.76
	Bronchitis	Under 1	1	1948	0.94	1.11	21.48	34.68	73704	-2913	102731	71.74	-2.84
			2	1851	0.96	1.18	19.92	31.37	33816	2519	64450	52.47	3.91
			3	1844	0.97	1.20	40.87	68.80	29008	1367	59737	48.56	2.29
		Over 1	1	1667	0.89	1.13	8.21	11.13	9882	-3070	25575	38.64	-12.00
			2	1592	0.92	1.17	5.96	7.61	6306	-309	22944	27.48	-1.35
			3	1588	0.92	1.19	6.04	7.64	5635	-171	22543	25.00	-0.76
GPIH	Acute Bronchitis	Under 1	1	1416	0.86	0.68	6.28	8.07	396	-420	7985	4.96	-5.26
			2	1414	0.89	0.53	4.07	5.17	-168	-259	7985	-2.10	-3.24
			3	1416	0.89	0.52	4.29	5.56	-383	-928	30358	-1.26	-3.06
		Over 1	1	1387	0.83	0.64	3.96	4.97	272	-336	6144	4.43	-5.47
			2	1369	0.85	0.56	3.59	4.30	-36	-152	6144	-0.59	-2.47
			3	1371	0.85	0.55	3.60	4.35	-160	-303	15234	-1.05	-1.99
	ILI	Under 1	1	1318	0.70	1.12	4.57	6.19	130	278	2699	4.82	10.30
			2	1309	0.71	1.15	4.44	5.78	130	356	2696	4.82	13.20
			3	1303	0.72	1.14	4.32	5.60	165	725	5735	2.88	12.64
		Over 1	1	1834	0.92	1.10	15.10	19.91	37	4365	14051	0.26	31.06
			2	1814	0.93	1.07	14.71	19.25	37	4508	14035	0.26	32.12
			3	1807	0.93	1.06	15.62	19.98	1031	8209	26194	3.94	31.34
	Pneumonia		1	934	0.39	1.04	4.41	5.72	105	41	706	14.87	5.80

Table 4.3: Impact of using three seasonality parameterisations on model fit, cross validation, and model predictions. One, two or three Fourier terms were introduced into the model, while all other model parameters were kept constant.

		Under 1	2	938	0.39	1.05	4.20	5.30	112	44	706	15.86	6.23
			3	941	0.39	1.06	4.07	5.06	224	85	1440	15.56	5.90
		Over 1	1	1420	0.74	1.11	3.65	4.75	808	145	3982	20.29	3.64
			2	1416	0.75	1.13	3.68	4.69	666	213	3981	16.73	5.35
			3	1418	0.75	1.16	3.77	4.80	1250	399	7489	16.69	5.33
	PSF	Under 1	1	2462	0.82	1.07	38.28	48.70	1444	-1116	65194	2.21	-1.71
			2	2420	0.85	1.09	33.32	41.06	940	792	65125	1.44	1.22
			3	2417	0.85	1.10	34.46	42.73	2278	1157	144263	1.58	0.80
		Over 1	1	3114	0.86	1.06	139.90	179.26	15545	4667	308234	5.04	1.51
			2	3094	0.87	1.06	135.55	170.80	5046	12382	308006	1.64	4.02
			3	3084	0.88	1.06	138.67	175.55	18353	27293	591438	3.10	4.61
	URTI	Under 1	1	3555	0.91	1.07	367.20	471.64	45948	-12126	817813	5.62	-1.48
			2	3510	0.93	1.08	313.77	396.10	24023	12427	816752	2.94	1.52
			3	3511	0.93	1.09	329.44	418.71	53716	20663	1811400	2.97	1.14
		Over 1	1	4143	0.90	1.05	1316.95	1734.42	235543	-26429	2453092	9.60	-1.08
			2	4110	0.92	1.06	1225.57	1561.91	93382	61036	2450100	3.81	2.49
			3	4109	0.92	1.06	1249.03	1604.44	240344	132155	4685317	5.13	2.82
	LRTI	Under 1	1	2848	0.96	1.09	107.64	145.54	43276	-8606	188197	22.99	-4.57
			2	2749	0.97	1.09	78.94	107.61	30413	1479	187255	16.24	0.79
			3	2755	0.97	1.09	83.52	113.03	65250	1603	415633	15.70	0.39
		Over 1	1	3290	0.93	1.06	243.40	330.68	72693	-17398	410051	17.73	-4.24
			2	3236	0.95	1.07	207.09	269.84	40980	3431	408754	10.03	0.84
			3	3236	0.95	1.08	217.15	287.12	83359	9093	781626	10.66	1.16
	Asthma		1	2021	0.76	1.10	15.22	18.68	1505	-497	18411	8.17	-2.70

Table 4.3: Impact of using three seasonality parameterisations on model fit, cross validation, and model predictions. One, two or three Fourier terms were introduced into the model, while all other model parameters were kept constant.

		Over 1	2	1994	0.79	1.11	13.62	16.83	329	155	18406	1.79	0.84
			3	1989	0.80	1.12	13.78	17.23	380	415	34073	1.12	1.22
	SF	Over 1	1	1378	0.84	0.92	4.15	5.07	25	178	4994	0.50	3.56
			2	1372	0.84	0.88	4.19	5.13	-56	272	4994	-1.12	5.45
			3	1372	0.85	0.87	4.36	5.37	-82	603	12286	-0.67	4.91

Table 4.3: Impact of using three seasonality parameterisations on model fit, cross validation, and model predictions. One, two or three Fourier terms were introduced into the model, while all other model parameters were kept constant.

Each syndromic indicator was modelled separately by age group, to allow for different syndromic data-pathogen leads by age group. Where syndromic indicators had a low number of counts (<500) over the study period, analysis was not undertaken for that age group, as a low number indicates that the syndromic indicator may not have been routinely used to describe disease for children in that age group. Data was first modelled using the Poisson distribution, if the model was over-dispersed, the negative binomial distribution was used. Over-dispersion was tested by comparing the sum of squared Pearson residuals to the residual degrees of freedom. Models were deemed as acceptable when the over-dispersion statistics <1.5 (Payne et al. 2018). A combination of 0-2 week leads for all pathogens was modelled, with the best combination of leads decided by the percentage of deviance explained. Leads of the laboratory data were included in the models to investigate if changes in trends in the syndromic data occurred before the laboratory data. We explored non-linearities in the relationship between the seasonal pathogens and syndromic indicators, by including the seasonal pathogens as quadratic polynomial functions, to account for possible threshold effects. However, this resulted in overfitting, consequently, the pathogens were included as linear terms. Model selection was based on Akaike Information Criterion (AIC), and percentage of deviance explained.

The algebraic definition of the model is given by:

$$Y_t \mid \mu_t, \sim F(\mu_t),$$

where μ_t is the is the age-specific number of presentations per syndromic indicator at time t, F is the likelihood of the model (Poisson or negative binomial). A logarithmic link function of the expected number of cases is modelled as:

$$\log(\mu_t) = \alpha + \log(P_t) + \rho R_{t-1} + \lambda t' + \sum_j \epsilon_j F_j + \tau H_t + \theta H_{t-1} + \sum_k \beta_k X_{t,k},$$

where α corresponds to the intercept; $\log(P_t)$ denotes the logarithm of the population P at risk for week t included as an offset to adjust counts by population; ρ is an auto-regressive coefficient of the residuals R lagged one week to account for potential temporal autocorrelation; t' represents a third degree polynomial of time with coefficient

λ; F is a matrix of j Fourier terms with coefficients ϵ to account for seasonal trends that may be related to factors other than RSV such as temperature and rainfall; H denotes a Boolean variable indicating the presence of public holidays on the current (t) or previous week ($t-1$) with coefficients τ or θ accordingly; X is a matrix of $k = 10$ infectious disease co-variates with regression coefficients β. Models were fitted in R version 3.5.2 using the *MASS* package (Venables and Ripley 2002).

Block time series cross validation was used to test model robustness and to identify overfitting in the models (Roberts et al. 2017). For the cross-validation analysis, each dataset was split into 10 folds, which are blocks of data of 24-25 continuous weeks, and data was trained and tested on future data in sequential linear time with nine iterations. With the cross-validation first iteration comprising of 10% training and 90% testing data, and the final iteration comprising of 90% training and 10% testing data. To assess overfitting mean absolute error (MAE) was used. A graphical representation of block time series cross validation is described in Figure 4.4.

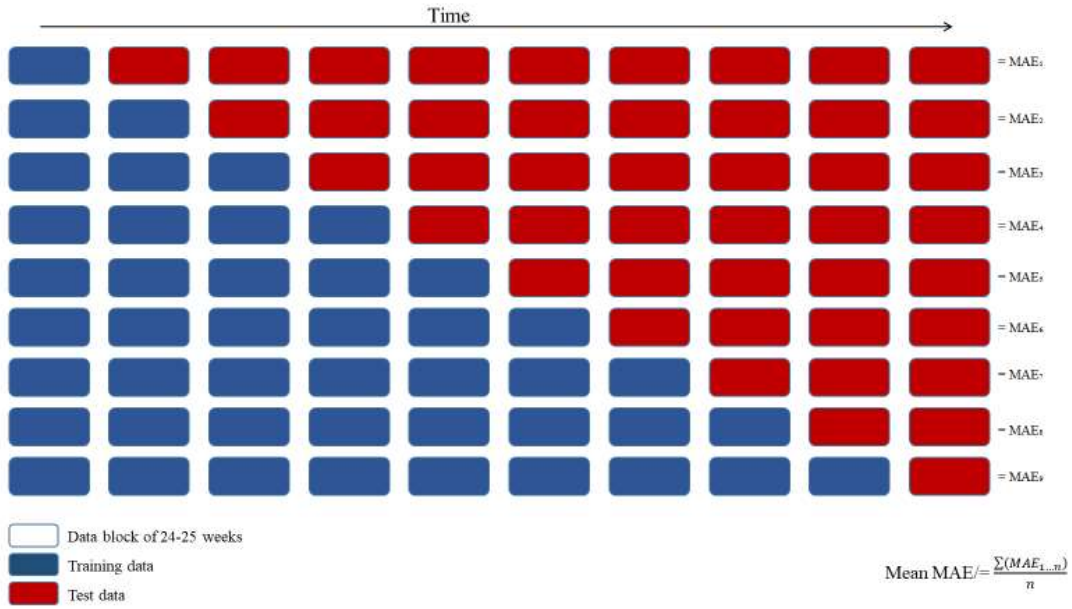

Figure 4.4: Graphical representation of the Block time series cross validation algorithm

4.5.4 Estimating RSV Burden

To estimate the attributable burden of RSV we estimated the relationship between each syndromic indicator and laboratory cases of RSV, while accounting for other effect modifiers described in Figure 4.3. The models were then used to estimate the casual relationship between laboratory cases of RSV and syndromic indicators. The key regression coefficients for RSV, influenza A and influenza B are presented in Table 4.4.

To estimate the attributable burden of RSV from the NHS111 and GPOOH syndromic indicators, first the number of weekly episodes for each syndromic indicator was predicted using the models, with the total study population in England. For the GPIH data, a linear extrapolation was undertaken of the predicted number of attributable cases of RSV with the total study population. This was done due to the poor surveillance coverage over some time periods and in some areas, using the model to predict the number of presentations using the population of the study population allows us to estimate the true absolute number presentations attributable to RSV. These predictions are the number of presentations to each service at total population level and will be referred to as the total predictions.

To estimate the causal relationship between RSV and the respiratory indicators we used a counterfactual approach, where we compare the outcome (number of presentations) when the variable of interest (RSV) was not present against when it was present. In this case we simulated an environment with zero RSV circulation (Emukule et al., 2017; Thompson et al., 2009; Yang et al., 2011), by setting the RSV term to zero. If the outcome does not differ when the variable of interest is absent, and other identified effect modifiers or confounding factors are accounted for, we can say there is no causal relationship between the number of presentations for the syndromic indicators and RSV.

The number of episodes were then predicted using a dataset where the RSV term was set to zero (baseline predictions), to simulate an environment with zero RSV circulation (Yang et al. 2011; Emukule et al. 2017). The estimate the number of presentations that could be attributed to RSV the difference between the total number of presentations in the original model and the number of presentations in the counterfactual model. This

gave us the total number of syndromic indicator presentations that could be attributable to RSV. The same method was used to estimate the attributable burden of influenza, where influenza A and influenza B terms were set to zero.

To calculate uncertainty ranges for the number of syndromic indicator presentations attributable to RSV and influenza, 95% confidence intervals were calculated for each set of predictions using the model, those with the pathogen of interest and those without the pathogen of interest in circulation. The difference was then taken between the upper bound of the two prediction datasets, and the lower bound of the two prediction datasets. Absolute number of episodes were calculated for each syndromic indicator, as well as presentation rate per 100,000 that were attributable to RSV and influenza.

Using these estimates, the percentage of presentations for each syndromic indicator attributable to RSV were estimated. To obtain results by healthcare service, we calculated the sum value of results from all individual indicators. In the case of GPOOH and GPIH services, multiple syndromic indicators were subsets of other broader indicators also analysed, for these systems, the broader syndromic indicators were used to estimate the healthcare service attributable burden (Figure 4.5). Burden estimates were calculated for each influenza season (1st October to 31st April), as this is the period when outbreaks of RSV occur in England (PHE, 2021). In the laboratory data, it was observed that 96.5% of RSV cases occurred within the influenza season. Henceforth, the data calculated for this period will be referred to as the annual figures.

			Under 1 year		1-4 years	
System	Indicator	Pathogen	Coefficient (95%CI)	Significance	Coefficient (95%CI)	Significance
NHS111	Difficulty Breathing	RSV	1.0022 (1.0018-1.0025)	***	1.0036 (1.0019-1.0053)	***
		Influenza A	1.0056 (1.0015-1.0098)	**	0.9988 (0.9957-1.0019)	NS
		Influenza B	0.9942 (0.9895-0.9991)	*	1.0018 (0.9976-1.006)	NS
	Cough	RSV	1.0020 (1.0016-1.0023)	***	1.0045 (1.0026-1.0064)	***
		Influenza A	1.0056 (1.0010-1.0102)	*	0.9998 (0.9964-1.0033)	NS
		Influenza B	1.0014 (0.9953-1.0076)	NS	1.0010 (0.9964-1.0058)	NS
	Cold/Flu	RSV	-	-	1.0005 (0.9985-1.0026)	NS
		Influenza A	-	-	1.00787 (1.0042-1.0116)	***
		Influenza B	-	-	1.0081 (1.0029-1.0134)	**
	Fever	RSV	0.9997 (0.9994-1.0000)	*	1.0005 (0.9992-1.0017)	NS
		Influenza A	1.0061 (1.0025-1.0096)	***	1.0054 (1.0035-1.0074)	***
		Influenza B	0.9972 (0.9930-1.0014)	NS	1.0023 (0.9995-1.0052)	NS
	Sore Throat	RSV	-	-	1.0013 (0.9998-1.0028)	NS
		Influenza A	-	-	1.0052 (1.0026-1.0079)	***
		Influenza B	-	-	0.9993 (0.9957-1.0030)	NS
GPIH	Acute Bronchitis	RSV	0.9998 (0.9993-1.0002)	NS	0.996 (0.9993-1.0002)	NS
		Influenza A	0.9955 (0.9887-1.0023)	NS	0.9833 (0.9887-1.0023)	NS
		Influenza B	0.9917 (0.9832-1.0003)	NS	0.9948 (0.9832-1.0003)	NS
	ILI	RSV	1.0006 (0.998-1.0015)	NS	1.0002 (0.9977-1.0026)	NS
		Influenza A	1.0206 (1.0082-1.0331)	**	1.0107 (1.0056-1.0157)	***
		Influenza B	1.0347 (1.0181-1.0514)	***	1.0294 (1.0225-1.0364)	***
	Pneumonia	RSV	1.0028 (1.0012-1.0044)	***	1.0099 (1.0071-1.0128	***
		Influenza A	1.0134 (0.9892-1.0378)	NS	1.0105 (1.0045-1.0165)	***
		Influenza B	1.0194 (0.9866-1.0522)	NS	0.9907 (0.9820-0.9933)	*
	URTI	RSV	1.0004 (1.0000-1.0007)	*	1.0020 (1.0006-1.0034)	**
		Influenza A	1.0056 (1.0013-1.0099)	*	1.0045 (1.0015-1.0074)	**
		Influenza B	0.9940 (0.9886-0.9994)	*	0.9965 (0.9924-1.0006)	NS
	LRTI	RSV	1.0018 (1.0014-1.0021)	***	1.0047 (1.0032-1.0062)	***
		Influenza A	1.0038 (0.9993-1.0082)	NS	1.0021 (0.9994-1.0047)	NS

Table 4.4: Regression coefficients of RSV, influenza A and influenza B with 95% confidence intervals for each syndromic indicator model. NS = not significant, * = ≤0.05, ** = ≤0.01, *** = ≤0.001, - = indicator not modelled.

		Influenza B	0.9936 (0.9881-0.9993)	*	0.9976 (0.9938-1.0014)	NS
GPOOH	ARD	RSV	1.0025 (1.0019 -1.0030)	***	1.0050 (1.0025 -1.0066)	***
		Influenza A	1.0083 (1.0011-1.0155)	*	1.0058 (1.0014 -1.0103)	**
		Influenza B	1.0087 (0.9981-1.0195)	NS	1.0091 (1.0021-1.0161)	*
	ARI	RSV	1.0024 (1.0018- 1.0029)	***	1.0046 (1.0027-1.0066)	***
		Influenza A	1.0098 (1.0026- 1.0171)	**	1.0067 (1.0023-1.0111)	**
		Influenza B	1.0092 (0.9986- 1.0200)	NS	1.0087 (1.0019-1.0157)	*
	ILI	RSV	1.0031 (1.0013-1.0049)	***	1.0012 (0.9954- 1.0071)	NS
		Influenza A	1.0113 (0.9851- 1.0380)	NS	1.0123 (0.9992-1.0255)	NS
		Influenza B	1.0406 (0.9983- 1.0832)	NS	1.0470(1.0258- 1.0691)	***
	DBWA	RSV	1.0045 (1.0038- 1.0052)	***	1.0055 (1.0024-1.0086)	***
		Influenza A	0.9976- (0.9875-1.0077)	NS	0.9999 (0.9930-1.0069)	NS
		Influenza B	1.0137 (0.9978- 1.0299)	NS	1.0035 (0.9928-1.0145)	NS
	Bronchitis	RSV	1.0053 (1.0044- 1.0061)	***	1.0112 (1.0083 -1.0140	***
		Influenza A	1.0145 (1.0018- 1.0274)	*	1.0016 (0.9940- 1.0092)	NS
		Influenza B	0.9786 (0.9602- 0.9973)	*	0.9912 (0.9797- 1.0027)	NS

Table 4.4: Regression coefficients of RSV, influenza A and influenza B with 95% confidence intervals for each syndromic indicator model. NS = not significant, * = ≤0.05, ** = ≤0.01, *** = ≤0.001, - = indicator not modelled.

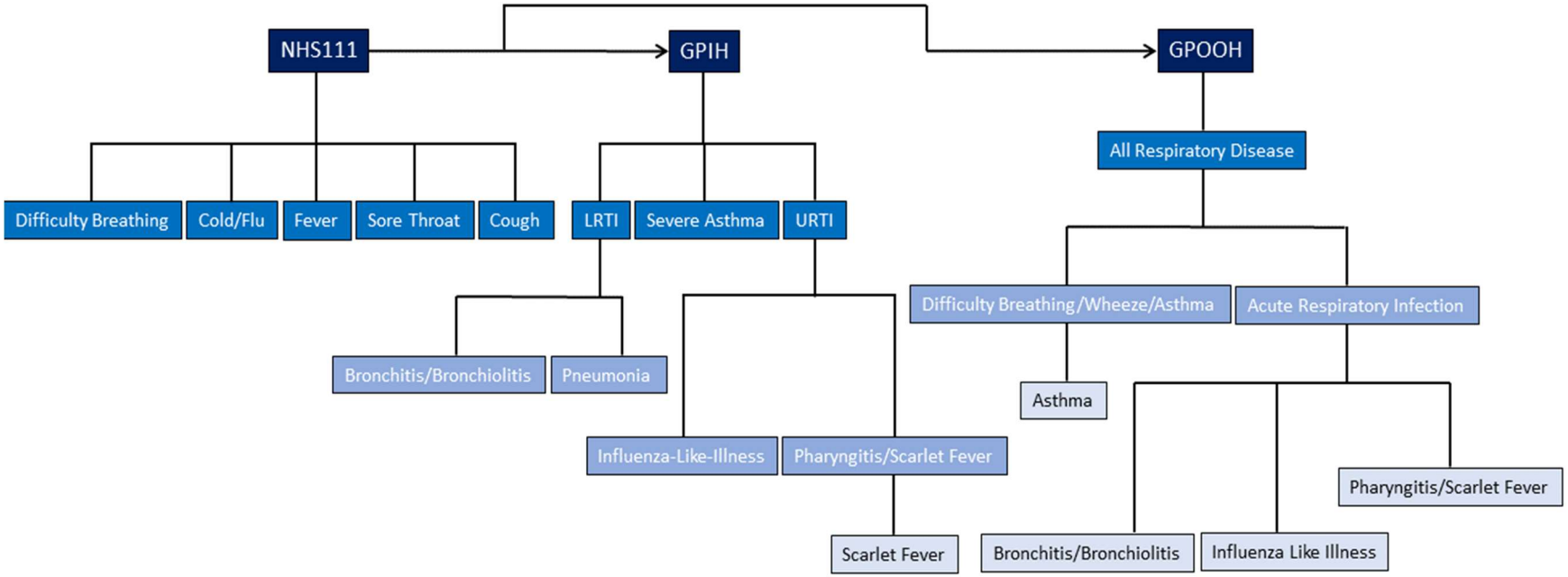

Figure 4.5: Schematic hierarchy of the syndromic surveillance indicators used in the study, and the services they relate to.

4.5.5 Estimating the Direct Cost of RSV

An estimated direct cost to the NHS was calculated using the number of presentations attributable to RSV for each service. Direct costs to each service per consultation were obtained from the literature, and only account for the consultation time and do not account for any additional resources such as prescriptions, practitioner travel time or parental loss of work. Calls to NHS111 were estimated to cost the NHS £12.26 per call (Turner et al. 2012) based on an average cost, but a high variability of cost has been observed across the call centre sites. Consultations to GPIH were estimated to cost £37.00 (Curtis 2013) based on a consultation time with a GP of 9.22 minutes for the financial year 2016/17. Previous estimates from the financial year 2013/14 have estimated this figure to be £46.00 based on a slightly longer consultation time of 11.7 minutes (Curtis 2018). Due to the study period covering between these years, cost estimates were calculated based on both of these figures to give an upper and lower bound of potential costs. Finally, consultations to GPOOH were estimated to cost £68.30 per case based on data for the financial year 2013/14, this figure is based on a report on Out-of-hours GP services in England by the Department of Health and NHS England (NAO 2014). The report noted the high level of variation in the cost of consultations for each case with 95% of costs between £28.30 and £134.30. Due to the high variability in cost to this service, estimates were calculated based on the average cost and the upper and lower estimates.

In this paper we used presentations to services monitored by syndromic surveillance to estimate community burden. Therefore, comparisons between the rate of laboratory testing and the rate of episodes attributed to RSV for each service were made.

4.5.6 Model Description and Fit

Full descriptions of model fit are provided in Table 4.5. Due to low counts (<500) over the study period; *Sore Throat, Cold/Flu, Asthma* (GPOOH), *Pharyngitis/Scarlet Fever* (GPOOH), *Scarlet Fever* (GPIH) and *Severe Asthma* (GPIH) syndromic indicators were not modelled for children under-1 year. Model distribution was chosen based on how adequately over-dispersion was controlled for. The Poisson distribution was used for

models that did not display high levels of over-dispersion (dispersion statistic >1.5). If a high level of overdispersion was observed, then the indicators were modelled using the negative binomial distribution.

Models were excluded if they displayed low deviance explained (<75%), as this indicated that the covariates did not explain the dependent data. Given this, models fitted to *Fever* (NHS111-under 1 year), *Influenza Like Illnes*s (GPOOH and GPIH– under 1 year), and *Pneumonia* (GPIH – under 1 year) syndromic indicators were not considered further.

A high percentage of deviance explained (>90%) was observed in the models for *Difficulty Breathing, Cough* (NHS111), *All Respiratory Disease, Acute Respiratory Infection, Bronchitis/Bronchiolitis* (GPOOH), *Upper Respiratory Tract Infection* and *Lower Respiratory Tract Infection* indicators for both age groups. In addition to these, there was also a high percentage of deviance explained in the *Difficulty Breathing/Wheeze/Asthma* (GPOOH) model in under-1's and *Cold/Flu* and *Influenza Like Illness* (GPIH) models in 1-4-year olds. In children under-1 year, the majority of models (11/15; 73%) had a best fit when the syndromic data led the RSV laboratory data by one week. However, in the 1-4-year age group, there was less agreement between the services, with leads ranging from 0-2 weeks. To assess overfitting, MAE was calculated using K-fold time series cross validation (Table 4.6). This analysis suggests that most of the models had low levels of overfitting, with MAE values <30% of mean observed values (28/31).

	Ages Modelled	Model Distribution	Deviance Explained		Over-dispersion		RSV Leads	
			Under 1 Year	1-4 Years	Under 1 Year	1-4 Years	Under 1 Year	1-4 Years
NHS111								
Difficulty Breathing	Both	negative binomial	0.96	0.91	1.11	1.13	1	1
Fever	Both	negative binomial	0.67	0.85	1.12	1.09	0	2
Cough	Both	negative binomial	0.97	0.94	1.10	1.10	1	1
Sore Throat	1-4 years	negative binomial	-	0.84	-	1.07	-	2
Cold/Flu	1-4 years	negative binomial	-	0.93	-	1.11	-	1
GPOOH		negative binomial						
All Respiratory Disease	Both	negative binomial	0.95	0.92	1.09	1.10	1	0
Difficulty Breathing/Wheeze/Asthma	Both	negative binomial	0.92	0.83	1.11	1.20	1	0
Asthma	1-4 years	negative binomial	-	0.79	-	1.14	-	0
Acute Respiratory Infection	Both	negative binomial	0.95	0.92	1.10	1.10	1	0
Influenza Like Illness	Both	negative binomial	0.54	0.75	1.11	1.15	1	2
Bronchitis/Bronchiolitis	Both	negative binomial	0.97	0.92	1.20	1.19	1	0
Pharyngitis/Scarlet Fever	1-4 years	negative binomial	-	0.51	-	1.14	-	2
GPIH		negative binomial						
Upper Respiratory Tract Infection	Both	negative binomial	0.93	0.92	1.09	1.06	1	1
Influenza Like Illness	Both	negative binomial	0.72	0.93	1.14	1.06	1	2
Pharyngitis or Scarlet Fever	Both	negative binomial	0.85	0.88	1.10	1.06	1	1
Scarlet Fever	1-4 years	Poisson	-	0.85	-	0.87	-	1
Lower Respiratory Tract Infection	Both	negative binomial	0.97	0.95	1.09	1.08	0	1
Pneumonia	Both	negative binomial	0.39	0.75	1.06	1.16	1	1
Severe Asthma	1-4 years	negative binomial	-	0.80	-	1.20	-	0
Bronchitis/Bronchiolitis	Both	Poisson	0.89	0.85	0.52	0.55	2	2

Table 4.5: Model distributions and summary statistics.

	Mean		MAE (% of Mean)	
	Under 1 Year	1-4 Years	Under 1 Year	1-4 Years
NHS111				
Difficulty Breathing	907.4	1441.8	77.2 (8.5)	171.3 (11.9)
Fever	-	1411.6	-	153.7 (10.9)
Cough	1143.0	2188.8	122.5 (10.7)	362.0 (16.5)
Sore Throat	-	723.3	-	86.6 (12.0)
Cold/Flu	-	370.1	-	69.1 (18.8)
GPOOH				
All Respiratory Disease	1323.5	3566.4	186.7 (14.1)	357.7 (10.0)
Difficulty Breathing/Wheeze/Asthma	119.6	284.0	27.2 (22.7)	39.4 (13.9)
Asthma	-	50.5	-	10.2 (20.2)
Acute Respiratory Infection	1038.3	2891.8	139.4 (13.4)	283.1 (9.8)
Influenza Like Illness	-	16.3	-	5.7 (35.0)
Bronchitis/Bronchiolitis	81.7	35.5	21.5 (26.3)	6.0 (16.9)
Pharyngitis/Scarlet Fever	-	10.8	-	3.3 (30.6)
GPIH				
Upper Respiratory Tract Infection	3383.9	10155.6	329.4 (9.7)	1249.0 (12.3)
Influenza Like Illness	-	57.9	-	15.6 (26.9)
Pharyngitis or Scarlet Fever	269.6	1275.1	34.5 (12.8)	138.7 (10.9)
Scarlet Fever	-	20.7	-	4.4 (21.3)
Lower Respiratory Tract Infection	775.4	1697.0	83.5 (10.8)	217.1 (12.8)
Pneumonia	2.9	16.5	4.1 (141.4)	3.8 (23.0)
Severe Asthma	-	76.2	-	13.8 (18.1)
Bronchitis/Bronchiolitis	33.0	25.5	4.3 (13.0)	3.6 (14.11)

Table 4.6: Time Series K Fold cross validation results.

4.6 Results

4.6.1 Estimating the Community Burden of RSV

Our models indicate that, during an average year, RSV-accounted for 21.7% and 20.8% of all *Difficulty Breathing* and *Cough* calls to NHS111 in children under-1 year, and 11.6% and 13.1% of calls in children 1-4 years (Table 4.7). RSV accounted for 52.3% of all *Bronchitis/Bronchiolitis* consultations to GPOOH in children under-1 year, and 13.0% of consultation in children 1-4 years. When looking at the broader GPOOH syndromic indicators, RSV accounted for 24.3% of *All Respiratory Disease* consultations in the under-1's, and 10.2% in children 1-4 years. In children under-1 year, RSV accounted for 19.0% of all presentations for *Lower Respiratory Tract Infection* to GPIH, and 13.3% in children 1-4 years. Further estimates of burden per syndromic indicator are presented in Table 4.8.

To estimate the attributable burden of RSV for each service, indicators were summed. In the case of NHS111, the indicators summed were *Difficulty Breathing* and *Cough* for the under-1 age group, and *Difficulty Breathing, Fever, Cough, Sore Throat* and *Cold/Flu* for those aged 1-4 years. Both GPOOH and GPIH data comprised of more specific indicators that were a subset of broader indicators of RSV. For both GPOOH and GPIH the broader indicators were found to detect a higher burden of RSV in comparison to the more specific indicators and were therefore used to calculate the total burden for the services. For GPOOH *All Respiratory Disease* was used and for GPIH *Upper Respiratory Tract Infection* and *Lower Respiratory Tract Infection* was used. Annual seasonal burden and a time series of burden for each service is presented in Figure 4.6 and Figure 4.7.

In total, we estimate that there were almost 200,000 presentations to NHS11, GPOOH and GPIH annually, in children under-5 years attributable to RSV over the five-year study period (Table 4.9); in children under-5 years there were a total of 9,047, 16,122, and 14,638 presentations per 100,000 population attributable to RSV to NHS11, GPOOH and GPIH, respectively. On average, there were 1,240, 2,051, and 2,544 presentations attributable to RSV per 100,000 residents per year to NHS111, GPOOH

and GPIH, respectively (Table 4.9). We estimate that, during the influenza season, 3.1% (1.8%-5.5%) of all presentations to NHS111 and 2.1% (1.0%-4.2%) of all presentations to GPOOH were attributable to RSV in children under-5 years (Table 4.9). The percentage of all GPIH consultations attributable to RSV could not be calculated because the total number of all consultations to GPIH was not available.

	Percentage (%) of All Events Attributable to RSV (Range)		Percentage (%) of All Events Attributable to Influenza (Range)		Ratio of RSV:Influenza	
	< 1 year	1-4 years	< 1 year	1-4 years	< 1 year	1-4 years
NHS111						
Difficulty Breathing	21.71 (20.88-22.55)	11.57 (10.81-12.33)	1.48 (1.10-1.86)	0.38 (0-0.91)	14.7:1	30.3:1
Cough	20.78 (19.77-21.79)	13.06 (12.19-13.94)	2.71 (2.36-3.06)	0.94 (0.43-1.56)	7.7:1	13.9:1
Fever	-	3.11 (2.56-3.66)	-	7.34 (6.92-7.76)	-	1:2.4
Cold/Flu	-	3.27 (2.17-4.38)	-	13.36 (11.94-14.79)	-	1:4.1
Sore Throat	-	5.63 (5.11-6.16)	-	6.15 (576-6.55)	-	1:1.1
GPOOH						
All Respiratory Disease	24.34 (20.05-28.64)	10.19 (8.35-12.02)	5.59 (3.67-7.51)	11.34 (8.42-14.27)	4.4:1	1:1.1
Difficulty Breathing/Wheeze/Asthma	43.36 (33.58-53.14)	13.02 (9.70-16.32)	0	2.50 (0.67-4.33)	309:1	5.2:1
Asthma	-	0.57 (0.04-2.09)	-	1.01 (0.36-2.09)	-	1:1.8
Acute Respiratory Infection	23.19 (19.08-27.30)	10.37 (8.56-12.18)	6.55 (4.48-8.62)	12.27 (9.18-15.35)	3.5:1	1:1.2
Influenza Like Illness	30.21 (11.44-48.74)	5.72 (3.13-8.39)	16.05 (0.79-31.23)	44.13 (16.06-72.21)	1.9:1	1:7.7
Bronchitis/Bronchiolitis	52.53 (38.42-66.61)	29.28 (23.16-35.44)	4.36 (0-6.54)	0 (0-0.52)	20.8:1	36.2:1
Pharyngitis/Scarlet Fever	-	-	-	-	-	1:2.9
GPIH						
Upper Respiratory Tract Infection	3.94 (3.06-4.81)	6.63 (5.89-7.37)	1.89 (1.28-1.89)	3.78 (3.44-4.21)	2.5:1	1.8:1
Influenza Like Illness	3.56 (1.04-6.30)	4.45 (3.06-5.85)	15.49 (12.47-18.05)	39.73 (34.81-44.74)	1:4.3	1:8.9
Pharyngitis or Scarlet Fever	2.19 (0.93-3.45)	4.28 (3.65-4.90)	1.16 (0.74-1.57)	6.44 (5.99-6.88)	1.9:1	1:1.5
Scarlet Fever	-	-0.97 (-2.55-0.31)	-	6.95 (6.10-7.97)	-	1:7.1
Lower Respiratory Tract Infection	18.95 (18.37-19.55)	13.30 (12.66-13.94)	0.50 (0.12-0.87)	1.50 (1.22-1.78)	37.9:1	8.8:1
Pneumonia	-	21.56 (19.73-23.76)	-	6.92 (5.70-8.09)	-	3.1:1
Severe Asthma	-	1.62 (-0.03-3.40)	-	1.96 (1.46-2.41)	-	1:1.2
Bronchitis/Bronchiolitis	-1.68 (-4.48-0.87)	-1.47 (-4.10-1.08)	-4.07 (-5.22- -2.91)	-2.76 (-3.70 - -1.79)	0.41:1	0.53:1

Table 4.7: Comparison of presentations attributable to RSV or Influenza during an average year (1st of October to 31st of April).

	Total No. of Presentations (Range)		Annual Average of Presentations (Range)		Episodes per 100,000 Population per Week (Range)	
	< 1 year	1-4 years	< 1 year	1-4 years	< 1 year	1-4 years
NHS111						
Difficulty Breathing	41,511 (39,919-43,109)	33,365 (31,180-35,559)	8,302 (7,984-8,622)	6,673 (6,236-7,112)	43 (41-45)	8 (8-9)
Cough	53,704 (51,085-56,313)	64,521 (60,197-68,835)	10,741 (10,217-11,263)	12,904 (12,039-13,767)	56 (53-58)	16 (15-17)
Fever	-	8,107 (6,664-9,541)	-	1,621 (1,333-1,908)	-	2.0 (1.6-2.4)
Cold/Flu	-	2,759 (1,827-3,690)	-	552 (366-738)	-	0.7 (0.5-0.9)
Sore Throat	-	7,600 (6,892-8,313)	-	1,520 (1,378-1,663)	-	1.9 (1.7-2.0)
GPOOH						
All Respiratory Disease	169,733 (139,779-199,677)	180,296 (147,837-212,748)	33,947 (27,956-39,935)	36,059 (29,567-42,550)	176. (145-207)	44 (37-53)
Difficulty Breathing/Wheeze/Asthma	26,258 (20,336-32,182)	14,736 (10,988-18,478)	5,252 (4,067-6,436)	2,947 (2,198-3,696)	27 (21-33)	3.7 (2.7-4.6)
Asthma	-	110 (7-403)	-	22 (1-81)	-	0.03 (0.00-0.10)
Acute Respiratory Infection	126,362 (103,960-148,762)	150,635 (124,321-176,956)	25,272 (20,792-29,752)	30,127 (24,864-35,391)	131 (108-154)	38 (31-44)
Influenza Like Illness	1,769 (670-2,854)	737 (403-1,081)	354 (134 -571)	147 (81-216)	1.8 (0.7-2.9)	0.2 (0.1-0.3)
Bronchitis/Bronchiolitis	28,881 (21,123-36,626)	5,579 (4,412-6,753)	5,776 (4,225-7,325)	1,116 (882-1,351)	30 (22-38)	1.4 (1.1-1.7)
Pharyngitis/Scarlet Fever	-	482 (278-6,84)	-	96 (56-137)	-	0.1 (0.1-0.2)
GPIH						
Upper Respiratory Tract Infection	52,728 (40,978-64,414)	234,864 (208,625-261,098)	10,546 (8,196-12,883)	46,973 (41,725-52,220)	55 (43-67)	59 (52-65)
Influenza Like Illness	165 (48-292)	1,023 (702-1,343)	33 (10-58)	205 (140-269)	0.2 (0.1-0.3)	0.3 (0.2-0.3)

Table 4.8: Presentations attributable to RSV for each syndromic indicator during the average year (1st of October to 31st of April).

Pharyngitis or Scarlet Fever	2,260 (957-3,563)	17,777 (15,157-20,374)	452 (191-713)	3,555 (3,031-4,075)	2 (1-4)	4.4 (3.8-5.1)
Scarlet Fever	-	-82 (-215-26)	-	-16 (-43-5)	-	-0.02 (-0.05-0.00)
Lower Respiratory Tract Infection	64,649 (62,653-66,684)	81,870 (77,925-85,833)	12,930 (12,531-13,337)	16,374 (15,585-17,167)	67 (65-69)	20 (19-21)
Pneumonia	-	1,247 (1,141-1,374)	-	249 (228-275)	-	0.31 (0.28-0.34)
Severe Asthma	-	380 (-8-799)	-	76 (-2-160)	-	0.10 (0-0.2)
Bronchitis/Bronchiolitis	-383 (-1020-198)	-160 (-447-118)	-77 (-2-40)	-32 (-89-24)	-0.4 (-1.0-0.20)	-0.04 (-0.11-0.03)

Table 4.8: Presentations attributable to RSV for each syndromic indicator during the average year (1st of October to 31st of April).

	Average Events Per Year (total)	Average Events Per Year (per 100,000 population)	Total Number of Events Over Study Period (per 100,000 population)	Percentage of All Presentations (%)*	Ratio of RSV:Influenza
NHS111					
< 1 year	19,043 (18,200-19,884)	2,882.5 (2,754.9-3,009.8)	13,867.3 (13,067.6-14,666.6)	4.2 (3.0-5.9)	7.2:1
1-4 years	23,720 (21,352-25,187)	861.8 (775.8-915.1)	4,227.5 (3,879.0-4,575.8)	2.6 (1.3-5.3)	2.6:1
<5 years	42,313 (39,552-45,071)	1,239.8 (1,158.9-1,320.6)	9,047.4 (8,473.3-9,621.2)	3.1 (1.8-5.5)	5.1:1
GPOOH					
< 1 year	33,947 (27,956-39,935)	5,138.5 (4,231.7-6,044.9)	25,692.2 (21,158.1-30,224.8)	2.9 (1.8-4.4)	4.3:1
1-4 years	36,059 (29,567-42,550)	1,310.2 (1,074.3-1,546.0)	6,550.8 (5,371.5-7,729.9)	1.8 (0.7-4.1)	1:1.1
<5 years	70,006 (57,523-82,485)	2,051.2 (1,685.5-2,416.9)	16,121.5 (13,264.8-18,977.4)	2.1 (1.0-4.2)	2.4:1
GPIH					
< 1 year	23,475 (20,726-26,220)	3,553.4 (3,137.3-3,969.9)	17,767.2 (15,686.5-19,844.1)	-	5.1:1
1-4 years	63,347 (57,310-69,386)	2,301.6 (2,082.3-2,521.5)	11,508.11 (10,411.4-12,605.3)	-	2.2:1
<5 years	86,822 (78,036-95,606)	2,543.9 (2,286.5-2,801)	14,637.7 (13,049.0-16,224.7)	-	3.4:1
All Systems					
< 1 year	76,465 (66,882-86,039)	11,574.4 (10,123.8-13,023.6)	57,326.7 (49,912.3-64735.5)	-	5.1:1
1-4 years	123,126 (108,229-137,123)	4,473.6 (3,932.4-4,982.2)	22,286.4 (19,661.9-24,911.0)	-	1.6:1
<5 years	199,591 (175,111-223,162)	5,848.1 (5,130.8-6,538.8)	39,806.6 (34,787.1-44,823.3)	-	3.1:1

Table 4.9: RSV attributable events to NHS111, GPOOH, GPIH and All Systems. *Total number of presentations only available for NHS111 and GPOOH.

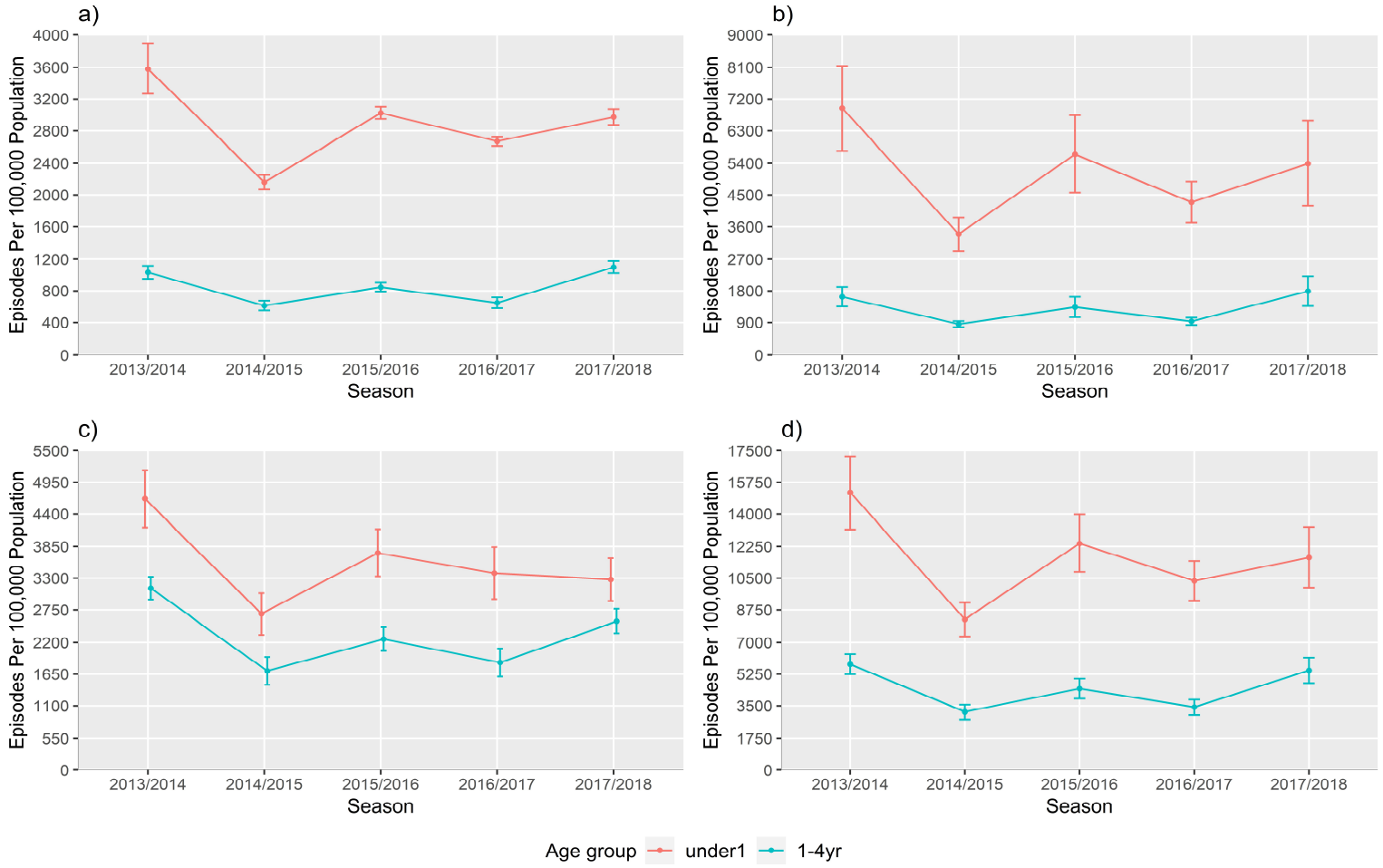

Figure 4.6: Estimated number of RSV episodes per 100,000 population per year by service monitored by syndromic surveillance; a) NHS111, b) GPOOH, c) GPIH, d) All services. Error bars indicate the error range based on upper and lower estimates of RSV.

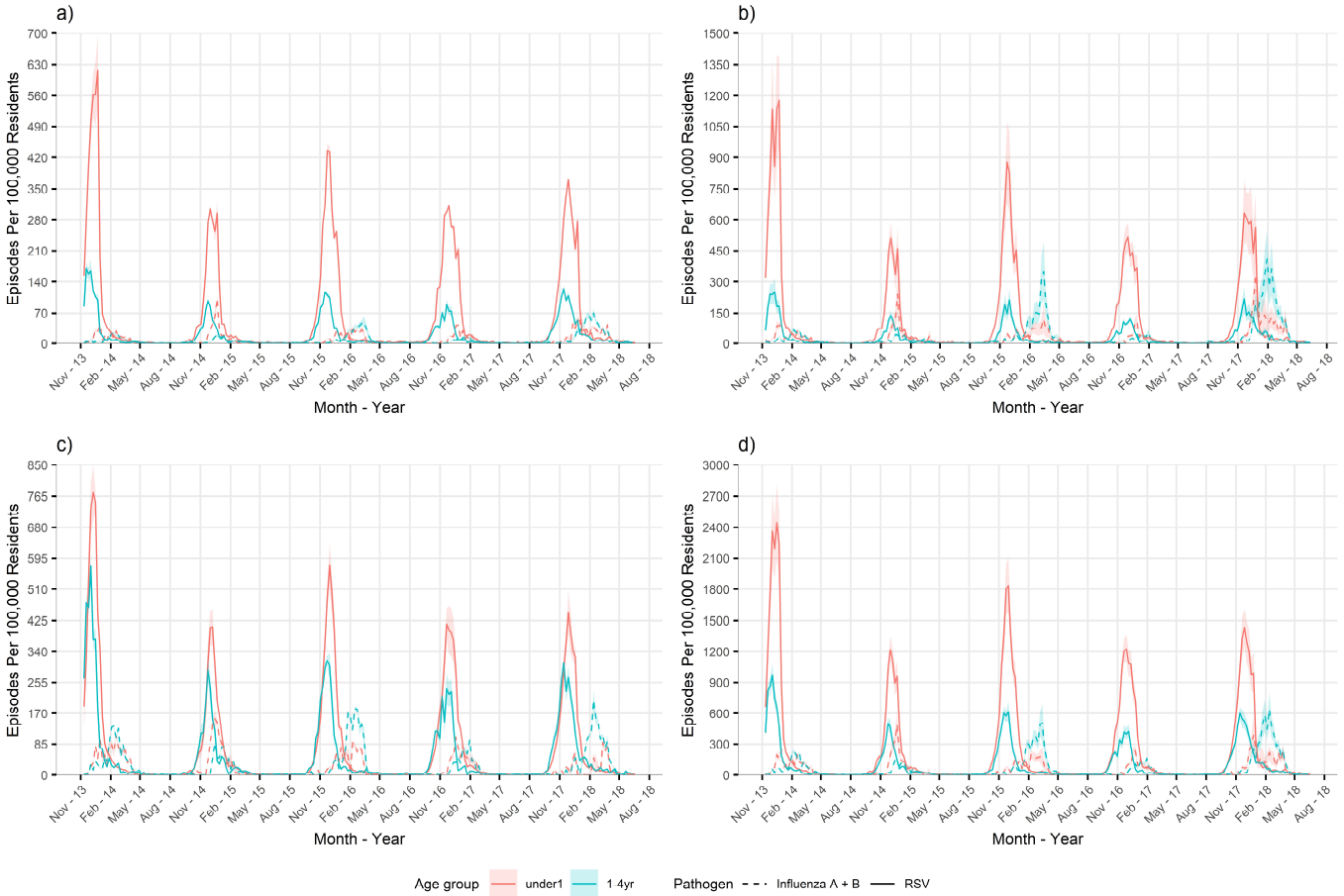

Figure 4.7: Time Series of RSV (solid lines) and influenza (dashed lines) episodes per 100,000 residents by service monitored by syndromic surveillance for the period 11th November 2013 and 18th June 2018; a) NHS111, b) GPOOH, c) GPIH, d) All services. Light shaded areas indicate the error range based on upper and lower estimates of RSV.

4.6.2 Community RSV Burden by Age Group

Across all three services, the burden was higher in children under-1 year compared to the 1-4 years age group. To NHS111 there were 3.3 times the rate of presentations in children under-1 year compared to those aged 1-4 years over the study period. To GPIH the presentation rate was 1.5 times greater in those under-1 year and to GPOOH the rate was 3.9 times higher in that under-1 year compared to those aged 1-4 years.

4.6.3 Comparing Community RSV Burden to that of Influenza

Compared to influenza A+B, there were more RSV presentations to NHS111, GPOOH and GPIH in children under-5 years, with 5.1, 2.4, and 3.4 the rate of presentations, respectively (Table 4.9). The largest disparity between RSV and influenza A+B was seen in the under-1 year age group, with 7.2 times more presentations attributable to RSV across all services.

4.6.4 Estimating the Direct Healthcare Cost of Community RSV Burden to the NHS

Calls to NHS111 attributable to RSV in children under-5 years were estimated to have a direct cost £520,000 per year (Table 4.10). The RSV attributable cost of GPIH consultations was estimated to be £4,073,000 per year for children under-5 years. RSV attributable consultations to GPOOH cost £5,327,000 in children under-5 years per year. The total direct attributable cost of RSV to these services amounts to £9,920,000 per year in all children under-5 years. Due to the variability of the costs for each system per episode further estimates have been made in Table 4.10, as well as cost breakdowns by age group.

	Cost Source	Reference	Cost per Episode	< 1 Year (£1,000s)	1-4 Years (£1,000s)	< 5 Years (£1,000s)
NHS111	Pilot study of four sites in 2012[1]	Turner et al. (2012)	£12.26	228 (215-241)	292 (268-316)	520 (483-558)
GPOOH	Report on Cost per case to GPOOH in 2013/14: Average (95% lower & upper cost)[2]	National Audit Office (2014)	£68.50 (£28.30-£134.30)	2,359 (975-4,626)	2,967 (1,226-5,818)	5,327 (2,201-10,780)
GPIH	Unit of care cost 2016/17 & 2013/14[3]	Curtis (2013; 2018)	£37.00 - £46.00	880-1,094	2,395-2,978	3,276-4,073
All Services	*NHS111: £12.26 GPOOH: £68.50 GPIH:* £46.00[4]			3,682	6,238	9,920

Table 4.10: Estimated direct cost (£) per year to NHS111, GPOOH and GPIH by age group. Estimates are based on central estimates of average annual number of RSV related episodes.[1] No range of values available for NHS111. [2] Value range based on the cost of 95% of consultations to GPOOH. [3] Values based on unit of care cost at start and end of study period. [4] 2013/14 GPIH chosen for final cost as this cost was based on a longer consultation time.

4.6.5 Comparison Between Syndromic and Laboratory Detections of RSV

The rate of episodes attributable to RSV to each service was compared to the rate of laboratory cases over the study period to estimate how many more episodes of RSV could be detected through syndromic surveillance (Table 4.11). For every one laboratory confirmed case in children under-5 years, 12.4, 20.7, 25.8 RSV related presentation was detected by NHS111, GPOOH and GPIH, respectively. The ratio between number of cases detected through syndromic surveillance to laboratory cases was highest in the 1-4 years age group, with the largest disparity in GPIH where 95 presentations were detected for every laboratory confirmed cases in this age group (Table 4.11). Time series of RSV attributable laboratory, NHS111, GPIH and GPOOH detections in children under-5 years is visualised in Figure 4.8.

System	Age Group	Ratio of Syndromic Episodes for One Laboratory Cases
NHS111	<1 Year	6.8 (6.4-7.2)
	1-4 Years	34.9 (32.1-37.8)
	<5 Years	12.4 (11.5-13.2)
GPOOH	<1 Year	12.6 (10.4-14.8)
	1-4 Years	53.8 (44.2-63.4)
	<5 Years	20.7 (17.1-24.4)
GPIH	<1 Year	8.7 (7.7-9.7)
	1-4 Years	94.8 (85.8-103.8)
	<5 Years	25.8 (23.2-28.4)

Table 4.11: Ratio of RSV attributable episodes to syndromic surveillance systems for one laboratory confirmed case of RSV by age group and system.

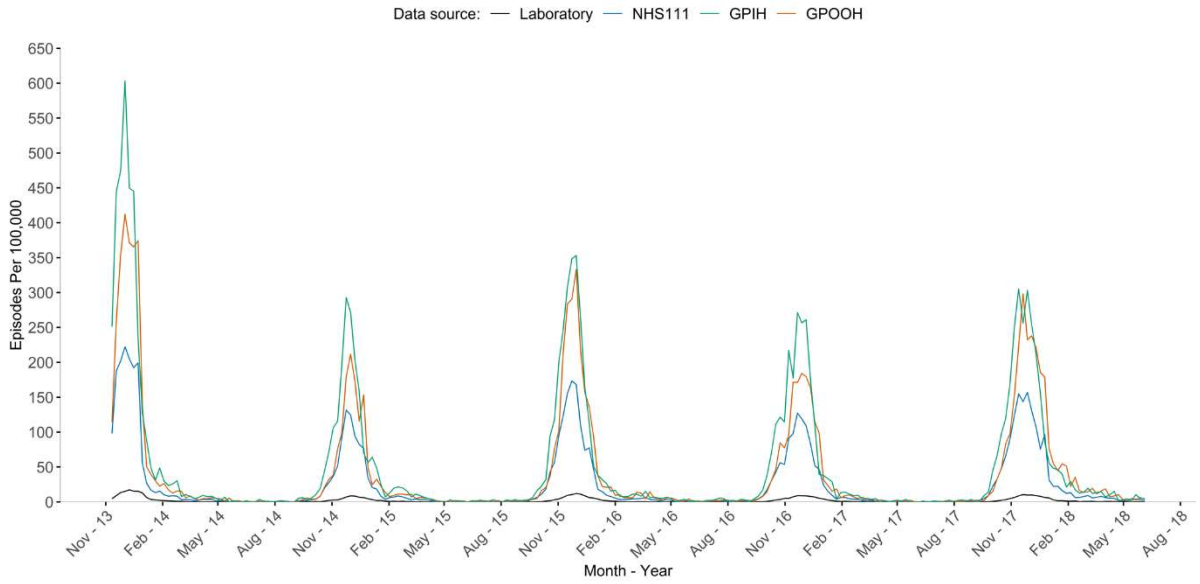

Figure 4.8: Time series of rate per 100,000 of RSV attributable episodes detected by syndromic surveillance and laboratory confirmed cases in children under-5 years for the period 11th November 2013 and 18th June 2018.

4.7 Discussion

4.7.1 Impact

This work comprises one of the most comprehensive analyses of respiratory syndromic surveillance data with over seven million healthcare presentations recorded. We used this rich data set to estimate the attributable burden of RSV from three healthcare services in England. These services focus upon healthcare at the community level and provide a broader estimate of attributable burden of disease compared to hospitalisations or laboratory-based surveillance. Our estimates of RSV risk were extrapolated to a full population providing an estimate for the community burden in children under-5 years in England. These estimates allow policy-makers and health professionals to make meaningful and informed decisions for targeted research, to identify and reduce health inequalities, control disease spread, plan and implement effective interventions and, ultimately, improve health outcomes. Our methods may be used to estimate other community burdens of disease in other settings.

4.7.2 Main Findings

We estimate that there were almost 200,000 RSV attributable presentations to NHS111, GPIH and GPOOH per year in children under-5 years over the five-year study period. Out of the three healthcare systems analysed, GPOOH had the highest rate of presentations in children under-1 year, and GPIH had the highest presentation rate in children aged 1-4 years. We observed that presentation rates were higher in children under-1 year compared to children ages 1-4 years across all three systems. When compared to influenza there were more presentations attributable to RSV, this was observed across the three systems.

