The use of linked electronic health data to investigate the burden and outcomes of community-acquired pneumonia among older individuals in the United Kingdom.

ELIZABETH RACHEL CLARE MILLETT

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy

University of London

OCTOBER 2017

Department of Infectious Disease Epidemiology

Faculty of Epidemiology and Population Health

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Declaration

I, Elizabeth Rachel Clare Millett, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Elizabeth Millett

October 2017

Abstract

The aim of this thesis was to use large linked electronic health datasets from primary and secondary care to better estimate the burden of community-acquired pneumonia (CAP) in older adults in the UK, and to identify the determinants of severe outcomes of these common infections. Hospitalisation for CAP is increasingly common in this older age group, and these patients remain at an elevated mortality risk for over a year after hospital discharge, making this an important area of study.

CAP incidence was estimated at 7.99 episodes/1000 person-years (IQR:7.92-8.07/1000); rates were higher in men than women and rose strikingly with age. CAP incidence generally increased between 1997 and 2011, but this growth was attenuated when the rates were age-standardised.

To separate trends in incidence from trends in treatment location, CAP episodes admitted to hospital within 28 days of diagnosis were compared to CAP episodes that were not hospitalised. A wide range of factors potentially associated with hospital admission were investigated, and 14 co-morbidities, five frailty factors, and four medications/vaccinations were identified. Despite adjusting for these factors, the average predicted probability of hospitalisation after CAP rose from 57% (1998-2000) to 86% (2009-2010), while duration of hospitalisation and 28-day mortality decreased.

Finally, prognostic models were developed with the aim of assisting GPs in identifying CAP patients with an unexpectedly high mortality risk in the year after hospital discharge. Among 17 factors identified, increasing age, dementia, congestive heart failure, low weight, residential care and leukaemia/lymphoma were the strongest positive predictors of mortality, while being female, an ex-smoker and pneumococcal vaccination received more than a year ago had the strongest negative effects. The model showed a reasonable ability to distinguish between patients who died and survived.

The linked data used in this study allowed greater capture of incident CAP episodes and thus better estimates of the burden of disease. They also provided enriched patient medical histories, enabling detailed examination of the determinants of hospitalisation or death after CAP. The results presented will be of use to both clinicians and health

planners as the UK's population ages, and burden of CAP increases. Further work is needed to fully understand the increasing hospitalisation trend seen, and to externally validate the prognostic models developed.

Acknowledgements

First and foremost I would like to thank my supervisor, Sara Thomas. I am immensely grateful that you encouraged me down this path, and for all your advice and support. I have learnt more from you than anyone else in my professional life, and it has been a joy being supervised by you. I shall miss our lively debates as well as many other aspects of our working relationship. Thank you.

I am also very grateful to the members of my advisory panel. The input (especially regarding clinical matters) of Jenni Quint and Liam Smeeth has been invaluable. Statistical advice was provided in a clear and wonderfully cheerful manner by Rhian Daniels, and later Bianca De Stavola. Thank you all for answering my questions clearly, patiently and with good humour.

Additional assistance was kindly provided for the prognostic modelling chapter by Adrian Root and David Prieto. Thank you both for your time and valuable input.

Many friends have supported me while I worked on this thesis; Sarah, Mary, Ruth, Helen, Charlotte, Clara, Caroline, Jenni, Lisa, Emily, Rohini, Cinda, Laura, Angela and my many office-mates in 113, thank you all. A special thank you to Katie, Rahul and Kirsty who provided support and humour during the difficult final year of this PhD.

I would like to express my deepest gratitude to my family. My Dad and brother have cheered me on or up whenever needed, and provided endless love and support. Thank you so very much for everything, it would have been much more difficult without you.

Finally, this thesis is dedicated to my Mum, an amazing role-model who was with me in spirit every step of the way.

Table of contents

List of Tables

List of Figures

Abbreviations

A&E	Accident and Emergency
ACSC	Ambulatory care sensitive conditions
BMI	Body mass index
BNF	British National Formulary
BTS	British Thoracic Society
CAP	Community-acquired pneumonia
CFR	Case fatality rate
COPD	Chronic obstructive pulmonary disease
CPRD	Clinical Practise Research Datalink
ERS	European Respiratory Society
ES	Enhanced Service
GP	General Practitioner
HAP	Hospital-acquired pneumonia
HCAP	Healthcare-associated pneumonia
HES	Hospital Episode Statistics
HR	Hazard ratio
ICD	International Classification of Disease
ICPC	International Primary Care Classification
ICS	Inhaled corticosteroids
ICU	Intensive care unit
IMD	Index of Multiple Deprivation
LSOA	Lower Super Output Area
LRTI	Lower respiratory tract infection
MeSH	Medical Subject Headings
MDR	Multidrug resistant
MI	Myocardial infarction
NHS	National Health Service
ONS	Office for National Statistics

OOH	Out of hours
OR	Odds ratio
PbR	Payment by results
PPV	Pneumococcal polysaccharide vaccine
QOF	Quality Outcomes Framework
RCGP	Royal College of General Practitioners
RR	Rate ratio
SE	Standard error
SES	Socioeconomic status
SHA	Strategic Health Authority
UK	United Kingdom
UTS	Up to standard (date)
VE	Vaccine effectiveness

Chapter 1   Background

This thesis uses linked electronic health records from primary and secondary care to obtain better estimates of the burden of community-acquired pneumonia (CAP) and lower respiratory tract infections (LRTI) among the UK's older adults, and to examine factors associated with serious outcomes after CAP. In the process of exploring these themes, methodological aspects of the use of routinely collected data are also considered.

This background chapter provides an overview of pneumonia; its characteristics, treatment and outcomes, and why, among the older population, infections such as this are increasingly important. I then summarise the objectives of this thesis, and provide an outline of each of the chapters.

1.1   Pneumonia

Pneumonia is a severe infection of the lower respiratory tract which particularly affects the very young (aged <5 years), older adults (≥65 years), those with underlying health conditions and the immunosuppressed. Morbidity and mortality due to pneumonia are high; between 22% and 44% of patients with CAP require hospitalisation,[1] and pneumonia is the leading cause of death due to infection in both Europe and the US.[2] The defining feature of pneumonia is inflammatory exudate of the lung parenchyma, which results in patients typically presenting with symptoms of cough, fever, breathlessness or rapid breathing and chest or pleuritic pain.[1]

1.1.1   The microbial ecology of the lung

Until recently it was thought that the lungs were a sterile environment,[3] and that pneumonia occurred when this sterile environment was compromised. However, an increasingly accepted theory is that the lungs have their own microbiome. The respiratory tract is an oxygen-rich environment, with a temperature that varies from that of the air outside (at the mouth) to the core temperature of the body at the alveoli, making the lung itself a warm, moist environment where, if unchecked, bacteria can thrive.[4] Microbes move into the lung via several sources; inhaled air, which contains 10^4-10^6 bacteria/mm^3, micro-aspiration of the contents of the upper respiratory tract, and dissemination of microbes along the mucosa. The microbiome of the lung is more

similar to that of the oropharynx than that of inhaled air, the nasopharynx or the lower GI tract. This is due to microaspiration of pharyngeal secretions, much of which may occur during sleep when people are lying down and the reflexes of the larynx and cough (which usually occurs once every two hours awake) are suppressed.[5] In healthy lungs, the most common phyla of bacteria are Bacteroidetes and Firmicutes, and genera are *Prevotella*, *Veillonella* and *Streptococcus*.[5] The composition of the microbiome is thought to be relatively uniform within a healthy respiratory tract, although its diversity decreases with increasing distance from the oropharyrnx. Compared to the digestive system, the bacterial density in the respiratory system is quite low.[4]

1.1.2 Defence mechanisms of the respiratory system

The configuration of the lung microbiome depends not just on the migration of microbes into the airways, but also on their rate of reproduction within the lung and their removal by the defences of the respiratory system. Prevention of microbial entry starts in the upper airways where the nasal passages are lined with hairs to trap large particles. Smaller particles that pass through these hairs and settle on the mucosa are cleared by columnar ciliated epithelium towards the oropharynx. The larynx plays an important role in protecting the airway as the epiglottis, a cartilaginous flap, closes over the trachea during swallowing and vomiting and so prevents aspiration. Microbes that pass through the larynx and trachea reach the bronchi and then the bronchioles. The bronchi contain submucosal mucus-secreting glands and are lined by an epithelium with cilia and goblet cells. The bronchioles also contain a single layer of ciliated cells, but have very few goblet cells.[6] The mucus-secreting glands and goblet cells both play an important role in defence, as they secrete globules of mucus which are fairly impermeable to water. The globules form an almost complete covering of mucus which sits on a layer of liquid found around the cilia of the epithelial cells. The tips of the cilia engage with the bottom of the mucus layer and the cilia move in a coordinated manner to push the mucus upwards toward the pharynx, leading to its expulsion from the airways by either coughing or swallowing. This is called the mucociliary escalator. Clearance of mucus from the large bronchi is relatively fast, taking only 30-60 minutes, however it can take several days for mucus to be cleared from the lower bronchioles.[6]

In addition to mucin, the mucus also contains other secretions from the cells in the airways, such as antimicrobial molecules (defensins, lysozymes), specific antibodies (IgA) and cytokines. These aid the immune system in identifying and killing the bacteria. Microbes and other particles thus get trapped in the layer of mucus, and are either inactivated or expelled from the airways before they cause any harm.[6]

Pathogens which manage to breach the physical barriers of the lungs are met by a coordinated defence provided by several cell types. Macrophages modify and control acute and inflammatory responses by secreting chemokines and other cytokines which promote or suppress the immune response. They can also work with dendritic cells to phagocytose infective material. Epithelial cells assist macrophages in producing chemokines and cytokines which leads to neutrophil accumulation and local inflammation. Cytokines perform many roles, including preventing microbes from adhering to the epithelial surface, inhibiting viral infections by disrupting their assembly processes, and triggering pro-inflammatory meditators which increase vascular permeability, bronchoconstriction and inflammatory cell infiltration.[7]

Neutrophil numbers in the airway are low, however a pool of these cells exists in the pulmonary circulation, and they can be recruited quickly to any site of infection. Neutrophils can travel out of pulmonary capillaries into the air spaces, where they phagocytose and kill microbes. They also produce mediators which activate B cells and dendritic cells. The latter facilitate antigen presentation to T and B lymphocytes, resulting in a more specific response to the pathogen and development of immunological memory.[7]

The mechanical and immunological defence systems of healthy lungs result in an environment that is not conducive to extensive bacterial growth, and thus levels of bacterial reproduction are relatively low. In healthy lungs the composition of the lung microbiome is therefore mainly determined by the balance of immigration and removal.[4] The pathogens known to cause pneumonia (further discussed in the section below) have been found in the lungs of healthy, pneumonia-free subjects.[5] This suggests the disease occurs due to the preferential reproduction of a single type of pathogen above all others in the microbiome. While research into this area is ongoing, there is in vitro evidence that features of the host inflammatory defences promote the

growth of certain potential pathogens via a positive feedback loop, leading to their domination of the lung microbiota.[5]

1.1.3 Classification and aetiology of pneumonia

Pneumonia can be classified into two main groups that reflect the setting in which infection was thought to be acquired: community-acquired or hospital-acquired pneumonia (HAP). This grouping is significant as CAP and HAP have different risk factors, mortality rates and treatment regimens, due in part to differing aetiologies.

CAP is caused by a variety of pathogens with bacteria the predominant cause; a recent review of 25 studies of the aetiology of CAP in European adults found that the most commonly isolated causative agent was *S. pneumoniae* (identified in 12-85% of patients in 19 studies).[8] Six of the studies stratified by age, and from these *S. pneumoniae*, *H. influenzae* and respiratory viruses were more frequently isolated among those aged ≥65 years, while *M. pneumoniae* was more frequently isolated from those aged <65.[8] The bacterial causes of CAP are typically sensitive to first-line antibiotics.[1] In practice, microbiological confirmation of the infectious aetiology of CAP is often not performed. In the primary care setting, microbiological investigations are not currently routinely recommended for patients with CAP, as knowledge of the causative organism typically does not alter the treatment plan.[1, 9] While this testing does occur in hospital, it can be difficult to identify the causative pathogen of CAP when antibiotic treatment has been initiated before sample acquisition. Some patients struggle to provide a sputum sample for culture, this is a particular problem among older adults. As a result, the majority of CAP cases have no information on aetiology recorded in their primary or secondary care records.

In contrast to CAP, HAP is more likely to be due to *Staphylococcus aureus* (including methicillin-resistant strains), or other bacteria which require broad-spectrum antibiotic treatment.[10] Consequently, CAP and HAP are considered separately to one another in both clinical and research settings. This thesis focuses on CAP, due to its higher incidence and the potential for it to affect a larger population than HAP. Episodes of HAP are excluded from all analyses, due to their differing aetiology, treatment and prognosis.

In 2005, the US introduced an additional category, health care-associated pneumonia (HCAP), which is used to describe pneumonia in patients who receive home wound care or infusion therapy, chronic dialysis, reside in a nursing home or care facility or have been hospitalised for ≥2 days in the past three months.[11] Early research in the US suggested that HCAP patients were more similar to HAP than CAP patients, with comparatively high levels of multidrug resistant (MDR) pathogens and mortality, leading to the recommendation of broad-spectrum antibiotic treatment for this group.[11, 12] However, the concept of HCAP remains controversial; a recent systematic review showed that across multiple international studies, HCAP classification was poorly predictive of MDR pathogens. Furthermore, HCAP and CAP mortality rates were comparable when age and co-morbidity were taken into account.[13] The inclusion or exclusion of HCAP as a separate entity is particularly important among older populations, who have a higher risk of developing pneumonia and who comprise 83% of those in residential care.[14] At present, evidence from Europe does not suggest that HCAP is microbiologically distinct from CAP, or that separate treatment guidelines or categorisation is warranted, and as such the term has not been widely adopted in European treatment guidelines.[13, 15, 16] Throughout this work, I use the current UK and European definition of CAP which includes those in residential or nursing facilities, and those who receive home care.

1.1.4   Diagnosis of CAP

Patients who have CAP can be diagnosed in primary care or in hospital, and the diagnostic criteria for CAP differ in these settings.   The gold standard for diagnosing pneumonia is a chest radiograph, which typically shows new shadowing which cannot be otherwise explained (caused by the inflammatory exudate).

In contrast, in a community setting (such as in UK primary care) it is not common practice for a chest radiograph to be performed to diagnose pneumonia. In the absence of a chest radiograph, the British Thoracic Society (BTS) guidelines define 'suspected CAP' as including the following four features:

> "Symptoms of an acute lower respiratory tract illness (cough and at least one other lower respiratory tract symptom).
>
> New focal chest signs on examination.

At least one systemic feature (either a symptom complex of sweating,

fevers, shivers, aches and pains and/or temperature ≥38°C).

No other explanation for the illness."[1]

Diagnosis of 'definite CAP' uses the same criteria as suspected CAP, with the addition of a chest radiograph showing lung shadowing that is thought to be new.

The combination of non-specific signs and symptoms used to diagnose pneumonia clinically can make it difficult to differentiate from other illnesses. For example, focal chest signs such as 'crackles' can also be present due to other conditions which cause lung stiffening.[17] More generally, there are a range of diseases other than pneumonia which can present with breathlessness, fatigue and/or cough, for instance less severe lower respiratory tract infections (LRTI, described in section 1.1.5), chronic obstructive pulmonary disease (COPD) or heart failure. This results in the diagnosis of pneumonia being subject to a degree of diagnostic inaccuracy.

Of particular relevance to the work in this thesis is that the difficulty in diagnosing pneumonia is magnified among older adults. Older patients have been shown to present with fewer symptoms than younger patients,[18] and their symptoms are more commonly non-specific, such as confusion, tiredness or loss of appetite without chest signs.[18, 19] Additionally, older patients have more complex underlying health issues, and the BTS highlights that CAP is particularly difficult to diagnose among older patients with COPD, other respiratory disease or left ventricular failure. To reflect the potentially challenging nature of diagnosing CAP in the older population, the BTS diagnostic criteria for this age group are slightly less specific:

"the presence of chest radiograph shadowing accompanied by acute

clinical illness (unspecified) without other obvious cause."[1]

The accuracy of a clinical CAP diagnosis made in general practice has been assessed in a number of countries. For example, a large multi-country European study (including England and Wales) was carried out between 2007-2010, comprising 2810 adult patients (mean age 50 years ± 17 years) presenting to primary care with acute cough; the GP diagnoses for these patients were compared with the results of subsequent chest radiographs, carried out in the week following diagnosis.[20] Overall, 5% of patients had radiological evidence of pneumonia. Despite the potential for misdiagnosis of other

conditions as CAP (as highlighted above), the study found that GP diagnoses of pneumonia had a high specificity (99%). However, clinical diagnoses had low sensitivity (29%, i.e. the GPs were only identifying 29% of radiologically-confirmed pneumonias). Patients who had radiological pneumonia without a GP clinical diagnosis had less severe symptoms and signs compared to those who were diagnosed clinically.

In hospital settings where chest radiographs are readily available, patients are also diagnosed as having CAP when a chest radiograph does not confirm the diagnosis. For example, a Canadian study comparing A&E discharge diagnoses of pneumonia to radiology reports found that of 671 A&E clinical pneumonia diagnoses, 45% of cases were confirmed by the radiology reports, while 43% were classified as 'normal' or 'non-pneumonia'.[21] In a second Canadian study, of 2706 patients who were thought on clinical examination to have pneumonia, 911 (34%) were not confirmed with chest radiographs.[22] Other studies have also reported low to moderate agreement of pneumonia diagnoses between hospital clinicians and radiologists.[23, 24]

However, the interpretation of chest radiographs is also subjective and there have been numerous reports of inter-observer variability in detecting infiltrates due to pneumonia on chest radiographs. In the large multi-country European study, appreciable inter-observer variation between radiologists was noted, with a kappa statistic of 0.45.[20] Similarly, moderate agreement between radiologists, or between radiologists and other hospital physicians, in the diagnosis of pneumonia from chest radiographs has been reported in studies from the UK, from elsewhere in Europe and from the USA .[24-26]

Interpretation of chest radiographs may be more complicated still among older adults, due to a higher prevalence of abnormal radiographic findings from pre-existing illnesses such as COPD, heart failure or previous episodes of severe LRTI. As a result, it is possible that the levels of discrepancy in interpretation of chest radiographs may be higher among older populations than in the studies reported above. In the second of the Canadian studies, patients with unconfirmed pneumonia were older and had more severe illness (defined using the pneumonia severity index (PSI), described in section 1.3.2.2) than those with chest radiograph confirmation, and the two groups had similar adjusted in-hospital mortality rates.[22] There were also differences in the underlying aetiology, identified by blood culture; patients with confirmed pneumonia were more

commonly infected with *S.pneumoniae* (64% v 14%), while gram-negative bacteria were more common in the unconfirmed pneumonia group (14% vs 40%).[22]   Thus, while consolidation on a chest radiograph is important to diagnose pneumonia, patients can experience severe disease and poor outcomes when they have a clinical diagnosis of pneumonia without evidence of an infiltrate.

1.1.5   Other lower respiratory tract infections

Pneumonia is a severe subset of a broader group of acute infections of the airways and/or lungs (the trachea and below) known as lower respiratory tract infections (LRTIs). The majority of LRTI are milder than pneumonia and are self-limiting, requiring treatment with antibiotics at most.[27]   However, less severe LRTI can deteriorate and become CAP.   When examined among primary care patients, the risk of progression from chest infection to pneumonia was particularly high in older adults.[28]

The aetiology of community-acquired LRTI in the UK has been investigated by two studies which showed that, among individuals of all ages, causes included (as percent of the total study population) influenza viruses (9%, 24%), *S. pneumoniae* (17%, 19%), *H. influenzae* (10%, 6%), rhinoviruses (4%, 33%) and multiple organisms (13%, 23%).[13, 14]   However, as for CAP, microbiological confirmation of the cause of LRTI is rarely performed as it does not alter the treatment plan.[9]

Generally, community-acquired LRTI short of CAP is diagnosed clinically, following the ERS definition:

> "cough as the main symptom, with at least one other lower respiratory
>
> tract symptom (sputum production, dyspnoea, wheeze or chest
>
> discomfort/pain) and no alternative explanation (e.g. sinusitis or
>
> asthma)".[9]

1.1.6   Burden of CAP

A 2012 review of studies of CAP incidence among adults in Europe has demonstrated wide ranging incidence estimates for CAP, with a sharp increase in rates with increasing age, variation between countries and a generally higher burden in men than women.[29] The older population had considerably higher CAP rates than younger adults.   Incidence

among older adults treated as outpatients ranged from 1.9/1000 population (among Spanish females ≥65 years) to 33.0/1000 (among Finnish adults ≥60 years), while for hospitalised older adults the range was 4.8/1000 population to 10.5/1000 person-years.[29] Incidence has been shown to be particularly high among those with specific co-morbidities such as chronic lung disease, cerebrovascular disease, dementia, liver or renal disease, cancer or immunosuppressive conditions.[30, 31]   Rates are also increased among those taking immunosuppressive medications,[32, 33] among smokers and ex-smokers,[30] those who are underweight and individuals with a history of alcohol abuse.[31]

As for CAP, LRTI as a whole are more common with increasing age; UK studies have provided LRTI incidence estimates varying from 45/1000 population (men aged ≥60) to 121/1000 (adults aged 70-79).[34, 35]

The results of a detailed literature review of CAP and LRTI incidence studies in Europe are presented in Chapter 4.

1.1.7   The importance of CAP in an ageing population

Infections such as pneumonia are a particular burden among older adults, where they are a major cause of morbidity, mortality and healthcare usage. The older population have both an increased risk of acquiring these illnesses, and subsequently suffering from more severe manifestations of disease. This increased susceptibility to CAP (and more broadly LRTI) is due in part to age related loss of functional reserves, and general deterioration of the immune system known as immunosenescence.[36] Ageing also has an effect on many of the mechanical defence systems of the lungs that are outlined in section 1.1.2, including reduction in the elasticity of the lungs, in cough function and in the ciliary mechanism, leading to less efficient clearing of mucus from the lungs.[3] In addition, improved treatment and management of major clinical events such as myocardial infarction (MI) or stroke, and chronic conditions such as diabetes or COPD have enabled patients increasingly to survive and live with diseases which affect their general health. The majority of the UK's older population are thought to be living with one or more long-term condition; the 2011 General Lifestyle Survey found that 58% of 65-74 year olds and 68% of those aged ≥75 years reported having a long standing illness.[37] In many cases, two or more illnesses (known as multi-morbidity) are present;

estimates from Scotland found multi-morbidity existed in 65% of 65-84 year olds and 82% of those aged ≥85 years.[38]

1.1.7.1   Frailty

In addition to the comparatively high prevalence of co-morbidities, older adults are also susceptible to a more general vulnerability in underlying health which is less easy to define. Frailty is typically characterised by declines in functional reserves across a range of physiological systems related to ageing.[39] These cumulative declines result in frail older adults having less resistance to stressors (such as infections), and being more vulnerable to many adverse outcomes ranging from dependency or falls to hospitalisation or death.[39] Frailty is made more complex by it being a dynamic state that can be reversible, although decline is more common than improvement.[40, 41] The ongoing discussion around the definition of frailty has resulted in a number of approaches to measuring it, the two most prominent of which are outlined in section 1.2.1.1.   A 2012 systematic review of the prevalence of frailty among community-dwelling older adults worldwide reported an overall weighted prevalence of frailty of 10.7%. The results from individual studies ranged from 4.0% to 59.1%,[42] with much of this variation attributed to differences in the frailty measures used.

1.1.7.2   The changing demography of the UK

The high burden of CAP among the older population is of particular importance as life expectancy in the UK is generally increasing, and is expected to continue to do so.[43] Recent estimates from the Office for National Statistics (ONS) suggest that the percentage of the UK population aged ≥65 years will increase from 17% in 2010 to 23% in 2035, with the proportion aged ≥85 years rising from 2% to 5% of the population over the same period.[44] This represents almost 7 million additional older adults in 20 years time, including more than two million aged ≥85 years.[45] Similar trends are being seen across Europe,[44] and these demographic changes will have wide-ranging consequences for future healthcare provision and usage.

1.1.7.3   Limitations of current UK CAP incidence studies

In order to adequately plan future resource use, it is imperative that we have accurate estimates of the burden of CAP among the older population, given its increasing size. As shown in detail in the literature review in Chapter 4, current UK incidence estimates of

CAP (and LRTI more generally) are from studies that were restricted to stand-alone data from either primary or secondary care settings, preventing the total burden of these infections among older adults from being calculated. Use of stand-alone GP records may not completely capture hospitalised cases, and use of hospital admissions data excludes patients well enough to receive treatment at home. Additionally, despite the majority of CAP cases occurring in older adults and the risk of infection increasing with age, studies rarely further stratify older populations by age to better understand trends in disease burden within this diverse and growing group.

1.1.8 Vaccination to prevent CAP among older adults

The most effective way to tackle CAP among an expanding older population is to prevent the illness from occurring. Prevention strategies for those aged 65 and over centre on routine vaccination against two pathogens which may be responsible for 50% of CAP in this population;[46] *S. pneumoniae*, and influenza virus.

Pneumococcal polysaccharide vaccine (PPV23) contains purified capsular polysaccharide from the 23 types of pneumococcus which account for around 96% of isolates that cause serious infection in the UK.[47] The introduction of PPV23 vaccination for older adults was staged between 2003 and 2005, and it has been offered to all those aged ≥65 since that time. Vaccination is only recommended once except for those with no spleen, splenic dysfunction or chronic renal disease, for whom vaccination is recommended every five years.[48] Cumulative vaccine coverage among older adults has risen from 29% in 2003/4 to 70.5% by the end of March 2011.[49, 50] However, research has shown a relative lack of long-term protection of PPV23. In England, vaccine effectiveness (VE) was found to wane over time from 48% (95% CI: 32%-60%) within two years of vaccination to 15% (95% CI: -3%-30%) after five or more years.[51] A 2008 Cochrane systematic review additionally highlighted the low VE among adults with underlying health conditions.[52]

Since 2000, influenza vaccine has been offered from September/October each year to all patients aged ≥65 years, as well as younger patients with underlying health conditions. Among older adults, uptake has remained around 73% since 2005/6, which is just below the European Union target of 75%.[53] The vaccine needs to be given each year as the circulating strains of influenza vary, thus the vaccine is tailored to the strains

predicted to be important each season. The changing composition of the vaccine leads to varying levels of VE which is lower in years when there is a mismatch between the vaccine and circulating strains. Regrettably VE has been found to be generally lower in those aged 65 years and older compared to those aged <65.[54] The effectiveness of influenza vaccine at protecting against pneumonia has ranged from 46% (30-58%) effective in well matched years, to having no effect in mismatched years.[55]

1.1.9   Treatment of CAP

As with diagnosis, management of CAP can occur in primary care, or in hospital (where patients may present directly via A&E or via a referral from their GP).

1.1.9.1   Treatment of CAP in primary care

Clinical scores can be used to assess severity of CAP, and thus guide treatment management. One such score is the CRB-65 score, which has been adapted from the CURB-65 score to enable its use in primary care (CURB-65 and CRB-65 are further described in section 1.3.2). For those who have a mild form of CAP and are able to be managed in the community, the BTS recommended treatment is amoxicillin 500mg three times daily for seven days, with doxycycline or clarithromycin as alternatives for patients sensitive to penicillins.[1] Among patients who have signs of a moderate to severe infection, urgent hospital admission should be considered.

For CAP patients treated in the community, the BTS recommend that patients are reviewed after 48 hours (or before if clinically indicated) to reassess disease severity after initiation of treatment. Patients who have not improved may require hospitalisation.[1]

1.1.9.2   Treatment in secondary care: hospitalisation

Patients can be hospitalised for a variety of reasons following onset of CAP, from clinical severity of disease (for moderate/severe CAP) or complications posed by co-existing illnesses, to social factors such as living alone and being unable to care for themselves when ill.

As in primary care, treatment of CAP in hospital depends on the level of severity of disease, again commonly assessed using the CURB-65 score. Antibiotics should be

commenced as soon as possible after diagnosis, and within four hours of presentation to hospital among those admitted.[56] Hospitalised patients may also require additional oxygen therapy or intravenous fluids.[1]

1.1.10 Treatment of other LRTI

In view of the potential for CAP to have evolved from a less severe LRTI, and the increased risk of this occurring among older adults, it is important to also consider LRTI as a broader group when thinking about CAP among those aged ≥65 years.

Current National Institute for Health and Clinical Excellence (NICE) guidelines for antibiotic prescribing in primary care for respiratory tract infections only recommend immediate antibiotic treatment for LRTI (short of CAP) among older patients considered high risk. Patients aged 65 to 79 years must have two of the following conditions to be considered in this group, while those aged 80 years and over must have at least one of; diabetes, a history of congestive heart failure, current steroid use, or hospitalisation in the previous year.[27] A delayed antibiotic prescribing strategy with advice can be considered for less severe cases.

Re-consultation in primary care for LRTI is common, with between 25% and 33% of patients presenting back to their GP within 28 days of their initial appointment.[35, 57-60] Re-consultation may be initiated by the GP or more commonly by the patient.[58]

1.2 Severe outcomes of CAP

In severe cases of CAP, patients may require hospitalisation and/or may die. As with incidence, the risk of severe outcomes following CAP also increases with age, with case fatality rates among older European adults ranging from 13-30%.[29]

1.2.1 Hospitalisation

Previous UK studies have found that between 22% and 44% of adults with CAP are hospitalised.[1] However, these estimates are over 15 years old and do not take into account the probable variability in treatment location by age. In recent years, hospitalisations for pneumonia among older adults have been examined in both the UK and other European countries, and several studies using large hospital admissions

databases to analyse trends in pneumonia hospitalisations over time have been published.

In England, increasing hospitalisations for CAP were initially reported by Trotter et al, who used hospital admissions data for the whole country (provided by Hospital Episode Statistics (HES)) to investigate the rate of first admissions for pneumonia each year between April 1997 and March 2005.[61] Over the study period the hospitalisation rate increased among all age groups, but in particular among older adults. Admissions increased 35% among 65-74 year olds, 28% among those aged 75-84 and 39% among those aged ≥85 years compared to 20% among the under 65's (from 2.63/1000 population, 6.84/1000, 15.99/1000 and 0.7/1000 respectively). The level of co-morbidity recording, measured using a summary co-morbidity score (the Charlson index, described in section 1.2.1.1) increased over the study period, but neither changes in co-morbidity score nor increasing age fully explained the rising number of hospital admissions.

More recently, Bardsley et al explored whether the number of hospitalisations in England for ambulatory care sensitive conditions (ACSC) had changed over time.[62] Again, stand-alone English HES records were used to investigate whether admissions for a range of conditions including pneumonia increased between April 2001 and March 2011.[62] The analysis provided more up to date admission rates for pneumonia, but lacked the level of detail provided by Trotter et al due to the broader scope of the research. Rates were not presented stratified by age, but were instead standardised to the European Standard Population. A 118% increase in the standardised pneumonia admission rate was found over the study period, which was among the highest relative increases for any condition in the study, along with influenza. In addition to the changes in hospitalisations over time, a separate report (also using HES data) highlighted that emergency admissions in England in 2012/13 disproportionately affected older adults, and that this varied by condition; 40% of emergency admissions were for older adults, but among emergency admissions for pneumonia, this rose to 70%.[63]

An increase in pneumonia hospitalisations over time has also been noted in several other European countries. Studies from Denmark,[64, 65] the Netherlands,[66] and

Portugal,[67] have all used routinely collected data to demonstrate rising admissions for pneumonia over the last 10 to 20 years.

1.2.1.1 Factors that may influence trends in CAP hospitalisation

Increasing hospitalisation levels for CAP have been well described, but what is driving these increases is less well understood. The results of a detailed literature review of risk factors for hospitalisation after a CAP episode are presented in Chapter 6. An overview of possible reasons for the increasing trends seen is provided below.

Increasing incidence

It is important to contextualise any trends in hospitalisation against the underlying trends in pneumonia incidence. For example, if CAP incidence increases over a period and the proportion of CAP patients hospitalised remains stable, the number of hospitalisations would also increase. Thus, increasing hospitalisations may simply reflect (at least in part) rising CAP rates. Use of stand-alone hospital admission data such as HES (used in the English hospitalisation studies) does not enable the distinction between increases in underlying incidence of a disease, and increases in hospital admissions over and above any increase in incidence.

Increasing prevalence of co-morbidities

Increasing levels of co-morbidity among the older population may also have played a key role in rising CAP admissions, as co-morbidities may affect both patients' susceptibility to CAP, and the severity of their illness. A commonly used method to summarise the extent of a patient's co-morbidities is to calculate their Charlson index score.

The Charlson Index

The Charlson index was originally published in 1987 as a new method to classify prognostic co-morbidity in longitudinal studies. The index was developed using the medical information of 604 patients admitted to the medical service at a New York hospital over a one-month period in 1984. The vital status of these patients was established at one year, including deaths during the initial hospitalisation and post-discharge. Cox regression was used to quantify the association between a range of

diseases and mortality at one year, and the resulting hazard ratios (HRs) were rounded and used to create the score that is still in use today. Conditions with HRs of 1.2 or less were not included in the final model, HRs ≥1.2 and < 1.5 were assigned a weight of 1; conditions with a HR ≥1.5 and < 2.5 a weight of 2; conditions with a HR of ≥2.5 to < 3.5 a weight of 3; and two conditions with weights of 6 or more were assigned a weight of 6.

Seventeen co-morbidities were included in the final score, two of which had scores which increased with increasing severity of disease (liver disease and diabetes), resulting in 19 factors to be considered. These co-morbidities and their associated scores are presented in Table 1-1.

Table 1-1 Scores assigned to co-morbidities when using the Charlson index

Score	Co-morbidity
1	Myocardial infarction
	Congestive heart failure
	Peripheral vascular disease
	Cerebrovascular disease
	Dementia
	Chronic lung disease
	Connective tissue disease
	Peptic ulcer disease
	Mild liver disease
	Diabetes
2	Hemiplegia
	Severe renal disease
	Diabetes with complications
	Solid cancer
	Leukaemia
	Lymphoma
3	Moderate/severe liver disease
6	Metastatic cancer
	AIDS

The total score was calculated for each patient by summing the scores for each co-morbid condition that they had, and their final co-morbidity score was then categorised as none (0), mild (1-2), moderate (3-4), or severe (≥5).[68] External validation of the score was performed in a cohort of 685 women first treated for breast cancer between 1962 and 1969 at Yale hospital.[68] In this cohort, 10-year mortality was the outcome of interest (although as the cohort were all breast cancer patients those who died of breast cancer were categorised as having left the study rather than having died). Formal

assessment of the model's performance (using methods outlined in section 7.3) was only assessed graphically in the original paper.

As mentioned previously, the majority of the UK's older population are now thought to be living with at least one long-term condition.[37] Of the European papers reporting increasing CAP hospitalisations above, only two studies investigated the possible contribution of rising co-morbidity levels through use of the Charlson index.[61, 64] When assessing the suitability of the Charlson index to adjust for co-morbidity status, it is important to consider its development and intended application as a score to predict mortality. While hospitalisation and death are both considered severe outcomes of disease, risk factors for these two events may differ, and therefore a score developed to predict mortality may not be the best tool to explain hospitalisation trends. The use of a score to adjust for co-morbidities also precludes identification of the risk of hospitalisation associated with individual co-morbidities. Among the growing older population it would be useful to better understand the role individual conditions such as dementia and chronic respiratory disease play in CAP hospitalisation trends, in order to inform clinicians and plan future resource allocation. Furthermore, stand-alone hospitalisation data (as used in the two English studies) have suboptimal recording of patients' co-morbidities, including only those pertinent to patients' care.

Medications and vaccinations

A consequence of the rising prevalence of co-morbidities is increasing prescription medication use. Some of these drugs, such as immunosuppressive medications used to treat conditions such as rheumatoid arthritis and chronic lung disease, increase patients' risk of CAP.[30] Conversely, some medicines such as statins may offer some protective effect,[32] as does prompt treatment of LRTI with antibiotics.[28] Interestingly, hospitalisations for CAP have risen despite the introduction of vaccinations for influenza and pneumococcal disease for older or at risk groups.[53, 48] There is evidence that influenza vaccine protects older individuals against hospitalisation after pneumonia,[69-71] however, PPV23 has not been shown to have such an effect.[72] The effect of medications or vaccinations on the rising hospitalisation levels has been relatively little studied. Vaccination and medication use are simply not recorded in HES, and as a result their effect on CAP hospitalisation levels in England has not been quantified.

Increasing prevalence of frailty

As highlighted in section 1.1.7, older adults are more prone to a general vulnerability in underlying health, called frailty. Frail patients are identified using a number of factors (due to the multiple systems involved), and many studies have developed models to this end. The inclusion of a range of factors, rather than simply summing the number of co-morbidities an older adult has, enables identification of frail older adults who do not have life threatening diseases but who have experienced physiological changes which make them more susceptible to adverse events. Two frequently discussed measures of frailty are the frailty phenotype and the frailty index.[73, 74] The 'phenotype' hypothesises that frailty can be recognised from a set of five deficits: measured slow walking speed, measured impaired grip strength, self-reports of declining activity levels, exhaustion and unintended weight loss. Patients with a score of three or more deficits are categorised as frail, and those with one or two deficits are pre-frail.[73] While the phenotype method is simple to use and extensively validated, the factors it contains are not routinely collected in primary or secondary care, limiting its usefulness in either setting.[39]

An alternative methodology is the 'index' (or cumulative deficit approach) whereby information on a large number of deficits is collected across co-morbidities, clinical signs and symptoms, disabilities or abnormal test findings. To be considered for inclusion, deficits must accumulate with age, be biologically plausible, and not saturate too early (i.e. the deficit cannot have a prevalence of 100% before older age).[74] Several indexes have been developed, and their ability to predict adverse outcomes found to be high as long as more than 30 deficits are included.[74] Patients' scores are calculated as the proportion of deficits they have, with higher proportions correlated with increasing susceptibility to adverse outcomes. Across these models, deficits have been found to accumulate within patients at on average 0.03/year.[74] The proportion of deficits tolerated by patients seems to be limited at around 0.67, after which no further accumulation is sustainable and death becomes likely.[75]

Despite the abundance of frailty measures available, frailty is not commonly included in studies assessing severe outcomes following CAP among older adults. However, it certainly plays a large role in clinicians' treatment location decisions for some patients.

1.2.2 Mortality

1.2.2.1 Short-term CAP mortality

CAP is the most common cause of death from infection in Europe and the US.[2, 76] In 2013 influenza/pneumonia was the fourth leading cause of death among women in England and Wales, and the sixth among men, accounting for 5.9% and 4.7% of deaths respectively.[77] Age specific pneumonia mortality rates from the ONS in men ranged from 3.99/1000 among 65-74 year olds to 258.99/1000 in those aged ≥90 and in women from 2.86/1000 among 65-74 year olds to 207.79/1000 in those aged ≥90.

However, these rates underestimate the total contribution of pneumonia to mortality in England and Wales. In 2001, the coding system used changed from the Ninth to the Tenth Revision of the International Classification of Diseases (ICD-10), and the rules for assigning cause of death were modified.[78] In brief, following changes to Selection Rule 3, the condition that led directly to the death was no longer assigned as the underlying cause when the disease thought to have started the fatal sequence was also recorded on the death certificate.[79] For example, a patient with dementia who died directly from pneumonia would, from 2001 onwards, have dementia coded as their underlying cause (rather than pneumonia pre-2001). Subsequent analysis showed that of deaths coded as due to pneumonia in 1999 using ICD-9, only 60% would have also been coded so using ICD-10; 13.7% changed coding to circulatory diseases, 10.1% dementia or Alzheimer's and 5.5% neoplasms.[78] Pneumonia is still recorded as a contributory cause; thus it is important to examine both underlying and contributory causes of death when investigating pneumonia mortality.

1.3 Severity assessment of CAP

To aid clinicians in deciding which patients are at increased risk of mortality, a severity assessment is usually undertaken. Severity assessment is one of the cornerstones of pneumonia management – as such several tools have been developed in order to inform this process (outlined in sections 1.3.2.1 to 1.3.2.3). Before discussing these tools I provide a brief introduction into the methods used to develop them.

1.3.1 Overview of prognostic modelling

Prognostic models (also known as risk scores) are used in many areas of medicine in order to aid clinical decision making. Patients' clinical and/or non-clinical characteristics are used to estimate the probability of experiencing a specified outcome, given the set of characteristics provided.

An advantage of prognostic models is that they incorporate several different factors which may influence the likelihood of an event, and so the prediction of risk does not simply depend on one factor. The models are developed using multivariable modelling. Commonly, the factors included in the model are weighted using the models' regression coefficients, which represent their importance as risk predictors. The sum of the weights for the factors present in a patient provides their total score. The score is then used to estimate the predicted absolute risk that the patient will experience the outcome.

1.3.2 Use of prognostic models to identify CAP patients at increased risk of mortality

There have been a large number of studies undertaken to enable identification of patients at short-term (usually 30 day) increased risk of mortality from CAP, most of which were designed to be used at the point of diagnosis. Two of the most commonly used and cited scores, CURB-65 and the Pneumonia Severity Index (PSI) are outlined below.[80, 81]

1.3.2.1 CURB-65 and associated scores

The CURB-65 and CRB-65 prognostic scores were developed by Lim et al in 2003.[80] They were based upon the modified British Thoracic Society (mBTS) rule, a score developed to identify patients with severe CAP and at high risk of mortality, which was modified from the existing BTS rules for assessing severity of CAP at point of hospital admission. The mBTS included confusion, urea >7 mmol/l, respiratory rate > 30/min and diastolic blood pressure < 60 mm Hg.[82] Lim et al expanded the mBTS score to create a more general rule with several categories which could separate patients into groups according to mortality risk, and suggest appropriate management strategies. The expanded score was developed and validated using prospectively collected CAP hospital admission data from 1068 patients in the UK, New Zealand and the Netherlands.[80]

New factors identified as independent prognostic indicators were only included in the final model if they were routinely available at the point of admission to hospital (serum albumin was excluded from the final model for this reason). Additionally, the final CURB-65 model was further adapted to include only clinical features and exclude laboratory results (CRB-65). The final CURB-65 and CRB-65 scores are outlined in Table 1-2.

Table 1-2 Factors included in CURB-65, CRB-65 and their scores

	Points assigned	
Factor	CURB-65	CRB-65
Confusion	1	1
Urea >7 mmol/l	1	N/A
Respiratory rate ≥30/min	1	1
Low **B**lood pressure (systolic <90 mm Hg or diastolic ≤60 mm Hg)	1	1
Age ≥**65** years	1	1
Risk category & recommended treatment	Score (Mortality risk, %)	
Mortality low, likely suitable for home treatment	0 or 1 (1.5%)	0 (1.2%)
Mortality intermediate, consider/likely to need hospital referral/assessment/treatment	2 (9.2%)	1 or 2 (8.15%)
Mortality high, manage in hospital as severe pneumonia	≥3 (22%)	3 or 4 (31%)

The removal of test results (the urea measurement) enabled CRB-65 to be used in primary care. However, the score was developed using hospitalised CAP patients, and a systematic review of studies using CRB-65 found that while the score performed well in hospitalised patients, it consistently over-predicted the probability of 30-day mortality in community-based patients.[83] One of the validation studies included in the systematic review included only primary care patients aged ≥65 years in their validation of CRB-65.[84] Due to their age, none of these patients were categorised as 'low risk' (score=0), and the authors suggested that among the older primary care population, changing the cut-off for 'high risk' from ≥3 to ≥2 may be more appropriate. It should be noted that neither the original CRB-65 derivation study, nor the validation of the score in older primary care patients included patients residing in nursing homes in their study populations, potentially limiting the accuracy of predictions of the scores in older, more frail populations.

1.3.2.2 The Pneumonia Severity Index (PSI)

The PSI was developed in the USA in 1997 to identify patients at low-risk of dying within 30 days of a pneumonia diagnosis, who may be suitable for ambulatory treatment.[81] The index was derived using a cohort of >14,000 hospitalised cases of pneumonia in the USA in 1989, and used in-hospital mortality within 30 days of diagnosis as the outcome. Of the three demographic factors, six co-morbidities, five abnormal physical findings and seven abnormal laboratory findings considered, 20 were included in the final index and each assigned a score (Table 1-3). Patients' total scores were then spilt into five risk groups, and the index externally validated in two separate cohorts.

Table 1-3 Factors included in the Pneumonia Severity Index and their scores

Type of factor	Factor	Points assigned
Demographic	Men	Age (years)
	Women	Age (years) – 10
	Nursing home resident	+ 10
Co-morbidities	Cancer	+ 30
	Liver disease	+ 20
	Congestive heart failure	+ 10
	Cerebrovascular disease	+ 10
	Renal disease	+ 10
Physical examination findings	Altered mental status	+ 20
	Respiratory rate ≥30/min	+ 20
	Systolic blood pressure <90mm Hg	+ 20
	Temperature <35°C or ≥40°C	+ 15
	Pulse ≥125/min	+ 10
Laboratory and radiographic findings	Arterial pH <7.35	+ 30
	Blood urea nitrogen ≥30mg/dl (11 mm/l)	+ 20
	Sodium <130mm/litre	+ 10
	Glucose ≥250mg/dl	+ 10
	Hematocrit <30%	+ 10
	Partial pressure of arterial oxygen <60 mm Hg	+ 10
	Pleural effusion	+ 10
Risk category	**Total score**	**Risk of death (%)**
I	≤50 (age ≤50, no co-morbidity, no abnormal physical findings)	0.1
II	≤70	0.6 – 0.9
III	71-90	0.9 – 2.8
IV	91-130	8.2 – 9.3
V	>130	27.0 – 29.2

Due to its more complex nature, the PSI is not routinely used in UK settings and the BTS recommends the use of the more simple CURB-65/CRB-65 to assess pneumonia severity.[1]

1.3.2.3   Other models

Prognostic models to predict mortality risk after CAP developed in and designed specifically for use in primary care are considerably less common than those for hospitalised CAP patients.   Models tailored specifically to the higher-risk, older population with their high co-morbidity burden would be of particular use in this setting. Bont et al developed a model for use in primary care to predict older adults' risk of hospitalisation or death in within 30 days of an LRTI, including CAP.[85] The model was developed using data from the database of the Utrecht GP research network, and validated using data from the Second Dutch National Survey of General Practice. It included factors that should be included in a patient's medical notes, so no additional tests were required for its use.   The final model included the type of LRTI diagnosis (bronchitis, COPD exacerbation, or pneumonia), age (65-79, ≥80 years,) congestive heart failure, diabetes, steroid use, hospitalisations in the last year (0, 1, ≥2) and use of antibiotics in the previous month. A further model was subsequently derived using the same development cohort, but for use specifically in patients aged ≥80 years.[86] The model was built with the same outcomes and aims, and included the same factors as the Bont et al model, with the exception of congestive heart failure.

A prediction rule specifically for use in older adults during the influenza season was developed by Hak et al,[87] in order to predict a combined outcome of hospitalisation for pneumonia or influenza, or death (all causes) during the influenza season. The model was limited to the influenza season as the authors' aim was to encourage influenza vaccination among the older population. Data from three health plans in the US were used to form large cohorts of older adults; patient records from one plan were used to develop the model, and the rest of the data used for its validation.   Fifteen clinical characteristics were considered as potential predictors. The final model consisted of age (<70, 70-74, 75-79, 80-89, ≥90), sex, outpatient visits in the previous year (0, 1-6, 7-12, ≥13), previous hospitalisation due to pneumonia/influenza, pulmonary disease, heart disease, renal disease/transplant, dementia/stroke and non-haematological/haematological cancer.

While these studies highlighted the important role of individual co-morbidities in post-LRTI mortality risk in older adults, none of the studies focussed solely on CAP and all used combined outcomes of hospitalisation or death. Risk factors for hospitalisation

may differ to those for death, and the use of a combined outcome does not allow these effects to be teased apart. Additionally, important predictors such as vaccination status and frailty were not considered for inclusion in the models.

1.3.2.4 Longer-term increased mortality post-CAP hospitalisation

Interestingly, it has also been shown that there is a longer-term effect of CAP on patients' risk of mortality, for a year or more after CAP diagnosis.[88] The first large study to demonstrate this was by Kaplan et al, who used US hospital discharge data from Medicare (a large database of American older adults' health insurance administrative claims) to compare mortality rates between older CAP patients (n>15,000) and age- sex- and race-matched controls (patients hospitalised for other conditions, n>75,000).[88] In-hospital mortality was higher in the CAP group compared to the hospitalised control group (11% vs 5.5%). Importantly, among the patients who survived the hospitalisation, one-year post-discharge mortality rates (adjusted for co-morbidity) remained higher among CAP patients compared to the controls (33.6% vs. 24.9%). The monthly risk of death in both groups decreased over the year-long period, but remained higher in the CAP group up to and including the twelfth month of follow-up (1.92% vs. 1.37%). When these results were standardised to the general US population, the mortality risk among the CAP cohort was considerably higher than that of the population in general. Increased long-term mortality post-CAP has since been shown in several additional studies among older adults, using mortality periods ranging from 90 days to several years.[89-92]

The exact mechanism behind the increased mortality risk after a CAP hospital discharge is not known – it could be that the CAP episode leads to increased mortality, or that the CAP is due to patients being less well generally, and thus it is a marker for underlying ill-health. Many studies have looked at physical markers during hospitalisation, but there is no universal agreement on what lies behind this excess mortality.[93]

Certainly, some effects of CAP remain after the infection clears. As described in section 1.1.2, infection with CAP triggers a rapid increase in pro-inflammatory cytokines, which can lead to worsening of disease and in severe cases lung damage. Inflammatory exudate enters the alveoli and can prevent the amount of air reaching the alveoli from equalling the amount of blood reaching the alveoli. Thus, a lower level of oxygen is

transferred to the blood (hypoxia), while removal of carbon dioxide remains stable. This ventilation/perfusion (V/Q) mismatch leads to type 1 respiratory failure. However, sub-clinical inflammation has been shown to continue even after clinical recovery.[94, 95] This has a range of effects on the body.

Many of these effects increase the short-term risk of acute cardiovascular events. Among patients with atherosclerosis, increased inflammatory activity within coronary atherosclerotic plaques can make them unstable and increases the possibility of their rupturing.[96] In addition, systemic inflammation temporarily disturbs endothelial function, which results in decreased vasodilation, and an increase in the procoagulant properties of the blood by promoting platelet activation and adhesion of leukocytes.[96, 95] Platelets can also be activated by bacterial products such as lipopolysaccharide, which induces mechanical stress.[95] Aggregation of platelets on the surface of a ruptured atherosclerotic plaque causes thrombus formation, increasing the risk of an acute ischaemic event in the coronary or cerebral vessels.[95]

Myocardial injury may result not only from the cytokines arising from the inflammatory response, but also via endotoxins, or infection of the cardiomyocytes with the pathogen which caused the pneumonia.[96] Additionally, demand ischaemia can occur as a result of the decreased ratio of metabolic supply to demand. This is caused by the combined effects of V/Q mismatch, and the increased heart rate which is a consequence of the systemic response to CAP.[96]

Furthermore, the systemic inflammatory response can result in acute kidney injury, and impaired sodium and water metabolism, which leads to volume overload. Medications with a high sodium content, such as some antibiotics, can exacerbate this effect.[96]

The results of a literature review of studies that investigated factors associated with long-term mortality after a CAP hospitalisation are presented in Chapter 7.

1.3.2.5 Longer-term risk predictions

CURB/CRB-65, PSI and the other models described previously are useful for short-term assessment of mortality and treatment decisions. However, they do not assist clinicians concerned about their patients' health after discharge from hospital for CAP, or their heightened long-term mortality risk. Increasing levels of CAP hospitalisation among a

growing older population, who may have a raised mortality risk post-hospital discharge, will result in a large burden on primary care. A risk-stratification tool to aid GPs in recognising their patients' longer-term mortality risk post CAP hospital-discharge would help decision making around resource allocation, and which patients are in need of extra care or additional check-ups.

1.4   The benefits of using linked data to investigate the burden and outcomes of CAP

Use of linked primary and secondary care data (described in detail in Chapter 2) can overcome many of the limitations of existing studies. The burden of CAP is more fully captured by the inclusion of both cases treated in the community, and those that required treatment in hospital.

Utilising linked data enables any increase in CAP hospitalisation to be investigated separately to any increase in CAP incidence, which has not been possible in existing studies (described in more detail in Chapter 6). The rich patient histories provided in combined primary and secondary care records allow extensive investigation of the effects of co-morbidities, medications, vaccinations, lifestyle and frailty factors on treatment setting over time, that would not be possible using stand-alone data sources. Additionally, the pathways of care for patients with CAP, from initial diagnosis and treatment in general practice to hospital admission and death can be reviewed. Without being able to separate increasing incidence from increasing hospitalisations, it is difficult to know where to direct healthcare resources – into slowing CAP incidence, or targeting use of appropriate treatment settings. Likewise, it would be of considerable use to better understand how patients hospitalised for CAP differ from those treated in an ambulatory setting. The use of linked data enables these questions to be explored.

Finally, these data facilitate the development of risk scores to be used in primary care to assess patients' longer-term mortality risk post CAP hospital-discharge. Hospital admission records provide diagnostic codes for the main and secondary conditions treated in each period of care, as well as dates of discharge (used to define cases), primary care records contain patients' co-morbidity status and general health profile (used to define potential predictors of mortality), and mortality records supply accurate

dates of death (used to define the outcome) as well as underlying and contributory causes of death (used to investigate the cause).

1.5 Rationale

In summary, this chapter has highlighted the particular importance of CAP among older populations. Existing studies of CAP and other LRTI in the UK have not focussed on this high risk, high burden population, and are limited by their use of stand-alone primary or secondary care data. By linking these data sources, it is possible to estimate more completely the incidence of these infections, and to differentiate between community- and hospital-acquired cases. Trends in hospitalisation can be separated from those for incidence, and risk factors that may also explain increasing hospitalisations be thoroughly investigated. In addition, factors that may predict longer-term mortality risk post-CAP can be assessed and combined to develop a prognostic score to aid clinical decision making.

1.6 Research objectives

In this thesis I have two main research aims, each with more than one objective.

1.6.1 Aim 1: To develop methods using linked electronic health records to better understand the burden of CAP and LRTI among older adults in the UK

Objective 1 – To use linked primary and secondary care records to better quantify the incidence of LRTI and CAP among older adults in the UK over time.

This is informed and supported by two supplementary objectives:

Objective 1a – To define an appropriate exclusion period for new patients' records at the start of follow-up, in order to ensure reports of historical episodes of disease are not included in incidence analyses.

Objective 1b – To compare CAP incidence estimates using stand-alone GP records to those from GP records linked to hospital admissions, to better understand the added value of using the linked data.

1.6.2 Aim 2: To identify risk factors for severe outcomes after CAP in the older population.

Objective 2 – Among older patients with CAP, to identify the risk factors for hospitalisation after a CAP episode, and to assess to what extent these factors contribute to the increasing hospitalisation trend over time in England.

Objective 3 - To estimate the risk of mortality in the week, month and year after discharge from hospital for CAP by developing prognostic models for each period, to attempt to aid GPs when making decisions about post-discharge care for these patients.

1.6.3 Outline of Chapters

In Chapter 2 I provide detail on the data sources utilised for these analyses, and the methods used to collate the multiple data sources into a single resource. The statistical methods used in each chapter are briefly outlined.

In Chapter 3 I investigate the appropriate point at which to start including new patients' records in analyses, in order to include incident rather than historical reports of CAP (Objective 1a). The findings of this chapter are used in all subsequent work.

In Chapter 4 I begin by providing the results of a review of the literature on the incidence of CAP and LRTI in older adults in Europe. This is followed by my own analysis of the incidence of LRTI and CAP among the UK's older population between 1997 and 2011 using linked data (Objective 1), which is presented as a paper published in PlosOne.[97]

The difference between CAP incidence estimates using stand-alone GP records and linked GP hospital admissions records (Objective 1b) is the focus of Chapter 5. Estimates from the same cohort of individuals are compared to better understand the potential added value of using the linked data.

The final two analyses focus on CAP, looking at factors associated with severe outcomes following a CAP diagnosis.

Firstly, in Chapter 6 I start by reviewing the literature on risk factors for hospitalisation after CAP. I then compare CAP patients treated in the community to those hospitalised within 28 days of diagnosis, in order to identify risk factors for hospital admission

(Objective 2). A range of co-morbidities, frailty and other factors are investigated, as is their contribution to the increasing level of hospitalisation after CAP between 1998 and 2011. In addition to the main analysis, presented as a paper published in BMJ Open, supplementary methods and an exploratory analysis are provided.

In Chapter 7 the focus is on patients who are discharged from hospital having been admitted for CAP, and their increased mortality risk in the year post-discharge. I first review the literature on factors associated with long-term mortality after a CAP hospitalisation. In an attempt to aid GPs when making decisions regarding these patients, I then develop a series of prognostic models to predict patients' risk of death in the first week, eight to thirty days and 31 to 365 days post-discharge (Objective 3). The models use information routinely collected in GP records, so that they could be incorporated into GP medical record software in the future for automatic risk calculation.

Finally, in Chapter 8 I discuss the findings of this thesis. The strengths and limitations of both the work I present, and more generally the use of linked electronic health records are considered. Further areas for research are outlined and the clinical implications of the work are discussed.

Chapter 2   Research methods

This chapter describes the data sources used in this thesis. The identification of the initial cohort of individuals included in the analyses is described, followed by the methods used to find CAP and other LRTI records and their subsequent management into episodes of illness (used to estimate incidence for objective 1). Nested subsets of these episodes then provided the data for the analysis of risk factors for hospitalisation post-CAP (objective 2) and the prognostic score for mortality risk after CAP hospitalisation (objective 3), as outlined in section 2.1.6.

2.1   Data sources

I mainly utilised anonymised data from three sources; primary care, secondary care and mortality records. For incidence analyses (presented in Chapter 5), linked data on Index of Multiple Deprivation (IMD) were also used.

2.1.1   Primary care data from CPRD GOLD

The Clinical Practice Research Datalink (CPRD, formerly known as the GPRD – General Practice Research Database) was started in 1987, and is now one of the world's largest databases of primary care electronic medical records. At the time of this study it contained anonymised information for >14 million patients who were broadly representative of the UK population with respect to age, sex and region.[98, 99] Patients' inclusion in CPRD is optional, and individual patients may opt out of being included in the database at any point. Opt-out rates are extremely low, and in 2013 were reported at only 1,000 out of 12 million patients.[100]

General practitioners (GPs) who contribute to CPRD record diagnoses, signs and symptoms as standard in clinical practice. Diagnoses are coded using Read codes, with additional fields and coding systems used where needed. Other relevant data including referrals, test results and prescription information are also provided.

2.1.1.1   Data quality

To ensure that the data remain of an appropriate quality, CPRD carry out internal checks on each collection of data submitted by every practice. Two assessments are undertaken; that there is continuity in data recording for a number of different aspects

of the patient's record, and that the number of deaths recorded over time is broadly within an expected range. If a meaningful gap in continuity or death recording is found, the date at which the gap ends is identified. The practice is assigned an 'up to standard' (UTS) date at the latest of these dates and the data deemed to be research quality from this point onwards. The generalisability and validity of both CPRD and HES data are discussed in section 2.1.5.

2.1.1.2 Data supplied by CPRD

CPRD provide data to researchers in several file types which include different categories of information. Patients are assigned a unique identifier which enables their records to be linked across the files, and a consultation identifier allows events from the same consultation to be connected. The file types utilised in this work are outlined below:

Patient file

Demographic details such as patients' sex and year of birth are provided in the Patient file, which also includes important dates such as the most recent date the patient joined the practice, and (when relevant) the date they transferred out of the practice.

Practice file

The Practice file gives information on the region where the practice is based, the date from when data from the practice were deemed to be of research quality (UTS date), and the date of the last data collection from the practice.

Consultation file

This file details the type of consultation, for example if it was a visit to the practice, a telephone call, night visit, discharge summary or hospital letter and so on.

Clinical file and Additional Clinical Details file

The Clinical file contains patients' medical history data including diagnoses, signs and symptoms, coded using Read codes. Diagnoses made during any admissions to hospital, provided via hospital letters or discharge summaries are also able to be recorded in this file. The clinical file can be linked to an 'Additional Clinical Details' file which is a

structured area for the GP to enter information on a wide range of specific topics, such as smoking status, alcohol consumption, weight, or problems with mobility or self-care.

Referral file

Patient referrals to external care centres are provided in this file. Diagnoses – the reason for the referral - are included, as is the referral speciality and type, such as day case, inpatient etc. and the level of urgency.

Test file

Tests ordered are entered as well as the results, and a Read code for the diagnosis or signs and symptoms.

Therapy file

Therapy data are entered from a Multilex Drug Dictionary. All prescriptions issued by the GP are recorded with the date of issue, and the British National Formulary (BNF) chapter and product code. Additionally, the number of tablets, numeric daily dose, number of days the medicine is to be taken, and whether the prescription is a repeat may be recorded.

Immunisation file

Vaccinations offered to the patient are detailed, along with whether they accept. Vaccines given are documented with the date, and whether the immunisation is routine.

2.1.2 Secondary care data from HES

HES data comprise anonymised data on admissions to NHS hospitals in England. Hospitalisations in HES have admission and discharge dates provided and this period of time is known as a 'spell'. Each spell may contain more than one episode (which denotes a period of consultant care) and each episode can contain up to 20 diagnoses. The primary diagnosis of each episode is defined in the NHS data dictionary as:

> "the main condition treated or investigated during the relevant episode of healthcare".[101]

Within the first episode of patient care, I took the primary diagnosis to represent the reason for the patient's admission. Demographic information such as year of birth and sex is also provided, as is supplementary information such as the method by which the patient was admitted to hospital (for example via A&E) and whether they were discharged to their usual home, to another facility, e.g. residential care, or if they died in hospital.

Unlike the diagnoses contained in CPRD, which are largely entered by the GP during a patient's consultation, diagnoses in HES data are entered by a team of clinical coders after a patient has been discharged from hospital. The clinical coders translate the diagnoses in a patient's medical notes into a series of diagnostic codes using the ICD-10 coding system.

2.1.3   ONS mortality data and IMD

Mortality data for individuals in England are available via the Office for National Statistics (ONS). Anonymised death certificate data are provided which include the patient's date of death, the underlying cause of death and up to 15 contributory causes coded using ICD-10.

Socioeconomic data are available via IMD quintile, a deprivation score calculated by combining a number of factors across seven domains (income, employment, health deprivation and disability, education, skills and training, barriers to housing and services, crime and living environment).[102] IMD quintiles are provided at individual patient level for patients in practices that have consented to linkage. Quintiles are calculated over Lower Super Output Areas (LSOA), small-level geographic regions defined by ONS which contain between 1000 and 3000 people.[103]

2.1.4   Data linkage

English CPRD practices can consent to having their patients' data linked to a number of supplementary data sources such as HES, ONS mortality statistics and IMD quintile. Data linkages are undertaken by a trusted third party to ensure patient anonymity. For HES data, deterministic linkage is undertaken using a patient's NHS Number, date of birth, sex and postcode which are run through a stepwise algorithm.[104]

Individual ONS mortality records are initially linked via HES, and subsequently via CPRD if no link via HES is established. Each record linkage is ranked according to the quality of the match (on NHS number, sex, date of birth, and postcode) between the data sources. The quality of ONS-HES links range from 1 (high quality) to 8 (low quality), but only records assigned a value of 4 or less are deemed to be of sufficient quality to be provided for research use. All records which cannot be linked via HES, but can be linked via CPRD practice information are assigned a score of 0 and are also provided to researchers. The matching scoring criteria for records made available for research use are shown in Table 2-1.

Table 2-1 ONS-HES, and ONS-CPRD record match quality values

Match quality	Matches on HES information
1	NHS number, sex, date of birth, and postcode
2	NHS number, sex, and date of birth
3	NHS number, sex, date of birth (partial match), and postcode
4	NHS number, sex, and date of birth (partial match)
	Matches on CPRD practice information (where no match to HES was established)
0	NHS number, sex, date of birth, and postcode

Patients are linked to IMD scores via their postcode. The scores are then divided into quintiles, and the quintiles provided to researchers.

Linked CPRD, HES and ONS data were available for >50% of English CPRD practices at the time the work on this thesis commenced.[104] Data linkages are not available for CPRD practices in Scotland, Wales or Northern Ireland due to the different versions of HES collected in these countries. HES data are available for linked English CPRD practices from 1st April 1997. For the work in this thesis I used records from ONS linkage for patients whose deaths were recorded from 2001 (the start of death coding using ICD-10), and IMD quintile data from 2007.

2.1.5 Generalisability and validity of the data

Both CPRD and the subset linked to HES have been shown to contain patients and practices that are broadly representative of the UK population with respect to age, sex and geographic region.[99, 105] CPRD has been used extensively for epidemiological

research across a range of conditions. A systematic review of 212 papers researching the validation and validity of 183 different diagnoses recorded within the data found that validity was high; respiratory diagnoses were confirmed in a median of 88% of diagnoses when validated internally (by checking against freetext or signs or symptoms also recorded in CPRD) or externally (by requesting additional anonymised information on the illness from the GP, for example hospital discharge letters). Overall a median of 89% of all diagnoses across a range of disease groups were found to have been confirmed by internal or external validation.[106]

It is important when using clinical data for research to consider the potential impact of external influences on trends in recording practices. One such influence is the introduction of the Quality and Outcomes Framework (QOF) in 2004. QOF is a voluntary annual programme for English general practices, in which they are scored against a series of indicators, and financially rewarded when they meet set targets.[107] Indictors are broadly categorised as clinical (measures pertaining to specific diseases such as COPD, or diabetes) and public health (measures such as the percentage of patients aged >15 who have a recorded smoking status in the previous 15 months). The number of clinical indictors varies per condition, and several indictors include the regularity of recording of important clinical measures such as lung function in COPD patients, or blood pressure measurements among diabetics.[108] These indicators have resulted in increased levels of recording among patients with QOF conditions (in order for indicator targets to be met). However, there is also some evidence that QOF has had a slightly detrimental effect on recording of some non-incentivised activities.[109]

The introduction of QOF has resulted in appreciable improvement in recording of some lifestyle factors. In particular, the recording of smoking status in CPRD has improved considerably over this period. In 1996, current and former smoking status were recorded at 79% and 29% of their respective rates when compared to the British National Household Survey.[110] By 2007-2011, CPRD prevalence of current smokers was within 1% of that from the Health Survey for England, although prevalence of former smoking was still up to 7% lower.[111] To a lesser extent, increasing recording in the last three years has also been reported for alcohol consumption and BMI (up to approximately 60% and 50% respectively by 2011).[99]

HES is less extensively validated than CPRD, as contact with care providers to check patient records and diagnoses is less easily achievable. The Audit Commission has been auditing clinical coding within hospitals since 2007/08, and has found that the accuracy of clinical coding is improving.[112] They found that the average clinical diagnosis error rate decreased from 17% in 2007/08 to 13% in 2009/10.[112] Information on the accuracy of specific diagnoses over this period was not provided.

2.1.6   Overview of the data used for each study objective

Figure 2-1 represents the populations used for each study objective.

In order to estimate the incidence of CAP (and LRTI as a whole, objective 1) I utilised both unlinked and linked CPRD data, including HES records when patients were eligible for linkage (all records included within the white box and those nested within it in Figure 2-1). Using these data, I identified records pertaining to LRTI, to pneumonia or to hospitalisation and created LRTI and pneumonia illness episodes which were subsequently categorised as community or hospital-acquired infections (described in detail in sections 2.4.1 and 2.4.2).

Patients in the initial cohort who were eligible for linked-HES data and who had a CAP episode were included in the analysis of objective 2, identification of risk factors for hospitalisation after CAP. (All records within the medium and dark purple boxes in Figure 2-1).

Patients from the hospitalisation cohort who were hospitalised with CAP and survived until at least the day after discharge were included in the analysis of objective 3, development of prognostic models to predict mortality after CAP hospitalisation. (All records in the dark purple box in Figure 2-1).

Figure 2-1 Hierarchy of study populations for study objectives 1, 2 and 3.
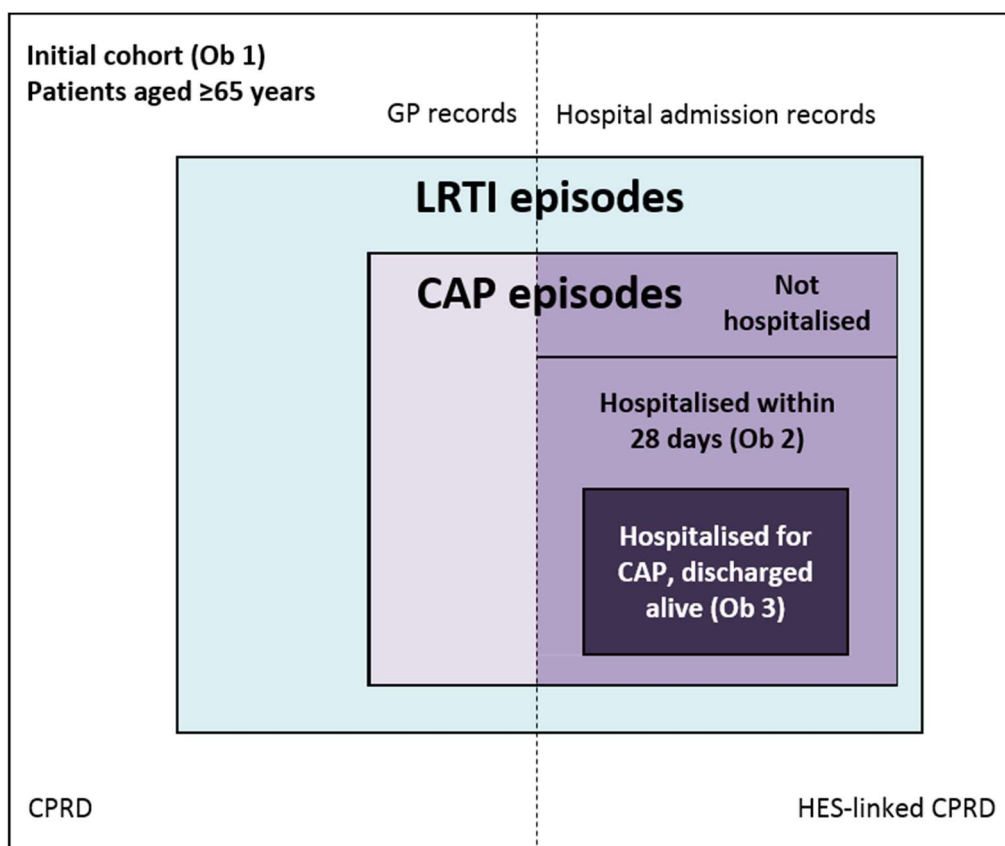
2.2   Initial cohort: eligibility criteria and study period

Patients records were processed through a series of steps as outlined in Figure 2-2 and described in detail below.