Estimating the Community Burden of RSV in England

Morbey et al. (2017a; 2018) investigated the association between respiratory pathogens and presentations to NHS111, GPIH and GPOOH. Although their studies used different methods and time periods, the results were broadly comparable to those presented in this paper, with RSV associated with *Difficulty Breath* and *Cough* calls to NHS111,

Lower respiratory Tract Infection consultations to GPIH and *All Respiratory Infection* consultations to GPOOH in children. However, the focus of both the Morbey et al. (2017a; 2018) studies was to explore the associations between respiratory pathogens to specific syndromic indicators. Our research is a notable advance on this work as we estimate the number of RSV-associated presentations for each syndromic indicator, and for each healthcare service. This has allowed us to produce burden estimates for RSV in the wider community, essential in public health terms, for understanding the current burden, exploring the potential impact of an RSV vaccine, and making informed decision on targeted research and public health interventions. These burden estimates also permit comparisons between age groups, comparisons to influenza and economic burden estimates to be produced.

Two studies have produced RSV burden estimates for England in children, Taylor et al. (2016) and Cromer et al. (2017). Both studies focused on data from clinical settings; GPIH, hospitalisation and deaths, whereas our research provides a comprehensive study of RSV in the community, but using NHS111, and GPOOH in addition to GPIH. Taylor et al. (2016) estimated that the RSV-attributable GP presentation rate annually for infants <6 months was 11,000-17,000 per 100,000, 10,000-15,000 per 100,000 in children 6-23 months and 5,000-10,000 per 100,000 in children 2-4 years over the 14-year study period. A similar study by Cromer et al. (2017) estimated 12,000 per 100,000 annual GPIH presentations in children under-5 years attributable to RSV in England. These estimates are higher than our estimates which are between 2,282-5,138 RSV related presentations per 100,000 in children under-1 year and 861-2,301 presentations per 100,000 in children ages 1-4 years. Although Taylor et al. and Cromer et al. used similar statistical methods, there are several reasons why our study may have produced lower estimates.

A key factor is that we accounted for more seasonal respiratory pathogens in our analysis than both Taylor et al. (2016) and Cromer et al. (2017). Taylor et al. (2016) only accounted for influenza in their model, and although Cromer et al. (2017) accounted for more pathogens (n = 8), backwards stepwise regression was used to remove the pathogens that were not significant in the model. Backwards stepwise

regression to remove non-significant variables is often criticised as it can result in explanatory variables with casual effects being removed from the analysis (Smith 2018). It is unclear which pathogens were accounted for in the final models used by Cromer et al. (2017). Rhinovirus is a common seasonal virus and is the second most common cause of viral bronchiolitis in children under-1 year after RSV (Lo et al. 2018; Kotaniemi-Syrjänen et al. 2008; Turunen et al. 2014), with 20-40% of children under one-year diagnosed with bronchiolitis are infected with rhinovirus (Jartti et al. 2009). It was unclear if Rhinovirus was accounted for in Cromer et al. (2017) and was it not accounted for by Taylor et al. (2016). The omission of other seasonal respiratory pathogens would overestimate the impact of RSV on presentations to NHS111, GPIH and GPOOH.

Our study included more Fourier terms to account for a higher degree of seasonality compared to Taylor et al. (2016). Cromer et al. (2017) did not account for seasonality when estimating the burden of RSV. Our results assume that less of the temporal variation in the data is explained by the respiratory pathogens, hypothesising other seasonal factors such as temperature and humidity might also contribute to presentations to the healthcare services studied. Therefore, Taylor et al. (2016) and Cromer et al. (2017) will produce higher burden estimates for RSV, by not account for seasonality adequately.

In Taylor et al. (2016), the study period was from 1995-2009; and in Cromer et al. (2017) the study period was from 2001 to 2008. GPOOH was introduced in 2004, and NHS111 in 2013. The introduction of these services has changed the way healthcare is sought in the community and may have led to a decrease in GPIH presentations, especially for acutely presenting illnesses. When estimating burden, it is an imperative to have recent data as this more accurately reflects current trends in both RSV and healthcare-seeking behaviours. There has been no previous research into the RSV-attributable burden to GPOOH and NHS111 services so a direct comparison to the literature could not be made.

Community RSV Burden by Age Group

Compared to 1-4-year olds, children under-1 year had a substantially higher attributable burden of RSV in all of the three healthcare services, with between 1.5 and 3.9 more presentations. This is in agreement with previous research (Reeves et al. 2017; Zhou et al. 2012; Ajayi-Obe et al. 2008; Taylor et al. 2016; Hall et al. 2009). The highest attributable burden of RSV in children under-1 year was observed in the GPOOH service and in those aged 1-4 years the highest burden was observed in the GPIH service. The high burden of RSV in GPOOH for children under 1 year may be due to the acute nature of RSV in young children (Glezen et al. 1986), highlighting the severity of illness in children under-1 year with the first infection associated with more severe lower respiratory tract infection, and reinfections often restricted to the upper respiratory tract. This may lead to enhanced use of GPOOH in children under-1 year in comparison to GPIH which is, generally, used for more routine illnesses, and NHS111 which is used as a non-emergency helpline.

Comparing Community RSV Burden to that of Influenza

In both age groups, the attributable burden of RSV to all three systems was larger than those attributable to influenza A and B. Higher burden of healthcare usage due to RSV compared to that of influenza has previously been described by several other European and North American studies (Schanzer, Langley and Tam 2006; Taylor et al. 2016; Heiden, Buchholz and Buda 2019; Bourgeois et al. 2009).

The disparity between RSV and influenza presentations was highest in children under-1 year compared to those aged 1-4 years in all three systems. Previous studies have estimated that 100% of children are infected with RSV by two years and up to 40% of children are infected with influenza before the age of three years (Glezen et al. 1986; El Guerche-Séblain et al. 2019), suggesting that there is more circulation of RSV than influence in young children. Influenza outbreaks are also less predictable, in terms of both magnitude and timing, compared to RSV, therefore it may be harder to estimate influenza burden using the methods described in this study. During the study period, all children aged 2-3 years were offered the influenza vaccine in primary care facilities and

there was a phased introduction of in school vaccination for all primary school children (4-10 years) (Green et al. 2015). By 2017/18 44.4% of children aged 3 years, and 59% of eligible school children had received the influenza vaccine. Vaccination in those aged two and over will reduce the influenza burden observed in this age group during the study period. It may also confer indirect protection from influenza to others, including, children under-2 years, through the vaccine herd effect, especially to those within the same household (Yin et al. 2017).

Estimating the Direct Healthcare Cost of Community RSV Burden to the NHS

The direct costs attributable to RSV related presentations to the three services were calculated to be almost £10 million per year. This consists of over £3.7 million per year in children under-1 year and over £6.2 million per year in children aged 1-4 years. These are based on average consultation costs but are subject to variability due to multiple factors including consultation length, and type and severity of illness (Turner et al. 2012; Curtis 2018; 2013; NAO 2014). A vaccine cost-effectiveness study estimated the direct cost of RSV to GPIH services for children under-5 years, as £16 million annually (Cromer et al. 2017), this figure is higher than the cost we estimated for GPIH services. However, we have already discussed, the methodology used in this study will lead to a greater number of RSV cases in comparison to the results presented here.

Comparison Between Syndromic and Laboratory Detections of RSV

By comparing the number of RSV episodes attributed to RSV for each healthcare service to the number laboratory confirmed cases of RSV, we can estimate the potential under-reporting (symptomatic cases who have sought healthcare but are not captured by surveillance systems) in laboratory surveillance. In children under-5 years every laboratory confirmed detection of RSV is estimated to result in 12, 21 and 26 presentations to NHS111, GPOOH and GPIH, respectively. This gives an overall ratio of laboratory to community healthcare presentations of between 1:26 and 1:69 depending upon the number of presentations to multiple systems. When stratified by age, more episodes were detected in the 1-4 year age group compared to the under-1

year age group across all services, in comparison to laboratory cases. This was particularly evident in GPIH, where 95 episodes were detected for every one laboratory detection. This age group difference reflects the much higher testing rate in children under-1 year due to disease severity, and the complication risk in infants. These ratios are unique and to our knowledge there are no other studies in the literature that measure the level of under-reporting of RSV in the community. This study highlights the high level of under-reporting of RSV in the community, with the level of under-reporting influenced by age. Further research is needed to understand the level of underreporting of RSV in other age groups.

4.7.3 Limitations

The syndromic data obtained for this research came from a real-time passive surveillance service. Consequently, there are periods where data is not received, or data reporting is incomplete. Coverage issues would likely increase the uncertainty around our estimates. We tried to reduce this effect by using change point analysis to identify and exclude areas with poor coverage and denominator populations were adjusted accordingly. Analysis was reliant on accurate classification of disease indicators by medical professionals. Although coding by medical professionals can lead to more specific diagnosis, coding practices are likely to vary between individuals, and by healthcare service provider. This limitation was accounted for by including broad coding terms in the analysis, such as *All Respiratory Disease* and *Upper Respiratory Tract Infection*, which will capture much of this variation. NHS111 is potentially less prone to variations in coding practises due to the structured coding system in place.

Although symptoms of RSV are predominantly respiratory, some non-respiratory symptoms, such as ear infections (otitis media), are associated with RSV. Although, several studies have observed the high prevalence of this symptom in young children with RSV; with 40-50% of hospitalised cases were observed to display this symptom (Heikkinen, Ojala and Waris 2017; Papenburg et al. 2012), most episodes of otitis media are thought to results from complications of URTI (Heikkinen and Chonmaitree 2003). From the surveillance systems utilised in this study we could only obtain respiratory

indicators that were associated with RSV. Hence, we may have underestimated the true RSV burden. However, it is not clear what proportion of RSV cases have otitis media only, and this will only influence our results if this is the only symptom that presented.

In a time-series study of any seasonal illness, the level of seasonal adjustment in the model can greatly influence the results. In this study seasonal adjustments (three sin/cosine pairs) were adopted to control for unobserved seasonal confounders. However, due to the highly cyclic nature of RSV in temperate regions, the addition of this component may have resulted in an underestimation of the RSV burden. This problem can be highlighted with the observations from the *Bronchitis* indicator from GPIH. When seasonality was accounted for, no RSV episodes attributable to *Bronchitis* were detected. Yet bronchitis is a common symptom of RSV. If seasonal trends were not accounted for in the model, we would have assumed that all the seasonal variation in the data is due to the respiratory infections included in the model, which may not be correct. Future research may evaluate the effects of using different seasonal terms on this data set.

In children under-1 year, periods of influenza burden were observed to be negative for GPIH, this was a particular problem for seasons 2015/16 and 2017/18. These negative predictions arise from the predicted number of episodes being higher if no influenza was in circulation compared to the dataset with influenza in circulation. These negative predictions could be due to the unpredictability of influenza activity and the relatively low number of confirmed cases on influenza in children under-1 year. The model relates the temporal trends of the syndromic data to that of the laboratory data, during these periods of negative predictions the temporal trends between the two datasets did not follow the same relationship as expected. This is only an issue with the GPIH prediction in children under-1 year, and the other services and age group are not affected by this issue. These negative estimates may result in a higher observed difference between RSV and influenza in this age group.

Only healthcare services in the community were included in the study. Therefore, children attending hospitals, and those who did not seek one of our three healthcare services (e.g., seeking advice from a pharmacist only) are excluded from our burden or

cost estimates. Our costs are also focussed on direct costs and other costs, such as prescription costs and costs resulting from parents having to take time of work to care for a sick child, are not included. Given the possibility that a patient can present to multiple services, or the same service more than once, one syndromic episode attributable to RSV does not necessarily equal a case of RSV detected through laboratory testing.

4.7.4 Conclusion

Here, we present a thorough investigation of the community RSV burden in England. We use one of the most comprehensive sets of respiratory syndromic surveillance data comprising seven million reports from a telehealth system and two general practitioner systems. Our results highlight that there is a substantial burden of RSV in children under-5 years in England and it exceeds that of influenza. Burden estimated in children under-1 year was found to be highest in GPOOH services, which could reflect the acute and severe nature of the illness in that age group. Compared to previous studies in the field, our inclusion of two novel data sources (i.e., GPOOH and NHS111), highlights the need to use additional forms of healthcare data sources to capture changes healthcare usage. This research provides an estimate for RSV-attributable burden from three community healthcare services, which can be used for vaccine impact studies. We argue that these estimates are meaningful to inform decision-making and planning public health processes.

Chapter 5

Identifying Differences in Regional RSV Burden and Seasonality in Children Under-One Year, England

5.1 What is Known About the Subject?

- RSV is a seasonal viral pathogen that causes respiratory tract infections and has a high health and economic burden in young children and infants.
- In temperate regions, such as England, RSV has distinct seasonal patterns with most infections occurring during the winter, between November and April.
- An association between RSV seasonality and latitude has previously been observed, with RSV activity peaking in the south and moving north. Peak timing has been observed to be later in more northernly latitudes.

5.2 What this Study Adds?

- We estimate the regional burden and activity timing of RSV in the community in children under-1 year in England.
- We observed spatial patterns in RSV burden, but not RSV activity timing, although some variation was observed. Further research is required to investigate the reasons behind these differences.
- This study provides a framework for further investigations of spatial variations in RSV burden and activity timing in the community.

5.3 Abstract

Respiratory syncytial virus (RSV) has a high health and economic burden on infants worldwide. Evidence from the literature suggests that RSV epidemics are associated with latitude and, in temperate countries, RSV has distinctive winter seasonal outbreaks. It is important to understand the timing of these epidemics in order to implement targeted public health measures. In England, seasonal epidemics occur between November and April. Here, were aim to describe the spatial variation of RSV burden and epidemic timing across England in children under-1 year.

Poisson regression models, using a Bayesian approach, were used to estimate the number of RSV attributable presentations from syndromic indicators to two community healthcare services; a telehealth service (NHS111) and family doctors (GPIH) at Public Health England (PHE) centre level (n=9). The models regressed syndromic indicator data against laboratory confirmed cases of RSV, while accounting for other seasonal respiratory pathogens, gender, public holidays, seasonality, and temporal autocorrelation. Model estimates were used to calculate seasonal burden of RSV, peak timing, start week, and epidemic length of RSV activity at each PHE centre for both NHS111 and GPIH. These burden estimates and timings were compared to identify spatial patterns of RSV.

RSV Burden per 100,000 people was observed to be highest in the southern regions compared to the rest of England for both services. Mean peak week of RSV activity was observed to vary, regionally, by 1.7 weeks for NHS111 presentations, and half a week for GPIH presentations. For both healthcare services RSV activity mean start week was observed to vary, regionally, by two weeks. Mean activity length had regional variations of a week, from epidemic activity lasting 9-10 weeks for both services. In the metrics used to measure RSV timing there was little distinctive evidence of spatial patterns of RSV activity.

We observed RSV burden was highest in the south of England, but we could not identify specific spatial patterns in the timing of RSV activity, although variations were

observed. Further research is required to investigate the reason behind these variations. This study provides a framework for future research.

5.4 Introduction

There is evidence of a seasonal component to global respiratory syncytial virus (RSV) activity and disease transmission which are thought to be driven by geographical and environmental factors (Obando-Pacheco et al. 2018). Outbreaks of RSV have distinct annual patterns of disease, which are determined by latitude. In tropical regions, typically warm and humid, RSV is endemic and occurs throughout the year with peaks in summer and early autumn (Yusuf et al. 2007). In temperate regions, RSV outbreaks have a seasonal trend and occur during the winter (Yusuf et al. 2007), with seasonal outbreaks of RSV in the northern hemisphere between November and April, and between March and October in the Southern hemisphere.

A global study of 27 countries investigated at the seasonality of RSV to understand the timing of epidemics worldwide. This study used data sources from official surveillance of laboratory confirmed cases of RSV (n=26), or official syndrome-based surveillance of Bronchiolitis diagnoses (n=1) when laboratory data was not available. RSV epidemics were observed to start in southern countries and move north. With activity starting in countries in the southern hemisphere between March and June and in the Northern Hemisphere between September and December. Seasons typically lasted 5-6 months, with shorter seasons seen in Spain, United Kingdom and Israel. In countries with wet/humid seasons, RSV activity lasted up to 10 months. Regional seasonality of RSV appeared constant, with RSV activity differing by only 1-3 weeks. Germany and Finland displayed different patterns in activity, with a late and early season observed in Germany, and a two-year cycle of alternating subgroups (A and B) in Finland (Obando-Pacheco et al. 2018). Another global study investigated RSV circulation in seven countries: Bangladesh, Guatemala, Thailand, China, Egypt, South Africa and Kenya. Each country experienced 1-2 epidemic periods each year and RSV peaked in wet months in countries with high precipitation, and during cooler months in countries

where the climate is hot and arid. There was no seasonal peak in countries on the equator (Haynes et al. 2013). A European study of 15 countries that reported RSV surveillance data to the European Centre for Disease Prevention and Control found similar results to the previous studies described with RSV activity peaking in southern countries first and moving north. RSV epidemics were observed to last longer in more northerly countries (Broberg et al. 2018). These studies suggest that RSV activity is linked to latitude gradient, which is supported by Bloom-Feshbach et al. (2013) who observed that peak RSV timing occurred later with increasing latitude.

Intra-country variation in RSV activity has been demonstrated in several studies. Mullins et al. (2003) used laboratory confirmed cases of RSV to determine the timings of RSV activity in the United States of America (USA). In southern regions, RSV activity began significantly earlier and lasted longer than other regions, and in the Midwest RSV, activity began significantly later and the seasons were shorter. Oren et al. (2018) used internet search engine queries associated with RSV to estimate spatial and temporal peak RSV activity across the USA. Seasonal peak of RSV related search queries was observed to move from South-East to North-West. A study conducted in Australia looked at the spatial spread of RSV laboratory confirmed cases and bronchiolitis hospitalisations across eight administrative regions that spanned several climatic zones. It was found that RSV activity, from both the RSV and bronchiolitis presentations, showed winter epidemic peaks in July/August in the southern regions which are of temperate climate. There were less identifiable seasonal peaks in the northern tropical regions (Hogan et al. 2016).

Although it is not entirely clear why these global and regional seasonal variations in RSV activity occur, RSV activity does appear to correlate with environmental conditions such as air pollution, temperature, rainfall and humidity. A study conducted in Greece by Sirimi et al. (2016) found a significant negative correlation between mean monthly temperature and RSV activity. Yusuf et al. (2007) observed that RSV activity is inversely related to UVB radiance and temperature in temperate regions, with RSV activity higher in cooler temperatures (in winter) in temperate climates. In more tropical climates, RSV activity was associated with high ambient temperatures and humidity.

Haynes et al. (2013) observed that RSV activity was associated with periods of high rainfall, humidity and temperature in Thailand, and with humid and warm months in Bangladesh and Thailand. RSV peaks occurred during cooler months in China and Egypt. Paynter (2015) observed that cold, dry conditions of temperate winters encourage RSV transmission by increasing the survival time of the virus on surfaces. In tropical regions, wet conditions are associated with increased RSV activity by increasing virus survival in droplets. A study by Belderbos et al. (2011) found an association between vitamin D deficiency in umbilical cord blood plasma of healthy neonates and lower respiratory tract infections due to RSV in the first year of life. Vitamin D deficiency has previously been associated with latitude, with higher risk in more northernly regions (Leary et al. 2017; Huotari and Herzig 2008). Although there is growing evidence highlighting the association between environmental variables and RSV activity, it is likely that human physiology, human behaviour and virus-environment interactions play a significant role in disease activity (Jackson et al. 2013; Sloan, Moore and Hartert 2011).

The majority of research investigating RSV activity within countries has focused on large countries with a large climatic spread, such as the USA and Australia (Hogan et al. 2016; Mullins et al. 2003). Although England is a small county, there can be substantial variations in regional climates due to topography, land use, latitude, proximity to the mid-latitude westerly wind belt in the Atlantic Ocean, and proximity to the continental influences of Europe (Met Office 2020) (Figure 5.1). The latitude of England ranges from around 50° north to 56° north, encompassing the same latitude as Germany, Poland, The Netherlands, Ireland and Denmark, where differences in RSV seasonality have been observed (Broberg et al. 2018). Given the associations with RSV activity and latitude, and subsequent environmental conditions, there could be substantial variation in burden and timing of RSV activity in England.

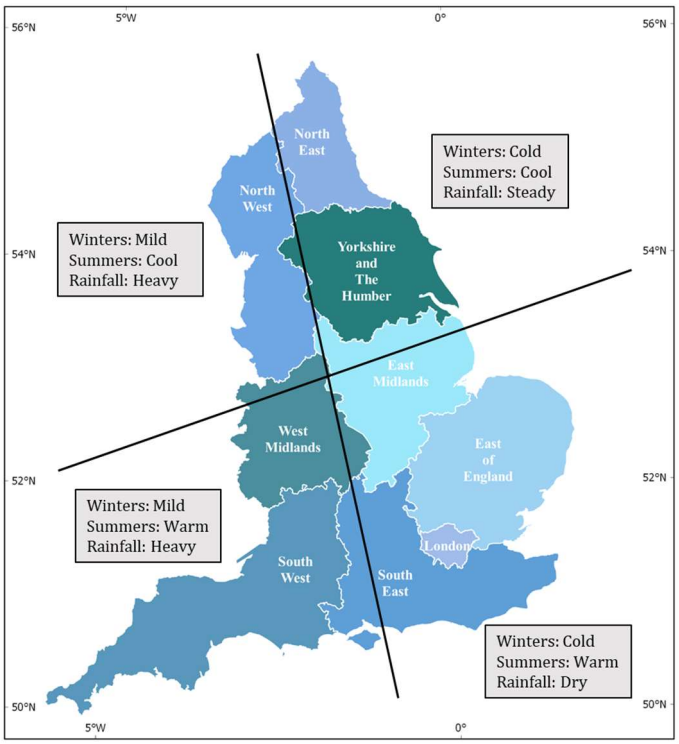

Figure 5.1: Regions of England with their generalised climatic conditions (Met Office 2020).

Data sources used in sub-national burden estimates have their limitations. Many studies have focussed on laboratory confirmed cases, but these data are highly influenced by hospital protocol, where hospitals will have different policies on whom to test for infection and when. In some cases, a patient will not be tested and diagnosis will be based on clinical findings. Using laboratory and hospitalisation data for estimating disease burden can also be skewed towards those who are most at risk of disease. This is particularly evident with RSV where premature birth, co-morbidities and immunosuppression can increase the risk of severe illness. Focusing on laboratory and hospitalisation data will limit the number of cases studied, which could lead to larger standard errors around estimates, increase the influence of anomalies in the data or bias the results to more severe cases. This makes it more difficult to detect seasonal or geographical trends. An alternative is the use of syndromic data, defined as non-specific pre-diagnostic-syndromic health data, which can be used as an indicator for disease and

is obtained from the surveillance of healthcare systems. By using a broader estimate of RSV and utilising data from healthcare systems more likely to be used earlier as the disease progresses, the spatial and temporal trends of its spread may be more likely to be observed.

Using syndromic indicators from two community healthcare services; NHS111 (a telehealth service) and GPIH (family doctors) this research aims to identify variations in RSV activity at regional level. Spatial-temporal analysis will be used to estimate the relationship between the syndromic data from NHS111 and GPIH services and laboratory confirmed cases of RSV, during the time period the week beginning the 30th June 2014 to 25th June 2018 in children under-1 year. Only data from children under-1 year were analysed in this research because, as observed in chapter four, RSV attributable burden was proportionally higher in this age group in the indicators used in this study. When analysis was conducted on the 1-4 year age group at a regional level, only small RSV signals were detected, making comparisons between regions difficult.

Here, we will build of the methods used in chapter four; by using the syndromic indicators that were most associated with RSV for NHS111, *Difficulty Breathing* and *Cough*, and GPIH, *lower respiratory tract infection* (*LRTI*). Using these indicators, we will produce a regional estimate of RSV burden using the methods described in chapter four; and this will allow us to investigate how burden and activity timing differs by region and if there is a latitudinal aspect to these differences. For this analysis we decided to run models that are able to capture the differences at regional level, rather than rely on using the proportion of each indicator associated with RSV for each region as estimated in chapter four because these proportions may be different by region. A further comparison between the methods and rationale of chapter four and this study is provided in Table 5.1.

Given the strong correlation with latitude and RSV activity reported in the literature, it might be expected that RSV activity in England would start in the south and move north. The regional variations of climate within England may also influence RSV activity. The observations from this research can be used for hypothesis generation of future research to investigate factors that influence variations in RSV activity. By focusing on data

sources from the community, we can detect cases before laboratory or hospital surveillance. Two services were included to allow us to assess consistency in the findings between systems. RSV activity will be assessed by looking at the regional burden, the start and peak of each seasonal outbreak and the length of outbreak.

The aims of this research are to:

1. Estimate the regional RSV attributable presentations to NHS111 and GPIH at Public Health England (PHE) centre level.
2. Detect the regional timing of RSV activity using the metrics; burden per 100,000, activity peak week, start week and activity length.
3. Compare the regional RSV burden and timing of RSV activity.

Study Title	Rationale	Aims	Metrics Estimated	Healthcare Systems	Indicators	Temporal Scale	Syndromic Spatial Scale	Laboratory Spatial Scale	Age Group
Estimating the Community Burden of Respiratory Syncytial Virus (RSV) in Children Under-Five Years, England.	Estimate community burden of RSV in children under-5 years in England to provide a better understanding of burden for targeted public health measures and to allow the evaluation of the potential benefits of a vaccine post-introduction.	1. Identify which syndromic indicators are associated with RSV. 2. Estimate the burden by age group 3. Compare burden to that of influenza. 4.Estimate the healthcare cost.	-Burden per 100,000. -Ratio to Influenza. -Cost to each healthcare service. -Comparison to number of laboratory detections.	NHS111, GPOOH, and GPIH	All respiratory indicators to NHS111 and GPIH	Week	National (England)	National (England)	Under-1 year and 1-4 years
Estimating Differences in Regional RSV Burden and Seasonality in Children Under-1 Year, England	Investigate if there are differences in the regional burden and timing of seasonality measures of RSV in the community in England in children under-1 year. RSV timing has been observed to have a strong latitude association; and this study aims to investigate if a difference in timings can be detected in England. This will provide a better understanding of RSV for targeted public health measures	1.Estimate the regional burden. 2. Estimate the regional seasonality of estimates. 3.Compare seasonality and burden between regions.	-Burden per 100,000. -Peak week of activity. -Start week of activity. -Length of activity.	NHS111, and GPIH	NHS111 – Difficulty Breathing + Cough; GPIH – Lower respiratory tract infection	Week	PHE Centre (nine areas in England)	National (England)	Under-1 year

Table 5.1: Comparison between chapter four and chapter five methods and rationale.

5.5 Methods

5.5.1 Data Sources

Epidemiological Data

Data was obtained from two healthcare systems monitored by PHE's real-time syndromic surveillance team (ReSST): NHS111 (a telephone helpline) and GPIH (general practitioner in-hours; a family doctor service).

Data comprised of weekly counts from 30th June 2014 to 25th June 2018, of syndromic indicator presentations that have previously been associated with RSV in children under-1 year. The syndromes included in the study were chosen based on the outcomes of chapter four, where the attributable burden of RSV for each indicator was estimated at a national level. In the case of NHS111, RSV was attributable to 20.8% and 21.7% of *Cough* and *Difficulty Breathing* presentations. The attributable proportion of RSV related presentations was similar between *Cough* and *Difficulty Breathing* indicators and only one indicator could be attributed to a healthcare-seeking contact at any time, therefore counts were summed to produce one set of counts from both indicators. *Lower Respiratory Tract Infection* (LRTI) was used to estimate RSV activity for the GPIH data, RSV was estimated to cause 19.0% of GPIH presentations coded as LRTI. Descriptive statistics of each indicator by each PHE centre are outlined in Table 5.2.

Count data was stratified by gender (male/female) and PHE centre, for both NHS111 and GPIH data. PHE centres are regional localities at which public health advice and support are provided from PHE. This geographical level consists of nine areas; with an average population of children under-1 year of 73,378 (range= 28,358-127,000).

System	PHE Centre	Total contact over study period (7/2014 to 6/2018)	Weekly Mean	Weekly Minimum	Weekly Maximum	Mean Weekly Population Over Study Period
NHS111 (Difficulty Breathing + Cough)	East Midlands	50,255	206.0	50	667	26,545
	East of England	44,872	183.9	34	703	28,195
	London	60,934	249.7	66	738	63,500
	North East	24,829	101.8	23	296	14,179
	North West	60,041	246.1	33	864	32,102
	South East	86,030	352.6	60	1,340	41,061
	South West	59,931	245.6	65	831	28,266
	West Midlands	59,293	243.0	52	766	34,330
	Yorkshire and the Humber	61,960	253.9	64	707	31,672
GPIH (LRTI)	East Midlands	10,118	40.6	8	152	8,683
	East of England	11,546	46.4	8	157	13,409
	London	31,752	127.5	20	470	31,442
	North East	7,798	31.3	3	117	5,855
	North West	43,970	176.6	21	594	23,194
	South East	31,383	126.0	15	485	21,986
	South West	15,191	61.0	5	260	11,710
	West Midlands	28,174	113.1	18	403	18,039
	Yorkshire and the Humber	11,726	47.1	11	152	12,870

Table 5.2: Summary statistics of syndromic data at each PHE Centre, in children under-1 year from 30th June 2014 to 25th June 2018.

Demographic Data

Population data used in the analysis of data from GPIH were based on the total number of registered patient's under-1 year at each GPIH practice monitored by syndromic surveillance at the PHE centre level, stratified by gender. The syndromic surveillance systems collect the daily number of registered patients, with the weekly number of registered patients calculated as the average number of patients for each week, accounting for the number of public holidays that occurred during that week. Population data for children under-1 year in England for NHS111 analysis, were obtained at the upper tier local authority (UTLA) level from mid-year estimates from the Office for National Statistics (ONS) (ONS 2014), and were stratified by gender. For the analysis,

population data was summed to PHE centre level. Population data was obtained at the UTLA level to account for areas with poor or no coverage by the surveillance system.

Estimating Periods of Poor Coverage

As described in chapter four, change point analysis was used to detect periods of poor coverage UTLA. This analysis was used to detect changes in the mean and variance of NHS111 syndromic data. Periods of poor coverage are due to periods were the surveillance system or healthcare systems had not been introduced and periods where data was not received by the surveillance systems. Weekly changes due to poor coverage were identified at the UTLA level, and consequently excluded from the analysis. Population data from ONS of corresponding UTLAs with poor coverage were also excluded, this allowed us to estimate the underlying study population. Syndromic and population data were summed to PHE centre level for analysis. As data on the underlying study population, the number of registered patients, was obtained through the GPIH surveillance system, we did not need to estimate periods of poor coverage this dataset. A full description of the methods used to estimate areas of poor coverage is provided in chapter four.

Pathogen Data

Weekly counts of laboratory confirmed cases of 10 seasonal respiratory pathogens (Table 5.3) were obtained from the PHE Second Generation Surveillance System (SGSS). This data was stratified by gender. Pathogen data was based on specimen collection date.

Pathogen	Total	Weekly Mean	Maximum in a Single Week
Respiratory Syncytial Virus	7,440	46.21	336.0
Rhinovirus	3,812	23.68	47.0
Parainfluenza	1,070	6.65	22.0
Influenza A	414	2.57	21
Human Metapneumovirus	530	3.29	19.0
Influenza B	141	0.88	13.0
Coronavirus	283	1.76	15.0
Streptococcus Pneumonia	166	1.03	5.0
Haemophilus Influenzae	48	0.30	3.0
Mycoplasma Pneumonia	37	0.23	4.0

Table 5.3: Summary statistics of laboratory confirmed pathogens in children under-1 year in England from 30th June 2014 to 25th June 2018

5.5.2 Statistical Analysis

A Bayesian Approach

There are several motivations for using a Bayesian approach in this study. Bayesian inference assigns probabilities as a measure of confidence in the occurrence of an event, whereas in the frequentist approach probabilities are based on the long-run frequency of events when experiments are repeated. These experiments are not repeated and therefore the bases of the frequentist approach are not based on reality. Fundamentally, Bayesian and frequentist models have different ways of fitting the data. Bayesian models fit the model to the data, whereas the frequentist approach fits the data to the model, therefore the Bayesian approach is seen as more data driven. With the frequentist approach there is a reliance on *p*-values and confidence intervals. The *p*-value, the probability of obtaining the observed test statistics as extreme as the under a null hypothesis, is of particular issue due to its common misuse or misinterpretation (Kim and Bang 2016). With Bayesian approaches prior knowledge of the data can be incorporated into analysis, although this has resulted in criticism of this approach as this prior knowledge can lead to models no longer being objective. Finally, data with

complicated spatial and temporal dependency structures, whereby observations are not independent and are related in systematically over space and/or time, can be hard to model using traditionally frequentist approaches and packages in R that can handle this data are limited. The primary motivation for using a Bayesian approach to the analysis in this study was due to the spatial and temporal structures of the syndromic data.

Integrated Nested Laplace Approximations (INLA)

The development of the Markov chain Monte Carlo (MCMC) methods for probability distribution sampling allowed for the development of Bayesian approaches to handle data with spatial, temporal and spatial-temporal dependencies (Metropolis et al. 1953; Hastings 1970; Gelfand and Smith 1990). Although MCMC has allowed for the growth of Bayesian statistics in spatial and spatial temporal modelling, this method has a high computational burden, with large or complex datasets taking several days to undertake Bayesian inference. The integrated nested Laplace approximations (INLA) for probability distribution sampling, developed by Rue, Martino and Chopin (2009), is capable of fast and accurate Bayesian inference, overcoming some of the computational limitations of the MCMC method. The Bayesian inference using the INLA method can be implemented using the R package *R-INLA* (Martins et al. 2012). Here we describe the use of a Bayesian approach, using INLA, to model data with a temporal dependency at multiple levels (regions).

Model Development

GPIH and NHS111 data was analysed at the PHE centre level. Given these regions are considerably large with a mean of 14,772 km^2 (range: 1,594-24,392 km^2), a mixed effects model was constructed with a temporal dependency without consideration of potential spatial dependencies.

Inclusion of the laboratory data at regional level (PHE centre) was explored, with analysis run using both national and regional laboratory counts. Due to the small number of positive cases of the pathogens analysed in certain areas, the regional laboratory did not detect RSV trends as well as data at the national level. This was most likely due to differences in laboratory testing policy at regional levels, rather than

differences in pathogen burdens. We calculated leads between the national laboratory data and the regional syndromic data to account for the syndromic data detecting pathogen activity before the laboratory data. For every combination of leads between 0 and 2 weeks for each pathogen and each region, models were run to estimate the combination of leads that fit the syndromic data best, based on deviance explained. The combination of leads found to have the highest deviance explained for NHS111 and GPIH data is described are Table 5.4.

	Pathogen									
System/PHE Centre	RSV	Influenza A	Influenza B	Coronavirus	Rhinovirus	Parainfluenza	*Streptococcus pneumoniae*	*Haemophilus influenzae*	*Mycoplasma pneumoniae*	HMPV
NHS111										
East Midlands	1	2	1	2	0	1	2	2	0	1
East of England	1	1	0	0	0	1	1	2	1	2
London	1	0	1	0	2	1	1	0	0	0
North East	0	0	2	2	2	2	2	2	0	1
North West	0	1	0	0	1	0	1	0	0	1
South East	1	0	1	1	0	1	1	0	1	1
South West	1	1	0	0	0	1	2	0	2	2
West Midlands	1	0	1	2	0	1	1	0	0	0
Yorkshire and the Humber	1	2	0	0	2	1	2	1	0	0
GPIH										
East Midlands	1	0	1	2	0	2	0	1	1	0
East of England	0	1	2	2	2	2	2	2	2	2
London	0	2	0	0	2	2	1	2	0	0
North East	0	2	1	0	2	2	1	0	0	2
North West	0	2	0	0	2	2	1	2	2	2
South East	0	0	2	2	2	2	2	0	1	0
South West	0	0	1	2	2	1	0	0	2	0
West Midlands	0	0	2	0	2	2	0	0	2	2
Yorkshire and the Humber	0	0	2	1	2	1	1	1	2	0

Table 5.4: Number of week-leads between the syndromic data and each pathogen.

Mixed effect models with a Poisson specification were constructed with PHE centre treated as a random effect to account for potential pseudoreplication, where by the are multiple observations of some independent variables from the same location due to the repeated observations by gender, and to account for regional differences that may influence presentations. An autoregressive effect of time (each week of the study period) was applied for each PHE centre, to account for temporal autocorrelation. Annual and weekly variations were accounted for through the inclusion of random effects for year and week. An interaction between RSV and PHE centre was included in the model to allow for varying RSV trends by PHE centre. The nine other respiratory pathogens, public holidays, and gender were also accounted for.

The general algebraic definition of the models is given by:

$$Y_{g,i,t} \mid \mu_{g,i,t}, \sim F(\mu_{g,i,t}),$$

Where $\mu_{g,i,t}$ is the is the gender g, specific number of contacts at each geographical location (PHE centre) i at time t, F is the Poisson likelihood of the model.

A logarithmic link function of the expected number of cases is modelled as:

$$\log(\mu_{g,i,t}) = \alpha + \log(P_{g,i,t}) + \rho R_{i,t-1} + \tau H_t + \theta H_{t-1} + \sum_k \beta_k X_{t,k} + u_w + \nu_y + \zeta g \cdot \gamma_i \cdot \chi_r$$

Where α corresponds to the intercept; $\log(P_{g,i,t})$ denotes the logarithm of the population at risk for gender g and geographical location i at time t included as an offset to adjust counts by population; ρ is an auto-regressive coefficient of the residuals R lagged one week $(t - 1)$ to account for potential temporal autocorrelation and to account for different time trends by location i; H denotes a Boolean variable indicating the presence of public holidays on the current (t) or previous week $(t - 1)$ with coefficients τ or θ accordingly; X is a matrix of k infectious disease co-variates with regression coefficients β. Unstructured random effects of week (u_w) and year (v_y) were included to account for seasonal trends that maybe related to factors other than RSV such as temperature and rainfall. $\zeta g \cdot \gamma_i \cdot \chi_r$ denotes the interaction between gender, geographic location and RSV

to account for different RSV trends by gender and region. Models were fitted in R version 3.5.2 (R Core Team 2018) using the *R-INLA* package (Martins et al. 2012).

Model Validation

To assess the predictive ability of the model and identify overfitting block cross-validation was used (Roberts et al. 2017). Each dataset was split into 10 equal sized blocks of continuous weeks. The model was trained and tested in linear time. The same method as described in chapter four was used for cross validation of the models used in this chapter.

RSV Estimation and Prediction

Using the methods described in chapter four, RSV was estimated by predicting the number of contacts to NHS11 and GPIH when RSV was in circulation compared to the number with zero RSV. The attributable burden of RSV was calculated as the difference between the predictions where RSV was present and those where RSV was not present. These predictions were calculated as rates per 100,000 residents for NHS111 and as rate per 100,000 registered population for GPIH predictions for every week over the study period.

Measures of RSV Activity

Using the estimated number of presentations attributable to RSV for each service, RSV activity and seasonality for each region was estimated. Four metrics for RSV activity were calculated to investigate possible regional differences; seasonal burden, peak week, activity start week, and activity length, a summary of these metrics is provided in Table 5.5. Seasonal burden was calculated as the regional number of presentations per 100,000 children under-1 year during the influenza season (October to April), as this is when most of the RSV activity occurs in the UK, for each year included in the study. Peak week was calculated as the week at which the rate of RSV was highest for each region. Activity start week and length were estimated from the identification of epidemic thresholds using the moving epidemic method (MEM) (Vega et al. 2015), which is used to approximate epidemic and intensity thresholds of seasonal respiratory

infections. This method was developed to further understand the intensity of influenza activity, but has previously been used to estimate thresholds for RSV activity (Harcourt et al. 2019; Vos et al. 2019). For each region, RSV activity thresholds were identified, using MEM, for each of the four seasons included in this study. MEM was applied using the MEM web application (Lozano 2018). Activity start week was estimated from the first week that RSV activity was above pre-epidemic levels, and activity length was calculated as the number of weeks at which RSV activity was at above the pre-epidemic threshold. All 4 metrics were then averaged to produce seasonal means over the study period.

RSV Activity	Definition	Method for Calculation
Burden per 100,000	Rate of indicator presentations attributable to RSV per 100,000 during influenza season for each region.	Predictions of indicators made where RSV was present and absent. To estimate weekly RSV burden, the difference between these predictions was taken. This burden estimate was summarised to influenza season and a rate per 100,000 was calculated based on the regional populations for each surveillance systems.
Peak Week	Week number at which rate of RSV was highest for each region.	From burden estimates the rate was plotted as a timeseries, and the week at which rate was highest for each year was estimated.
Start Week	First week number at which RSV rate was above the epidemic threshold for each region.	Epidemic thresholds identified using the moving epidemic method of burden estimates (Vega et al. 2015).
Activity Length	Length of time, in weeks, at which the rate of RSV was above the epidemic threshold for each region.	Epidemic thresholds identified using the moving epidemic method of burden estimates (Vega et al. 2015).

Table 5.5: RSV activity metrics, definition and method for calculation.

5.6 Results

5.6.1 Population Coverage During Study Period

Prior to modelling the data from the two healthcare services it was essential to understand the denominator population. This can vary because, due to the passive nature of the surveillance systems monitoring these healthcare services, there are periods where data transfer does not occur, resulting in variations in the underlying study population. The denominator population for the GPIH data is recorded through the surveillance system as the number of registered patients to the GPIH services submitting data, and therefore takes account of varying underlying population and data supply issues. For the NHS111 surveillance system this was estimated using change point analysis at the UTLA level. Over the study period there were variations in population coverage of this surveillance system due the system not being introduced in some areas during the early study period and data transfer issues. UTLA areas that were identified as having poor coverage were excluded from the analysis, with corresponding population estimates from ONS and were then summarised at the PHE centre level.

The PHE centre population coverage of both surveillance systems over the study period can be seen in Figure 5.2, and the national figures are described in Table 5.6. GPIH had a much lower level of coverage over the study period with 44.6% of the population monitored by this surveillance system, compared to NHS111 which had a total coverage of 92.3% (Table 5.6). Over time, NHS111 had increasing population coverage, going from 83.6% in the first year of the study to 95.0% in the final year. Whereas, population coverage decreased from 47.0% to 36.0% of the population over the study period for the GPIH surveillance system (Table 5.6).

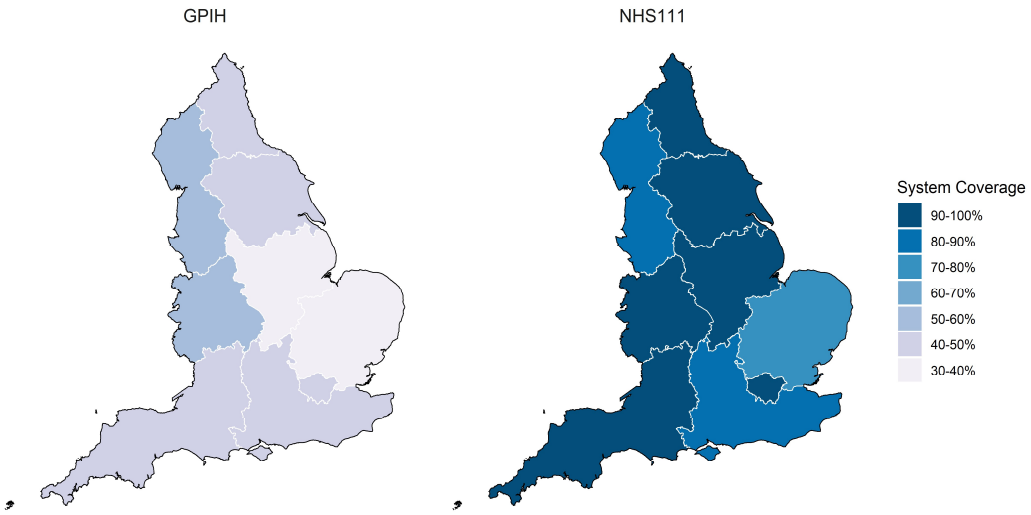

Figure 5.2: Average surveillance coverage of the two systems monitored by syndromic surveillance; GPIH and NHS111 coverage is the estimated total population monitored by the syndromic surveillance systems from 30th June 2014 to 25th June 2018.

	Year				
System	2014/15	2015/16	2016/17	2017/18	Total Study Period
NHS111	83.6	95.9	94.7	95.0	92.3
GPIH	47.0	48.4	46.7	36.0	44.6

Table 5.6: Percentage coverage (%) of the two systems monitored by syndromic surveillance system by year included in the study.

5.6.2 Model Development and Validation

Both NHS111 and GPIH datasets were modelled using the Poisson distribution, as the data was under-dispersed (dispersion statistic <1) (Table 5.7). The predictive ability of the models was evaluated using mean absolute error (MAE), which is an estimate of the difference between observed and predicted values, with smaller values indicative of a smaller difference. MAE values for NHS111 and GPIH were estimated using block cross-validation. There were small discrepancies between the observed and fitted values with an average MAE value of 7.5 presentations per fold for the NHS111 model and 8.43 presentations per fold for the GPIH model. These values are small compared to the mean number of *Cough* and *Difficulty Breathing* presentations to NHS111 of 115.7 and *LRTI* presentations to GPIH of 43.1 over the study period.

Model	Dependent Variable	DIC	Dispersion	Mean number of contacts	MAE (number of contacts)
NHS111	Cough + Difficulty Breathing	33891.8	0.78	115.7	7.45
GPIH	Lower Respiratory Tract Infection	28500.2	0.83	43.10	8.43

Table 5.7: Model metrics of NHS111 and GPIH models.

5.6.3 Seasonal Burden

For each influenza season included in the study, burden maps (burden per 100,000 residents) were created to identify and compare the regional distribution of RSV across England for NHS111 and GPIH over multiple seasons. The average burden per 100,000 residents for the total study period was also mapped to identify the overall regional distribution of RSV (Figure 5.3). Over the four-year study period there were clear patterns of areas with consistently higher and lower burdens of RSV attributable presentations to both NHS111 and GPIH. For GPIH there was an east-west divide, with higher burdens observed in the west coast compared to the east coast, where the burden was consistently low. High burdens were also observed in the southern and northern regions. Over the four-year study period the southern regions had a consistently higher

burden compared to the other regions. Although there was some level of consistency over the study period, NHS111 did not show the same level of consistency as the GPIH data. Higher burdens were observed in the southern regions, the North East and Yorkshire and the Humber, consistently during the study period. London had the lowest burden during the study period. There were some similarities between the two systems, with the southern regions and the North East consistently showing the highest burden in both healthcare services.

Table 5.8 describes the actual burden for each year and region during the study period, and a total for the study period. From this table it can be observed that the same regions are consistent in where they have high or low burdens of RSV, especially observations from GPIH. RSV attributable presentations to GPIH were consistently highest in the South East and South West, and lowest in the East Midlands, East of England and Yorkshire and the Humber. There was slightly more variation in which region had the highest burden of RSV attributable presentations to NHS111, with the South West, South East and North East having the highest burdens of the study period; London consistently had the lowest burden.

	Year				
System	2014/15	2015/16	2016/17	2017/18	Mean for All Study Period
NHS111					
East Midlands	947	1,285	1,146	1,278	1,164
East of England	902	1,254	1,252	1,364	1,168
London	445	608	558	592	551
North East	1,092	1,466	1,366	1,928	1,462
North West	812	1,279	1,024	1,115	1,098
South East	1,127	1,525	1,490	1,373	1,410
South West	1,264	1,742	1,437	1,518	1,491
West Midlands	944	1,246	1,082	1,226	1,126
Yorkshire and the Humber	1,116	1,409	1,229	1,407	1,290
GPIH					
East Midlands	921	840	740	711	808
East of England	828	786	681	602	726
London	1,459	1,387	1,355	982	1,285
North East	1,809	1,624	1,598	1,276	1,566
North West	1,757	1,756	1,700	1,232	1,600
South East	2,167	2,037	1,970	1,448	1,895
South West	2,583	2,374	2,361	1,643	2,235
West Midlands	1,773	1,651	1,616	1,235	1,550
Yorkshire and the Humber	644	608	520	468	562

Table 5.8: Estimated burden per 100,000 residents of RSV attributable presentations to NHS111 and GPIH. Colour depicts whether burden was high or low for that year (red = high, green = low), and was assigned based on the ranking of each region for each year. NHS111 and GPIH were ranked separately.

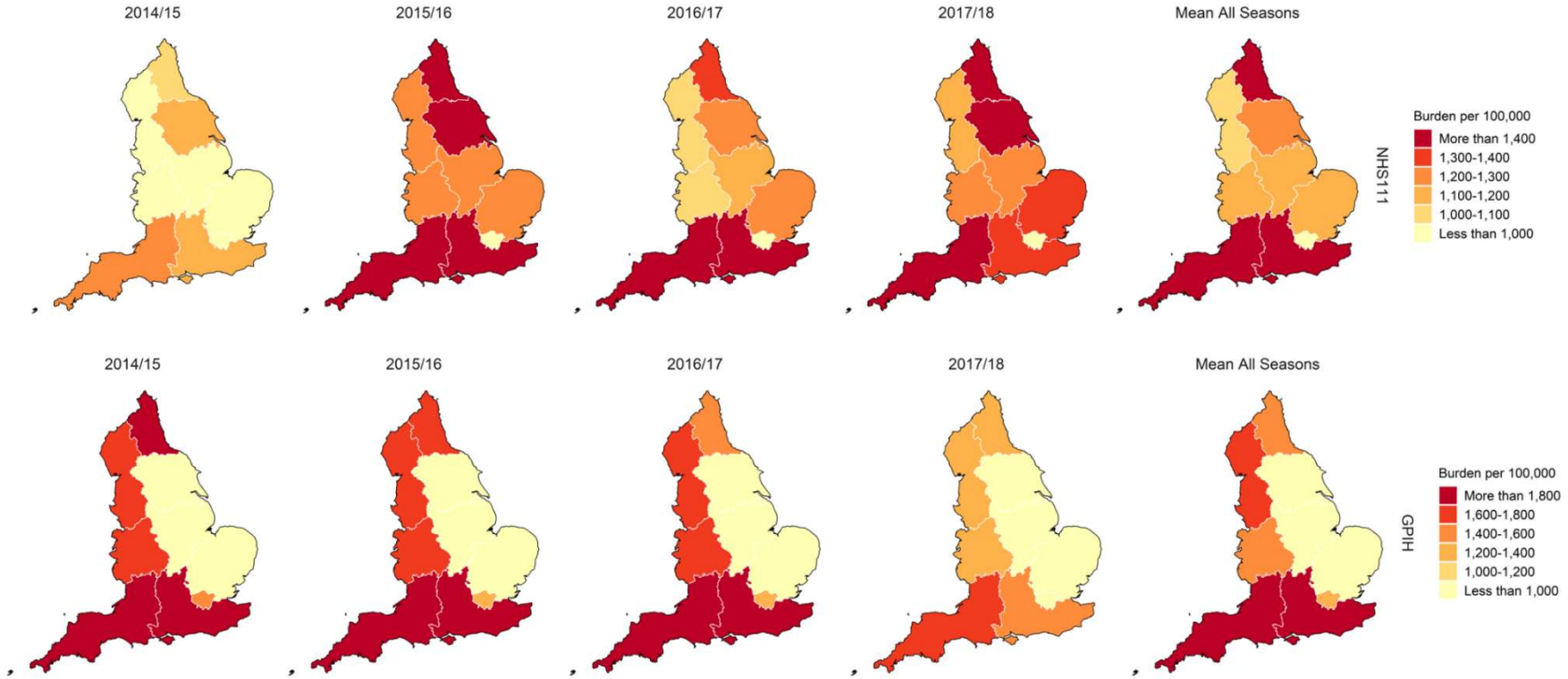

Figure 5.3: Estimated burden per 100,000 residents of RSV attributable presentations to NHS111 (top) and GPIH (bottom). Maps presented for each year of study and a mean for the total study period.

5.6.4 Activity Peak Week

Peak week, the week number at which the rate of RSV attributable presentations was highest, for each season included in the study was mapped by region to identify if there were spatial patterns of peak RSV activity (Figure 5.4). For both NHS111 and GPIH there are no clear spatial patterns of when activity was at its highest between the season, with very little variation between regions in when the peak occurred. When looking at the average peak week across all years, there was little variation in RSV attributable presentations to GPIH, with the average peak week between week 50 and 51. There was slightly more variation in the peak week of activity of RSV attributable presentations to NHS111, with the peak slightly later in the South West and Northern regions.

Table 5.9 describes the actual peak week for each season, and the average over the study period. From this you can see there is very little variation in peak week by season, this is especially true for GPIH where peak week only varied by half a week. There are no consistent spatial patterns in peak for RSV activity in GPIH, whereas some consistencies can be observed from NHS111 peak week with North West consistently later, and the West Midlands earlier compared to the other regions. Average peak week was comparable between the two services.

	Year				
System	2014/15	2015/16	2016/17	2017/18	Mean for All Study Period
NHS111					
East Midlands	51	49	48	49	49.3
East of England	50	49	48	49	49.0
London	50	49	49	49	49.3
North East	51	50	49	49	49.8
North West	52	50	49	49	50.0
South East	50	49	48	48	48.8
South West	51	50	48	49	49.5
West Midlands	49	48	48	48	48.3
Yorkshire and the Humber	50	49	48	49	49.0
GPIH					
East Midlands	51	50	48	50	49.8
East of England	51	50	49	50	50.0
London	50	50	49	49	49.5
North East	51	51	49	49	50.0
North West	51	50	49	49	49.8
South East	51	50	49	49	49.8
South West	51	51	49	49	50.0
West Midlands	51	50	49	49	49.8
Yorkshire and the Humber	51	50	49	50	50.0

Table 5.9: Estimated peak week of RSV attributable presentations to NHS111 and GPIH. Colour depicts whether peak week was earlier or later in the year (red = later, green = earlier), and was assigned based on the ranking of each region for each year. NHS111 and GPIH were ranked separately.

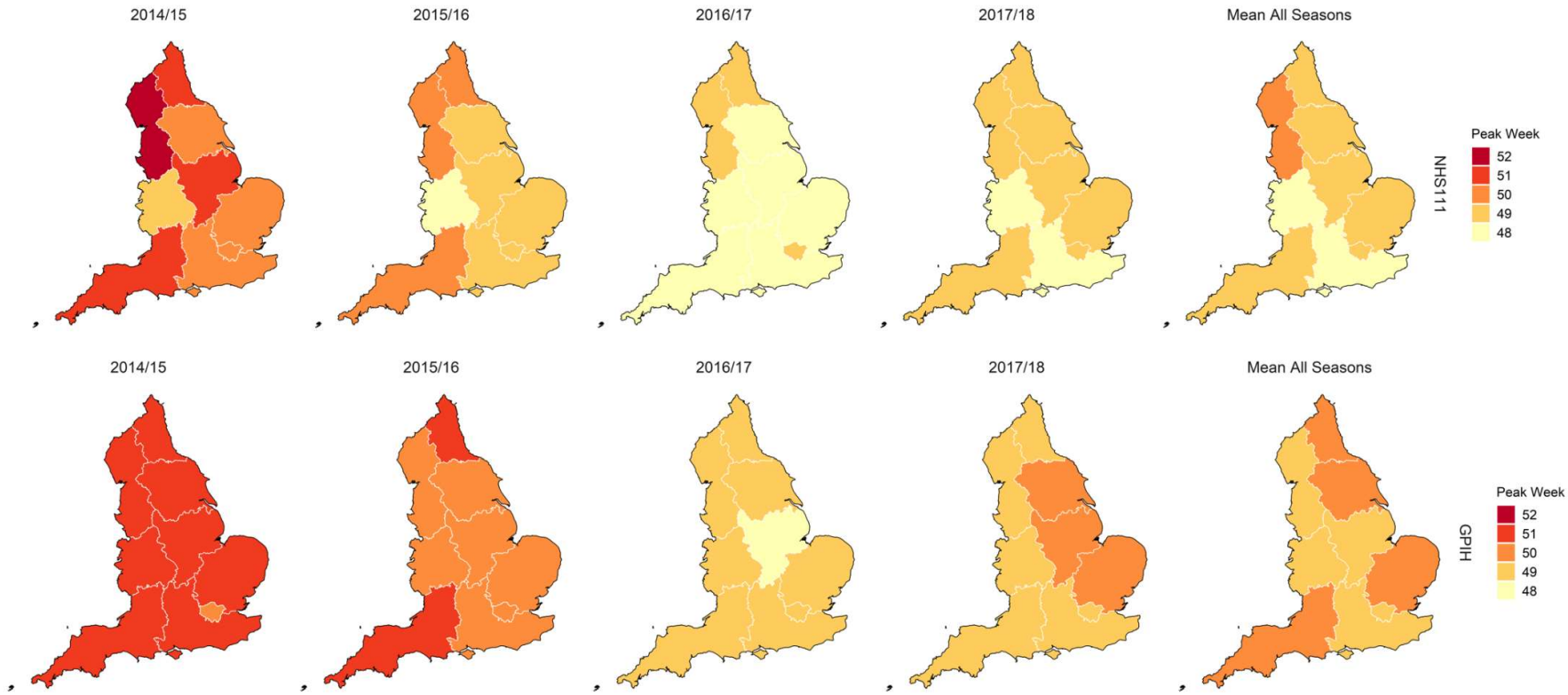

Figure 5.4: Estimated peak week of RSV attributable presentations to NHS111 (top) and GPIH (bottom). Maps presented for each year of study and a mean for the total study period.