Figure 2-2 Process of defining the initial cohort
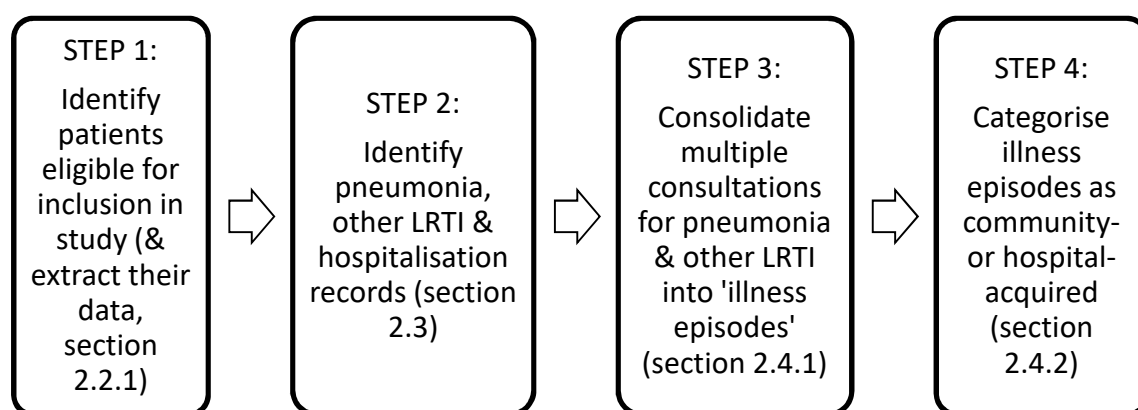
All patients aged ≥65 years who were registered with an up to standard CPRD practice for ≥1 day between 1st April 1997 and 31st March 2011 were included in the initial cohort. Within this group, patients eligible for linkage to HES, ONS or IMD records were identified (via CPRD), and their linked data retrieved.

2.2.1 Data extraction

Detailed CPRD records were obtained for all patients who had one or more pneumonia or other LRTI consultation in CPRD or any LRTI hospitalisation in HES (as defined in section 2.3.2) within the study period. Patients who met the inclusion criteria but had no LRTI records during the study period were included as the denominator in incidence analyses (Chapters 3, 4 and 5). Limited data such as demographic details were provided for these patients in the CPRD denominator files.

2.2.2 Period of time patients' records were used for pneumonia and LRTI diagnoses

When establishing the initial dataset of pneumonia and LRTI illness episodes, patients' records were included from the latest date of their 65th birthday, their current registration date (CRD) (plus the time established in the analysis in Chapter 3), 1st April 1997 and the practice's up-to-standard (UTS) date.

Records were included until the earliest of a patient's transfer out date (if they moved GP surgery), death date (defined using CPRD or the death date from linked-ONS records, described in section 2.6), their 116th birthday (to limit patients with spurious year of birth, as there are no records of a person living to this age in the UK) and the practice's last data collection date.

Once the dataset of illness episodes was established, specific start and end dates within these broad limits were used in each study. These varied between studies and are specified in each chapter of the thesis.

2.3 Defining LRTI, CAP and hospitalisations

The focus of this thesis is on CAP. However, as described previously, less severe LRTI can progress to CAP, and thus it is important to also consider LRTIs as a whole. In order to prevent multiple consultations for the same illness being included as separate incident events, within each patient's data, records for LRTI in general and pneumonia specifically were identified and consolidated into illness episodes. Due to pneumonia being a subset of LRTI, I initially defined LRTI illness episodes, and then defined pneumonia episodes within these LRTI illnesses. For this reason, I describe the methods in the rest of this chapter first for LRTI, and then for pneumonia. These illness episodes

were subsequently assigned as community- or hospital-acquired, depending on their proximity to any prior hospital admission. The section below explains this process in detail.

2.3.1 LRTI and CAP illness-episode definitions

A community-acquired LRTI of any type was defined as the presence of a diagnosis of LRTI in CPRD, or as the primary diagnostic code in the first episode of a hospitalisation (which I assumed to represent the reason for admission) in HES, with no LRTI code in CPRD/HES in the previous 28 days or record of hospitalisation in the previous 14 days. Hospitalisation was defined differently in HES-linked and unlinked data, as outlined in section 2.3.3.

CAP was defined as for LRTI, but using a restricted subset of codes specifically for pneumonia.

2.3.2 LRTI and pneumonia records

The LRTI code list was developed by three clinical epidemiologists, who searched the CPRD Medical Browser for any relevant Read codes (Appendix A). Codes denoting an acute infection of the trachea and the airways below were included, such as those (with or without a specific aetiology) for tracheitis, bronchitis, lower respiratory tract infection, chest infection and pneumonia. As outlined in section 1.1.3, the causative pathogen of LRTI is rarely investigated in general practice and so a broad code list was used. COPD exacerbations which mentioned infection were also included. Codes for potentially non-acute conditions such as chronic bronchitis, tuberculosis and abscesses of the lung and trachea were excluded. Codes for aspiration pneumonitis were also excluded in line with previous studies of CAP.[1]

Within the LRTI code list, codes which stipulated or otherwise indicated pneumonia, (e.g. 'lung consolidation') were identified so that pneumonia could additionally be analysed separately. Codes that stipulated a post-operative infection were flagged so that they could be identified as hospital-acquired, and 'history of pneumonia' codes were also labelled. This process was repeated using the ICD-10 code list for use in the HES data (Appendix A).

2.3.3 Identifying hospitalisations

The focus of this thesis was community-acquired pneumonia, thus hospital-acquired infections were excluded. To do this I needed to identify hospitalisations both in the HES-linked and CPRD stand-alone data. The different recording systems used necessitated different approaches which are described below.

2.3.3.1 Hospitalisation recorded in HES (for HES-linked patients)

HES data were provided from 1st April 1997 to 31st October 2011. Among patients eligible for HES linkage, hospital admission was defined using the HES admission and discharge dates. Occasionally HES spells are not clearly delineated, and in these cases the spells were summarised and the last of the discharge dates used in analyses. These 'multiple spells' took three forms:

- Sequential spells: the patient was admitted and discharged on the same date, and then subsequently readmitted for more than one day.

- Nested spells: a second spell started and ended within a patient's earlier spell.

- Overlapping spells: a second spell started during a patient's first spell, but ended later than the first spell.

The three scenarios are represented in Figure 2-3.

Figure 2-3 Methods of summarising multiple HES hospitalisations that were not clearly delineated

61
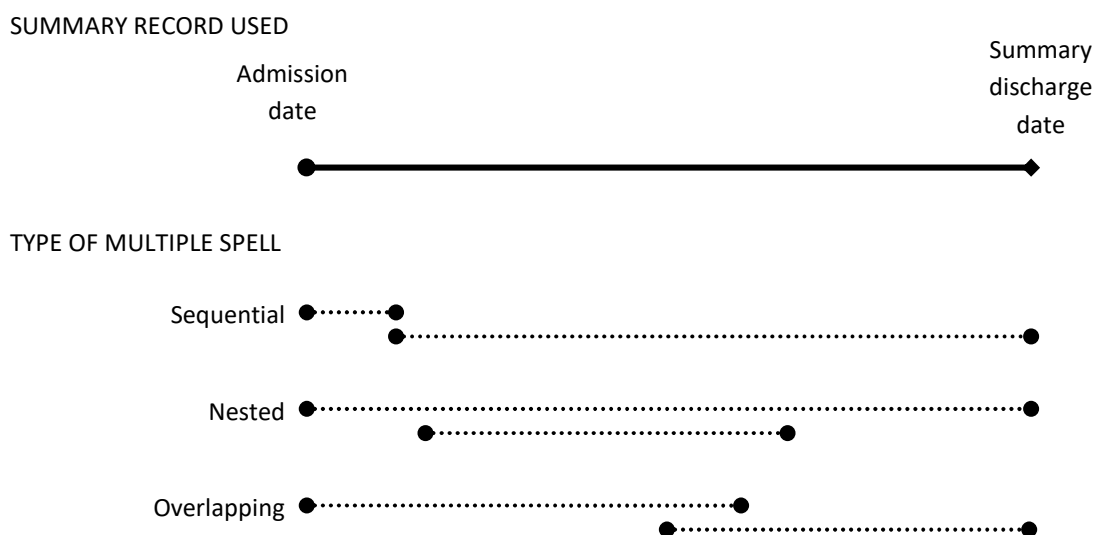
When more than one of these situations arose in parallel, the summary discharge date took the value of the latest discharge date in the sequence. The primary code of the first episode in the sequence was used as the admitting condition – if there were two admissions in a day (as in scenario one, above) the spell with admission and discharge on the same day was deemed the first hospitalisation on that date, and so the primary code of the first episode in that spell used as the admission reason for that cluster of hospitalisations.

For patients who were eligible for HES but who had no hospitalisation records, it was assumed they had not been admitted to hospital within the study period.

2.3.3.2 Hospitalisation records in CPRD (for non HES-linked patients)

There is no single recording system for patients' hospitalisations within CPRD, resulting in hospital admissions, discharges and referrals being recorded in several file types: the clinical, referral, test and consultation files. Within these files a range of codes and coding systems are used, which range from specific information such as 'inpatient' for hospital admission, to more general categories such as 'discharge notice received' which may apply to a discharge from a hospital admission or outpatient department.

Read codes pertaining to hospitalisation in general (rather than specific reasons for hospitalisation, such as hip replacement) were identified in the medical browser by Sara Thomas. Records relating to hospitalisation in the additional coding systems were also identified in the consultation, referral and clinical files by Sara Thomas and myself.

Due to the complexity of the data, records were categorised (by Sara Thomas and myself) into eight levels of hospitalisation; hospitalised as an inpatient, history of hospitalisation, referral for possible hospital admission, A&E record, hospital day-case, non-urgent hospitalisation, hospice referral/admission and 'possibly hospitalised' (for less definitive codes). Records with stronger evidence of inpatient care were prioritised over other categories of hospitalisation recorded on the same day. These fine levels of categorisation enabled different combinations of hospitalisation categories to be used for different scenarios as laid out in sections 2.4.2 and 2.5.1.

2.4 Creation of LRTI and CAP illness episodes

2.4.1 LRTI episode structure

In CPRD, all records for LRTI during the study period were extracted from the clinical, referral and test files. Within these records, pneumonia codes were flagged and multiple LRTI records on one date combined into a single record. In HES all LRTI and pneumonia codes were flagged, along with their position within a hospital episode and spell.

All CPRD LRTI codes and the HES records which had an LRTI code in the primary code of the first episode of a spell were then used to derive the incident date of an episode of illness. Patients may consult multiple times for an ongoing LRTI, and so LRTI records within 28 days of a prior LRTI consultation were considered a continuation of the previous episode. The 28 day re-consultation period was used in line with previous research.[35, 57-60] Thus one LRTI episode could contain several consultations, and a new illness was only recorded when there were more than 28 days between two LRTI consultations (Figure 2-4). For example if a patient had consultations for LRTI on 1st January, 3rd January, 17th January, 8th March, and 7th April then three episodes of LRTI were defined with incident dates of 1st January , 8th March, and 7th April.

Figure 2-4 Defining incident LRTI consultations

KEY

↓ LRTI consultation

↓ Incident LRTI consultation

Days between consultations
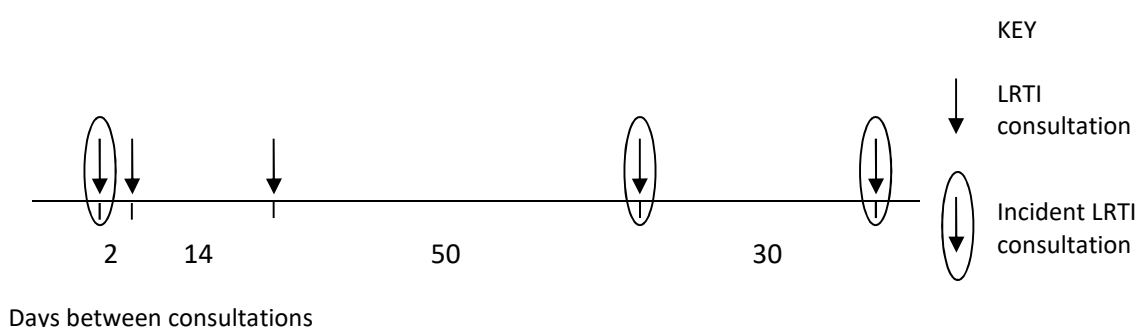
Pneumonia episodes could start on the LRTI incident date (i.e. the patient presented with pneumonia) or later within an LRTI episode as long as there was no hospitalisation record between the LRTI and pneumonia codes (as the pneumonia may then have been hospital-acquired, discussed further in section 2.4.2). As for LRTI as a whole, pneumonia

codes within 28 days of each other were considered part of the same episode (Figure 2-5).

Figure 2-5 Defining incident pneumonia consultations
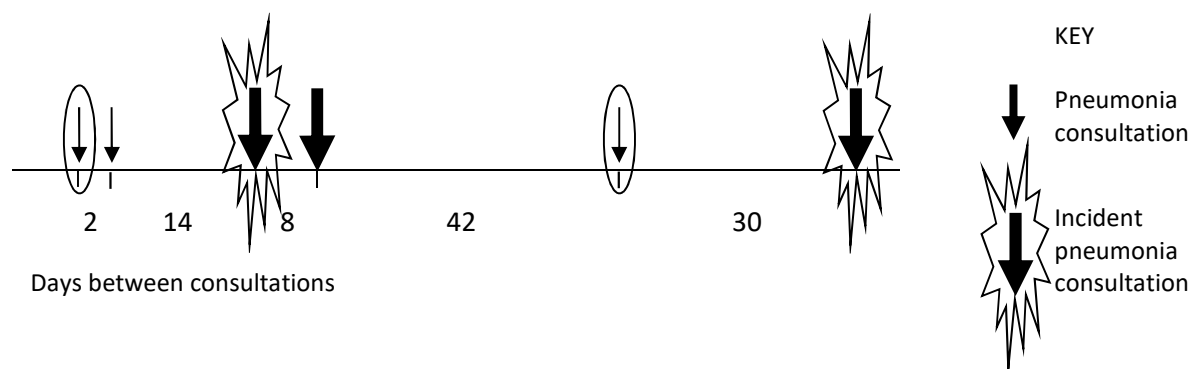
2.4.2 Defining community-acquired episodes

In order to minimise the number of hospital-acquired cases included in the analysis, LRTI cases that occurred ≤14 days after a patient was discharged from hospital were considered potentially hospital-acquired (Figure 2-6). This period was extended to 28 days if the previous hospitalisation included an LRTI or pneumonia code (in keeping with the LRTI episode structure above). Records denoting hospital discharge were by necessity defined differently in CPRD-only and HES-linked data. In CPRD-only data hospitalisation records suggesting the patient had been in hospital for at least one day were used (hospitalisation categories; 'inpatient', 'day-case', 'non-urgent' or 'hospice'), whereas in HES-linked data the date of hospital discharge was used.

When LRTI episodes included a hospitalisation which was followed within 14 days by a pneumonia code, the LRTI was included but the pneumonia was considered hospital-acquired. This was due to an inability to ascertain whether the pneumonia infection was a worsening of the LRTI, or a hospital-acquired infection.

Figure 2-6 Defining potentially hospital-acquired LRTI

KEY

⌂ Hospitalisation record

✕ Potentially hospital-acquired LRTI

Days between consultations: 2 14 43 7 30
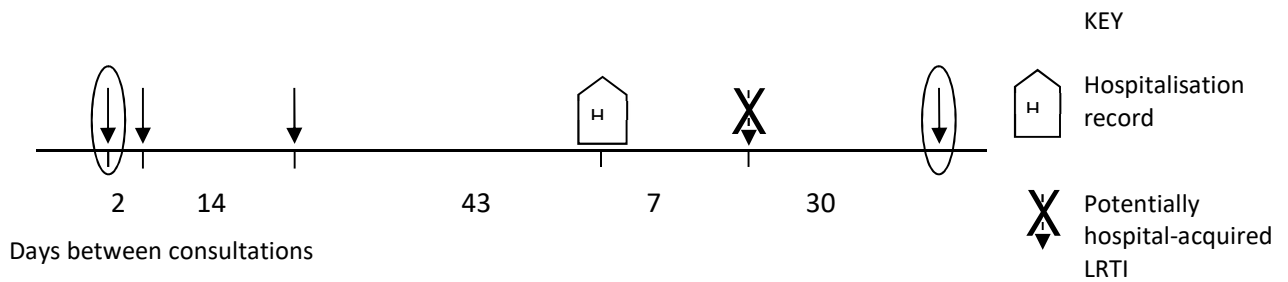
2.4.3 Discrepant recordings for pneumonia and other LRTI in the linked data

A hierarchy was developed to manage CPRD and HES records with the same event date, but conflicting pneumonia or other LRTI codes in the two data sources. CPRD records for pneumonia or other LRTI occurring on the same day as a HES admission with no HES pneumonia or other LRTI code at any point in the hospitalisation were not included as illness episodes as it was reasoned that they may have been misdiagnosed by the GP. Records as above that did have a pneumonia or other LRTI code at some point during the HES spell were included as the patient may have been admitted with a more serious condition (e.g. a stroke) and the co-existing infection not coded until later. When the primary code in a HES record was LRTI and the CPRD code was for pneumonia (and vice versa) the HES diagnosis was used, as diagnostic tools such as radiography are widely available in hospital settings and therefore more likely to have been used in hospital diagnoses.

2.5 Further management of the data to identify hospitalisation post-LRTI, and define person-time at risk of community-acquired infection

2.5.1 Hospitalisation after an LRTI episode

Patients were considered to have been hospitalised in the 28 days after an incident CAP record if they had a HES admission date in that time (linked data, objectives 1, 2 & 3). In the unlinked data, a CPRD hospitalisation record for a non-routine admission (i.e. 'inpatient' or 'hospice') after a CAP episode and more widely for any LRTI was examined briefly in descriptive analyses of patients with incident infections in Chapter 4 (Figure 2-7).

Figure 2-7 Hospitalisation within 28 days of an LRTI episode
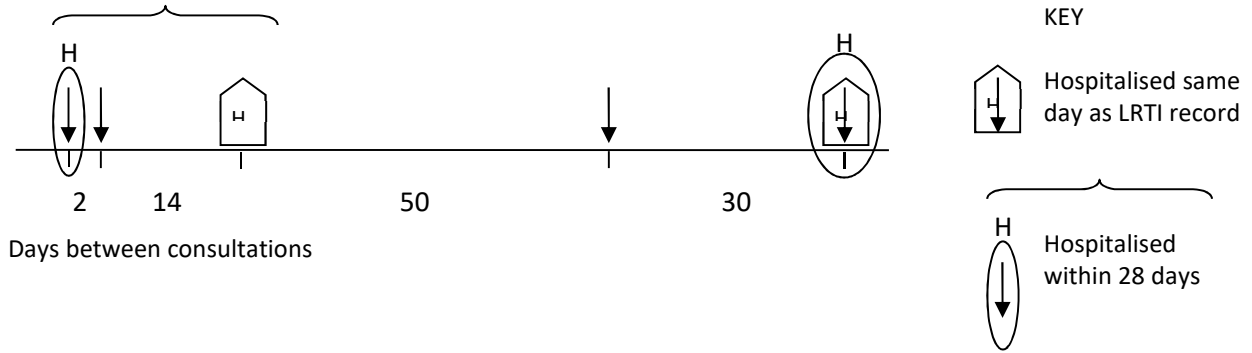
2.5.2   Defining person-time at risk in incidence analyses (Chapters 3, 4 & 5)

2.5.2.1   Time not at risk of community-acquired LRTI

For incidence analyses, patients were deemed not at risk of an incident LRTI during an episode of LRTI (whether community or hospital-acquired) and for 28 days after the last LRTI consultation in the episode (Figure 2-8) (due to the 28 day re-consultation period explained above). Hospital inpatients were not considered at risk of community-acquired infections, and so person-time was excluded from the denominator during all HES-recorded hospitalisations and for the 14 days after discharge. If there was a hospitalisation within an LRTI episode, time at risk of a new LRTI began at the latter of these two times (Figure 2-9, Figure 2-10).

Figure 2-8 Defining person-time not at risk due to an LRTI episode
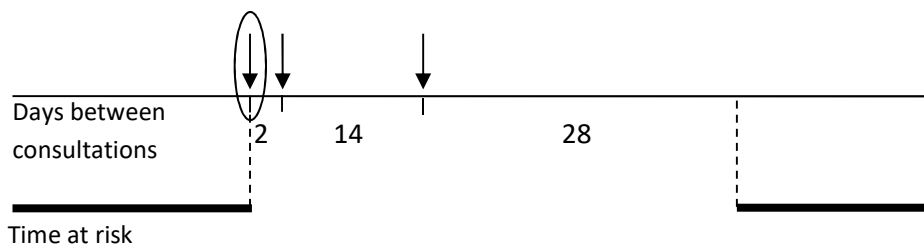
This patient was not at risk over the three LRTI consultations in the episode or for 28 days after their last LRTI record.

Figure 2-9 Defining person-time not at risk due to an LRTI episode or hospital admission
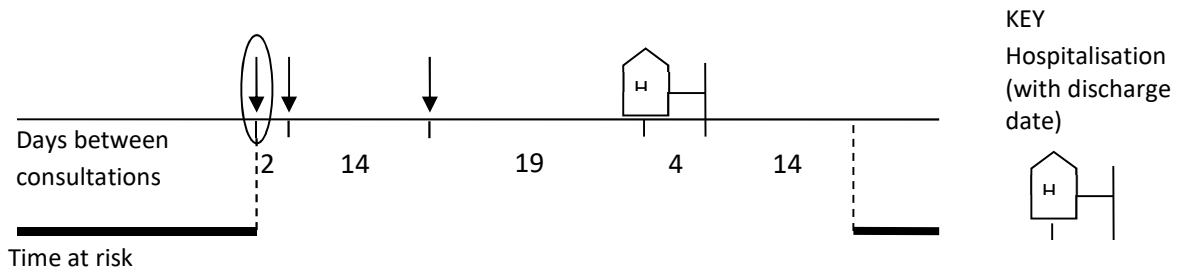
This patient was not at risk over the three LRTI consultations in the episode, the hospitalisation that occurred within the 28 day 'wash out' period or for 14 days after discharge from hospital.

Figure 2-10 Defining person-time not at risk due to a hospital-acquired LRTI episode
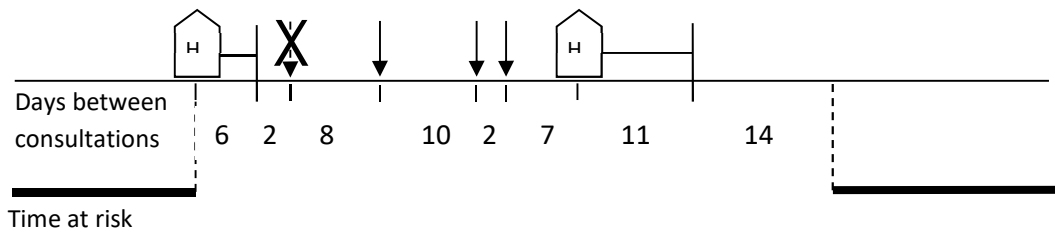
This patient was not at risk during their initial hospitalisation, the subsequent potentially hospital-acquired LRTI episode, the hospitalisation that followed or for 14 days after discharge from hospital.

Time was not excluded for the duration of hospitalisations in the stand-alone CPRD patients' data as no admission or discharge dates were recorded in CPRD. There has been no validation of CPRD hospitalisation codes, nor the timing of recording of these codes (for example these could be recorded on the patient's date of admission, date of discharge or the date the practice received the discharge summary), and so the 14 day post-discharge period was not excluded either when using the stand-alone CPRD data. Further details of how person-time was handled in the stand-alone data, the reasoning behind this decision, and possible implications on estimates from stand-alone data are discussed in section 4.5.

2.5.2.2   Adaptations for time not at risk of CAP

Cases of less severe LRTI remained at risk of pneumonia until they had a pneumonia-coded record or were hospitalised. Cases that developed pneumonia within an LRTI episode (prior to any hospitalisations) had person-time excluded from the denominator from the date of the pneumonia code, until 28 days after the last consultation within that episode (Figure 2-11). LRTI episodes that contained a hospitalisation had person-time excluded from the date of hospitalisation to the latter of 14 days after discharge or 28 days after the end of the episode (Figure 2-12).

Figure 2-11 Defining person-time not at risk of pneumonia
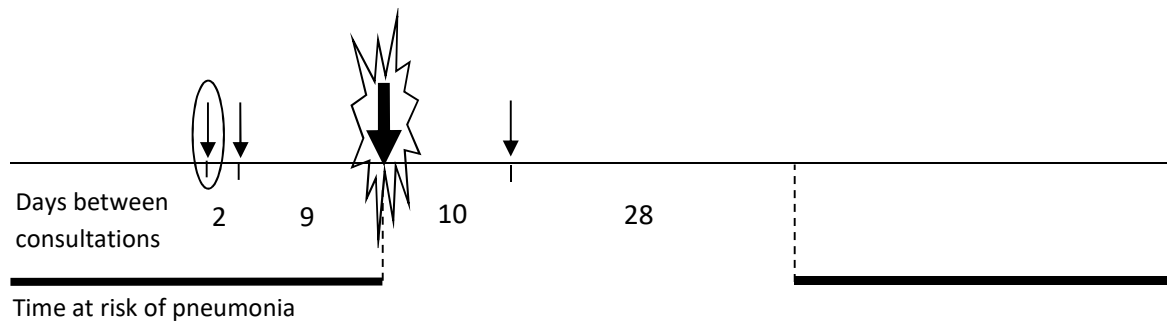
This patient was at risk of pneumonia through their incident LRTI record until the pneumonia code.

Figure 2-12 Defining person-time not at risk of CAP due to hospitalisation
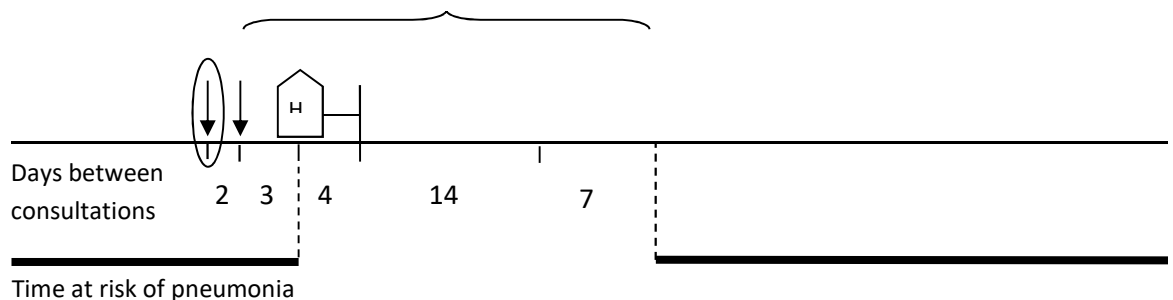
This patient was at risk of pneumonia during the first 5 days of their LRTI, until they were hospitalised. The hospitalisation then made them ineligible for being at risk of community-acquired infection, including pneumonia.

2.6 Defining date of death

Patients' date of death (as described below) was utilised when deriving the date of end of follow-up in all analyses. ONS-provided date of death was further used as an outcome in the analyses of hospitalisation after CAP and prognostic models (Chapter 6 and Chapter 7).

2.6.1 Date of death in ONS unlinked patients

For patients ineligible for linked data, I used the date of death field contained within CPRD. CPRD derived this date from an algorithm they developed which took the earliest of; the transfer out date if reason of transfer was death, the event date of a statement of death Read code, and the event date recorded in the 'death administration' structured data area. If no date of death was recorded the patient was assumed to have survived until the end of their follow-up.

2.6.2 Date of death in ONS linked patients (objectives 2 and 3)

Date of death from linked death certificate data was available for all patients with linked data. In an extremely small number of patients (n=87), more than one death record for a patient was identified by the linkage process. In 40 of the 87 patients the multiple records had the same match quality value (described in section 2.1.4), and so a different mechanism to establish which record to include was required. For reference, I compared the CPRD and ONS dates of death for patients who only had one ONS-linked date of death, and almost all (98%) of these dates were found to be within 28 days of each other. Given the similarities between the criteria used for the match quality scores (see Table 2-1, above) I decided to use the ONS death date that was closest to the CPRD date (when patients had both dates recorded) rather than use the match quality when patients had multiple ONS death record matches. When patients had multiple ONS death records but no CPRD date of death, I used the best quality ONS match.

2.7 Covariates

Several key variables were used in incidence analyses, and a wide range of variables were used in subsequent analyses on hospitalisation (objective 2) and prognostic modelling (objective 3), and these are outlined below.

2.7.1   Year of illness

Time was categorised using the financial year structure of 1ˢᵗ April to 31ˢᵗ March the following calendar year. CAP and other LRTI are caused by pathogens which predominantly circulate in the autumn and winter, and this categorisation ensured that winter peaks of illness were not split across two time periods.

2.7.2   Age

In order to preserve patient anonymity CPRD only supply researchers with patients' year of birth. All patients were given an estimated date of birth of 1ˢᵗ July in their birth year, resulting in a maximum margin of error of six months. Age was categorised into five-year groups from 65 to 89 and then ≥90 years.

2.7.3   Sex

Patients were included in the analyses if their sex was coded as male or female. A small number of patients with an indeterminate gender were excluded from the studies, as there were too few of them to be included in multivariable analyses.

2.7.4   Region

Regions in England were defined using the ten Strategic Health Authorities (SHAs) in place at the start of this study, and additionally Wales, Scotland and Northern Ireland. Smaller level geographical groupings are not available within CPRD in order to preserve patient anonymity.

2.7.5   Additional co-variables used in objectives 2 and 3

The analyses involved in objective 2 and 3 required the definition of a range of additional co-variables that were thought to be potentially associated with either hospitalisation or death after a CAP episode. These co-morbidities, medications, vaccinations, and lifestyle and frailty factors are defined in detail in section 6.4 but are summarised briefly here.

2.7.5.1 Co-morbidities

Twenty two co-morbidities of interest were defined. All the constituent co-morbidities in the Charlson co-morbidity index were included.[68]   In addition I considered additional cardiac, neurological, and immune disorders that could increase the risk of severe CAP or mortality from non-respiratory causes after CAP.

Among the Charlson co-morbidities, lung disease is an important risk factor for severe CAP, as it results in injury to the respiratory tract, reducing the effectiveness and efficiency of the mucociliary escalator and creating pockets of environment which are more accommodating of some species of bacteria.[4]   Several of the Charlson co-morbidities also result in patients having an increased risk of severe infection due to a weakened immune response (the importance of which as a defence against CAP was outlined in section 1.1.2). I added to the Charlson co-morbidities additional disorders of the immune mechanism, such as aplastic anaemia; the full list is given in section 6.4.1.3. Similarly, I considered neurological co-morbidities such as Parkinson's disease and multiple sclerosis, which can result in damage to the mechanical barriers to infection, such as a reduced cough reflex or ability to swallow. As highlighted in section 1.3.2.4, systemic infection can increase the short-term risk of acute cardiovascular events, particularly among individuals with existing cardiovascular disease. The Charlson co-morbidities include cardiovascular conditions such as myocardial infarction and cerebrovascular disease; I added ischaemic heart disease to this list.

2.7.5.2 Frailty factors

A range of frailty factors were defined, based on those in the frailty index described in section 1.2.1.1. These were the presence of; mobility issues, a patient's inability to care for themselves (self-care), a bedsore or ulcer, tiredness, anxiety or depression, low weight or poor nutrition, incontinence or catheterisation, a history of falling, visual impairment and recent need of a carer. Additionally their living arrangements (such as living alone, in sheltered accommodation or residential care) were recorded.

2.7.5.3 Medications and vaccinations

Medications used to treat CAP/other LRTI (antibiotics) as well as those that may increase the risk of severe outcomes after CAP such as inhaled corticosteroids, oral steroids, immunosuppressants other than steroids, and those that may be protective such as

statins were investigated. Influenza and pneumococcal vaccination status were also included to examine their effect on the severity of infection.

2.7.5.4 Lifestyle factors

Patients' smoking status (current, ex- or never smoker) and if they had a history of excess alcohol consumption were explored as these factors can also increase a patient's risk of CAP. A major long-term effect of cigarette smoking is a reduction in mucociliary transport, resulting in an increased susceptibility to respiratory tract infections.[6] Alcohol has immunosuppressive effects that can last for several months, and drinking can also increase the risk of aspiration pneumonia.

2.8 Statistical methods

The statistical methods used in each analysis are described in detail in individual chapters. Common statistical methodology, and the analyses themselves are outlined below.

2.8.1 Clustering of illness episodes

Due to the high incidence of CAP and other LRTI within the older population under study, some patients experienced more than one episode of illness in a year. These episodes are said to be clustered or correlated within a patient. Standard model fitting methods assume that all observations (in this case CAP episodes) are independent of one another with respect to exposure status. In clustered data this assumption may not be true, and a correction needs to be made for any within-subject correlations. Ignoring the violation of the independence assumption leads to standard errors (SEs) which are too small, confidence intervals around exposure effect estimates which are too narrow and p-values which are too small.[113] This increases the risk of Type 1 error, in which an association between an outcome and exposure is reported but is actually due to chance. The different methods used in this thesis for adapting models to the clustered data and taking the clustering into account, resulting in valid standard errors and p-values, are outlined below.

2.8.1.1 Robust standard errors

When parameters such as rate ratios or odds ratios are estimated using the likelihood approach, their SEs are estimated using the variability assumed by the underlying statistical model. When data are clustered, the precise probability model underlying the likelihood may not be correct due to the violation of the independence assumption. Robust standard errors take a different approach and are estimated using the variability in the data itself, which is measured by the residuals (the difference between each outcome and the predicted value of the outcome from the regression model).[113] Only the SEs are affected and the parameter estimates (odds ratios (ORs), rate ratios (RRs) or hazard ratios (HR)) do not differ from those computed without robust SEs. One adaptation needed when using this method is the use of Wald tests rather than likelihood ratio tests (LRTs) for hypothesis testing, as likelihood is not affected by robust SEs and so LRTs do not take account of the clustering.[113]

I used this simplest approach to clustering in Chapter 7, where Cox regression with robust SEs was used to develop a series of prognostic models to predict mortality risk after CAP hospital discharge.

2.8.1.2 Generalised estimating equations, or population averaged approach

Generalised estimating equations (GEE) extend the approach of robust SEs while also taking account of the correlations when calculating the effect estimates. The assumed correlations of the residuals are known as the working correlation matrix, and the appropriate matrix is specified when analysing the data. The most commonly used is the 'exchangeable' correlation matrix which assumes that any two observations within a cluster are equally correlated.[114] The working correlation matrix is then used to adjust the model parameter estimates and SEs for the correlation in the data (in effect by giving relatively more weight to smaller clusters). The regression coefficient produced is the average of the individual regression lines for each cluster and is interpreted as the mean effect across the population, which is why GEE is also known as a population averaged approach.[114] As when using the robust standard errors approach, likelihood ratio tests cannot be performed, and Wald tests should be used instead.

In Chapter 6 I used population averaged logistic regression to estimate the percentage of patients with CAP who were hospitalised within 28 days, between 1998 and 2011.