5.6.5 Activity Start Week

Start week, the week at which RSV activity was estimated to be above pre-epidemic levels, was mapped to identify spatial patterns at each region and season (**Error! Reference source not found.** 5.5). In NHS111 services, RSV activity appeared to start at the same time for most regions during the time period, around week 43, apart from in the North East where activity started later for most years and on average was two weeks after the rest of England. Start week was more varied in GPIH services, with activity starting later in the southern regions and the North East. There was little consistency between the two services, although in the first year of the study, 2014/15, RSV activity appeared to start later than average for all regions, this finding was observed from both NHS111 and GPIH services.

Table 5.10 describes the activity start week for each year and region, as well as the average over the study period. From this table we can observe that there is some consistency in timing in the North East, where start week is later compared to the other regions in both GPIH and NHS111 systems. For the most part activity start week in all regions tends to be the same, with activity starting earlier in NHS111 services, compared to GPIH.

	Year				
System	2014/15	2015/16	2016/17	2017/18	Mean for All Study Period
NHS111					
East Midlands	46	43	43	43	43
East of England	45	43	44	43	43
London	45	43	43	43	43
North East	47	45	43	44	45
North West	47	44	43	43	43
South East	46	43	43	43	43
South West	46	43	43	43	43
West Midlands	46	43	43	43	43
Yorkshire and the Humber	46	43	43	43	43
GPIH					
East Midlands	46	43	43	43	43
East of England	46	43	44	44	44
London	46	44	44	44	44
North East	47	45	44	44	45
North West	46	44	43	43	43
South East	46	45	44	44	45
South West	46	45	44	44	45
West Midlands	47	44	43	44	44
Yorkshire and the Humber	46	44	43	44	44

Table 5.10: Estimated start week of epidemic activity from RSV attributable presentations to NHS111 and GPIH. Colour depicts whether start week was earlier or later in the year (red = later, green = earlier), and was assigned based on the ranking of each region for each year. NHS111 and GPIH were ranked separately.

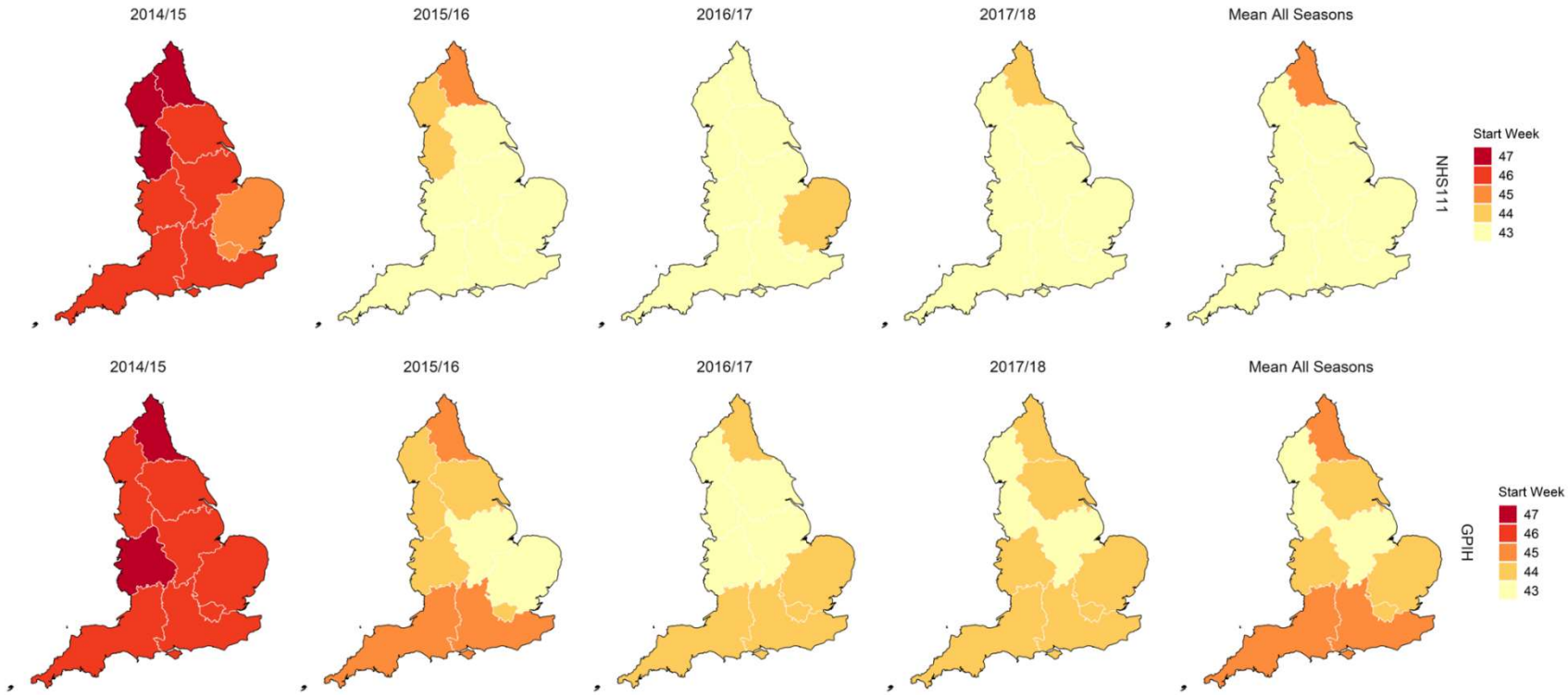

Figure 5.5: Estimated start week of epidemic activity from RSV attributable presentations to NHS111 (top) and GPIH (bottom). Maps presented for each year of study and a mean for the total study period.

5.6.6 Activity Length

Activity length, the number of weeks at which RSV activity was above the pre-epidemic level, is mapped in Figure 5.6 to illustrate the spatial and yearly patterns of activity length. There does not appear to be any consistent spatial patterns in activity length, for both NHS111 and GPIH. It does appear that there are similar lengths between seasons for each region and this can be observed in both NHS111 and GPIH.

From Table 5.11 it can be observed that there is very little variation in outbreak length by year and between the two services, with the majority of outbreaks occurring for 8-11 weeks. There is consistency of activity length between years, with similar lengths observed between NHS111 and GPIH for each year.

	Year				
System	2014/15	2015/16	2016/17	2017/18	Mean for All Study Period
NHS111					
East Midlands	8	9	11	11	10.0
East of England	9	9	10	10	9.5
London	9	9	11	10	9.5
North East	8	8	11	10	9.0
North West	8	9	11	11	10.0
South East	8	9	11	10	9.5
South West	8	9	11	10	9.5
West Midlands	8	9	11	10	9.5
Yorkshire and the Humber	8	9	11	11	10.0
GPIH					
East Midlands	9	9	11	10	9.5
East of England	9	9	10	10	9.5
London	9	9	10	10	9.5
North East	8	8	10	10	9.0
North West	9	9	11	11	10.0
South East	9	8	10	10	9.5
South West	9	8	10	10	9.5
West Midlands	8	9	11	10	9.5
Yorkshire and the Humber	9	9	11	10	9.5

Table 5.11: Estimated outbreak length of epidemic activity from RSV attributable presentations to NHS111 and GPIH. Colour depicts whether outbreak length was shorter or longer in the year (red = longer, green = shorter), and was assigned based on the ranking of each region for each year. NHS111 and GPIH were ranked separately.

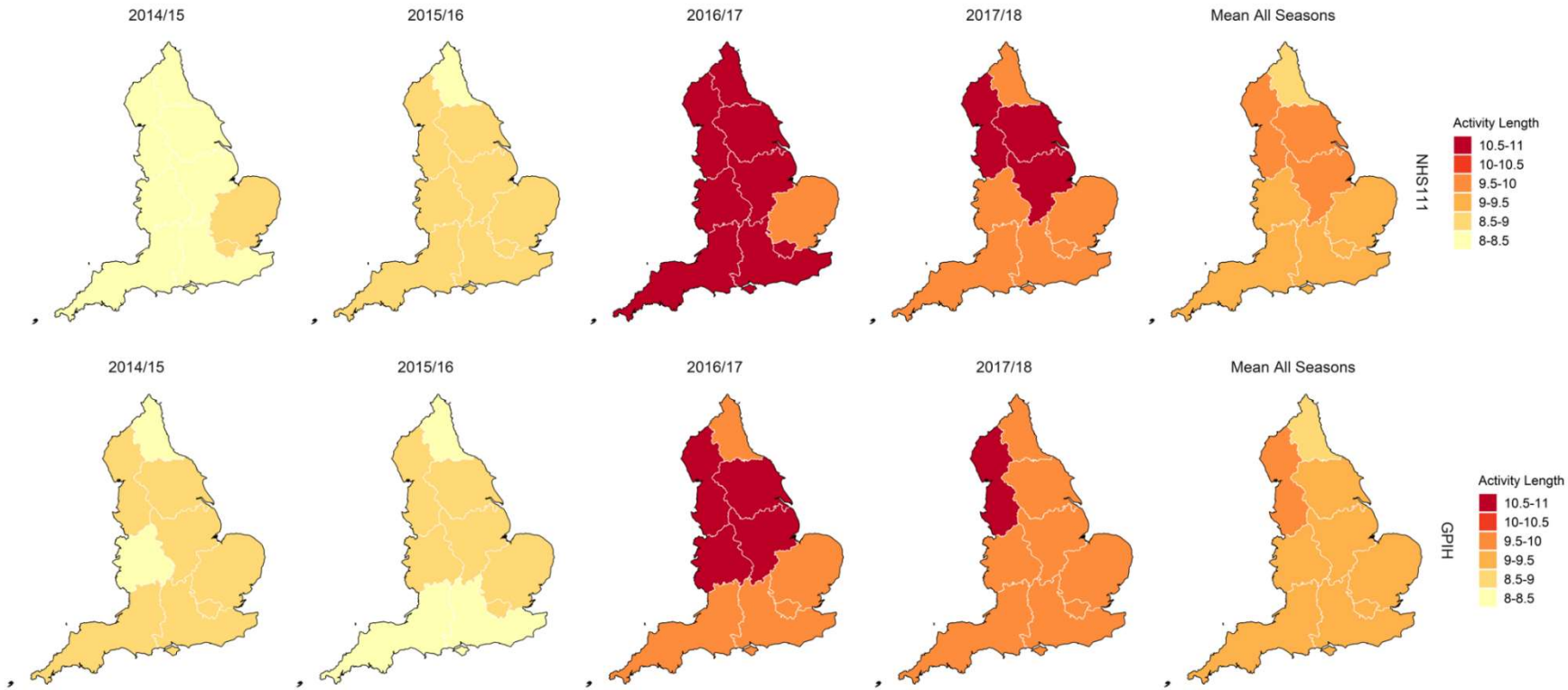

Figure 5.6: Estimated epidemic activity length of epidemic activity from RSV attributable presentations to NHS111 (top) and GPIH (bottom). Maps presented for each year of study and a mean for the total study period.

5.7 Discussion

5.7.1 Impact

Here, we present an analysis of the regional burden and timing of RSV activity in children under-1 year in England from two community healthcare services, using a Bayesian approach. This is the first study to investigate the within country variation of RSV activity timing in England. The two services included in the study, NHS111 and GPIH, are based in the community, therefore provide a larger number of cases for analysis compared to traditional data sources (hospitalisations and laboratory cases) and reflects the community burden of disease, rather than the more severe cases captured by traditional sources. Using two healthcare services has allowed us to compare and contrast the finding from this study, adding further robustness. The surveillance of NHS111 and GPIH was estimated to monitor around 92.3% and 44.6% of the population over the study period. We used four metrics in this study to look at the timing and spatial differences of RSV activity; burden per 100,000, peak week, start week and outbreak length. This has allowed us to fully investing the possible regional differences in RSV activity. The methods and findings used in this study can allow policy-makers to make targeted decisions on where resources are most effective in order to improve healthcare outcomes. This research has laid the framework for further research on spatial variations in RSV activity or of other pathogens.

5.7.2 Main Findings

There appeared to be consistent patterns in regional variation of RSV burden, with, on average, a higher rate of presentations in the southern regions, this finding was observed in both the GPIH and NHS111 services. Over the four-year study period there were very constant patterns from the GPIH services, with more presentations in the south and a clear east-west divide in presentation rate. Although, more presentations to NHS111 were observed in the south, there was not as much consistency in the other regions compared in presentations to GPIH. This inconstancy in NHS111 presentations may be due genuine difference in healthcare-seeking behaviours or the introduction of NHS111

as a new healthcare service. A study by Lewis, Stavola and Hardelid (2020) investigated the spatial and seasonal variation in hospitalisations due to bronchiolitis, a common complication of RSV, in children under-1 year in England. Similar to our observations, there was a large spatial variation in rate of admission with London having the lowest rate and the North West of England had the highest. This is similar to our observations with London having the one of lowest rates per 100,000 presentations attributable to RSV for both NHS111 and GPIH. Although we observed the southern regions (South East and South West) as having the highest rate per 100,000, high rates were also observed in the northern regions (North East and North West) in our study. Higher rates of hospitalisation were associated with a higher deprivation score, and a higher log of population density was associated with a lower rate of admission. The Lewis, Stavola and Hardelid (2020) study was conducted at a lower level of geography, clinical commissioning group, which allowed the authors to explore the effects of deprivation and population density on rate of hospital admission. Due to the large area of the geography used in our analysis, PHE centre, population density and deprivation data was too aggregated to be explored. Therefore, further research is needed to explore these factors using our data.

It is not clear why higher presentation rates were observed in southern regions, from the observations of other metrics of RSV activity there was no evidence to suggest that timing of activity is different in these regions. In the literature, lower population density and higher deprivation has been associated with a higher rate of hospitalisations due to bronchiolitis in England (Lewis, Stavola and Hardelid 2020). This is the opposite associated as would be expect through our observations, as the southern regions of England tend to have a higher population density and lower levels of deprivation. The Lewis study focused on hospitalisations, whereas our study used presentations from the community, which may explain differences in the observations. Environmental factors, such a meteorological or social condition, may also contribute to the higher burden of RSV in the south. In temperate regions with low humidity and low temperatures in winter, RSV has been observed to survive better in both aerosols and on surfaces (Yusuf et al. 2007; Paynter 2015). Other environmental factors such as deprivation (Foley et al. 2019) and household crowding/siblings (Simoes 2003) have been associated with

severity of RSV infection. Further research needs to be conducted to investigate whether these factors contribute to differences in RSV burden across England.

Mean peak week of RSV activity was observed to vary by about two weeks for NHS111 presentations and one week for GPIH presentations. For both healthcare services, RSV activity mean start week was observed to vary by about two weeks and mean activity length had mean regional variations of a week. But for these three metrics there was little distinctive evidence of consistent regional patterns of RSV activity timing. From the literature, peak week of bronchiolitis hospitalisations in children under-1 year in England was also observed to only vary, on average, by a week; with activity peaking first in London and last in the North East (Lewis, Stavola and Hardelid 2020). We observed earlier peak activity timings for both NHS111 and GPIH compared to those observed in hospitalisations, but this will be due to the timeliness of outbreak detection when using syndromic data in comparison to hospitalisation or laboratory data (Ziemann et al. 2016). For NHS111 peak week was observed to vary by 1.7 weeks between the first peak, on average, in the West Midland and the last peak, on average, in the North West. There was much less variation in peak timing for the GPIH data, with peak week only varying by 0.5 weeks. Although our findings differ from those from Lewis, Stavola and Hardelid (2020) in terms of where the variation in peak timing occurred, there was agreement in how little variation in average peak week there was. Differences observed in the geographical patterns could be due to the different time period used. Lewis, Stavola and Hardelid (2020), observed that earlier peak week was associated with areas with higher population densities, and both high and low deprivation scores. Lewis, Stavola and Hardelid (2020) observations are supported by an American study that investigated the local variation in peak timing of RSV hospitalisations in children under-2 year; earlier peak timing was associated with higher population density (Noveroske et al. 2016). Average Peak timing was first in the West Midlands, an area of high population density, and last in the north, where there are areas of low population density for the NHS111 data. From the GPIH there was too little variation in the peak timing to make conclusions regarding variations in peak timing. But these observations require further investigation using smaller geographical areas to make more conclusive observations.

From the literature, it would be expected that in a temperature country like England, that RSV activity would start and peak earlier in the south of the country and later in the North, as RSV start and peak has been observed to follow a northwards movement in temperate climates (Broberg et al. 2018). RSV activity length had also been observed to last longer in more northernly locations (Broberg et al. 2018; Bloom-Feshbach et al. 2013). These trends cannot conclusively be observed from our data, with the start week of epidemic activity where start week was observed to be later, on average, in the south compared to other regions in the GPIH data, and no difference, on average, between regions was observed in the NHS111 data. A European study explored the geographical variations in RSV activity timing (peak week, start week and length) in 15 European countries, including the UK and countries of similar latitude: Germany, Poland, Netherlands, Ireland and Denmark. Between these countries median start week varied from week 44 to week two (10 weeks), median peak week varied from week 50 to week 6 (eight weeks) and activity length varied from 15 to 20 weeks (five weeks). Typically, the UK was observed to have an earlier start and peak week compared to these other countries. The UK, Netherlands and Ireland also showed little variation in timings, and these countries are the most similar in terms of size, and latitudinal range (Obando-Pacheco et al. 2018). Although the European study observed an association between latitude and RSV activity timing these observations could not be replicated in our data at a region level within England, this could be because there is not enough latitudinal variation within England or there may be other factors that explain differences in RSV activity. This European study used data from the UK rather than just England which could influence these observations. This European study also primality used surveillance of acute respiratory infections (ARI) and influenza like illness (ILI) which could pick up other infections in circulation at the same time as RSV.

This work provides a framework for further exploration of regional timing and burden of RSV activity in England when data is available for a longer time period and at a finer level of geography. This will provide a much more definite answer to regional differences of RSV activity. From the literature there is evidence of a latitudinal effect on RSV timing in temperate regions, which we could not replicate using data from England. However, some potential regional differences were observed but there was no

spatial pattern to these differences; further research is required to explore the possible other reasons (such as deprivation, population density and meteorological conditions) that may explain some regional differences in RSV activity.

5.7.3 Limitations

The syndromic data obtained for this research came from the real-time passive surveillance of NHS111 and GPIH. Consequently, there were periods of no or incomplete data recording. For GPIH the underlying population is reported, for NHS111 this was estimated using change point analysis It was estimated that the NHS111 surveillance systems monitored a very high proportion (92%) of the population over the study period. Although the underlying population for the GPIH surveillance system was known and covered around 45% of the English population of children under-1 year, GPIH data was not available at a geography smaller than PHE centre. Due to this we could not determine how representative the populations monitored by the GPIH surveillance system was of the population in each region. NHS111 data was available at a small geographical level (UTLA), and from this we could determine that the coverage was even across England.

In a time-series analysis of any seasonal illness, the degree of seasonal adjustment in the model can influence the results. In this study, seasonal adjustments, through the inclusion of a random effect for week and year, were adopted to control for unobserved seasonal confounders. However, due to the highly cyclic nature of RSV in temperate regions, the addition of this component may have resulted in an underestimation of the RSV burden. If seasonal trends were not accounted for in the model, we would have assumed that all the seasonal variation in the data is due to the respiratory infections included in the model, which may not be correct. Due to computational power limitations, seasonal adjustments were applied at the national level and not the regional level. Therefore, we assume that regional trend in RSV and other respiratory pathogen activity follow the national trend.

In this analysis we applied a timeseries regression between the national laboratory pathogen data and the regional syndromic data. This assumes that regional RSV trends

follow the same national trend, only differencing by timing of RSV activity. We tried to use regional laboratory in the analysis, but due to a low number of laboratory confirmed cases of both RSV and the other respiratory pathogens we accounted for, the output had large errors around the predictions. We attempted to reduce the impact of using the national laboratory trends by allowing the laboratory data to lag behind the syndromic data to different degrees by region. This will have allowed us to detect if the timing of the national trends differed by region.

Previous studies that have investigated the timing of RSV activity, both globally and within countries have, typically, had over 10 years of data (Obando-Pacheco et al. 2018; Haynes et al. 2013; Broberg et al. 2018; Oren et al. 2018; Hogan et al. 2016), whereas our study only included four years. Although we managed to detect some differences in the timing of RSV activity, the findings were generally inconclusive. If we had analysed data for a longer period of time, more distinct patterns may have emerged. The small study period used in this analysis is because NHS111 was fully introduced in 2014. Consequently, this study can be seen as a frame work for future analysis when more data is available.

Similar studies from the literature use either laboratory confirmed cases of RSV (Haynes et al. 2013; Broberg et al. 2018; Noveroske et al. 2016), or syndromic data of indicators associated with RSV, such as acute respiratory infections (Broberg et al. 2018), infleunza-like-illness (Broberg et al. 2018), and bronchiolitis (Lewis, Stavola and Hardelid 2020; Noveroske et al. 2016). Our approach estimated the number of presentations attributable to RSV from syndromic indicators (*Difficulty Breathing* + *Cough* and *LRTI*) that have previously been associated with RSV infection. Although our approach relied on modelling the data to provide estimates of RSV, and therefore will have introduced some error into our estimates; the syndromic indicators used have previously been associated influenza and rhinovirus (Morbey et al. 2017a; 2017b), therefore using this approach will reduce the influence of these other pathogens on our observations.

The data used in this study was analysed at a highly aggregated level of geography, PHE centre, with only nine separate areas in England. A similar study which

investigates the geospatial and seasonal variation in bronchiolitis timing in England analysed data at both regional level (similar regions to those used in this study) and clinical commissioning group, with 209 areas in England (Lewis, Stavola and Hardelid 2020). This allowed the authors to explore the possible reasons for different regional timings of activity. While we were able to explore the differences, due to the highly aggregated nature of the geography used, we could not explore the potential reasons behind any variations observed. GPIH data was only available at PHE centre, therefore could not be analysed at a lower level of geography. However, NHS111 data was available at a lower level of geography, upper tier local authority level (UTLA). We estimated the number of presentations that were attributable to RSV were estimated at UTLA level, of which there are 149 areas in England, for the NHS111 data, however this produced estimates with large errors when using the methods described. We also explored using the unadjusted syndromic data, but at this level there was a lot more variation within the data, and often peaks from the *Difficulty Breathing + Cough* syndromes were observed to be during times at which RSV was not circulation. Which can be attributed to the lower specificity of these indicators, and the detection of illnesses that occur at different times of the year. Further research is required to investigate the possible reasons behind the geographical variation in RSV timing in the community.

5.7.4 Conclusion

Here were have presented an investigation into the geographical variation in RSV timing in children under-1 year in England, using data from two community healthcare services. We hypothesised that due to the latitudinal association with RSV timing, RSV activity would move northwards. Our results could not support this hypothesis. Although we could not identify specific geographical patterns in RSV activity, some variation was identified which required further research to investigate the reason behind these variations. This study provides a framework for future research when more data is available for a longer timeframe, to provide more conclusive observations and when the data can be analysed at a lower level of geography to explore the reasons behind any variations observed.

Chapter 6

Exploring the Effect of Meteorological Conditions on Acute Respiratory Infections in Children Under-Five Years, England

6.1 What is Known About the Subject?

- The meteorological conditions temperature and humidity play an important role in the transmission of acute respiratory infections through aiding virus stability and changing social behaviours.
- Children are particularly vulnerable to both acute respiratory infections and fluctuations in meteorological conditions.

6.2 What this study adds?

- Many studies that estimate the relationship between meteorological conditions and respiratory infections focus on hospitalisations, here we estimate relationships using wider burden of respiratory presentations from healthcare services in the community.
- The influence of temperature and humidity on respiratory infections is described in the literature; here we extend on what is known to explore the influence of rainfall and wind speed.
- We explore the spatial relationships between the meteorological conditions and respiratory presentations to identify areas that are particularly vulnerable to changing conditions; an aspect not explored in the literature.
- We also explored the monthly relationships between the meteorological conditions and respiratory presentations. Temperature and humidity have an

influence on respiratory presentations throughout the year, suggesting similar drivers of transmission.

6.3 Abstract

Acute respiratory infections (ARI) are a major contributor to the burden of disease and mortality in children under-5 years, globally. In children, a variety of pathogens cause ARI, with the most common being respiratory syncytial virus, influenza viruses and rhinovirus. ARI display seasonal patterns in temperate regions with the number of infections increasing in the autumn and winter, and decreasing in the spring. Here we explore the relationship between respiratory presentations to three community healthcare services and meteorological conditions in children under-5 years in England. These results can inform decision-making in public health processes and the observations can be used to understand how a changing climate may impact respiratory infections in the future.

A Bayesian spatial-temporal Poisson regression was used to estimate the relationship between respiratory presentations in children under-5 years to the three healthcare services: telehealth (NHS111), family doctors (general practitioner in-hours) and unscheduled care (general practitioner out of-hours) services from 11/11/2013 to 18/06/2018, and average meteorological conditions in the current week and previous week: temperature, rainfall, humidity and wind speed. We explored both linear and non-linear relationships, as well as the spatial and monthly associations.

Most respiratory indicators studied had a negative relationship with temperature and rainfall, and positive relationship with humidity and wind speed. The *Cold/Flu* had the strongest relationship to temperature with a 5.6% (95%CI: 5.0-6.2%) increase in weekly presentation rate for every 1°C decrease in average temperature over the previous two weeks. The *Bronchitis* indicator had the strongest relationship with wind speed, with an increase in weekly presentation rate of 4.7% (95%CI: 2.4-7.0%) for every 1m/s^2 increase in average wind speed. Rainfall was observed to have a small positive association with respiratory presentations, with a 0.03% - 2.2% increase in weekly presentation rates for every 1mm decrease in average rainfall. Humidity was observed to have between 0.03% - 0.1% increase in weekly respiratory presentation rates for every 1% increase in average humidity. The monthly influence of these meteorological

conditions was explored, we observed that for many of the indicators, temperature and humidity had an influence of presentations all year round.

These observations are important in understanding the relationship between respiratory presentations in the community in children under-5 years and meteorological conditions. In terms of the relationship between respiratory presentations and the metrological conditions studied, the association were in line with the literature. When monthly relationships were explored were found that these conditions have an influence all year round. We propose these observations are due to similar social behaviours in response to changes in meteorological conditions in both summer and winter months. We could not account for those who did not present to healthcare services, or those who presented to services not included in the analysis, such as hospitals or pharmacies. Due to the nature of the data used in the study we could not account for co-morbidities or ethnicity.

6.4 Introduction

Acute respiratory tract infections (ARI), are caused by a variety of pathogens, and are classified into upper respiratory tract infections (URTI) and lower respiratory tract infections (LRTI). URTI are typically mild and include symptoms such as coughing, sneezing, nasal congestion, and fever (Dasaraju and Liu 1996). Often URTI infections are self-limiting but can become more serious and progress to LRTI. Manifestations of LRTI include pneumonia, and bronchiolitis, with symptoms including cough, fever, and difficulty breathing (Dasaraju and Liu 1996).

Previous research has shown that rates of ARI peak in children aged 1-2 years (Monto and Sullivan 1993). In terms of URTI, children aged 0-4 years have on average 6-8 episodes per year and adults 2-4 episodes per year (Monto 1994; Monto and Ullman 1974). Males aged 0-3 years have been observed to have a higher number ARI compared to females of the same age (Monto 1994), as well as more severe infection (de Lusignan et al. 2018).

In children aged 0-4 years, the highest proportion of ARI are caused by rhinoviruses, followed by respiratory syncytial virus (RSV) and parainfluenza viruses, which have been observed to account for 39%, 19% and 18% of isolates, respectively (Monto and Sullivan 1993). In children, a variety of viruses can cause LRTI infections, including; RSV, influenza viruses, parainfluenza viruses, and rhinovirus, (Dasaraju and Liu 1996; van Woensel, van Aalderen and Kimpen 2003). Bacterial causes of LRTI include *S. pneumoniae*, *Mycoplasma pneumoniae* and *H. influenzae* (Dasaraju and Liu 1996). RSV is the most common cause of LRTI in children (Simoes 1999; Anderson et al. 1985), accounting for around 75% of hospital admissions due to bronchiolitis (Izurieta et al. 2000). Children are particularly vulnerable to complications of respiratory infections, with LRTI causing significant morbidity and mortality in children (Murdoch and Howie 2018; Troeger et al. 2018). In 2016, LRTI was estimated to result in over 68 million episodes, 5 million hospitalisations and 650,000 deaths in children under-5 years globally, the majority (75%) of these deaths occurred in children under-1 year. LRTI accounted for 13.1% of deaths in children under-5 years in 2016 (Troeger et al. 2018).

ARI display clear seasonal patterns, in temperature regions, with the number of infections increasing in the autumn and winter and decreasing in the spring. However, the different viral pathogens responsible for ARI display different seasonal oscillations. Influenza, coronavirus and RSV typically peak in winter months (Tamerius et al. 2011; Midgley et al. 2017; Landes et al. 2013; Morikawa et al. 2015; Killerby et al. 2018). Adenovirus, Human metapneumovirus (HMPV), and rhinovirus can be detected throughout the year (Landes et al. 2013; Morikawa et al. 2015), although the severity of rhinovirus increases in winter (Lee et al. 2012; Monto 2002). Some enteroviruses have increased occurrence during the summer (Abedi et al. 2018) and the seasonality of parainfluenza varies according to type (Abedi et al. 2016). An illustration of the seasonal patterns of respiratory syncytial virus, influenza A, adenovirus, and parainfluenza virus is provided in Figure 6.1 for the Netherlands. Even when pathogens are prevalent at the same time, peaks occur at different times and some pathogens can display negative associations with each other (Anestad 1982; Linde et al. 2009; Casalegno et al. 2010). Metrological conditions have been shown to drive these seasonal

patterns both directly, by increasing virus stability, and indirectly through changes in human behaviour and immunity (Moriyama, Hugentobler and Iwasaki 2020).

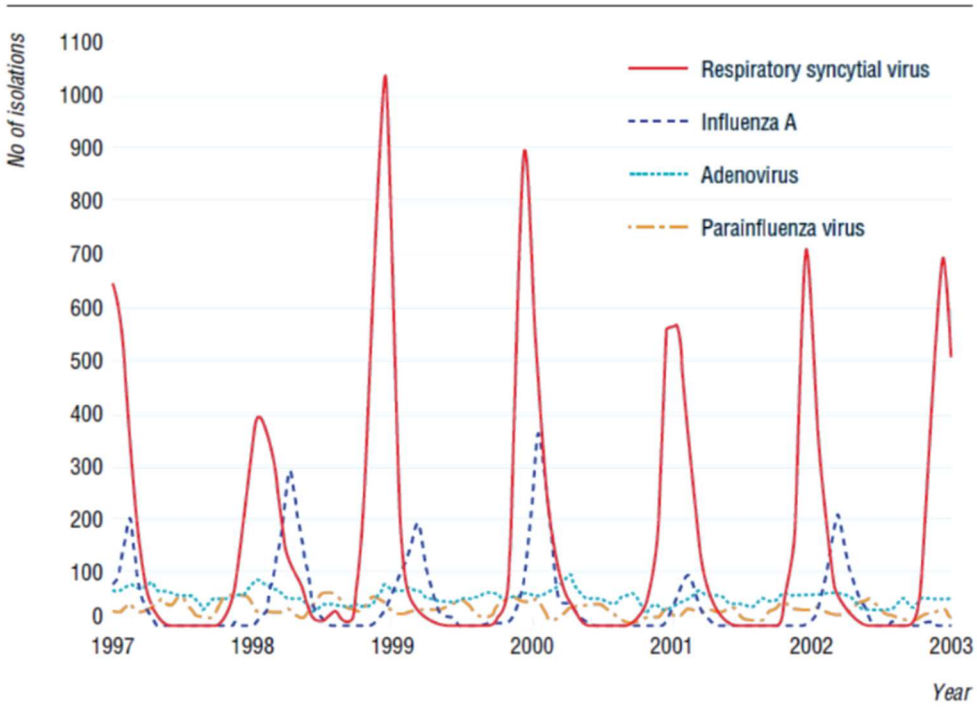

Figure 6.1: Epidemiology of respiratory syncytial virus, influenza A, adenovirus, and parainfluenza virus in the Netherlands, 1997-2003. Taken from van Woensel, van Aalderen and Kimpen (2003).

du Prel et al. (2009) explored the influence of climate on hospitalisations due to ARI in children (<16 years) in Mainz, Germany. Hospitalisations due to ARI were negatively associated with temperature and positively associated with relative humidity, wind velocity and atmospheric pressure. A study conducted in Mexico investigated the relationship between weekly meteorological conditions and primary care presentations due to ARI in the Monterrey metropolitan area. A positive relationship between relative humidity and accumulated rainfall, and negative relationship with temperature and

presentations due to ARI was observed (Costilla-Esquivel et al. 2014). As previously discussed, ARI can be caused by multiple pathogens. Using more specific indicators for disease, the relationship between pathogens and meteorological conditions may be estimated. Studies in Italy, Australia and New Zealand have investigated the association between hospitalisations due to bronchiolitis in children under-1 year and the meteorological conditions. These studies observed a decrease in temperature and an increase in humidity was associated with hospitalisations due to bronchiolitis (Nenna et al. 2017; Hoeppner et al. 2017). Price, Graham and Ramalingam (2019) investigated the associations between meteorological conditions and viral respiratory pathogens in Edinburgh, Scotland. Adenovirus, influenza viruses, RSV and HMPV were associated with low temperatures; RSV and influenza A virus were associated with a narrow humidity-range, and Human parainfluenza viruse (HPIV) type 3 was associated with seasons with lower humidity. du Prel et al. (2009) investigated the relationship between meteorological conditions and children hospitalised due to specific pathogens. RSV, influenza A, rhinovirus and adenovirus were inversely correlated with temperature, and were positively correlated with humidity. Several studies have focused on the association between confirmed RSV infection and meteorological conditions in children. RSV has been consistently inversely associated with temperature and positively associated with humidity in temperate climates (du Prel et al. 2009; Hervás, Reina and Hervás 2012; Meerhoff et al. 2009; Yusuf et al. 2007; Welliver 2007). A summary of the associations between the meteorological variables and respiratory pathogens or syndromic indicators is provided in Appendix Table A1.

Here, we present an in-depth study of the associations between presentations due to ARI to three community healthcare services in children under-5 years and meteorological conditions in England between 21st October 2013 to 25th June 2018. We explore both broad syndromic indicators (such as *ARI*, *URTI* and *LRTI*) and specific indicators (such as *Cold/Flu, Influenza-Like-Illness* and *Bronchitis*). This allows us to explore the relationship between meteorological conditions and presentations due to all ARI and symptoms that are associated with specific pathogens. The three community healthcare services are NHS111, a non-emergency medical telephone helpline; general practitioners in-hours (GPIH), a routine family doctor service; and general practitioners

out-of-hours (GPOOH), an unscheduled care service. By focusing on these services, we can get an understanding of the relationship between respiratory infections (proxied by respiratory presentations) and meteorological conditions in the community. Previous studies have primarily focused on data from hospitalisations and laboratory surveillance (Price, Graham and Ramalingam 2019; du Prel et al. 2009; Hervás, Reina and Hervás 2012; Meerhoff et al. 2009; Yusuf et al. 2007; Welliver 2007), which capture a fraction of the true number of cases and often only detect the most severe cases. The focus of this study is children under-5 years given their vulnerability to ARI (Nair et al. 2013; WHO and UNICEF 2013) and changes in meteorological conditions (Xu et al. 2012; CSMF 2011; Xu, Hu and Tong 2014). A better understanding of the relationship between ARI and meteorological conditions will allow us to make reliable predictions on the incidence of presentations related to paediatric respiratory infections in response to changes in meteorological conditions. Our results can aid policy makers and health professionals prepare and target public health measures. With observable changes in our weather related to climate change, this research can also be used to understand how a changing climate may impact ARI in the future.

Most studies presented in this review have assumed a linear relationship between meteorological conditions and respiratory presentations. Several studies have demonstrated a more complex relationship when non-linear relationships are explored (Nyoka et al. 2017; Liu et al. 2015; Mäkinen et al. 2009). We will explore both linear and non-linear relationships. The majority of studies previously discussed have focused on the relationship between ARI and meteorological conditions in localised areas such as cities (Xu et al. 2012; CSMF 2011; Xu, Hu, and Tong 2014). Given that our data is available from regions across the whole of England, we can explore if there are spatial variations in the associations between ARI and meteorological conditions. Different respiratory pathogens circulate at different times of the year (Moriyama, Hugentobler and Iwasaki 2020). Therefore, we will explore how the relationship between ARI and meteorological conditions changes throughout the year. For this analysis gender, deprivation, urbanicity and public holidays, will be accounted for as they have previously been associated with presentations due to respiratory illness (chapter three).

The aims of this study are:

1. To characterise the relationship between ARI syndromic indicators to NHS111, GPOOH and GPIH healthcare services and the meteorological conditions: temperature, relative humidity, rainfall, and wind speed, in England in children under-5 years.
2. To investigate if the associations between ARI and the meteorological conditions differ spatially across England.
3. To evaluate if the associations between ARI and the meteorological conditions change over the year.

6.5 Methods

6.5.1 Data Sources

Epidemiological Data

Syndromic surveillance data was obtained from three healthcare systems monitored by Public Health England (PHE) real-time syndromic surveillance team (ReSST): NHS111 (a telephone helpline), GPIH (general practitioner in-hours; a family doctor service) and GPOOH (general practitioner out-of-hours; an unscheduled care service). Data comprised of respiratory indicators (Table 6.1) that have previously been associated with ARI and its causative pathogens including RSV, influenza viruses, rhinovirus and HMPV in young children (Morbey et al. 2017a; 2017b; 2018; chapter four). Summary statistics and descriptions of each indicator are provided in Table 6.1. Time-series of each indicator at the national level are provided in Figures 6.2, 6.3 and 6.4, where the seasonal trends of each indicator can be observed.

The syndromic data comprised of weekly counts, to remove day of the week effects, of each indicator in children under-5 years between 21st October 2013 and 25th June 2018. The count data was stratified by week, gender, and geographical location. Data was obtained at the lowest possible geography. For the GPIH data this was PHE centre and

for NHS111 and GPOOH data this was upper tier local authority (UTLA). Due to issues with small populations (<1,900 residents under-5 years), data could not be obtained for three UTLA: City of London, the Isles of Scilly, and Rutland.

	Indicator	Mean weekly count (Min - Max)	Total number of presentations	Percentage of weeks with Zero counts	Average weekly population under 5 years	Location (*n*)
NHS111	Difficulty Breathing	8.8 (0-169)	573,197	5.46	3,405,10	UTLA (149)
	Cough	12.4 (0-284)	812,968	4.15	3,405,10	UTLA (149)
	Cold/Flu	1.38 (0-28)	90,440	43.49	3,405,10	UTLA (149)
GPOOH	All Respiratory Disease (ARD)	31.6 (0-346)	1,193,130	1.74	2,664,454	UTLA (149)
	Acute Respiratory Infection (ARI)	25.5 (0-303)	959,181	2.80	2,664,454	UTLA (149)
	Difficulty Breathing with Asthma (DBWA)	2.7 (0-44)	69,304	29.03	2,664,454	UTLA (149)
	Bronchitis	0.8 (1-21)	28,611	64.89	2,664,454	UTLA (149)
GPIH	Upper Respiratory Tract Infection (URTI)	752.2 (46-3,969)	3,303,633	0.0	1,723,424*	PHE Centre (9)
	Lower Respiratory Tract Infection (LRTI)	137.3 (5-951)	603,272	0.0	1,723,424*	PHE Centre (9)
	Influenza Like Illness (ILI)	3.8 (0-74)	16,863	33.9	1,767,734*	PHE Centre (9)
	Acute Bronchitis	3.25 (0-10)	14,267	18.8	1,239,448*	PHE Centre (9)

Table 6.1: Summary and descripton of weekly syndromic indicator data included in the study, between 21st October 3013 and 25th June 2018. * Number of registered populations differ for GPIH indicators because GP practices submit different syndromic indicator data depending on the system they use.

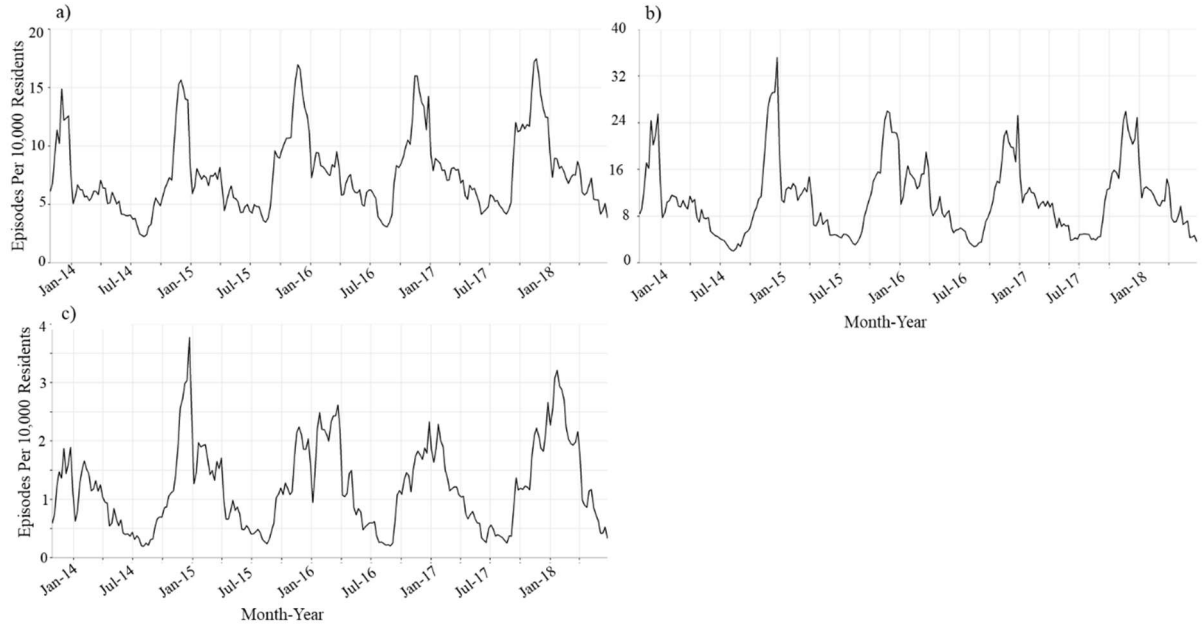

Figure 6.2: National time-series of NHS111 syndromic indicators over the period 21st October 2013 to 25th June 2018 in England: a) *Difficulty Breathing*, b) *Cough* and, c) *Cold/Flu.*

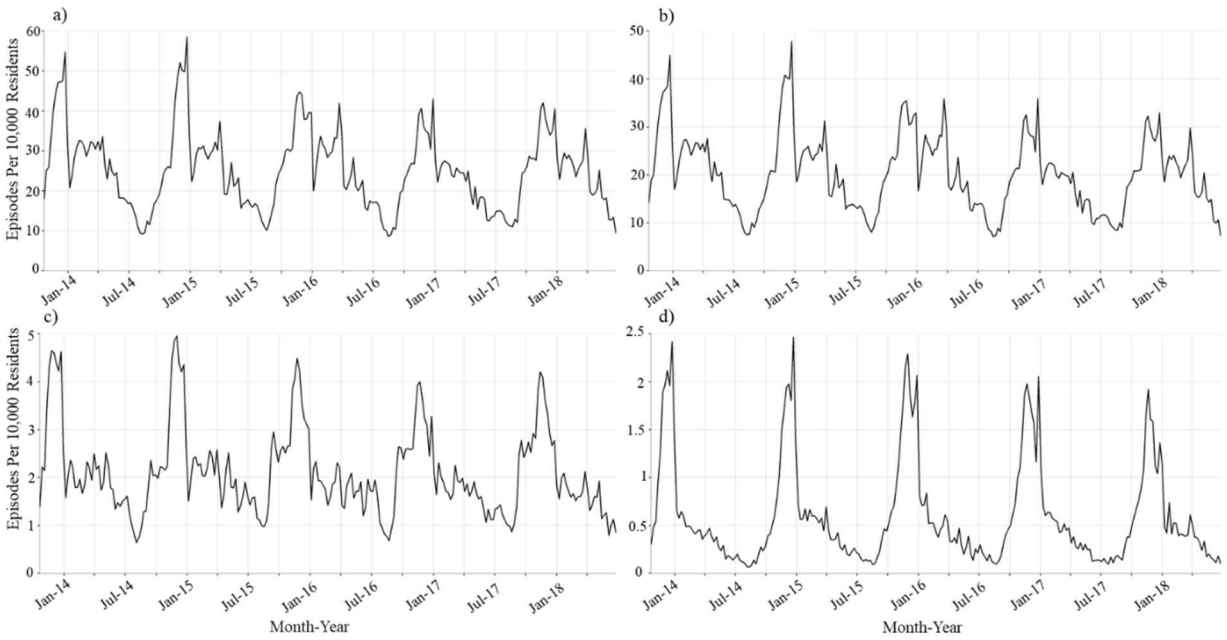

Figure 6.3: National time-series of GPOOH syndromic indicators over the period 21st October 2013 to 25th June 2018 in England: a) *Acute Respiratory Disease*, b) *Acute Respiratory Infection*, c) *Difficulty Breathing with Asthma*, and d) *Bronchitis.*

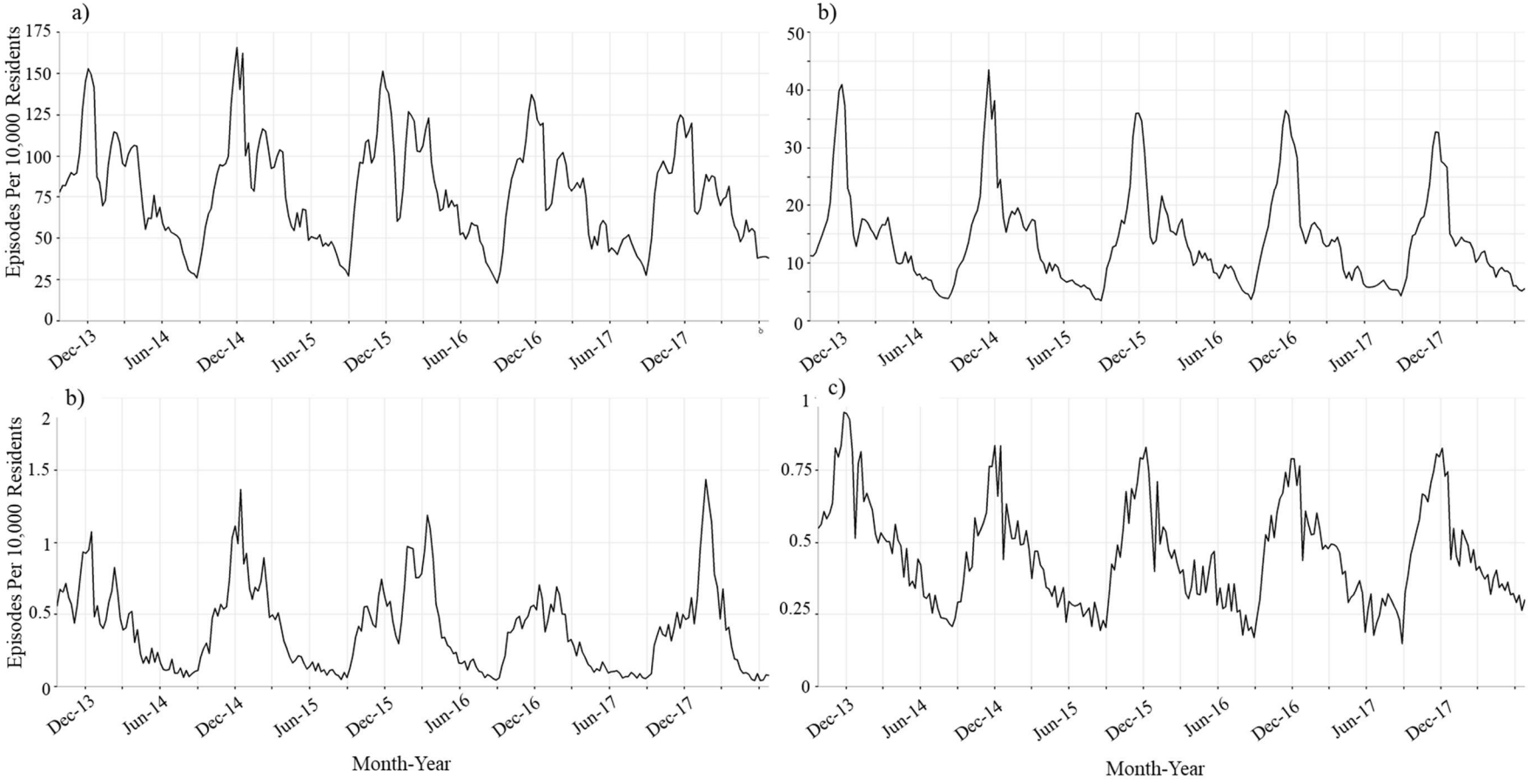

Figure 6.4: National time-series of GPIH syndromic indicators over the period 21st October 2013 to 25th June 2018 in England: a) *Upper Respiratory Tract Infection,* b) *Lower Respiratory Tract Infection*, c) *Influenza-Like-Illness*, and d) *Acute Bronchitis*.

Estimating Areas with Poor Coverage

There were periods of time when no or very low numbers of respiratory presentations were observed for individual UTLA for NHS111 and GPOOH. This is due to periods where there were data transfer issues between the providers and PHE (ReSST, personal communication, July 2019). Therefore, the true number of cases was unknown and these data needed to be excluded. The periods of poor coverage were detected using the pruned exact linear time method (PELT) of change-point analysis (Killick, Fearnhead and Eckley 2012). PELT estimates periods of poor coverages based upon changes in mean and variance of the data. PELT was conducted using the R package *changepoint* (Killick and Eckley 2014). If an UTLA was found to have poor or no coverage, it was excluded from the analysis for that period of time. A more thorough description of how we estimated periods of poor coverage can be found in chapter four.

Demographic Data

Population data for children under-5 years in England used for NHS111 and GPOOH analysis were obtained at the UTLA level from mid-year population estimates stratified by gender (ONS 2014). Population used in GPIH analysis was the average number of daily registered patients each week, stratified by gender. Populations for the GPIH data were obtained directly from the surveillance system. Further explanation of the demographic data used in this study is provided in chapter four.

Socioeconomic Data

Deprivation scores for each UTLA were obtained from the English Indices of Deprivation 2015 (DCLG 2015). The Office for National Statistics Rural Urban Classification were used to obtain the percentage of UTLA classified as urban (DEFRA 2016). Due to the highly aggregated nature of deprivation and urbanicity data at PHE centre these variables were not accounted for in the analysis of GPIH data. Further descriptions of the deprivation and urbanicity data are provided in Table 6.2.

Metrological Data

Based on their association with ARI as described from previous studies: temperature (Price, Graham and Ramalingam 2019; du Prel et al. 2009; Hervás, Reina and Hervás 2012; Meerhoff et al. 2009; Yusuf et al. 2007; Welliver 2007; Hoeppner et al. 2017; Nenna et al. 2017; Costilla-Esquivel et al. 2014; Sennerstam and Moberg 2004), relative humidity (Welliver 2007; Mäkinen et al. 2009; Lowen et al. 2007; Shaman and Kohn 2009), rainfall (Costilla-Esquivel et al. 2014; Agrawal et al. 2009; Brooks et al. 2007) and wind speed (Hoeppner et al. 2017) were selected as independent variables of interest. Hourly relative humidity (%), data were obtained from the ERA5-hourly gridded dataset on pressure levels, at 31km^2 resolution, and hourly rainfall (mm), temperature (°C) and wind speed (m/s^{-1}) data was obtained from the ERA5-Land hourly dataset, at 9km^2 resolution, from the Copernicus Climate Data Store (C3S 2019; 2017). Hourly data was averaged to obtain a weekly average, and then a spatially weighted average of values in each UTLA and PHE centre were calculated. This produced a weekly mean of relative humidity, rainfall, temperature and wind speed for each spatial (UTLA or PHE centre) area. For the purposes of this thesis, they will be referred to as humidity, rainfall, temperature and wind speed. Details of these variables and how they were generated in presented in Table 6.2. Spatial variation and time-series of each weather variable over the study period is illustrated in Appendix 2. A moving average (MA) of the weather variables of the current week and the previous week were calculated and used for the analysis as we hypothesised that the weather from the current week and previous week would have an effect on ARI presentations and to account for a lag time between exposure and development of symptoms.

	Original Data				Data Used in Analysis			
Covariate	Data Source	Definition of Original Data	Time Resolution	Space Resolution	Definition of Data Used in Analysis	Time Resolution	Space Resolution	Method to obtain data for analysis
Mean weekly temperature (°C)	Copernicus atmospheric reanalysis of the global climate – ERA5-Land hourly dataset (C3S 2019).	Hourly temperature at 2m above the surface of land.	Hourly	9km grid	Weekly mean of daily mean temperature at 2m above the surface for each area (UTLA or PHE Centre) in °C.	Weekly	UTLA and PHE Centre	1. Using Python 3 (Van Rossum and Drake 2009), climate data was downloaded using the Climate Data Store API. 2. Using the Python package *xarray* (Hoyer and Hamman 2017), hourly data was resampled to a daily time period to calculate the daily mean of temperature. 3. In R version 3.5.2 (R Core Team 2018), for each grid point the average weekly daily temperature was calculated. 4. A weighted mean for each area was calculated. Weighted means were based on the proportion of each raster grid square in the UTLA5. Data was converted from K to °C.
Mean weekly humidity (%)	Copernicus atmospheric reanalysis of the global climate – ERA5- hourly gridded dataset on pressure levels (C3S 2017).	Hourly water pressure as a percentage at which the air becomes saturated. This data was obtained at 1000hPa pressure levels which is around the	Hourly	31km grid	Weekly mean of daily mean of water pressure as a percentage at which the air becomes saturated for each area.	Weekly	UTLA and PHE Centre	1. Using Python 3 (Van Rossum and Drake 2009), climate data was downloaded using the Climate Data Store API. 2. Using the Python package *xarray* (Hoyer and Hamman 2017), hourly data was resampled to a daily time period to calculate the daily mean of humidity. 3. In R version 3.5.2 (R Core Team 2018), for each grid point the average weekly daily humidity was calculated. 4. A weighted mean for each area was calculated. Weighted means were based on

Table 6.2: Description of independent data included in analysis.

		average surface pressure level.						the proportion of each raster grid square in the UTLA.
Mean weekly rainfall (mm/day)	Copernicus atmospheric reanalysis of the global climate – ERA5-Land hourly dataset (C3S 2019).	Hourly total liquid fall to the earth's surface. This includes rain and snow.	Hourly	9km grid	Weekly mean of daily sum of total liquid fall to the earth's surface for each area in mm.	Weekly	UTLA and PHE Centre	1. Using Python 3 (Van Rossum and Drake 2009), climate data was downloaded using the Climate Data Store API. 2. Using the Python package *xarray* (Hoyer and Hamman 2017), hourly data was resampled to a daily time period to calculate the daily sum of rainfall. 3. In R version 3.5.2 (R Core Team 2018), for each grid point the average weekly daily rainfall was calculated. 4. A weighted mean for each area was calculated. Weighted means were based on the proportion of each raster grid square in the UTLA. 5. Rainfall was converted from meters to millimetres.
Mean weekly wind speed (m/s^{-1})	Copernicus atmospheric reanalysis of the global climate – ERA5-Land hourly dataset (C3S 2019)	Data obtained from 10m u- and 10m v-component of wind – Eastward and northward components of the 10m wind. Horizontal	Hourly	9km grid	Weekly mean of daily wind speed at 10m above the earth's surface in m/s^{-1}.	Weekly	UTLA and PHE Centre	1. Using Python 3 (Van Rossum and Drake 2009), climate data was downloaded using the Climate Data Store API. 2. In Python using the u and v components of wind, wind speed was calculated (uv): $(u^2 + v^2)^{1/2}$. 3.Using the Python package *xarray* (Hoyer and Hamman 2017), hourly data was resampled to a daily time period.

Table 6.2: Description of independent data included in analysis.

		speed of air moving towards the east and north, at a height of ten metres above the surface of the Earth.						4. In R version 3.5.2 (R Core Team 2018), for each grid point the average weekly daily winds peed was calculated. 5. A weighted mean for each area was calculated. Weighted means were based on the proportion of each raster grid square in the UTLA
Deprivation	Department Communities & Local Government (DCLG 2015)	Mean index of multiple deprivation score at UTLA (2015).	-	UTLA	Mean index of multiple deprivation score at UTLA.	Whole study period	UTLA	As obtained from DCLG
Percentage (%) of UTLA that is urban	Department for Environment, Food & Rural Affairs (DEFRA 2016)	Percentage of UTLA population living in urban environments (2011).	-	UTLA	Percentage of UTLA population living in urban environments .	Whole study period	UTLA	As obtained from DEFRA

Table 6.2: Description of independent data included in analysis.

6.5.2 Data Exploration

The underlying aim of this study is whether the meteorological variables; temperature, humidity, wind speed and rainfall, are associated with the respiratory indicators. Before we could test for this aim, we had to identify whether collinearity between these variables exists. Collinearity was investigated using the variance inflation factors (VIF), with a threshold > 2.5 indicating collinearity (Dormann et al. 2013; Johnston, Jones and Manley 2018). Pairwise correlations between each variable were also calculated, with a value higher than 0.5 indicating collinearity (Dormann et al. 2013). All VIF values were calculated to be <2 (Table 6.3) and no issues with high correlations were observed (Figure 6.5).