2.8.1.3 Random effects and multilevel models, or subject-specific approach

Random effects models take a different approach to the issue of clustered data. Rather than treating the correlation between outcomes as a nuisance which can be adjusted for (as in GEE), random effects models explicitly include the clustering in the model. Each cluster is assumed to have been randomly drawn from the population, and individual clusters are allowed to vary from the regression model, producing cluster level residuals (the difference between the predicted value of the outcome from the cluster-specific regression and the predicted value of the outcome from the overall regression model). These residuals are treated as a random unobserved variable distributed with a mean of zero, and so only the standard deviation of the distribution (the group-effect) needs to be estimated.[114]

The inclusion of the cluster level group-effect in the model results in changes to both the log effect estimate and the SEs produced by the model – although the change to the log effect estimate is smaller than that of the SEs. Comparison of the log-likelihood of a random effects model with a standard regression model can be undertaken using a LRT to assess whether the addition of the random effects term provides a better fit for the data. Estimates produced from a random effects model are interpreted differently to those from population averaged models above; instead of representing the mean effect of the variable of interest across the population, they represent the effect of the variable on the mean subject or cluster. Thus the effect estimates produced by multilevel models are said to be subject- or cluster-specific.

Poisson regression with random effects was used to calculate the incidence of CAP in two chapters. In Chapter 4, I estimated the burden of CAP (and LRTI more broadly) among older adults in the UK between 1997 and 2011. Subsequently, in Chapter 5 I compared the incidence estimates of CAP derived from stand-alone CPRD data to those from linked CPRD-HES in order to assess the added value of using linked-data when estimating CAP incidence.

Multilevel models can be extended to have as many levels as appropriate. Each additional level of clustering added to the model results in a new cluster-specific group effect being added. The model testing process above can be extended to test three levels of clustering against two, and so on to check that the appropriate number of levels

of clustering are being included. In Chapter 6 I used multilevel logistic regression, and investigated whether in addition to CAP episodes being clustered within patients, treatment decisions were clustered within general practices, which is an example of a three-level model. The multilevel model was then used to investigate risk factors for hospitalisation in the 28 days after a CAP diagnosis.

2.8.1.4 Comparison of results using the population-averaged and multilevel approaches

Coefficients produced by models using a subject-specific approach are usually larger than those produced using a population-averaged approach. However, the ratio of the parameter estimate to its standard error is generally similar for the two models, resulting in equivalent results from significance tests.[114] Both approaches are equally valid, and the most appropriate method to answer the research question should be chosen.

I have mainly used subject-specific approaches (random effects and multilevel models) for the work in this thesis, as these are considered the most satisfactory (due to their being based on a full probability model for the data),[113] and I was generally interested in establishing effects at the individual CAP episode level.

In order to investigate the trend in hospitalisation following CAP over time (Chapter 6), a population-averaged approach was more appropriate as I wanted to examine this trend over the population. I therefore converted the results from the multilevel risk factor model into population-averaged results for this part of the analysis.

2.8.2 Predictive and causal model building

It is important to choose the appropriate modelling strategy for the type of analysis being performed. In general, strategies can be categorised as predictive or causal and these are used to answer different types of question, as outlined below.

2.8.2.1 Predictive modelling

Predictive modelling is used when trying to predict an outcome in a given set of circumstances. The aim is to use the minimum number of variables to explain most of the variation in the outcome.[115] The number of variables in the model is particularly

important in prognostic scores, as a balance must be struck between including enough variables to predict the outcome of interest as accurately as possible but not including so many as to make the score unwieldly or underpowered, thus limiting its real-world use. Commonly, variables are 'selected' into a regression model using a stepwise process; variables are either added to a model (forwards stepwise selection) or removed from a model initially containing all variables of interest (backwards stepwise selection). Decisions on whether the change should be retained are made by comparing the models with and without the variable of interest (using a predefined cut-off of the p-value from either a LRT or Wald test as appropriate) to assess the fit of the new model compared to the old. The model with the better fit is then used as the comparison model, and a new variable added (forward) or removed (backward) until all variables of interest have been considered. The end result is a parsimonious model. A predictive modelling approach was used in Chapter 7 to develop the prognostic models to predict mortality risk after CAP hospital discharge.

2.8.2.2 Causal modelling

An alternative method is causal (or explanatory) modelling, which aims to estimate the effect of the risk factors of interest, while also controlling for any potential confounders. The model is in effect being used to test a causal explanation, and any variable which may be a risk factor or confounder should be included.[116] If the study population is large enough, all potential risk factors and confounders can be included in the model as there is no need to aim for parsimony. However, if there are a large number of variables of interest, this may result in unstable or exaggerated estimates due to sparsity of data or multicollinearity. In such cases restricting the number of variables included in the model should be considered. A further important consideration is the hierarchical relationship between the factors of interest and whether the effects of a factor (or group of factors) are direct, or mediated through other factors. If it is possible that some factors are mediators, models should be built sequentially in order to be able to assess the effects of distal variables without adjusting for the more proximate mediators. The mediators are then added to the model in order of increasing proximity to the outcome.[116] The estimates produced for the mediators show their effect on the outcome, adjusted for the effects of the more distal variables, and the effects of the distal variables in this model represent those not mediated through the new group of

variables. A hierarchical framework of how potential risk factors are related should be considered before analysis.

A causal modelling approach was used for the analyses of risk factors for hospitalisation after CAP (presented in Chapter 6). Mediation analysis was used to investigate the effect of distal and proximate risk factors on hospitalisation, as well as the extent which trends in hospitalisation after CAP over time were explained by these factors.

2.9 Power calculations

A major advantage of the use of CPRD and linked HES is the very large size of the datasets. This enables precise estimation of incidence, and exploration of a wide range of potential risk factors for hospitalisation and death following CAP. Due to the large size, formal power calculations were not performed for the majority of analyses. The sample size requirements when developing prognostic scores are discussed in Chapter 7.

2.10 Ethics approval

Ethics approval for the study was provided by the London School of Hygiene and Tropical Medicine ethics committee (6116), and the Independent Scientific Advisory Committee (11_033).

Chapter 3 Determining start of follow-up for analyses using incident episodes of community-acquired LRTI and pneumonia in older individuals.

In the previous chapter I described the internal data quality checks that CPRD undertake to ensure that records from each practice are fit for research use (section 2.1.1.1). The date at which the records meet this standard is known as the 'up to standard' date (UTS). While the checks performed by CPRD protect against large-scale errors, individual patient-level data quality issues also need to be considered when assigning patients' start of follow-up. In this chapter I outline these issues, and report an analysis undertaken to minimise these problems.

3.1 Background

3.1.1 Incorrect date recording for historical illnesses in general practice

There can be inconsistencies in the recording of illness in the first few months after patient registration, for example when recording patients' medical histories. Upon registering with a new GP, patients provide their medical history and report any previous instances of severe or chronic illness. This information is commonly obtained during a new-patient health check (which may not occur immediately after patient registration) or during a consultation with a GP. GPs are able to record historical diagnoses on the date the diagnosis was originally made (or a close approximation to this date), but these diagnoses are also sometimes recorded on the date of the current consultation. Reasons for this include patients not remembering the date of onset of a previous illness, GPs noting elsewhere that diagnoses are historical (for example in the freetext field, which is not routinely provided with CPRD data), or due to time pressures in a short consultation preventing information from being entered on a number of dates. Historical diagnoses coded using the current appointment date incorrectly appear to researchers to be a new illness, and if included in analyses lead to overestimation of the incidence of disease.

However, this over-reporting only affects a subgroup of the CPRD population. Patients who registered before the practice became 'up to standard' do not begin follow-up in research studies until on or after the UTS date, and thus any historical illnesses reported

when the patient first registers with the practice will be automatically excluded from incidence analyses (Figure 3-1). Conversely, patients who registered after their practice became UTS do not have this exclusion period if the UTS date is taken as the start of follow-up, and so pose a problem for researchers – are illnesses recorded early in a new patient's follow-up incident events or prevalent/historical episodes that have been recorded retrospectively and which need to be excluded from incidence analyses?

Figure 3-1 Inclusion of past medical history records in analysis, by whether a patient registers with a GP pre-UTS or post-UTS
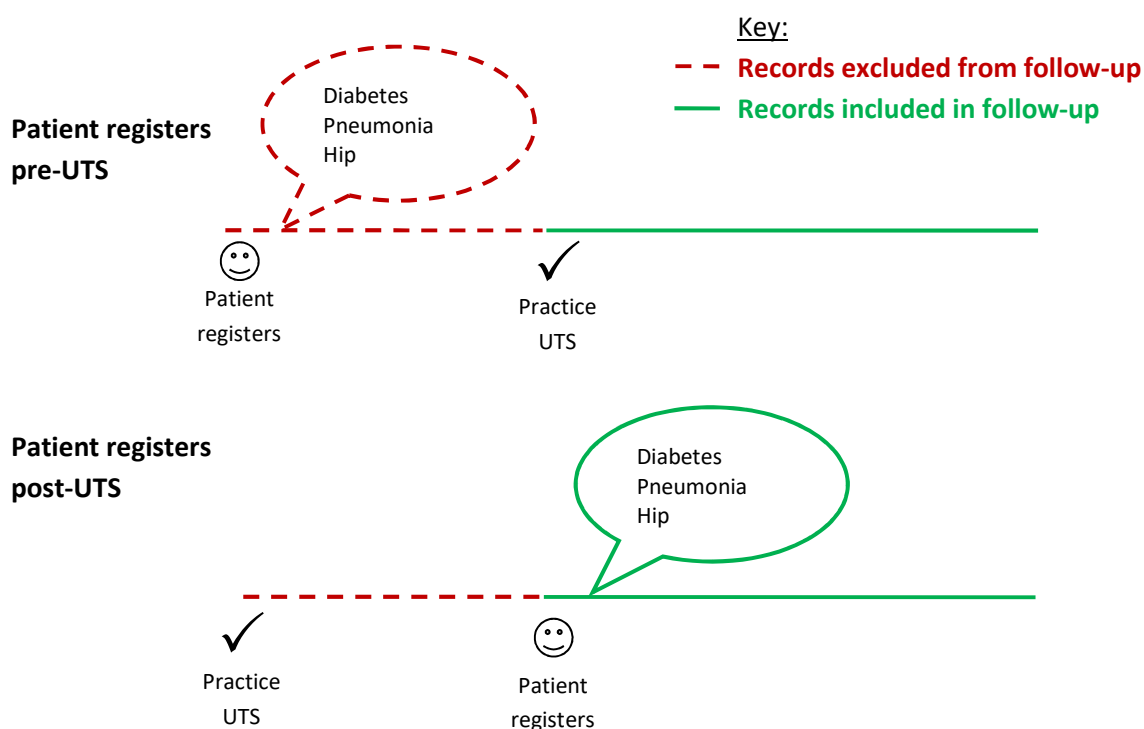
3.1.2 Assessment of the length of the over-recording period

The relationship between time since registration and measured incidence rates in CPRD data has previously been examined for several chronic and acute conditions (including pneumonia) by Lewis et al.[117] Patients were stratified by whether they had registered with the practice pre- or post- the practice's UTS date. Within these groups incidence rates were calculated for each condition over the first three years of follow-up. To quantify when the period of over-reporting ended, incidence rates in three-month periods for the first year were each compared to the rate in months 13-36 and incidence rate ratios (IRR) produced. The rate in months 13-36 was considered the baseline, as by this point it was assumed any over-reporting would have ended. IRRs greater than 1.2 were considered to be overestimations. It was shown that rates were uniformly

higher at the start of follow-up in the group of patients who registered with a practice post-UTS. The length of this overestimation varied both by disease type (acute or chronic) and to a lesser extent by different diseases of the same type. Chronic relapsing diseases appeared to have incidence rates which were overestimated for 12 months or longer at the start of follow-up, while neoplastic disease incidence was overestimated in the first 6 months. Acute conditions varied from over-reporting in the first three months (myocardial infarction, wrist fracture, deep vein thrombosis/pulmonary embolism), to six months (urinary tract infections) and nine months (pneumonia).

3.1.3 Tailoring of these analyses to specific populations

The Lewis analysis provided an important and interesting overview of the problem of over-reporting in the first months after registration within CPRD data. In order to make as accurate choice of exclusion period as possible, the authors suggested that similar analyses were undertaken by researchers on their diagnoses of interest, to further refine the methodology Lewis developed. However, in addition to varying by disease type, the period of over-reporting may vary by other factors such as age, sex or year of registration. A better understanding of the heterogeneity of over-reporting by these factors would enable further refinement of the follow-up exclusion period needed. These additional factors are discussed in more detail below.

3.1.3.1 Additional factors which may influence the period of over-reporting

Age

Lewis' investigation included patients of all ages, which is a heterogeneous population. The post-UTS group is made up of patients who have recently changed GP, and the reasons behind this may differ between younger and older adults. If different age groups also report their medical history to their GPs at different times, this would affect the exclusion period required. For example, declining health or an inability to live independently may lead to older patients to move into supported or residential accommodation, and consequently also to change GP. Patient behaviour upon registration may also differ with age; underlying illness and the need for repeat prescriptions may motivate older patients to provide their medical history to a new GP in a timelier manner than younger, healthier patients (and thus require a smaller exclusion period). Finally, older patients may visit the GP more frequently than the

general population which may affect the timeframe of any over-reporting of previous illness.

Sex

Men have been shown to consult their GP slightly less than women.[118] If this extends to taking longer to make initial contact with their GP, it could result in different lengths of over-reporting periods between the sexes.

Year of start of follow-up

The analysis by Lewis included records from 1987 to 2003. Health service provision and strategy has changed since the first practices became up to standard in 1987, and electronic health record use has proliferated from being rare to the norm over this period. More recently the introduction of QOF in 2004 has led to an increase in recording of QOF-incentivised conditions and a more complete reporting of their detail.[109] While CAP is not included in the QOF scheme, other common co-morbidities in older adults which are risk factors for CAP (such as diabetes and COPD) are.[107] The proliferation of electronic health records over time and introduction of QOF may have had an effect on recording of medical histories, and for investigators using a restricted time-period of records it may be possible to further improve the exclusion window needed post-UTS.

Health checks

New-patient health checks can be noted in patients' records using a variety of Read codes, as can additional health checks which are offered to patients at certain ages, for example the over-75 health check. Excluding pneumonia records on the same date as these checks could further refine the window of record exclusion, resulting in the incidence rate returning to the baseline more quickly.

Person-time at risk

Lewis calculated incidence over three month windows, excluding participants from contributing person-time from the date of the illness and, for acute conditions, restarting contributions in the second three-month period following the event date. Defining episodes of illness using start and end dates, and excluding person-time at risk

during this period is a more exact way of determining who should be contributing to the analysis at any one time.

Width of incidence windows

Dividing follow-up into smaller incidence windows of four weeks (rather than three months) enables more accurate evaluation of when the incidence reaches baseline, and thus allows a more precise period of time to be removed from start of follow-up.

Repeat episodes

Lewis included a maximum of one acute illness per six months. Inclusion of more frequent repeat episodes of infections is possible when using a smaller incidence window, and is particularly important for common seasonal diseases.

3.1.4   Aims

This analysis aimed to refine the exclusion period required at the beginning of patient records when examining CAP episodes in patients who registered with a GP after the UTS date. I utilised the episode structure for CAP events outlined in section 2.4 to enable accurate inclusion of repeat events. Incidence was calculated over 28 day windows (rather than Lewis' original 3-month window) to increase the precision of the estimate of period of exclusion. I investigated the length of time it took for the incidence of CAP to return to baseline in older patients who registered pre- and post-UTS, further stratified by age, sex, and year of start of follow-up. Additionally, records for CAP on the date of health checks were excluded to examine their contribution to over-reporting.

3.2   Methods

3.2.1   Study population and follow-up period

As this analysis was concerned with over-reporting in GP records, the data used were limited to CPRD stand-alone data. In order to investigate whether over-reporting had changed over time, records from the beginning of CPRD in 1987 were eligible for inclusion in the study.

Patients were classified into two groups – those who registered before the practice's UTS date (the 'pre-UTS' group), and those who registered after or on the practice's UTS

date ('post-UTS'). Patient follow-up began at the later of the registration and UTS dates. Patients who were younger than 65 years of age when they started follow-up were excluded, as it was specifically the level of over-reporting within the older population that was of interest. (In the pre-UTS group, patients aged <65 when they registered with their GP were excluded; in the post-UTS group patients aged <65 when the practice became UTS were excluded). Patients aged ≥65 years who did not have a three year period between registering and UTS (i.e. those who registered with their GP less than three years before the practice became UTS) were not included in this analysis, as they did not clearly belong in either group (those registered pre- or post-UTS, Figure 3-2).

Follow-up ended at the earliest of the patient's date of death, the date they transferred out of the GP practice, three years after the start of follow-up and the date of last data collection from the practice.

Figure 3-2 Patients aged ≥65 years included and excluded from these analyses

Key:
Records excluded from analyses
Records included in analyses

Included: Pre-UTS group

Patient registers — ≥3 years — Practice UTS ✓

Included: Post-UTS group

Practice UTS ✓ — Any time period — Patient registers

Excluded from these analyses

Patient registers — <3 years — Practice UTS ✓
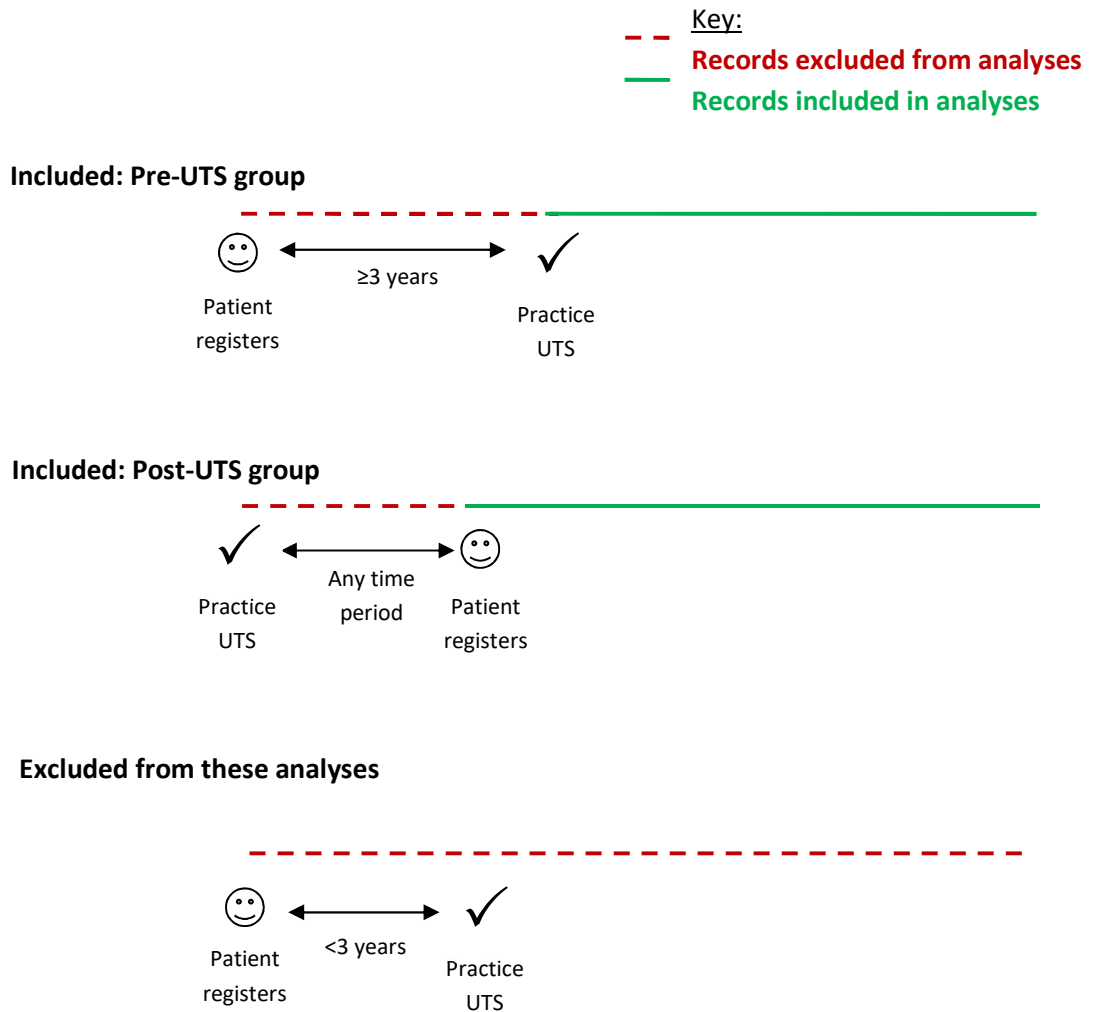
3.2.2 Data management and structure

Within the maximum three years of follow-up for this analysis, all pneumonia (or wider LRTI) codes were ordered into illness episodes as defined for stand-alone CPRD data in section 2.4.1. These episodes were then deemed community- or hospital-acquired depending on whether there was a CPRD code for hospitalisation in the 14 days prior to the incident date (section 2.4.2). Hospital-acquired episodes were not included in subsequent analyses.

Time was divided into 28 day intervals from the start of follow-up. Due to the illness-episode structure used (with episodes at least 29 days apart), it was only possible for patients to have a maximum of one event per 28 day period.

3.2.3 Covariates of interest

Analyses within the pre and post-UTS groups were further stratified by several factors.

3.2.3.1 Age and sex

Age was categorised into 5-year groups from 65 to 84 and ≥85 years. Sex was classified as male or female.

3.2.3.2 Year of start of follow-up

To investigate whether over-reporting had changed over time, patients were divided into three groups depending on their start of follow-up: 9th September 1987 to 31st March 1997, 1st April 1997 to 31st March 2004, 1st April 2004 to 26th August 2011 (the end of data collection). Cut points were chosen based on dates useful to researchers who use CPRD: the start of HES-CPRD linkage (1st April 1997) and the introduction of QOF (1st April 2004).

Additionally, the month of start of follow-up was examined for patients in both the pre- and post-UTS groups, to see if any seasonal trends were evident.

3.2.3.3 Health checks

A Read code list for health checks was devised to explore whether CAP records coded on the day of a health check or medical screen could be identified and excluded. Health checks were divided into two sets of codes; a 'strict' list, which included items offered to newly registered patients, such as new-patient checks, screens and histories, and a 'general' list which also included other routine screens and checks such as geriatric screening, retirement and insurance medical exams.

All records containing a strict or general health check code within the first year of follow-up were flagged. If a health check was on the same date as the incident code of a CAP episode, that episode was assumed to be historic and was excluded from the analysis.

3.2.4 Statistical analyses

The baseline characteristics of the pre- and post-UTS groups were examined, including age at start of follow-up, sex, year-group and month of start of follow-up.

Yearly rates of CAP were calculated for each 28-day period of the three years of follow-up, stratified by pre/post-UTS groups using Equation 3-1. Patients could not contribute person-time at risk to the analyses during any CAP episode, whether community or hospital-acquired. Person-time was excluded from the date the episode started until 29 days after the last CAP code within an episode (as outlined in section 2.5.2).

Equation 3-1 Yearly incidence rate for CAP

$$\text{Yearly rate for 28 day period} = \frac{\text{Number of incident CAP diagnoses}}{\text{Person time at risk (days)}} \text{ x } 365.25$$

Rates were further stratified by age, sex, year group and health checks, and the pre- and post-UTS rates compared graphically.

Incidence rate ratios (IRRs) were calculated using Poisson regression, comparing rates in each of the 13-four week periods of the first year of follow-up to those of weeks 53-156 combined. This latter period was used to estimate the baseline incidence rate and (as in Lewis,[117]) the IRRs were deemed to be 'at baseline' when they were within 20% of the baseline rate (IRR<1.2). As patients could not have more than one CAP per 28-day incidence window, it was not necessary to take clustering of illness-episodes within patients into account in this analysis.

3.2.5   All LRTI

The primary analyses of interest were on the effect of over-reporting on CAP estimates. However, I also repeated the analyses for all LRTI (using the methods described above) in order to better understand the exclusion period required for a generally less severe group of infections. As CAP was the primary analysis which informed future work, results for LRTI are presented briefly with supplementary results provided in Appendix B.

3.3   Results

Within CPRD there were 1,149,386 patients aged ≥65 when they joined their practice or on the date when their practice was deemed UTS. Of these, 9.7% (111,103) were excluded from the analysis due to not having three years of follow-up before the

practice became UTS. The remaining 1,038,283 patients were included in the analysis, 55.5% of which were in the pre-UTS group and 44.5% in the post-UTS group (Table 3-1).

Table 3-1 Baseline characteristics of the study population

	Registered pre-UTS n (%)	Registered post-UTS n (%)
Number of patients	576175 (55.5)	462108 (44.5)
Sex		
Male	240144 (41.7)	178607 (38.7)
Female	336031 (58.3)	283501 (61.3)
Age at start of follow-up		
Median (IQR)	74 (69-80)	76 (70-84)
65-69	162078 (28.1)	107157 (23.2)
70-74	144976 (25.2)	91503 (19.8)
75-79	121739 (21.1)	82963 (18.0)
80-84	80629 (14.0)	76449 (16.5)
85+	66753 (11.6)	104036 (22.5)
Year follow-up started		
1987 - 1996	253658 (44.0)	116197 (25.1)
1997 - 2003	280766 (48.7)	157026 (34.0)
2004 - 2011	41751 (7.2)	188885 (40.9)
Month follow-up started		
January	83071 (14.4)	39224 (8.5)
February	23670 (4.1)	33601 (7.3)
March	51116 (8.9)	36246 (7.8)
April	40472 (7.0)	39431 (8.5)
May	35099 (6.1)	36313 (7.9)
June	42406 (7.4)	38316 (8.3)
July	40130 (7.0)	45431 (9.8)
August	29332 (5.1)	37826 (8.2)
September	84418 (14.7)	42573 (9.2)
October	56957 (9.9)	43936 (9.5)
November	52354 (9.1)	36499 (7.9)
December	37150 (6.4)	32712 (7.1)

The patients who registered post-UTS were slightly older than those registered pre-UTS (median ages 76 (IQR: 70-84) and 74 (IQR: 69-80) respectively). The largest age category in both pre- and post-UTS groups was 65-69 year olds (28.1% and 23.2% respectively). Within the pre-UTS group, the proportion of patients decreased as age increased; this trend was also seen in the post-UTS group until the ≥85 year category which accounted for 22.5% of post-UTS patients. Over 90% of those who registered pre-UTS did so before 2004, while 40% of patients who registered post-UTS did so after 2004.

The month of start of follow-up for patients in the post-UTS group (for whom follow-up started at their registration date) was relatively evenly distributed, while in the pre-UTS group (who started follow-up on the UTS date) there were peaks in January (14.4%) and September (14.7%) representing times when large numbers of practices join the CPRD.

3.3.1 Period of over-reporting of CAP incidence and unstratified IRRs

Before any stratification, CAP incidence was higher in those registered post-UTS compared to those registered pre-UTS throughout the three years of follow-up.

Incidence of CAP in the pre-UTS group was relatively stable over follow-up. The incidence of CAP among post-UTS patients reached a plateau at around window 8 (weeks 29-32, Figure 3-3) and the IRRs were within 20% of the baseline (weeks 53-156 combined) in all age groups and year groupings by week 29 (Table 3-2).

Figure 3-3 Comparison of CAP incidence in those who registered pre and post-UTS over the first three years of follow-up
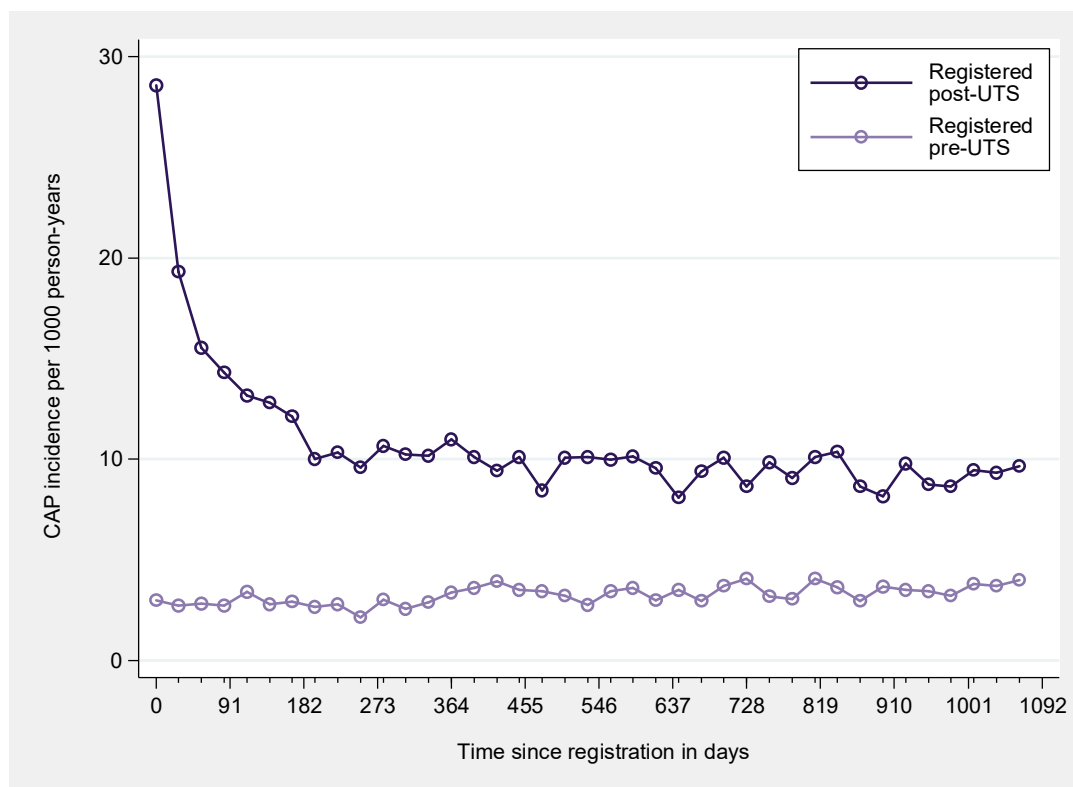
(The unlabelled markers on the x-axis represent each 28 day period, time is days is also provided for ease of interpretation).

Table 3-2 Incidence rate ratios (IRRs) comparing incidence of CAP in four weekly periods of the first year with the incidence in the second and third years of follow-up, for patients who registered post-UTS, stratified by age, sex and year of start of follow-up

Time (weeks)	Unstratified IRR (95%CI)	Age at start of follow-up IRR (95%CI)					Sex IRR (95%CI)		Year of start of follow-up IRR (95%CI)		
		65-69	70-74	75-79	80-84	85+	Male	Female	1987 - 1996	1997 - 2003	2004 - 2011
1-4	3 (2.44-3.68)	3.01 (1.45-6.28)	3.02 (1.62-5.63)	2.17 (1.43-3.29)	2.51 (1.73-3.63)	2.47 (1.93-3.17)	3.58 (2.49-5.16)	2.62 (2.06-3.35)	0.76 (0.59-0.98)	3.29 (2.37-4.58)	3.27 (2.36-4.54)
5-8	2.03 (1.71-2.4)	1.48 (0.9-2.45)	1.54 (1-2.38)	1.62 (1.13-2.32)	1.83 (1.33-2.51)	1.79 (1.45-2.22)	2.1 (1.6-2.76)	1.98 (1.6-2.45)	0.7 (0.55-0.89)	2.08 (1.6-2.69)	2.31 (1.76-3.04)
9-12	1.63 (1.4-1.9)	**1.13 (0.73-1.75)**	1.39 (0.92-2.11)	1.39 (1-1.95)	1.69 (1.24-2.31)	1.33 (1.11-1.6)	1.75 (1.36-2.26)	1.55 (1.28-1.87)	0.53 (0.43-0.66)	1.82 (1.43-2.33)	1.74 (1.37-2.2)
13-16	1.5 (1.3-1.74)	1.32 (0.82-2.13)	**1.11 (0.77-1.62)**	1.29 (0.93-1.79)	1.38 (1.04-1.84)	1.32 (1.09-1.6)	1.68 (1.31-2.15)	1.39 (1.16-1.66)	0.61 (0.48-0.78)	1.71 (1.34-2.17)	1.53 (1.22-1.92)
17-20	1.38 (1.2-1.59)	0.87 (0.59-1.28)	1.23 (0.83-1.83)	**1.16 (0.85-1.59)**	**1.09 (0.85-1.4)**	1.35 (1.11-1.64)	1.5 (1.18-1.9)	1.3 (1.09-1.56)	0.46 (0.38-0.57)	1.51 (1.2-1.89)	1.52 (1.21-1.9)
21-24	1.35 (1.17-1.55)	0.88 (0.59-1.3)	0.9 (0.64-1.26)	1.08 (0.8-1.47)	1.29 (0.97-1.7)	1.31 (1.08-1.6)	1.33 (1.06-1.67)	1.36 (1.13-1.63)	0.7 (0.55-0.91)	1.31 (1.06-1.62)	1.55 (1.23-1.96)
25-28	1.27 (1.11-1.46)	0.93 (0.62-1.41)	0.98 (0.68-1.4)	0.82 (0.63-1.07)	1.35 (1.01-1.81)	1.22 (1-1.48)	1.32 (1.05-1.66)	1.24 (1.04-1.48)	0.71 (0.55-0.93)	1.54 (1.22-1.95)	**1.15 (0.94-1.41)**
29-32	**1.05 (0.92-1.19)**	0.81 (0.55-1.19)	0.66 (0.49-0.89)	0.97 (0.72-1.3)	0.98 (0.77-1.26)	**1.01 (0.85-1.21)**	1.1 (0.89-1.35)	**1.02 (0.87-1.19)**	0.48 (0.39-0.6)	**1.18 (0.97-1.45)**	1.08 (0.89-1.32)
33-36	1.09 (0.95-1.24)	0.68 (0.48-0.97)	0.94 (0.66-1.34)	1.23 (0.88-1.73)	0.96 (0.75-1.23)	0.98 (0.82-1.17)	1.09 (0.89-1.35)	1.08 (0.92-1.28)	0.51 (0.41-0.64)	1.22 (0.99-1.51)	1.12 (0.91-1.37)
37-40	1 (0.89-1.14)	0.79 (0.54-1.15)	0.61 (0.46-0.81)	0.68 (0.53-0.88)	1.11 (0.85-1.46)	1 (0.84-1.21)	0.98 (0.81-1.2)	1.02 (0.87-1.2)	0.69 (0.53-0.9)	1.12 (0.91-1.36)	0.97 (0.8-1.18)
41-44	1.12 (0.98-1.28)	0.61 (0.43-0.85)	0.79 (0.57-1.1)	0.84 (0.63-1.1)	1.35 (1-1.84)	1.07 (0.89-1.3)	1.1 (0.89-1.37)	1.13 (0.95-1.35)	0.92 (0.68-1.25)	1.07 (0.88-1.3)	1.22 (0.98-1.51)
45-48	1.08 (0.94-1.23)	0.9 (0.6-1.36)	0.87 (0.61-1.23)	0.8 (0.61-1.06)	0.89 (0.69-1.14)	1.17 (0.95-1.44)	1.16 (0.93-1.44)	1.02 (0.87-1.21)	0.83 (0.62-1.11)	0.92 (0.76-1.1)	1.31 (1.04-1.65)
49-52	1.07 (0.93-1.22)	1.1 (0.69-1.75)	0.92 (0.64-1.31)	1.01 (0.74-1.38)	0.98 (0.75-1.28)	1 (0.82-1.21)	1.25 (0.99-1.58)	0.95 (0.81-1.11)	0.82 (0.61-1.09)	1.13 (0.92-1.39)	1.07 (0.87-1.32)

(IRR in bold represent the first time the IRR fell to below the threshold of 1.2)

3.3.1.1 Results stratified by age

Pre-UTS patients had lower incidence of CAP than patients in the same age group who registered post-UTS, and this discrepancy increased with increasing age (Figure 3-4 and Table 3-2). The period of over-reporting also increased with increasing age group, ranging from 8 weeks among those aged 65-69, to 28 weeks for those aged ≥85 years. In all age groups the IRRs oscillated after their initial decrease below the threshold, with most returning to an IRR>1.2 once after their initial decrease (Table 3-2).