A high number of zeros were observed in some of the syndromic indicators of interest (Table 6.1), syndromic indicators that displayed a high number of zeros were modelled using both Poisson and zero-inflated Poisson distributions to ensure the models were accounted for zeros adequately. Distribution was chosen based on deviance information criterion (DIC) and how well the zeros were accounted for. To identify outliers in the model predictions, residual plots were examined.

Variable	VIF Value (UTLA)	VIF Value (PHE Centre)
Rainfall	1.18	1.24
Humidity	1.40	1.39
Temperature	1.40	1.41
Wind Speed	1.19	1.25
Percentage Urban	1.32	-
Average Deprivation Score	1.31	-

Table 6.3: VIF values for variables included in analysis. VIF values provided for both analysis at UTLA and PHE centre.

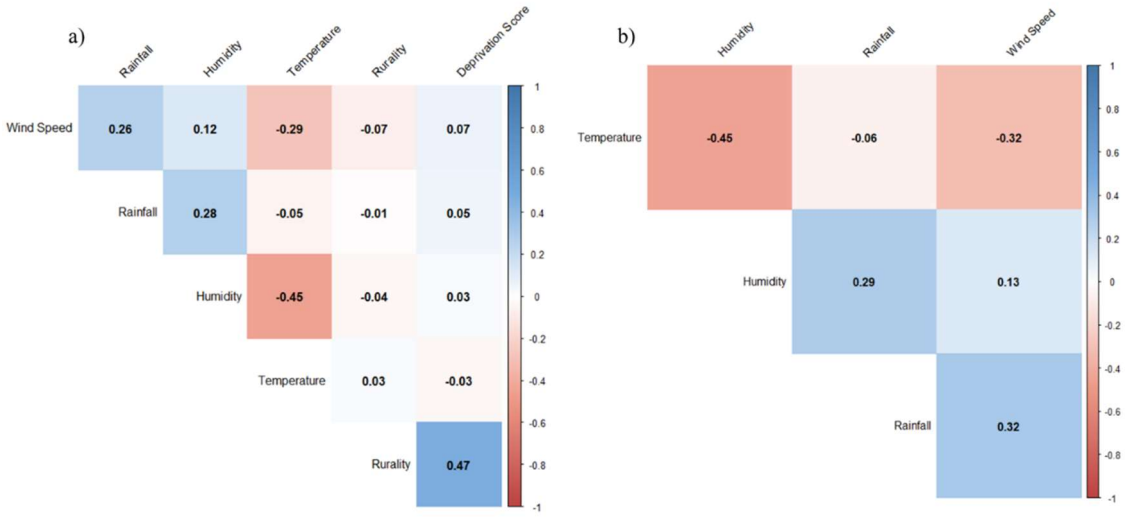

Figure 6.5: Pairwise correlations of variables included in analysis; a) variables at UTLA and b) variables at PHE centre.

6.5.3 Statistical Analysis

A Bayesian approach was used for statistical inference in this study using the R package *R-INLA* (Martins et al. 2012). This approach, using the *R-INLA* package, allows us to account for both spatial and temporal dependency as part of the analysis; other packages that allow us to account for these dependency structures are limited in R. *R-INLA* uses the integrated nested Laplace approximation (INLA) for probability distribution sampling which is capable of much faster Bayesian inference compared to other Bayesian approaches for distribution sampling. The Bayesian inference using the INLA approach through the package *R-INLA* has successfully been demonstrated in several epidemiological studies where this methodology has allowed complex spatial and temporal studies to be undertaken at national and subnational levels (Dwyer-Lindgren et al. 2019; Sparks 2015; Librero et al. 2017).

The aims of this study are to explore the relationships between the respiratory indicators and weather variables, and to investigate how the relationship with these weather variables vary spatially and monthly. To meet these aims we followed a series of steps to understand the effect of the observed covariates (temperature, humidity, wind speed

and rainfall), on respiratory presentations in the community in children under-5 years, while controlling for gender, deprivation, urbanicity, and seasonality. These models controlled for deprivation, rurality and public holidays to reduce the impact of other causes of health seeking behaviours on our observations.

Accounting for Spatial Autocorrelation in NHS111 and GPOOH Data

Spatial autocorrelation was accounted for when modelling NHS111 and GPOOH syndromic indicators. Through the neighbourhood structure, information on an area and its nearest neighbours are used in the analysis to control for spatial autocorrelation. Nearest neighbours are those that share a border with the given area, the neighbourhood structure of the 149 UTLA in this analysis are illustrated in Figure 6.6. The Besag-York-Mollie (BYM) model was used to model this spatial dependency using this neighbourhood structure (Besag, York and Mollié 1991). This model estimates the conditional mean of the random effects as an average of an area's neighbours to account for spatial correlations, while also accounting for variation that is not spatially correlated.

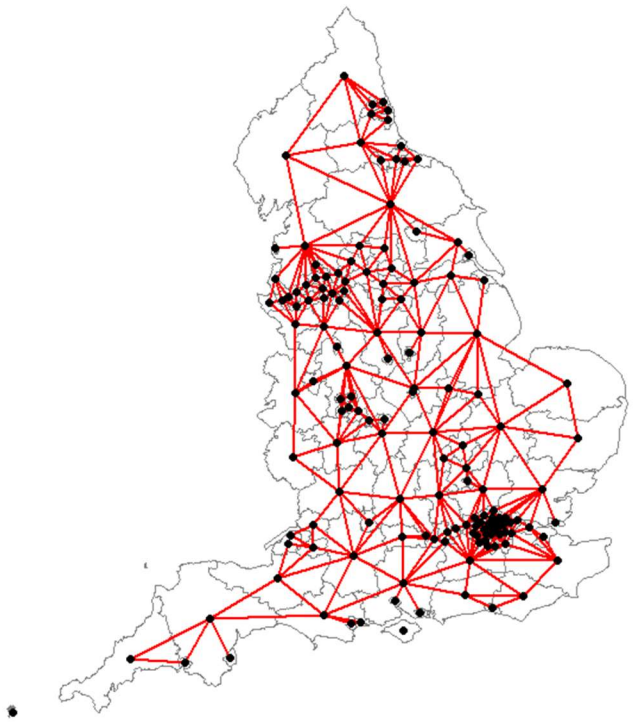

Figure 6.6: Nearest neighbours at UTLA level, England.

The Base Model

This model explores the influences of the fixed continuous covariates (i.e., temperature, wind speed, humidity, and rainfall) on the dependent variables (respiratory syndromic indicators), while accounting for deprivation, urbanicity, gender, public holidays, seasonality, and spatially and temporally structured random effects. The meteorological covariates (temperature, wind speed, humidity, and rainfall) vary by both space and time, deprivation and urban percentage vary spatially, and public holidays vary by time. Given that GPIH data was only available at a high level of geography (nine coarse separate areas) spatial dependency was not accounted for.

To allow for different time trends by space (UTLA or PHE Centre), an interaction between space and time (week) was included in the model. The interaction comprised of an unstructured spatial effect and a structured temporal effect, with a first order autoregressive term for the residuals of time. By including a structured temporal effect, temporal dependency could be accounted for. An autocorrelation term of the residuals was included when accounting for these time trends to adjust for the communicable, person-to-person, nature of respiratory infections (Allard 1998). Annual and weekly variation was accounted for by including a random effect for year and week to account for unobserved temporal effects and seasonality of exposures (Fisman 2012). Public holidays, as well as a one-week lag of public holidays, were included in the model to account for the closure of medical facilities during the study period, and changes in healthcare-seeking behaviours, as well as their lag effects. Gender, deprivation, and urbanicity were also accounted for as these were identified as possible confounding factors. The variables of interest (an average the current week and previous week of temperature, humidity, wind speed and rainfall) were included in the model as continuous fixed covariates.

The general algebraic definition of all the NHS111, GPOOH and GPIH models is given by:

$$Yg,i,t \mid \mu g,i,t, \sim F(\mu g,i,t),$$

Where $\mu_{g,i,t}$ is the is the gender g, specific number of contacts at each UTLA i at time t, F is the Poisson likelihood of the model.

A logarithmic link function of the expected number of cases for NHS111 and GPOOH data is modelled as:

$$\log(\mu g,i,t) = \alpha + \log(Pg,i,t) + \beta Xi,t + \zeta di + \eta ki + \gamma g + \tau Hi,t + \theta Hi,t-1 + \pi Ri,t-1 + uw + nuy + \omega i + \psi i$$

where α corresponds to the intercept; $\log(P_{g,i,t})$ denotes the logarithm of the gender g and geographical location i specific population P at risk for each time point (week) t. This was included as an offset to adjust counts by population; X is a matrix of moving average of the mean weekly meteorological variables: temperature, rainfall, humidity and wind speed, at each time point t and geographical location i with regression coefficients β to account for their delayed effects on the syndromic indicators; d denotes a continuous variable for deprivation score with coefficient ζ, at each geographical location i; k denotes a continuous variable to represent the percentage of geographical location i that is classified as urban with coefficient η. g denotes a categorical variable gender with coefficient γ; H denotes a Boolean variable indicating the presence of public holidays on the current (t) or previous week ($t-1$) with coefficients τ or θ accordingly; π is a random effect of the auto-regressive coefficient of the residuals R lagged one week at each geographical location i to account for potential temporal autocorrelation; unstructured random effects of week (u_w) and year (v_y) were included to account for unobserved seasonal trends. Unknown spatial confounding factors and spatial dependency structures were incorporated using structured (ω_i), and unstructured (ψ_i) spatial random effects for each geographic location i. Spatial random effects were specified using a Besag-York-Mollie model (Besag, York and Mollié 1991).

A logarithmic link function of the expected number of cases for GPIH data is modelled as:

$$\log(\mu g,i,t) = \alpha + \log(Pg,i,t) + \beta Xi,t + \gamma g + \tau Hi,t + \theta Hi,t-1 + \pi Ri,t-1 + uw + nuy$$

where α corresponds to the intercept; $\log(P_{g,i,t})$ denotes the logarithm of the gender g and geographical location (PHE centre) i specific population P at risk for time t. This was included as an offset to adjust counts by population; X is a matrix of the MA of the mean weekly meteorological variables: temperature, rainfall, humidity and wind speed, at each time point t and geographical location i with regression coefficients β to account for their delayed effects on the syndromic indicators; g denotes a categorical variable gender with coefficient γ; H denotes a Boolean variable indicating the presence of public holidays on the current (t) or previous week ($t-1$) with coefficients τ or θ accordingly; π is a random effect of the auto-regressive coefficient of the residuals R lagged one week at each geographical location i to account for potential temporal autocorrelation; unstructured random effects of week (u_w) and year (v_y) were included to account for unobserved seasonal trends.

Within the model non-linearities were also explored but the results indicated that, although small non-linearities between the respiratory indicators and meteorological conditions were observed, linear models are adequate at describing the associations.

Spatially Varying Covariates

To investigate whether the relationships between the respiratory indicators and meteorological conditions varied by location an interaction term between each covariate of interest and space was introduced into the base model. The NHS111 and GPIH base models were extended to allow spatially varying relationships with the meteorological conditions by including the term $\beta X_t \cdot \delta i_t$ which denotes an interaction between each of the meteorological conditions (temperature, rainfall, humidity and wind speed) and a categorical variable for each geographical location i. Given the inconsistent spatial availability of data from the GPOOH surveillance system (Table 6.4), we did not analyse the data using this method for GPOOH.

Monthly Varying Covariates

By investigating how the relationship between the response variable and the covariates of interest vary by month, we can observe if there is any difference in the relationships

across the year. The NHS111, GPOOH and GPIH base models were extended to allow monthly varying relationships with the meteorological conditions by including the term $\beta X_t \cdot \delta m_t$ which denotes an interaction between each of the meteorological conditions (temperature, rainfall, humidity and wind speed) and a categorical variable for each month m.

Model Diagnostics

Models were chosen based on the deviance information criterion (DIC), a method described by Spiegelhalter and Best *et al.* (2002) (Spiegelhalter et al. 2002), with the best fitting model having the lowest DIC value. Models were described as different when the DIC value difference was greater than seven, weakly different with a value between three and seven, and not different when the difference in DIC values was below.

6.6 Results

6.6.1 Population Coverage During Study Period

The coverage of each surveillance system during the study period is described in Table 6.4. NHS111 had a high level of surveillance coverage, with 90.7% of the total under-5 population in England during the study period, and GPOOH monitored 61.1% of the total under-5 population in England. The population coverage of the GPIH surveillance system ranged from 37.6-52.2% depending on the respiratory indicator. Different practices would submit information on different indicators; therefore, the underlying number of registered patients was different.

Surveillance System	Percentage (%) of Population Monitored	Percentage (%) of Spatial Areas with Surveillance Coverage
NHS111	90.7	77.2
GPOOH	61.1	43.6
GPIH	37.6-52.2[1]	100[2]

Table 6.4: Coverage by the surveillance systems over the study period, 21st October 2013 to 25th June 2018.
[1]Syndromic indicators had different populations depending on the GPIH system
[2]All GPIH PHE centres had coverage over the study period, but we do not know how representative of the population it is

6.6.2 Model Selection

The variables included in the models were chosen prior to analysis based on hypothesis generation and biological plausibility. Spatial autocorrelation was accounted for in the NHS111 and GPOOH through the inclusion of a BYM model, and temporal autocorrelation was accounted for by including an autoregressive term for time in each spatial area. Temporal autocorrelation was assessed in all models using partial autocorrelation function plots which were found to be satisfactory. Residual plots were also inspected to ensure there were no underlying patterns in the error terms and these were observed to be satisfactory.

NHS111 indicator *Cold/Flu*, GPOOH indicators *DBWA* and *Bronchitis*, and GPIH indicators *ILI* and *Acute Bronchitis* were modelled using both zero-inflated and non-zero inflated Poisson distributions, with models selected based on the lowest DIC (Table 6.5). In all cases, the Poisson distribution was found to be the best model, with the zero-inflated models either no different or slightly worse based on DIC. Dispersion adequately accounted for in all models, with a dispersion statistic < 1 (Table 6.5).

	Indicator	Model	DIC	Dispersion
NHS111	Difficulty Breathing	Poisson	309812.9	0.93
	Cough	Poisson	333139.8	0.90
	Cold/Flu	Poisson	168588.4	0.91
		Zero-inflated Poisson	168597.7	-
GPOOH	ARD	Poisson	247440.3	0.92
	ARI	Poisson	236698.7	0.92
	DBWA	Poisson	131388.5	0.95
		Zero-inflated Poisson	131394.5	-
	Bronchitis	Poisson	69970.5	0.82
		Zero-inflated Poisson	69973.8	-
GPIH	URTI	Poisson	42566.7	0.80
	LRTI	Poisson	33744.7	0.71
	ILI	Poisson	14238.4	0.87
		Zero-inflated Poisson	14280.0	-
	Acute Bronchitis	Poisson	14329.4	0.57
		Zero-inflated Poisson	14333.7	-

Table 6.5: Summary statistics of linear models.

6.6.3 The Base Model

When exploring linear relationships there was a large amount of agreement between the respiratory indicators as to which meteorological conditions were identified as important by the models (Rate ratios (RR) and 95% credible intervals (CI) are presented in Table 6.6). Bayesian analysis does not infer statistical significance as no hypothesis testing is undertaken, instead analysis identifies the variables that have an important relationship with the dependant variable. A relationship is considered important when the 95% CI does not cross one. Ten out of 11 of the respiratory indicators had an important relationship with temperature, with all displaying an inverse relationship indicating that as temperature decreases, respiratory presentations increase. Eight out of 11 respiratory indicators had an important relationship with rainfall, with all displaying an inverse relationship. Nine out of 11 respiratory indicators had an important relationship with humidity and wind speed, all displaying a positive relationship.

In terms of the size of the relationship between the meteorological conditions and respiratory indicators most estimates were similar between the respiratory indicators (Table 6.6 and Figure 6.7). Temperature RR were between 0.944-0.993 indicating a 0.7-5.6% decrease in presentation rates for every 1°C increase in average temperature over

the previous two weeks, with *Cold\Flu* having the strongest relationship with temperature compared to the other indicators (Table 6.6 and Figure 6.7). Again, rainfall estimates were similar among respiratory indicators where rainfall was important within the models, with most estimates showing negative associations between 0.978-0.997, indicating a 0.3-2.2% decrease in weekly presentation rate for every 1mm/day increase in average rainfall over the previous two weeks. Again, *Cold/Flu* had the strongest relationship with rainfall compared to the other indicators (Table 6.6 and Figure 6.7). All estimates of humidity were between 1.003- 1.010, indicating an increase in weekly presentation rate of 0.3-1.0% for every 1% increase in average humidity over the previous two weeks (Table 6.6 and Figure 6.7). Estimates for wind speed were between 1.009-1.053, indicating an increase in weekly presentation rate of 0.9-5.3% for every $1m/s^2$ increase in average wind speed over the previous two weeks. The GPOOH indicator *Bronchitis* had the strongest relationship with wind speed (Table 6.6 and Figure 6.7).

	Indicator	Temperature (°C)	Rainfall (mm)	Humidity (%)	Wind Speed (m/s)
NHS111	Difficulty Breathing	**0.975 (0.972-0.978)**	0.999 (0.996-1.002)	**1.004 (1.004-1.005)**	**1.016 (1.010-1.022)**
	Cough	**0.979 (0.976-0.982)**	**0.991 (0.989-0.994)**	**1.003 (1.003-1.004)**	**1.016 (1.011-1.022)**
	Cold/Flu	**0.944 (0.938-0.950)**	**0.978 (0.972-0.984)**	**1.008 (1.006-1.010)**	**1.042 (1.030-1.055)**
GPOOH	ARD	**0.979 (0.976-0.981)**	**0.995 (0.993-0.997)**	**1.006 (1.005-1.006)**	**1.018 (1.013-1.023)**
	ARI	**0.980 (0.977-0.983)**	**0.995 (0.992-0.997)**	**1.005 (1.005-1.006)**	**1.018 (1.013-1.023)**
	DBWA	**0.965 (0.960-0.971)**	0.995 (0.990-1.001)	**1.009 (1.008-1.011)**	**1.030 (1.018-1.041)**
	Bronchitis	**0.976 (0.965-0.987)**	**0.984 (0.974-0.994)**	**1.009 (1.006-1.011)**	**1.053 (1.030-1.076)**
GPIH	URTI	**0.993 (0.990-0.995)**	**0.995 (0.993-0.997)**	1.000 (1.000-1.001)	1.001 (0.996-1.005)
	LRTI	0.999 (0.996-1.003)	**0.997 (0.994-0.999)**	1.000 (0.999-1.001)	1.001 (0.994-1.007)
	ILI	**0.972 (0.956-0.988)**	0.989 (0.977-1.002)	**1.010 (1.004-1.017)**	**1.039 (1.008-1.071)**
	Acute Bronchitis	**0.971 (0.959-0.984)**	0.998 (0.983-1.012)	**1.006 (1.003-1.009)**	**1.028 (1.001-1.055)**

Table 6.6: Rate ratios with 95% credible intervals. Variables that were considered important in the models are in bold.

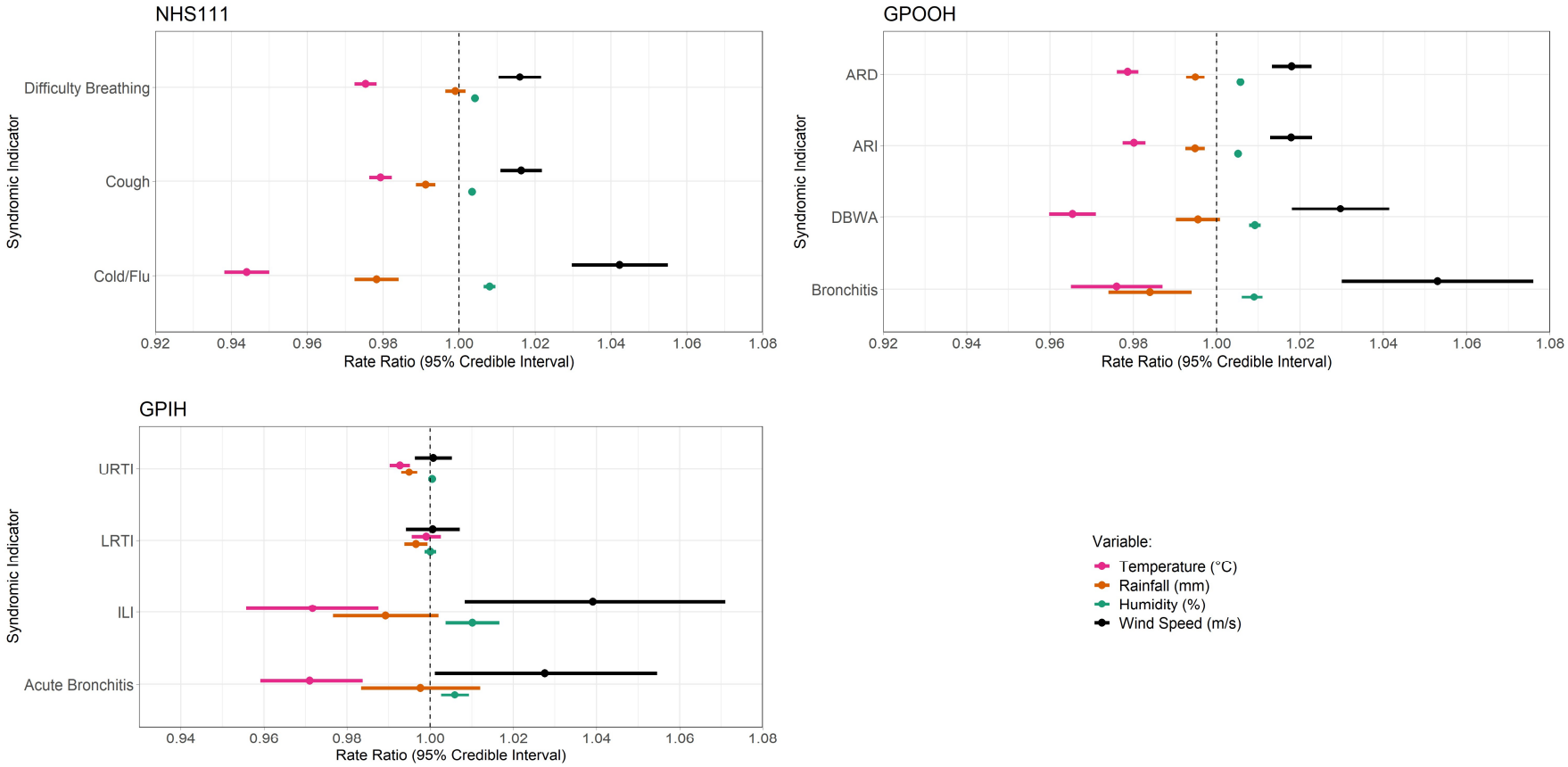

Figure 6.7: Rate ratios and 95% credible intervals.

6.6.4 Spatially Varying Covariates

The spatial distribution of the relationships between the meteorological conditions and indicators from NHS111 and GPIH was explored to identify areas that are vulnerable to changes in meteorological conditions. This spatial distribution could not be explored in GPOOH data due to the inconsistent and low surveillance systems coverage in some areas.

The spatial distribution of the relationship between each respiratory variable and meteorological condition is described in Figures 6.8-6.11. Here, maps are presented for the importance of the relationship within each area. A relationship is considered important when the 95% credible interval did not cross one, the direction of the relationship is also described in the importance map, which highlights the areas where the models identified the relationships between the meteorological conditions and respiratory indicator as important.

The spatial distribution of the relationship between temperature and all the NHS111 indicators was important in the models with a negative relationship in the majority of UTLAs in England (Figure 6.8). In most areas in England, there was a negative relationship between the NHS111 indicator *Cough* and temperature, except from the north where temperature was not important in the model. The relationship between temperature and GPIH indicators *URTI, ILI* and *Acute Bronchitis* was important in most PHE centres (Figure 6.8). The relationship between temperature and the GPIH indicator *URTI*, was not identified as important by the model in the north and west, but there was an negative relationship in the south and east.

Rainfall was not identified as important by the models in the majority of England for the four NHS111 indicators, but clusters of negative relationships were observed in the northwest, west and south of England (Figure 6.9). Rainfall was not identified as important by the models for most regions for the GPIH indicators *LRTI* and *ILI* and *Acute Bronchitis*. Rainfall was identified as important by the model in the western, middle, and southern areas of England for *URTI*, this relationship was negative (Figure 6.9).

Although clusters of positive relationships between humidity and NHS111 indicators can be observed, there are few clear spatial patterns (Figure 6.10). Most English areas have a positive relationship between humidity and *Cold/Flu*, with large clusters in the north. For the GPIH indicators there was a clear spatial pattern with humidity, with an positive relationship in the west and south. *URTI* and *LRTI* also had a negative relationship on the east coast (Figure 6.10).

For the majority of English areas, we were unable to find a relationship that was identified as important by the models between wind speed and NHS111 and GPIH indicators (Figure 6.11).

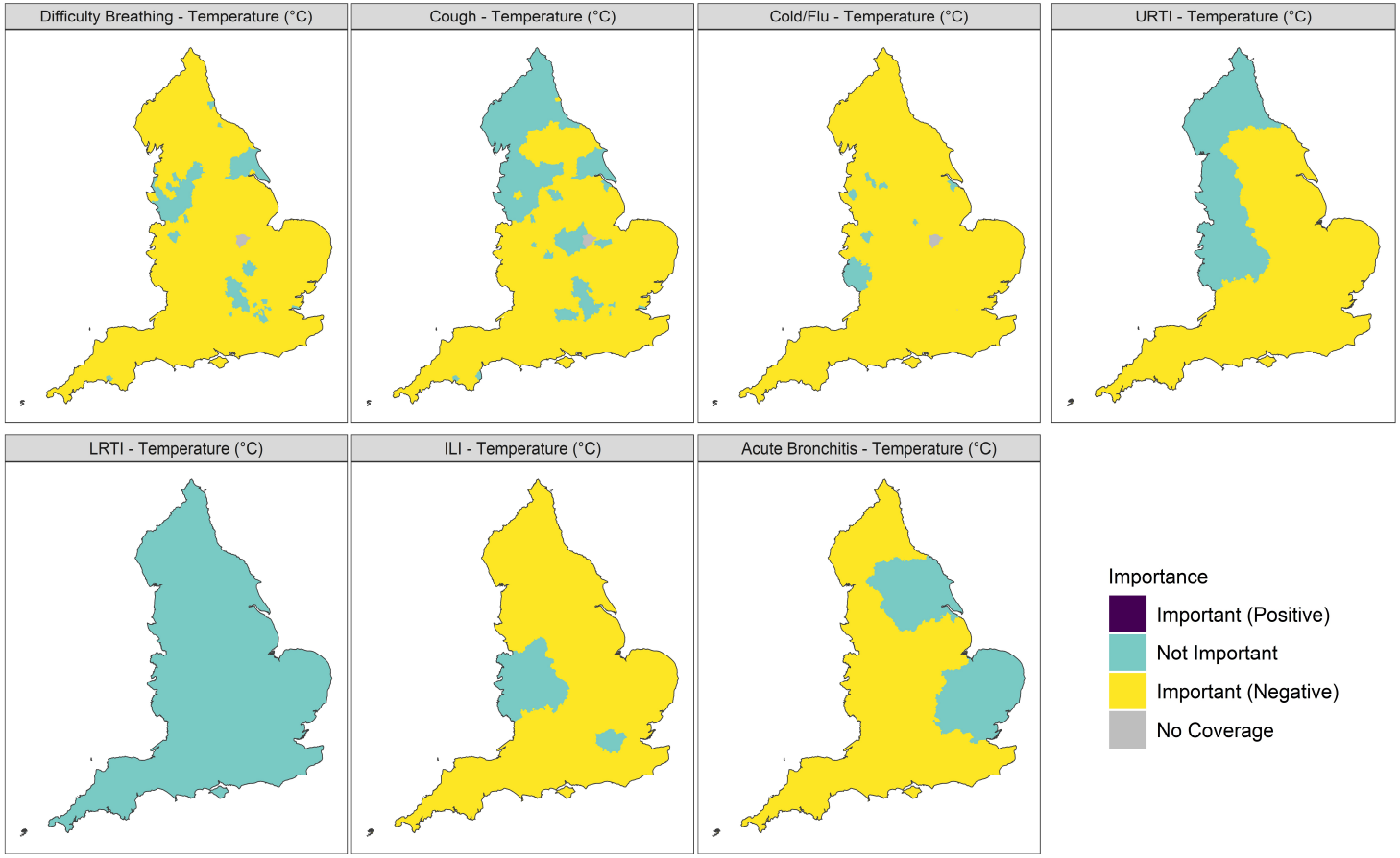

Figure 6.8: Spatial distribution of the relationship between temperature and NHS111 and GPIH syndromic indicators.

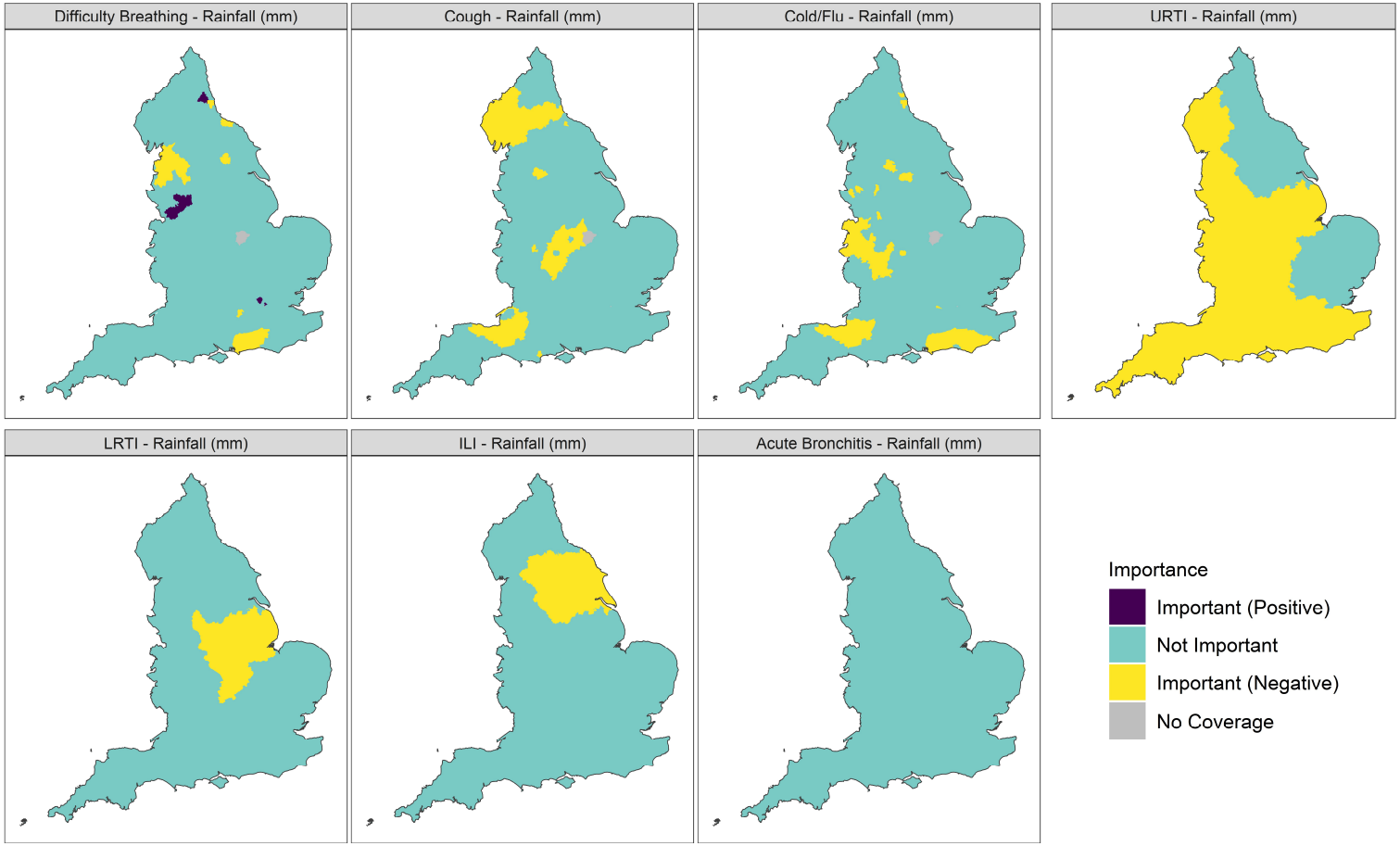

Figure 6.9: Spatial distribution of the relationship between rainfall and NHS111 and GPIH syndromic indicators.

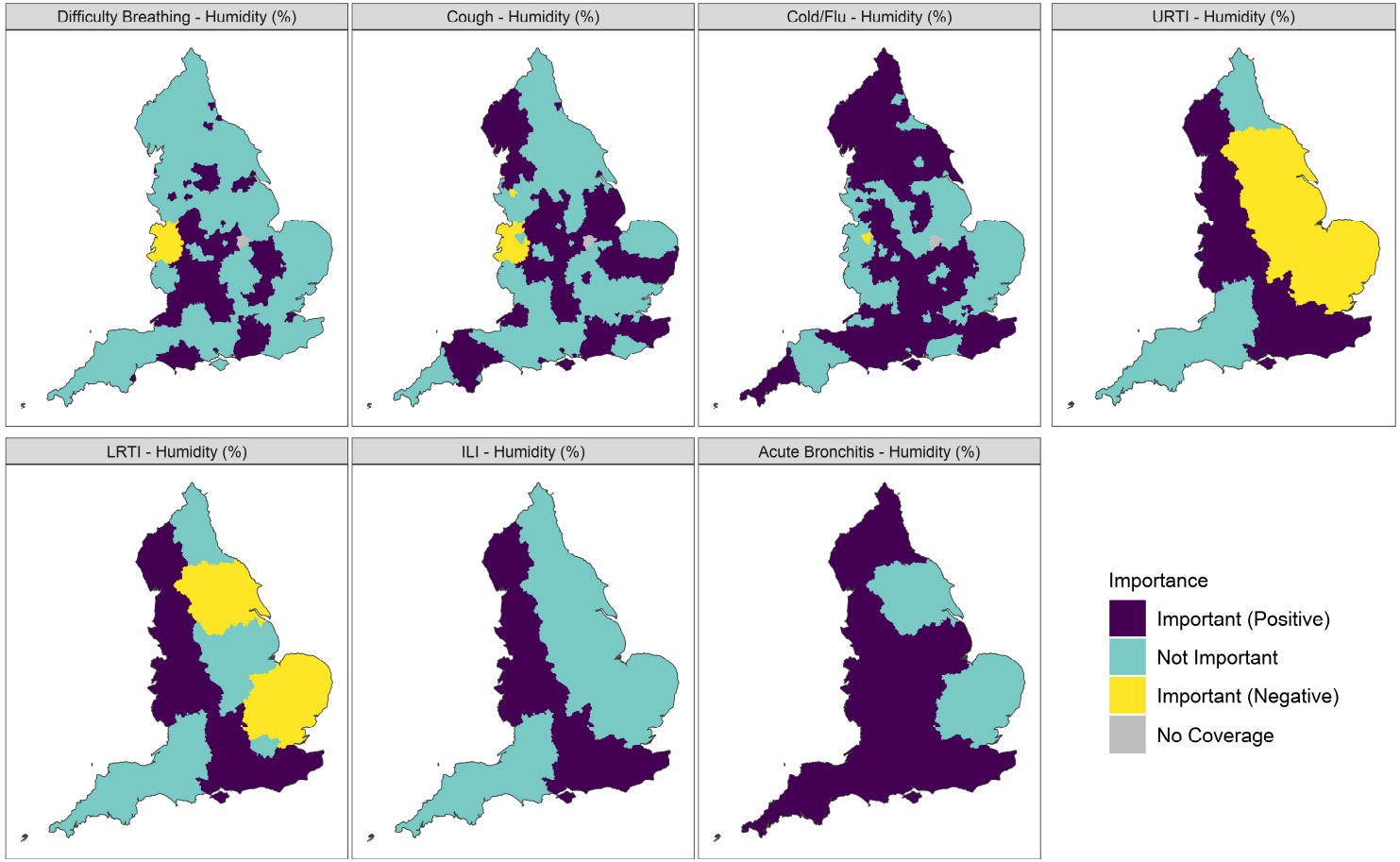

Figure 6.10: Spatial distribution of the relationship between humidity and NHS111 and GPIH syndromic indicators.

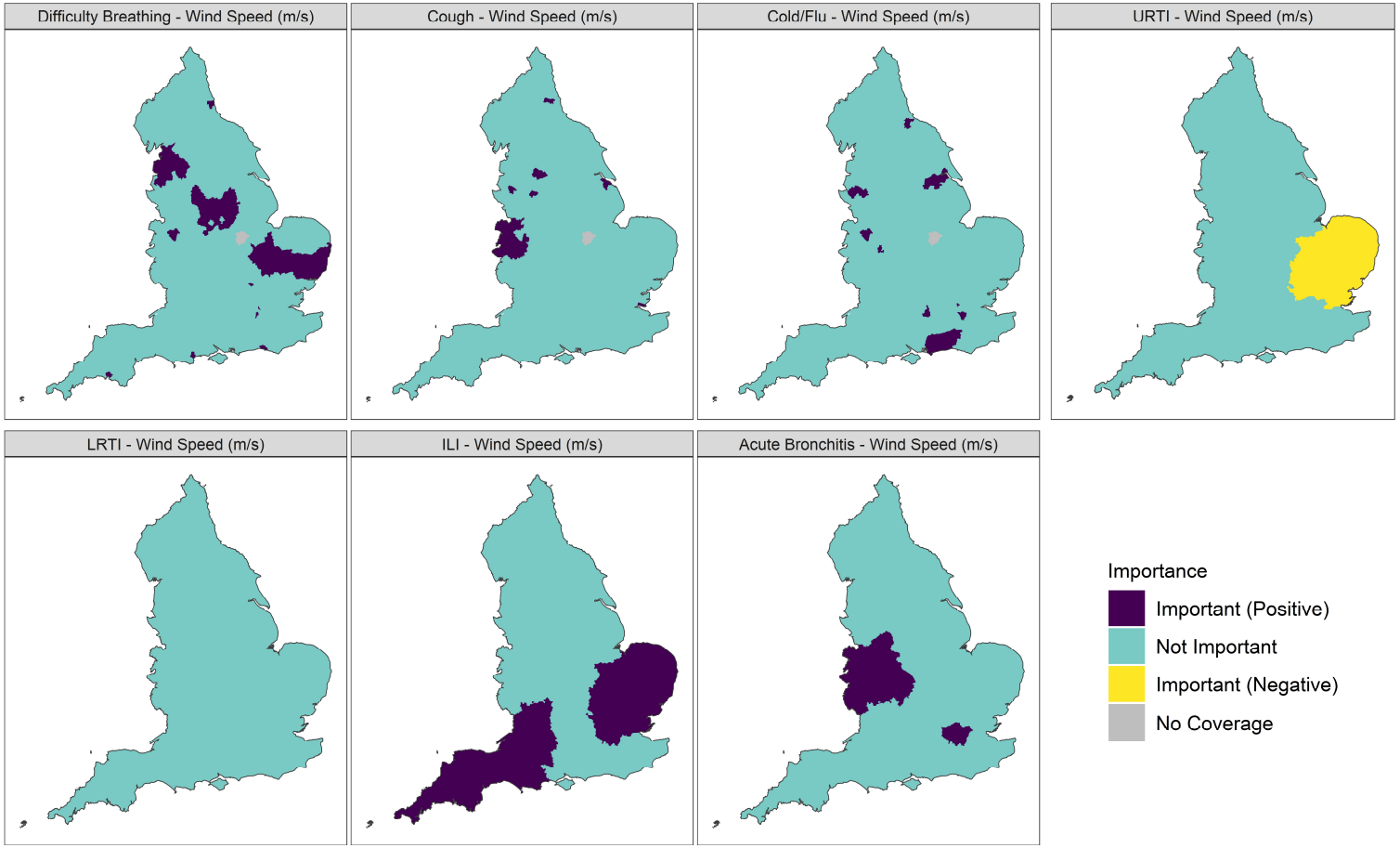

Figure 6.11: Spatial distribution of the relationship between wind speed and NHS111 and GPIH syndromic indicators.

6.6.5 Monthly Varying Covariates

The relationship between month and meteorological conditions was explored for each syndromic indicator to investigate if the relationship between these factors changed over the year.

The monthly relationship between the NHS111 respiratory indicators and temperature was similar (Figure 6.12), with a, mostly, inverse relationship throughout the year but a stronger relationship with temperature was observed in the late spring and summer months, indicating that a 1°C decrease in temperature during this time period led to a higher change in presentation rate compared to a 1°C decrease in temperature during the winter months. The relationship between *Cold/Flu* presentations and humidity was consistent over the year, with a positive relationship. Similar monthly relationships between rainfall and the three NHS111 indicators was observed, with a negative relationship in the winter and a relationship that was not identified as important by the models during the summer months. The monthly relationship between humidity and the NHS111 indicators *Difficulty Breathing* and *Cough* was similar, with humidity having a greater positive relationship on presentation rates in the summer months (June-August) and lower positive relationship during the spring (March – May). The monthly relationship between wind speed and the three NHS111 indicators was largely similar, with a strong positive relationship in the late autumn/winter months and no relationship in the other months (Figure 6.12).

A similar monthly relationship between temperature and the respiratory indicators to GPOOH was observed (Figure 6.13) with a stronger negative relationship in the summer, although for *Bronchitis* this relationship was mostly not identified as important by the models. The monthly relationship between rainfall and the four GPOOH indicators was similar and highly variable throughout the year, with an inverse relationship in early winter. The observed monthly relationship between GPOOH indicators *ARD, ARI,* and *DBWA* and humidity had a stronger positive relationship in the summer months compared to winter and spring. The monthly relationship between *Bronchitis* presentations and humidity was largely not identified as important except

from in late autumn/early winter where it was positive. The respiratory indicators to GPOOH had a similar monthly relationship with wind speed, with a non-important relationship throughout the year apart from in early winter (Figure 6.13).

For the GPIH indicators *URTI* and *LRTI* (Figure 6.14), the monthly relationship with temperature was negative during the summer months, for *ILI* and *Acute Bronchitis* this relationship with temperature was largely not identified as important over the year. For the four GPIH indicators no monthly relationship with rainfall was observed. *URTI, LRTI, ILI* and *Acute Bronchitis* display similar monthly trends in their relationship with humidity, with a negative or non-important relationship for most of the year, and a positive relationship in early winter. No monthly relationship between wind speed and *Acute Bronchitis* and *ILI* was observed. Although the relationship between *URTI*, and *LRTI* and wind speed was not important for most of the year, positive relationships were observed in March and April. A summary of all the results is described in Appendix Table A2.

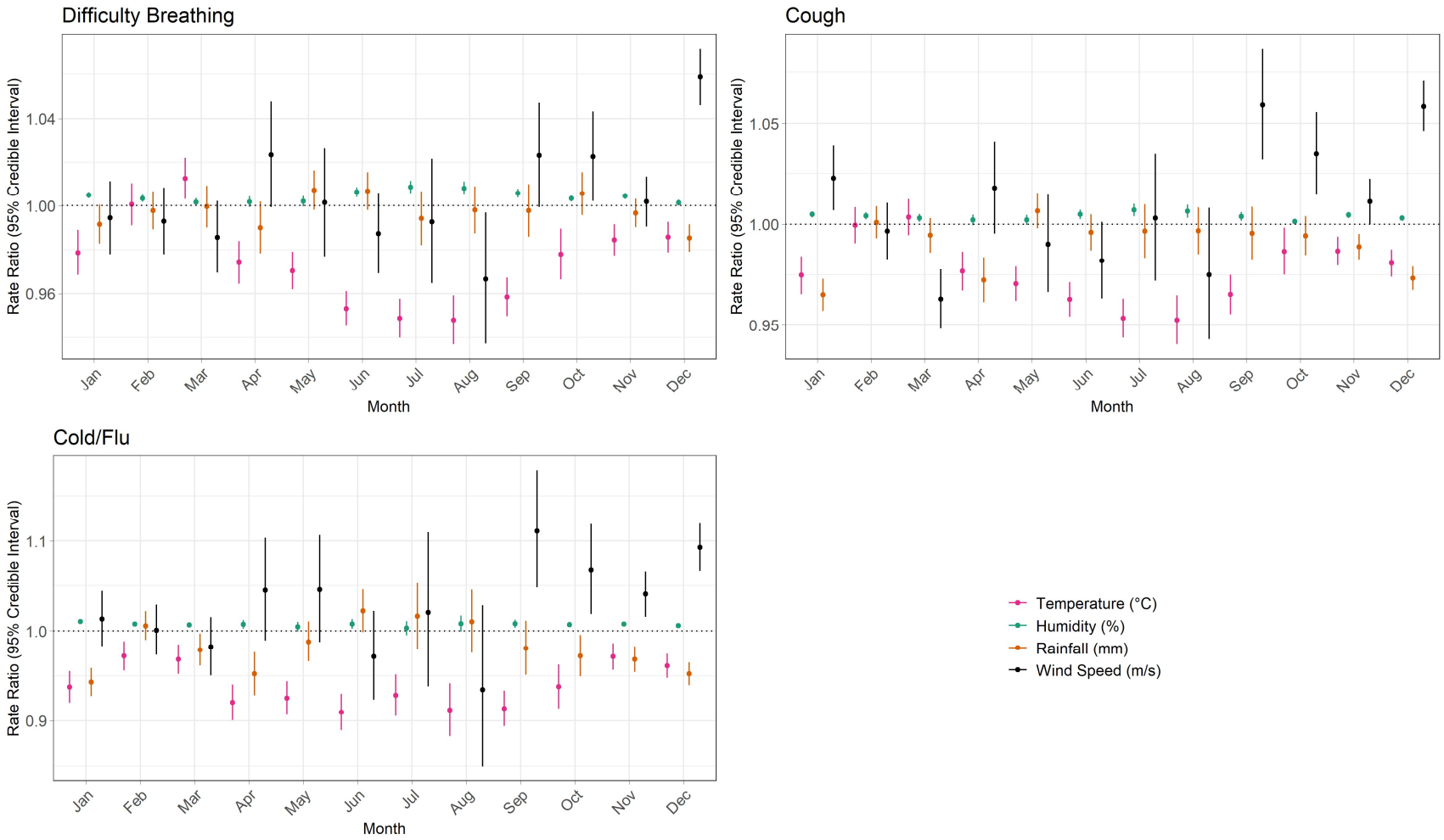

Figure 6.12: Monthly relationship between meteorological conditions and NHS111 syndromic indicators.

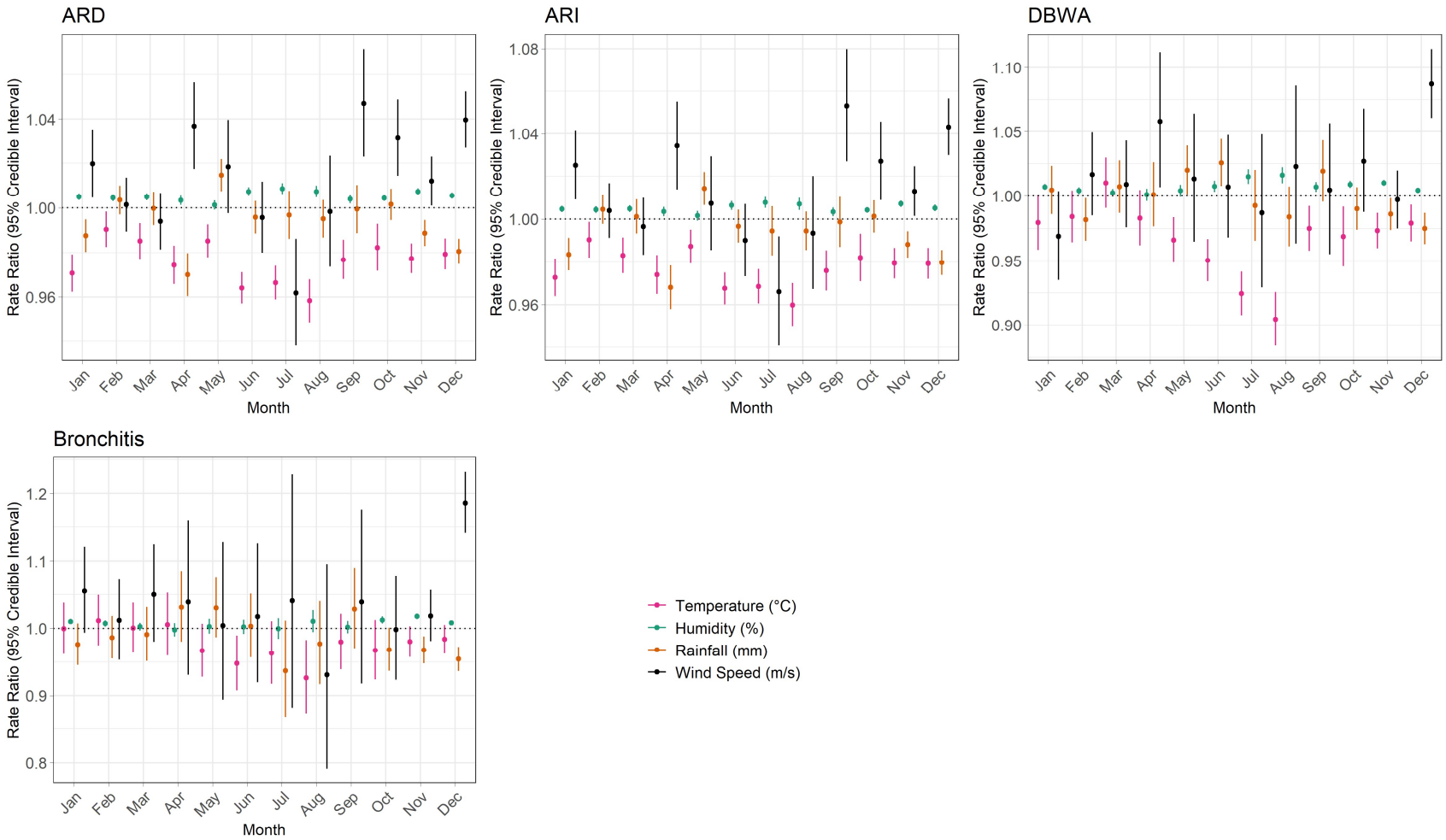

Figure 6.13: Monthly relationship between meteorological conditions and GPOOH syndromic indicators.

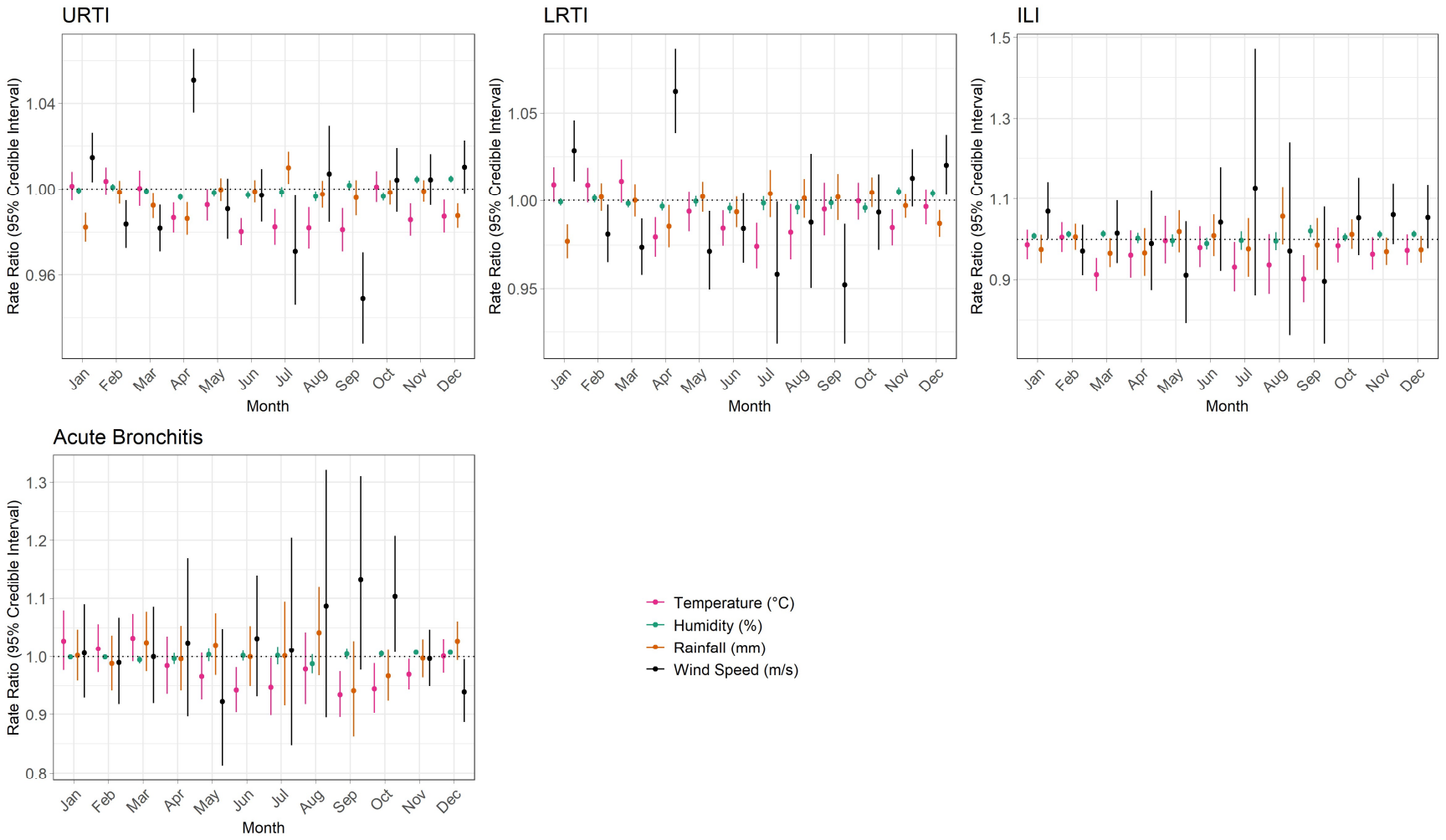

Figure 6.14: Monthly relationship between meteorological conditions and GPIH syndromic indicators.

6.7 Discussion

6.7.1 Impact

Here, we present an exploration of the relationship between meteorological conditions and respiratory presentations to three community healthcare services in England in children under-5 years, using a spatial-temporal Bayesian approach. Three community healthcare services were analysed to allow us to fully explored these relationships in the community oppose to the more severe cases that are captured by more traditional surveillance programmes of laboratory cases and hospitalisations. We also captured healthcare-seeking behaviours from a much larger proportion (between 38% and 91%) depending on the surveillance system) of the English population compared to traditional data sources. Utilising data from three healthcare services has allowed us to compare our findings, adding to the robustness of our observations. Similar previous research has focused on localised areas when exploring the relationship between meteorological conditions and respiratory infections; our use of syndromic data has allowed us to obtain data from the whole of England. We have explored the spatial and monthly distribution of the relationship between meteorological conditions and respiratory presentations associated with infections. The findings from this study, which used data from syndromic surveillance, were similar to research that has used hospitalisation or laboratory data. This highlights how data from syndromic sources may be used when other data is unavailable. Given that syndromic data was available over a large geographic area, while covering a large population, it also allows us to draw wider conclusions compared to data with less population and geographical coverage. The observations from this study can allow policy makers to make targeted decisions on where resources are most effective in relation to meteorological conditions, to improve health outcomes.

6.7.2 Main Findings

The Base Model

The relationship between respiratory presentations to NHS111, GPIH and GPOOH and meteorological conditions were observed to be similar when comparing different respiratory indicators. All respiratory indicators included in the analysis, except *LRTI*, was identified by the models as having an important inverse relationship with temperature. Most previous studies have focussed on hospitalisations, here we estimate relationships using wider burden of respiratory presentations in the community. Our observations are similar to these studies and comparable relationships have been well described in the literature. A decrease in temperature is associated with an increase in the incidence of influenza (Price, Graham and Ramalingam 2019; du Prel et al. 2009), and adenovirus (Price, Graham and Ramalingam 2019; du Prel et al. 2009), HMPV (Price, Graham and Ramalingam 2019; du Prel et al. 2009), coronavirus (du Prel et al. 2009), RSV (Price, Graham and Ramalingam 2019; du Prel et al. 2009; Meerhoff et al. 2009; Yusuf et al. 2007), and presentations due to all ARI (du Prel et al. 2009; Costilla-Esquivel et al. 2014) and bronchiolitis (Hoeppner et al. 2017; Nenna et al. 2017). A decrease in temperature has also been associated with an earlier onset of influenza epidemic timing (Sundell et al. 2016; Jaakkola et al. 2014). There are several factors that may explain the association between temperature and respiratory presentations. Reduced temperatures may extend the period of which an infectious droplet remains airborne, increasing its ability to infect a new host (Sundell et al. 2016; Harper 1961). Low temperatures also appear to be critical for viral stability, with the lipid encasing of viruses, such as influenza, remaining intact for longer at low temperatures, aiding airborne transmission (Polozov et al. 2008). Eccles (2002) proposed that breathing in cold air results in a fall in respiratory epithelium temperature and therefore causes a decrease in the effectiveness of respiratory defences. Changes in social contact behaviour are also known to be an important factor in infectious disease transmission (Mikolajczyk et al. 2008; Mossong et al. 2008; Hens et al. 2009; Eames et al. 2012) as people spend longer time indoors during periods of cold weather (Mccurdy and Graham 2003; Graham and McCurdy 2004) and have longer indoor contact times (Willem et al.