Figure 3-4 Comparison of age-stratified CAP incidence in those who registered pre and post-UTS over the first three years of follow-up
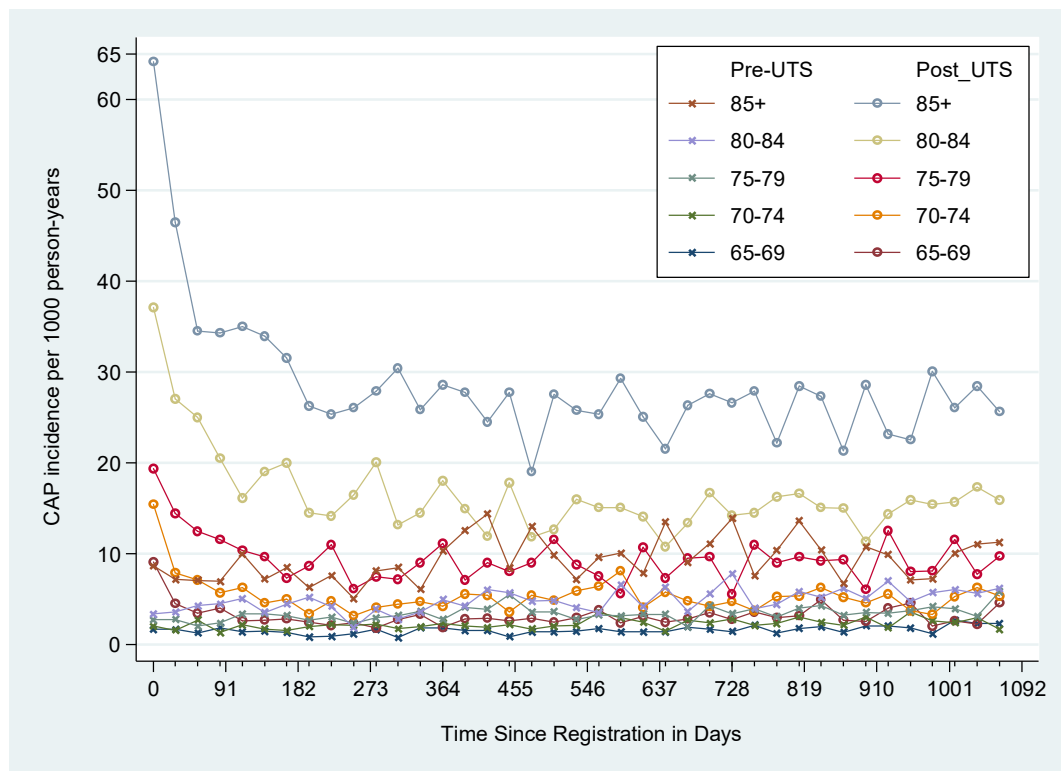
3.3.1.2 Results stratified by sex

There was no clear difference in length of over-reporting of CAP incidence by sex (the IRRs for were within 20% of the baseline at weeks 29-32 weeks for both men and women), although men had slightly higher incidence of CAP in the initial periods of investigation (Table 3-2, Figure 3-5).

Figure 3-5 Comparison of CAP incidence stratified by sex in those who registered pre and post-UTS over the first three years of follow-up
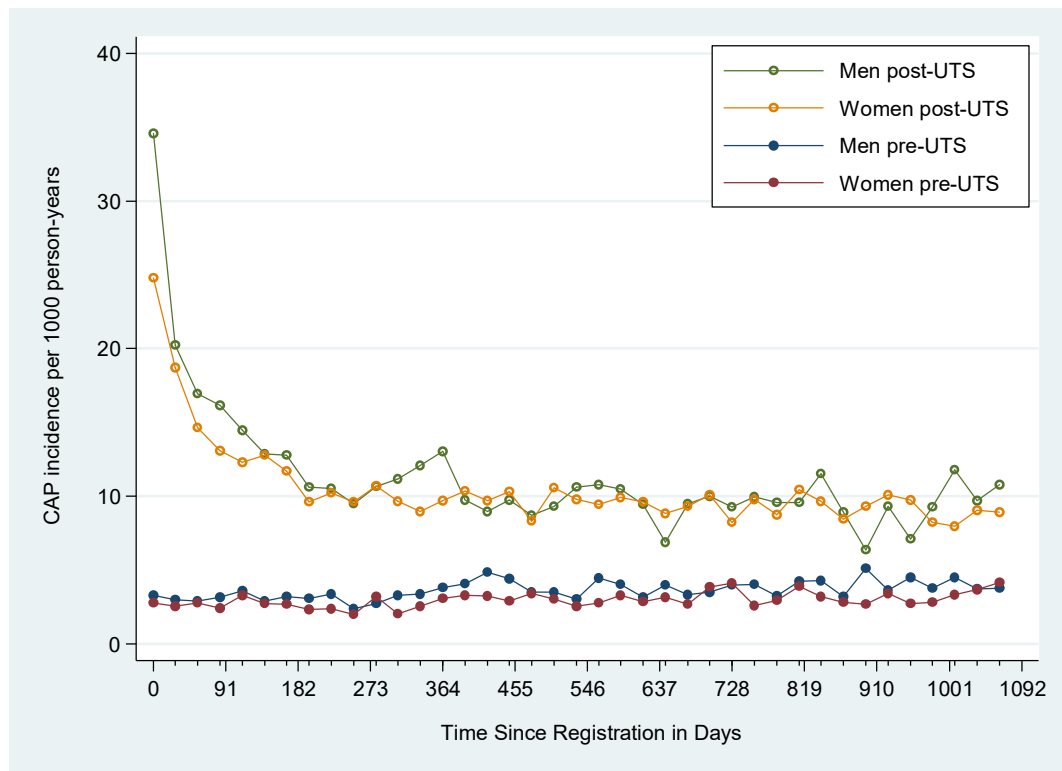
3.3.1.3 Results stratified by time period of registration

Over-reporting was difficult to ascertain in the 1987-1996 period. The CAP trend over time among the post-UTS group was strikingly different to that in the other two periods under study. Both pre- and post-UTS groups were lower than those in other time periods and showed a gradual general increase in incidence over time, although the IRRs were consistently less than one throughout the first year of follow-up (Table 3-2, Figure 3-6).

The two later time periods showed similar patterns to one another, with a high initial peak of over-reporting. The 1997-2003 post-UTS group took four weeks longer to reach the 20% threshold than the 2004-2011 post UTS-group (29-32 weeks and 25-28 weeks respectively).

Figure 3-6 Comparison of CAP incidence stratified by year of start of follow-up in those who registered pre and post-UTS over the first three years of follow-up
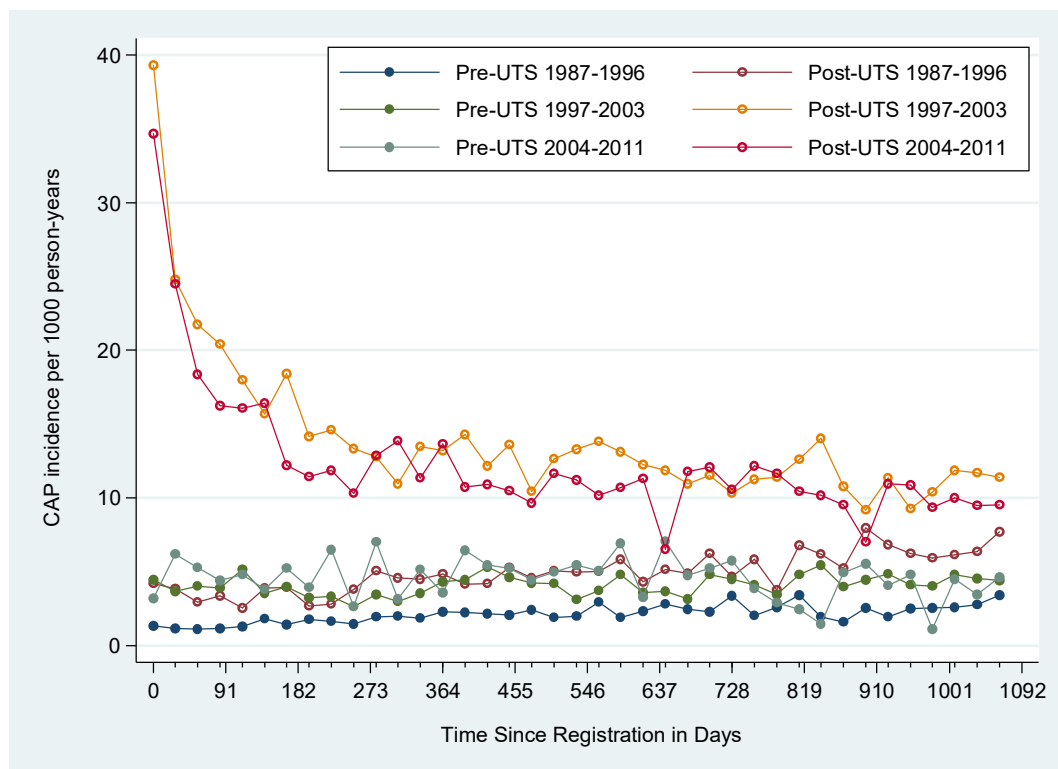
3.3.1.4 Results stratified by health check status

Excluding CAP episodes which started on the same day as a record for a health check did not reduce the period of over-reporting, irrespective of whether the general or strict health check code list was used. There was a very slight decrease in incidence over the first two four-week periods, but exclusion of neither strict nor general health check codes reduced the exclusion period required from follow-up (Figure 3-7).

Figure 3-7 Comparison of CAP incidence over the first year of follow-up in the post-UTS group stratified by health check exclusion
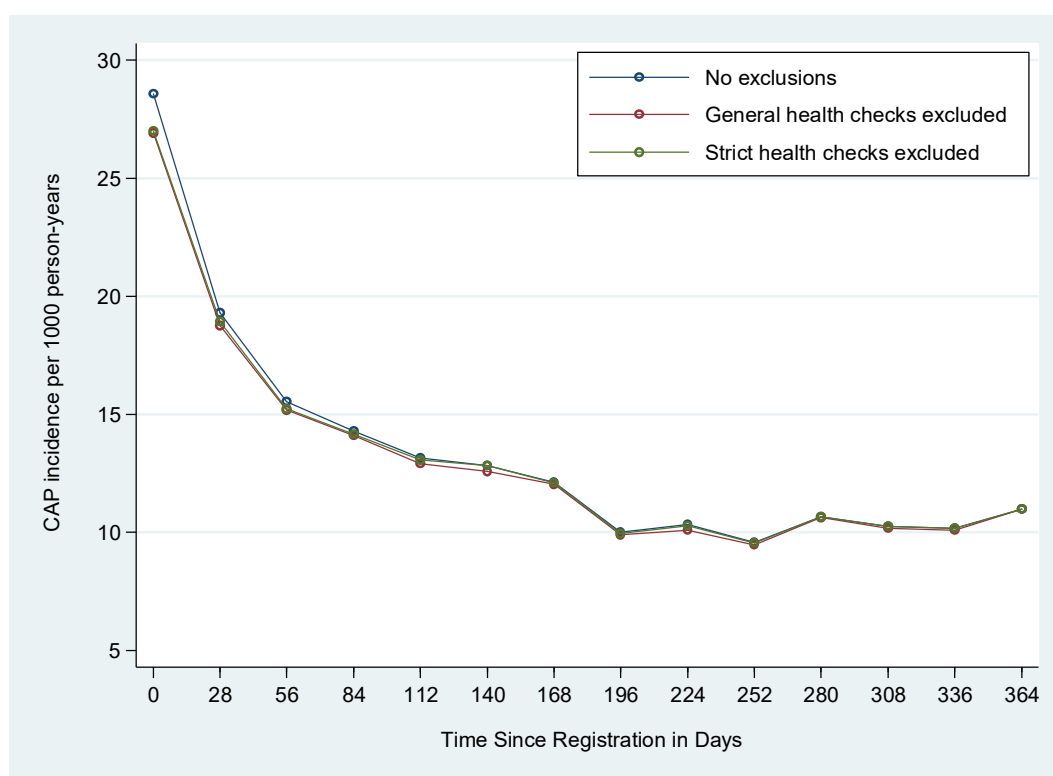
3.3.2 Period of over-reporting for LRTI as a whole

The period of over-reporting of LRTI incidence as a whole was considerably lower than that of CAP, overall and across all stratifications undertaken. In all but three groups, LRTI incidence had returned to baseline by the second risk period (4-8 weeks), with the exceptions of those aged ≥85 and the year group 2004-2011 which both returned to baseline at 9-12 weeks, and the 1987-1996 period which (as for CAP) was consistently lower than the baseline period (Table 3-3 and Appendix B).

Table 3-3 Incidence rate ratios (IRRs) comparing incidence of any LRTI in four weekly periods of the first year with the incidence in the second and third years of follow-up for patients who registered post-UTS, stratified by age, sex and year of start of follow-up

Time (weeks)	Unstratified IRR (95%CI)	Age at start of follow-up IRR (95%CI)					Sex IRR (95%CI)		Year of start of follow-up IRR (95%CI)		
		65-69	70-74	75-79	80-84	85+	Male	Female	1987 - 1996	1997 - 2003	2004 - 2011
1-4	1.57 (1.51-1.64)	1.34 (1.23-1.46)	1.36 (1.24-1.49)	1.31 (1.2-1.43)	1.52 (1.39-1.66)	1.65 (1.53-1.77)	1.69 (1.58-1.8)	1.5 (1.43-1.58)	0.8 (0.74-0.85)	1.55 (1.46-1.65)	1.73 (1.63-1.84)
5-8	**1.15 (1.11-1.19)**	**0.93 (0.86-1)**	**0.94 (0.88-1.02)**	**1.06 (0.98-1.14)**	**1.15 (1.07-1.25)**	1.21 (1.14-1.29)	**1.2 (1.13-1.27)**	**1.12 (1.07-1.17)**	0.66 (0.62-0.7)	**1.15 (1.09-1.22)**	1.24 (1.18-1.31)
9-12	1.11 (1.07-1.15)	0.96 (0.89-1.03)	0.95 (0.88-1.03)	1.05 (0.97-1.13)	1.13 (1.04-1.22)	**1.1 (1.04-1.17)**	1.15 (1.09-1.22)	1.08 (1.04-1.13)	0.65 (0.61-0.69)	1.14 (1.08-1.2)	**1.17 (1.11-1.24)**
13-16	1.07 (1.04-1.11)	1 (0.92-1.07)	0.93 (0.86-1)	1.01 (0.94-1.09)	0.98 (0.91-1.05)	1.11 (1.04-1.18)	1.11 (1.05-1.17)	1.05 (1-1.09)	0.71 (0.67-0.76)	1.09 (1.03-1.15)	1.12 (1.06-1.18)
17-20	1.06 (1.03-1.1)	0.96 (0.89-1.03)	0.95 (0.88-1.02)	1 (0.93-1.09)	1.05 (0.97-1.13)	1.07 (1-1.13)	1.07 (1.01-1.13)	1.06 (1.01-1.1)	0.69 (0.64-0.74)	1.05 (1-1.11)	1.14 (1.08-1.2)
21-24	1.04 (1.01-1.08)	0.98 (0.91-1.06)	0.82 (0.76-0.88)	0.94 (0.87-1.02)	1.07 (0.98-1.15)	1.1 (1.03-1.17)	1.07 (1.01-1.13)	1.02 (0.98-1.07)	0.72 (0.67-0.77)	1.03 (0.98-1.09)	1.11 (1.05-1.17)
25-28	1.03 (0.99-1.06)	0.91 (0.85-0.98)	0.99 (0.91-1.07)	1 (0.93-1.08)	0.97 (0.9-1.05)	1.03 (0.97-1.1)	1.02 (0.97-1.08)	1.03 (0.99-1.08)	0.75 (0.7-0.81)	1.03 (0.98-1.09)	1.07 (1.02-1.13)
29-32	1 (0.97-1.04)	0.9 (0.83-0.96)	0.94 (0.87-1.02)	1 (0.92-1.08)	0.98 (0.91-1.06)	1 (0.94-1.07)	1.05 (0.99-1.12)	0.97 (0.93-1.02)	0.76 (0.71-0.82)	0.96 (0.91-1.01)	1.08 (1.03-1.14)
33-36	1.02 (0.98-1.05)	0.95 (0.88-1.03)	0.91 (0.84-0.98)	1.01 (0.93-1.09)	0.98 (0.91-1.06)	1.03 (0.97-1.1)	1.02 (0.97-1.08)	1.01 (0.97-1.06)	0.81 (0.75-0.87)	0.98 (0.92-1.03)	1.09 (1.03-1.15)
37-40	1.01 (0.97-1.04)	0.95 (0.88-1.03)	0.96 (0.88-1.04)	0.97 (0.9-1.05)	1.02 (0.94-1.11)	0.96 (0.9-1.03)	1.05 (0.99-1.12)	0.98 (0.93-1.02)	0.82 (0.76-0.88)	0.96 (0.91-1.01)	1.07 (1.02-1.13)
41-44	1 (0.96-1.03)	0.91 (0.84-0.98)	0.92 (0.85-0.99)	0.89 (0.83-0.97)	1.06 (0.97-1.15)	1.03 (0.96-1.1)	0.99 (0.94-1.05)	1 (0.96-1.05)	0.87 (0.8-0.94)	0.98 (0.92-1.03)	1.03 (0.97-1.09)
45-48	1.04 (1-1.08)	0.99 (0.92-1.07)	0.98 (0.91-1.07)	0.97 (0.89-1.05)	1.01 (0.92-1.09)	1.07 (0.99-1.14)	1.05 (0.99-1.11)	1.03 (0.98-1.08)	0.9 (0.83-0.98)	1.01 (0.96-1.07)	1.07 (1.01-1.13)
49-52	1.01 (0.98-1.05)	1.01 (0.93-1.09)	0.98 (0.9-1.06)	0.91 (0.84-0.99)	0.99 (0.91-1.08)	1.02 (0.95-1.09)	1 (0.95-1.06)	1.02 (0.97-1.07)	0.89 (0.82-0.96)	1.04 (0.98-1.1)	1 (0.95-1.05)

IRR in bold represent the first time the IRR fell to below the threshold of 1.2

3.4 Discussion

The overall period of over-reporting of CAP incidence among older adults who registered with their GP post-UTS was 28 weeks. The duration of increased incidence in the first few months after patient registration varied by age and year of start of follow-up, but not by sex.

The narrow (28 day) incidence windows used in this analysis, coupled with the use of episodes of illness (including start and end dates) permitted a more precise approach to the problem of over-reporting of historical illness than that used by Lewis et al. They used three month incidence windows, and recommended a nine month period of exclusion at the start of follow-up for post-UTS patients when examining pneumonia incidence. The data I have presented demonstrate that for studies of CAP among older adults this period can be reduced to 28 weeks (around 6.5 months), or even less in patients aged less than 85 years.

The longer over-reporting among the oldest old after joining a practice may be due to a proportion of these patients moving into residential or nursing care, necessitating a change of GP. These frail patients are more likely to have had a previous episode of CAP, as well as being at higher risk of CAP in the future than the younger-old. The additional 2.5 months of follow-up gained from using a 28 week rather than nine month period of exclusion is particularly valuable among these patients with high CAP incidence and probable shorter survival than their younger counterparts. Starting follow-up slightly earlier after registration should enable more of these vulnerable and high-risk patients to be included in analyses.

Results prior to 1997 were surprising – unlike in other time-periods, the incidence increased over time and was lower in the first year of follow-up than the second and third combined. While this was an interesting finding, it does not impact on any of the later work in this thesis, which only included records from 1997 onwards, (the point from which HES-linked data are available). Incidence of CAP reached within 20% of baseline four weeks faster in 2004-2011 (post-QOF) than in 1997-2003, although there is no evidence that this is attributable to QOF itself (pneumonia has never been included as a QOF indicator). This may simply reflect improved timeliness of GP recording of historical events among the older population over time.

Exclusion of episodes of illness recorded on the same day as a health check did not enable any further refinement of the exclusion period. For a record to be recognised as a health check a relevant code must be used by the GP. It is possible that GPs who record health check codes are the same GPs who record historical events with historical dates. Alternatively, historical illnesses recorded during these health checks could be being entered as 'freetext' by some GPs, and thus not coded using Read codes or captured in this analysis.

Those who registered when the practice was already contributing data to CPRD (the post-UTS group) consistently had higher incidence of CAP than the pre-UTS group. The post-UTS group was generally older than the pre-UTS group, and this difference was most notable among those aged ≥85 who comprised >22% of the post-UTS group but <12% of the pre-UTS group. Additionally, a higher proportion of the data contributed by the post-UTS group was in the latter years of the study, whereas the pre-UTS group contributed the vast majority before 2004 (post-UTS: 40.9% patients entered the study in or after 2004; pre-UTS: 92.7% entered before 2004). The combination of increasing incidence of CAP with age, ad improved use of electronic health records over time are likely the underlying reasons for the higher overall rate in the post-UTS group than the pre-UTS group.

LRTI as a broad group was less susceptible to over-reporting than CAP, and had both a smaller and shorter period of raised incidence. Patients visiting a new GP almost certainly report historic episodes of severe, less common illness such as pneumonia more than illnesses they contract yearly such as other LRTI. Those who are not registered with a GP may be prompted to do so if they become ill and require medical attention, and thus the small, short increase in LRTI may in part be documenting incident events.

As highlighted above, this refinement of Lewis' methods has many advantages, such as the use of shorter incidence windows, analysis of a specific sub-population, and additional stratification of results by age and study period. However, there are some potential limitations which should be considered. It is probable that in addition to historical illnesses, some of the CAP events recorded early in follow-up represent truly incident illnesses. As for LRTI as a whole, they may have represented patients' who

registered once ill with CAP, or alternatively they could denote expected end of life events in patients who have moved to palliative or enhanced nursing care. It is impossible to distinguish between the historical and incident events in these data (hence the need for this analysis). Given this, the exclusion of both the events and the person-time contributed by post-UTS registered patients in the first few months of follow-up avoids overestimation of incidence and provides conservative but more robust estimates.

Patients who registered with their GP less than three years before the practice became UTS were not included in these analyses, as they did not clearly fit into either of the pre- or post-UTS groups. There is no obvious reason to expect that these 9% of the study population would have a different level of over-reporting of CAP to those who were included in the analyses, and so generalising these results to the entire pre-UTS group (including the 9% excluded here) should be acceptable.

3.5   Implications of this work

This work informs all subsequent analyses in this thesis, the majority of which concern CAP. Those analyses concerning LRTI also include pneumonia diagnoses within the broader LRTI umbrella. For consistency, I have used a 28 week period of exclusion (that of CAP among those aged ≥85 years) in all ensuing analyses to safeguard against probable over-reporting of any type of LRTI in any age group. Rather than adding this period solely to those who registered post-UTS, I added 28 weeks (196 days) to the current registration date of patients who registered either after or less than 28 weeks before their practice became UTS (thus also capturing the group of patients excluded from these analyses). Any historical illnesses in the early period of these patients' follow-up that were recorded shortly after the start of UTS were thus also excluded. This should have ensured that historical reports of CAP and other LRTI were not included in my analyses.

The reduction of the period of exclusion from nine months to 28 weeks also permitted the inclusion of 74 additional days of follow-up per patient registered post-UTS. More than 340,000 such patients were in CPRD from 1997 onwards (when HES-linked data became available), enabling over 70,000 person-years to additionally be utilised in the incidence analyses of CAP and LRTI, which are outlined in the next Chapter.

Chapter 4 Incidence of community-acquired pneumonia and of lower respiratory tract infections in general among older adults in the United Kingdom

In this Chapter I first describe the existing literature around the burden of CAP and LRTI in general among older adults within the UK and the rest of Europe. I then detail a descriptive analysis of the incidence of community-acquired LRTI and specifically CAP among older adults in the UK (objective 1). The main analyses in this Chapter are presented as a paper published in PlosOne in 2013.

4.1 Literature review

4.1.1 Aim of review

The aim of the literature review was to summarise existing evidence for the burden of CAP, and of all LRTI, among older adults in Europe. A broad review of the clinical and economic burden of CAP in Europe was published in 2012.[29] This paper included information on several aspects of CAP, including aetiology, antibiotic resistance, morbidity and mortality, economic costs, treatment and prevention in addition to incidence. By necessity, the incidence aspect of the review was short and the strengths and limitations of each paper not thoroughly discussed. The CAP incidence section of this review extends that work, focusing specifically on CAP incidence estimates in the older population. To the best of my knowledge, the incidence of LRTI more generally (not restricted to CAP) has not previously been summarised within the European older population.

The primary focus of this review was on studies from the UK, in order to summarise existing knowledge about burden of disease among the UK's expanding older population (and to highlight the methodological limitations of existing research). Studies from Europe were included for context, but those from the US were not included due to their differing categorisation of pneumonia types (as described in section 1.1.3).

4.1.2 Methods

4.1.2.1 Search strategy

Records in Medline were searched from 1980 to the present day. A list of Medical Subject Headings (MeSH) and free text words was developed to look for articles which explored the incidence of community-acquired pneumonia or LRTI in Europe among older adults (either as a focus or a subgroup). Terms for pneumonia, LRTI, incidence, older adults and European countries were combined; the search strategy is provided in Appendix C. The term 'community-acquired' was not included in the search strategy as LRTI is less commonly categorised as such than pneumonia. Instead, whether the study was regarding community-acquired pneumonia was assessed during the screening process.

I also examined annual reports from the Royal College of General Practitioners (RCGP), a large network of GPs across England and Wales who report new episodes of selected illness including pneumonia and LRTI as a whole via a weekly returns service.[119]

4.1.2.2 Inclusion criteria

I searched for articles published between 1980 to March 2015 (the date of the final search). Studies which presented original data for patients aged ≥50 years on the incidence of CAP/LRTI, or which included information from which the incidence in older age groups could be calculated, were included. In order to include as many studies as possible, a minimum age of 50 rather than 65 years was used due to the considerable variability in age categorisation between studies. Case series, case reports, articles which were not available in English and review articles which did not present original research were excluded. In order to investigate the complete burden of CAP (and of all LRTI), two further restrictions were placed on results.

a) Studies which presented only pathogen-specific (e.g. *Streptococcus pneumoniae*) or LRTI subset-specific (such as bronchitis or influenza) rates were not included, as it is difficult to estimate how much these subsets contribute to the total burden of disease.

b) CAP and other LRTI are treated in primary and/or secondary care settings. Studies which are set only in a secondary care setting, whether they include both inpatients and/or outpatients will not capture patients who present to primary care, and thus are not a good measure of the total burden of these infections. For this reason, studies set solely in hospital settings were not included in this review. In contrast, studies set in primary care may include both GP consultations with patients with pneumonia, and retrospectively recorded hospitalised pneumonia episodes. This is possible, because (as highlighted in section 2.1.1.2), GPs are provided discharge summaries of their patients' hospital admissions, which can be added to their general practice records (although the completeness of this recording has not been examined). Thus studies based on GP data were considered eligible for this review.

I screened all articles based on the title and abstract. The full text was obtained for articles which potentially met the inclusion criteria, for further assessment of eligibility. When I was uncertain about the eligibility of a paper, I discussed it with my supervisor and we came to a decision together. The reference lists of included studies were scanned for further papers which met the search criteria but were not included in the Medline results.

4.1.2.3   Data extraction

I extracted the data from each paper using a standardised data extraction form in Excel. Study characteristics of interest included the location, year and duration of the study, the data sources utilised, number of cases, age categorisation used, study design, case ascertainment, case definitions and methods of calculating incidence. For ease of comparison, rates/risks were converted to per 1000 population/person-years where necessary.

Key features of each study were summarised narratively, and are described below. I assessed the quality of the studies in terms of the case definitions used, case ascertainment, and how multiple events and person-time at risk were handled. No formal quality assessment was performed.

4.1.3 Results

After de-duplication the Medline search generated 1412 citations. A flow diagram of included studies is presented in Figure 4-1. I deemed 1347 papers to be not relevant to the aims of this review: reasons included not reporting incidence of CAP/LRTI, not reporting incidence specifically among older adults, reporting incidence of hospital-acquired illness, and reporting only hospitalisation rates or results for specific pathogens/subsets of LRTI. I obtained the full text of 65 papers, of which 17 were found to meet the inclusion criteria. Four papers reported different aspects of the same study; the results from the most recent of these publications are presented below, and the other three papers were excluded (after the additional information on the methods reported in these papers was extracted).[120] One study known to the authors was not identified by the search due to the keywords for the paper not including a geographical region; this paper was also included, as were the data from the RCPG Weekly Returns Service annual reports.[34, 121]

Figure 4-1 Flow chart of study selection

```
┌─────────────────────────────┐        ┌─────────────────────────────┐
│ Records identified through  │        │ Additional records identified│
│ Medline search              │        │ (n=2, one additional paper, │
│ (n=1432)                    │        │ and RCGP Returns data)      │
└──────────────┬──────────────┘        └──────────────┬──────────────┘
               │                                        │
               └──────────────┬─────────────────────────┘
                              ▼
              ┌──────────────────────────────────┐
              │ Records after duplicates removed │
              │ (n=1414)                         │
              └──────────────┬───────────────────┘
                             ▼
              ┌──────────────────────┐         ┌──────────────────────┐
              │ Records screened     │────────▶│ Records excluded     │
              │ (n=1414)             │         │ (n=1347)             │
              └──────────┬───────────┘         └──────────────────────┘
                         ▼
              ┌──────────────────────┐         ┌──────────────────────┐
              │ Full-text articles   │────────▶│ Full-text articles   │
              │ assessed for         │         │ excluded             │
              │ eligibility          │         │ (n=30)               │
              │ (n=50)               │         └──────────────────────┘
              └──────────┬───────────┘
                         ▼
              ┌──────────────────────┐
              │ Eligible studies     │
              │ (CAP: n=13, LRTI: n=6)│
              └──────────────────────┘
```
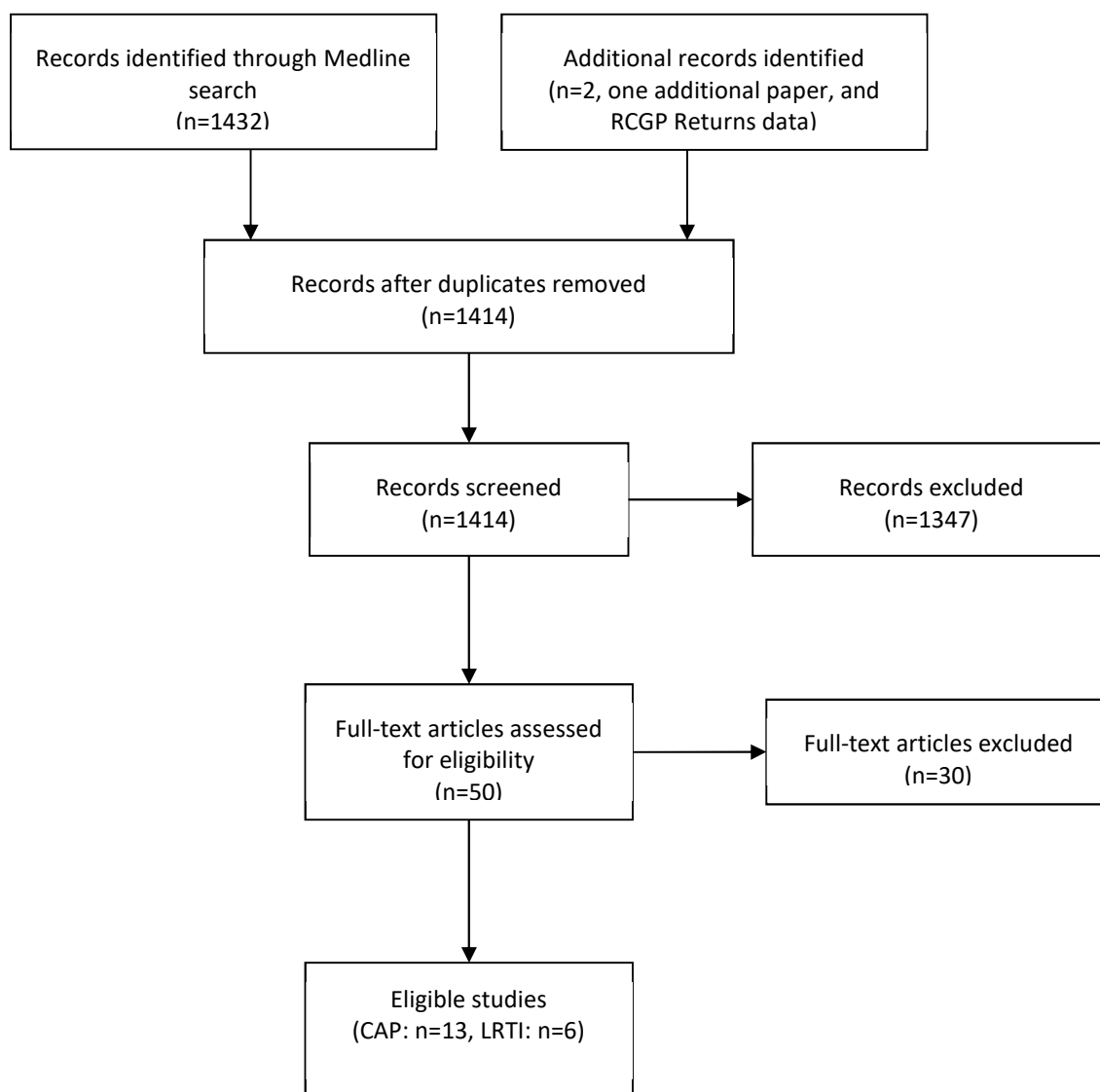
4.1.3.1 Included studies

Incidence estimates for older adults were provided for CAP in 13 studies,[66, 120-131] and LRTI in six studies,[34, 35, 121, 131-133] (with two studies reporting both). Results are presented in Table 4-1, Table 4-2 and Table 4-3.