2012). Prolonged social contact has been observed to be dependent on weather conditions, and this type of contact is important for disease transmission and it tends to be more intrusive and involves closer contact (Smieszek 2009).

Rainfall was identified as having an important inverse relationship with most respiratory indicators, except *Difficulty Breathing, DBWA, ILI* and *Acute Bronchitis*. Although the relationship between rainfall and ARI has been previously investigated (Hoeppner et al. 2017; Costilla-Esquivel et al. 2014), but not in studies exploring the wider burden of respiratory presentations in the community, there is contradicting evidence of its relationship with respiratory infections. A study conducted in Mexico has previously linked accumulated rainfall to an increase in ARI consultations (Costilla-Esquivel et al. 2014). Other studies have proposed a link between increased rainfall and the respiratory infections RSV and influenza (Chan et al. 2002; Robertson et al. 2004; Brooks et al. 2007). Although study conducted in India observed an inverse relationship between rainfall and RSV, but a positive relationship with influenza (Agrawal et al. 2009). Most of these studies have focused on countries with tropical climates, therefore their relevance to the UK may be uncertain. It is unclear why we have observed an inverse relationship between rainfall and respiratory presentations, research has proposed that increased rainfall leads to more time spent indoors, increasing prolonged exposure to others (E. Murray et al. 2012), but this does not explain our findings. Further research is required to understand this relationship. More research needs to be conducted in temperate countries to explore the relationship between ARI and rainfall to explore if there are differences due to geography.

A positive and important relationship was identified by the models between humidity and wind speed, and all respiratory indicators apart from *URTI* and *LRTI*. It is unexpected that *LRTI* and *URTI* were not associated with humidity and wind speed, as the other respiratory indicators were associated with these conditions. These observations could be due to the lack of specificity of these indicators or the highly aggregated nature of the data at PHE level. A similar relationship with humidity has previously been described in the literature for all ARI (du Prel et al. 2009), RSV (du Prel et al. 2009; Hervás, Reina and Hervás 2012; Meerhoff et al. 2009; Yusuf et al.

2007; Welliver 2007) and its associated symptoms (Hoeppner et al. 2017; Nenna et al. 2017), influenza, rhinovirus, and adenovirus (Price, Graham and Ramalingam 2019). Higher relative humidity is associated with increased survival of RSV and influenza aerosols. Higher humidity promotes virus survival in aerosols by slowing the evaporation of droplets (Paynter 2015). There is less research into the relationship between respiratory infections and wind speed, compared to those that investigate temperature and humidity. But previous research has described a similar relationship between wind speed and bronchiolitis hospitalisations (Hoeppner et al. 2017) and hospitalisations due to ARI (du Prel et al. 2009). A study by Feng et al. (2020) investigated the influence of wind on COVID-19 airborne transmission, they observed that microdroplets of virus can transport in the air further due to wind convection. Higher wind speeds could influence the airborne distribution and spread of respiratory pathogens, but further research is required to explore this mode of transmission. Periods of high wind may also influence social behaviour, with people more likely to spend prolonged periods inside, where pathogen transmission is more likely to occur (Smieszek 2009).

A notable consistency with the literature is the meteorological conditions that were associated with the RSV manifestation bronchitis. In our study, both bronchitis indicators from GPOOH and GPIH services were negatively associated with temperature and rainfall, and positively associated with humidity. These associations have been previously described in the literature for both RSV and bronchitis (Price, Graham and Ramalingam 2019; du Prel et al. 2009; Hervás, Reina and Hervás 2012; Meerhoff et al. 2009; Hoeppner et al. 2017; Nenna et al. 2017).

In summary, seasonal patterns of respiratory infections appear to be influenced by several contributing factors including changes in temperature, absolute humidity, sunlight, vitamin status, and host behaviour (Dowell 2001; Fisman 2012; Cannell et al. 2006; Sloan, Moore and Hartert 2011; Shaman and Kohn 2009; Azziz Baumgartner et al. 2012; Tamerius et al. 2011), it is likely that each of these factors play a modulating influence on disease transmission. Further research is required to determine the causative pathway between meteorological conditions and respiratory infections. Here

we estimated relationships using wider burden of respiratory presentations in the community, but the results were comparable to other studies which have estimated the relationship between meteorological conditions and respiratory infections.

Spatially Varying Covariates

The spatial distribution of the relationship between the respiratory indicators and meteorological conditions were explored to investigate whether specific areas were more susceptible to changes in meteorological conditions. There were spatial patterns between temperature and respiratory indicators from both NHS111 and GPIH services. Yusuf et al. (2007) explored the relationship between temperature and RSV in ten cities spanning different latitudes and climatic regions. A strong relationship between temperature and RSV was observed in eight of the cities, with all but one displaying a negative relationship. This highlights how temperature can have different influences in different climatic regions. Here we observe a number of spatial patterns in the relationship between the respiratory indicators and temperature than may be useful for hypothesis generation. However, the spatial patterns do not appear consistent between indicators making overall conclusions challenging.

URTI was the only indicator to have a clear pattern in spatial distribution with rainfall; which was observed to have an important inverse relationship in most regions apart from those in the east. There is very little literature describing region differences in the relationship between rainfall and respiratory infections, therefore it is difficult to interpret why these relationships were observed.

In contrast to the other meteorological indicators, spatial patterns were observed between humidity and respiratory indicators. However, for GPIH data, there was a consistent pattern of apositive association with humidity in the west and south east for all indicators. The patterns were less consistent for NHS111. In England, the west coast tends to be more humid compared to the east (Met Office 2020) which could explain the observations for the GPIH data.

Previously, we observed that windspeed had a positive relationship with all respiratory indicators except *URTI* and *LRTI,* however, no spatial patterns between wind speed and

the respiratory indicators were observed. This suggest that no area is particularly vulnerable to higher presentations of respiratory infections as a result of wind speed.

The spatial distribution of the relationship between respiratory infections and meteorological conditions is not well explored in the literature. Many patterns were observed in our data, but interpreting these was challenging as they often varied by indicator with patterns observed, however there was no consistency. This made it difficult to draw conclusions about the patterns were observed. These spatial patterns may be emerging due to regional differences in social behaviour, regional climatic differences or differences in meteorology across England. More research is required to explore why these spatial differences are being observed.

Monthly Varying Covariates

Most respiratory indicators demonstrated an inverse relationship with temperature in the autumn and/or winter. Surprisingly, the more specific indicators; *ILI,* and *Bronchitis* was observed to have no relationship with temperature in the autumn and winter. It is unclear why we have observed this as it would be expected that there would be a negative relationship with *ILI* and *bronchitis* in autumn winter. These observations could be due to small number of presentations (with 16,863 *ILI* and 14,267 *Acute Bronchitis* presentations over the study period compared to 3.3 million *URTI* presentations) for these indicators resulting in large uncertainty around the estimates. Another surprising observation was that for the indicators *Difficulty Breathing, Cough, Cold/Flu, ARD, ARI, DBWA*, *URTI* and *LRTI* the negative relationship with temperature was strongest in summer. Respiratory syndromes that are associated with the summer have been related to higher temperatures, such as hay fever or asthma (Soneja et al. 2016; Bodaghkhani et al. 2019; Upperman et al. 2017; Ziska et al. 2019). This observation may reflect host behaviours with people more likely to stay inside when it is colder, where they may be exposed to more allergens (Gaffin and Phipatanakul 2009; Sheehan and Phipatanakul 2016) or are able to spread disease through close contact (Zhang et al. 2020; Pica and Bouvier 2012).

Difficulty Breathing, Cough, Cold/Flu, ARD, ARI, DBWA, Bronchitis, URTI, and *LRTI* were observed to have an inverse relationship with rainfall, mainly in the autumn and winter. There is little in the literature regarding the relationship between rainfall and respiratory syndromes. Given there is very in the literature on this topic it is difficult to hypothesis why this relationship was observed.

Cold/Flu was observed to have a consistent positive relationship with humidity throughout the year. For *Difficulty Breathing, Cough, ARD, ARI* and *DBWA* humidity was observed to be positive for most months, but the relationship was stronger in the summer, with a higher change in presentations rate for each 1% increase in humidity. This could be because these indicators are detecting non-infectious respiratory symptoms that are worse in the summer due to high humidity, such as asthma (Lam et al. 2016; Romaszko-Wojtowicz et al. 2020). This contrasts with *URTI, LRTI, Bronchitis, Acute Bronchitis*, and *ILI* where an positive relationship in the autumn and winter was observed. The positive autumn and winter relationship for *Acute Bronchitis, Bronchitis, URTI* and *LRTI* may be due to the strong association between these indicators and RSV, influenza and other respiratory infections. The relationship between humidity and both the onset of outbreaks, and the number of presentations of respiratory infections such as RSV and influenza has been well described in the literature (Price, Graham and Ramalingam 2019; du Prel et al. 2009; Hervás, Reina and Hervás 2012; Meerhoff et al. 2009; Yusuf et al. 2007; Welliver 2007). In temperate climates such as England these viruses circulate in the late autumn and winter (Janet, Broad and Snape 2017; Fleming, Cross, and Pannell 2005b). The less specific indicators: *Difficulty Breathing, Cough ARD, ARI* and *DBWA* demonstrated a positive relationship with humidity in both winter and summer, but this relationship was stronger in summer.

Wind speed was observed to have a positive relationship with *Difficulty Breathing, Cough, Cold/Flu, ARD, ARI, DBWA*, and *Bronchitis* in the late autumn and winter months. Spikes of importance were observed in Spring for *URTI, LRTI ARD*, and *ARI.* There has been little research into the relationship between wind and respiratory pathogens. Especially in regards to when the relationship between these factors is

strongest. We observed that the relationship between respiratory indicators and wind was strongest in the late autumn and winter. We hypothesis that during this period respiratory pathogens such as influenza and RSV are in circulation, and during period of high wind people may congregate indoors, where higher rates of spread of these pathogens can occur (Zhang et al. 2020; Pica and Bouvier 2012).

Respiratory season pathogens that have a large health burden in children and have been observed to circulate at different times during the year in temperate regions and have different relationships with meteorological conditions, with influenza, RSV and coronavirus circulating in the winter, rhinovirus and HMPV circulating in the spring and autumn, non-rhinovirus enteroviruses circulating in the summer and adenovirus circulating all year round (Moriyama, Hugentobler and Iwasaki 2020). One of the interesting observations from this study is that for many of the respiratory variables the relationship with temperature and humidity had a stronger influence in the summer months. There could be several factors that contribute to this observation. The viruses that circulate in the summer and spring, HMPV, rhinovirus and non-rhinovirus enteroviruses may be more influenced by meteorological changes compared to those that circulate at different times of the year (Tamerius et al. 2011; Midgley et al. 2017; Landes et al. 2013; Morikawa et al. 2015; Lee et al. 2012; Monto 2002; Abedi et al. 2018; 2016; Moriyama, Hugentobler and Iwasaki 2020). The syndromic indicators used in this study are associated with different diseases at different times of the year, adding to the complexity in interpreting these findings. For example, *Difficulty Breathing*, an indicator from the NHS111 system, is associated with RSV and influenza in the winter (Morbey et al. 2017a; chapter four), but when these viruses are not circulating in the summer this indicator may be more associated with other illnesses such as asthma, heat related illness and hay fever. Fluctuations in meteorological conductions could result in similar social behaviours in both summer and winter, for example a drop in temperature may result in people spending more time indoors where disease transmission is more likely (Zhang et al. 2020; Pica and Bouvier 2012).

6.7.3 Limitations

The data obtained from the healthcare services monitored by the syndromic surveillance service are pre-diagnostic and do not directly relate to specific pathogens. Although from previous research we can infer which pathogens are associated with the syndromic indicators used in this study, we cannot make a direct inference about the relationships observed and respiratory pathogens. The literature also states that respiratory pathogens circulate at different times. Therefore, the relationships observed between the syndromic data and meteorological conductions may result from different pathogens at different times of the year. Some of the syndromic indicators used are more likely to be associated with pathogens, for example *Bronchitis* is highly correlated with RSV in children under-5 years. Whereas other syndromic indicators are less specific and are likely to detect multiple pathogens; for example, *URTI*, or may detect both respiratory pathogens and non-infections respiratory symptoms such as *Difficulty Breathing*. Therefore, inference of the relationship between the indicators and meteorological conditions needs to be taken in the context of each syndromic indicator. However, irrespective of aetiology respiratory presentations are still experienced by individuals and put pressures on healthcare systems, therefore it is important to explore the factors that can influence these presentations.

At the community level there are multiple pathways to obtain healthcare and people may seek it in different ways and for different reasons. We have tried to counteract these effects by including three commonly used healthcare services in the community, to capture different pathways of access. We could not account for those who did not present to healthcare services, or those who presented to services not included in the analysis, such as hospitals or pharmacies.

GPIH data was only available at PHE centre, a geography of only nine very coarse regions, and so. the meteorological conditions included in the analysis at this geography were highly aggregated. This situation made comparisons to the NHS111 and GPOOH data difficult and may have impacted the relationship between the GPIH data and meteorological conditions. Although the underlying population for GPIH was known

(the number of registered patients), because GPIH data was not available at a geography smaller than PHE centre, we could not determine how representative the study population was of each region. We could not investigate potential non-linear relationships between GPIH data and the meteorological conditions, likely due to the highly aggregated nature of the data at PHE centre level. Although GPOOH data was available at a low level of geography, the coverage of this system was not equally distributed over England. Consequently, we could not investigate the spatial distribution of the relationship between GPOOH data and the meteorological conditions.

Seasonal adjustments were made through the inclusion of a random effect for year and week to control for unobserved seasonal confounders. These components may have resulted in an underestimation of the relationships with the meteorological conditions. Other studies that have explored the relationship between ARI and meteorological conditions have not accounted for unobserved seasonal effects (Price, Graham and Ramalingam 2019; du Prel et al. 2009; Meerhoff et al. 2009; Yusuf et al. 2007; Hoeppner et al. 2017). However, if seasonal trends were not accounted for in the model, we would have to assume that all the seasonal variation in the data is due to the known variables included in the model, which may not be correct. Given the inclusion of these seasonal covariates, we still observed similar relationships between ARI and metrological conditions to what is described in the literature. It requires a careful balance when accounting for seasonality as to not over or underestimate its influence. Despite these limitations, this study provides evidence to support the influence of meteorological conditions on respiratory presentations to community healthcare services in children under 5 years.

This study has focused on children under-5 years due to their vulnerability to meteorological conditions and respiratory infections. Observations may reflect associations that are unique to this age group. Therefore, further research is required to investigate if these relationships hold true in other age groups.

6.7.4 Conclusion

We present a comprehensive analysis of the influence of the meteorological conditions: temperature, humidity, rainfall and wind speed on respiratory presentations to three community healthcare services in England in children under-5 years. We use one of the most comprehensive sets of respiratory syndromic surveillance data assembled for England comprising seven million reports from a telehealth system and two General Practitioner systems. Analysis was conducted using a spatial-temporal Bayesian approach which allowed for the inclusion of spatial and temporal dependency structures. Our results highlight the influence of meteorological conditions on respiratory presentations across England. Most respiratory indicators studied have a negative relationship with temperature and rainfall, and a positive relationship with humidity and wind speed. We present a novel aspect of research by exploring changes in the spatial and temporal associations between these indicators and the meteorological conditions. This allows us to identify areas that are of particular risk to fluctuations in meteorological conditions and the time periods where changes in meteorological conditions have the strongest influence. These results can guide policymakers and health professionals in their preparation and targeting of public health measures and to inform decision-making and planning in public health processes. These observations may also be used to understand how a changing climate may impact respiratory infections in the future.

Chapter 7

Discussion

7.1 Chapter Findings

In chapter three we investigated the demographic and socioeconomic patterns (age, gender and deprivation) of healthcare-seeking behaviour due to respiratory illnesses in England, using over 13 million respiratory related presentations to NHS111 (telehealth service), GPIH (general practitioner consultations in hours), and GPOOH (general practitioner consultations out-of-hours) services. It was observed that those from more deprived backgrounds and young males (under-5 years) were most likely to present to the three healthcare services studied. These observations were compared to non-respiratory presentations to establish if the findings were specific to respiratory diseases, or due to a wider trend in healthcare-seeking behaviours for all illnesses. Comparable results, in terms of which factors and their magnitude were observed between respiratory and non-respiratory presentations, suggesting similar influences on healthcare-seeking behaviour between respiratory and non-respiratory presentations.

Given the high burden of respiratory syncytial virus (RSV) in children under-5 years and the lack of estimates of disease burden in the community, in chapter four, we used syndromic surveillance data from three healthcare services (NHS111, GPIH and GPOOH) to estimate the attributable RSV burden in England in the community. We estimated there was almost 200,000 RSV attributable presentations to NHS111, GPIH and GPOOH per year in children under-5 years over the study period. Higher presentation rates were observed in children under-1 year compared to those of 1-4 years. In comparison to influenza, more presentations were attributable to RSV. For every confirmed case of RSV notified by laboratories we estimated there was up to 69 healthcare presentations to NHS111, GPIH or GPOOH in children under-5 years.

In temperate countries, such as England, RSV has distinctive winter activity. The timing of RSV activity is associated with latitude, as well as environmental conditions such as air pollution, temperature, rainfall and humidity. Given this, in chapter five we attempted to establish if there were regional differences in RSV activity (burden, and outbreak peak week, start week and length) in England by using data from syndromic surveillance of NHS111 and GPIH. With the metrics we used there was little evidence for regional patterns in the timing of RSV activity, although burden was observed to be highest in the southern regions of England.

In chapter six we estimated the relationship between the meteorological conditions: temperature, rainfall, humidity and windspeed, and acute respiratory infection related presentations to NHS111, GPIH and GPOOH in children under-5 years. To fully explore these relationships, we investigated both linear and non-linear relationships, as well as the spatial and monthly associations. We observed that the majority of respiratory indicators studied had a statistically important negative relationship with temperature and rainfall (presentations increased with decreasing temperature and rainfall), and statistically important positive relationships with humidity and wind speed (presentations increased as humidity and wind speed increased). The monthly influence of these meteorological conditions was explored and we observed that for many of the indicators, temperature and humidity had an influence on presentations all year round. Although some spatial relationships were observed between the respiratory syndromic indicators and meteorological conditions, we concluded these were, currently, of little public health importance.

7.2 Recurring Themes in Syndromic Surveillance

Healthcare-seeking behaviour vs healthcare need

Syndromic surveillance primarily detects healthcare-seeking behaviours which is not the same as healthcare need, as we cannot establish a confirmed diagnosis or the severity of illness from the data. This is highlighted in chapter three where we observed similar

social and demographic patterns in healthcare presentations in both respiratory and non-respiratory presentations. This suggested that the demographics of healthcare-seeking behaviours may have an influence on the way healthcare is sought, irrespective of symptoms.

Healthcare-seeking behaviour is the process by which a person perceives themselves or a dependant to have a health problem, and the action they undertake to find an appropriate healthcare solution. Healthcare need is where an individual's health would benefit from healthcare. Individuals may have different perceptions of what their healthcare need is, leading to different healthcare-seeking behaviours (MacKian 2003; Latunji and Akinyemi 2018; Olenja 2003; Poortaghi et al. 2015). This decision-making process is influenced by two main factors: i) how severe or potentially severe the illness is perceived to be, and ii) how effective the healthcare-seeking behaviour is perceived to be (Oberoi et al. 2016). Other factors not related to illness including sex, age, race, socio-economic status, cultural background and their lived experiences of health and healthcare may also influence healthcare-seeking behaviour (Oberoi et al. 2016). A review by Zeng and Wagner (2002) proposed that healthcare-seeking behaviours comprised of four phases: recognition of symptoms, interpretation of symptoms, cognitive representation of illness, and seeking treatment. During each of these phases the final outcome of seeking treatment can be influenced by demographic, cultural, economic and psychological factors, and can vary over time. Given the anonymous and passive nature of syndromic surveillance the only factors we could explore in relations to healthcare-seeking behaviour were sociodemographic (age, gender and deprivation).

External factors can also influence if and how healthcare is sought by an individual. For example, in 2015, media reporting of a possible exposure of *Cryptosporidium* in drinking water in North West England led to a 28.5% increase in rate of presentations due to diarrhoea to GP services, despite no laboratory cases linked to drinking water being detected (Elliot et al. 2016b). The media's impact was also observed during a mumps outbreak in England and Wales between 2003 and 2004. The incidence rate of mumps appeared to be influenced by the level of newspaper coverage, leading the authors to hypothesis that the increased coverage led to an increased awareness of both

the public and clinicians (Olowokure et al. 2007). The recent COVID-19 pandemic has also highlighted the impact of external influences on healthcare usage and healthcare-seeking behaviour. There was increase in media reporting, vast changes to the advice on how to seek healthcare, and stay-at-home orders were put in place to prevent the spread of disease. At the same time there were large changes in national health service (NHS) activity, changes in patient behaviour, and a reduction in other infectious diseases (The Health Foundation 2020b). Changes in healthcare usage were highlighted by a 57% reduction in accident and emergency visits in April 2020 compared to April 2019 in England (The Health Foundation 2020b). Although this reduction will in part be due to changing patterns of need, reductions in infections due to stay-at-home orders and patients being directed elsewhere, there are concerns people are delaying seeking healthcare to avoid being infected with COVID-19, which may lead to poor long-term outcomes particularly in cancer and cardiovascular disease (The Health Foundation 2020a). These examples emphasise how complex the decision to seek healthcare is, what the influences are, and how it can be difficult to determine at what point individuals will seek healthcare.

Given that the data used in this thesis is aggregated and anonymised we cannot ascertain why healthcare is being sought. Individuals' presenting with the same symptom, for example *Cough,* might have different levels of disease severity and perception of the severity of their disease. Their perceived illness severity, as well as a complex mix of other factors, lead to the decision to seek healthcare. Therefore, we were unable to separate healthcare need from healthcare-seeking behaviours in our data. This is highlighted in our observations from chapter three, where similar demographics (young males, all females and those from more deprived backgrounds) had higher presentation rates for both respiratory and non-respiratory illness. This was a somewhat surprising result and could be due to similar aetiologies for disease in their demographics or similar healthcare-seeking behaviours, or both. Sex differences in morbidity and mortality in children has been observed to be due to genetic and biological factors, with young males seen as biologically weaker (Waldron 1983). Adult females may be more likely to seek healthcare due to sex specific concerns, such as pregnancy and gynaecological issues, or be more likely to present to healthcare for the same illness compared to males due to

gender specific factors (Thompson et al. 2016; Mackenzie, Gekoski and Knox 2006; Matheson et al. 2014; Nabalamba and Millar 2007; Galdas, Cheater and Marshall 2005; Cornally and McCarthy 2011; Deeks et al. 2009). A review by Galdas, Cheater and Marshall (2005) found there was a growing number of studies highlighting delayed healthcare-seeking behaviour in men which leads to worse health outcomes, and they hypothesised that was due to masculine beliefs about seeking help. Those from more deprived backgrounds are more likely to suffer ill health as a result of their environment, but may be less likely to seek healthcare compared to their less deprived counterparts (Fiscella and Holt 2007; Haroon, Barbosa and Saunders 2011; Edwards and Pill 1996). The research in this thesis comprises of ecological studies, whereby associations have been observed between certain demographics and symptoms of disease, further research is required to investigate the specific behaviours and decision that drive individuals to seek healthcare from the services monitored by syndromic surveillance.

Complementing Laboratory Surveillance

Surveillance of diseases can be separated into two categories, disease-specific and syndromic (Abat et al. 2016). Disease-specific surveillance is typically diagnostic and includes: laboratory data, statutory infectious disease notifications, hospital episode statistics and discharge data, and it monitors specific diseases or public health threats (Abat et al. 2016; Koski 2011; Wagner, Moore and Aryel 2011). Syndromic surveillance is, typically pre-diagnostic, and data sources can range from medical records, telehealth systems, pharmacies, web searches and absenteeism from work or school (Triple-S Project 2011; H. Chen, Zeng and Yan 2010). Both these types of surveillance play a vital role in the surveillance of public health threats, but it important to recognise their strengths and limitations to understand the role they play. Disease-specific surveillance, typically, uses standardised tools and diagnostic criteria for diagnosis of the target disease, allowing for global trends and comparisons of trends between different geographical locations. Prior to starting disease-specific surveillance, pathogens and diseases need to be clearly defined in order to develop diagnostic criteria and tools to target these diseases. The use of laboratories is costly, limiting its use in low-income countries, they require highly trained staff to run effectively and the

diagnostic ability is limited to the capacity of the laboratory. In disease-specific surveillance good reporting systems need to be in place once a diagnosis in made in order to capture cases of disease. In the United Kingdom (UK) there is a statutory duty to notify public health authorities of suspected cases of specified notifiable diseases and organisms (PHE 2021). However, these cases need to be actively reported to the notification system, and although statutory, under-reporting can occur, negatively impacting the effectiveness of these surveillance systems (Brabazon et al. 2008; Davison et al. 2003; Herbert et al. 2015). Limited laboratory capacity, lack of trained staff, and poor reporting systems can result in the true prevalence of the disease of interest being underestimated (Abat et al. 2016). Although syndromic surveillance is not disease-specific, steps can be taken to capture and utilise symptoms that have strong associations with specific disease. When the surveillance systems that are operated by the real-time syndromic surveillance team (ReSST) in Public Health England (PHE) were first developed, syndromic indicators of interest were selected by trained medical personnel based on common symptoms of diseases of interest. For example, the indicator *Influenza-Like-Illness* from the GPIH surveillance system was selected for the suitability to detect cases of influenza. As previously discussed, many diseases have similar symptoms, therefore these indicators will be associated with multiple diseases. Analysis can be conducted to estimate the association between syndromic indicators and specific diseases. In chapter four, we used regression analysis to identify which syndromic indicators were most associated with RSV and influenza in children under-5 years. Morbey et al. (2017a:2017b; 2018) used similar statistical techniques to estimate which respiratory indicators were most associated with respiratory infections. Although these steps can help us infer which specific diseases might be causing trends in the data, care needs to be taken as other infectious diseases or non-infectious diseases can also cause similar symptoms, and these symptoms can differ by age. In some cases, it may not be possible to distinguish between specific diseases due to their similar symptoms.

In chapter four we used the example of RSV in children to explore how syndromic surveillance can be used to estimate disease burden. Data were obtained from three community healthcare services (NHS111, GPIH and GPOOH). This allowed us to gain

a wider estimate of RSV burden in the community compared to the other data sources such as GP and hospital attendances, and laboratory notifications (PHE 2019). Although these estimates provided a wider burden of disease, we could not account for those who did not seek healthcare. The majority of children are infected with RSV by two years (Andeweg et al. 2021), with a high burden of disease in children under-5 years (Fleming et al. 2015; Goldstein et al. 2018; Shi et al. 2017), there are several risk factors that increase the risk for hospitalisation and severe outcomes due to RSV. These risk factors include being aged under-1 year, underlying respiratory and cardiovascular disorders, low birth weight, and preterm birth (Cai et al. 2020). With the majority of laboratory reports coming from hospitals, this introduced significant bias into the laboratory surveillance system, with severe illness being over represented in laboratory surveillance data for RSV. This is highlighted in our research where the ratio between laboratory cases and presentations to NHS111, GPIH and GPOOH was higher in the 1-4 age group compared to the under-1 age group who are more likely to be captured by traditional surveillance. Syndromic surveillance can be used to detect and monitor disease in populations that might be overlooked by traditional surveillance systems.

One of the main strengths of syndromic surveillance is its timeliness, it utilises data sources which are often used before an official diagnosis is made, therefore public health threats can be detected earlier (Triple-S Project 2011). Given it does not require high-cost laboratory facilities, syndromic surveillance can be deployed in low-income countries provided they have the human and technological resources required. Syndromic surveillance has low specificity and high sensitivity, as it deploys a broad-brush approach of using non-specific indicators of disease (Lateef 2012). This results in surveillance detecting those with the targeted disease, as well as those who do not have the targeted disease. The level of these false positive detections will depend on the disease of interest (with those with more specific symptoms having lower false positive rates), and the amount of information available and the analysis undertaken (Abat et al. 2016; Lateef 2012).

Under-ascertainment occurs when someone does not seek healthcare for their illness, this can be due to lack of or mild symptoms, a self-limiting illness or difficulty accessing

services. Under-ascertainment can also be influenced by health literacy, healthcare availability and culture (Gibbons et al. 2014). The level of under-ascertainment varies considerably based on the type of illness, as well as population affected; although the level of under-ascertainment varies by disease, a significant proportion of people with infections do not seek healthcare. These undetected cases will not be notified by either syndromic or diagnostic surveillance, consequently a proportion of the true burden of disease will be missed by routine surveillance.

Under-reporting of disease occurs when those with an illness seek healthcare, but the event is not captured by surveillance. Under-reporting can occur due to multiple factors; a disease might not be reported due to lack of knowledge by healthcare professionals, budget constraints or restrictions in laboratory usage. Furthermore, the surveillance systems might not be in place for reporting certain diseases. During disease outbreaks, reporting may be more likely due to heightened awareness and need for diagnosis (Sethi et al. 1999; Gibbons et al. 2014; Hardnett et al. 2004; MacDougall et al. 2008; O'Brien et al. 2010). In chapter four we observed that for every laboratory confirmed case reported in children under-5 years, 12.4, 20.7, 25.8 RSV related presentations were detected by NHS111, GPOOH and GPIH syndromic surveillance systems, respectively. This observation highlights how syndromic surveillance can be used to complement other forms of surveillance to reduce the level of under-reporting.

Both diagnostic and syndromic surveillance have their strengths and weaknesses, but both can be used together to strengthen public health surveillance which monitors and detects public health threats.

What are we detecting?

By definition, syndromic surveillance monitors people seeking healthcare information, such as through internet searches, or those presenting to healthcare services, such as family practitioners, with signs and symptoms of disease rather than confirmed clinical or laboratory diagnosis (Triple-S Project 2011). Although this can result in earlier detection of outbreaks, or the detection of more cases of disease, these signs and

symptoms may not be disease-specific. This can make it difficult to ascertain which disease is causing changes in syndromic indicator trends.

The inclusion of broad indicators in syndromic surveillance (for example *Acute Respiratory Infection* (GPOOH)) is part of the design of syndromic surveillance, in order to capture as many cases of the disease of interest as possible. More specific indicators are available (for example *Bronchitis/Bronchiolitis* (GPIH)) but due to the lower number of presentations related to these indicators there is more noise in this data. Although similar syndromic indicators occur across the systems (for example *Difficulty Breathing* (NHS111) and *Difficulty Breathing/Wheeze/Asthma* (GPOOH)) they are not directly comparable and will detect slightly different diseases. This is because different factors, such as disease severity and demographics will lead to different healthcare-seeking behaviours to different services, and due to differences within the healthcare systems, for example GPs have a wider range of tools available to diagnose in comparison to NHS111. This can be observed from the results in chapter four, where 30% of *Influenza-Like-Illness* presentations to GPOOH in children under-1 year were estimated to be associated with RSV, compared to only 4% of *Influenza-Like-Illness* presentations to GPIH.

In chapter four we observed that *Acute Respiratory Infection* (GPOOH), and *Influenza-Like-Illness* (GPOOH), are both associated with RSV and influenza in children under-5 years. This is also highlighted by Morbey et al. (2017a), where influenza and RSV were found to be associated with both *Cold/Flu* and *Difficulty Breathing* presentations. Human Metapneumovirus, influenza and RSV were associated with *Cough* presentations to NHS111 in all age groups. In chapter four and five we related trends in respiratory indicators to laboratory confirmed cases of respiratory pathogens, allowing us to identify which pathogens were most likely to be associated with presentations for specific indicators. This method to identify with pathogens are associated with specific syndromic indicators is useful when retrospective studies are being conducted, but cannot be used for routine surveillance as trends in the syndromic data occur before they can be confirmed by laboratory data (Cooper et al. 2009).

To further obscure the interpretation of syndromic indicator trends, different age groups can present with different clinical manifestations for the same disease and some diseases are more likely to affect certain age groups. This can be observed in chapter four where RSV was associated with 43% of *Difficulty Breathing/Wheeze/Asthma presentations* to GPOOH in children under-1 year, but only 13% of presentations in children aged 1-4 years. Similar results were obtained by Morbey et al. (2017a), with rhinovirus associated with 18% of *Cold/Flu* presentations to NHS111 in those over 64 years, but was not associated with *Cold/Flu* presentations in any other age group. By understanding how different pathogens affect different age groups prior to analysis, accounting for this effect by analysing a variety of syndromic indicators and conducting sub-analysis by age, the impact of differing symptoms can be alleviated. For example, although RSV can infect all age groups, children under-5 years are at higher risk of severe disease and lower respiratory tract involvement (Shi et al. 2017). When estimating the burden of RSV using syndromic surveillance in chapter four, we limited the analysis to those under-5 years due to the more distinct symptoms and the higher burden of disease. In older age groups, it may have been more difficult to differentiate RSV trends from other pathogens using syndromic data.

Different seasonal pathogens, which manifest with similar symptoms typically circulate at different times of the year. For, example RSV, influenza and coronavirus typically circulate in the winter, adenovirus, rhinovirus, metapneumovirus and parainfluenza circulates all year, and enterovirus circulating in the summer (Moriyama, Hugentobler and Iwasaki 2020). Applying biological plausibility to the interpretation of trends can help identify responsible pathogens, for example, relating influenza and RSV to changes in respiratory indicator trends in winter when those viruses circulate, as we did in chapter four and five.

The ultimate aim of the ReSST suite of syndromic surveillance systems is to provide early warning of seasonal increases of disease, situational awareness during incidents, and reassurance of a lack of impact of risks, by primarily focusing on indicators related to respiratory and gastrointestinal infectious diseases, air pollution and heat waves (Elliot et al. 2017). Given the broad nature of some syndromic indicators, for example

Cough, and *Difficulty Breathing*, inevitably syndromic surveillance will detect diseases not within their remit. For example, *Cough* and *Difficulty Breathing* could be related to influenza (Monto et al. 2000), chronic obstructive pulmonary disease (COPD), or seasonal allergies (J. Smith and Woodcock 2006; Molinari, Colombo and Celenza 2014). In addition to this, the specific illness that is associated with a syndromic indicator will vary throughout the year.

Currently, only one syndromic indicator is received by ReSST per presentation, even if a patient presents with more than one symptom. In the case of NHS111 only one chief complaint is allocated per call, whereas the other healthcare systems monitored by syndromic surveillance (GPIH, GPOOH, ambulance and emergency departments) can have more than one clinical finding, although this is not reported to ReSST. Combining clinical findings may help differentiate between diseases. For example, if GPIH presentations for *Difficulty Breathing* could be combined with *Acute Respiratory Infection,* there would be the potential to differentiate presentations for respiratory infections from those for other illnesses that are not of interest.

Although steps can be taken to relate specific diseases to specific indicators, ultimately there will be uncertainty around which disease is causing changes in indicator trends. This was highlighted by G. E. Smith et al. (2019) who stated there needs to be "increased clarity" about the capabilities of syndromic surveillance due to the non-specific nature of some of the syndromic indicators.

Data Quality

Data quality is paramount to a successful public health surveillance system. Inadequate data quality can lead to poor understanding of disease epidemiology, and can undermine the surveillance systems ability to meet the aims of the programme and detect outbreaks of disease (Venkatarao et al. 2012; CDC 1988). It is imperative to continuously monitor and evaluate these systems to ensure data quality and performance are at optimum levels (CDC 1988).

Several factors that may result in fluctuations in data quality were identified by the Triple-S Project and include: sites failing to report data, lack of training, a reduction in the incidence of a presenting symptom, lack of motivation or involvement by the data providers and lack of data from specific areas resulting in a system that is not representative of the population and the diseases being monitored (Triple-S Project 2013). If syndromic surveillance remains a passive service, then in terms of ensuring data quality, there is little syndromic surveillance services can do, other than monitor data quality and provide feedback to data providers. Steps taken to monitor data quality should include estimating whether the number of records received is expected, estimating percentage completion of each field within a record, and checking the structure of the data (Triple-S Project 2013). As data from syndromic surveillance is used for other epidemiological purposes (such as spatial analysis chapters three and six) data quality issues become increasingly apparent. The data used in syndromic surveillance can be inconsistent and incomplete, although these issues in data quality can be accounted for in routine purposes of syndromic surveillance, especially when data is analysed at the national level. However, any data quality issues can be problematic for surveillance systems and can result in missing outbreaks, especially at the local level. These issues with the data do not result in the data being unusable, but they must be understood for the data to be used its full capability.

In terms of monitoring data quality, ReSST will evaluate the data for gross errors, such as data missing from whole regions. However due to the passive and timely nature of syndromic surveillance little can be done to correct these quality issues apart from trying to account for them in the analysis or acknowledging them when disseminating the data. For example, when analysing data from NHS111 the total number of calls was used as the denominator to account for the instability of the system as the healthcare service was rolled out across England. However, in chapter three we observed that while data on the total number of calls were submitted for several regions, indicator data was not, meaning disease trends or outbreaks could not be observed in those regions. Under the definition described by the Triple-S Project, syndromic surveillance data should already be collected for purposes other than surveillance and is generated as to not create an additional burden to healthcare providers (Triple-S Project 2011). This passive nature

can make data quality hard to ensure, because any data quality issues identified by ReSST cannot be corrected retrospectively by the data providers. The Triple-S Project highlighted that unspecific data with diagnostic mistakes is an inherent feature of syndromic surveillance, but as long as the quality of the data is stable, trends can still be identified (Triple-S Project 2013). However, it is important to highlight that it is only the transfer of data between the data providers and ReSST that is passive, the access to this data is not mandated, therefore these systems were established and are maintained through continuous cooperation between the data providers and ReSST providing some mechanism for feedback.

Little research has been conducted into whether data received by ReSST surveillance have been stable over time, primarily not to put addition burden on the healthcare services, and because over the period that ReSST have been operational there have been continuous changes to the healthcare services monitored. As the frequency of data submitted by the data providers in each geographical location has changed over time, this has created some problems when comparing across geographies. This is highlighted in chapter three, where large areas were excluded from analysis due to poor surveillance coverage, and large confidence intervals around the estimates were observed due to uncertainty in the data. However, this would only result in errors in the estimates if areas of poor surveillance coverage are correlated with the variables of interest, of which there was no evidence. In chapter five, data from GPOOH was excluded from the analysis due to the inconsistencies in the frequency of data submission which resulted in large geographical areas with no data, and variable coverage in some areas.

Evidence of Data Quality Issues in the Syndromic Surveillance Systems.

Representativeness of the population is a key aspect in surveillance; without which a disease's incidence and its distribution by person, place and time, is unknown (CDC 1988). Representativeness can be assessed by comparing the characteristics of the event to all events, and is based on the knowledge of population characteristics (age, location, gender and socioeconomic status), aetiology of the disease of interest, and healthcare services being monitored (CDC 1988). Comparing multiple sources of data can help identify whether a surveillance system is representative of the population (CDC 1988).

In chapter three, for the period between 1st January 2015 and 31st December 2016, 3.0% of NHS111 reports, where a chief symptom and age were present, were excluded due to lack of information on geographic location and gender. When the NHS111 was first developed between 2013 and 2014, it was observed that 20-30% of reports did not have syndromic information, and less than 2% lacked geographical location (Harcourt et al. 2016). Data issues based on population characteristics comprised a small proportion (2-3%) of received reports from NHS111, and this has remained stable from when the surveillance system was first introduced in 2013, to 2016 the last year of the data that was analysed for this thesis (Harcourt et al. 2016; chapter three). For the same period 9.8% of GPOOH reports where a chief complaint was present were excluded due to lack of information on age, geographic location or gender (chapter three). The data issues observed in the GPOOH system would only have an impact if they were clustered in a specific area, or this proportion fluctuated over time; however, we could not assess whether these exclusions occurred randomly. When surveillance of GPOOH was first introduced in 2010, problems with data quality were observed in between 10% and 90% of reports containing read codes, the number received depended on the GPOOH providers (Harcourt et al. 2012b). Further observation of the GPOOH surveillance system is required to monitor changes in the missing data. When a patient registers with a GP practice, they are required to provide demographic information, including age and gender, therefore all data from GP services included corresponding age and gender information. Due to the large volume of data and high proportion of population monitored by the three healthcare services, the surveillance systems are likely to be representative of the population in terms of age and gender despite the missing data based on these characteristics from NHS111 and GPOOH.

It was difficult to assess the representativeness of the surveillance coverage of the GPIH surveillance system. The denominator used from this system is currently the total daily number of registered patients in each PHE region, which comprise of nine large areas in England. It is not known where the practices that submit data are located, therefore, it is unknown if they are clustered in one area within the region or spread out. This has implications when interpretating observations from spatial analysis of the GPIH system. Data from GPIH services is available at a lower level of geography (postcode district)

from another data provider, but this data is not aggregated by age and gender, making it impractical for the epidemiological purposes of this thesis. Again, due to the passive nature of syndromic surveillance and the aim not to create additional pressure on the healthcare services, ReSST cannot retrospectively change their data agreements to collect this data. However, it is important to highlight that it is only geographical representativeness that is unknown, the GPIH surveillance system monitors over 40% of the population in England (chapter three), and therefore the data is highly likely to be representative in terms of age and gender. In contrast to the ReSST GPIH surveillance system, the Royal College of General Practitioners Research and Surveillance Centre has a surveillance network (RCGP 1957) of 1,700 GP practices across England and Wales, which was designed to monitor a representative sample of the population (Correa et al. 2016). This sentinel network collects syndromic data on presentations due to communicable and respiratory diseases, as well as conducting influenza virology screening, where those who are suspected of having influenza are tested. Data from this sentinel network is reported at weekly time points, and timely information at subnational level is not available, unlike with the ReSST GPIH system (G. E. Smith et al. 2007). This means it will be slower to detect public health threats, and temporal analysis will be limited compared to GPIH system operated by ReSST. Despite the uncertainty in the representativeness of GPIH data at the local level, we still included regional data from this service in our spatial analysis. However, it is important to highlight that the results observed at the regional level might not hold true at a finer geography, and the observations may contain bias toward certain areas. This uncertainty cannot be quantified.

In contrast to a report that is received with missing spatial information, poor coverage occurs where there the healthcare service is not operating in a geographical location, the surveillance systems are not operational, or surveillance services are operational but the data is not received due to issues with data transfer. In chapter four the coverage of the surveillance systems was estimated using change point analysis to identify periods where there were large changes in the mean and variance in the quality of data received by the NHS111 and GPOOH surveillance systems at the UTLA geographical level. In total, 38 out of the 149 UTLA studied had at least one week where NHS111 surveillance

system was not operational (where there was no data submitted from the providers or the healthcare system was not operational in an area) between the 11th November 2013 and 18th June 2018. Most of the issues with coverage were clustered at the beginning of the study period, when the service was first introduced. More specifically, coverage of the population in England was 83.6% in 2013/14, and this rose to 95% in 2017/18. NHS111 was the most stable system, with a consistent proportion of the population monitored over the study period. There were larger issues with surveillance coverage with the GPOOH surveillance system, with 54 of the 149 UTLA studied having periods of poor surveillance coverage. This poor coverage occurred throughout the study period and while most of the areas of poor coverage occurred randomly, there were large areas, such at the North East of England, where the coverage was constantly poorer than the rest of the country. Some areas were also inconsistent, and would drop in and out of providing data. This is particularly true for the GPOOH services, where some providers were inconsistent in transferring data to ReSST, and the landscape of GPOOH provisions changed over the study period. Data from the GPOOH surveillance system was not used to estimate the burden of RSV at the regional level in chapter five or when estimating the effect on meteorological conditions on respiratory indicators at the subnational level, in chapter six, due to inconsistencies in frequency of data submission from some geographical areas. Data from GPOOH was used in chapter three because there was no evidence to suggest that the areas of poor coverage were clustered by socioeconomic status, and the results were similar to what was observed from NHS111 data. This not only highlights the importance of representative geographical coverage in surveillance systems, but also the passiveness of the surveillance systems. One of the main principles of syndromic surveillance is that it does not create additional burden on healthcare services (Triple-S project 2013), therefore, they cannot retrospectively request data issues to be resolved. Research using syndromic surveillance has to be conducted within the limits of the data and the principles of syndromic surveillance.

Here, we have discussed how issues in poor coverage in the GPOOH surveillance system and lack of knowledge of geographical representativeness in the GPIH surveillance system have resulted in limitations in spatial analysis in this thesis and also led to questions about their representativeness when evaluating their use in routine

surveillance. This was emphasised by Colón-González et al. (2018), who noted that the effectiveness of data streams varied nationally, and an increase in coverage was required to improve the ability to detect local outbreaks of cryptosporidiosis. Although observations may not be as influenced by missing data at the national level, due to data quality issues there are limitations in the spatial analysis that can be conducted at the local level using data from syndromic surveillance.

Gaps in What we are Detecting

In chapter three we observed that those from more deprived backgrounds and young males (under-5 years) were the demographics that were most likely to present to the three healthcare services studied. This observation was true for those presenting with respiratory symptoms but also for all presentations. By identifying populations characteristics that are more likely to exhibit healthcare-seeking behaviours, targeted public health interventions can be aimed at these subgroups. Understanding who uses these services can also be used to identify baseline demographic patterns of healthcare-seeking behaviour. It is important to understand these underlying patterns to allow us to detect changes in healthcare-seeking behaviour, changes in the population at risk from certain diseases, and determine whether disease burden is increasing or decreasing in certain demographics (Soucie 2012; Nsubuga et al. 2006; Thacker et al. 1996). When making comparisons to historical data to detect changes in trends in the daily syndromic data, the data are, typically, aggregated into age subgroups. Although our analysis has identified that deprivation and gender are also important characteristics when looking at trends in the data, this may not be practical as part of routine surveillance. Our data was analysed at an annual time period, with reports from ReSST reported at a daily or weekly time period. This approach results in lower number of presentations and more noise in the data, further subdividing of the data by gender and deprivation would add additional noise which could result in false alerts to possible public health threats. This highlights how observations from academic research may not be practical in the real world.

Ethnicity has been observed to play in important role in disease (CDC 2021; Wang et al. 2020; Apea et al. 2021), although it is not common for surveillance systems to

monitor differences in disease trends by ethnicity or race. Differences in disease trends and outcomes can occur due to genetic factors linked to race, or environmental, social, religious or cultural factors linked to ethnicity. Ethnic disparities in disease prevalence and health outcomes have previously been well observed in chronic illness. In the UK, South Asian populations have been observed to have a higher prevalence of diabetes (Hanif Wasim and Susarla 2018), and cardiovascular diseases (Zaman and Mangtani 2007; Bhopal et al. 2005) such as angina, myocardial infarction (Bansal et al. 2013) and heart disease, and a lower prevalence of cancer (Harding and Rosato 1999) compared to the general population. The incidence of stroke is highest among black populations (Stewart et al. 1999), as well as the risk of dementia poststroke (Shiekh et al. 2020) compared to white populations in the UK. Those from black and minority ethnic backgrounds have also been observed to present later with cancer symptoms, leading to poorer survival when compared to white ethnic groups (Department of Health 2007). In terms of infectious diseases, a study in Scotland found Pakistani and African ethnic groups had a higher rate of infection-related hospitalisations/deaths compared to white and Chinese ethnic groups, even when accounting for socioeconomic status (Gruer et al. 2021). Ethnic disparities in health and disease came to the forefront during the COVID-19 pandemic. After taking other confounding factors into account, such as age, sex, income, education, housing tenure, and deprivation, those from ethnic minority groups were more likely to contract severe COVID-19, and twice as likely to die from COVID-19 compared to white British ethnic groups in England (Raisi-Estabragh et al. 2020; PHE 2020; Apea et al. 2021).

The reasons why those from non-white ethnic groups have a higher risk of disease is multifactorial, with a complex combination of biological, religious, cultural, and societal factors. Although the research highlights higher rates of chronic diseases in black and minority ethnic (BAME) groups, few studies have associated these differences to genetic factors linked to race (Karter 2003). Environmental factors that can lead to higher rates of disease in BAME groups include; vitamin deficiencies, higher rates of smoking and comorbidities, more likely to live in urban deprived areas and in overcrowded housing, and are more likely to working in lower paid jobs. Societal factors linked to health disparities include poor access to healthcare, poor experience of

healthcare and racial discrimination and marginalisation (Hanif Wasim and Susarla 2018; Razai et al. 2021; PHE 2018; Mindell et al. 2014; Wang et al. 2020).

Therefore, the collection of data on ethnicity in public health surveillance is important to identify and reduce health inequalities and improve health outcomes. Ethnicity has been shown to play an important role in disease, ReSST does not receive this data primality because many of the healthcare services monitored do not collect this data. Although it has been identified why it is important to collect information on ethnicity, barriers can make this data difficult to obtain for surveillance systems. Healthcare services where information is obtained may not collect ethnicity data, patients may not be willing to disclose their ethnicity or service providers may not be willing to provide this information.

Currently routine data disseminated from ReSST is not subdivided by gender, deprivation or ethnicity. However, information on gender is collected and if trends are identified, the information is disseminated appropriately. When conducting epidemiological research, studies using data from ReSST include both gender and age information. Deprivation cannot, currently, be directly obtained from patient information. However, an ecological measure can be provided using the geographical location provided. Again, deprivation data is not included in routine surveillance, but has been included in some epidemiological studies from ReSST (Todkill et al. 2017a; Adams et al. 2019; Morrison et al. 2020). Ethnicity data is not collected by ReSST and cannot be estimated using the information provided. In practical terms, as part of routine daily surveillance with ReSST, having data stratified by age, gender, deprivation and ethnicity has the potential to create too much noise in the data, making trends difficult to detect. Analysing ethnicity (if collected in the future) and deprivation data may be more suited to analyses using longer time periods (weeks, months or months) or for ad hoc epidemiological analysis.

7.3 Reflections on Syndromic Surveillance

Future of Syndromic Surveillance

Adapting to Changes Post-COVID-19

Since the start of the COVID-19 pandemic in 2020 there have been vast changes to the way people live and seek healthcare. Although some of these changes in behaviour may revert to as they were before the pandemic started, there could be equally permanent changes across the health and social care sectors in the UK, as well as changes in society. One of the recent changes that could have the biggest impact on the syndromic surveillance systems is way people are seeking healthcare, and the way services are run. In order to protect patients and staff, less face-to-face consultations were held, therefore more telehealth technologies were introduced. As well as the additional telehealth telephone service NHS111, a new NHS111 online service was introduced. This online service allows users to find the right healthcare in their area, get advice on self-help, or speak to a medical professional if needed (NHS Digital 2021b). In addition to this, GP services used telehealth technologies by offering telephone or video consultations, and specialist services also used video consultations where appropriate (Car et al. 2020). This adapting model of care and advancement in technologies used in healthcare will undoubtably impact the data collected by ReSST. Syndromic surveillance services need to constantly adapt to changes in the way healthcare is sought and to new healthcare services that are introduced. The flexibility of the syndromic surveillance infrastructure has allowed it to adapt to the pandemic. A new surveillance system was developed to monitor NHS111 online live, a healthcare service which was introduced within two months of the start of the pandemic, and new COVID-19 specific indicators were monitored from the existing services (Elliot et al. 2020). In addition to these organisational changes, there have been large changes within society in response to the pandemic, with people working from home where possible, using face masks in public and reducing social contact. These changes have not only reduced the transmission of COVID-19, but also other infectious diseases (Iacobucci 2020). Although some of these practises may end post-pandemic, changes to the way people work are expected to stay.

A survey of 4,933 adults in the UK found that 57% of respondents wanted to continue working from home in some way (YouGov 2020). With infections spread more easily indoors (Zhang et al. 2020), and allowing people to work from home when ill, future changes in work practice could reduce the long-term burden of infectious diseases and change the way people seek healthcare. Although, inherently, syndromic surveillance is able to adapt to changing behaviours, any change has a large impact on the ability to detect trends, as this crucially depends on making comparisons to historic data. In this thesis we analysed data at weekly time points of multiple years where the indicators were used routinely, this provided us with more data and therefore gave us more certainty in our observations. Elliot et al. (2010) highlighted how changes in healthcare-seeking behaviours resulted in some syndromic indicators being unusable during the COVID-19 pandemic. Given the flexibility of syndromic surveillance and that ReSST monitors a broad range of healthcare services (NHS111 online, NHS111, GP in and out-of-hours, ambulances, and emergency departments), surveillance should adapt to the changing behaviours in the way people seek healthcare post-pandemic and to changes in society. In comparison, hospitalisation and laboratory diagnostics and surveillance is less flexible, and as it might only detect the most serious cases it may not adapt to healthcare-seeking behaviours or changes in health of wider society that may come from changing work practices.

The primary advantage of syndromic surveillance is its timeliness in comparison to surveillance of hospitalisations and laboratory data. Its timeliness comes at the cost of confirmatory ability, with syndromic surveillance detecting signs and symptoms associated with disease. Home or self-test allow users to collect and test their own specimens without intervention from health professionals. In recent years the use of home testing kits in diagnosing infectious diseases, such as STIs, HIV and HPV, has been growing (Ibitoye et al. 2014; Kpokiri et al. 2020; Mahase 2021), but their use during the COVID-19 pandemic has brought them to the forefront of disease surveillance. The COVID-19 rapid lateral flow test kits allow people to test themselves at home and receive results within 30 minutes. These tests were developed to help detect cases in those who do not show symptoms, and prevent the transmission of the virus. From April 2021, the UK government made millions of these tests available to the

public even if they did not have symptoms of coronavirus (UK Government 2021). A study examining the usability and acceptability of COVID-19 self-test kits, observed a high degree of acceptability with 91.5% and 94.4% of users obtaining a valid result, depending on which test used (Atchison et al. 2021). Although currently home testing kits are not available for other common seasonal illnesses, such as influenza and RSV, their use during the pandemic raises questions on how they can be used in the future. Their rapid and diagnostic nature could provide timely and confirmatory results either at home or a healthcare setting, However, the targeted disease would have to be severe enough for the patient to seek healthcare and get tested, therefore, like syndromic surveillance, rapid testing might not reduce under-ascertainment in disease surveillance. They could also be deployed to both symptomatic and non-symptomatic populations, allowing the detection of previously undetectable cases, reducing under-ascertainment in disease surveillance. In real terms, home testing is currently not available for other seasonal infections, a large-scale disease surveillance network based on home testing would be expensive and currently there is no infrastructure to capture data from these kits. Syndromic surveillance also provides vital information on severity of disease, and who is seeking healthcare, which may not be available from rapid testing. During the 2009 influenza A/H1N1 pandemic syndromic surveillance was used in conjunction with respiratory virus self-sampling to monitor community transmission of influenza. Patients who called NHS Direct (pre-cursor to NHS111) with cold or flu like symptoms were sent self-sampling kits which were then tested by microbiology laboratories. This study provided a less biased estimate of community transmission of influenza A/H1N1, as well as providing a reliable indication of local transmission (Elliot et al. 2009b). This highlights how self-testing can be used to compliment syndromic and laboratory surveillance.

Syndromic Surveillance in Epidemiological Research

Routine and Ad Hoc Analysis

When syndromic surveillance was first introduced its primary focus was to detect bioterrorism events. Since then, syndromic surveillance systems have developed to encompass many more public health threats. Currently, the underlying aim of ReSST is

to provide early warning of seasonal increases of disease, situational awareness and reassurance of a lack of impact during incidents, mass gatherings and public health threats (Elliot et al. 2017). Previous examples of routine surveillance include monitoring extreme weather events (Elliot et al. 2014; S. Smith et al. 2016a; S. Smith et al. 2016b; Hughes et al. 2014), mass gatherings (Severi et al. 2012; Todkill et al. 2016), air pollution (Elliot et al. 2016a) and norovirus activity. It has also played a vital role in monitoring influenza and other respiratory viruses' activity along with other public health surveillance programmes (PHE 2019). Recently there has been a move to conduct ad hoc analysis on the data from syndromic surveillance for epidemiological purposes not in the current remit of ReSST. This includes monitoring the impact of vaccinations on subsequent morbidity using GP consultations, and estimate the association between socioeconomic status and healthcare-seeking behaviours (Todkill et al. 2017a; Adams et al. 2018). In this thesis we have also highlighted how this data can be successfully used in ad hoc epidemiological studies outside the current remit of ReSST. It was stated by G. E. Smith et al. (2019) that the primary purpose of syndromic surveillance should "remain on public health utility", "that public health action needs to drive the scope and outputs of syndromic surveillance" and "syndromic surveillance should complement and augment a variety of traditional surveillance systems in order to provide wider intelligence about a public health issue or incident". Data from ReSST are from healthcare services not routinely monitored by traditional forms of surveillance providing a unique insight into healthcare-seeking behaviours. Traditional forms of surveillance, such as hospitalisation episodes and laboratory notifications, capture a small proportion of disease prevalence, typically the more severe cases who have sought hospital care, or those with underlying health condition which puts them at more risk of severe diseases. Whereas the healthcare services monitored by ReSST which are described in this thesis, capture cases of disease in the community and less severe cases of disease. These syndromic surveillance systems monitor a large proportion of the population of England, as well as data availability at small levels of geography, allowing analysis and comparisons at a subnational level. This is a vast, unique and rich data source. In this thesis we used data from over 39.5 million presentations to NHS111, GPIH and GPOOH, from several geographical areas including postcode district, upper

tier local authority and PHE centre. In addition, data is available from many symptoms of disease allowing the exploration of a variety of diseases and illnesses. This has allowed us to utilise this data in diverse ad hoc epidemiological studies to estimate RSV burden (chapter four) and activity in the community at national and regional level (chapter five), estimate the relationship between socioeconomic status (chapter three) and respiratory presentations, and estimate the association between meteorological conditions and respiratory presentations (chapter six).