Community-acquired pneumonia

Study populations: Studies reporting CAP incidence (n=13) were available across a wide geographical area, reporting results from eight countries from northern, central and southern Europe between 1981 and 2012 (Table 4-1). Multiple studies took place in Spain (four), the UK (two) and the Netherlands (two). Methods of case ascertainment varied; five studies which included primary care data only used pre-recorded pneumonia

diagnoses from medical record review (including both UK studies,[66, 121, 128, 130, 131] and five used patients' primary and secondary care reports (one from health insurance reimbursement claims).[120, 123-125, 129] Three studies collected patients' data prospectively from both primary and secondary care.[122, 126, 127] Only two studies were set specifically among older adults, one among those aged ≥50 years, and one among those aged ≥65 years.[122, 129]

Case definitions: Seven studies identified pneumonia cases using a pre-specified code list.[66, 121, 123, 125, 128, 129, 131]   The codes used to define pneumonia were provided in the majority of studies (see Table 4-1); one UK study additionally included some non-specific LRTI codes such as "Acute Lower Respiratory Tract Infection".[128] Five studies based pneumonia diagnoses on the presence of predefined clinical signs and symptoms, four of these required new radiological findings,[120, 122, 124, 126] and one included them as optional (Table 4-1).[127] The Italian prospective study was based on GPs' clinical opinion with no mention of a standardised case definition.[130] One Spanish paper expanded their case definition for older patients to include investigation of those presenting with confusion and other non-specific symptoms.[126] The criteria used to differentiate CAP from HAP were described in five papers, and a range of criteria were used; pneumonia developed >3 days after admission to hospital,[127] after admission at any time,[134] hospital discharge <7 days before symptom onset,[126] and discharge <14 days before pneumonia diagnosis.[122, 124] Four of the remaining eight studies described the included pneumonias as CAP without supplying any exclusion criteria or definition,[120, 125, 129, 130] and four did not explicitly state that the pneumonias included in their analysis were community-acquired (but only cited and compared their results to other papers on CAP).[66, 123, 128, 131] Despite HCAP not being a commonly adopted classification of pneumonia in Europe (see section 1.1.3), patients residing in nursing homes were excluded from two Spanish studies.[124, 126]

Multiple events: Approaches to multiple CAP events were only outlined in three papers; one Dutch study using electronic health records excluded re-consultations within 90 days,[66] and one Dutch and one Spanish study included only the first episode in the study period.[120, 131] Repeat CAP events were not discussed by the other papers, with the exception of Myles et al who stated that 79% of cases had one pneumonia

diagnosis in the study period, but did not adjust for any clustering of episodes within patients in analyses.[128]

Person-time at risk: A single study calculated incidence as a rate, including the time from patients entering the study to their first pneumonia or the study end.[120]   The remaining studies used population-based denominators, calculating risks per 1000 population.[66, 121-131] None of the CAP studies removed person-time not at risk (due to patients' hospitalisations or the duration of their CAP) from the denominator.

Table 4-1 Studies reporting pneumonia incidence

Country, Author [ref]	Population (n if provided)	Study period	Case ascertainment	Case definition	Cases (n)	Rate /1000 population (95%CI if provided)
UK Myles [128]	Patients from >300 general practices	1991-2003	Electronic health record review	All recorded diagnoses (Read codes) of pneumonia , (including some non-specific acute LRTI codes)	All ages: 56,322	≥65y 8 60-69y: 3 70-79y: 7 ≥80y: 16
UK RCGP [121]	Patients from ~100 general practices	2005-2010	Diagnoses/consultations for new illness/ exacerbations	READ codes for pneumonia and pneumonitis mapped to ICD codes for analysis (ICD9 480-486)	N/R	Range among ≥65y: 0.9 to 3.6 Full detail provided in Table 4-2
Finland Jokinen [127]	Individuals living in 4 districts (n=46,979, ≥60y: n=8373)	1981 – 1982	All patients in study area with clinically suspected pneumonia, reported by health care units, university hospital & autopsy.	Predefined clinical signs & symptoms or chest X-ray findings. Diagnosis reviewed & confirmed at follow-up 2/4w later. Excluded if onset >3d after admission to hospital	M F 60-74y: 64 34 ≥75y: 42 27	Total:60-74y: 15.4 (12.4-18.5) ≥75y: 34.2 (26.2-42.1) ≥60y: 19.9 (17.0-22.9) M: 60-74y: 25.0 (18.9-31.0) ≥75y: 65.2 (46.1-84.3) ≥60y: 33.0 (26.9-39.2) F: 60-74y: 9.0 (6.0-12.0) ≥75y: 19.6 (12.3-27.0) ≥60y: 11.8 (8.9-14.8)
Netherlands Hak [131]	Patients from 90 general practices (n=358,008)	2000-2002	Electronic health record review	ICPC code R81 (pneumonia) Only first episode in study period used	N/R	65-74y: 12.5 ≥75y: 21.6

ICPC: International Primary Care Classification. ICD: International Classification of Disease

Table 4-1 Studies reporting pneumonia incidence – continued

Country, Author [ref]	Population (n if provided)	Study period	Case ascertainment	Case definition	Cases (n)	Rate /1000 population (95%CI if provided)
Netherlands van Gageldonk-Lafeber [66]	Patients from ~85 general practices (~350,000 patients)	2001 – 2007	Electronic health record review	GP consultations for pneumonia coded ICPC R81 Excluded repeat consultations <90d after 1st pneumonia consultation.	N/R	2001/2 2006/7 65-74y: 11. 3 17.5 ≥75y: 21.2 31.4 Age/sex adjusted increase/year: 65-74y: 13.6% (9.8-17.5) ≥75y: 11.1% (6.9-15.5)
Poland Patrzalek [123]	Kielce population (n>200,000)	2005, 2007 – 2010	Inpatient & outpatient diagnoses registered by GPs & specialists in internal medicine	ICD-10 codes for pneumonia (J18 and J15).	≥65y: 2005: 535 2007: 581 2008: 488 2009: 309 2010: 330	≥65y: 2005: 19.4 2007: 20.5 2008: 16.9 2009: 10.6 2010: 11.0
Hungary Tichopad [129]	National population (of Hungary) ≥65y: n=1,589,248	2006-2010	Review of health insurance reimbursement claims	Not reported	≥65y: 22,470 inpatient 56,813 outpatient (including primary care)	≥65y: 35.8 60-74y: 34.8 75-84y: 35.5 ≥85y: 45.1
Spain Almirall [126]	Community-dwelling residents aged >14y in one region (annual pop. 74,368)	1993 – 1995	Patients referred by GP or presenting to hospital: diagnosed by hospital specialists.	Comprehensive list of symptoms/signs + chest x-ray (less specific symptoms/signs also investigated for older patients) Excluded: aspiration pneumonia, active pulmonary tuberculosis, hospital discharge <7d pre symptoms	All ages: 241 71 primary care 170 inpatient	≥65y (all): 3.2 M ≥65y: 5.2* F ≥65y: 1.9*

*read off graph. . ICD: International Classification of Disease

Table 4-1 Studies reporting pneumonia incidence - continued

Country, Author [ref]	Population (n if provided)	Study period	Case ascertainment	Case definition	Cases (n)	Rate /1000 population (95%CI if provided)
Spain Vila-Corcoles 2009 [120, 135, 136, 134]	Community dwelling older adults (≥65 years) from 8 General Practices (n=11,240)	2002 - 2005	Medical record review of hospital discharge summaries & primary care databases.	Identified using ICD-9 code in discharge diagnosis/GP databases Case definition: Comprehensive list of symptoms/signs + chest x-ray. Excluded: re-hospitalisation <30d after in-patient CAP treatment & pneumonia post hospitalisation. First episode only	**Care setting** 1° 2° ≥65y 355 118 65-74y 145 50 75-84y 50 43 ≥85y 60 25	≥65y: 14.0 (12.7-15.3) 65-74y: 10.0 (8.6-11.4) 75-84y: 16.9 (14.6-19.4) ≥85y: 29.4 (23.5-36.2) M≥65y: 19.2 (17.1-21.6) F ≥65y: 10.0 (8.6-11.5)
Spain Capelastegui [124]	Adult patients (≥18 years) from one hospital & 150 GPs (n=254,523)	2006 - 2007	GPs alerted study organisers to CAP cases. Hospital discharge diagnoses also checked	New pulmonary infiltrate on CXR + symptoms consistent with pneumonia Excluded: discharged from hospital/ palliative care in previous 14d, or acquired in hospital/nursing home	All ages: 787	65-74y: 4.8* ≥75y: 9.9
Spain Sicras-Mainar [125]	Adults from 6 General Practices & 2 hospitals (n=90315)	2008 & 2009	Electronic health record review	Codes for pneumonia (ICPC-2 codes R81, 480-487 & ICD-9-CM 481) Excluded: tuberculosis, lung cancer or 'from other sanitary areas'	All ages: 581 340 outpatients 241 inpatients	65-74y: 5.1 ≥75y: 8.1
Italy Viegi [130]	Patients from 287 practices (around 410,000 patients)	1999 - 2000	GPs reported suspected CAPs & hospital discharges.	Not reported	≥65y: 324	≥65y: 3.34
Crete Bertsias [122]	Residents aged ≥50 in rural area of one district (n=45,300)	2011 - 2012	Cases identified at 6 general practices, and 2 pulmonary hospital clinics.	Acute LRTI confirmed with lung infiltrate on chest X-ray. Not hospitalised <14d pre diagnosis	≥50y: 124	≥50y: 27.35

*read off graph. ICPC: International Primary Care Classification. . ICD: International Classification of Disease

Table 4-2 Mean yearly pneumonia and LRTI data from RCGP Weekly Returns Service over time[121]

	Pneumonia				LRTI			
	65-74 years		≥75 years		65-74 years		≥75 years	
	Men	Women	Men	Women	Men	Women	Men	Women
2005	0.9	1.5	3.4	2.9	91.4	106.0	144.5	133.1
2006	1.2	1.2	3.2	2.6	81.6	97.3	120.7	120.7
2007	1.5	0.8	2.3	2.7	90.7	104.3	134.7	134.9
2008	1.1	1.2	3.3	2.3	84.1	97.4	130.1	127.7
2009	1.2	0.9	2.7	2.5	78.3	90.2	124.9	121.7
2010	1.4	1.2	3.6	3.0	77.3	92.5	124.9	126.6

Incidence: Within the UK, CAP incidence estimates ranged from 0.8/1000 among women aged 65-74 in 2007 to 16/1000 among both men and women aged ≥80 between 1991 and 2003.[121, 128] Estimates from the rest of Europe were generally higher, peaking at 45.1/1000 among those aged ≥85 in Hungary.[129] Interestingly, rates also varied within countries (Table 4-1 and Table 4-2) with the greatest variation demonstrated in Spain, where incidence rose more than threefold over a decade among patients aged ≥65 in similar geographical regions.[120, 126]

Older age was not finely classified: one broad category was used in four studies, two categories were used in five studies, and three groups in four studies (Table 4-1 and Table 4-2). Of the seven studies (including both of those from the UK) which included two or more age groups within the older population, all showed rising incidence with rising age.[66, 120, 121, 124, 125, 127-129] The age categorisation used varied between studies making direct comparison of age-specific estimates difficult. Of the studies which stratified results within the older population by sex, five out of six found that men had higher incidence of CAP than women.[120, 124, 126-128] This was also the case in most years for the ≥75 group from the RCGP (Table 4-2).[121]

Incidence of CAP over time among the older population was reported by one UK, one Dutch and one Polish study with differing trends. UK data from the RCGP fluctuated over time and showed no clear trend between 2005 and 2010.[121] Over the same period, incidence in Poland was found to have decreased 43.5% among those aged ≥65 years from 19.39/1000 CAP per population to 10.95/1000.[123] Contrastingly, an earlier Dutch study reported an increase in CAP of 13.5% per year among those aged 65-74 and

11.1% among those aged ≥75 between 2001/2 and 2006/7 (adjusted for population changes in age and sex).[66]

All Lower Respiratory Tract Infections

Study populations: The six studies of LRTI incidence were equally distributed between the UK [34, 35, 121] and the Netherlands [131-133] in periods between 1990 and 2010 (Table 4-3). The UK studies included adults of all ages; two were small GP-based prospective cohort studies,[34, 35] and one used information provided by a large national general practice database.[121] Two papers from the Netherlands utilised large general practice databases, one included all adults [131] while one was restricted to those aged ≥65 years.[132] The third Dutch paper was a prospective cohort study of 85-90 year olds in the Leiden region.[133]

Case definitions: Specific case definitions for the diagnosis of LRTI were only supplied for two studies, both from the UK (Table 4-3).[34, 35] The other studies used GP-diagnosed LRTI, in one case with the additional requirement of an antibiotic prescription or microbiological/radiological confirmation.[132] The RCGP data were based on reports of new episodes of illness, with no further definition given.

Multiple events: None of the UK studies provided information on their approach to including multiple events (e.g. creation of illness episodes, or use of first LRTI only), or the statistical methods used to account for multiple episodes per patient in analyses (if these were included).[34, 35, 121] Among the Dutch studies, two restricted analyses to only the first episode of LRTI during the study period,[132, 133] while the other included multiple illnesses (again, with no discussion of adjustments made for clustering).[131]

Person-time at risk: Only two (Dutch) studies used individual person-time at risk to estimate incidence rates,[132, 133] with other studies estimating risks,[34, 35] rates using the mid-year population as the denominator,[131] or providing no information on the denominator used.[121] As with the CAP studies, no adjustment was made to the rates in any study to account for person-time not at risk of LRTI.

Incidence: Overall LRTI rates ranged from 52/1000 among 65-74 year olds to 193.7/1000 in ≥80 year olds.[131, 132] Similar to the CAP studies, broad age categories were used

for in most analyses. Rates of LRTI (as a whole) increased consistently with age, but varied considerably between studies, even among similar age groups. For example, among those aged 65-74 years, incidence ranged from 52 to 106/1000 (Table 4-2 and Table 4-3).[121, 131] Among older age groups, estimates of first LRTI ranged from 93.8 to 193.7/1000.[132, 133] Unlike CAP, there was no clear evidence of higher LRTI incidence rates among men, with higher incidence among women in the UK studies in at least some age groups (Table 4-2),[121, 34] and conflicting evidence of higher rates in men in the two Dutch studies (Table 4-3).[132, 131]

Trends in incidence over time were solely available from the UK via the multiple yearly reports from the RCGP (Table 4-2). These data showed an overall decrease in LRTI incidence between 2005 and 2010 with some yearly fluctuation.[121]

Table 4-3 Studies reporting LRTI incidence

Country, Author [ref]	Population (n if provided)	Study period	Case ascertainment	Case definition	Cases (n)	Rate /1000 population (95%CI if provided)
England Macfarlane [35]	One Nottingham general practice (adult patients only) (60-69y n=1110 70-79y n=783)	1990 - 1991	Surgery consultations & home visits	Provided comprehensive list of symptoms/signs + antibiotics prescribed + patient had not received antibiotics within the last 14 days	All ages: 480 60-69: 85 70-79: 111	60-69y: 66.2 70-79y: 121.5
England Macfarlane [34]	Two Nottingham general practices (adult patients only) (n=14453, not further reported by age)	1997 - 1998	Records of previously well adults consulting with LRTI	Provided comprehensive list of symptoms/signs + no consultation for LRTI in previous month	All ages: 638 (<25% of cases aged ≥60)	All ≥60y: 56 M ≥60y: 45 F ≥60y: 64
UK RCGP [121]	Patients from ~100 general practices	2005-2010	Diagnoses/consultations for new illness/ exacerbations	READ codes for LRTI and pneumonia mapped to ICD codes for analysis	N/R	Range among ≥65y: 77.3 to 144.5. Full detail provided in Table 4-2

Table 4-3 Studies reporting LRTI incidence - continued

Country, Author [ref]	Population (n if provided)	Study period	Case ascertainment	Case definition	Cases (n)	Rate /1000 population (95%CI if provided)
Netherlands Voordouw [132]	General Practice Research Database. Patients aged ≥65 + ≥1 year recorded medical history (≈150 GPs, (n=26,701)	1996 - 2002	Identified from medical chart of patient	LRTI (pneumonia, acute bronchitis, or exacerbation of chronic bronchitis). + antibiotics prescribed OR + X-ray/microbiological confirmation. First LRTI in study period only.	65-69: 956 70-79: 1478 ≥80: 978	65-69y: 91.1[†] 70-79y: 140.5[†] ≥80y: 193.7[†]
Netherlands Sliedrecht [133]	All consenting residents aged ≥85 living in Leiden, Netherlands (n=587)	Enrolled 1997-1999, followed for earlier of 5 years or death	Annual interviews of participating physicians, and deaths due to pneumonia (if no prior clinical diagnosis of LRTI)	GP/nursing home physician diagnosis of LRTI, based on history taking, physical exam and clinical judgement. First LRTI in study period only.	173 first LRTI	85-89y: 93.8 (79.8-107.7)
Netherlands Hak [131]	Patients from 90 computerised general practices throughout Netherlands (n=358,008)	12 months within 2000 – 2002	Electronic health record review	Codes for; acute bronchitis, influenza, pneumonia or asthma/COPD exacerbations Sensitivity analysis incl: dyspnoea, wheezing, other respiratory problems, coughing, abnormal sputum.	N/R	65-74y: 52* ≥75y: 70

† calculated, not provided in paper, *read off graph.

4.1.4 Discussion

This literature review highlights the paucity of information on incidence of CAP and LRTI as a whole among older adults in the UK, and describes considerable variation between estimates from the rest of Europe.

Over a 35 year period, only two CAP and three LRTI studies were found which contained information on incidence among older adults in the UK. These studies confirmed that incidence of CAP (and LRTI as a whole) increases markedly with age among the older population. UK incidence trends over time were only available from the RCGP data, which suggested no clear trend for CAP rates between 2005 and 2010, but a slight decrease in LRTI rates.[121]

In general, UK studies provided estimates which were either broadly comparable to or lower than those from other European countries. This may have been in part due to the UK studies' reliance on stand-alone GP records (thus underestimating overall incidence), whereas many other countries also included hospital admission records.

In addition to the variation seen between countries, there was also a notable range of CAP and LRTI incidence estimates within countries. Differing age structures (and a lack of consistency in age-stratification of rates) would undoubtedly account for some of the variation displayed. Additionally, there may be regional differences in important risk factors such as co-morbidities and smoking status, or preventative measures such as pneumococcal and influenza vaccine programmes.

A further contributing factor to the diverse estimates presented is the range of methodologies used by the different studies. Methods of ascertaining cases of CAP (and of all LRTI) varied considerably between studies, as did the case definitions used. Of the five CAP studies with specified case definitions, only one included a modified set of criteria to capture the sometimes complex or atypical presentation of CAP in older adults (such as weakness or confusion).[126] Additionally, one prospective study included GP-suspected CAP, which may have enabled capture of these patients.[130] The exclusion of older patients with atypical CAP symptoms would likely have led to an underestimation of the disease burden among the older population. The coded clinical diagnosis used by studies utilising electronic health records may have included atypical

presentations, and thus not been subject to this particular method of under ascertainment.

Standalone GP records were used to calculate incidence in five of the CAP studies and all of those for LRTI as a whole. Both of the UK CAP estimates solely used GP data; estimates from the RCGP were among the lowest presented,[121] possibly in part due to the under-diagnosis of pneumonia by GPs when chest radiograph confirmation was unavailable.[20] Estimates from Myles et al were probably inflated due to the inclusion of non-specific codes for LRTI, some cases of which will not have been severe enough to be classed as pneumonia.[128] In contrast, eight of the non-UK CAP studies included both primary and secondary care data, enabling capture of both patients who initially presented to their GP, and those who presented to A&E. If GP recording of hospital discharge summaries was incomplete, this would have led to an underestimate of incidence in primary care data due to under-recording of these more severe cases. Conversely, both stand-alone and linked data are vulnerable to producing over-estimates of incidence if repeat consultations are not managed appropriately, and such an approach was reported in only a single CAP study.[66] This potential for both over- and underestimation when using stand-alone GP records without an episode structure makes the incidence estimates provided by studies using only primary care data difficult to interpret.

One CAP,[131] and two LRTI papers [132, 133] limited inclusion to the first episode of illness. Patients who have experienced a CAP/LRTI more broadly are known to be at higher risk of subsequent episodes of illness. Therefore excluding these recurrent episodes will probably have led to an underestimate of the disease burden.

Only five of the 13 CAP studies provided exclusion criteria for potential HAP, and these definitions varied from onset >3 days after admission to hospital to an apparently indefinite period after hospitalisation.[122, 124, 126, 127, 134] Two of these studies excluded patients residing in nursing care, which is not in line with current European definitions of CAP, and will have led to the exclusion of older, frailer patients at higher risk of the infections under study.[126, 124] In both studies, while patients in nursing facilities were excluded from the numerator if they had pneumonia, there is no mention that all nursing home residents were also excluded from the denominator. This

reduction in cases and inflation of the population at risk will have both led to an underestimation of the incidence of CAP, in addition to making the results from these studies less generalisable to the older population as a whole.

Overall, the number of studies which focussed on older adults was disappointing considering their high burden of disease, and none of the UK studies of either CAP or of all LRTI specialised in this population. In two of the three UK LRTI studies, less than 25% of the study populations were aged ≥60 years, resulting in a small number of older participants.[34, 35] Among all studies, the categorisation of age was fairly broad despite large study sizes. The older population is a diverse group including a range of patients, from those who work full-time to those who require round the clock care. This broad age categorisation offers little insight into the true burden among those most at risk.

Information regarding incidence trends over time was also sparse, and was only presented by one UK and one non-UK source. As the proportion of the European population aged ≥65 years continues to grow, it is crucial that we develop a more thorough knowledge of the incidence of LRTI and particularly CAP over time, including estimates finely stratified by age to enable accurate service planning and provision of resources.

The CAP incidence results presented demonstrate a high degree of variability, even when examined by time, country, primary or secondary care setting or study design. Some of the disparity between study findings is likely to be attributable to the different sources of information used to make the CAP diagnosis. As discussed in section 1.1.4, there is evidence that GPs tend to underdiagnose pneumonia in patients with acute cough when they do not have access to a chest radiograph,[20] while A&E clinicians tend to over-diagnose pneumonia compared to radiology reports.[21] Thus, studies in which the diagnoses were made largely in primary care may have underestimated overall incidence, whereas inclusion of diagnoses from A&E (whether recorded in secondary care records or reported back to GPs and included in patients' primary care notes) may have led to higher estimates of CAP incidence. Even when chest radiographs were used to make diagnoses, the known inter-observer variability between radiologists across Europe in detecting infiltrates due to pneumonia on chest radiographs is also likely to

explain some of the differences in reported incidence between countries and over time.[20] The availability of chest CT scans may have further contributed to between-study differences; whether this would have led to increased estimates due to the ability to identify chest infiltrates too small to be seen on chest radiographs,[137] or decreased estimates due to better imaging and exclusion of some CAP previously diagnosed by radiographs, is difficult to ascertain.[138]

The methods used to conduct the review also need consideration. While I conducted the literature search, assessment of eligibility and data extraction in a systematic manner, time and resource constraints necessitated restricting the search to the Medline database, and to papers reported in English. This may have resulted in some non-UK European studies being missed from this review. Additionally, I selected studies and performed the data extraction single-handedly, without the benefit of a second reviewer. This could have led to errors in the data extracted, and exclusion of some eligible studies, although I did seek advice on papers whenever I was unsure about their eligibility for inclusion in the review.

4.1.5   Rationale for incidence study

As detailed above, at the time of this study there was very little detailed information available on the burden of CAP (or of LRTI as a whole) within the older UK population. The UK studies identified in this review lagged behind those of other European countries due to their use of stand-alone GP data, rather than linked primary and secondary sources. The UK studies were also small in size,[35, 34] included non-specific LRTI codes when calculating pneumonia incidence,[128] or did not provide results further stratified by age after 75 years.[121] None of the UK studies differentiated between community- and hospital-acquired disease, and none excluded person-time not at risk from their denominators. Accurate categorisation of person-time at risk is particularly important among the older population, who spend more time in hospital (and therefore not at risk of community-acquired infections) than the younger population. The additional use of linked data to estimate the burden of these important infections would both enable more complete capture of pneumonia events, and allow better distinction between hospital- and community-acquired infections. Furthermore, the UK studies did not clearly address the issue of multiple consultations for one episode of illness. A clearly

defined episode structure to manage repeat recording of an ongoing illness would improve UK estimates of these common infections.

4.2 Introduction to research paper 1

A retrospective cohort study was used to estimate the burden of CAP, and of all community-acquired LRTI, among UK older adults. The aim of this work was to use large general practice and (when available) linked hospital admissions data to provide better estimates of CAP and community-acquired LRTI in older adults in the UK. As CAP is nested within LRTI, I present the incidence of LRTI first, followed by that of the CAP subset.

I used the methods outlined in Chapter 2 to derive a cohort of patients using both the stand-alone CPRD and CPRD HES-linked data. This enabled the inclusion of a large number of patients, and captured more events than if only the linked-data had been used. Episodes of LRTI and of pneumonia were defined as explained in section 2.4.1 and community-acquired infections differentiated from those acquired in hospital as set out in section 2.4.2. Patients' person-time at risk was excluded from follow-up as discussed in section 2.5.2.

In addition to the results presented graphically in the paper, tables of these results are included in Appendix D (supplied as online appendices to the article).

4.3   Research paper cover sheet

LONDON
SCHOOL *of*
HYGIENE
&TROPICAL
MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED <u>FOR EACH</u> RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Elizabeth Millett
Principal Supervisor	Sara Thomas
Thesis Title	The use of linked electronic health data to investigate the burden and outcomes of community-acquired pneumonia among older individuals in the United Kingdom.

<u>If the Research Paper has previously been published please complete Section B, if not please move to Section C</u>

SECTION B – Paper already published

Where was the work published?	Plos One		
When was the work published?	11th September 2013		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A		
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	This study was conceived by Sara Thomas, who also obtained funding, ethical approval and the data for the study from CPRD. I developed the detailed study design, supervised by Sara Thomas and with advice from Rhian Daniels. The Read and ICD-10 codelists for LRTI and pneumonia were

	devised by Sara Thomas and two other clinical epidemiologists. Sara Thomas derived the hospitalisation Read codelist, which we each independently categorised into the eight hospitalisation groups. We compared categorisations and discussed those which did not match until we reached a consensus. I designed the methods to derive illness-episodes, differentiate community- from hospital-acquired infections and person-time at risk, with input from Sara Thomas and contributions from Helen McDonald. Clinical advice was provided by Jennifer Quint and Liam Smeeth. I conducted all data management, analysis, led the interpretation of results (with input from Sara Thomas) and I wrote the first draft of the paper. All co-authors contributed revisions, which I then incorporated. After peer-review, I further adapted the manuscript to include the reviewers' comments, with advice from all co-authors.

Student Signature: _Lizzie Millett_ **Date:** 04 / 12 / 15

Supervisor Signature: _San Thomas_ **Date:** 4/12/15

This work was published in PlosOne which is an open access journal which allows all authors to retain ownership of the copyright for their article. More information is available at http://www.plosone.org/static/license.

Incidence of Community-Acquired Lower Respiratory Tract Infections and Pneumonia among Older Adults in the United Kingdom: A Population-Based Study

Elizabeth R. C. Millett[*]**, Jennifer K. Quint, Liam Smeeth, Rhian M. Daniel, Sara L. Thomas**

Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

Abstract

Community-acquired lower respiratory tract infections (LRTI) and pneumonia (CAP) are common causes of morbidity and mortality among those aged ≥65 years; a growing population in many countries. Detailed incidence estimates for these infections among older adults in the United Kingdom (UK) are lacking. We used electronic general practice records from the Clinical Practice Research Data link, linked to Hospital Episode Statistics inpatient data, to estimate incidence of community-acquired LRTI and CAP among UK older adults between April 1997-March 2011, by age, sex, region and deprivation quintile. Levels of antibiotic prescribing were also assessed. LRTI incidence increased with fluctuations over time, was higher in men than women aged ≥70 and increased with age from 92.21 episodes/ 1000 person-years (65-69 years) to 187.91/1000 (85-89 years). CAP incidence increased more markedly with age, from 2.81 to 21.81 episodes/1000 person-years respectively, and was higher among men. For both infection groups, increases over time were attenuated after age-standardisation, indicating that these rises were largely due to population aging. Rates among those in the most deprived quintile were around 70% higher than the least deprived and were generally higher in the North of England. GP antibiotic prescribing rates were high for LRTI but lower for CAP (mostly due to immediate hospitalisation). This is the first study to provide long-term detailed incidence estimates of community-acquired LRTI and CAP in UK older individuals, taking person-time at risk into account. The summary incidence commonly presented for the ≥65 age group considerably underestimates LRTI/CAP rates, particularly among older individuals within this group. Our methodology and findings are likely to be highly relevant to health planners and researchers in other countries with aging populations.

Introduction

Pneumonia and lower respiratory tract infections (LRTI) are major causes of morbidity and mortality among those aged 65 years and over in the UK and other European countries [1–3]. The UK's population is aging; recent estimates suggest that in 2035, 23% of the UK will be aged ≥65 years and 5% will be ≥85, compared to 17% and 2% respectively in 2010 [4]. The 'oldest old' (≥85 years) are at particularly high risk of infections due to co-morbidities and waning immune function. Community-acquired pneumonia (CAP) in older individuals is a particular concern, as it can aggravate underlying co-morbidities and have serious consequences [5]. Thus, the need

has been highlighted for new population-based studies of the incidence of these infections among older adults in different European locations [6].

There are few longitudinal studies on the burden of these infections specifically amongst older adults in the UK. This is a disparate group, including people working full-time and those that require round-the-clock care. Available incidence estimates vary, partly due to different age categorisations and methods used [1,7,8], and community- and hospital-acquired infections are rarely differentiated. Existing studies of regional and socio-economic variations in incidence have not age-stratified further after 65 years. Therefore there is a paucity of information for this important and growing subsection of the

population; it is essential that their LRTI burden is better understood to enable planning and provision of health care. The extent of antibiotic prescribing in general practice for both LRTI and CAP among older adults also needs ongoing assessment.

The aim of this study was to describe the incidence of community-acquired LRTI and CAP among individuals aged ≥65 years between April 1997 and March 2011, using linked electronic health records from primary and secondary care. We describe how the incidence of these common infections varied over time by age, sex, region and socioeconomic deprivation and the extent of antibiotic prescribing in this group.

Methods

Data sources

The Clinical Practice Research Datalink (CPRD, formerly known as GPRD) is a large UK-based electronic database of primary care records which currently includes around 8% of the UK population. The age, sex and regional distribution of patients from contributing practices are representative of the UK overall [9]. Anonymised patient-level information including diagnoses (coded using Read codes), referrals to specialist care, prescriptions, hospitalisations, demographic and lifestyle details are included. Data from a practice are only used for research after they have met a series of quality checks and have been deemed 'up to standard' by CPRD.

Over 50% of English practices that contribute to CPRD consent to linkage of their patients' records to Hospital Episode Statistics (HES) data. These contain information on all NHS inpatient hospitalisations in England since 1997, with diagnoses coded using ICD-10. Hospitalisations include one or more 'episodes', each denoting a period of consultant care.

Study population and follow-up time

Patients aged ≥65 years between 1st April 1997 and 31st March 2011 were eligible for inclusion. Follow-up began at the latest of the study start date, patients' 65th birthday, the date CPRD deemed the practice 'up to standard' or 28 weeks after patients' registration (to exclude reports of historical illness that are often recorded when a patient first joins a practice; the 28-week period was chosen after analyses based on existing methods) [10]. Follow-up ended at the earliest of the study end date, death, transfer out of the CPRD or the practices' last data collection date. Patients who contributed at least one day of follow-up were included in the study.

Codes used to define LRTI, hospitalisations and antibiotics

Read and ICD-10 code lists for LRTI and within this list, pneumonia were developed by three clinical epidemiologists, including a consultant respiratory physician and a GP. An estimated 50-70% of chronic obstructive pulmonary disease (COPD) exacerbations are due to LRTI [11]; COPD exacerbation codes that did not mention infection were also identified for sensitivity analyses.

We examined evidence of hospitalisations to distinguish between potentially hospital-acquired and community-acquired infections. In the CPRD data, we identified hospitalisation codes and relevant fields indicating hospitalisation in the consultation, referral and clinical files (code lists available on request).

Antibiotic therapy codes were identified and categorised by their British National Formulary (BNF) subchapter, excluding antituberculosis and antileprotic drugs [12–15].

LRTI/pneumonia illness-episode structure

Records containing LRTI or pneumonia codes were identified in both CPRD and HES. In order to accommodate multiple consultations for one illness, CPRD or primary HES LRTI/pneumonia records within 28 days of each other were regarded as part of the same illness-episode. The first record was deemed the index date and the illness-episode finished 28 days after its last LRTI code. Within HES, only LRTI/pneumonia codes recorded as the primary code of the first episode of a hospitalisation (the condition the patient was admitted for) were used when defining the index date, to avoid including hospital-acquired infections.

As pneumonia formed a subset of LRTI, pneumonia illness-episodes could start on the same date as an LRTI episode (if the patient initially presented with pneumonia) or at some point within an LRTI episode (if an LRTI had worsened).

Defining community-acquired illness

Incident cases of LRTI were regarded as hospital-acquired if in the previous 14 days the patient had been discharged from hospital (using HES records for any illness) or there was a CPRD hospital code (using the unlinked data) [14,16–19].

Other variables

We used the 'financial year' definition of April–March to ensure the winter peak of LRTI was not split across two years. Age was grouped in five-year bands from 65 to 89 years, then as ≥90 years. English regions were defined by Strategic Health Authority (SHA). Index of Multiple Deprivation (IMD) quintile (2007) was available at Office for National Statistics (ONS) small area level (100 houses) for >50% of CPRD patients. The IMD is calculated using seven domains: income, employment, health deprivation and disability, education, skills and training, barriers to housing and services, crime and living environment.

Analyses

At their simplest, incidence rates are calculated by dividing the number of new episodes of illness by the person-time at risk. Patients were considered not at risk of a community-acquired LRTI during an LRTI illness-episode (whether community or hospital-acquired), during a HES hospitalisation or for the 14 days after any HES hospitalisation or CPRD hospital code. This person-time was excluded from the denominator when calculating incidence.

Incidence rates were calculated by year, sex, age, region of England and IMD quintile. Poisson regression with random effects was used to account for multiple illness-episodes per

person. Incidence rates were directly age-standardised using the ONS mid-year UK population estimates from 2004, and 95% confidence intervals (CI) were calculated.

We identified how many patients had been prescribed an antibiotic on the illness index date, and the type of antibiotic prescribed. Among those without a prescription on the index date, we calculated the percentage that were hospitalised or died as possible reasons for not having received a GP prescription. For remaining patients, analyses were repeated sequentially for the week after, and then two to four weeks after the index date. Finally, records in the week before the index date were examined to assess how many patients had received an antibiotic prescription in this time.

Statistical analysis was conducted using Stata version 11.2 and Microsoft Excel.

Ethics information

All data were anonymised prior to receipt by the authors. Ethics approval for the study was given by the Independent Scientific and Advisory Committee (of CPRD), and the London School of Hygiene and Tropical Medicine Ethics Committee.

Results

The study population comprised 1,534,443 patients from 625 practices across the UK (Table 1). Over half the participants were aged 65-69 at the start of the study and 56% were female. Median period of patient follow-up was 5.1 years (interquartile range (IQR): 2.3-9.2). HES-linked information was available for 59.7% of patients, whose characteristics were largely similar to those of the whole cohort (Table 1).

Incidence by age and sex

Over the 14-year study period, 974,121 episodes of community-acquired LRTI were identified in 448,469 patients (median number of episodes=1, IQR:1-2). The median age at diagnosis was 76 (IQR:70-82) years. Crude overall LRTI incidence was 122.93 episodes/1000 person-years (IQR: 122.49-123.37/1000 person-years); incidence generally increased over time with some fluctuations, from a low of 100.96 (1997) to a peak of 148.04/1000 person-years (2008), and was similar in men and women (Figure 1a & Table S1). After standardising for age the increase over time was less marked, and a higher rate in men than women was revealed (Figure 1a & Table S2).

The difference between the sexes was also apparent in the age-stratified rates (Figure 2a,b), particularly in older age groups. Incidence increased with age, doubling between the 65-69 and 85-89 age groups, and fluctuations over time were more marked at older ages. Age-stratified results were not presented graphically for those aged 90+, as the age structure of this group varied over time, making the results difficult to interpret (Table S1).

Inclusion of COPD exacerbation codes did not change the overall pattern of LRTI incidence over time, but increased the rates in both sexes by around 7% (Table S3 & Figure S1).

A total of 64,978 CAP episodes were identified in 58,772 patients. CAP patients were generally older than those with

Table 1. Characteristics of the study population, for all patients and for patients with HES-linked data.

	Entire study population		HES-linked patients
	n (%)		n (%)
Number of patients	1534443		916128 (59.7)
Median years follow-up (IQR)	5.1 (IQR:2.3-9.2)		5.3 (IQR:2.3-9.6)
Sex			
Male	672858 (43.9)		402474 (43.9)
Female	861585 (56.1)		513654 (56.1)
Age at start of follow-up			
65-69	819333 (53.4)		491205 (53.6)
70-74	237349 (15.5)		140743 (15.4)
75-79	197400 (12.9)		117298 (12.8)
80-84	142077 (9.3)		84533 (9.2)
85-89	89292 (5.8)		53559 (5.8)
90+	48992 (3.2)		28790 (3.1)
Region practice is based[a]		England only %	
North East	29432 (1.9)	2.4	20615 (2.3)
North West	178011 (11.6)	14.5	145665 (15.9)
Yorkshire & The Humber	71012 (4.6)	5.8	46482 (5.1)
East Midlands	60824 (4.0)	5.0	32744 (3.6)
West Midlands	130162 (8.5)	10.6	106627 (11.6)
East of England	150977 (9.8)	12.3	115749 (12.6)
South West	144749 (9.4)	11.8	128592 (14.0)
South Central	165094 (10.8)	13.5	110107 (12.0)
London	155219 (10.1)	12.7	106431 (11.6)
South East Coast	139681 (9.1)	11.4	103116 (11.3)
Northern Ireland	43633 (2.8)		N/A
Scotland	121428 (7.9)		N/A
Wales	144221 (9.4)		N/A
Index of Multiple Deprivation (IMD) quintiles relative to country as a whole			
Unavailable	712107 (46.4)		99150 (10.8)
0 (least deprived)	190874 (23.2)		189466 (23.2)
1	204589 (24.9)		203345 (24.9)
2	171462 (20.9)		170527 (20.9)
3	150185 (18.3)		149164 (18.3)
4 (most deprived)	105226 (12.8)		104476 (12.8)

[a] Hospital Episode Statistics (HES) data available for England only

doi: 10.1371/journal.pone.0075131.t001

LRTI (median age=81 years, IQR:75-87). Overall incidence of CAP was 7.99/1000 person-years (IQR:7.92-8.07/1000 person-years), was somewhat higher in men than women and increased slightly over time (Figure 1b, Table S4). After standardising for age, the increase was no longer apparent and the higher rate in men than women was accentuated (Figure 1b & Table S2). CAP rates increased with age with the rate in the 85-89 years group over seven times that of the 65-69 year olds
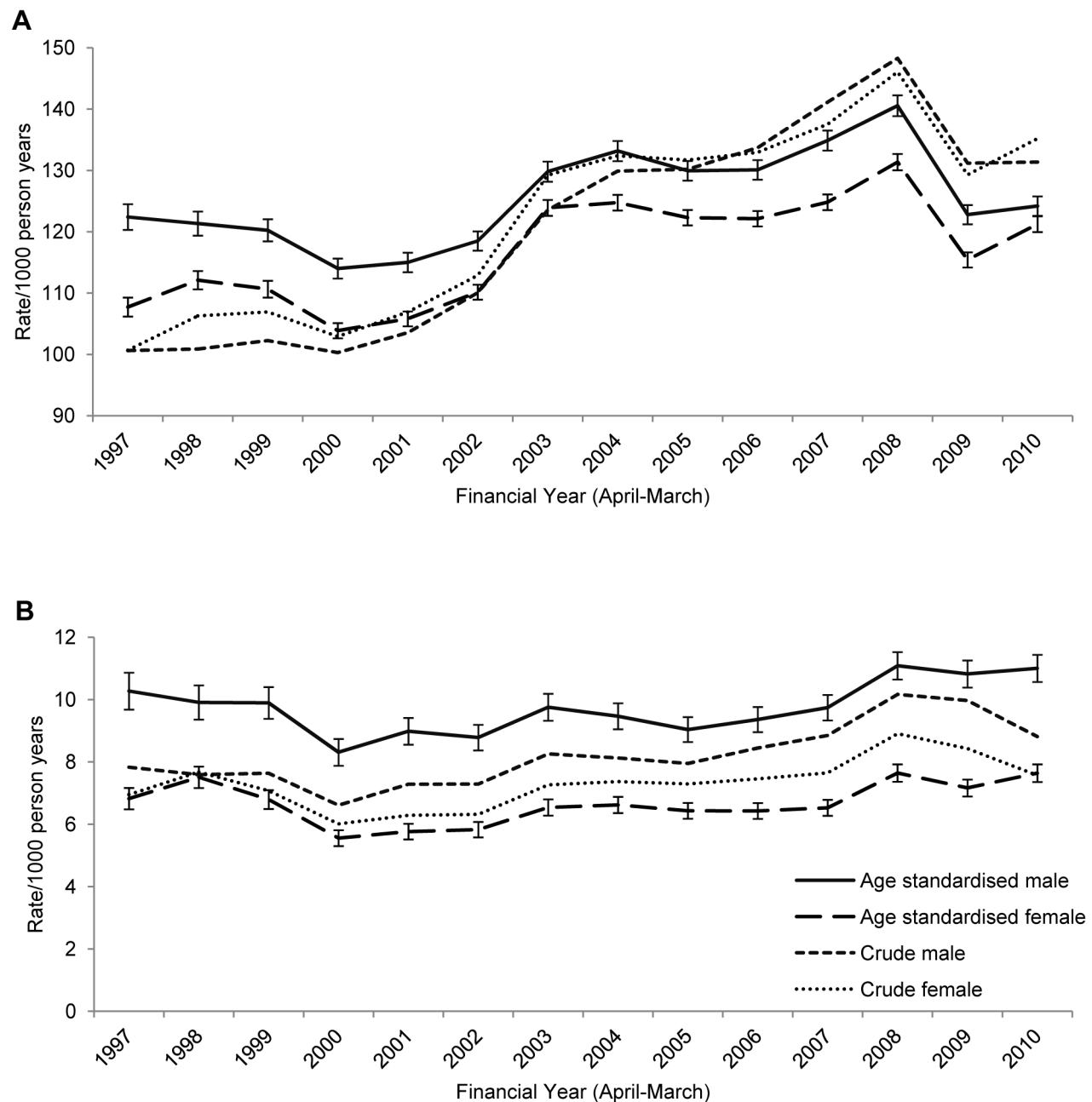
Figure 1. Incidence of LRTI and CAP by sex over time. Crude and age-standardised incidence of a) LRTI and b) CAP by sex over time. Standardised to UK population, mid-year 2004.

doi: 10.1371/journal.pone.0075131.g001

(Figure 2c,d). Women's CAP incidence was comparable to that of men aged five years younger.

Incidence by region and deprivation

Incidence of both LRTI and CAP varied markedly by English region (Figure 3). Age-standardised incidence of LRTI was higher among the North and Midland regions than the South, and was highest in the North West. For CAP, high rates were

seen in the North East, but also in the South Central region. For both conditions, the lowest rates were in London and the South East Coast. Rates are only presented for England; rates for Scotland, Northern Ireland and Wales were not comparable, due to lack of linked HES data outside England.

Incidence of both LRTI and CAP increased with increasing deprivation, with a marked difference between IMD quintiles three and four (the most deprived, Figure 4). This pattern was
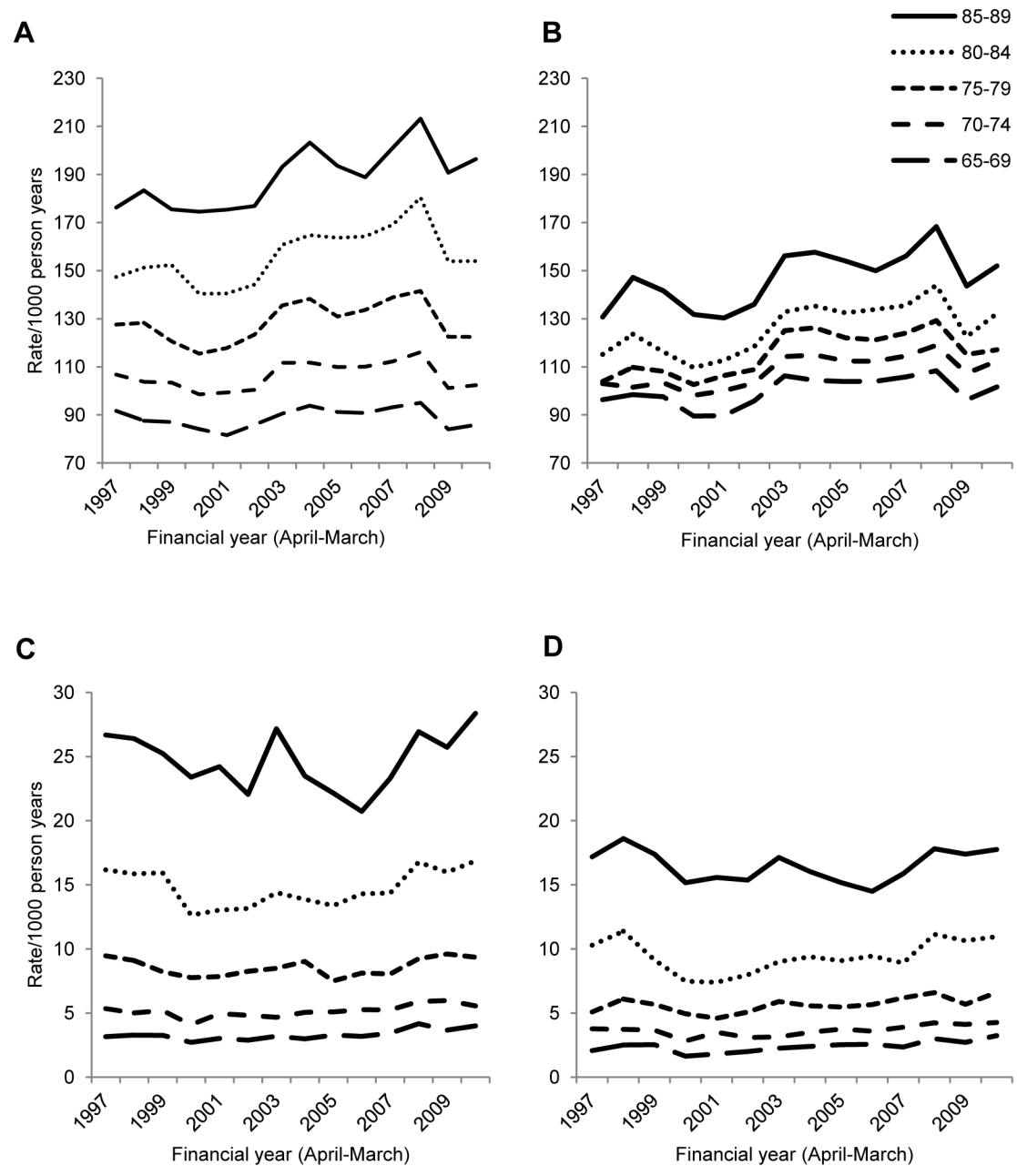
Figure 2. Incidence of LRTI and CAP by age and sex over time. Incidence by age of LRTI in a) men, b) women and CAP in c) men, d) women over time.

present for both men and women, and remained after standardising for age.

Antibiotic treatment of LRTI and CAP

More than three-quarters of LRTI patients were prescribed antibiotics by their GP on the day their illness-episode was diagnosed; 7.8% did not receive antibiotics but were hospitalised (Table 2). Over half of CAP patients were hospitalised on the day of diagnosis without a GP-prescribed antibiotic on that day (58.2%), with death on the index date without antibiotics or hospitalisation (13.0%) more common than antibiotic receipt (9.9%). A larger percentage of CAP (12.7%) than LRTI episodes (11.0%) had no antibiotic or hospitalisation records in the 29 days after the index date (Table 2). Of these CAP episodes, only 10% had received antibiotics in the previous week. Penicillins, macrolides and cephalosporins were the most commonly prescribed antibiotics on the day for both conditions (Table 3).
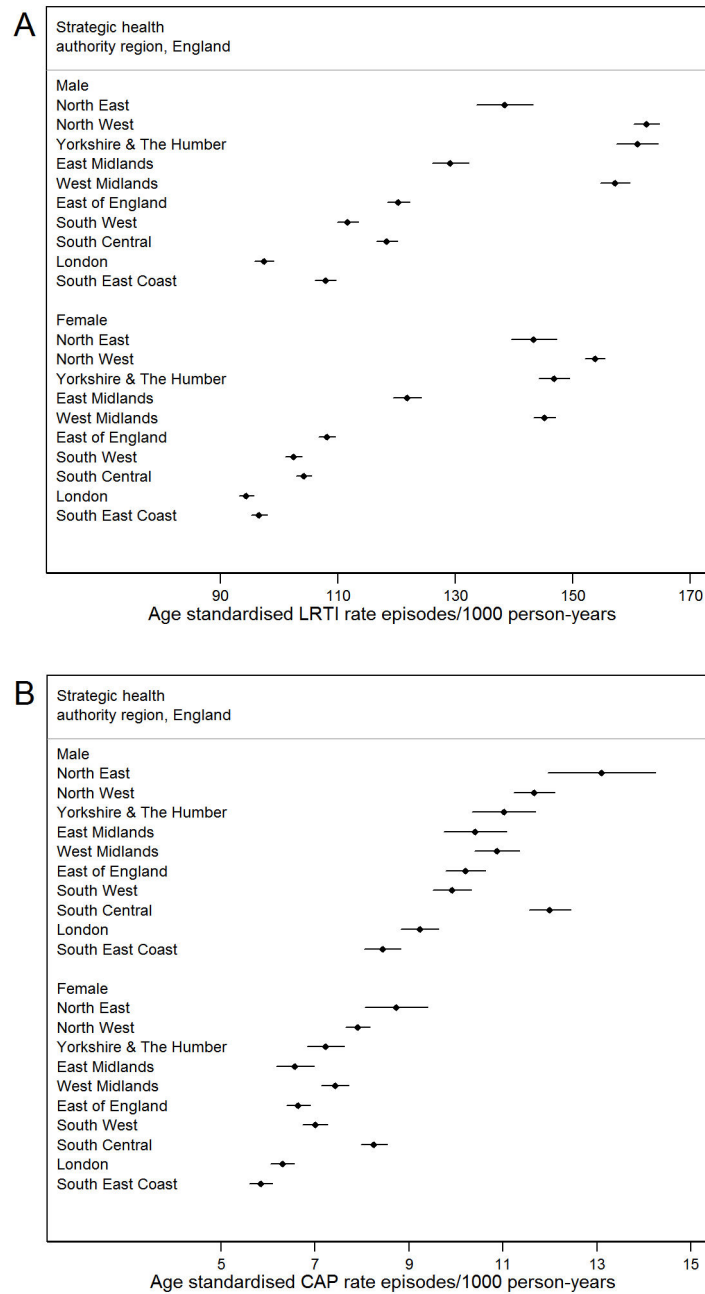
Figure 3. Age-standardised incidence of LRTI and CAP by region and sex. Age-standardised incidence of a) LRTI and b) CAP by region and sex. Standardised to UK population, mid-year 2004.

Discussion

This is the first study to provide detailed estimates of the burden of community-acquired LRTI and CAP among older UK individuals over a prolonged time period. We have shown that the incidence of LRTI and CAP increases markedly with age within this older population. Those aged 85-89 years had double the rate of LRTI and seven times more CAP illness-episodes than those aged 65-69 years, and rates were predominantly higher in men than women of the same age group. Incidence varied between regions of England, with rates in the North generally higher than the South. We also found striking differences by IMD quintile, with incidence in the most deprived quintile around 70% higher than the lowest quintile for both LRTI and CAP. There was an increase in incidence of both diseases over the study period, although rates did fluctuate somewhat. The increase was attenuated after age-standardisation, indicating that the rise was largely due to
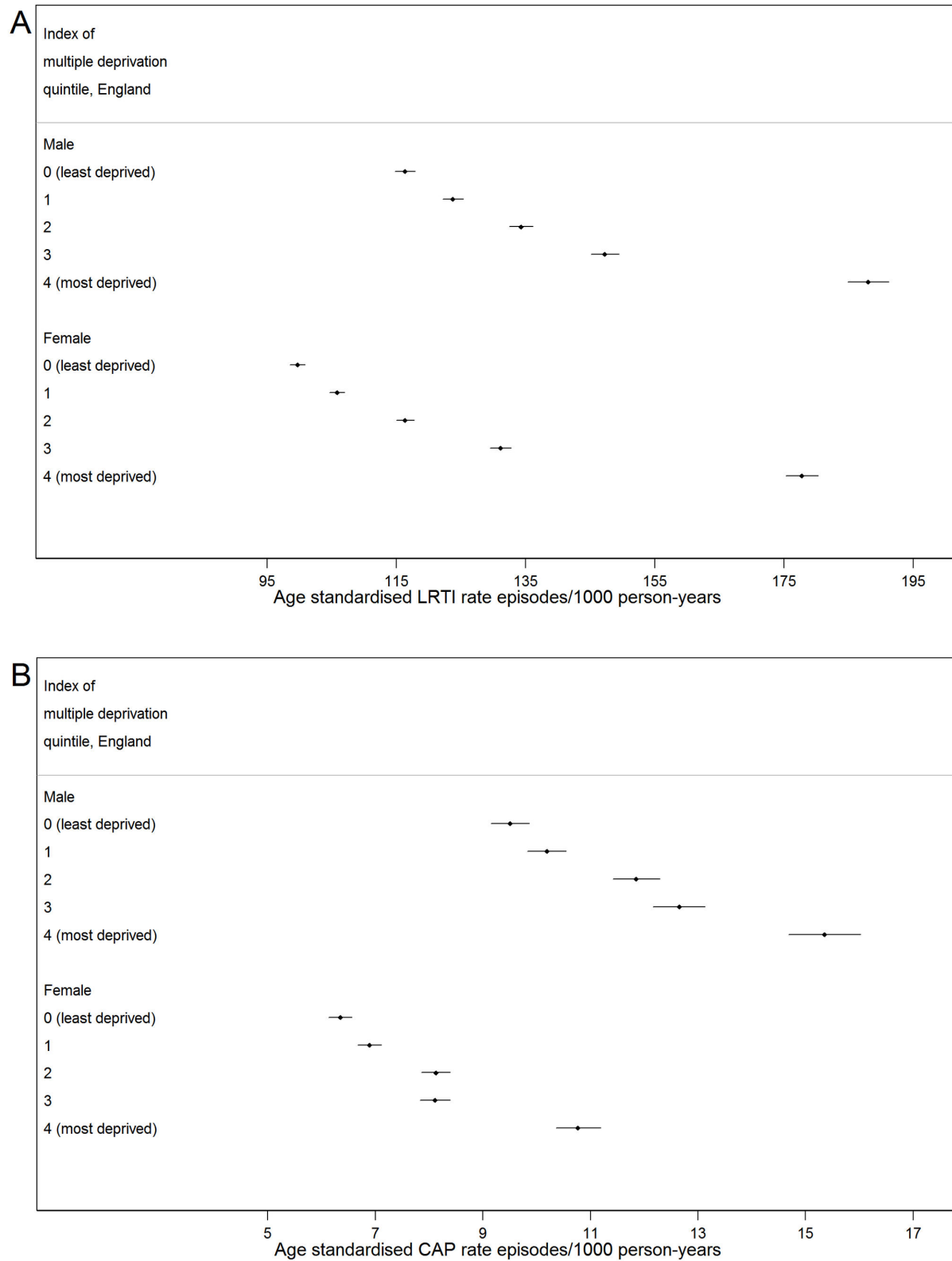
Figure 4. Age-standardised incidence of LRTI and CAP by IMD quintile and sex. Age-standardised incidence of a) LRTI and b) CAP by index of multiple deprivation quintile and sex.
Standardised to UK population, mid-year 2004.

Table 2. Timing of antibiotic prescriptions and other outcomes around the index date of LRTI and CAP illness-episodes.

Timing of treatment	Treatment received (all mutually exclusive)	LRTI		CAP	
			Antibiotics prescribed in 7 days before index date		Antibiotics prescribed in 7 days before index date
		Total (%)	n (%)	Total (%)	n (%)
On index date	Antibiotics prescribed	738794 (75.9)	22541 (3.1)	6412 (9.9)	777 (12.1)
	Patient hospitalised (no GP antibiotics)	75453 (7.8)	8013 (10.6)	37830 (58.2)	4980 (13.2)
	Patient died (no GP antibiotics/ hospitalisation)	8094 (0.8)	937 (11.6)	8476 (13.0)	1450 (17.1)
1-7 days after index date	Antibiotics prescribed	16116 (1.7)	1776 (11.0)	788 (1.2)	157 (19.9)
	Patient hospitalised (no GP antibiotics)	5461 (0.6)	671 (12.3)	961 (1.5)	168 (17.5)
	Patient died (no GP antibiotics/ hospitalisation)	722 (<0.1)	123 (17.0)	399 (0.6)	102 (25.6)
8-28 days after index date	Antibiotics prescribed	17406 (1.8)	2198 (12.6)	1155 (1.8)	208 (18.0)
	Patient hospitalised (no GP antibiotics)	4497 (0.5)	27 (9.3)	610 (0.9)	72 (11.8)
	Patient died (no GP antibiotics/ hospitalisation)	291 (<0.1)	423 (9.1)	89 (0.1)	21 (23.6)
No treatment recorded on index date or subsequent 28 days		107287 (11.0)	7355 (6.9)	8258 (12.7)	855 (10.4)
Total		974121 (100.0)	44064 (4.5)	64978 (100.0)	8790 (13.5)

doi: 10.1371/journal.pone.0075131.t002

Table 3. Variety of antibiotic (by BNF sub-chapter) prescribed on the LRTI/CAP index date.

Antibiotic variety (by BNF[a] sub-chapter)	Number of varieties prescribed	
	LRTI	CAP
	n (%)	n (%)
Penicillins	511253 (69.2)	3662 (57.1)
Cephalosporins	64609 (8.8)	610 (9.5)
Tetracyclines	36785 (5.0)	218 (3.4)
Aminoglycosides	18 (<0.1)	1 (<0.1)
Macrolides	125343 (17.0)	1665 (26.0)
Clindamycin	22 (<0.1)	3 (<0.1)
Others	130 (<0.1)	4 (<0.1)
Sulphonamides	10577 (1.4)	80 (1.3)
Metronidazole	542 (<0.1)	16 (0.3)
Quinolones	28812 (3.9)	522 (8.1)
Nitrofurantoin & Methenamine	692 (0.1)	21 (0.3)
Two or more antibiotics	25 (<0.1)	0 (0.0)
Total varieties of Antibiotic	778808 (105.4*)	6802 (106.0*)
Patients prescribed to	738794	6412

[a] British National Formulary

* Total is more than 100% as some patients were prescribed more than one variety of antibiotic

doi: 10.1371/journal.pone.0075131.t003

The derivation of illness-episodes allowed us to combine repeat consultations for one illness, giving a better measure of new illness compared to studies which included all consultations or restricted analyses to the first consultation in a year [7,8]. The incidence of LRTI we present is up to double that reported in previous UK studies [1,8]. A key factor in the age-adjusted increase in incidence in our study is likely to be the improved survival of patients with co-morbidities, resulting in a higher prevalence of patients at increased risk of infection over time. Our CAP incidence rates are almost 20% higher than previously reported in a similar UK GP population during a slightly earlier period, although with similar trends by sex and IMD [7].

Comparison of our findings with those from elsewhere in Europe are restricted by the paucity of large European studies of either LRTI or CAP set specifically in an older population, by methodological differences and limitations of some studies, and by real variation in the underlying risk profile of the populations studied. Among LRTI studies, incidence of first LRTI among 85-90 year olds in a municipality in the Netherlands during the first half of our study period estimated a considerably lower rate (93.8/1000 person years) than that we present [20]. This small study only included patients' first episode of LRTI in incidence estimates and did not exclude person-time spent in hospital, which may explain much of the difference. Estimates from the Second Dutch National Survey of General Practice (2000-2002) were also lower than ours (70/1000 person-years among those ≥75); again, person-time not at risk was not excluded, and a different coding system for LRTI was used [21].

population aging. LRTIs were most commonly treated by an antibiotic prescription from a GP, whereas over half of CAP patients did not receive a GP antibiotic prescription but were hospitalised on the illness index date.

Up to now there has been relatively little detailed information on the incidence of these infections among the UK's older population. Importantly, our exclusion of person-time not at risk, and identification of potentially hospital-acquired illness provides more accurate estimates of community-acquired infection than previously presented for UK older adults [1,7].

Rates of CAP among older adults in Europe in the last 30 years have varied widely, both between and within countries. Some of these were small regional studies, and/or were restricted to either hospital or primary care settings. For example, a Spanish cohort study (2002-2005) of individuals aged ≥65 years set in the Tarragona region which did include both outpatient and hospitalised cases estimated CAP incidence at 14/1000 person years, twice that reported in this paper; higher incidence was largely among 65-74 year olds [6]. An active surveillance program for CAP was established before the start of the Spanish study, with primary care physicians encouraged to register all CAP cases confirmed radiographically. This could have changed primary care physicians' diagnostic practices for CAP. In addition, a high proportion of individuals sought care directly from hospital and not from their general practitioner. This may have resulted in a somewhat more frequent categorisation of LRTI cases as CAP compared to our study population, in which some younger less severe cases would have been diagnosed and treated in primary care (where radiological investigations for suspected CAP are uncommon). In contrast, our CAP rates are considerably higher than those presented by another small Spanish study set in the Barcelona region in a slightly earlier study period, which estimated CAP incidence as 3.16/1000 among those aged ≥65 [22]. Neither Spanish study removed person-time not at risk of CAP nor age-standardised their overall rates. A large Italian study gave a lower CAP incidence estimate of 4.8/1000 population among those aged ≥65 years, but was restricted to hospitalised cases [23]. Interestingly, a large German study that also included only hospitalised CAP patients reported CAP incidence in those aged ≥60 of 7.65/1000 population, similar to our findings [24]. As with our study, the German and Italian studies reported consistently higher rates in men and sharply rising incidence with increasing age. Incidence of CAP was higher still in a small Finnish study of both hospitalised and non-hospitalised CAP in 1981/2 at 19.9/1000 population [25]. The difference between this finding and ours may be due in part to an earlier study period and climatic differences. Comparisons of our CAP incidence findings with those from the USA are less meaningful, as in the US patients in long-term residential care with pneumonia are not included with CAP but classified separately as having healthcare-associated pneumonia, which is not recommended practice in Europe [26].

During our study period routine vaccination for all older adults against seasonal influenza and pneumococcal disease was introduced, in 2000 and 2003 respectively. Yearly uptake of influenza vaccine has increased from 65% in 2000 and has remained between 71% and 75% since 2003 [27,28]. Coverage of pneumococcal vaccination (PPV23) has increased steadily from 29% of ≥65 year olds in 2003 to an estimated 70.5% by 31 March 2011 [29,30]. Thus our findings show rising LRTI and CAP incidence despite increasing levels of influenza and pneumococcal vaccine coverage. This might be in part because effectiveness of the PPV23 vaccine among older individuals appears to be limited [31].

As lower respiratory tract infections are caused by a number of pathogens whose circulation and severity of resulting disease varies from year to year, it is no surprise that LRTI incidence fluctuates somewhat over time. However, the peak LRTI incidence shown here in 2008 has not been reported elsewhere. National reported laboratory data for England and Wales do not show peaks in isolates of *Mycoplasma pneumoniae*, respiratory syncitial virus or influenza A or B in 2008 [32,33]. It should be noted that this peak is emphasised by the decrease in 2009, which could have resulted from low susceptibility among older individuals to the pandemic strain of influenza circulating in the 2009/10 season [34].

A high proportion of individuals with LRTI were given an antibiotic prescription on the index date, similar to a previous report from 1995–2000 [13]. It has been shown that among patients aged ≥65 presenting with an LRTI, 4% of those not prescribed antibiotics on the index date were diagnosed with pneumonia in the next month, compared to 1.5% of those prescribed antibiotics [35]. Higher antibiotic prescribing rates in older people may be to prevent worsening of LRTI or deterioration of co-morbidities, in line with recently issued guidance [36].

The increase in LRTI and CAP with age is unsurprising given age-related immunosenescence, and the growing prevalence of co-morbidities within older groups. The previously unreported regional differences between the North and South of England shown in this study may be due to interrelated socioeconomic and other factors, such as smoking and nutritional habits, as well as extent of co-morbidities and lower winter temperatures. It is notable that the regions with the highest age-standardised LRTI rates (North West, Yorkshire and the Humber, West Midlands and females in the North East) also have high reported prevalence of smoking [37]. The high level of CAP (but not LRTI) found in the South Central region was unexpected, and we cannot currently explain this; it does not appear to be due to a higher proportion of 'oldest old'. We cannot exclude the possibility that some of the regional variation in CAP is due to different diagnostic preferences geographically in categorising LRTI as CAP. However, we think it unlikely that this would explain all of the marked variation seen.

Hospitalisation without a GP antibiotic prescription was the primary intervention for 58% of CAP episodes in this study. Previous reports estimated that a third of all pneumonia patients are treated in hospital [38]; we would expect this to be higher among our older population. Penicillins and macrolides were the most prescribed antibiotics, in line with British Thoracic Society guidelines issued at the time [39]. CAP cases who died on the index date (13%) will have included some death notifications received by the GP. However, 12.7% of CAP cases had no record of treatment or death on the index date or the following four weeks, and only 10% of these patients had received antibiotics in the preceding week. Reasons for this may include high-risk patients taking previously prescribed prophylactic antibiotics, or prescription of antibiotics during a home visit that were incompletely captured in the electronic record.

Our study has many strengths, being a large, population-based study of over 1.4 million patients' primary care records, with additional information on hospital admissions for 59% of

128

patients. The addition of linked HES admission/discharge dates allowed better differentiation between potentially hospital- and community-acquired illness-episodes. It also enabled exclusion of person-time not at risk from the incidence calculation. Older adults spend more time in hospital than their younger counterparts, making this an important consideration. HES linkage was not available for the whole study population, and so we could not remove person-time at risk from the denominator of all patients. This may have led to a slight underestimation of incidence. However, we did use hospitalisation codes recorded in CPRD to exclude potentially hospital-acquired infections from the CPRD-only subset of the data.

In primary care settings in the UK, GPs often diagnose pneumonia without an x-ray, which may have led to some CAP cases being categorised as LRTI and vice-versa. Misclassification between other conditions (e.g. chronic respiratory disease) and LRTI may have occurred in a minority of patients over time, but is not likely to have favoured one condition over the other. Clinical guidelines for the diagnosis of LRTI did not change substantially during the study period but we cannot exclude the possibility that increased awareness and variation in clinical practice could have contributed in part to some of the upward trend observed. Thus our estimates reflect those of GP clinical opinion, in line with previous studies [1,7,35,40].

We used an episode structure for illnesses due to the high consultation rate among our study population [41]. The 28-day period free from LRTI consultations specified as necessary before a new episode could begin was chosen to be similar to previous UK studies, which excluded patients if they had an LRTI diagnosis up to 28 days before the index date [8,40]. It is possible that using this period excluded a few new illnesses from the numerator of our rate, and also excluded some person-time at risk. The 14-day exclusion period we placed after any hospitalisation is commonly used [14,16–19], but again may have excluded some new community-acquired episodes.

Conclusions

Community-acquired LRTI and CAP are important causes of morbidity and mortality in the aging UK population. Our new estimates show that the summary incidence of LRTI and CAP commonly presented for the ≥65 age group considerably underestimates UK disease rates in the higher ages within this group. It is important that variations in LRTI and CAP incidence in older individuals by age, region and IMD are taken into account in future health planning in the UK. Routine data such as these are used in many countries to assess disease burden. Given our findings, our methodology is likely to be highly relevant to other countries with aging populations, so that they can obtain more accurate incidence estimates of these important infections.

Supporting Information

Table S1. Community-acquired LRTI incidence rates overall and over time by sex, age, region and IMD quintile. (DOC)

Table S2. Age-standardised incidence of LRTI and CAP by year, region and IMD quintile. Standardised to UK population, mid-year 2004. (DOC)

Table S3. Community-acquired LRTI incidence including COPD exacerbations over time by sex, age, region and IMD quintile. (DOC)

Figure S1. Age standardised incidence of LRTI including COPD exacerbation codes by sex over time. Standardised to UK population, mid-year 2004. (DOCX)

Table S4. Community-acquired pneumonia incidence rates overall and over time by sex, age, region and IMD quintile. (DOC)

Author Contributions

Conceived and designed the experiments: SLT ERCM JKQ LS RD. Analyzed the data: ERCM. Contributed reagents/materials/analysis tools: SLT RD JKQ LS ERCM. Wrote the manuscript: ERCM. Interpretation of data: ERCM SLT JKQ LS RD. Critical revision of manuscript for important intellectual content: ERCM SLT JKQ LS RD. Final approval of version to be published: ERCM SLT JKQ LS RD.

References

1. Davies SC (2012) Annual Report of the Chief Medical Officer, Volume One, 2011, On the State of the Public's Health. London: Department of Health.
2. Ashton C, Bajekal M, Raine R (2010) Quantifying the contribution of leading causes of death to mortality decline among older people in England, 1991-2005. Health Stat Q: 100-127. PubMed: 20383167.
3. Eurostat (2013/3/1) Causes of death statistics - people over 65. Statistics Explained.
4. Office for National Statistics (2011) Statistical Bulletin - Older People's Day 2011.
5. Kaplan V, Clermont G, Griffin MF, Kasal J, Watson RS et al. (2003) Pneumonia: still the old man's friend? Arch Intern Med 163: 317-323. doi:10.1001/archinte.163.3.317. PubMed: 12578512.
6. Ochoa-Gondar O, Vila-Córcoles A, de Diego C, Arija V, Maxenchs M et al. (2008) The burden of community-acquired pneumonia in the elderly: the Spanish EVAN-65 study. BMC Public Health 8: 222. doi:10.1186/1471-2458-8-222. PubMed: 18582392.
7. Myles PR, McKeever TM, Pogson Z, Smith CJ, Hubbard RB (2009) The incidence of pneumonia using data from a computerized general practice database. Epidemiol Infect 137: 709-716. doi:10.1017/S0950268808001428. PubMed: 18840320.
8. Macfarlane J, Holmes W, Gard P, Macfarlane R, Rose D et al. (2001) Prospective study of the incidence, aetiology and outcome of adult lower respiratory tract illness in the community. Thorax 56: 109-114. doi:10.1136/thorax.56.2.109. PubMed: 11209098.

129

9. Office for National Statistics (2000) Key Health Statistics from General Practice 1998 London. Office for National Statistics.

10. Lewis JD, Bilker WB, Weinstein RB, Strom BL (2005) The relationship between time since registration and measured incidence rates in the General Practice Research Database. Pharmacoepidemiol Drug Saf 14: 443-451. doi:10.1002/pds.1115. PubMed: 15898131.

11. Sapey E, Stockley RA (2006) COPD exacerbations . 2: aetiology. Thorax 61: 250-258. doi:10.1136/thx.2005.041822. PubMed: 16517585.