Previously, when estimating the burden of RSV in children in England, the data has primarily originated from hospitalisations or positive samples from laboratories (Müller-Pebody et al. 2002; Deshpande and Northern 2003; Heikkinen, Ojala and Waris 2017; Reeves et al. 2017; Chavez et al. 2019). Although, recent burden estimates have been made using additional data from GP practices (Taylor et al. 2016; Cromer et al. 2017; C. Murray 2013), these data sources can introduce bias into estimates, capturing only those with severe disease or underlying health conditions that puts them at risk of severe disease. By utilising other data sources, such as those monitored by syndromic surveillance, not only can more of the burden be estimated, but the impact of this bias can potentially also be reduced. The high level of geographical coverage of some of the syndromic surveillance systems may also help reduce bias in burden estimates caused by differences in hospital reporting and laboratory testing. In this thesis we look specifically at the burden of RSV in children under-5 year, but there is scope to utilise this method on other age groups or diseases. Although the highest burden of RSV is among children under-5 years, Fleming et al. (2015) highlighted that there is still substantial disease burden in adults, which is poorly characterised. The burden of influenza in the UK has been well explored and characterised, but again estimates primarily focus on primary care and hospitalisations (Cromer et al. 2014; Fleming et al. 2016). Gastrointestinal diseases are also a potential area of research to estimate burden of disease using the method described in this thesis. Two large-scale studies on infectious diseases in the community in the UK were published in 1999 and 2012, and these were designed to estimate the true number of infections in the community of several gastrointestinal infectious (Tam et al. 2012; Wheeler et al. 1999). Although of robust design and analysis, the large-scale nature of the studies meant they required a

lot of resources (Sethi et al. 1999). Data from syndromic surveillance could be an efficient way to continually estimate and monitor gastrointestinal diseases rates in the community in conjunction with these studies.

In this thesis, we have highlighted how syndromic surveillance data can be successfully used in ad hoc epidemiological studies outside the current remit of ReSST. When estimating the relationship between respiratory presentations and meteorological conditions using syndromic surveillance, observations were similar to those estimated using other, more traditional forms of data (Z. Chen et al. 2014; du Prel et al. 2009; Hervás, Reina and Hervás 2012; Jaakkola et al. 2014; Lam et al. 2016; Liu et al. 2015; Oliveira-Santos et al. 2016; Price, Graham and Ramalingam 2019; Sirimi et al. 2016). This validates the observations and highlights how data from syndromic surveillance is a viable data source for these study types. Given the volume and resolution of the data, we were able to investigate regional and temporal relationships, which is something not previously done. There is a growing interest about how syndromic surveillance can be used to develop forecasting models (Elliot et al. 2017; G. E. Smith et al. 2019) to predict future presentations to healthcare services, this research highlights the importance of meteorological conditions on presentations to healthcare services. It also highlights the potential for disease forecasting at the regional level.

Big Data

Data obtained through ReSST can be described as “Big Data”. Although there is no widely accepted definition of Big Data in epidemiology, Roski, Bo-Linn and Andrews (2014) emphasised there were three generally accepted features: volume, variety, and velocity.

Volume refers to the amount of data, and is a key characteristic of Big Data (Roski, Bo-Linn and Andrews 2014). Data from the healthcare services monitored by ReSST has the potential to receive data covering the whole English population; with NHS111, which has the largest coverage, monitoring over 95% of the population in the last year of data available for this thesis. Data is received at daily time points and is sub-divided by age, gender, and depending on the system, data is available at a low level spatial and

temporal resolution. This has the potential to create huge amounts of data. As highlighted by Mooney, Westreich and El-Sayed (2015), with this larger volume of data comes the increased use of statistical and computational techniques and programmes. With larger datasets, developing directed acyclic graphs can be used to help visualise theorized data relations and possible causal relationships. With the use of DAGitty, a browser-based environment for creating, editing, and analysing directed acyclic graphs, such acyclic graphs were created to visualise possible causal relationships in the data used in this thesis (Textor et al. 2016). With the large samples size that accompany data from syndromic surveillance, significant or important results may be accompanied by low p-values and small effect sizes. Expert knowledge is required to separate these findings from a highly precise finding, with little significance to wider public health and one with potential importance (Mooney, Westreich and El-Sayed 2015; Robins 2001; Poole 2001; Siontis and Ioannidis 2011). This was highlighted in chapter six when looking at the spatial patterns in the relationship between meteorological conditions and respiratory presentations in children under-5 years. Although important relationships were observed, the lack of spatial clustering and small credible intervals led to the conclusion that they were of little importance to wider public health. However, this information may be more useful in the future with the growing interest in using syndromic surveillance in forecasting presentations to healthcare services. Big data requires flexible and easily expandable data storage (Roski, Bo-Linn and Andrews 2014), but also appropriate hardware to analyse the data to its full capabilities. In this thesis we attempted to undertake spatial and temporal analysis on the data provided by ReSST. Due to the data sharing agreements between ReSST and the data providers, data for this thesis could only be analysed on PHE infrastructure. Due to these constraints, analysis could not be conducted to its full potential. With the large capabilities of the data from ReSST, the ability to conduct comprehensive analysis will depend of the IT infrastructure available.

Variety refers to incorporating data from different sources into one for combined analysis (Mooney, Westreich and El-Sayed 2015), for example linking of patient data between healthcare services, or linking area characteristics to patient data. The data from the surveillance systems are not linked, research could be conducted to investigate

if combining data streams of specific indicators from different services could increase the sensitivity and specificity of the detection capabilities of the systems (G. E. Smith et al. 2019). Using multiple data sources in epidemiological studies, as done in this thesis, can add robustness of findings due to similar observations. The availability of these unique data sources from syndromic surveillance has allowed us to explore healthcare usage in populations not routinely captured by traditions forms of surveillance. G. E. Smith et al. (2019) highlighted that as new potential sources of data become available (such as twitter or school absences), there must be a clear need for them and they must add value within syndromic surveillance. In addition, they established that data sources should be dis-established if they add little to public health intelligence. Suggesting that data should only be collected for a comprehensive and cost-effective syndromic surveillance service that adds value to public health.

Velocity refers to data generation, compilation and analysis in real time (Mooney, Westreich and El-Sayed 2015). In the case of ReSST, data is available at daily time steps, from the previous day, and analysed on the same day using algorithms, with results interpreted by epidemiologists. Although data in this thesis was analysed at annual and weekly time steps, data availability at a daily time point gives huge potential for rapid epidemiological analysis, such as, when researching events that happen over a short period of time or have immediate effects such as heatwaves, or air pollution (S. Smith et al. 2016a:2016b; Elliot et al. 2016a). Not only does timely data generation and analysis allow for the early detection of public health threats, but it also allows for the quick implementation and evaluation of interventions (Mooney, Westreich and El-Sayed 2015).

Although typically messy, Big Data has enormous potential in a variety epidemiological research designs to identify population health threats, targets and evaluating intervention targets (Mooney, Westreich and El-Sayed 2015). Previous research and results from this thesis have demonstrated the potential use of syndromic data in a variety of epidemiological studies, with more potential capabilities with further expansion of ReSST and IT infrastructure.

Data Sharing and Continuing Work with Data Providers

The relationships between ReSST, the data providers, public health users and external organisations were highlighted by G. E. Smith et al. (2019) as being imperative to the sustainability and success of syndromic surveillance, but there is a delicate balance, with the most important relationship being that between the data providers and ReSST. The relationship is often based upon the 'public good' of syndromic surveillance and is underpinned by trust, with providers needing to be aware of any release of raw data, or interpretation of data (G. E. Smith et al. 2019). In addition to this, a fundamental responsibility of syndromic surveillance is to not create additional burden on front-line healthcare services by requesting additional data once an agreement has been made, therefore it is imperative this relationship is strong from the offset. Given the considerable amount of data available, ReSST are often asked by external organisations, such as academia, for raw data (G. E. Smith et al. 2019). There is an increased trend towards making and ensuring transparency of data, which includes making datasets freely available, to allow for reproducible research to verify published findings and toughen scientific rigor (Coughlin 2017). Elliot et al. (2017) emphasised how collaboration between academia research and public health organizations can allow the development of new ideas, methods and can bring external research funding. However, this collaboration needs to occur under clear data governance arrangements between the data providers and ReSST. Under such agreements in place in England at the moment, data cannot be removed from PHE data servers, due to the IT infrastructure available this can limit the possibilities in research capabilities. Although working within the constraints of the data agreement between ReSST and the data providers can be difficult, it is imperative for ReSST to maintain this relationship for continued access to the data.

7.4 Final Thoughts

Main Contributions and How We Filled Research Gaps

In this thesis we aimed to quantify how data from syndromic surveillance can be used outside its principle aims of detecting public health threats, situational awareness and for reassurance to the lack of threat from large events (G. E. Smith et al. 2019). We

primarily focused on spatial and temporal analysis, because data from ReSST was available at multiple geographical (postcode district, UTLA and PHE centre) and temporal (annual, monthly and weekly) resolutions, allowing us to explore healthcare-seeking behaviours over time and at the local level. Previously, these aspects of the data have not been fully explored in epidemiological studies, with the majority of the ReSST research output focusing on their aims and disease trends at the national level. All analysis included in this thesis incorporated a spatial or temporal element, and we highlighted how this data can be used successfully for this type of analysis, while emphasising the considerations that need to be made, especially in terms of issues with inconsistent coverage.

The primary aim of this thesis was to highlight how this data can be used in observational epidemiological studies. Much of the literature on observational epidemiological studies primarily focuses on diagnostic data, such as hospitalisations or laboratory notifications. The use of these datasets can introduce significant bias in these studies towards more severe disease in acute healthcare settings. The healthcare services used in this thesis: NHS111, GPIH and GPOOH operate in the community, and will therefore capture a higher proportion, and less severe cases of disease. These cases may not be included in diagnostic surveillance figures but they place a significant burden on healthcare, and society; therefore, it is important to capture and quantify these previously unreported cases. In addition to this, ReSST monitor healthcare services which are not commonly used in epidemiological research. In this thesis the inclusion of NHS111 and GPOOH data provided a unique insight into healthcare-seeking behaviour in the community.

Here, we successfully explored how data from syndromic surveillance can be successfully used in observational epidemiological studies, while also expanding the literature on specific public health issues. In chapter three we explored the demographic and socioeconomic factors in healthcare-seeking behaviour to NHS111, GPIH and GPOOH for respiratory illnesses. This research allowed us to understand who is more likely to use these services, and how this compared to more acute services and other diseases. In chapter four we estimated the burden of RSV that can be attributable to

NHS111, GPIH and GPOOH. Previous research on RSV burden typically uses data from hospitalisations or laboratory notifications, here we were able to quantify the community burden. In this chapter we also observed higher presentations in younger children, and RSV had a higher burden than influenza in children under-5 years, which is similar to what has been observed using diagnostic data, but have not previously been described using syndromic data. In chapter five we attempted to identify differences in regional seasonality and burden of RSV in children under-1 year. Although higher burdens were observed in the south of England in both NHS111 and GPIH services, no regional differences in RSV seasonally were observed. This chapter highlighted the limitations of the data, particularly in terms of surveillance coverage issues of the GPOOH surveillance system and the lack of geographical granularity of the GPIH data. Finally, in chapter six we explored the association between meteorological conditions (temperature, humidity, rainfall and wind speed) and NHS111, GPIH and GPOOH presentations due to acute repository illness in children under-5 years. The observations were similar to what has been described in the literature using diagnostic data, highlighting similar meteorological effects on presentations in the community and in acute healthcare setting. Due to the unique features of the data, we were also able to explore spatial and temporal relationships between presentations and the meteorological data, demonstrating that meteorological conditions have an effect all year, and have higher significance in some areas.

Sources of Uncertainty

The focus of this thesis is the utility of data from syndromic surveillance in observational epidemiological research, previously we have discussed the possible sources of uncertainty from syndromic data sources. However, it is important to highlight other possible sources of uncertainty in our research.

Although the inclusion of data from the healthcare services monitored by syndromic surveillance can reduce the level of under-reporting of disease, by capturing cases of disease that would not have previously been identified through laboratory notifications, hospitalisation data or notifiable disease reports, the use of syndromic data does not reduce the rate of under-ascertainment in the surveillance data. Under-ascertainment

occurs when an individual with the disease of interest does not attend healthcare services, and therefore are not captured by surveillance systems that monitor healthcare usage (Gibbons et al. 2014). The level of under-ascertainment varies by disease, population subgroups, and geographical area (Gibbons et al. 2014), and is a source of uncertainty in most epidemiological datasets (Delgado-Rodríguez and Llorca 2004). Reasons an individual might not seek healthcare include: no or mild symptoms, self-limiting illness, lack of access to healthcare or religious or cultural factors (Gibbons et al. 2014). Under-ascertainment can introduce bias into the data if the cases captured by the surveillance systems do not represent those that occur in the population, for example those who might not be captured due to lack of access to healthcare, or population subgroup who are less likely to seek healthcare (Delgado-Rodríguez and Llorca 2004). Lake et al. (2009) highlighted that under-ascertainment can vary geographically, with differences in healthcare attendance (Farmer et al. 2006) and diagnosis (Sethi et al. 1999) in more urban areas compared to more rural areas, and diagnostic protocols in laboratories differing by geography. This can lead to bias in our observations, however we attempted to alleviate this by including geographical level fixed effects in our spatial analysis.

The level of spatial precision of the data can have an impact of how accurately the trends and relationship can be identified (Jeffery, Ozonoff and Pagano 2014; Jeffery et al. 2009). Due to the inherent design of the surveillance systems or to reduce noise in the datasets, the data used in this thesis was aggregated to multiple geographical resolutions. The lowest level of geography available for NHS111 and GPOOH data was postcode district, and for the GPIH data it was PHE centre. PHE centre is a very large geographical area with only nine regions in England, each with millions of residents. In this thesis, NHS111 and GPOOH data was also analysed at UTLA and PHE centre geographical levels. Corresponding independent variables were also aggregated to the geographical levels that the syndromic data were analysed at. These aggregations may have resulted in loss of information from the data and the masking of relationships between the dependent and independent variables (Jeffery, Ozonoff and Pagano 2014). This is particularly true of the meteorology data aggregated to PHE centre in chapter

six, there is likely to be climatic variations within each PHE centre which will not be captured in this dataset.

The use of data aggregation gives rise to ecological fallacy, whereby inferences about an individual are based upon observations from the group from which they belong (Robinson 1950). Portnov, Dubnov and Barchana (2007) emphasised that the use of aggregated data in epidemiological research should not be avoided, primarily because aggregated data is more readily available in this field, but it is important for the researcher to identify when ecological fallacy might occur and address it where possible. In the analysis conducted in this thesis we tried to reduce the effect of confounding factors, by denying them through the use of directed acyclic graphs, and including them in the analysis where possible. By analysing several data sources at different geographies can also identify similar observations. However, it is important to acknowledge that aggregated data may result in the identification of relationships that are not present at the individual level, this is particularly true of the highly aggregated GPIH data.

Seasonality is an important feature of infectious diseases, with many having cyclic change in occurrence over a year (Dowell 2001; Grassly and Fraser 2006; Christiansen et al. 2012). In this thesis we focus on respiratory diseases, selecting syndromic indicators that have been previously associated with respiratory infections (Morbey et al. 2017a;2017b; 2018), which are known to have a strong seasonal element (Moriyama, Hugentobler and Iwasaki 2020). It was important for us to control for seasonality in our analysis in chapters four to six, because there was a seasonal component to both the outcome (syndromic indicators) and the exposures (respiratory pathogens, meteorological condition and unknown seasonal factors), therefore seasonality was a confounding factor (Christiansen et al. 2012). However, the level of seasonal adjustment in the model can greatly influence the results, not accounting for it properly can overestimate the relationship between the exposure and outcome variables, and over accounting for seasonality can mask the relationship between the exposure and outcome variables. To further complicate matters, as syndromic indicators can be associated with more than one pathogen or disease, and this relationship can vary overtime, the

syndromic data often have more than one seasonal cycle over a year, which needs to be accounted for the statistical analysis.

Future Research

In chapter three we demonstrated that young males and those from the most deprived areas were more likely to seek healthcare in the community for respiratory illnesses, with similar findings observed for all non-respiratory presentations. The approach here could be applied to other surveillance systems or disease groups because the observations from this chapter are an important step in understanding who uses these services and are important for public health decision making. In chapter four we successfully estimated the healthcare burden attributable to RSV in the community in children under-5 years. Fleming et al. (2015) highlighted that RSV also had a substantial burden in adults, especially in the elderly. Therefore, the methods described in chapter four could be used to estimate the community burden of RSV in adults. The surveillance systems have been designed to detect trends in other infectious illness such influenza, norovirus and rotavirus, this method could be used to estimate and monitor trends for other infectious diseases. In chapter four we demonstrated that different degrees of seasonality can influence observations. Although we explored the impact of different parameterisations on our findings, we highlight this as an area for further research. Future research could be conducted to explore the impacts of different seasonality approaches and accounting for different degrees of seasonality in time-series epidemiological modelling, while considering the most appropriate methods for different diseases. The observation we found in chapter five were inconclusive. We had only four RSV seasons worth of data in this analysis, therefore when more data is available the research could be conducted again. In chapter six we focused on the relationship between meteorological conditions and respiratory presentations in children under-5 year, this work could be extended to look at other age groups or diseases.

Supplementary Material

Pathogen/Indicator	Meteorological Condition	Author	Age of Cases	Clinical Setting	Place of Study	Association
Rhinovirus	Temperature (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant negative association
	Dew point (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
	Relative humidity (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association
	Humidity-range (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
	Wind velocity	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
	Relative atmospheric pressure	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
Adenovirus	Temperature (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant negative association
	Dew point (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
	Relative humidity (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association
	Humidity-range (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
	Wind velocity	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association

Table A1: Summary of respiratory pathogen or syndromic indicator associations with meteorological variables from the literature.

	Relative atmospheric pressure	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
Human Metapneumovirus	Temperature (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant negative association
	Dew point (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
	Relative humidity (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
		Du Prel et al. (du Prel et al. 2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association
	Humidity-range (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
	Wind velocity	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association
	Relative atmospheric pressure	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
Influenza A	Temperature (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant negative association
	Dew point (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
	Relative humidity (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant positive association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association
	Humidity-range (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
	Wind velocity	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association
	Relative atmospheric pressure	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association

Table A1: Summary of respiratory pathogen or syndromic indicator associations with meteorological variables from the literature.

Influenza B	Temperature (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant negative association
	Dew point (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
	Relative humidity (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
	Humidity-range (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
	Wind velocity	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association
	Relative atmospheric pressure	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
Parainfluenza-1	Temperature (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
	Dew point (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
	Relative humidity (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant positive association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association
	Humidity-range (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
	Wind velocity	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant negative association
	Relative atmospheric pressure	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
Parainfluenza-2	Temperature (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association

Table A1: Summary of respiratory pathogen or syndromic indicator associations with meteorological variables from the literature.

	Dew point (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
	Relative humidity (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
	Humidity-range (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
Parainfluenza-3	Temperature (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant negative association
	Dew point (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	No significant association
	Relative humidity (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
	Humidity-range (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant positive association
	Wind velocity	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant negative association
	Relative atmospheric pressure	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
Enterovirus	Temperature (°C)	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
	Relative humidity (%)					Significant positive association
	Wind velocity					No significant association
	Relative atmospheric pressure					No significant association
Coronavirus	Temperature (°C)	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant negative association
	Relative humidity (%)					Significant positive association

Table A1: Summary of respiratory pathogen or syndromic indicator associations with meteorological variables from the literature.

	Wind velocity					Significant positive association
	Relative atmospheric pressure					Significant positive association
Respiratory Syncytial Virus	Temperature (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant negative association
		Yusuf et al. (2007)	All ages	Multiple laboratories	Mexico City (Mexico), Miami (USA), Delhi (India), Houston (USA), Tucson (USA), Santiago (Chile), Buffalo (USA), Winnipeg (Canada), Bethel (USA)	Significant positive association in Miami. Significant negative association in Delhi, Houston, Tucson, Buffalo, Winnipeg, Bethel and Santiago.
	Dew point (°C)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
		Yusuf et al. 2007)	All ages	Multiple laboratories	Mexico City (Mexico), Miami (USA), Delhi (India), Houston (USA), Tucson (USA), Santiago (Chile), Buffalo (USA), Winnipeg (Canada), Bethel (USA)	Significant positive association in Mexico City, and Miami. Significant negative association in Delhi, Houston, Tucson, Buffalo, Winnipeg, and Bethel.
	Relative humidity (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
		du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association

Table A1: Summary of respiratory pathogen or syndromic indicator associations with meteorological variables from the literature.

		Yusuf et al. 2007)	All ages	Multiple laboratories	Mexico City (Mexico), Miami (USA), Delhi (India), Houston (USA), Tucson (USA), Santiago (Chile), Buffalo (USA), Winnipeg (Canada), Bethel (USA)	Significant positive association in Mexico City, Miami, and Santiago. No significant association at other sites
	Humidity-range (%)	Price, Graham, and Ramalingam (2019)	All ages	Samples from 8 hospitals and 126 primary care centres	Edinburgh, Scotland	Significant negative association
	Wind velocity	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association
		Meerhoff et al. (2009)	All ages	Data from 11 laboratories	Netherlands	No significant association
	Relative atmospheric pressure	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
	Prevailing wind direction (°)	Meerhoff et al. (2009)	All ages	Data from 11 laboratories	Netherlands	No significant association
	Minimum temperature (°C)	Meerhoff et al. (2009)	All ages	Data from 11 laboratories	Netherlands	Significant negative association
	Maximum sunshine duration (%)					No significant association
	Precipitation (mm)	Meerhoff et al. (2009)	All ages	Data from 11 laboratories	Netherlands	Respiratory Syncytial Virus
		Yusuf et al. 2007)	All ages	Multiple laboratories	Mexico City (Mexico), Miami (USA), Delhi (India), Houston (USA), Tucson (USA), Santiago (Chile), Buffalo (USA), Winnipeg	Significant positive association in Miami only. No significant association at other sites

Table A1: Summary of respiratory pathogen or syndromic indicator associations with meteorological variables from the literature.

					(Canada), Bethel (USA)	
	Mean surface air pressure (hPa)	Meerhoff et al. (2009)	All ages	Data from 11 laboratories	Netherlands	No significant association
	Cloud cover (octants)	Meerhoff et al. (2009)	All ages	Data from 11 laboratories	Netherlands	Significant positive association
	Mean relative atmospheric humidity (%)	Meerhoff et al. (2009)	All ages	Data from 11 laboratories	Netherlands	Significant positive association
	UVB radiance ($W/m^2/nm$)	Yusuf et al. 2007)	All ages	Multiple laboratories	Mexico City (Mexico), Miami (USA), Delhi (India), Houston (USA), Tucson (USA), Santiago (Chile), Buffalo (USA), Winnipeg (Canada), Bethel (USA)	Significant negative association (only measured in Miami, Buffalo, Winnipeg)
Acute Respiratory Infections	Temperature (°C)	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant negative association
		Mäkinen et al. (2009)	Military recruits (17-19 years)	Those who presented with respiratory symptoms	Kajaani garrison in northern Finland	Significant negative association
	Relative humidity (%)	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association
		Costilla-Esquivel et al. (2014)	All ages	11 Primary care centres	Apodaca and Guadalupe, Mexcio	Significant negative association
	Absolute humidity	Mäkinen et al. (2009)	Military recruits (17-19 years)	Those who presented with respiratory symptoms	Kajaani garrison in northern Finland	Significant positive association
	Wind velocity	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	Significant positive association

Table A1: Summary of respiratory pathogen or syndromic indicator associations with meteorological variables from the literature.

	Relative atmospheric pressure	du Prel et al. (2009)	< 16 years	Hospitalisations at one children's hospital	Mainz, Germany	No significant association
	Maximum Temperature (°C)	Mäkinen et al. (2009)		Those who presented with respiratory symptoms	Kajaani garrison in northern Finland	Significant positive association
	Minimum temperature	Costilla-Esquivel et al. (2014)	All ages	11 Primary care centres	Apodaca and Guadalupe, Mexcio	Significant positive association
	Rainfall	Costilla-Esquivel et al. (2014)	All ages	11 Primary care centres	Apodaca and Guadalupe, Mexcio	Significant positive association
URTI	Temperature (°C)	Mäkinen et al. (2009)	Military recruits (17-19 years)	Those who presented with respiratory symptoms	Kajaani garrison in northern Finland	Significant negative association
	Maximum Temperature (°C)	Mäkinen et al. (2009)	Military recruits (17-19 years)	Those who presented with respiratory symptoms	Kajaani garrison in northern Finland	Significant negative association
	Absolute humidity	Mäkinen et al. (2009)	Military recruits (17-19 years)	Those who presented with respiratory symptoms	Kajaani garrison in northern Finland	Significant positive association
	Temperature Change	Liu et al. (2015)	Children <16 years	Children who visited the Department of Paediatrics at Guangzhou Women and Children's Medical Centre for symptoms of an RTI.	Guangzhou, China	Significant positive association (at 3.5 °C temperature change and with no lag)
LRTI	Temperature (°C)	Mäkinen et al. (2009)	Military recruits (17-19 years)	Those who presented with respiratory symptoms	Kajaani garrison in northern Finland	Significant negative association
	Maximum Temperature (°C)	Mäkinen et al. (2009)	Military recruits (17-19 years)	Those who presented with respiratory symptoms	Kajaani garrison in northern Finland	Significant positive association
	Absolute humidity	Mäkinen et al. (2009)	Military recruits (17-19 years)	Those who presented with respiratory symptoms	Kajaani garrison in northern Finland	Significant positive association

Table A1: Summary of respiratory pathogen or syndromic indicator associations with meteorological variables from the literature.

	Temperature Change	Liu et al. (2015)	Children <16 years	ll children who visited the Department of Paediatrics at Guangzhou Women and Children's Medical Centre for symptoms of an RTI.	Guangzhou, China	Significant positive association (at -6.2, -3.5. 2.9 and 3.8°C temperature change and with lag of 5, 10 and 15 days)
Common Cold	Temperature (°C)	Mäkinen et al. (2009)	Military recruits (17-19 years)	Those who presented with respiratory symptoms	Kajaani garrison in northern Finland	Significant negative association
	Maximum Temperature (°C)	Mäkinen et al. (2009)	Military recruits (17-19 years)	Those who presented with respiratory symptoms	Kajaani garrison in northern Finland	Significant negative association
	Absolute humidity	Mäkinen et al. (2009)	Military recruits (17-19 years)	Those who presented with respiratory symptoms	Kajaani garrison in northern Finland	Significant positive association
Bronchiolitis	Minimum temperature	Hoeppner et al. (2017)	Patient aged 2month – 2 years	Hospitalisations due to bronchiolitis at seven paediatric centres	Australia and New Zealand	Significant negative association
	Wind speed	Hoeppner et al. (2017)	Patient aged 2month – 2 years	Hospitalisations due to bronchiolitis at seven paediatric centres	Australia and New Zealand	Significant positive association
		Nenna et al. (2017)	< 1 year	Hospitalisations due to bronchiolitis at University Hospital of the "Sapienza" University	Rome, Italy	No significant association
	Relative humidity	Hoeppner et al. (2017)	Patient aged 2month – 2 years	Hospitalisations due to bronchiolitis at seven paediatric centres	Australia and New Zealand	No significant association
		Nenna et al. (2017)	< 1 year	Hospitalisations due to bronchiolitis at University Hospital of the "Sapienza" University	Rome, Italy	Significant positive association
	Rainfall	Hoeppner et al. (2017)	Patient aged 2month – 2 years	Hospitalisations due to bronchiolitis at seven paediatric centres	Australia and New Zealand	No significant association

Table A1: Summary of respiratory pathogen or syndromic indicator associations with meteorological variables from the literature.

	Temperature (°C)	Nenna et al. (2017)	< 1 year	Hospitalisations due to bronchiolitis at University Hospital of the "Sapienza" University	Rome, Italy	Significant negative association

Table A1: Summary of respiratory pathogen or syndromic indicator associations with meteorological variables from the literature.

System	Indicator	Meteorological Condition	Relationship Direction	Spatial Pattern	Monthly Pattern
NHS111	Difficulty Breathing	Temperature	Negative	Majority of England with important negative association. Association appears stronger around the south coast	Negative important relationship except in February and March. Relationship stronger in summer.
		Rainfall	Not important	Not important in the majority of England, apart from small clusters in the North West and South East, where there is an important negative relationship.	Important negative relationship in December only.
		Humidity	Positive	Clusters of important positive relationships, but no clear spatial pattern.	Important positive relationship every month except March-May. Relationship strongest in the summer.
		Wind Speed	Positive	Clusters of important positive relationships, but no clear spatial pattern.	Important positive relationship in December.
	Cough	Temperature	Negative	Majority of England with important negative association. Relationship not important in the North of England. Association appears stronger around the South West coast.	Negative important relationship except in February and March. Relationship stronger in summer.
		Rainfall	Negative	Not important in the majority of England, apart from clusters in the North, midlands and South West, where there is an important negative relationship.	Important negative relationship in November, December January and April.
		Humidity	Positive	Clusters of important positive relationships, but no clear spatial pattern.	Important positive relationship every month except April, May and October. Relationship strongest in the summer.
		Wind Speed	Positive	Majority of England with important negative association. Association appears stronger around the South West coast and in the north.	Important positive relationship in September, October and December. Important negative relationship in March.
	Cold/Flu	Temperature	Negative	Majority of England with important negative association. Association appears stronger around the south coast	Negative important relationship all year. Relationship stronger in summer.
		Rainfall	Negative	Not important in the majority of England, apart from clusters in the West Midlands and south, where there is an important negative relationship.	Important negative relationship in autumn, winter and spring.

Table A2: Summary of results from chapter six.

		Humidity	Positive	Important positive relationship in large areas on England, primarily in the north, midlands and south of England.	Important positive relationship all year except May, July and August. Relationship similar all year.
		Wind Speed	Positive	Clusters of important positive relationships, but no clear spatial pattern.	Important positive relationship from September-December.
GPOOH	ARD	Temperature	Negative	-	Negative important relationship all year. Relationship stronger in summer.
		Rainfall	Negative	-	Important negative relationship in November and December.
		Humidity	Positive	-	Important positive relationship all year except May. Relationship strongest in the summer.
		Wind Speed	Positive	-	Important positive relationship from September-January, and in April. Important negative relationship in July.
	ARI	Temperature	Negative	-	Negative important relationship all year. Relationship stronger in summer.
		Rainfall	Negative	-	Important negative relationship in November, December January and April. Important positive relationship in May.
		Humidity	Positive	-	Important positive relationship all year except May. Relationship strongest in the summer.
		Wind Speed	Positive	-	Important positive relationship from September-January, and in April. Important negative relationship in July.
	DWBA	Temperature	Negative	-	Negative important relationship except January-April. Relationship stronger in summer.
		Rainfall	Not important	-	Important negative relationship in November and December.
		Humidity	Positive	-	Important positive relationship all year except March-May. Relationship strongest in the summer.

Table A2: Summary of results from chapter six.

		Wind Speed	Positive	-	Important positive relationship in December.
	Bronchitis	Temperature	Negative	-	Important negative relationship in June and August.
		Rainfall	Negative	-	Important negative relationship in October, November and December.
		Humidity	Positive	-	Relationship only important in the winter (October – January) where it was positive.
		Wind Speed	Positive	-	Important positive relationship in December.
GPIH	URTI	Temperature	Negative	Relationship not important in the north and west; but important in the south and east of England. Relationship strongest in the East of England.	Important negative relationship in summer only.
		Rainfall	Negative	Important in all areas of England with an important negative relationship, expect North East, Yorkshire and the Humber and the East of England. Relationship strongest in the South East.	Important negative relationship in December, January March and April. Important positive relationship in July.
		Humidity	Not important	Positive relationship in the west and South East of England, negative relationship in the east.	Important positive relationship in November and December. Important negative relationship in April, June, and October.
		Wind Speed	Not important	Important negative relationship in the East of England only.	Important positive relationship in January and April. Important negative relationship in September.
	LRTI	Temperature	Not important	No important relationship in any region of England	Important negative relationship in summer only.
		Rainfall	Negative	Relationship only important in the East Midlands, where the relationship is negative.	Important negative relationship in December, January, and April.
		Humidity	Not important	Positive relationship in the west and South East of England, negative relationship in the east.	Important positive relationship in November and December. Important negative relationship in April, June, and October.
		Wind Speed	Not important	No important relationship in any region of England	Important positive relationship in December, January and April.

Table A2: Summary of results from chapter six.

					Important negative relationship in February, March, May and September.
	ILI	Temperature	Negative	Important negative relationship in all areas except London and West Midlands. Relationship strongest in the North East.	Important negative relationship in March, July and September only.
		Rainfall	Not important	Relationship only important in Yorkshire and the Humber, where the relationship is negative.	No important relationship for any month.
		Humidity	Positive	Positive relationship in the west and South East of England.	Important positive relationship from November-March.
		Wind Speed	Positive	Important positive relationship in the East of England ad South West only.	No important relationships all year.
	Acute Bronchitis	Temperature	Negative	Important negative relationship in all areas except the East of England and Yorkshire and the Humber. Relationship strongest in the North West and south of England.	Important negative relationship in June, July, September and October only.
		Rainfall	Not important	Relationship not important in any region.	No important relationship for any month.
		Humidity	Positive	Important positive relationship in all areas except Yorkshire and the Humber and the East of England.	Important positive relationship in November and December.
		Wind Speed	Positive	Important positive relationship in London and the West Midlands only.	No important relationships all year.

Table A2: Summary of results from chapter six.

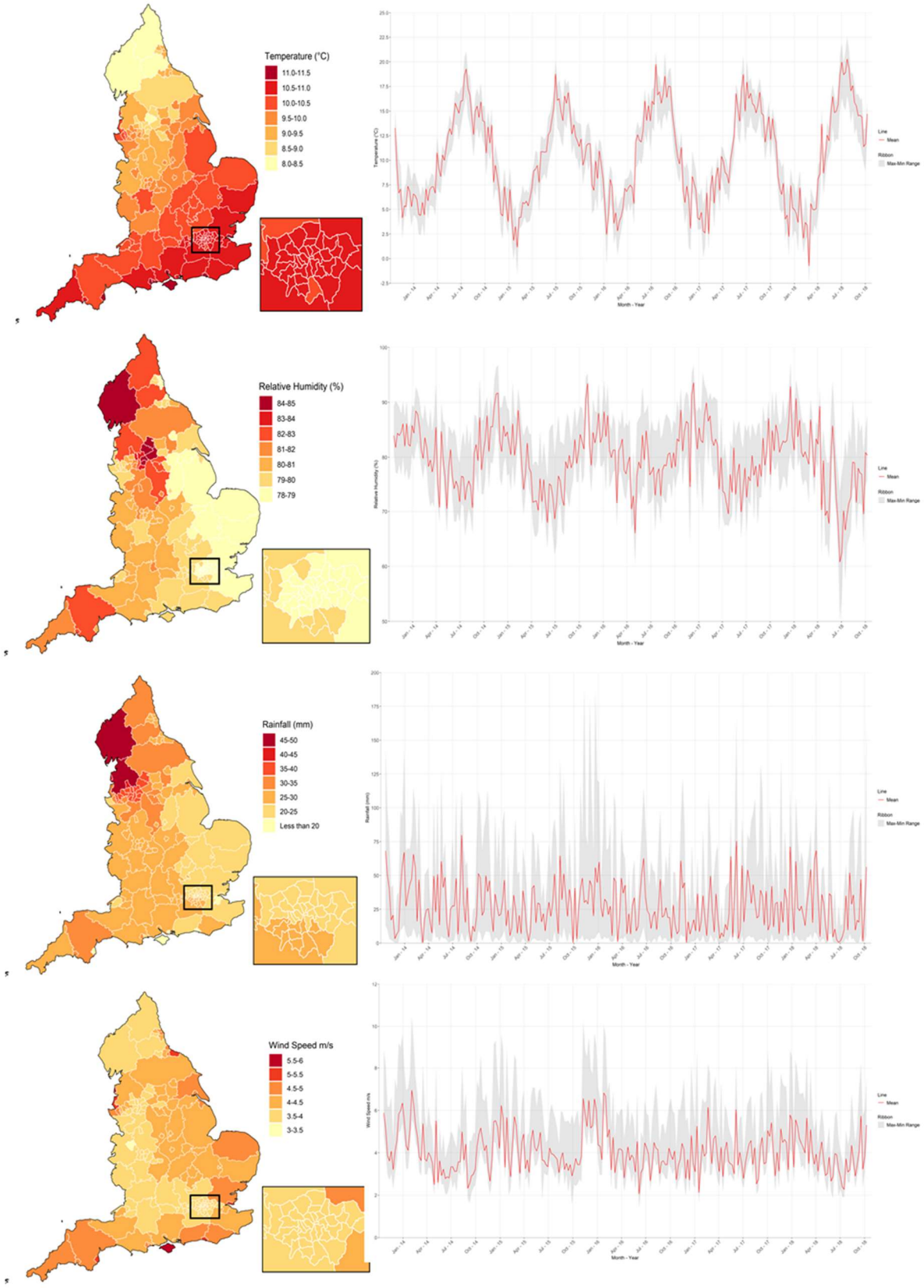

Figure A1: Map and time-series of the weekly mean of the meteorological conditions at UTLA and as a time series throughout study period (21st October 2013 to 25th June 2018

References

Abat, C., H. Chaudet, J. Rolain, P. Colson, and D. Raoult. 2016. "Traditional and Syndromic Surveillance of Infectious Diseases and Pathogens." *International Journal of Infectious Diseases* 48: 22–28. https://doi.org/10.1016/j.ijid.2016.04.021.

Abedi, G. R, M. M Prill, G. E Langley, M. E Wikswo, G. A Weinberg, A. T Curns, and E. Schneider. 2016. "Estimates of Parainfluenza Virus-Associated Hospitalizations and Cost Among Children Aged Less Than 5 Years in the United States, 1998-2010." *Journal of the Pediatric Infectious Diseases Society* 5 (1): 7–13. https://doi.org/10.1093/jpids/piu047.

Abedi, G. R., J. T. Watson, W. Nix, M. Oberste, and S. I. Gerber. 2018. "Enterovirus and Parechovirus Surveillance - United States, 2014-2016." *MMWR. Morbidity and Mortality Weekly Report* 67 (18): 515–18. https://doi.org/10.15585/mmwr.mm6718a2.

Adams, N. L., T. C. Rose, A. J. Elliot, G. E Smith, R. A. Morbey, P. Loveridge, J. Lewis, et al. 2018. "Social Patterning of Telephone Health-Advice for Diarrhoea and Vomiting: Analysis of 24 Million Telehealth Calls in England." *Journal of Infection*, 78(2): 95–100. https://doi.org/10.1016/j.jinf.2018.09.008.

Agrawal, A. S., M. Sarkar, S. Chakrabarti, K. Rajendran, H. Kaur, A. C. Mishra, M. K. C., Trailokya N. Naik, M. S. Chadha, and M. Chawla-Sarkar. 2009. "Comparative Evaluation of Real-Time PCR and Conventional RT-PCR during a 2 Year Surveillance for Influenza and Respiratory Syncytial Virus among Children with Acute Respiratory Infections in Kolkata, India, Reveals a Distinct Seasonality of Infection." *Journal of Medical Microbiology* 58 (12): 1616–22. https://doi.org/10.1099/jmm.0.011304-0.

Ajayi-Obe, E. K., P. G. Coen, R. Handa, K. Hawrami, C. Aitken, E. D. G. McIntosh, and R. Booy. 2008. "Influenza A and Respiratory Syncytial Virus Hospital Burden in Young Children in East London." *Epidemiology and Infection* 136 (8): 1046–58. https://doi.org/10.1017/S0950268807009557.

Al Sallakh, M. A. Al, S. E. Rodgers, R. A. Lyons, A. Sheikh, and G. A. Davies. 2017. "Socioeconomic Deprivation and Inequalities in Asthma Care in Wales." *The Lancet* 390: S19. https://doi.org/10.1016/S0140-6736(17)32954-9.

Allard, R. 1998. “Use of Time-Series Analysis in Infectious Disease Surveillance.” *Bulletin of the World Health Organization* 76 (4): 327–33.

Anderson, L. J., J. C. Hierholzer, C. Tsou, R. M. Hendry, B. F. Fernie, Y. Stone, and K. McIntosh. 1985. “Antigenic Characterization of Respiratory Syncytial Virus Strains with Monoclonal Antibodies.” *The Journal of Infectious Diseases* 151 (4): 626–33.

Andersson, T., P. Bjelkmar, A. Hulth, J. Lindh, S. Stenmark, and M. Widerström. 2014. “Syndromic Surveillance for Local Outbreak Detection and Awareness: Evaluating Outbreak Signals of Acute Gastroenteritis in Telephone Triage, Web-Based Queries and over-the-Counter Pharmacy Sales.” *Epidemiology & Infection* 142 (2): 303–13. https://doi.org/10.1017/S0950268813001088.

Andeweg, S. P., R. M. Schepp, J. van de Kassteele, L. Mollema, G. A. M. Berbers, and M. van Boven. 2021. “Population-Based Serology Reveals Risk Factors for RSV Infection in Children Younger than 5 Years.” *Scientific Reports* 11 (1): 8953. https://doi.org/10.1038/s41598-021-88524-w.

Anestad, G. 1982. “Interference between Outbreaks of Respiratory Syncytial Virus and Influenza Virus Infection.” *The Lancet* (8270): 502. https://doi.org/10.1016/s0140-6736(82)91466-0.

Apea, V. J., Yize I. W., Rageshri D., Z. A. Puthucheary, R. M. Pearse, C. M. Orkin, and J. R. Prowle. 2021. “Ethnicity and Outcomes in Patients Hospitalised with COVID-19 Infection in East London: An Observational Cohort Study.” *BMJ Open* 11 (1): e042140. https://doi.org/10.1136/bmjopen-2020-042140.

Asten, L. van, C. van den Wijngaard, W. van Pelt, H. van Vlie, and M. Koopmans. 2007. “Understanding the Dynamics of Gastro-Intestinal Syndrome: General Practioner and Hospital Data vs Laboratory Surveillance.” *Advances in Disease Surveillance* 2 (215). http://faculty.washington.edu/lober/www.isdsjournal.org/htdocs/articles/949.pdf.

Atchison, C., Philippa Pristerà, E. Cooper, V. Papageorgiou, R. Redd, M. Piggin, B. Flower, et al. 2021. “Usability and Acceptability of Home-Based Self-Testing for Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Antibodies for Population Surveillance.” *Clinical Infectious Diseases* 72 (9): e384–93. https://doi.org/10.1093/cid/ciaa1178.

Ayala, A., V. Berisha, K. Goodin, K. Pogreba-Brown, C. Levy, B. McKinney, L. Koski, and S. Imholte. 2016. “Public Health Surveillance Strategies for Mass

Gatherings: Super Bowl XLIX and Related Events, Maricopa County, Arizona, 2015." *Health Security* 14 (3): 173–84. https://doi.org/10.1089/hs.2016.0029.

Azziz B., C. N. Dao, S. Nasreen, M. U. Bhuiyan, S. Mah-E-Muneer, A. Al Mamun, M. A. Y. Sharker, et al. 2012. "Seasonality, Timing, and Climate Drivers of Influenza Activity Worldwide." *The Journal of Infectious Diseases* 206 (6): 838–46. https://doi.org/10.1093/infdis/jis467.

Bahk, G. J., Y. Soo Kim, and M. Su Park. 2015. "Use of Internet Search Queries to Enhance Surveillance of Foodborne Illness." *Emerging Infectious Diseases* 21 (1): 1906–12. https://doi.org/10.3201/eid2111.141834

Baker, M. G, L. Telfar Barnard, A. Kvalsvig, A. Verrall, J. Zhang, M. Keall, N. Wilson, T. Wall, and P. Howden-Chapman. 2012. "Increasing Incidence of Serious Infectious Diseases and Inequalities in New Zealand: A National Epidemiological Study." *The Lancet* 379 (9821): 1112–19. https://doi.org/10.1016/S0140-6736(11)61780-7.

Bansal, N., C. M. Fischbacher, R. S. Bhopal, H. Brown, M. FC Steiner, S. Capewell, and on behalf of the Scottish Health and Ethnicity Linkage Study. 2013. "Myocardial Infarction Incidence and Survival by Ethnic Group: Scottish Health and Ethnicity Linkage Retrospective Cohort Study." *BMJ Open* 3 (9): e003415. https://doi.org/10.1136/bmjopen-2013-003415.

Bawa, Z., A. J. Elliot, R. A. Morbey, S. Ladhani, N. A. Cunliffe, S. J. O'Brien, M. Regan, and G. E. Smith. 2015. "Assessing the Likely Impact of a Rotavirus Vaccination Program in England: The Contribution of Syndromic Surveillance." *Clinical Infectious Diseases* 61 (1): 77–85. https://doi.org/10.1093/cid/civ264.

Belderbos, M. E., M. L. Houben, B. Wilbrink, E. Lentjes, E. M. Bloemen, J. L. L. Kimpen, M. Rovers, and L. Bont. 2011. "Cord Blood Vitamin D Deficiency is Associated with Respiratory Syncytial Virus Bronchiolitis." *Pediatrics* 127 (6): e1513–20. https://doi.org/10.1542/peds.2010-3054.

Berger, M., R. Shiau, and J. M. Weintraub. 2006. "Review of Syndromic Surveillance: Implications for Waterborne Disease Detection." *Journal of Epidemiology and Community Health* 60 (6): 543–50. https://doi.org/10.1136/jech.2005.038539.

Besag, J., J. York, and A. Mollié. 1991. "Bayesian Image Restoration, with Two Applications in Spatial Statistics." *Annals of the Institute of Statistical Mathematics* 43 (1): 1–20. https://doi.org/10.1007/BF00116466.

Besculides, M., R. Heffernan, F. Mostashari, and D. Weiss. 2005. “Evaluation of School Absenteeism Data for Early Outbreak Detection, New York City.” *BMC Public Health* 5: 105. https://doi.org/10.1186/1471-2458-5-105.

Bhaskaran, Krishnan., Gasparrini, Antonio., Hajat, Shakoor., Smeeth, Liam., Armstrong, Ben., 2013. “Time series regression studies in environmental epidemiology.” *International Journal of Epidemiology* 42, 1187–1195. https://doi.org/10.1093/ije/dyt092

Bhopal, R., C. Fischbacher, E. Vartiainen, N. Unwin, M. White, and G. Alberti. 2005. “Predicted and Observed Cardiovascular Disease in South Asians: Application of FINRISK, Framingham and SCORE Models to Newcastle Heart Project Data.” *Journal of Public Health* 27 (1): 93–100. https://doi.org/10.1093/pubmed/fdh202.

Bloom-Feshbach, K., W. J. Alonso, V. Charu, J. Tamerius, L. Simonsen, M. A. Miller, and C. Viboud. 2013. “Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review.” *PloS One* 8 (2): e54445. https://doi.org/10.1371/journal.pone.0054445.

Bodaghkhani, E., M. Mahdavian, C. MacLellan, A. Farrell, and S. Asghari. 2019. “Effects of Meteorological Factors on Hospitalizations in Adult Patients with Asthma: A Systematic Review.” Review Article. *Canadian Respiratory Journal*. https://doi.org/10.1155/2019/3435103.

Bourgeois, F. T., C. Valim, A. J. McAdam, and K. D. Mandl. 2009. “Relative Impact of Influenza and Respiratory Syncytial Virus in Young Children.” *Pediatrics* 124 (6): e1072–80. https://doi.org/10.1542/peds.2008-3074.

Brabazon, E. D., A. O’Farrell, C. A. Murray, M. W. Carton, and P. Finnegan. 2008. “Under-Reporting of Notifiable Infectious Disease Hospitalizations in a Health Board Region in Ireland: Room for Improvement?” *Epidemiology and Infection* 136 (2): 241–47. https://doi.org/10.1017/S0950268807008230.

Broberg, E. K., M. Waris, K. Johansen, R. Snacken, P. Penttinen, and European Influenza Surveillance Network. 2018. “Seasonality and Geographical Spread of Respiratory Syncytial Virus Epidemics in 15 European Countries, 2010 to 2016.” *Eurosurveillance* 23 (5): 17–00284. https://doi.org/10.2807/1560-7917.ES.2018.23.5.17-00284.

Brooks, M. E., K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Maechler, and B. M. Bolker. 2017. “GlmmTMB

Balances Speed and Flexibility Among Packages for Zero-Inflated Generalized Linear Mixed Modelling." *The R Journal* 9 (2): 378–400.

Brooks, W. A., P. Terebuh, C. Bridges, A. Klimov, D. Goswami, A. Tahia Sharmeen, T. Azim, et al. 2007. "Influenza A and B Infection in Children in Urban Slum, Bangladesh." *Emerging Infectious Diseases* 13 (10): 1507–8. https://doi.org/10.3201/eid1310.070368.

Buckingham-Jeffery, E., R. A. Morbey, T. House, A. J. Elliot, S. Harcourt, and G. E. Smith. 2017. "Correcting for Day of the Week and Public Holiday Effects: Improving a National Daily Syndromic Surveillance Service for Detecting Public Health Threats." *BMC Public Health* 17 (1): 477. https://doi.org/10.1186/s12889-017-4372-y.

Buehler, J., R. Berkelman, D. Hartley, and C. Peters. 2003. "Syndromic Surveillance and Bioterrorism-Related Epidemics." *Emerging Infectious Diseases* 22 (11). https://doi.org/10.3201/eid0910.030231.

Buehler, J., A. Sonricker, M. Paladini, P. Soper, and F. Mostashari. 2008. "Syndromic Surveillance Practice in the United States: Findings from a Survey of State, Territorial, and Selected Local Health Departments." *Advances in Diseases Surveillance* 6: 1–20.

Bundle, N., N. Q. Verlander, R. A. Morbey, O. Edeghere, S. Balasegaram, S. de Lusignan, G. E. Smith, and A. J. Elliot. 2019. "Monitoring Epidemiological Trends in Back to School Asthma among Preschool and School-Aged Children Using Real-Time Syndromic Surveillance in England, 2012–2016." *Journal of Epidemiology and Community Health* 73 (9): 825–831. https://doi.org/10.1136/jech-2018-211936.

C3S (Copernicus Climate Change Service). 2017. "ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate. Copernicus Climate Change Service Climate Data Store (CDS)." https://cds.climate.copernicus.eu/cdsapp#!/home.

C3S (Copernicus Climate Change Service). 2019. "C3S ERA5-Land Reanalysis. Copernicus Climate Change Service." https://cds.climate.copernicus.eu/cdsapp#!/home.

Cai, W., S. Buda, E. Schuler, S. Hirve, W. Zhang, and W. Haas. 2020. "Risk Factors for Hospitalized Respiratory Syncytial Virus Disease and Its Severe Outcomes." *Influenza and Other Respiratory Viruses* 14 (6): 658–70. https://doi.org/10.1111/irv.12729.

Cameron, A.C, and P.K Trivedi. 2013. *Regression Analysis of Count Data*. New York: Cambridge University Press.

Cannell, J. J., R. Vieth, J. C. Umhau, M. F. Holick, W. B. Grant, S. Madronich, C. F. Garland, and E. Giovannucci. 2006. "Epidemic Influenza and Vitamin D." *Epidemiology and Infection* 134 (6): 1129–40. https://doi.org/10.1017/S0950268806007175.

Car, J., G. C-H. Koh, P. S. Foong, and C. J. Wang. 2020. "Video Consultations in Primary and Specialist Care during the Covid-19 Pandemic and Beyond." *BMJ* 371: m3945. https://doi.org/10.1136/bmj.m3945.

Care Quality Commission. 2014. "Our New Approach to the Inspection of NHS GP Out-of-Hours Services: Findings from the First Comprehensive Inspections." https://www.cqc.org.uk/sites/default/files/20140924_gp_out_of_hours_final.pdf.

Carneiro, H. A., and E. Mylonakis. 2009. "Google Trends: A Web-Based Tool for Real-Time Surveillance of Disease Outbreaks." *Clinical Infectious Diseases* 49 (10): 1557–64. https://doi.org/10.1086/630200.

Casalegno, J. S., M. Ottmann, M. Bouscambert Duchamp, V. Escuret, G. Billaud, E. Frobert, F. Morfin, and B. Lina. 2010. "Rhinoviruses Delayed the Circulation of the Pandemic Influenza A (H1N1) 2009 Virus in France." *Clinical Microbiology and Infection* 16 (4): 326–29. https://doi.org/10.1111/j.1469-0691.2010.03167.x.

CDC (Centers for Disease Control and Prevention). 1988. "Guidelines for Evaluating Surveillance Systems." *MMWR* 37 (S-5): 1–18.

CDC (Centers for Disease Control). 1998. "Preventing Emerging Infectious Diseases: A Strategy for the 21st Century Overview of the Updated CDC Plan." *CDC MMWR Recommendations and Reports* 47 (RR15) (September): 1–14.

CDC (Centers for Disease Control). 2005. "Surveillance for Illness and Injury After Hurricane Katrina, New Orleans, Louisiana, September 8--25, 2005." *CDC MMWR Weekly* 54 (40): 1018–21.

CDC (Centers for Disease Control). 2006a. "Increased Antiviral Medication Sales before the 2005-06 Influenza Season--New York City." *MMWR. Morbidity and Mortality Weekly Report* 55 (10): 277–79.

CDC (Centers for Disease Control). 2006b. “Injury and Illness Surveillance in Hospitals and Acute-Care Facilities After Hurricanes Katrina and Rita --- New Orleans Area, Louisiana, September 25--October 15, 2005.” *CDC MMWR Weekly* 50 (2): 35–38.

CDC (Centers for Disease Control). 2010. *National Biosurveillance Strategy for Human Health. Version 2.0*. Washington, DC: Department of Health and Human Services. https://stacks.cdc.gov/view/cdc/35002.

CDC (Centers for Disease Control and Prevention). 2021. “Hospitalization and Death by Race/Ethnicity.” May 26, 2021. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html.

Chan, P. W. K., F. T. Chew, T. N. Tan, K. B. Chua, and P. S. Hooi. 2002. “Seasonal Variation in Respiratory Syncytial Virus Chest Infection in the Tropics.” *Pediatric Pulmonology* 34 (1): 47–51. https://doi.org/10.1002/ppul.10095.

Charland, K. M., J. S. Brownstein, A. Verma, S. Brien, and D. L. Buckeridge. 2011. “Socio-Economic Disparities in the Burden of Seasonal Influenza: The Effect of Social and Material Deprivation on Rates of Influenza Infection.” *PLOS ONE* 6 (2): e17207. https://doi.org/10.1371/journal.pone.0017207.

Chavez, D., V. Gonzales-Armayo, E. Mendoza, R. Palekar, R. Rivera, A. Rodriguez, C. Salazar, A. Veizaga, and A. Añez. 2019. “Estimation of Influenza and Respiratory Syncytial Virus Hospitalizations Using Sentinel Surveillance Data—La Paz, Bolivia. 2012–2017.” *Influenza and Other Respiratory Viruses* 13 (5): 477–483. https://doi.org/10.1111/irv.12663.

Chen, H., D. Zeng, and P. Yan. 2010. *Infectious Disease Informatics: Syndromic Surveillance for Public Health and Biodefense*. Springer Science & Business Media.

Chen, Z., Y. Zhu, Y. Wang, W. Zhou, Y. Yan, C. Zhu, X. Zhang, H. Sun, and W. Ji. 2014. “Association of Meteorological Factors with Childhood Viral Acute Respiratory Infections in Subtropical China: An Analysis over 11 Years.” *Archives of Virology* 159 (4): 631–39. https://doi.org/10.1007/s00705-013-1863-8.

Cheng, C. K.Y., B. J. Cowling, E. H.Y. Lau, L. Ming Ho, G. M. Leung, and D. K.M. Ip. 2012. “Electronic School Absenteeism Monitoring and Influenza Surveillance, Hong Kong.” *Emerging Infectious Diseases* 18 (5): 885–87. https://doi.org/10.3201/eid1805.111796.

Cheng, C. K. Y., H. Channarith, and B. J. Cowling. 2013. “Potential Use of School Absenteeism Record for Disease Surveillance in Developing Countries, Case Study in Rural Cambodia.” *PLoS ONE* 8 (10). https://doi.org/10.1371/journal.pone.0076859.

Cherry, J., G. Demmler-Harrison, S. Kaplan, W. Steinbach, and P. Hotez. 1998. *Respiratory Syncytial Virus. In: Feigin RD, Cherry JD. (Eds) Textbook of Pediatric Infectious Diseases*. Vol. 1. 2 vols. Philadelphia, PA: WB Saunders. https://www.elsevier.com/books/feigin-and-cherrys-textbook-of-pediatric-infectious-diseases/cherry/978-1-4557-1177-2.

Chew, C., and G. Eysenbach. 2010. “Pandemics in the Age of Twitter: Content Analysis of Tweets during the 2009 H1N1 Outbreak.” *PloS One* 5 (11): e14118. https://doi.org/10.1371/journal.pone.0014118.

Chisholm, R. H., P. T. Campbell, Y. Wu, S. Y. C. Tong, J. McVernon, and N. Geard. 2018. “Implications of Asymptomatic Carriers for Infectious Disease Transmission and Control.” *Royal Society Open Science* 5 (2): 172341. https://doi.org/10.1098/rsos.172341.

Chretien, J., H. S Burkom, E. R. Sedyaningsih, R. P. Larasati, A. G. Lescano, C. C. Mundaca, D. L. Blazes, et al. 2008. “Syndromic Surveillance: Adapting Innovations to Developing Settings.” *PLoS Medicine* 5 (3). https://doi.org/10.1371/journal.pmed.0050072.

Christiansen, C. F., L. Pedersen, H. T. Sørensen, and K. J. Rothman. 2012. “Methods to Assess Seasonal Effects in Epidemiological Studies of Infectious Diseases—Exemplified by Application to the Occurrence of Meningococcal Disease.” *Clinical Microbiology and Infection* 18 (10): 963–69. https://doi.org/10.1111/j.1469-0691.2012.03966.x.

Colón-González, F. J., I. R. Lake, R. A. Morbey, A. J. Elliot, R. Pebody, and G. E. Smith. 2018. “A Methodological Framework for the Evaluation of Syndromic Surveillance Systems: A Case Study of England.” *BMC Public Health* 18 (1): 544. https://doi.org/10.1186/s12889-018-5422-9.

Cooper, D. L., G. E. Smith, W. J. Edmunds, C. Joseph, E. Gerard, and R. C. George. 2007. “The Contribution of Respiratory Pathogens to the Seasonality of NHS Direct Calls.” *Journal of Infection* 55 (3): 240–48. https://doi.org/10.1016/j.jinf.2007.04.353.

Cooper, D. L., N. Q. Verlander, A. J. Elliot, C. A. Joseph, and G. E. Smith. 2009. “Can Syndromic Thresholds Provide Early Warning of National Influenza

Outbreaks?" *Journal of Public Health* 31 (1): 17–25. https://doi.org/10.1093/pubmed/fdm068.

Corley, C. D., D. J. Cook, A. R. Mikler, and K. P. Singh. 2010. "Using Web and Social Media for Influenza Surveillance." *Advances in Experimental Medicine and Biology* 680: 559–64. https://doi.org/10.1007/978-1-4419-5913-3_61.

Cornally, N., and G. McCarthy. 2011. "Help-Seeking Behaviour for the Treatment of Chronic Pain." *British Journal of Community Nursing* 16 (2): 90–98. https://doi.org/10.12968/bjcn.2011.16.2.90.

Correa, A., W. Hinton, A. McGovern, J. van Vlymen, I. Yonova, S. Jones, and S. de Lusignan. 2016. "Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) Sentinel Network: A Cohort Profile." *BMJ Open* 6 (4): e011092. https://doi.org/10.1136/bmjopen-2016-011092.

Costilla-Esquivel, A., F. Corona-Villavicencio, J. G. Velasco-Castañón, C. E. Medina-DE LA Garza, R. T. Martínez-Villarreal, D. E. Cortes-Hernández, L. E. Ramírez-López, and G. González-Farías. 2014. "A Relationship between Acute Respiratory Illnesses and Weather." *Epidemiology and Infection* 142 (7): 1375–83. https://doi.org/10.1017/S0950268813001854.

Coughlin, S. S. 2017. "Reproducing Epidemiologic Research and Ensuring Transparency." *American Journal of Epidemiology* 186 (4): 393–94. https://doi.org/10.1093/aje/kwx065.

Cromer, D., A. Jan van Hoek, M. Jit, W. J. Edmunds, D. Fleming, and E. Miller. 2014. "The Burden of Influenza in England by Age and Clinical Risk Group: A Statistical Analysis to Inform Vaccine Policy." *Journal of Infection* 68 (4): 363–71. https://doi.org/10.1016/j.jinf.2013.11.013.

Cromer, D., A. Jan van Hoek, A. T. Newall, A. J. Pollard, and M. Jit. 2017. "Burden of Paediatric Respiratory Syncytial Virus Disease and Potential Effect of Different Immunisation Strategies: A Modelling and Cost-Effectiveness Analysis for England." *The Lancet Public Health* 2 (8): e367–e374. https://doi.org/10.1016/S2468-2667(17)30103-2.

CSMF (Committee on Sports Medicine and Fitness). 2011. "Climatic Heat Stress and Exercising Children and Adolescents." *Pediatrics* 128 (3): e741–47. https://doi.org/10.1542/peds.2011-1664.

Curtis, L. 2013. "Unit Costs of Health and Social Care. Canterbury: Personal Social Services Research Unit." Personal Social Services Research Unit, University of Kent.

Curtis, L. 2018. “Unit Costs of Health and Social Care.” Community-Based Health Care Staff. Canterbury: Personal Social Services Research Unit, University of Canterbury.

D’Amato, Maria., Molino, Antonio., Calabrese, Giovanna., Cecchi, Lorenzo., Annesi-Maesano, Isabella., D’Amato, Gennaro., 2018. “The impact of cold on the respiratory tract and its consequences to respiratory health.” *Clinical and Translational Allergy* 8, 20. https://doi.org/10.1186/s13601-018-0208-9

Dasaraju, P. V., and C. Liu. 1996. “Infections of the Respiratory System.” In *Medical Microbiology*, edited by Samuel Baron, 4th ed. Galveston (TX): University of Texas Medical Branch at Galveston. http://www.ncbi.nlm.nih.gov/books/NBK8142/.

Davison, K. L., N. S. Crowcroft, M. E. Ramsay, D. W.G. Brown, and N. J Andrews. 2003. “Viral Encephalitis in England, 1989–1998: What Did We Miss?” *Emerging Infectious Diseases* 9 (2): 234–40. https://doi.org/10.3201/eid0902.020218.

DCLG (Department for Communities and Local Government). 2015. “English Indices of Deprivation 2015: Technical Report. Available from: Https://Assets.Publishing.Service.Gov.Uk/Government/Uploads/System/Uploads/Attachment_data/File/464485/English_Indices_of_Deprivation_2015_-_Technical-Report.Pdf.”

de Lusignan, S., A. Correa, S. Pathirannehelage, R. Byford, I. Yonova, A.J. Elliot, T. Lamagni, et al. 2016. “RCGP Research and Surveillance Centre Annual Report 2014–2015: Disparities in Presentations to Primary Care.” *British Journal of General Practice* 67 (654). https://doi.org/10.3399/bjgp16X688573.

de Lusignan, S., A. Correa, R. Pebody, I. Yonova, G. E. Smith, R. Byford, S. Rankiri Pathirannehelage, et al. 2018. “Incidence of Lower Respiratory Tract Infections and Atopic Conditions in Boys and Young Male Adults: Royal College of General Practitioners Research and Surveillance Centre Annual Report 2015-2016.” *JMIR Public Health and Surveillance* 4 (2): e49. https://doi.org/10.2196/publichealth.9307.

Deeks, A., C. Lombard, J. Michelmore, and H. Teede. 2009. “The Effects of Gender and Age on Health-Related Behaviours.” *BMC Public Health* 9 (1): 213. https://doi.org/10.1186/1471-2458-9-213.

DEFRA (Department for Environment, Food & Rural Affairs). 2016. “2011 Rural-Urban Classification of Local Authorities and Other Geographies - Lookup for 2011 Rural Urban Classification of Local Authorities.”

Delgado-Rodríguez, M., and J. Llorca. 2004. “Bias.” *Journal of Epidemiology & Community Health* 58 (8): 635–41. https://doi.org/10.1136/jech.2003.008466.

Department of Health. 2007. “Cancer Reform Strategy.” https://www.nhs.uk/NHSEngland/NSF/Documents/Cancer%20Reform%20Strategy.pdf.

Desai, R., A. J. Hall, B. A. Lopman, Y. Shimshoni, M. Rennick, N. Efron, Y. Matias, M. M. Patel, and U. D. Parashar. 2012. “Norovirus Disease Surveillance Using Google Internet Query Share Data.” *Clinical Infectious Diseases* 55, (8): e75–e78. https://doi.org/10.1093/cid/cis579.

Deshpande, S, and V Northern. 2003. “The Clinical and Health Economic Burden of Respiratory Syncytial Virus Disease among Children under 2 Years of Age in a Defined Geographical Area.” *Archives of Disease in Childhood* 88 (12): 1065–69. https://doi.org/10.1136/adc.88.12.1065.

Ding, Y., R. Sauerborn, B. Xu, N. Shaofa, W. Yan, V. K. Diwan, and H. Dong. 2015. “A Cost-Effectiveness Analysis of Three Components of a Syndromic Surveillance System for the Early Warning of Epidemics in Rural China.” *BMC Public Health* 15 (1127). https://doi.org/10.1186/s12889-015-2475-x.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. García Marquéz, et al. 2013. “Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance.” *Ecography* 36 (1): 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x.

Dowell, S. F. 2001. “Seasonal Variation in Host Susceptibility and Cycles of Certain Infectious Diseases.” *Emerging Infectious Diseases* 7 (3): 369–74. https://doi.org/10.3201/eid0703.010301.

du Prel, J-B., W. Puppe, B. Gröndahl, M. Knuf, F. Weigl, F. Schaaff, F. Schaaff, and H-J. Schmitt. 2009. “Are Meteorological Parameters Associated with Acute Respiratory Tract Infections?” *Clinical Infectious Diseases* 49 (6): 861–68. https://doi.org/10.1086/605435.

Dwyer-Lindgren, L., M. A. Cork, A. Sligar, K. M. Steuben, K. F. Wilson, N. R. Provost, B. K. Mayala, et al. 2019. “Mapping HIV Prevalence in Sub-Saharan Africa between 2000 and 2017.” *Nature* 570 (7760): 189–93. https://doi.org/10.1038/s41586-019-1200-9.

Eames, K. T. D., N. L. Tilston, E. Brooks-Pollock, and W. J. Edmunds. 2012. "Measured Dynamic Social Contact Patterns Explain the Spread of H1N1v Influenza." *PLOS Computational Biology* 8 (3): e1002425. https://doi.org/10.1371/journal.pcbi.1002425.

Earp, B. D., J. T. Monrad, M. LaFrance, J. A Bargh, L. L. Cohen, and J. A. Richeson. 2019. "Featured Article: Gender Bias in Pediatric Pain Assessment." *Journal of Pediatric Psychology* 44 (4): 403–414. https://doi.org/10.1093/jpepsy/jsy104.

Eccles, R. 2002. "An Explanation for the Seasonality of Acute Upper Respiratory Tract Viral Infections." *Acta Oto-Laryngologica* 122 (2): 183–91. https://doi.org/10.1080/00016480252814207.

Edwards, A., and R. Pill. 1996. "Patterns of Help-Seeking Behaviour for Toddlers from Two Contrasting Socio-Economic Groups: New Evidence on a Neglected Topic." *Family Practice* 13 (4): 377–81. https://doi.org/10.1093/fampra/13.4.377.

Egger, J. R., A. G. Hoen, J. S. Brownstein, D. L. Buckeridge, D. R. Olson, and K. J. Konty. 2012. "Usefulness of School Absenteeism Data for Predicting Influenza Outbreaks, United States." *Emerging Infectious Diseases* 18 (8): 1375–77. https://doi.org/10.3201/eid1808.111538.

El Guerche-Séblain, C., A. Moureau, C. Schiffler, M. Dupuy, S. Pepin, S. I. Samson, P. Vanhems, and F. Schellevis. 2019. "Epidemiology and Burden of Influenza in Healthy Children Aged 6 to 35 Months: Analysis of Data from the Placebo Arm of a Phase III Efficacy Trial." *BMC Infectious Diseases* 19 (1): 308. https://doi.org/10.1186/s12879-019-3920-8.

Elliot, A. J. 2009a. "Syndromic Surveillance: The next Phase of Public Health Monitoring during the H1N1 Influenza Pandemic?" *Eurosurveillance* 14 (44). http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19391.

Elliot, A. J, C. Powers, A. Thornton, C. Obi, C. Hill, I. Simms, P. Waight, et al. 2009b. "Monitoring the Emergence of Community Transmission of Influenza A/H1N1 2009 in England: A Cross Sectional Opportunistic Survey of Self Sampled Telephone Callers to NHS Direct." *BMJ* 339: b3403. https://doi.org/10.1136/bmj.b3403.

Elliot, A. J., N. Singh, P. Loveridge, S. Harcourt, S. Smith, R. Pnaiser, K. Kavanagh, et al. 2010. "Syndromic Surveillance to Assess the Potential Public Health Impact of the Icelandic Volcanic Ash Plume across the United

Kingdom, April 2010." *Eurosurveillance* 15 (23): 19583. https://doi.org/10.2807/ese.15.23.19583-en.

Elliot, A. J., R. A. Morbey, H. E. Hughes, S. E. Harcourt, S. Smith, P. Loveridge, O. Edeghere, et al. 2013. "Syndromic Surveillance - a Public Health Legacy of the London 2012 Olympic and Paralympic Games." *Public Health* 127 (8): 777–81. https://doi.org/10.1016/j.puhe.2013.05.007.

Elliot, A. J., A. Bone, R. A. Morbey, H. E. Hughes, S. Harcourt, S. Smith, P. Loveridge, et al. 2014. "Using Real-Time Syndromic Surveillance to Assess the Health Impact of the 2013 Heatwave in England." *Environmental Research* 135 (November): 31–36. https://doi.org/10.1016/j.envres.2014.08.031.

Elliot, A. J., S. Smith, A. Dobney, J. Thornes, G. E. Smith, and S. Vardoulakis. 2016a. "Monitoring the Effect of Air Pollution Episodes on Health Care Consultations and Ambulance Call-Outs in England during March/April 2014: A Retrospective Observational Analysis." *Environmental Pollution* 214: 903–11. https://doi.org/10.1016/j.envpol.2016.04.026.

Elliot, A. J., H. E Hughes, J. Astbury, G. Nixon, K. Brierley, R. Vivancos, T. Inns, et al. 2016b. "The Potential Impact of Media Reporting in Syndromic Surveillance: An Example Using a Possible Cryptosporidium Exposure in North West England, August to September 2015." *Eurosurveillance* 21 (41). https://doi.org/10.2807/1560-7917.ES.2016.21.41.30368.

Elliot, A. J., R. A. Morbey, O. Edeghere, I. R. Lake, F. J. Colón-González, R. Vivancos, G. J. Rubin, S. J. O'Brien, and G. E. Smith. 2017. "Developing a Multidisciplinary Syndromic Surveillance Academic Research Program in the United Kingdom: Benefits for Public Health Surveillance." *Public Health Reports* 132 (1): 111S–115S. https://doi.org/10.1177/0033354917706953.

Elliot, A. J., S. E. Harcourt, H. E. Hughes, P. Loveridge, R. A. Morbey, S. Smith, A. Soriano, et al. 2020. "The COVID-19 Pandemic: A New Challenge for Syndromic Surveillance." *Epidemiology & Infection* 148. https://doi.org/10.1017/S0950268820001314.

Emukule, G. O., P. Spreeuwenberg, S. S. Chaves, J. A. Mott, S. Tempia, G. Bigogo, B. Nyawanda, et al. 2017. "Estimating Influenza and Respiratory Syncytial Virus-Associated Mortality in Western Kenya Using Health and Demographic Surveillance System Data, 2007-2013." *PLOS ONE* 12 (7): 1–15. https://doi.org/10.1371/journal.pone.0180890.

Falagas, M. E., E. G. Mourtzoukou, and K. Z. Vardakas. 2007. “Sex Differences in the Incidence and Severity of Respiratory Tract Infections.” *Respiratory Medicine* 101 (9): 1845–63. https://doi.org/10.1016/j.rmed.2007.04.011.

Farmer, J., L. Iversen, N. C. Campbell, C. Guest, R. Chesson, G. Deans, and J. MacDonald. 2006. “Rural/Urban Differences in Accounts of Patients’ Initial Decisions to Consult Primary Care.” *Health & Place* 12 (2): 210–21. https://doi.org/10.1016/j.healthplace.2004.11.007.

Feng, Y., T. Marchal, T. Sperry, and H. Yi. 2020. “Influence of Wind and Relative Humidity on the Social Distancing Effectiveness to Prevent COVID-19 Airborne Transmission: A Numerical Study.” *Journal of Aerosol Science* 147: 105585. https://doi.org/10.1016/j.jaerosci.2020.105585.

Fiscella, Kevin, and Kathleen Holt. 2007. “Impact of Primary Care Patient Visits on Racial and Ethnic Disparities in Preventive Care in the United States.” *Journal of the American Board of Family Medicine* 20 (6): 587–97. https://doi.org/10.3122/jabfm.2007.06.070053.

Fisman, D. 2012. “Seasonality of Viral Infections: Mechanisms and Unknowns.” *Clinical Microbiology and Infection* 18 (10): 946–54. https://doi.org/10.1111/j.1469-0691.2012.03968.x.

Fleming, D., R. Pannell, and K. Cross. 2005a. “Mortality in Children from Influenza and Respiratory Syncytial Virus.” *Journal of Epidemiology and Community Health* 59 (7): 586–90. https://doi.org/10.1136/jech.2004.026450.

Fleming, D., K. W. Cross, and R. S. Pannell. 2005b. “Influenza and Its Relationship to Circulatory Disorders.” *Epidemiology and Infection* 133 (2): 255–62. https://doi.org/10.1017/s0950268804003231.

Fleming, D., R. J. Taylor, R. L. Lustig, C. Schuck-Paim, F. Haguinet, D. J. Webb, J. Logie, G. Matias, and S. Taylor. 2015. “Modelling Estimates of the Burden of Respiratory Syncytial Virus Infection in Adults and the Elderly in the United Kingdom.” *BMC Infectious Diseases* 15: 443. https://doi.org/10.1186/s12879-015-1218-z.

Fleming, D. M., R. J. Taylor, F. Haguinet, C. Schuck-Paim, J. Logie, D. J. Webb, R. L. Lustig, and G. Matias. 2016. “Influenza-Attributable Burden in United Kingdom Primary Care.” *Epidemiology and Infection* 144 (3): 537–547. https://doi.org/10.1017/S0950268815001119.

Foley, D., E. Best, N. Reid, and M. J. Berry. 2019. “Respiratory Health Inequality Starts Early: The Impact of Social Determinants on the Aetiology and

Severity of Bronchiolitis in Infancy." *Journal of Paediatrics and Child Health* 55 (5): 528–32. https://doi.org/10.1111/jpc.14234.

Ford, E. S., D. M. Mannino, A. G. Wheaton, W. H. Giles, L. Presley-Cantrell, and J. B. Croft. 2013. "Trends in the Prevalence of Obstructive and Restrictive Lung Function among Adults in the United States: Findings from the National Health and Nutrition Examination Surveys from 1988-1994 to 2007-2010." *Chest* 143 (5): 1395–1406. https://doi.org/10.1378/chest.12-1135.

Frosst, G. O., S. E. Majowicz, and V. L. Edge. 2006. "Factors Associated with the Use of Over-the-Counter Medications in Cases of Acute Gastroenteritis in Hamilton, Ontario." *ResearchGate* 97 (6): 489–93.

FSA (Food Standards Agency). 2000. "A Report of the Study of Infectious Intestinal Disease in England." London HMSO. https://www.food.gov.uk/sites/default/files/media/document/iid1_study_final_report.pdf.

Gaffin, J. M., and W. Phipatanakul. 2009. "The Role of Indoor Allergens in The Development of Asthma." *Current Opinion in Allergy and Clinical Immunology* 9 (2): 128–35.

Galdas, Paul M., Francine Cheater, and Paul Marshall. 2005. "Men and Health Help-Seeking Behaviour: Literature Review." *Journal of Advanced Nursing* 49 (6): 616–23. https://doi.org/10.1111/j.1365-2648.2004.03331.x.

Gartner, A., D. Farewell, F. Dunstan, and E. Gordon. 2008. "Differences in Mortality between Rural and Urban Areas in England and Wales, 2002-04." *Health Statistics Quarterly* 39: 6–13.

Gault, G., S. Larrieu, C. Durand, L. Josseran, B. Jouves, and L. Filleul. 2009. "Performance of a Syndromic System for Influenza Based on the Activity of General Practitioners, France." *Journal of Public Health* 31(2), 286–292. https://doi.org/10.1093/pubmed/fdp020.

GBDCN (Global Burden of Disease Collaborative Network). 2012. "Global Burden of Disease Study 2010 (GBD 2010) Results by Risk Factor 1990-2010." Institute for Health Metrics and Evaluation (IHME), Seattle, USA.

Gelfand, A. E., and A. F. M. Smith. 1990. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* 85 (410): 398–409. https://doi.org/10.2307/2289776.

Geyer, S., R. Peter, and J. Siegrist. 2002. "Socioeconomic Differences in Children's and Adolescents' Hospital Admissions in Germany: A Report

Based on Health Insurance Data on Selected Diagnostic Categories." *Journal of Epidemiology and Community Health* 56 (2): 109–14. https://doi.org/10.1136/jech.56.2.109.

Gibbons, C. L., M. J. Mangen, D. Plass, A. H. Havelaar, R. J. Brooke, P. Kramarz, K. L. Peterson, et al. 2014. "Measuring Underreporting and Under-Ascertainment in Infectious Disease Datasets: A Comparison of Methods." *BMC Public Health* 14: 147. https://doi.org/10.1186/1471-2458-14-147.

Gibson, M., M. Petticrew, C. Bambra, A. J. Sowden, K. E. Wright, and M. Whitehead. 2011. "Housing and Health Inequalities: A Synthesis of Systematic Reviews of Interventions Aimed at Different Pathways Linking Housing and Health." *Health & Place* 17 (1): 175–84. https://doi.org/10.1016/j.healthplace.2010.09.011.

Ginsberg, J., M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant. 2009. "Detecting Influenza Epidemics Using Search Engine Query Data." *Nature* 457 (7232): 1012–14. https://doi.org/10.1038/nature07634.

Glezen, W. P., L. H. Taber, A. L. Frank, and J. A. Kasel. 1986. "Risk of Primary Infection and Reinfection with Respiratory Syncytial Virus." *American Journal of Diseases of Children* 140 (6): 543–46.

Goldstein, E., H. H. Nguyen, P. Liu, C. Viboud, C. A. Steiner, C. J. Worby, and M. Lipsitch. 2018. "On the Relative Role of Different Age Groups During Epidemics Associated with Respiratory Syncytial Virus." *The Journal of Infectious Diseases* 217 (2): 238–44. https://doi.org/10.1093/infdis/jix575.

Graham, S. E., and T. McCurdy. 2004. "Developing Meaningful Cohorts for Human Exposure Models." *Journal of Exposure Analysis and Environmental Epidemiology* 14 (1): 23–43. https://doi.org/10.1038/sj.jea.7500293.

Grassly, N. C, and C. Fraser. 2006. "Seasonal Infectious Disease Epidemiology." *Proceedings of the Royal Society B: Biological Sciences* 273 (1600): 2541–50. https://doi.org/10.1098/rspb.2006.3604.

Green, H. K., N. Andrews, L. Letley, A. Sunderland, J. White, and R. Pebody. 2015. "Phased Introduction of a Universal Childhood Influenza Vaccination Programme in England: Population-Level Factors Predicting Variation in National Uptake during the First Year, 2013/14." *Vaccine* 33 (22): 2620–28. https://doi.org/10.1016/j.vaccine.2015.03.049.

Gruer, L D, G I Cézard, L A Wallace, S J Hutchinson, A F Douglas, D Buchanan, S V Katikireddi, et al. 2021. "Complex Differences in Infection Rates between Ethnic Groups in Scotland: A Retrospective, National Census-Linked

Cohort Study of 1.65 Million Cases." *Journal of Public Health*. https://doi.org/10.1093/pubmed/fdaa267.

Hair, J., W. Black, B. Babin, and R. Anderson. 2010. *Multivariate Data Analysis*. 7th ed. Harlow, Essex: Pearson Education Limited. https://www.pearson.com/us/higher-education/program/Hair-Multivariate-Data-Analysis-7th-Edition/PGM263675.html.

Hall, C., and C. A. McCarthy. 2000. "Respiratory Syncytial Virus." In *Mandell GL, Bennett JE, Dolin R, Eds. Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*, 5th ed., 2:1782–1801. Philadelphia: Churchill Livingstone.

Hall, C. Breese, G. A. Weinberg, M. K. Iwane, A. K. Blumkin, K. M. Edwards, M. A. Staat, P. Auinger, et al. 2009. "The Burden of Respiratory Syncytial Virus Infection in Young Children." *The New England Journal of Medicine* 360 (6): 588–98. https://doi.org/10.1056/NEJMoa0804877.

Hanif W., and R. Susarla. 2018. "Diabetes and Cardiovascular Risk in UK South Asians: An Overview." *The British Journal of Cardiology* 25 (2). https://doi.org/10.5837/bjc.2018.s08.

Harcourt, S. E., G. E. Smith, A. J. Elliot, R. Pebody, A. Charlett, S. Ibbotson, M. Regan, and J. Hippisley-Cox. 2012a. "Use of a Large General Practice Syndromic Surveillance System to Monitor the Progress of the Influenza A(H1N1) Pandemic 2009 in the UK." *Epidemiology & Infection* 140 (1): 100–105. https://doi.org/10.1017/S095026881100046X.

Harcourt, S. E., J. Fletcher, P. Loveridge, A. Bains, R. A. Morbey, A. Yeates, B. McCloskey, et al. 2012b. "Developing a New Syndromic Surveillance System for the London 2012 Olympic and Paralympic Games." *Epidemiology & Infection* 140 (12): 2152–56. https://doi.org/10.1017/S0950268812001781.

Harcourt, S. E., R. A. Morbey, P. Loveridge, L. Carrilho, D. Baynham, E. Povey, P. Fox, et al. 2016. "Developing and Validating a New National Remote Health Advice Syndromic Surveillance System in England." *Journal of Public Health*, March, fdw013. https://doi.org/10.1093/pubmed/fdw013.

Harcourt, S., R. A. Morbey, C. Bates, H. Carter, S. N. Ladhani, S. de Lusignan, G. E. Smith, and A. J. Elliot. 2018. "Estimating Primary Care Attendance Rates for Fever in Infants after Meningococcal B Vaccination in England Using National Syndromic Surveillance Data." *Vaccine* 36 (4): 565–71. https://doi.org/10.1016/j.vaccine.2017.11.076.

Harcourt, S. E., R. A. Morbey, G. E. Smith, P. Loveridge, H. K. Green, R. Pebody, J. Rutter, F. A. Yeates, G. Stuttard, and A. J. Elliot. 2019. "Developing Influenza and Respiratory Syncytial Virus Activity Thresholds for Syndromic Surveillance in England." *Epidemiology and Infection* 147: e163. https://doi.org/10.1017/S0950268819000542.

Hardelid, P., R. Pebody, and N. Andrews. 2013. "Mortality Caused by Influenza and Respiratory Syncytial Virus by Age Group in England and Wales 1999-2010." *Influenza and Other Respiratory Viruses* 7 (1): 35–45. https://doi.org/10.1111/j.1750-2659.2012.00345.x.

Harding, S., and M. Rosato. 1999. "Cancer Incidence among First Generation Scottish, Irish, West Indian and South Asian Migrants Living in England and Wales." *Ethnicity & Health* 4 (1–2): 83–92. https://doi.org/10.1080/13557859998218.

Hardnett, F. P., R. M. Hoekstra, M. Kennedy, L. Charles, F. J. Angulo, and Emerging Infections Program Food Net Working Group. 2004. "Epidemiologic Issues in Study Design and Data Analysis Related to FoodNet Activities." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 38 (3): S121-126. https://doi.org/10.1086/381602.

Haroon, S. M.M., G. P. Barbosa, and P. J. Saunders. 2011. "The Determinants of Health-Seeking Behaviour during the A/H1N1 Influenza Pandemic: An Ecological Study." *Journal of Public Health* 33 (4): 503–10. https://doi.org/10.1093/pubmed/fdr029.

Harper, G. J. 1961. "Airborne Micro-Organisms: Survival Tests with Four Viruses." *The Journal of Hygiene* 59 (4): 479–86.

Hastings, W. K. 1970. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika* 57 (1): 97–109. https://doi.org/10.1093/biomet/57.1.97.

Hawker, J. I, B. Olowokure, F. Sufi, J. Weinberg, N. Gill, and R. C Wilson. 2003. "Social Deprivation and Hospital Admission for Respiratory Infection: An Ecological Study." *Respiratory Medicine* 97 (11): 1219–24. https://doi.org/10.1016/S0954-6111(03)00252-X.

Hay, S. I, A. Alemu Abajobir, K. Hassen Abate, C. Abbafati, K. M. Abbas, F. Abd-Allah, R. Suliankatchi Abdulkader, et al. 2017. "Global, Regional, and National Disability-Adjusted Life-Years (DALYs) for 333 Diseases and Injuries and Healthy Life Expectancy (HALE) for 195 Countries and

Territories, 1990–2016: A Systematic Analysis for the Global Burden of Disease Study 2016." *The Lancet* 390 (10100): 1260–1344. https://doi.org/10.1016/S0140-6736(17)32130-X.

Haynes, A. K., A. P. Manangan, M. K. Iwane, K. Sturm-Ramirez, N. Homaira, W. Abdullah Brooks, S. Luby, et al. 2013. "Respiratory Syncytial Virus Circulation in Seven Countries with Global Disease Detection Regional Centers." *The Journal of Infectious Diseases* 208 (3): S246-254. https://doi.org/10.1093/infdis/jit515.

Heffernan, R., F. Mostashari, D. Das, A. M. Karpati, M Kulldorfft, and D. Weiss. 2004. "Syndromic Surveillance in Public Health Practice, New York City - Volume 10, Number 5—May 2004 - Emerging Infectious Disease Journal - CDC." *Emerging Infectious Diseases* 10. https://doi.org/10.3201/eid1005.030646.

Heiden, M. an der, U. Buchholz, and S. Buda. 2019. "Estimation of Influenza- and Respiratory Syncytial Virus-Attributable Medically Attended Acute Respiratory Infections in Germany, 2010/11-2017/18." *Influenza and Other Respiratory Viruses* 13 (5): 517–21. https://doi.org/10.1111/irv.12666.

Heikkinen, T., and T. Chonmaitree. 2003. "Importance of Respiratory Viruses in Acute Otitis Media." *Clinical Microbiology Reviews* 16 (2): 230–41. https://doi.org/10.1128/CMR.16.2.230-241.2003.

Heikkinen, T., E. Ojala, and M. Waris. 2017. "Clinical and Socioeconomic Burden of Respiratory Syncytial Virus Infection in Children." *The Journal of Infectious Diseases* 215 (1): 17–23. https://doi.org/10.1093/infdis/jiw475.

Henning, K.. 2003. *Microbial Threats to Health: Emergence, Detection, and Response. Appendix B: Syndromic Surveillance*. Washington (DC): National Academies Press (US).

Hens, N., N. Goeyvaerts, M. Aerts, Z. Shkedy, P. Van Damme, and P. Beutels. 2009. "Mining Social Mixing Patterns for Infectious Disease Models Based on a Two-Day Population Survey in Belgium." *BMC Infectious Diseases* 9 (1): 5. https://doi.org/10.1186/1471-2334-9-5.

Herbert, S., G. Leong, K. Hewitt, and J. Cassell. 2015. "Do Genitourinary Physicians Report Notifiable Diseases? A Survey in South East England." *International Journal of STD & AIDS* 26 (3): 173–80. https://doi.org/10.1177/0956462414531932.

Hervás, D., J. Reina, and J. A. Hervás. 2012. “Meteorologic Conditions and Respiratory Syncytial Virus Activity.” *The Pediatric Infectious Disease Journal* 31 (10): e176-181. https://doi.org/10.1097/INF.0b013e31825cef14.

Higgins, D., C. Trujillo, and C. Keech. 2016. “Advances in RSV Vaccine Research and Development - A Global Agenda.” *Vaccine* 34 (26): 2870–75. https://doi.org/10.1016/j.vaccine.2016.03.109.

Hoeppner, T., M. Borland, F. E. Babl, J. Neutze, N. Phillips, D. Krieser, S. R. Dalziel, et al. 2017. “Influence of Weather on Incidence of Bronchiolitis in Australia and New Zealand.” *Journal of Paediatrics and Child Health* 53 (10): 1000–1006. https://doi.org/10.1111/jpc.13614.

Hogan, A. B., R. S. Anderssen, S. Davis, H. C. Moore, F. J. Lim, P. Fathima, and K. Glass. 2016. “Time Series Analysis of RSV and Bronchiolitis Seasonality in Temperate and Tropical Western Australia.” *Epidemics* 16: 49–55. https://doi.org/10.1016/j.epidem.2016.05.001.

Hoyer, S., and J. Hamman. 2017. “Xarray: N-D Labeled Arrays and Datasets in Python.” *Journal of Open Research Software* 5 (1): 10. https://doi.org/10.5334/jors.148.

Hughes, H. E., R. A. Morbey, T. C. Hughes, T. E. Locker, T. Shannon, C. Carmichael, V. Murray, et al. 2014. “Using an Emergency Department Syndromic Surveillance System to Investigate the Impact of Extreme Cold Weather Events.” *Public Health* 128 (7): 628–35. https://doi.org/10.1016/j.puhe.2014.05.007.

Hughes, H. E., R. A. Morbey, T. C. Hughes, T. E. Locker, R. Pebody, H. K. Green, J. Ellis, G. E. Smith, and A. J. Elliot. 2016. “Emergency Department Syndromic Surveillance Providing Early Warning of Seasonal Respiratory Activity in England.” *Epidemiology & Infection* 144 (5): 1052–64. https://doi.org/10.1017/S0950268815002125.

Hulth, A., G. Rydevik, and A. Linde. 2009. “Web Queries as a Source for Syndromic Surveillance.” *PLOS ONE* 4 (2): e4378. https://doi.org/10.1371/journal.pone.0004378.

Huotari, A., and K-H. Herzig. 2008. “Vitamin D and Living in Northern Latitudes--an Endemic Risk Area for Vitamin D Deficiency.” *International Journal of Circumpolar Health* 67 (2–3): 164–78. https://doi.org/10.3402/ijch.v67i2-3.18258.

Iacobucci, Gareth. 2020. “Covid Lockdown: England Sees Fewer Cases of Colds, Flu, and Bronchitis.” *BMJ* 370: m3182. https://doi.org/10.1136/bmj.m3182.

Ibitoye, M., T. Frasca, R. Giguere, and A. Carballo-Diéguez. 2014. “Home Testing Past, Present and Future: Lessons Learned and Implications for HIV Home Tests (A Review).” *AIDS and Behaviour* 18 (5): 933–49. https://doi.org/10.1007/s10461-013-0668-9.

Imai, C., B. Armstrong, Z. Chalabi, P. Mangtani, and M. Hashizume. 2015. “Time Series Regression Model for Infectious Disease and Weather.” *Environmental Research* 142: 319 – 327. https://doi.org/10.1016/j.envres.2015.06.040.

Iversen, L., P. C. Hannaford, D. B. Price, and D. J. Godden. 2005. “Is Living in a Rural Area Good for Your Respiratory Health? Results from a Cross-Sectional Study in Scotland.” *Chest* 128 (4): 2059–67. https://doi.org/10.1378/chest.128.4.2059.

Izurieta, H. S., W. W. Thompson, P. Kramarz, D. K. Shay, R. L. Davis, F. DeStefano, S. Black, H. Shinefield, and K. Fukuda. 2000. “Influenza and the Rates of Hospitalization for Respiratory Disease among Infants and Young Children.” *The New England Journal of Medicine* 342 (4): 232–39. https://doi.org/10.1056/NEJM200001273420402.

Jaakkola, K., A. Saukkoriipi, J. Jokelainen, R. Juvonen, J. Kauppila, O. Vainio, T. Ziegler, E. Rönkkö, J. Jaakkola, and T. M Ikäheimo. 2014. “Decline in Temperature and Humidity Increases the Occurrence of Influenza in Cold Climate.” *Environmental Health* 13: 22. https://doi.org/10.1186/1476-069X-13-22.

Jackson, S., K. H. Mathews, D. Pulanić, R. Falconer, I. Rudan, H. Campbell, and H. Nair. 2013. “Risk Factors for Severe Acute Lower Respiratory Infections in Children – a Systematic Review and Meta-Analysis.” *Croatian Medical Journal* 54 (2): 110–21. https://doi.org/10.3325/cmj.2013.54.110.

Janet, S., J. Broad, and M. D. Snape. 2017. “Respiratory Syncytial Virus Seasonality and Its Implications on Prevention Strategies.” *Human Vaccines & Immunotherapeutics* 14 (1): 234–44. https://doi.org/10.1080/21645515.2017.1403707.

Jartti, T., P. Lehtinen, T. Vuorinen, and O. Ruuskanen. 2009. “Bronchiolitis: Age and Previous Wheezing Episodes Are Linked to Viral Etiology and Atopic Characteristics.” *The Pediatric Infectious Disease Journal* 28 (4): 311–17. https://doi.org/10.1097/INF.0b013e31818ee0c1.

Jeffery, C., A. Ozonoff, L. F. White, M. Nuño, and M. Pagano. 2009. “Power to Detect Spatial Disturbances under Different Levels of Geographic Aggregation.” *Journal of the American Medical Informatics Association* 16 (6): 847–54. https://doi.org/10.1197/jamia.M2788.

Jeffery, C., A. Ozonoff, and M. Pagano. 2014. “The Effect of Spatial Aggregation on Performance When Mapping a Risk of Disease.” *International Journal of Health Geographics* 13 (1): 9. https://doi.org/10.1186/1476-072X-13-9.

Johnston, R., K. Jones, and D. Manley. 2018. “Confounding and Collinearity in Regression Analysis: A Cautionary Tale and an Alternative Procedure, Illustrated by Studies of British Voting Behaviour.” *Quality & Quantity* 52 (4): 1957–76. https://doi.org/10.1007/s11135-017-0584-6.

Jordan, R., N. Verlander, B. Olowokure, and J. I. Hawker. 2006. “Age, Sex, Material Deprivation and Respiratory Mortality.” *Respiratory Medicine* 100 (7): 1282–85. https://doi.org/10.1016/j.rmed.2005.10.014.

Josseran, L., J. Nicolau, N. Callere, P. Astagneau, and G. Brucker. 2006. “Syndromic Surveillance Based on Emergency Department Activity and Crude Mortality: Two Examples.” *Eurosurveillance* 11 (12). http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=668.

Karter, A. J. 2003. “Commentary: Race, Genetics, and Disease— in Search of a Middle Ground.” *International Journal of Epidemiology* 32 (1): 26–28. https://doi.org/10.1093/ije/dyg033.

Kelly, J. S., H. Piercy, R. Ibbotson, and S. V. Fowler Davis. 2018. “Who Attends Out-of-Hours General Practice Appointments? Analysis of a Patient Cohort Accessing New out-of-Hours Units.” *BMJ Open* 8 (6): e020308. https://doi.org/10.1136/bmjopen-2017-020308.

Keramarou, M., and M. R. Evans. 2012. “Completeness of Infectious Disease Notification in the United Kingdom: A Systematic Review.” *The Journal of Infection* 64 (6): 555–64. https://doi.org/10.1016/j.jinf.2012.03.005.

Khan, A., A. Levitt, and M. Sage. 2000. “Biological and Chemical Terrorism: Strategic Plan for Preparedness and Response.” *CDC MMWR Recommendations and Reports* 49 (RR04) (April): 1–14.

Killerby, M. E., H. M. Biggs, A. Haynes, R. M. Dahl, D. Mustaquim, S. I. Gerber, and J. T. Watson. 2018. “Human Coronavirus Circulation in the United States 2014-2017.” *Journal of Clinical Virology: The Official Publication of the Pan*

American Society for Clinical Virology 101: 52–56. https://doi.org/10.1016/j.jcv.2018.01.019.

Killick, R., P. Fearnhead, and I. A. Eckley. 2012. "Optimal Detection of Changepoints with a Linear Computational Cost." *Journal of the American Statistical Association* 107 (500): 1590–98. https://doi.org/10.1080/01621459.2012.737745.

Killick, R., and I. A. Eckley. 2014. "Changepoint: An R Package for Changepoint Analysis." *Journal of Statistical Software* 58 (1): 1–19. https://doi.org/10.18637/jss.v058.i03.

Kim, J., and H. Bang. 2016. "Three Common Misuses of P Values." *Dental Hypotheses* 7 (3): 73–80. https://doi.org/10.4103/2155-8213.190481.

Kom M., C. Aïcha, G. De Serres, M. Douville Fradet, G. Lebel, S. Toutant, R. Gilca, M. Ouakki, N. Zafar Janjua, and D. M. Skowronski. 2012. "School Absenteeism as an Adjunct Surveillance Indicator: Experience during the Second Wave of the 2009 H1N1 Pandemic in Quebec, Canada." *PLoS ONE* 7 (3). https://doi.org/10.1371/journal.pone.0034084.

Koski, E. 2011. "Clinical Laboratory Data for Biosurveillance." In *Infectious Disease Informatics and Biosurveillance: Research, Systems and Case Studies*, edited by Carlos Castillo-Chavez, Hsinchun Chen, William B. Lober, Mark Thurmond, and Daniel Zeng, 67–87. Integrated Series in Information Systems. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4419-6892-0_4.

Kotaniemi-Syrjänen, A., T. M. Reijonen, K. Korhonen, M. Waris, R. Vainionpää, and M. Korppi. 2008. "Wheezing Due to Rhinovirus Infection in Infancy: Bronchial Hyperresponsiveness at School Age." *Pediatrics International: Official Journal of the Japan Pediatric Society* 50 (4): 506–10. https://doi.org/10.1111/j.1442-200X.2008.02620.x.

Kpokiri, Eneyi E, Gifty Marley, Weiming Tang, Noah Fongwen, Dan Wu, Sima Berendes, Bhavana Ambil, et al. 2020. "Diagnostic Infectious Diseases Testing Outside Clinics: A Global Systematic Review and Meta-Analysis." *Open Forum Infectious Diseases* 7 (10). https://doi.org/10.1093/ofid/ofaa360.

Lajoie, J. 2013. "Understanding the Measurement of Global Burden of Disease." National Collaborating Centre for Infectious Diseases. https://nccid.ca/publications/understanding-the-measurement-of-global-burden-of-disease/.

Lake, I. R., G. Nichols, F. C. D. Harrison, G. Bentham, R. S. Kovats, C. Grundy, and P. R. Hunter. 2009. "Using Infectious Intestinal Disease Surveillance Data to Explore Illness Aetiology; a Cryptosporidiosis Case Study." *Health & Place* 15 (1): 333–39. https://doi.org/10.1016/j.healthplace.2008.06.005.

Lam, H. C-Y., A. M. Li, E. Y-Y. Chan, and W. B. Goggins. 2016. "The Short-Term Association between Asthma Hospitalisations, Ambient Temperature, Other Meteorological Factors and Air Pollutants in Hong Kong: A Time-Series Study." *Thorax* 71 (12): 1097–1109. https://doi.org/10.1136/thoraxjnl-2015-208054.

Landes, M. B., R. Brock Neil, S. S. McCool, B. P. Mason, A. M. Woron, R. L. Garman, and D. L. Smalley. 2013. "The Frequency and Seasonality of Influenza and Other Respiratory Viruses in Tennessee: Two Influenza Seasons of Surveillance Data, 2010-2012." *Influenza and Other Respiratory Viruses* 7 (6): 1122–27. https://doi.org/10.1111/irv.12145.

Lateef, F. 2012. "Syndromic Surveillance: A Necessary Public Health Tool." *Journal of Acute Disease* 1 (2): 90–93. https://doi.org/10.1016/S2221-6189(13)60022-0.

Latunji, O.O., and O.O. Akinyemi. 2018. "Factors Influencing Health-Seeking Behaviour Among Civil Servants in Ibadan, Nigeria." *Annals of Ibadan Postgraduate Medicine* 16 (1): 52–60.

Lazer, D., R. Kennedy, G. King, and A. Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis." *Science* 343 (6176): 1203–5. https://doi.org/10.1126/science.1248506.

Leader, S., and K. Kohlhase. 2003. "Recent Trends in Severe Respiratory Syncytial Virus (RSV) among US Infants, 1997 to 2000." *The Journal of Pediatrics* 143 (5): S127-132.

Leary, P. F., I. Zamfirova, J. Au, and W. H. McCracken. 2017. "Effect of Latitude on Vitamin D Levels." *The Journal of the American Osteopathic Association* 117 (7): 433–39. https://doi.org/10.7556/jaoa.2017.089.

Lee, W-M., R. F. Lemanske, M. D. Evans, F. Vang, T. Pappas, R. Gangnon, D. J. Jackson, and J. E. Gern. 2012. "Human Rhinovirus Species and Season of Infection Determine Illness Severity." *American Journal of Respiratory and Critical Care Medicine* 186 (9): 886–91. https://doi.org/10.1164/rccm.201202-0330OC.

Lewis, K. M., B. De Stavola, and P. Hardelid. 2020. “Geospatial and Seasonal Variation of Bronchiolitis in England: A Cohort Study Using Hospital Episode Statistics.” *Thorax* 75 (3): 262–68. https://doi.org/10.1136/thoraxjnl-2019-213764.

Leynaert, B., J. Sunyer, R. Garcia-Esteban, C. Svanes, D. Jarvis, I. Cerveri, J. Dratva, et al. 2012. “Gender Differences in Prevalence, Diagnosis and Incidence of Allergic and Non-Allergic Asthma: A Population-Based Cohort.” *Thorax* 67 (7): 625–31. https://doi.org/10.1136/thoraxjnl-2011-201249.

Librero, J., B. Ibañez, N. Martínez-Lizaga, S. Peiró, E. Bernal-Delgado, and on behalf of the Spanish Atlas of Medical Practice Variation Research Group. 2017. “Applying Spatio-Temporal Models to Assess Variations across Health Care Areas and Regions: Lessons from the Decentralized Spanish National Health System.” *PLOS ONE* 12 (2): e0170480. https://doi.org/10.1371/journal.pone.0170480.

Linde, A., M. Rotzén-Ostlund, B. Zweygberg-Wirgart, S. Rubinova, and M. Brytting. 2009. “Does Viral Interference Affect Spread of Influenza?” *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin* 14 (40).

Liptzin, D. R., L. I. Landau, and L. M. Taussig. 2015. “Sex and the Lung: Observations, Hypotheses, and Future Directions.” *Pediatric Pulmonology* 50 (12): 1159–69. https://doi.org/10.1002/ppul.23178.

Liu, Y., Y. Guo, C. Wang, W. Li, J. Lu, S. Shen, H. Xia, J. He, and X. Qiu. 2015. “Association between Temperature Change and Outpatient Visits for Respiratory Tract Infections among Children in Guangzhou, China.” *International Journal of Environmental Research and Public Health* 12 (1): 439–54. https://doi.org/10.3390/ijerph120100439.

Lo, D., J. L. Kennedy, R. C. Kurten, R. A. Panettieri, and C. J. Koziol-White. 2018. “Modulation of Airway Hyperresponsiveness by Rhinovirus Exposure.” *Respiratory Research* 19 (1): 208. https://doi.org/10.1186/s12931-018-0914-9.

Lombardo, J, and D Ross. 2006. “Part I: System Design and Implementation.” In *Disease Surveillance: A Public Health Informatics Approach*. Hoboken, New Jersey: John Wiley & Sons inc. http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470068124.html.

Lowen, A. C., S. Mubareka, J. Steel, and P. Palese. 2007. “Influenza Virus Transmission Is Dependent on Relative Humidity and Temperature.” *PLOS Pathogens* 3 (10): e151. https://doi.org/10.1371/journal.ppat.0030151.

Lozano, J. E. 2018. *Second Release of the MEM Shiny Web Application R Package*.

MacDougall, L., S. E. Majowicz, K. Doré, J. Flint, K. Thomas, S. Kovacs, and P. Sockett. 2008. “Under-Reporting of Infectious Gastrointestinal Illness in British Columbia, Canada: Who Is Counted in Provincial Communicable Disease Statistics?” *Epidemiology and Infection* 136 (2): 248–56. https://doi.org/10.1017/S0950268807008461.

Mackenzie, C. S., W. L. Gekoski, and V. J. Knox. 2006. “Age, Gender, and the Underutilization of Mental Health Services: The Influence of Help-Seeking Attitudes.” *Aging & Mental Health* 10 (6): 574–82. https://doi.org/10.1080/13607860600641200.

MacKian, Sara. 2003. “A Review of Health Seeking Behaviour: Problems and Prospects.” University of Manchester: Department of International Development. https://assets.publishing.service.gov.uk/media/57a08d1de5274a27b200163d/05-03_health_seeking_behaviour.pdf.

Magruder, Steven. 2003. “Evaluation of Over-the-Counter Pharmaceutical Sales as a Possible Early Warning Indicator of Human Disease.” *Johns Hopkins APL Technical Digest* 24 (4): 349–53.

Mahase, Elisabeth. 2021. “Home Testing HPV Kits Will Be Offered to Women in England as Part of Screening Trial.” *BMJ* 372: n537. https://doi.org/10.1136/bmj.n537.

Mäkinen, T. M., R. Juvonen, J. Jokelainen, T. H. Harju, A. Peitso, A. Bloigu, S. Silvennoinen-Kassinen, M. Leinonen, and J. Hassi. 2009. “Cold Temperature and Low Humidity Are Associated with Increased Occurrence of Respiratory Tract Infections.” *Respiratory Medicine* 103 (3): 456–62. https://doi.org/10.1016/j.rmed.2008.09.011.

Mandl, K. D., J. M. Overhage, M. M. Wagner, W. B. Lober, P. Sebastiani, F. Mostashari, J. A. Pavlin, et al. 2004. “Implementing Syndromic Surveillance: A Practical Guide Informed by the Early Experience.” *Journal of the American Medical Informatics Association* 11 (2): 141–50. https://doi.org/10.1197/jamia.M1356.

Manierre, M. J. 2015. “Gaps in Knowledge: Tracking and Explaining Gender Differences in Health Information Seeking.” *Social Science & Medicine (1982)* 128: 151–58. https://doi.org/10.1016/j.socscimed.2015.01.028.

Martins, T., G, D. Simpson, F. Lindgren, and H. Rue. 2012. *Bayesian Computing with INLA: New Features*. Computational Statistics and Data Analysis. www.r-inla.org.

Marx, M. A., C. V. Rodriguez, J. Greenko, D. Das, R. Heffernan, A. M. Karpati, F. Mostashari, S. Balter, M. Layton, and D. Weiss. 2006. “Diarrheal Illness Detected Through Syndromic Surveillance After a Massive Power Outage: New York City, August 2003.” *American Journal of Public Health* 96 (3): 547–53. https://doi.org/10.2105/AJPH.2004.061358.

Matheson, F. I., K. L. W. Smith, G. S. Fazli, R. Moineddin, J. R. Dunn, and R. H. Glazier. 2014. “Physical Health and Gender as Risk Factors for Usage of Services for Mental Illness.” *Journal of Epidemiology and Community Health* 68 (10): 971–78. https://doi.org/10.1136/jech-2014-203844.

May, L., J-P. Chretien, and J. A. Pavlin. 2009. “Beyond Traditional Surveillance: Applying Syndromic Surveillance to Developing Settings – Opportunities and Challenges.” *BMC Public Health* 9: 242. https://doi.org/10.1186/1471-2458-9-242.

McCurdy, T., and S. E. Graham. 2003. “Using Human Activity Data in Exposure Models: Analysis of Discriminating Factors.” *Journal of Exposure Science & Environmental Epidemiology* 13 (4): 294–317. https://doi.org/10.1038/sj.jea.7500281.

McNulty, C., T. Nichols, D. P. French, P. Joshi, and C. C. Butler. 2013. “Expectations for Consultations and Antibiotics for Respiratory Tract Infection in Primary Care: The RTI Clinical Iceberg.” *British Journal of General Practice* 63 (612): e429–36. https://doi.org/10.3399/bjgp13X669149.

Medina, S., M. Sala-Soler, D. Cooper, M. Kanieff, C. Caserio Schonemann, C. Dupuy, A. J. Elliot, et al. 2014. “Guidelines to Implement or Improve Syndromic Surveillance Systems.” *Online Journal of Public Health Informatics* 6 (1). https://doi.org/10.5210/ojphi.v6i1.5110.

Meerhoff, T. J., J. W. Paget, J. L. Kimpen, and F. Schellevis. 2009. “Variation of Respiratory Syncytial Virus and the Relation with Meteorological Factors in Different Winter Seasons.” *The Pediatric Infectious Disease Journal* 28 (10): 860–66. https://doi.org/10.1097/INF.0b013e3181a3e949.

Melbye, H., L. Joensen, M. Bech Risør, and P. A. Halvorsen. 2012. “Symptoms of Respiratory Tract Infection and Associated Care-Seeking in Subjects with and without Obstructive Lung Disease; The Tromsø Study: Tromsø 6.” *BMC Pulmonary Medicine* 12 (1): 51. https://doi.org/10.1186/1471-2466-12-51.

Met Office. 2020. “UK Regional Climates.” Accessed August 14, 2020. https://www.metoffice.gov.uk/research/climate/maps-and-data/regional-climates/index.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. “Equation of State Calculations by Fast Computing Machines.” *Journal of Chemical Physics* 21: 1087–92. https://doi.org/10.1063/1.1699114.

Midgley, C. M., Amber K. Haynes, J. L. Baumgardner, C. Chommanard, S. W. Demas, M. M. Prill, G. R. Abedi, A. T. Curns, J. T. Watson, and S. I. Gerber. 2017. “Determining the Seasonality of Respiratory Syncytial Virus in the United States: The Impact of Increased Molecular Testing.” *The Journal of Infectious Diseases* 216 (3): 345–55. https://doi.org/10.1093/infdis/jix275.

Mikolajczyk, R. T., M. K. Akmatov, S. Rastin, And M. Kretzschmar. 2008. “Social Contacts of School Children and the Transmission of Respiratory-Spread Pathogens.” *Epidemiology and Infection* 136 (6): 813–22. https://doi.org/10.1017/S0950268807009181.

Milinovich, Gabriel J., Gail M. Williams, Archie C. A. Clements, and Wenbiao Hu. 2014. “Internet-Based Surveillance Systems for Monitoring Emerging Infectious Diseases.” *The Lancet Infectious Diseases* 14 (2): 160–68. https://doi.org/10.1016/S1473-3099(13)70244-5.

Mindell, J. S., C. S. Knott, L. S. Ng Fat, M. A. Roth, O. Manor, V. Soskolne, and N. Daoud. 2014. “Explanatory Factors for Health Inequalities across Different Ethnic and Gender Groups: Data from a National Survey in England.” *Journal of Epidemiology and Community Health* 68 (12): 1133–44. https://doi.org/10.1136/jech-2014-203927.

Molinari, Giuliano, Giselda Colombo, and Cinzia Celenza. 2014. “Respiratory Allergies: A General Overview of Remedies, Delivery Systems, and the Need to Progress.” *ISRN Allergy* 2014: 326980. https://doi.org/10.1155/2014/326980.

Monto, A. S., and B. M. Ullman. 1974. “Acute Respiratory Illness in an American Community. The Tecumseh Study.” *JAMA* 227 (2): 164–69.

Monto, A. S., and K. M. Sullivan. 1993. “Acute Respiratory Illness in the Community. Frequency of Illness and the Agents Involved.” *Epidemiology and Infection* 110 (1): 145–60. https://doi.org/10.1017/s0950268800050779.

Monto, A. S. 1994. “Studies of the Community and Family: Acute Respiratory Illness and Infection.” *Epidemiologic Reviews* 16 (2): 351–73. https://doi.org/10.1093/oxfordjournals.epirev.a036158.

Monto, A. S., S. Gravenstein, M. Elliott, M. Colopy, and J. Schweinle. 2000. “Clinical Signs and Symptoms Predicting Influenza Infection.” *Archives of Internal Medicine* 160 (21): 3243–47. https://doi.org/10.1001/archinte.160.21.3243.

Monto, A. S. 2002. “The Seasonality of Rhinovirus Infections and Its Implications for Clinical Recognition.” *Clinical Therapeutics* 24 (12): 1987–97. https://doi.org/10.1016/s0149-2918(02)80093-5.

Mooney, Stephen J, Daniel J Westreich, and Abdulrahman M El-Sayed. 2015. “Epidemiology in the Era of Big Data.” *Epidemiology* 26 (3): 390–94. https://doi.org/10.1097/EDE.0000000000000274.

Moore, Kieran. 2004. “Real-Time Syndrome Surveillance in Ontario, Canada: The Potential Use of Emergency Departments and Telehealth.” *European Journal of Emergency Medicine: Official Journal of the European Society for Emergency Medicine* 11 (1): 3–11.

Morbey, R. A., A. J. Elliot, A. Charlett, S. Ibbotson, N. Q. Verlander, S. Leach, I. Hall, et al. 2014. “Using Public Health Scenarios to Predict the Utility of a National Syndromic Surveillance Programme during the 2012 London Olympic and Paralympic Games.” *Epidemiology & Infection* 142 (5): 984–93. https://doi.org/10.1017/S095026881300188X.

Morbey, R. A., A. J. Elliot, A. Charlett, N. Andrews, N. Q. Verlander, S. Ibbotson, and G. E. Smith. 2015a. “Development and Refinement of New Statistical Methods for Enhanced Syndromic Surveillance during the 2012 Olympic and Paralympic Games.” *Health Informatics Journal* 21 (2): 159–69. https://doi.org/10.1177/1460458213517577.

Morbey, R. A., A. J. Elliot, A. Charlett, N. Q. Verlander, N. Andrews, and G. E. Smith. 2015b. “The Application of a Novel ‘Rising Activity, Multi-Level Mixed Effects, Indicator Emphasis’ (RAMMIE) Method for Syndromic Surveillance in England.” *Bioinformatics* 31 (22): 3660–65. https://doi.org/10.1093/bioinformatics/btv418.

Morbey, R. A., S. Harcourt, R. Pebody, M. Zambon, J. Hutchison, J. Rutter, H. Thomas, G. E. Smith, and A. J. Elliot. 2017a. "The Burden of Seasonal Respiratory Infections on a National Telehealth Service in England." *Epidemiology and Infection* 145 (9): 1922–32. https://doi.org/10.1017/S095026881700070X.

Morbey, R. A., A. J. Elliot, M. Zambon, R. Pebody, and G. E. Smith. 2017b. "Interpreting Specific and General Respiratory Indicators in Syndromic Surveillance." *Online Journal of Public Health Informatics* 9 (1). https://doi.org/10.5210/ojphi.v9i1.7597.

Morbey, R. A., A. J. Elliot, S. Harcourt, S. Smith, S. de Lusignan, R. Pebody, A. Yeates, M. Zambon, and G. E. Smith. 2018. "Estimating the Burden on General Practitioner Services in England from Increases in Respiratory Disease Associated with Seasonal Respiratory Pathogen Activity." *Epidemiology and Infection* 146 (11): 1389–1396. https://doi.org/10.1017/S0950268818000262.

Morikawa, S., U. Kohdera, T. Hosaka, K. Ishii, S. Akagawa, S. Hiroi, and T. Kase. 2015. "Seasonal Variations of Respiratory Viruses and Etiology of Human Rhinovirus Infection in Children." *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology* 73 (December): 14–19. https://doi.org/10.1016/j.jcv.2015.10.001.

Moriyama, M., W. J. Hugentobler, and A. Iwasaki. 2020. "Seasonality of Respiratory Viral Infections." *Annual Review of Virology*, March. https://doi.org/10.1146/annurev-virology-012420-022445.

Morrison, K. E, F. J Colón-González, R. A Morbey, P. R Hunter, J. Rutter, G. Stuttard, S. de Lusignan, et al. 2020. "Demographic and Socioeconomic Patterns in Healthcare-Seeking Behaviour for Respiratory Symptoms in England: A Comparison with Non-Respiratory Symptoms and between Three Healthcare Services." *BMJ Open* 10 (11): e038356. https://doi.org/10.1136/bmjopen-2020-038356.

Mossong, J., N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, et al. 2008. "Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases." *PLOS Medicine* 5 (3): e74. https://doi.org/10.1371/journal.pmed.0050074.

Mufson, M. A., H. D. Levine, R. E. Wasil, H. E. Mocega-Gonzalez, And H. E. Krause. 1973. "Epidemiology of Respiratory Syncytial Virus Infection Among Infants and Children in Chicago." *American Journal of Epidemiology* 98 (2): 88–95. https://doi.org/10.1093/oxfordjournals.aje.a121542.

Mukherjee, M., A, Stoddart, R. P. Gupta, B. I. Nwaru, A. Farr, M. Heaven, D. Fitzsimmons, et al. 2016. “The Epidemiology, Healthcare and Societal Burden and Costs of Asthma in the UK and Its Member Nations: Analyses of Standalone and Linked National Databases.” *BMC Medicine* 14 (1): 113. https://doi.org/10.1186/s12916-016-0657-8.

Müller-Pebody, B., W. J. Edmunds, M. C. Zambon, N. J. Gay, and N. S. Crowcroft. 2002. “Contribution of RSV to Bronchiolitis and Pneumonia-Associated Hospitalizations in English Children, April 1995-March 1998.” *Epidemiology and Infection* 129 (1): 99–106.

Mullins, J. A., A. C. Lamonte, J. S. Bresee, and L. J. Anderson. 2003. “Substantial Variability in Community Respiratory Syncytial Virus Season Timing.” *The Pediatric Infectious Disease Journal* 22 (10): 857–62. https://doi.org/10.1097/01.inf.0000090921.21313.d3.

Murdoch, D. R., and S. R. C. Howie. 2018. “The Global Burden of Lower Respiratory Infections: Making Progress, but We Need to Do Better.” *The Lancet. Infectious Diseases* 18 (11): 1162–63. https://doi.org/10.1016/S1473-3099(18)30407-9.

Murray, C., M. A. Richards, J. N. Newton, K. A. Fenton, H. Ross Anderson, C. Atkinson, D. Bennett, et al. 2013. “UK Health Performance: Findings of the Global Burden of Disease Study 2010.” *The Lancet* 381 (9871): 997–1020. https://doi.org/10.1016/S0140-6736(13)60355-4.

Murray, E. L., M. Klein, L. Brondi, J. E. McGowan, C. van Mels, W. A. Brooks, D. Kleinbaum, D. Goswami, P. B. Ryan, and C. B. Bridges. 2012. “Rainfall, Household Crowding, and Acute Respiratory Infections in the Tropics.” *Epidemiology and Infection* 140 (1): 78–86. https://doi.org/10.1017/S0950268811000252.

Murray, J. C. 2013. “The Clinical Burden of Respiratory Syncytial Virus (RSV) Bronchiolitis among Infants in the United Kingdom (UK).” Ph.D., Imperial College London. http://hdl.handle.net/10044/1/12717.

Nabalamba, Alice, and Wayne J. Millar. 2007. “Going to the Doctor.” *Health Reports* 18 (1): 23–35.

Nair, H., J. Nokes, B. D. Gessner, M. Dherani, S. A. Madhi, R. J. Singleton, K. L. O’Brien, et al. 2010. “Global Burden of Acute Lower Respiratory Infections Due to Respiratory Syncytial Virus in Young Children: A Systematic Review and Meta-Analysis.” *Lancet* 375 (9725): 1545–55. https://doi.org/10.1016/S0140-6736(10)60206-1.

Nair, H., E. AF Simões, I. Rudan, B. D. Gessner, E. Azziz-Baumgartner, J. S. F. Zhang, D. R. Feikin, et al. 2013. "Global and Regional Burden of Hospital Admissions for Severe Acute Lower Respiratory Infections in Young Children in 2010: A Systematic Analysis." *The Lancet* 381 (9875): 1380–90. https://doi.org/10.1016/S0140-6736(12)61901-1.

NAO (National Audit Office). 2014. "Out-of-Hours GP Services in England." Department of Health and NHS England.

Napoli, C., F. Riccardo, S. Declich, M. G. Dente, M. G. Pompa, C. Rizzo, M. C. Rota, and A. Bella. 2014. "An Early Warning System Based on Syndromic Surveillance to Detect Potential Health Emergencies among Migrants: Results of a Two-Year Experience in Italy." *International Journal of Environmental Research and Public Health* 11 (8): 8529–41. https://doi.org/10.3390/ijerph110808529.

Nenna, R., M. Evangelisti, A. Frassanito, C. Scagnolari, A. Pierangeli, G. Antonelli, A. Nicolai, et al. 2017. "Respiratory Syncytial Virus Bronchiolitis, Weather Conditions and Air Pollution in an Italian Urban Area: An Observational Study." *Environmental Research* 158: 188–93. https://doi.org/10.1016/j.envres.2017.06.014.

Newitt, S., A. J. Elliot, R. Morbey, H. Durnall, M. E. Pietzsch, J. M. Medlock, S. Leach, and G. E. Smith. 2016. "The Use of Syndromic Surveillance to Monitor the Incidence of Arthropod Bites Requiring Healthcare in England, 2000–2013: A Retrospective Ecological Study." *Epidemiology & Infection* 144 (11): 2251–59. https://doi.org/10.1017/S0950268816000686.

Newton, J. N, A. Briggs, C. Murray, D. Dicker, K. Foreman, H. Wang, M. Naghavi, et al. 2015. "Changes in Health in England, with Analysis by English Regions and Areas of Deprivation, 1990–2013: A Systematic Analysis for the Global Burden of Disease Study 2013." *The Lancet* 386 (10010): 2257–74. https://doi.org/10.1016/S0140-6736(15)00195-6.

NHS (National Health Service). 2005. "Pandemic Flu: UK Health Departments' Influenza Pandemic Contingency Plan." London: Department of Health Publications.

NHS (National Health Service). 2017. "NHS 111." December 22, 2017. https://www.nhs.uk/using-the-nhs/nhs-services/urgent-and-emergency-care/nhs-111/.

NHS (National Health Service). 2021. "Using the NHS and Other Health Services during Coronavirus." April 15, 2021.

https://www.nhs.uk/conditions/coronavirus-covid-19/social-distancing/using-the-nhs-and-other-health-services/.

NHS Digital. 2018. “Read Codes.” 2018. https://digital.nhs.uk/services/terminology-and-classifications/read-codes#:~:text=Read%20Codes%20are%20a%20coded,3%20(CTV3%20or%20v3).

NHS Digital 2020. “NHS Pathways.” 2020. https://digital.nhs.uk/services/nhs-pathways.

NHS Digital 2021a. “SNOMED CT.” January 15, 2021. https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct.

NHS Digital. 2021b. “NHS 111 Online.” March 25, 2021. https://digital.nhs.uk/services/nhs-111-online.

Nilsson, P., and M. H. Laurell. 2005. “Impact of Socioeconomic Factors and Antibiotic Prescribing on Penicillin- Non-Susceptible Streptococcus Pneumoniae in the City of Malmö.” *Scandinavian Journal of Infectious Diseases* 37 (6–7): 436–441. https://doi.org/10.1080/00365540510037795.

Noveroske, D. B., J. L. Warren, V. E. Pitzer, and D. M. Weinberger. 2016. “Local Variations in the Timing of RSV Epidemics.” *BMC Infectious Diseases* 16: 674. https://doi.org/10.1186/s12879-016-2004-2.

Nsubuga, P., M. E. White, S. B. Thacker, M. A. Anderson, S. B. Blount, C. V. Broome, T. M. Chiller, et al. 2006. “Public Health Surveillance: A Tool for Targeting and Monitoring Interventions.” In *Disease Control Priorities in Developing Countries*, 2nd ed. Washington (DC): World Bank. http://www.ncbi.nlm.nih.gov/books/NBK11770/.

Nyoka, R., J. Omony, S. M. Mwalili, T. N. O. Achia, A. Gichangi, and H. Mwambi. 2017. “Effect of Climate on Incidence of Respiratory Syncytial Virus Infections in a Refugee Camp in Kenya: A Non-Gaussian Time-Series Analysis.” *PLOS ONE* 12 (6): e0178323. https://doi.org/10.1371/journal.pone.0178323.

O’Brien, S. J., G. Rait, P. R. Hunter, J. J. Gray, F. J. Bolton, D. S. Tompkins, J. McLauchlin, et al. 2010. “Methods for Determining Disease Burden and Calibrating National Surveillance Data in the United Kingdom: The Second Study of Infectious Intestinal Disease in the Community (IID2 Study).” *BMC Medical Research Methodology* 10: 39. https://doi.org/10.1186/1471-2288-10-39.

Obando-Pacheco, PP., A. J. Justicia-Grande, I. Rivero-Calle, C. Rodríguez-Tenreiro, P. Sly, O. Ramilo, A. Mejías, et al. 2018. "Respiratory Syncytial Virus Seasonality: A Global Overview." *The Journal of Infectious Diseases* 217 (9): 1356–64. https://doi.org/10.1093/infdis/jiy056.

Oberoi, S., N. Chaudhary, S. Patnaik, and A. Singh. 2016. "Understanding Health Seeking Behavior." *Journal of Family Medicine and Primary Care* 5 (2): 463–64. https://doi.org/10.4103/2249-4863.192376.

Ogra, P. L. 2004. "Respiratory Syncytial Virus: The Virus, the Disease and the Immune Response." *Paediatric Respiratory Reviews* 5: S119–26. https://doi.org/10.1016/S1526-0542(04)90023-1.

Okoromah, C. N., and O. Oviawe. 2002. "Is Childhood Asthma Underdiagnosed and Undertreated?" *The Nigerian Postgraduate Medical Journal* 9 (4): 221–25.

Olenja, J. 2003. "Health Seeking Behaviour in Context." *East African Medical Journal* 80 (2): 61–62. https://doi.org/10.4314/eamj.v80i2.8689.

Oliveira-Santos, M., J. A. Santos, J. Soares, A. Dias, and M. Quaresma. 2016. "Influence of Meteorological Conditions on RSV Infection in Portugal." *International Journal of Biometeorology* 60 (12): 1807–17. https://doi.org/10.1007/s00484-016-1168-1.

Olowokure, Babatunde, Lilian Clark, Alex J Elliot, Douglas Harding, and Ann Fleming. 2007. "Mumps and the Media: Changes in the Reporting of Mumps in Response to Newspaper Coverage." *Journal of Epidemiology and Community Health* 61 (5): 385–88. https://doi.org/10.1136/jech.2005.042598.

ONS (Office for National Statistics). 2014. "Estimates of the Population for the UK, England and Wales, Scotland and Northern Ireland: Mid-Year Estimates 2013-2018."

Oren, E., J. Frere, E. Yom-Tov, and E. Yom-Tov. 2018. "Respiratory Syncytial Virus Tracking Using Internet Search Engine Data." *BMC Public Health* 18 (1): 445. https://doi.org/10.1186/s12889-018-5367-z.

Papenburg, J., M. Hamelin, N. Ouhoummane, J. Carbonneau, M. Ouakki, F. Raymond, L. Robitaille, et al. 2012. "Comparison of Risk Factors for Human Metapneumovirus and Respiratory Syncytial Virus Disease Severity in Young Children." *The Journal of Infectious Diseases* 206 (2): 178–189. https://doi.org/10.1093/infdis/jis333.

Parrott, R. H., H. Wha Kim, J. O. Arrobio, D. S. Hodes, B. R. Murphy, C. D. Brandt, E. Camargo, And R. M. Chanock. 1973. "Epidemiology of Respiratory Syncytial Virus Infection in Washington, D.C.: Ii. Infection And Disease with Respect to Age, Immunologic Status, Race and Sex." *American Journal of Epidemiology* 98 (4): 289–300. https://doi.org/10.1093/oxfordjournals.aje.a121558.

Paterson, B. J., and D. N. Durrheim. 2013. "The Remarkable Adaptability of Syndromic Surveillance to Meet Public Health Needs." *Journal of Epidemiology and Global Health* 3 (1): 41–47. https://doi.org/10.1016/j.jegh.2012.12.005.

Payne, E. H., M. Gebregziabher, J. W. Hardin, V. Ramakrishnan, and L. E. Egede. 2018. "An Empirical Approach to Determine a Threshold for Assessing Overdispersion in Poisson and Negative Binomial Models for Count Data." *Communications in Statistics: Simulation and Computation* 47 (6): 1722–1738. https://doi.org/10.1080/03610918.2017.1323223.

Paynter, S. 2015. "Humidity and Respiratory Virus Transmission in Tropical and Temperate Settings." *Epidemiology & Infection* 143 (6): 1110–18. https://doi.org/10.1017/S0950268814002702.

Pebody, R. G., H. K. Green, N. Andrews, N. L. Boddington, H. Zhao, I. Yonova, J. Ellis, et al. 2015. "Uptake and Impact of Vaccinating School Age Children against Influenza during a Season with Circulation of Drifted Influenza A and B Strains, England, 2014/15." *Euro Surveillance: Bulletin Européen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin* 20 (39). https://doi.org/10.2807/1560-7917.ES.2015.20.39.30029.

Pendarvis, J., R. Miramontes, J. Schlegelmilch, A. Fleischauer, J. Gunn, L. Hutwagner, and A. Barry. 2007. "Analysis of Syndrome Definitions for Gastrointestinal Illness with ICD9 Codes for Gastroenteritis during the 2006-07 Norovirus Season in Boston." *Advances in Disease Surveillance* 4 (263).

Peppa, M., W. J. Edmunds, and S. Funk. 2017. "Disease Severity Determines Health-Seeking Behaviour amongst Individuals with Influenza-like Illness in an Internet-Based Cohort." *BMC Infectious Diseases* 17: 238. https://doi.org/10.1186/s12879-017-2337-5.

PHE (Public Health England). 2015. "Syndromic Surveillance: Systems and Analyses - GOV.UK." 2015. https://www.gov.uk/government/collections/syndromic-surveillance-systems-and-analyses.

PHE (Public Health England). 2018. “Local Action on Health Inequalities Understanding and Reducing Ethnic Inequalities in Health.” https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/730917/local_action_on_health_inequalities.pdf.

PHE (Public Health England). 2019. “Surveillance of Influenza and Other Respiratory Viruses in the UK Winter 2018 to 2019.” https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/839350/Surveillance_of_influenza_and_other_respiratory_viruses_in_the_UK_2018_to_2019-FINAL.pdf.

PHE (Public Health England). 2020. “Disparities in the Risk and Outcomes of COVID-19.” https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/908434/Disparities_in_the_risk_and_outcomes_of_COVID_August_2020_update.pdf.

PHE (Public Health England). 2021. “Guidance: Notifiable Diseases and Causative Organisms: How to Report.” 2021. https://www.gov.uk/guidance/notifiable-diseases-and-causative-organisms-how-to-report.

PHE, (Public Health England), 2021. Guidance: Respiratory syncytial virus (RSV): symptoms, transmission, prevention, treatment. URL https://www.gov.uk/government/publications/respiratory-syncytial-virus-rsv-symptoms-transmission-prevention-treatment/respiratory-syncytial-virus-rsv-symptoms-transmission-prevention-treatment#:~:text=In%20temperate%20climates%20such%20as,less%20predictable%20in%20its%20timing. (accessed 3.18.21).

Pica, N., and N. M Bouvier. 2012. “Environmental Factors Affecting the Transmission of Respiratory Viruses.” *Current Opinion in Virology* 2 (1): 90–95. https://doi.org/10.1016/j.coviro.2011.12.003.

Pivette, M. 2014a. “Surveillance of Gastrointestinal Disease in France Using Drug Sales Data.” *Epidemics* 8: 1–8. https://doi.org/10.1016/j.epidem.2014.05.001.

Pivette, M., J. E. Mueller, P. Crépey, and A. Bar-Hen. 2014b. “Drug Sales Data Analysis for Outbreak Detection of Infectious Diseases: A Systematic Literature Review.” *BMC Infectious Diseases* 14: 604. https://doi.org/10.1186/s12879-014-0604-2.

Pleasants, R. A, I. L Riley, and D. M Mannino. 2016. “Defining and Targeting Health Disparities in Chronic Obstructive Pulmonary Disease.” *International*

Journal of Chronic Obstructive Pulmonary Disease 11: 2475–96. https://doi.org/10.2147/COPD.S79077.

Polgreen, P. M., Y. Chen, D. M. Pennock, and F. D. Nelson. 2008. "Using Internet Searches for Influenza Surveillance." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 47 (11): 1443–48. https://doi.org/10.1086/593098.

Polozov, I. V., L. Bezrukov, K. Gawrisch, and J. Zimmerberg. 2008. "Progressive Ordering with Decreasing Temperature of the Phospholipids of Influenza Virus." *Nature Chemical Biology* 4 (4): 248–55. https://doi.org/10.1038/nchembio.77.

Poole, C. 2001. "Low P-Values or Narrow Confidence Intervals: Which Are More Durable?" *Epidemiology (Cambridge, Mass.)* 12 (3): 291–94. https://doi.org/10.1097/00001648-200105000-00005.

Poortaghi, S., A. Raiesifar, P. Bozorgzad, S. E. J. Golzari, S. Parvizy, and F. Rafii. 2015. "Evolutionary Concept Analysis of Health Seeking Behavior in Nursing: A Systematic Review." *BMC Health Services Research* 15 (11): 523. https://doi.org/10.1186/s12913-015-1181-9.

Portnov, B. A., J. Dubnov, and M. Barchana. 2007. "On Ecological Fallacy, Assessment Errors Stemming from Misguided Variable Selection, and the Effect of Aggregation on the Outcome of Epidemiological Study." *Journal of Exposure Science & Environmental Epidemiology* 17 (1): 106–21. https://doi.org/10.1038/sj.jes.7500533.

Price, R. H. Macgregor, C. Graham, and S. Ramalingam. 2019. "Association between Viral Seasonality and Meteorological Factors." *Scientific Reports* 9 (1): 929. https://doi.org/10.1038/s41598-018-37481-y.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. Vienna, Austria. https://www.R-project.org/.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R *Foundation for Statistical Computing*. Vienna, Austria: https://www.R-project.org.

R Studio Team. 2015. *RStudio: Integrated Development for R.* Boston, MA: RStudio, Inc. http://www.rstudio.com/.

Raghavan, D., and R. Jain. 2016. “Increasing Awareness of Sex Differences in Airway Diseases.” *Respirology (Carlton, Vic.)* 21 (3): 449–59. https://doi.org/10.1111/resp.12702.

Raisi-Estabragh, Z., C. McCracken, M. S. Bethell, J. Cooper, C. Cooper, M. J. Caulfield, P. B. Munroe, N. C. Harvey, and S. E. Petersen. 2020. “Greater Risk of Severe COVID-19 in Black, Asian and Minority Ethnic Populations Is Not Explained by Cardiometabolic, Socioeconomic or Behavioural Factors, or by 25(OH)-Vitamin D Status: Study of 1326 Cases from the UK Biobank.” *Journal of Public Health* 42 (3): 451–60. https://doi.org/10.1093/pubmed/fdaa095.

Razai, R. S., H. K. N. Kankam, A. Majeed, A. Esmail, and D. R. Williams. 2021. “Mitigating Ethnic Disparities in Covid-19 and Beyond.” *BMJ* 372 (1): m4921. https://doi.org/10.1136/bmj.m4921.

RCGPP (Royal College of General Practitioners). 1957. “RCGP Research and Surveillance Centre (RSC).” 1957. https://www.rcgp.org.uk/clinical-and-research/our-programmes/research-and-surveillance-centre.aspx.

Reeves, R. M., P. Hardelid, R. Gilbert, F. Warburton, J. Ellis, and R. G. Pebody. 2017. “Estimating the Burden of Respiratory Syncytial Virus (RSV) on Respiratory Hospital Admissions in Children Less than Five Years of Age in England, 2007-2012.” *Influenza and Other Respiratory Viruses* 11(2): 122–129. https://doi.org/10.1111/irv.12443.

Riaz, S. P., M. Horton, J. Kang, V. Mak, M. Lüchtenborg, and H. Møller. 2011. “Lung Cancer Incidence and Survival in England: An Analysis by Socioeconomic Deprivation and Urbanization.” *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer* 6 (12): 2005–10. https://doi.org/10.1097/JTO.0b013e31822b02db.

Riccardo, F., C. Napoli, A. Bella, C. Rizzo, M. C. Rota, M. G. Dente, S. De Santis, and S. Declich. 2011. “Syndromic Surveillance of Epidemic-Prone Diseases in Response to an Influx of Migrants from North Africa to Italy, May to October 2011.” *Eurosurveillance* 16 (46). http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=20016.

Rice, R. E. 2006. “Influences, Usage, and Outcomes of Internet Health Information Searching: Multivariate Results from the Pew Surveys.” *International Journal of Medical Informatics* 75 (1): 8–28. https://doi.org/10.1016/j.ijmedinf.2005.07.032.

Roberts, D. R., V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, et al. 2017. “Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure.” *Ecography* 40 (8): 913–29. https://doi.org/10.1111/ecog.02881.

Roberts, S. E., L. A. Button, J. M. Hopkin, M. J. Goldacre, R. A. Lyons, S. E. Rodgers, A. Akbari, and K. E. Lewis. 2012. “Influence of Social Deprivation and Air Pollutants on Serious Asthma.” *European Respiratory Journal* 40 (3): 785–88. https://doi.org/10.1183/09031936.00043311.

Robertson, S. E., A. J. Suleiman, F. R. Mehta, S. H. al-Dhahry, and M. S. el-Bualy. 1994. “Poliomyelitis in Oman: Acute Flaccid Paralysis Surveillance Leading to Early Detection and Rapid Response to a Type 3 Outbreak.” *Bulletin of the World Health Organization* 72 (6): 907–14.

Robertson, S. E., A. Roca, P. Alonso, E. A. F. Simoes, C. B. Kartasasmita, D. O. Olaleye, G. N. Odaibo, et al. 2004. “Respiratory Syncytial Virus Infection: Denominator-Based Studies in Indonesia, Mozambique, Nigeria and South Africa.” *Bulletin of the World Health Organization* 82 (12): 914–22. https://doi.org//S0042-96862004001200007.

Robins, J. M. 2001. “Data, Design, and Background Knowledge in Etiologic Inference.” *Epidemiology* 12 (3): 313–20. https://doi.org/10.1097/00001648-200105000-00011.

Robinson, D., E. Schulz, P. Brown, and C. Price. 1997. “Updating the Read Codes: User-Interactive Maintenance of a Dynamic Clinical Vocabulary.” *Journal of the American Medical Informatics Association: JAMIA* 4 (6): 465–72. https://doi.org/10.1136/jamia.1997.0040465.

Robinson, W. S. 1950. “Ecological Correlations and the Behaviour of Individuals.” *American Sociological Review* 15 (3): 351–57. https://doi.org/10.2307/2087176.

Rolland, E., K. M. Moore, V. A. Robinson, and D. McGuinness. 2006. “Using Ontario’s ‘Telehealth’ Health Telephone Helpline as an Early-Warning System: A Study Protocol.” *BMC Health Services Research* 6: 10. https://doi.org/10.1186/1472-6963-6-10.

Romaszko-Wojtowicz, A., I. Cymes, E. Dragańska, A. Doboszyńska, J. Romaszko, and K. Glińska-Lewczuk. 2020. “Relationship between Biometeorological Factors and the Number of Hospitalizations Due to Asthma.” *Scientific Reports* 10 (1): 9593. https://doi.org/10.1038/s41598-020-66746-8.

Roski, J., G. W. Bo-Linn, and T. A. Andrews. 2014. “Creating Value in Health Care Through Big Data: Opportunities and Policy Implications.” *Health Affairs* 33 (7): 1115–22. https://doi.org/10.1377/hlthaff.2014.0147.

Rue, H., S. Martino, and N. Chopin. 2009. “Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71 (2): 319–92. https://doi.org/10.1111/j.1467-9868.2008.00700.x.

Salisbury, C., M. Trivella, and S. Bruster. 2000. “Demand for and Supply of out of Hours Care from General Practitioners in England and Scotland: Observational Study Based on Routinely Collected Data.” *BMJ* 320 (7235): 618–621. https://doi.org/10.1136/bmj.320.7235.618.

Schanzer, D. L., J. M. Langley, and T W. S. Tam. 2006. “Hospitalization Attributable to Influenza and Other Viral Respiratory Illnesses in Canadian Children.” *The Pediatric Infectious Disease Journal* 25 (9): 795–800. https://doi.org/10.1097/01.inf.0000232632.86800.8c.

Schmidt, W, R. Pebody, and P Mangtani. 2010. “School Absence Data for Influenza Surveillance: A Pilot Study in the United Kingdom.” *Eurosurveillance* 15 (3). http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19467.

Sennerstam, R. B., and K. Moberg. 2004. “Relationship between Illness-Associated Absence in Day-Care Children and Weather Parameters.” *Public Health* 118 (5): 349–53. https://doi.org/10.1016/j.puhe.2003.10.011.

Sethi, D., J. Wheeler, L. C. Rodrigues, S. Fox, and P. Roderick. 1999. “Investigation of Under-Ascertainment in Epidemiological Studies Based in General Practice.” *International Journal of Epidemiology* 28 (1): 106–12. https://doi.org/10.1093/ije/28.1.106.

Severi, E., E. Heinsbroek, C. Watson, M. Catchpole, and Collective HPA Olympics Surveillance Work Group. 2012. “Infectious Disease Surveillance for the London 2012 Olympic and Paralympic Games.” *Eurosurveillance* 17 (31): 20232. https://doi.org/10.2807/ese.17.31.20232-en.

Shaman, J., and M. Kohn. 2009. “Absolute Humidity Modulates Influenza Survival, Transmission, and Seasonality.” *Proceedings of the National Academy of Sciences* 106 (9): 3243. https://doi.org/10.1073/pnas.0806852106.

Sheehan, W. J., and W. Phipatanakul. 2016. “Indoor Allergen Exposure and Asthma Outcomes.” *Current Opinion in Pediatrics* 28 (6): 772–77. https://doi.org/10.1097/MOP.0000000000000421.

Shi, T., D. A. McAllister, K. L. O’Brien, E. A. F. Simoes, S. A. Madhi, B. D. Gessner, F. P. Polack, et al. 2017. “Global, Regional, and National Disease Burden Estimates of Acute Lower Respiratory Infections Due to Respiratory Syncytial Virus in Young Children in 2015: A Systematic Review and Modelling Study.” *The Lancet* 390 (10098): 946–58. https://doi.org/10.1016/S0140-6736(17)30938-8.

Shiekh, Suhail Ismail, Harriet Forbes, Rohini Mathur, Liam Smeeth, Neil Pearce, and Charlotte Warren-Gash. 2020. “Ethnicity and Risk of Diagnosed Dementia after Stroke: A Cohort Study Using the Clinical Practice Research Datalink.” *Journal of Epidemiology and Community Health* 74 (2): 114–19. https://doi.org/10.1136/jech-2019-212825.

Simoes, E. AF. 1999. “Respiratory Syncytial Virus Infection.” *The Lancet* 354 (9181): 847–52. https://doi.org/10.1016/S0140-6736(99)80040-3.

Simoes, E. A. F. 2003. “Environmental and Demographic Risk Factors for Respiratory Syncytial Virus Lower Respiratory Tract Disease.” *The Journal of Pediatrics* 143 (5): S118-126. https://doi.org/10.1067/s0022-3476(03)00511-0.

Siontis, G. C. M., and J. P. A. Ioannidis. 2011. “Risk Factors and Interventions with Statistically Significant Tiny Effects.” *International Journal of Epidemiology* 40 (5): 1292–1307. https://doi.org/10.1093/ije/dyr099.

Sirimi, N., M. Miligkos, F. Koutouzi, E. Petridou, T. Siahanidou, and A. Michos. 2016. “Respiratory Syncytial Virus Activity and Climate Parameters during a 12-Year Period.” *Journal of Medical Virology* 88 (6): 931–37. https://doi.org/10.1002/jmv.24430.

Sloan, C., M. L. Moore, and T. Hartert. 2011. “Impact of Pollution, Climate, and Sociodemographic Factors on Spatiotemporal Dynamics of Seasonal Respiratory Viruses.” *Clinical and Translational Science* 4 (1): 48–54. https://doi.org/10.1111/j.1752-8062.2010.00257.x.

Sloane, P. D., J. K. MacFarquhar, E. Sickbert-Bennett, C. Madeline Mitchell, R. Akers, D. J. Weber, and K. Howard. 2006. “Syndromic Surveillance for Emerging Infections in Office Practice Using Billing Data.” *The Annals of Family Medicine* 4 (4): 351–58. https://doi.org/10.1370/afm.547.

Smieszek, T. 2009. “A Mechanistic Model of Infection: Why Duration and Intensity of Contacts Should Be Included in Models of Disease Spread.” *Theoretical Biology and Medical Modelling* 6 (1): 25. https://doi.org/10.1186/1742-4682-6-25.

Smith, G. 2018. “Step Away from Stepwise.” *Journal of Big Data* 5 (1): 32. https://doi.org/10.1186/s40537-018-0143-6.

Smith, G. E., J. Hippisley-Cox, S. Harcourt, M. Heaps, M. Painter, A. Porter, and M. Pringle. 2007. “Developing a National Primary Care-Based Early Warning System for Health Protection—a Surveillance Tool for the Future? Analysis of Routinely Collected Data.” *Journal of Public Health* 29 (1): 75–82. https://doi.org/10.1093/pubmed/fdl078.

Smith, G. E., Z. Bawa, Y. Macklin, R. A. Morbey, A. Dobney, S. Vardoulakis, and A. J. Elliot. 2015. “Using Real-Time Syndromic Surveillance Systems to Help Explore the Acute Impact of the Air Pollution Incident of March/April 2014 in England.” *Environmental Research* 136 (1): 500–504. https://doi.org/10.1016/j.envres.2014.09.028.

Smith, G. E., A. J. Elliot, S. Ibbotson, R. A. Morbey, O. Edeghere, J. Hawker, M. Catchpole, T. Endericks, P. Fisher, and B. McCloskey. 2016. “Novel Public Health Risk Assessment Process Developed to Support Syndromic Surveillance for the 2012 Olympic and Paralympic Games.” *Journal of Public Health* 39(3): e111-e117. https://doi.org/10.1093/pubmed/fdw054.

Smith, G. E., A. J. Elliot, I. Lake, O. Edeghere, R. A. Morbey, M. Catchpole, D. L. Heymann, et al. 2019. “Syndromic Surveillance: Two Decades Experience of Sustainable Systems – Its People Not Just Data!” *Epidemiology and Infection* 147 (1). https://doi.org/10.1017/S0950268819000074.

Smith, J., and A. Woodcock. 2006. “Cough and Its Importance in COPD.” *International Journal of Chronic Obstructive Pulmonary Disease* 1 (3): 305–14.

Smith, S., A. J. Elliot, S. Hajat, A. Bone, C. Bates, G. E. Smith, and S. Kovats. 2016a. “The Impact of Heatwaves on Community Morbidity and Healthcare Usage: A Retrospective Observational Study Using Real-Time Syndromic Surveillance.” *International Journal of Environmental Research and Public Health* 13 (1). https://doi.org/10.3390/ijerph13010132.

Smith, S., A. J. Elliot, S. Hajat, A. Bone, G. E. Smith, and S. Kovats. 2016b. “Estimating the Burden of Heat Illness in England during the 2013 Summer

Heatwave Using Syndromic Surveillance." *Journal of Epidemiology and Community Health* 70 (5): 459–65. https://doi.org/10.1136/jech-2015-206079.

Smith, S., R. A. Morbey, R. G. Pebody, T. C. Hughes, S. de Lusignan, F. A. Yeates, H. Thomas, S. J. O'Brien, G. E. Smith, and A. J. Elliot. 2017. "Retrospective Observational Study of Atypical Winter Respiratory Illness Season Using Real-Time Syndromic Surveillance, England, 2014-15." *Emerging Infectious Diseases* 23 (11): 1834–42. https://doi.org/10.3201/eid2311.161632.

Snell, N., D. Strachan, R. Hubbard, J. Gibson, K. Gruffydd-Jones, and I. Jarrold. 2016. "S32 Epidemiology of Chronic Obstructive Pulmonary Disease (COPD) in the Uk: Findings from the British Lung Foundation's 'Respiratory Health of the Nation' Project." *Thorax* 71 (3): A20–A20. https://doi.org/10.1136/thoraxjnl-2016-209333.38.

Sofianopoulou, E., S. P. Rushton, P. J. Diggle, and T. Pless-Mulloli. 2013. "Association between Respiratory Prescribing, Air Pollution and Deprivation, in Primary Health Care." *Journal of Public Health (Oxford, England)* 35 (4): 502–9. https://doi.org/10.1093/pubmed/fdt107.

Soneja, S., C. Jiang, J. Fisher, C. Romeo Upperman, C. Mitchell, and A. Sapkota. 2016. "Exposure to Extreme Heat and Precipitation Events Associated with Increased Risk of Hospitalization for Asthma in Maryland, U.S.A." *Environmental Health* 15 (1): 57. https://doi.org/10.1186/s12940-016-0142-z.

Soriano, J. B, A. Alemu Abajobir, K. Hassen Abate, S. Ferede Abera, A. Agrawal, M. Beshir Ahmed, A. Nidhal Aichour, et al. 2017. "Global, Regional, and National Deaths, Prevalence, Disability-Adjusted Life Years, and Years Lived with Disability for Chronic Obstructive Pulmonary Disease and Asthma, 1990–2015: A Systematic Analysis for the Global Burden of Disease Study 2015." *The Lancet Respiratory Medicine* 5 (9): 691–706. https://doi.org/10.1016/S2213-2600(17)30293-X.

Soucie, J. M. 2012. "Public Health Surveillance and Data Collection: General Principles and Impact on Hemophilia Care." *Hematology* 17 (1): S144–46. https://doi.org/10.1179/102453312X13336169156537.

Sparks, C. 2015. "An Examination of Disparities in Cancer Incidence in Texas Using Bayesian Random Coefficient Models." *PeerJ* 3: e1283. https://doi.org/10.7717/peerj.1283.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. "Bayesian Measures of Model Complexity and Fit." *Journal of the Royal*

Statistical Society: Series B (Statistical Methodology) 64 (4): 583–639. https://doi.org/10.1111/1467-9868.00353.

Stewart, J. A., R. Dundas, R. S. Howard, A. G. Rudd, and C. D. Wolfe. 1999. "Ethnic Differences in Incidence of Stroke: Prospective Study with Stroke Register." *BMJ (Clinical Research Ed.)* 318 (7189): 967–71. https://doi.org/10.1136/bmj.318.7189.967.

Stoto, M., M. Schonlau, and L. Mariano. 2004. "Syndromic Surveillance: Is It Worth the Effort?" *Chance* 17 (1): 19–24. http://dx.doi.org/10.1080/09332480.2004.10554882.

Sundell, N., L-M. Andersson, R. Brittain-Long, M. Lindh, and J. Westin. 2016. "A Four-Year Seasonal Survey of the Relationship between Outdoor Climate and Epidemiology of Viral Respiratory Tract Infections in a Temperate Climate." *Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology* 84: 59–63. https://doi.org/10.1016/j.jcv.2016.10.005.

Suzue, T., Y. Hoshikawa, S. Nishihara, A. Fujikawa, N. Miyatake, N. Sakano, T. Yoda, A. Yoshioka, and T. Hirao. 2012. "The New School Absentees Reporting System for Pandemic Influenza A/H1N1 2009 Infection in Japan." *PLoS ONE* 7 (2). https://doi.org/10.1371/journal.pone.0030639.

Tam, C. C, L. C Rodrigues, L. Viviani, J. P Dodds, M. R Evans, P. R Hunter, J. J Gray, et al. 2012. "Longitudinal Study of Infectious Intestinal Disease in the UK (IID2 Study): Incidence in the Community and Presenting to General Practice." *Gut* 61 (1): 69–77. https://doi.org/10.1136/gut.2011.238386.

Tamerius, J., M. I Nelson, S. Z Zhou, C. Viboud, M. A Miller, and W. J Alonso. 2011. "Global Influenza Seasonality: Reconciling Patterns across Temperate and Tropical Regions." *Environmental Health Perspectives* 119 (4): 439–45. https://doi.org/10.1289/ehp.1002383.

Taylor, S., R. J. Taylor, R. L. Lustig, C. Schuck-Paim, F. Haguinet, D. J. Webb, J. Logie, G. Matias, and D. M. Fleming. 2016. "Modelling Estimates of the Burden of Respiratory Syncytial Virus Infection in Children in the UK." *BMJ Open* 6 (6): e009337. https://doi.org/10.1136/bmjopen-2015-009337.

Textor, J., B. van der Zander, M. S. Gilthorpe, M. Liśkiewicz, and G. Ellison. 2016. "Robust Causal Inference Using Directed Acyclic Graphs: The R Package 'Dagitty.'" *International Journal of Epidemiology* 45 (6): 1887–94. https://doi.org/10.1093/ije/dyw341.

Thacker, S. B, D. F. Stroup, R. G. Parrish, and H. A. Anderson. 1996. "Surveillance in Environmental Public Health: Issues, Systems, and Sources." *American Journal of Public Health* 86 (5): 633–38.

The Health Foundation. 2020a. "Healthcare Usage during the COVID-19 Crisis."

The Health Foundation. 2020b. "How Is COVID-19 Changing the Use of Emergency Care?" 2020. https://www.health.org.uk/news-and-comment/charts-and-infographics/how-is-covid-19-changing-the-use-of-emergency-care.

The King's Fund. 2011. "Quality of Care in General Practice, Independent Inquiry Report, Chapter Two: The Evolving Role and Nature of General Practice in England." London: The King's Fund.

Thompson, A. E., Y. Anisimowicz, B. Miedema, W. Hogg, W. P. Wodchis, and K. Aubrey-Bassler. 2016. "The Influence of Gender and Other Patient Characteristics on Health Care-Seeking Behaviour: A QUALICOPC Study." *BMC Family Practice* 17. https://doi.org/10.1186/s12875-016-0440-0.

Todkill, D., Paul Loveridge, Alex James Elliot, Roger Morbey, Simon de Lusignan, Obaghe Edeghere, and Gillian Smith. 2017. "Socioeconomic and Geographical Variation in General Practitioner Consultations for Allergic Rhinitis in England, 2003–2014: An Observational Study." BMJ Open 7 (8): e017038. https://doi.org/10.1136/bmjopen-2017-017038.

Todkill, D., H. E. Hughes, A. J. Elliot, R. A. Morbey, O. Edeghere, S. Harcourt, T. Hughes, et al. 2016. "An Observational Study Using English Syndromic Surveillance Data Collected During the 2012 London Olympics – What Did Syndromic Surveillance Show and What Can We Learn for Future Mass-Gathering Events?" *Prehospital and Disaster Medicine*, January, 1–7. https://doi.org/10.1017/S1049023X16000923.

Todkill, D., P. Loveridge, A. J. Elliot, R. A. Morbey, S. de Lusignan, O. Edeghere, and G. E. Smith. 2017a. "Socioeconomic and Geographical Variation in General Practitioner Consultations for Allergic Rhinitis in England, 2003–2014: An Observational Study." *BMJ Open* 7 (8): e017038. https://doi.org/10.1136/bmjopen-2017-017038.

Todkill, D., P. Loveridge, A. J. Elliot, R. A. Morbey, O. Edeghere, T. Rayment-Bishop, C. Rayment-Bishop, J. E. Thornes, and G. E. Smith. 2017b. "Utility of Ambulance Data for Real-Time Syndromic Surveillance: A Pilot in the West Midlands Region, United Kingdom." *Prehospital and Disaster Medicine* 32 (6): 667–72. https://doi.org/10.1017/S1049023X17006690.

Travers, D., C. Barnett, A. Ising, and A. Waller. 2006. “Timeliness of Emergency Department Diagnoses for Syndromic Surveillance.” *AMIA Annual Symposium Proceedings* 2006: 769–73.

Triple-S Project. 2011. “Assessment of Syndromic Surveillance in Europe.” *The Lancet* 378 (9806): 1833–34. https://doi.org/10.1016/S0140-6736(11)60834-9.

Triple-S Project. 2013. “Guidelines for Designing and Implementing a Syndromic Surveillance System.” European Union: Triple-S Project.

Troeger, C., B. Blacker, I. A Khalil, P. C Rao, J. Cao, S. R M Zimsen, S. B Albertson, et al. 2018. “Estimates of the Global, Regional, and National Morbidity, Mortality, and Aetiologies of Lower Respiratory Infections in 195 Countries, 1990–2016: A Systematic Analysis for the Global Burden of Disease Study 2016.” *The Lancet Infectious Diseases* 18 (11): 1191–1210. https://doi.org/10.1016/S1473-3099(18)30310-4.

Trueman, D., F. Woodcock, and E. Hancock. 2017. *Estimating the Economic Burden of Respiratory Illness in the UK*. British Lung Foundation. https://www.probonoeconomics.com/sites/default/files/files/British%20Lung%20Foundation%20full%20report%2015032017_0.pdf.

Tsai, S., T. Hamby, A. Chu, J. A. Gleason, G. M. Goodrow, H. Gu, E. Lifshitz, and J. A. Fagliano. 2016. “Development and Application of Syndromic Surveillance for Severe Weather Events Following Hurricane Sandy.” *Disaster Medicine and Public Health Preparedness* 10 (3): 463–71. https://doi.org/10.1017/dmp.2016.74.

Tsui, F. C., M. M. Wagner, V. Dato, and C. C. Chang. 2001. “Value of ICD-9 Coded Chief Complaints for Detection of Epidemics.” *Proceedings. AMIA Symposium*, 711–15.

Turner, J., A. O’Cathain, E. Knowles, J. Nicholl, J. Tosh, F. Sampson, P. Coleman, and J. Coster. 2012. “Evaluation of NHS 111 Pilot Sites.” Final Report. University of Sheffield.

Turunen, R., A. Koistinen, T. Vuorinen, B. Arku, M. Söderlund-Venermo, O. Ruuskanen, and T. Jartti. 2014. “The First Wheezing Episode: Respiratory Virus Etiology, Atopic Characteristics, and Illness Severity.” *Pediatric Allergy and Immunology: Official Publication of the European Society of Pediatric Allergy and Immunology* 25 (8): 796–803. https://doi.org/10.1111/pai.12318.

UK Data Service. Office for National Statistics, National Records of Scotland, and Northern Ireland Statistics and Research Agency. 2017. “2011 Census Aggregate Data.” *UK Data Service*, February. https://doi.org/DOI: http://dx.doi.org/10.5257/census/aggregate-2011-2.

UK Government. 2021. “Order Coronavirus (COVID-19) Rapid Lateral Flow Tests.” 2021. https://www.gov.uk/order-coronavirus-rapid-lateral-flow-tests.

Unkel, S., C. P. Farrington, P. H. Garthwaite, C. Robertson, and N. Andrews. 2012. “Statistical Methods for the Prospective Detection of Infectious Disease Outbreaks: A Review.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 175 (1): 49–82. https://doi.org/10.1111/j.1467-985X.2011.00714.x.

Upperman, C. R., J. D. Parker, L. J. Akinbami, C. Jiang, X. He, R. Murtugudde, F. C. Curriero, L. Ziska, and A. Sapkota. 2017. “Exposure to Extreme Heat Events Is Associated with Increased Hay Fever Prevalence among Nationally Representative Sample of US Adults: 1997–2013.” *The Journal of Allergy and Clinical Immunology. In Practice* 5 (2): 435-441.e2. https://doi.org/10.1016/j.jaip.2016.09.016.

van Rossum, G, and F. Drake. 2009. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

van Woensel, J. B. M., W. M. C. van Aalderen, and J. L. L. Kimpen. 2003. “Viral Lower Respiratory Tract Infection in Infants and Young Children.” *BMJ (Clinical Research Ed.)* 327 (7405): 36–40. https://doi.org/10.1136/bmj.327.7405.36.

Vandentorren, S., A. Paty, E. Baffert, P. Chansard, and C. Caserio-Schönemann. 2016. “Syndromic Surveillance during the Paris Terrorist Attacks.” *The Lancet* 387 (10021): 846–47. https://doi.org/10.1016/S0140-6736(16)00507-9.

Vega, T., J. E. Lozano, T. Meerhoff, R. Snacken, J. Beauté, P. Jorgensen, R. Ortiz de Lejarazu, et al. 2015. “Influenza Surveillance in Europe: Comparing Intensity Levels Calculated Using the Moving Epidemic Method.” *Influenza and Other Respiratory Viruses* 9 (5): 234–46. https://doi.org/10.1111/irv.12330.

Velsko, S., and T. Bates. 2016. “A Conceptual Architecture for National Biosurveillance: Moving Beyond Situational Awareness to Enable Digital Detection of Emerging Threats.” *Health Security* 14 (3): 189–201. https://doi.org/10.1089/hs.2015.0063.

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. 4th ed. Statistics and Computing. Venables,W.N.:Statistics w.S-PLUS. New York: Springer-Verlag. https://www.springer.com/gp/book/9780387954578.

Venkatarao, Epari, Rajan R Patil, Deepa Prasad, Anita Anasuya, and Reuben Samuel. 2012. "Monitoring Data Quality in Syndromic Surveillance: Learnings from a Resource Limited Setting." *Journal of Global Infectious Diseases* 4 (2): 120–27. https://doi.org/10.4103/0974-777X.96778.

Vilain, P., F. Pagès, X. Combes, P. M. Dit Cassou, KK.atia Mougin-Damour, Y. Jacques-Antoine, and L. Filleul. 2015. "Health Impact Assessment of Cyclone Bejisa in Reunion Island (France) Using Syndromic Surveillance." *Prehospital and Disaster Medicine* 30 (2): 137–44. https://doi.org/10.1017/S1049023X15000163.

Vink, N. M., D. S. Postma, J. P. Schouten, J. G. M. Rosmalen, and H. Marike Boezen. 2010. "Gender Differences in Asthma Development and Remission during Transition through Puberty: The TRacking Adolescents' Individual Lives Survey (TRAILS) Study." *Journal of Allergy and Clinical Immunology* 126 (3): 498-504.e6. https://doi.org/10.1016/j.jaci.2010.06.018.

Vos, L. M., A. C. Teirlinck, J. E. Lozano, T. Vega, G. A. Donker, A. Im Hoepelman, L. J. Bont, J. Jelrik Oosterheert, and A. Meijer. 2019. "Use of the Moving Epidemic Method (MEM) to Assess National Surveillance Data for Respiratory Syncytial Virus (RSV) in the Netherlands, 2005 to 2017." *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin* 24 (20). https://doi.org/10.2807/1560-7917.ES.2019.24.20.1800469.

Wagner, M. M., A. W. Moore, and R. M. Aryel. 2011. *Handbook of Biosurveillance*. Elsevier.

Waldron, I. 1983. "Sex Differences in Human Mortality: The Role of Genetic Factors." *Social Science & Medicine* 17 (6): 321–33. https://doi.org/10.1016/0277-9536(83)90234-4.

Wang, L., M. F. Ramoni, K. D. Mandl, and P. Sebastiani. 2005. "Factors Affecting Automated Syndromic Surveillance." *Artificial Intelligence in Medicine* 34 (3): 269–78. https://doi.org/10.1016/j.artmed.2004.11.002.

Wang, Y. J., X. P. Chen, W. J. Chen, Z. L. Zhang, Y. P. Zhou, and Z. Jia. 2020. "Ethnicity and Health Inequalities: An Empirical Study Based on the 2010 China Survey of Social Change (CSSC) in Western China." *BMC Public Health* 20 (1): 637. https://doi.org/10.1186/s12889-020-08579-8.

Welliver, R. C. Sr. 2007. “Temperature, Humidity, and Ultraviolet B Radiation Predict Community Respiratory Syncytial Virus Activity.” *Journal Seasonal and Geographic Variation in Respiratory Syncytial Virus Outbreaks across the United States* 26 (11). https://doi.org/10.1097/INF.0b013e318157da59.

Wheeler, J. G., D. Sethi, J. M. Cowden, P. G. Wall, L. C. Rodrigues, D. S. Tompkins, M. J. Hudson, and P. J. Roderick. 1999. “Study of Infectious Intestinal Disease in England: Rates in the Community, Presenting to General Practice, and Reported to National Surveillance. The Infectious Intestinal Disease Study Executive.” *BMJ (Clinical Research Ed.)* 318 (7190): 1046–50.

WHO (World Health Organisation), and UNICEF (United Nations International Children’s Fund). 2013. “Ending Preventable Child Deaths from Pneumonia and Diarrhoea by 2025: The Integrated Global Action Plan for Pneumonia and Diarrhoea (GAPPD).” WHO/UNICEF. https://www.who.int/maternal_child_adolescent/documents/global_action_plan_pneumonia_diarrhoea/en/.

WHO (World Health Organisation). 2014. “WHO | Global Status Report on Noncommunicable Diseases 2014.” WHO. http://www.who.int/nmh/publications/ncd-status-report-2014/en/.

Wijngaard, C. C. van den, W. van Pelt, N. J. Nagelkerke, M. Kretzschmar, and M. P. Koopmans. 2011. “Evaluation of Syndromic Surveillance in the Netherlands: Its Added Value and Recommendations for Implementation.” *Euro Surveillance: Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin* 16 (9).

Willem, L., K. Van Kerckhove, D. L. Chao, N. Hens, and P. Beutels. 2012. “A Nice Day for an Infection? Weather Conditions and Social Contact Patterns Relevant to Influenza Transmission.” *PLOS ONE* 7 (11): e48695. https://doi.org/10.1371/journal.pone.0048695.

Wilson, K., and J. S. Brownstein. 2009. “Early Detection of Disease Outbreaks Using the Internet.” *CMAJ : Canadian Medical Association Journal* 180 (8): 829–31. https://doi.org/10.1503/cmaj.090215.

Xu, Z., R. A. Etzel, H. Su, C. Huang, Y. Guo, and S. Tong. 2012. “Impact of Ambient Temperature on Children’s Health: A Systematic Review.” *Environmental Research* 117 (1): 120–31. https://doi.org/10.1016/j.envres.2012.07.002.

Xu, Z., W. Hu, and S. Tong. 2014. “Temperature Variability and Childhood Pneumonia: An Ecological Study.” *Environmental Health: A Global Access Science Source* 13 (1): 51. https://doi.org/10.1186/1476-069X-13-51.

Yang, L., S. S. Chiu, K-P. Chan, K-H. Chan, W. Hing-Sang Wong, J. S. Malik Peiris, and C-M. Wong. 2011. “Validation of Statistical Models for Estimating Hospitalization Associated with Influenza and Other Respiratory Viruses.” *PLOS ONE* 6 (3): 1–8. https://doi.org/10.1371/journal.pone.0017882.

Yih, W. K., S. Deshpande, C. Fuller, D. Heisey-Grove, J. Hsu, B. A. Kruskal, M. Kulldorff, et al. 2010. “Evaluating Real-Time Syndromic Surveillance Signals from Ambulatory Care Data in Four States.” *Public Health Reports* 125 (1): 111–20.

Yin, J. K., A. E. Heywood, M. Georgousakis, C. King, C. Chiu, D. Isaacs, and K. K. Macartney. 2017. “Systematic Review and Meta-Analysis of Indirect Protection Afforded by Vaccinating Children Against Seasonal Influenza: Implications for Policy.” *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 65 (5): 719–28. https://doi.org/10.1093/cid/cix420.

YouGov. 2020. “Most Workers Want to Work from Home after COVID-19.” September 22, 2020. https://yougov.co.uk/topics/economy/articles-reports/2020/09/22/most-workers-want-work-home-after-covid-19.

Yusuf, S., G. Piedimonte, A. Auais, G. Demmler, S. Krishnan, P. Van Caeseele, R. Singleton, et al. 2007. “The Relationship of Meteorological Conditions to the Epidemic Activity of Respiratory Syncytial Virus.” *Epidemiology and Infection* 135 (7): 1077–90. https://doi.org/10.1017/S095026880600776X.

Zaman, M., and P. Mangtani. 2007. “Changing Disease Patterns in South Asians in the UK.” *Journal of the Royal Society of Medicine* 100 (6): 254–55.

Zeng, X., and M. Wagner. 2002. “Modeling the Effects of Epidemics on Routinely Collected Data.” *Journal of the American Medical Informatics Association: JAMIA* 9 (6): s17–22. https://doi.org/10.1197/jamia.M1219.

Zhang, N., W. Chen, P-T. Chan, H-L. Yen, J. Wei-Tze Tang, and Y. Li. 2020. “Close Contact Behavior in Indoor Environment and Transmission of Respiratory Infection.” *Indoor Air* 30 (4): 645–61. https://doi.org/10.1111/ina.12673.

Zhou, H., W. W. Thompson, C. G. Viboud, C. M. Ringholz, P-Y. Cheng, C. Steiner, G. R. Abedi, L. J. Anderson, L. Brammer, and D. K. Shay. 2012. "Hospitalizations Associated with Influenza and Respiratory Syncytial Virus in the United States, 1993–2008." *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 54 (10): 1427–1436. https://doi.org/10.1093/cid/cis211.

Ziemann, A., and T. Krafft. 2013. "Guidelines for Assessment of Data Sources." Triple S: Syndromic Surveillance Systems in Europe. http://www.syndromicsurveillance.eu/triple-s_guidelines_datasources.pdf.

Ziemann, A., N. Rosenkötter, L. Garcia-Castrillo Riesgo, M. Fischer, A. Krämer, F. K. Lippert, G. Vergeiner, H. Brand, and T. Krafft. 2015. "Meeting the International Health Regulations (2005) Surveillance Core Capacity Requirements at the Subnational Level in Europe: The Added Value of Syndromic Surveillance." *BMC Public Health* 15 (2). https://doi.org/10.1186/s12889-015-1421-2.

Ziemann, A., A. Fouillet, H. Brand, and T. Krafft. 2016. "Success Factors of European Syndromic Surveillance Systems: A Worked Example of Applying Qualitative Comparative Analysis." *PLoS ONE* 11 (5). https://doi.org/10.1371/journal.pone.0155535.

Ziska, L. H., L. Makra, S. K. Harry, N. Bruffaerts, M. Hendrickx, F. Coates, A. Saarto, et al. 2019. "Temperature-Related Changes in Airborne Allergenic Pollen Abundance and Seasonality across the Northern Hemisphere: A Retrospective Data Analysis." *The Lancet Planetary Health* 3 (3): e124–31. https://doi.org/10.1016/S2542-5196(19)30015-4.