12. Gulliford M, Latinovic R, Charlton J, Little P, van Staa T et al. (2009) Selective decrease in consultations and antibiotic prescribing for acute respiratory tract infections in UK primary care up to 2006. J Public Health (Oxf) 31: 512-520. doi:10.1093/pubmed/fdp081. PubMed: 19734168.

13. Ashworth M, Charlton J, Latinovic R, Gulliford M (2006) Age-related changes in consultations and antibiotic prescribing for acute respiratory infections, 1995-2000. Data from the UK General Practice Research Database. J Clin Pharm Ther 31: 461-467. doi:10.1111/j.1365-2710.2006.00765.x. PubMed: 16958824.

14. Ashworth M, Latinovic R, Charlton J, Cox K, Rowlands G et al. (2004) Why has antibiotic prescribing for respiratory illness declined in primary care? A longitudinal study using the General Practice Research Database. J Public Health (Oxf) 26: 268-274. doi:10.1093/pubmed/fdh160.

15. Ashworth M, Charlton J, Ballard K, Latinovic R, Gulliford M (2005) Variations in antibiotic prescribing and consultation rates for acute respiratory infection in UK general practices 1995-2000. Br J Gen Pract 55: 603-608. PubMed: 16105368.

16. Schnoor M, Hedicke J, Dalhoff K, Raspe H, Schäfer T (2007) Approaches to estimate the population-based incidence of community acquired pneumonia. J Infect 55: 233-239. doi:10.1016/j.jinf.2007.04.355. PubMed: 17599417.

17. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N et al. (2003) Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax 58: 377-382. doi:10.1136/thorax.58.5.377. PubMed: 12728155.

18. Bont J, Hak E, Hoes AW, Schipper M, Schellevis FG et al. (2007) A prediction rule for elderly primary-care patients with lower respiratory tract infections. Eur Respir J 29: 969-975. doi: 10.1183/09031936.00129706. PubMed: 17215313.

19. Gutiérrez F, Masiá M, Mirete C, Soldán B, Rodríguez JC et al. (2006) The influence of age and gender on the population-based incidence of community-acquired pneumonia caused by different microbial pathogens. J Infect 53: 166-174. doi:10.1016/j.jinf.2005.11.006. PubMed: 16375972.

20. Sliedrecht A, den Elzen WPJ, Verheij TJM, Westendorp RGJ, Gussekloo J (2008) Incidence and predictive factors of lower respiratory tract infections among the very elderly in the general population. The Leiden 85-plus Study. Thorax 63: 817-822. doi: 10.1136/thx.2007.093013. PubMed: 18388206.

21. Hak E, Rovers MM, Kuyvenhoven MM, Schellevis FG, Verheij TJ (2006) Incidence of GP-diagnosed respiratory tract infections according to age, gender and high-risk co-morbidity: the Second Dutch National Survey of General Practice. Fam Pract 23: 291-294. doi:10.1093/fampra/cmi121. PubMed: 16464869.

22. Almirall J, Bolíbar I, Vidal J, Sauca G, Coll P et al. (2000) Epidemiology of community-acquired pneumonia in adults: a population-based study. Eur Respir J 15: 757-763. doi:10.1034/j.1399-3003.2000.15d21.x. PubMed: 10780770.

23. Giorgi Rossi P, Agabiti N, Faustini A, Ancona C, Tancioni V et al. (2004) The burden of hospitalised pneumonia in Lazio, Italy, 1997-1999. Int J Tuberc Lung Dis 8: 528-536.

24. Ewig S, Birkner N, Strauss R, Schaefer E, Pauletzki J et al. (2009) New perspectives on community-acquired pneumonia in 388 406 patients.

Results from a nationwide mandatory performance measurement programme in healthcare quality. Thorax 64: 1062-1069. doi:10.1136/thx.2008.109785. PubMed: 19454409.

25. Jokinen C, Heiskanen L, Juvonen H, Kallinen S, Karkola K et al. (1993) Incidence of Community-Acquired Pneumonia in the Population of Four Municipalities in Eastern Finland. Am J Epidemiol 137: 977-988. PubMed: 8317455.

26. Woodhead M, Blasi F, Ewig S, Garau J, Huchon G et al. (2011) Guidelines for the management of adult lower respiratory tract infections - Summary. Clin Microbiol Infect 17: 1-24. doi:10.1111/j.1469-0691.2010.03448.x.

27. Joseph C, Goddard N (2003) Influenza vaccine uptake in the elderly: results from a rapid assessment of the effectiveness of new government policy in England for the winters 2000/2001 and 2001/2002. [Erratum appears in Vaccine (2003 Sep 8);21(25-26): 1137. Vaccine 21: 1148. PubMed: 12559791.

28. Public Health England (2013) Influenza vaccination uptake monitoring on behalf of the Department of Health. Available: http://www.hpa.org.uk/web/HPAweb&HPAwebStandard/HPAweb_C/1195733756886. Accessed 24 July 2013

29. Noakes K, Pebody RG, Gungabissoon U, Stowe J, Miller E (2006) Pneumococcal polysaccharide vaccine uptake in England, 1989–2003, prior to the introduction of a vaccination programme for older adults. J Public Health (Oxf) 28: 242-247. doi:10.1093/pubmed/fdl017.

30. Begum F, Pebody R (2012) Pneumococcal Polysaccharide Vaccine (PPV) uptake report. Department of Health.

31. Joint Committee on Vaccination and Immunisation (2011) JCVI statement on the routine pneumococcal vaccination programme for adults aged 65 years and older 20 July 2011.

32. Public Health England (2013) Six major respiratory pathogens reported from PHE and NHS laboratories (LabBase) in England and Wales between weeks 01/2004 and 30/2013 (3-week moving average). Available: http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1247728930779. Accessed 24 July 2013

33. Public Health England (2013) Laboratory reports to HPA Colindale of infections due to Mycoplasma pneumoniae England and Wales by date of report 1991-2013 (4 weekly). Available: http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1194947359371. Accessed 24 July 2013

34. Health Protection Agency (2010) Epidemiological report of pandemic (H1N1) 2009 in the UK.

35. Petersen I, Johnson AM, Islam A, Duckworth G, Livermore DM et al. (2007) Protective effect of antibiotics against serious complications of common respiratory tract infections: retrospective cohort study with the UK General Practice Research Database. BMJ 335: 982. doi:10.1136/bmj.39345.405243.BE. PubMed: 17947744.

36. NICE Short Clinical Guidelines Technical Team (2008) Respiratory tract infections – antibiotic prescribing. Prescribing of antibiotics for self-limiting respiratory tract infections in adults and children in primary care. London: National Institute for Health and Clinical Excellence.

37. Office for National Statistics (2012) eneral lifestyle survey overview: A report on the 2010 general lifestyle survey.

38. Holmes WF, Woodhead M (1999) Providing better care for patients who may have pneumonia. Thorax 54: 925-928. doi:10.1136/thx.54.10.925. PubMed: 10491457.

39. British Thoracic Society Standards of Care Committee (2001) BTS Guidelines for the Management of Community Acquired Pneumonia in Adults. Thorax 56 (Suppl 4): IV1-I64. PubMed: 11713364.

40. Winchester CC, Macfarlane TV, Thomas M, Price D (2009) Antibiotic prescribing and outcomes of lower respiratory tract infection in UK primary care. Chest 135: 1163-1172. doi:10.1378/chest.07-2940. PubMed: 19420194.

41. Hippisley-Cox J, Vinogradova Y (2009) Trends in Consultation Rates in General Practice - 1995-2009.

130

4.5 Further discussion around the use of combined stand-alone CPRD and HES-linked CPRD data in one study

These analyses utilised a combination of stand-alone CPRD records (40.3% of cohort) and HES-linked (59.7%) CPRD records. Rather than limiting the analyses to using HES-linked data only, it was decided to also include information from unlinked patients in order to provide a very large study population, enabling the calculation of detailed incidence estimates finely stratified by age group and sex over time.

The two data sources contained a wealth of information, but neither were simple to use, as expected when using data whose primary purpose is for clinical rather than research use. In addition to the data management considerations (outlined in sections 2.3.3 and 2.4), the possible consequences of using the two data sources in a single analysis should not be overlooked. Below I first outline the different strategies used, followed by how these could have affected the estimates I have presented.

4.5.1 Different methods of capturing hospitalisation records

The use of stand-alone CPRD data assumed accurate and timely recording by general practices of hospitalisations, thus enabling differentiation between HAP and CAP. Inspection of completeness and timeliness of recording of hospitalisations in CPRD data is being carried out as part of a separate ongoing investigation (Sara Thomas, personal communication). Findings so far indicate that hospitalisations were recorded in CPRD on the HES date of admission, the HES date of discharge, the date the discharge summary was received by the practice, on another date or not at all.

My analyses above also assumed that GP records included the reason patients were admitted to hospital (as notified via hospital discharge summaries), enabling hospitalisations due to pneumonia to be identified. In contrast, all inpatient hospitalisations are recorded in HES, and cases of pneumonia are identifiable in HES data by looking at the reason for admission (which I assumed to be captured in the primary code of the first episode in a spell).

4.5.2 Different methods for calculating person-time at risk

The methods used to define person-time at risk were applied slightly differently in the linked and unlinked data. As outlined in section 2.5.2, the inclusion of dates of hospital admission and discharge for HES-linked patients allowed the exclusion of the time these patients were in hospital, as well as the 14 days after discharge when patients were ineligible for a community-acquired infection. The exclusion of time patients spent admitted to hospital was not possible in stand-alone CPRD data, as admission and discharge dates were not recorded. Additionally, since the codes used to record hospitalisation in CPRD data and the recorded dates of these hospitalisations had not been validated at the time of the analyses, I decided not to exclude 14 days from the denominator after a CPRD hospitalisation code (as described in section 2.5.2.1). My reasoning for this was if the codes for hospitalisation in stand-alone CPRD were incorrect or recorded on the wrong date, episodes of community-acquired illness would have been wrongly labelled as hospital-acquired and excluded, resulting in an underestimate of the incidence. Excluding the 14 days after a CPRD hospital record from the person-time included in the denominator would have decreased the size of the denominator, thus increasing the size of the estimate. The resulting rates would have been difficult to interpret due to the contrasting direction of the effects of these mis-categorisations. Instead, I decided to be cautious and to exclude only the events but retain the person-time, and produce conservative estimates of incidence.

4.5.3 Potential implications of these differences between stand-alone and linked data

The exclusion of person-time from linked data and extent of completeness of recording of hospitalisations by general practices are important when considering the impact of using both stand-alone and linked data in a single analysis. The potential implications of these differences in the two data sources are summarised in Figure 4-2 below.

In a best case scenario (box 1 in Figure 4-2), we assume that hospital admissions for CAP (when the patient presented directly to hospital, and had not initially consulted their GP) were fully recorded by the GP via information received in a hospital discharge summary. In this scenario, CAP incidence estimates from linked and from stand-alone data would differ only by the amount of person-time that could be excluded during and

in the 14 days after a hospitalisation when using the linked data. The inability to exclude this time when using the stand-alone data would have resulted in the denominator of the incidence calculation being too large, and thus a slight underestimation of incidence.

Figure 4-2 Use of stand-alone CPRD data when calculating incidence estimates: effect of four scenarios
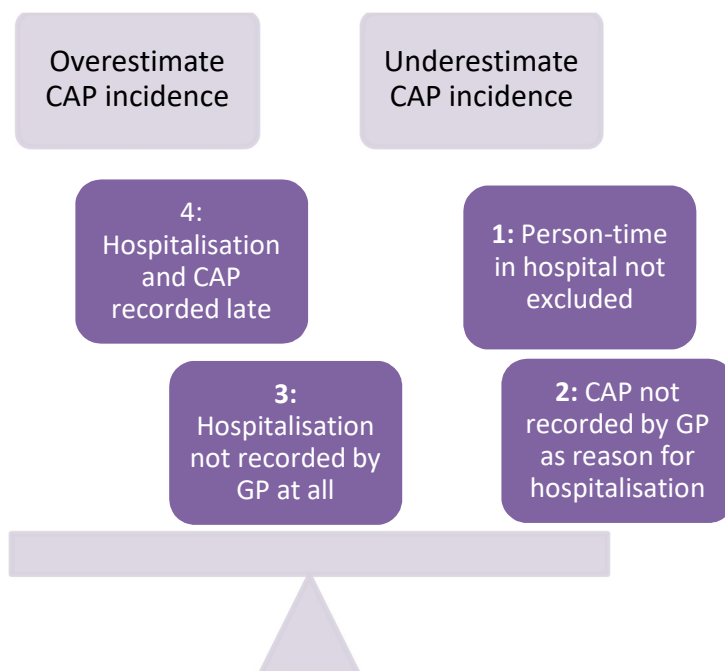
In scenario 2, GPs recorded that patients had been admitted to hospital (allowing differentiation of community- and hospital-acquired illness) but did not record pneumonia as the reason for the hospitalisation. This would have led to a reduction in the number of pneumonia episodes recorded in the stand-alone data (box 2 Figure 4-2), reducing the numerator in the incidence calculation. This, in addition to the over-inclusion of person-time at risk (in scenario 1) would result in further underestimation of CAP incidence.

Thirdly, if a patient's discharge summaries were not recorded in their GP records at all (scenario 3, box 3 Figure 4-2), it would also not be possible to tell whether GP-diagnosed pneumonias were community- or hospital-acquired. This would result in the inclusion of some HAP episodes as CAP, inflating the rates previously underestimated due to lack of removal of person-time.

Finally, a single CAP episode could be recorded twice due to the patient consulting their GP, subsequently being hospitalised and the hospital admission (and CAP diagnosis) being recorded post-discharge in the patient's general practice record more than 28 days after the initial GP consultation. This would also inflate the incidence rates produced (scenario 4, box 4 Figure 4-2).

In reality the recording of hospitalisation and reasons for admission are likely to vary between general practices, resulting in a mix of scenarios one to four above. It is probable that the exclusion of person-time and the more complete recording of events in the linked data would lead to the incidence rates calculated using these data being consistently higher than those produced using the stand-alone CPRD data. Consequently, although the characteristics of the HES-linked patients were similar to those of the cohort overall (Table 1 Paper 1), it does not follow that the incidence estimates produced by stand-alone and linked data would be equal. The different method of excluding person-time not at risk and possible inaccuracies in recording of hospitalisations in GP data are likely to have affected CAP estimates more than those for all LRTI, as the majority of LRTIs are not severe enough to require hospitalisation and so are solely seen and recorded by a GP. However, more than 58% of CAP episodes had a hospital admission record in either CPRD (stand-alone data) or HES (linked data) on the index date (Table 2 Paper 1).

4.5.4 Further questions raised by the use of linked data

It is unclear how much added value is gained by using linked primary and secondary care data compared to primary care data alone. Do the incidence estimates from the linked data justify its use given the more complex and time consuming analysis required? If the ratio of rates from the two data sources remained constant over time, would it be possible to approximate results from the linked-data using the stand-alone CPRD?

I therefore devised an investigation to try to answer these questions, by comparing CAP incidence estimates from HES-linked and stand-alone CPRD data. A full account of this is presented in the next Chapter.

Chapter 5   A comparison of CAP incidence estimates from stand-alone CPRD records versus CPRD HES-linked data in the same cohort of patients

5.1   Background

In the previous chapter I presented CAP incidence estimates which were estimated using records from CPRD HES-linked data for patients eligible for linkage, and from stand-alone CPRD data for patients who were not. I then discussed the possible ramifications of using both stand-alone and linked records in one analysis. In this chapter I set out to answer some of the questions raised.

While the value of linked over stand-alone data has been investigated for conditions such as cardiovascular events, asthma, diabetes, and upper gastrointestinal bleeding,[139-142] the benefits of using linked data to examine the burden of important infections such as pneumonia are less well known. In order to investigate the disparity between estimates produced using primary care data and those from linked primary-secondary records, I undertook a direct comparison of the incidence of CAP in a single group of patients using each of the two data sources in turn. Using records for the same group of patients ensured that any differences in CAP incidence were due to the data type used, rather than underlying differences in the characteristics of the cohort.

In this chapter I present the results of this comparison. A version of this work which focuses more on the implications of the findings for future work on infections in older adults, rather than on the effect on CAP estimates has been submitted for publication to the Journal of Clinical Epidemiology (see Appendix E). In the submitted version, population averaged models were used to adjust for the clustering of CAP episodes within patients. Here, I use a random effects analysis to take account of the clustering, so that the methods were consistent with those in Chapter 4. As discussed in section 2.8, both methods are commonly used for these types of analyses, and the method of adjusting for clustering is the only change between the two analyses presented – the remaining methods, discussion and overall conclusions do not differ.

5.2 Methods

5.2.1 Study populations and period of follow-up

Both practices and patients joined CPRD throughout the study period, and thus analyses were performed on dynamic cohorts. In order for the two data sources to be comparable, a near-identical group of patients were used in both analyses. Patients included in the study were eligible for record linkage, were aged ≥65 years and contributed ≥1 day of follow-up. Follow-up started at the latest of the study start date (1st April 1997), the patient's 65th birthday, the UTS date or 28 weeks after patient registration. Follow-up ended at the earliest of the study end date (31st March 2011), death, last collection date of information from CPRD or the date the patient left the practice.

Total incidence of CAP included both GP-diagnosed and hospital-diagnosed cases (either identified in CPRD or HES). I did not include a 'HES only' group, as that would examine the rate of hospitalised CAP rather than CAP incidence.

5.2.2 Defining CAP episodes

Pneumonia illness episodes, community-acquired infections and exclusion of person-time were identified as outlined in section 2.4, and as used in the analyses in presented in Chapter 4.

In both the analysis of the stand-alone CPRD data, and the analysis of the linked CPRD-HES data, pneumonia episodes with no record of hospitalisation in the 14 days before the incident date were classed as community-acquired. As with the incidence analyses in Chapter 4, episodes which started within 14 days of a hospitalisation were assumed to be hospital-acquired (HAP) and were not included as incident events. The method for defining hospitalisation, and thus distinguishing between CAP and HAP, differed between the two analyses as outlined in section 2.4.2. In the stand-alone data, records for hospitalisation were defined using Read codes and relevant fields in the consultation, referral and clinical GP files, and the date of these records were used as the start of the 14 day period for categorising a pneumonia episode as HAP. In the linked cohort the 14 day period started at the discharge date of any hospital admission.

5.2.3 Defining person-time at risk

Patients were not considered 'at-risk' of pneumonia during any pneumonia episode or for the 28 days after the last code in the episode, and this time was excluded from the denominator in both cohorts (see section 2.5.2). As described in Chapter 4, a key difference in the analysis of the linked data was the capacity to also exclude the duration of any hospital admission and the subsequent 14 days from person-time at risk of a community-acquired infection, which is not replicated in the stand-alone data due to the absence of admission and discharge dates in these records.

5.2.4 Statistical methods

As in Chapter 4, Poisson models with random effects were used to calculate the incidence of CAP across clusters of CAP episodes per patient.

Rates were calculated stratified by financial year, age group and sex. The percentage increase in estimates when using the linked data was calculated using Equation 5-1.

Equation 5-1 Percentage increase in CAP incidence estimates when using linked data:

$$\left(\frac{CAP\ incidence\ rate\ in\ linked\ data}{CAP\ incidence\ rate\ in\ standalone\ data} \times 100 \right) - 100$$

Within the linked data I was also able to attempt to examine pathways of care. To do this, I examined whether patients had consulted with a GP (either face to face or by telephone) on the CAP incident date, using the 'constype' field in the CPRD consultation file. I additionally checked whether there were any consultations for LRTI (excluding CAP) in the 28 days prior to the CAP incident date.

5.3 Results

5.3.1 Study population and number of CAP episodes identified in each dataset

The study population included 917,852 patients in the stand-alone data from 351 practices across England. The linked analysis included 916,128 of these patients who had ≥1 day of follow-up after additionally excluding person-time in hospital (which resulted in exclusion of <0.2% of individuals from the linked analysis, Table 5-1). In both analyses the majority of patients were aged 65-69 years at their start of study (53%) and

over half (56%) were female (Table 5-1). When only GP records were used, 31,575 CAP episodes were identified between 1997 and 2010. Using linked GP and hospital admission data identified 45,285 CAP episodes (Table 5-2). The distribution of CAP episodes by age group and sex did not differ particularly between the two analyses. However, while the number of CAP episodes recorded increased over time in both data sources, the increase was considerably more pronounced in the linked data (Table 5-2).

Table 5-1 Characteristics of the study populations in each analysis

	Number of patients (%)	
	CPRD only	CPRD HES-linked
Total	917852	916128
Male	403191 (43.9)	402474 (43.9)
Age group		
65-69	487509 (53.1)	487189 (53.2)
70-74	141600 (15.4)	141398 (15.4)
75-79	117973 (12.9)	117693 (12.8)
80-84	85183 (9.3)	84828 (9.3)
85-89	54853 (6.0)	54531 (6.0)
≥90	30734 (3.3)	30489 (3.3)
Number of CAP episodes		
1	26805	37040
2	1848	3054
3	253	479
4	48	106
5	11	32
6	10	12
7	0	4
8	1	2

Table 5-2 Number of CAP events recorded in each analysis

	Number of CAP episodes (%)	
	CPRD only	CPRD HES-linked
Total	31575	45285
Male	14551 (46.1)	21085 (46.6)
Age group		
65-69	3250 (10.3)	4742 (10.5)
70-74	4280 (13.6)	6216 (13.7)
75-79	5677 (18.0)	8008 (17.7)
80-84	6632 (21.0)	9468 (20.9)
85-89	6159 (19.5)	9027 (19.9)
≥90	5577 (17.7)	7824 (17.3)
Year		
1997	1611 (5.1)	1801 (4.0)
1998	1894 (6.1)	2174 (4.8)
1999	2055 (6.5)	2422 (5.3)
2000	1929 (6.1)	2330 (5.1)
2001	2148 (6.8)	2760 (6.1)
2002	2264 (7.2)	2908 (6.4)
2003	2540 (8.0)	3272 (7.2)
2004	2522 (8.0)	3431 (7.6)
2005	2442 (7.7)	3421 (7.6)
2006	2302 (7.3)	3578 (7.9)
2007	2363 (7.5)	3860 (8.5)
2008	2571 (8.1)	4536 (10.0)
2009	2491 (7.9)	4351 (9.6)
2010	2443 (7.7)	4441 (9.8)

5.3.2 Incidence estimates

Incidence estimates using linked data were higher than those using stand-alone data (Figure 5-1). Overall, the incidence was 48.9% higher using the linked-data, and the difference increased markedly over time from 13% higher in 1997, to 101% higher in 2010.

Figure 5-1 Incidence of CAP amongst older adults by data source over time
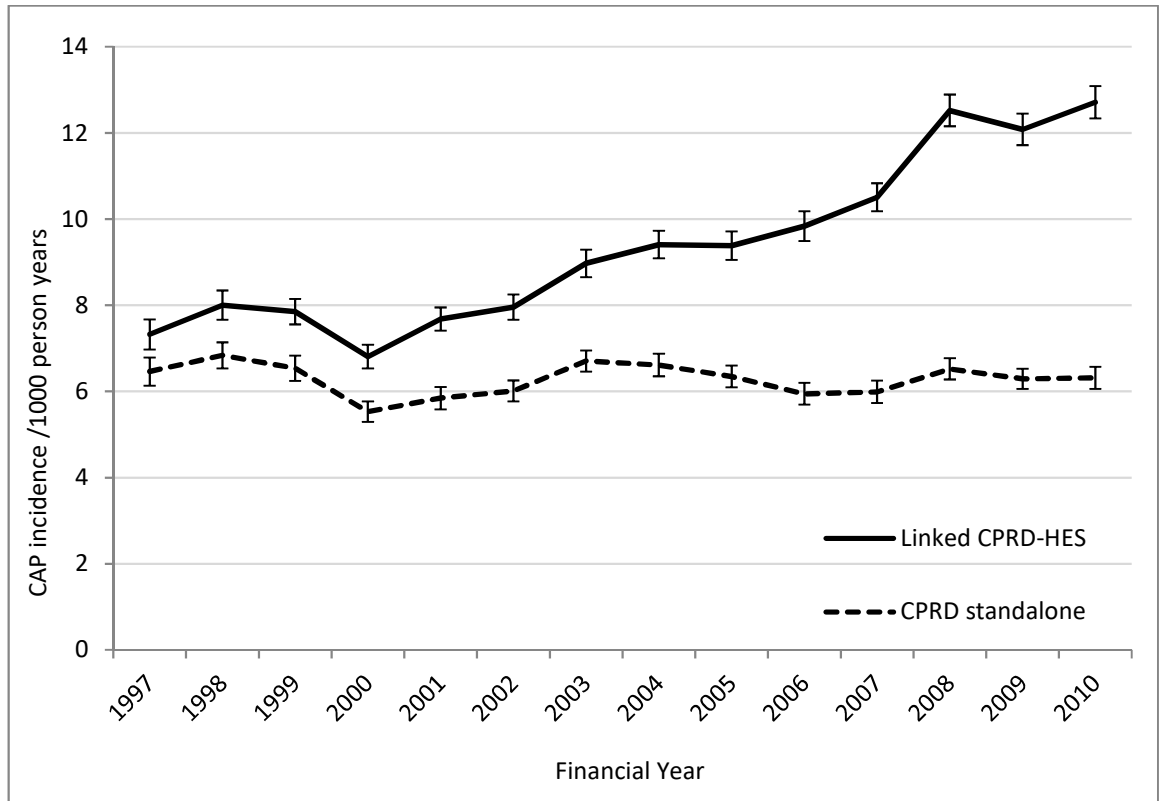
While CAP rates rose with age in both data sources, the relative increase in CAP estimates using the linked compared to GP stand-alone data was comparable across age groups, therefore the disparity between the two data sources could not be attributed to a specific age group (Figure 5-2). CAP incidence was higher in men than women in both analyses, but the divergence between estimates was observed at similar levels in both sexes (Figure 5-3).

Figure 5-2 Percentage increase in CAP estimates when using linked-data compared to stand-alone GP records, by age group
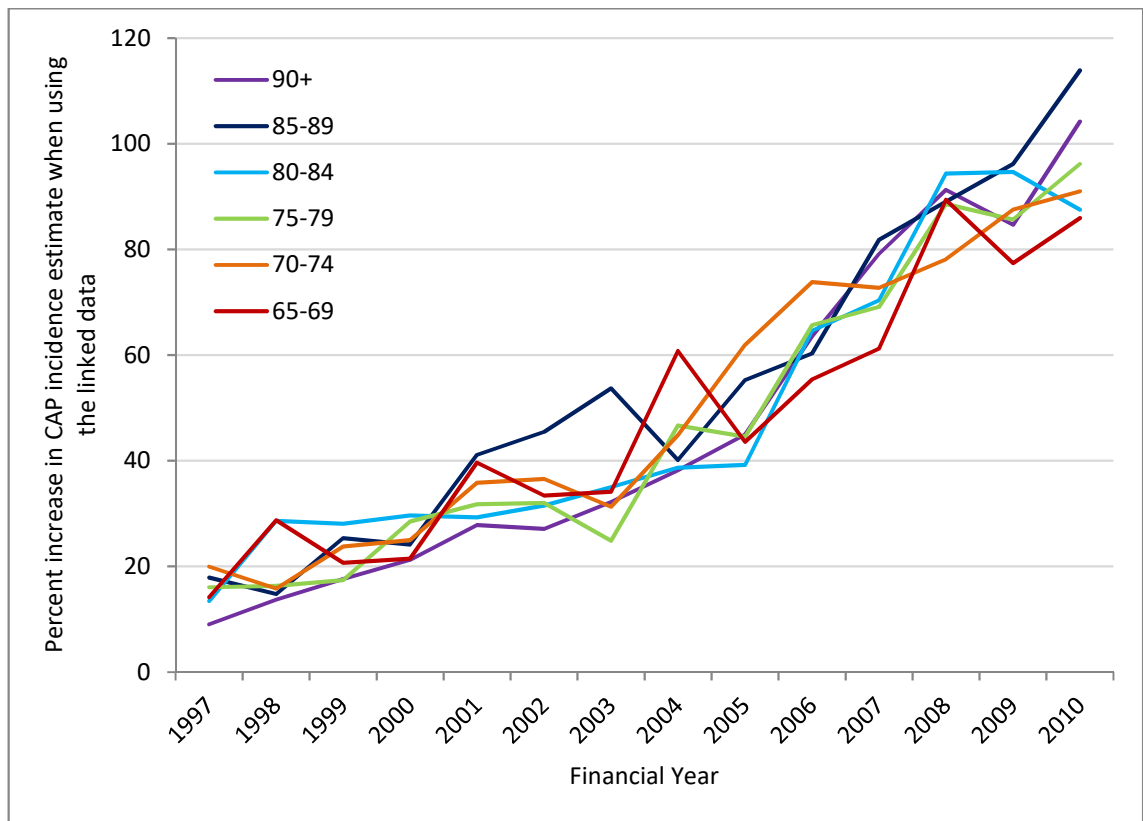
Figure 5-3 Percentage increase in CAP estimates when using linked-data compared to stand-alone GP records, overall and by sex
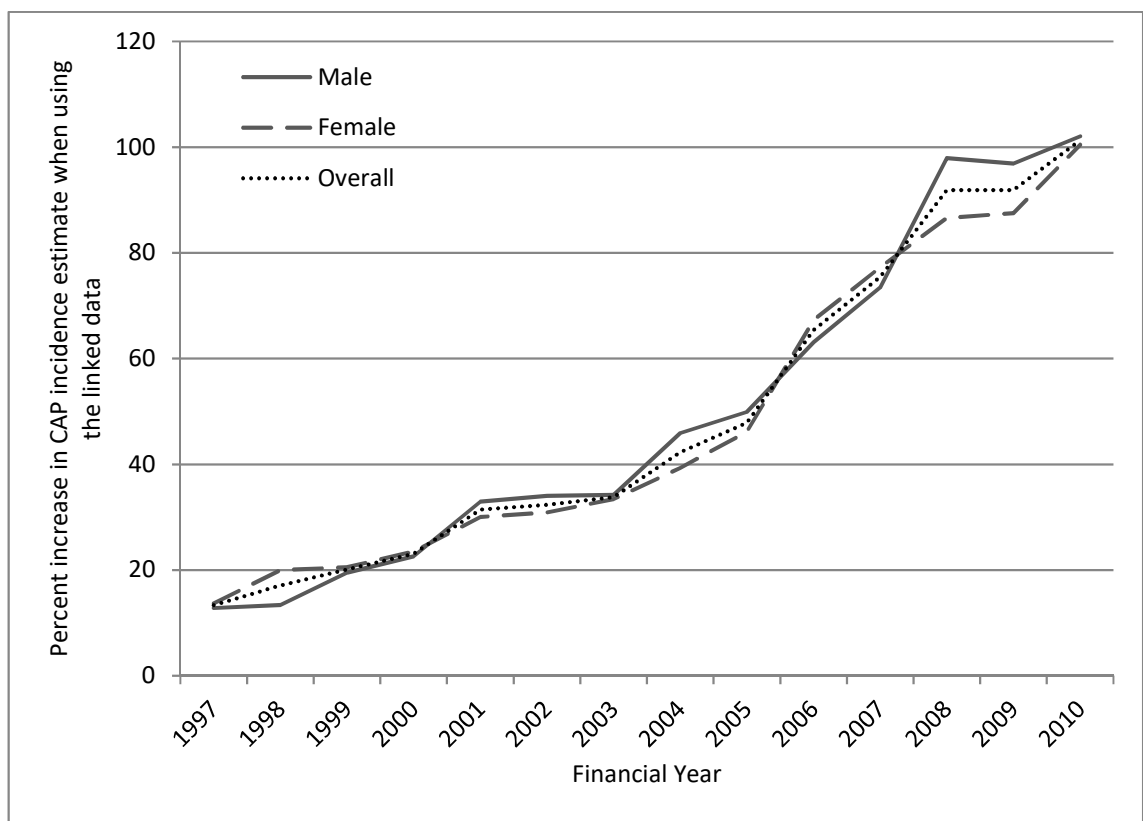
Due to the dynamic nature of the cohort, the number of patients contributing to each analysis increased over the study period, increasing the person-time included. However, while the increase in person-time within each analysis was similar (91% increase in linked vs. 93% in stand-alone), the increase in CAP episodes was substantially larger in the linked data (147% vs 52%, Table 5-2).

Between 1997 and 2010, the percentage of patients who had consulted with their GP on the day of the CAP diagnosis decreased from 82% to 43%. Over the same period, consultation with a GP for an LRTI in the 28 days prior to the CAP diagnosis also decreased, from 15% to 10% of CAP episodes.

5.4 Discussion

CAP incidence estimates from linked primary and secondary data were nearly 50% higher overall than those from primary care data alone. The divergence between the estimates increased appreciably over the 14 year study period, and CAP incidence using linked data was double the estimate from stand-alone GP records by March 2011.

As discussed in Chapter 4, the analysis using linked data permitted a different approach to defining person-time at risk of community-acquired infection, as I could discount the person-time patients were admitted to hospital. This also led to the exclusion of 1724 patients (<0.2%) whose HES records revealed that they were admitted to hospital or recently discharged from hospital (within the prior 14 days) for their entire period of follow-up, and so were not considered at risk of a community-acquired infection. However, from examining the change in crude number of CAP events and in the person-time included in each analysis, it seems that the higher number of CAP episodes recorded in the linked data explains the divergence. As previously discussed, all pneumonia events recorded in GP records are included in the linked data, but pneumonias from hospital admissions are only included in the stand-alone data if they are retrospectively recorded by the patient's general practice. It is apparent that not all CAP identified in hospital are then recorded using a Read code for pneumonia in GP records. However, this under-recording of hospitalised CAP in primary care data could not explain the divergence shown on its own. Additionally, hospital admissions due to pneumonia would have needed to increase over the study period (and be captured in the linked-data) in order to explain the divergence seen over time.

There may be more than one explanation for increasing hospital admissions for CAP in our study population. For example, patients may be increasingly presenting directly to hospital rather than consulting their GP, due to changes in service provision or perceived severity of illness. This is compatible with the decreasing trend in CAP episodes that had a GP consultation on the day of the CAP diagnosis. Alternatively, the threshold for admission for older patients with CAP presenting to Accident and Emergency Departments may have decreased. Both of these scenarios would be consistent with the larger increase in CAP episodes in the HES records. Increasing CAP hospitalisations over the study period and the reasons behind this trend are the focus of the next Chapter, and are discussed further there.

5.4.1 Study strengths

The analyses used large, nationally representative datasets which contained ≥900,000 patients. Over 99.8% of patients were included in both analyses, thus virtually excluding any other factors from having an effect on the discrepancy reported, and enabling examination of the differences in estimates of CAP solely due to the type of data and methodology used.

5.4.2 Potential limitations

The two data sources use different coding systems, and coding practices may have varied over time within each data source. Read codes (used in CPRD) include 'tentative' pneumonia codes such as 'Influenza or pneumonia', but equivalent codes do not exist in the ICD-10 coding system (used in HES). The tentative pneumonia codes were not included as pneumonia in this study. Patients who consulted their GP and were given this code, but were subsequently hospitalised with a CAP diagnosis would have been included as a CAP episode in the linked data analysis, but not in that of the stand-alone data. However, to have contributed to the disparity, these tentative diagnoses would have needed to be increasingly used over time, whereas general consultations with a GP on the date of a CAP diagnosis, or for less serious LRTI within the previous 28 days both decreased over the study period. Alternatively, if there was an increasing tendency over time for hospital physicians to diagnose and label older patients as having pneumonia (for example due to increasing use of CT scans), this would contribute to the divergent trends. There is no direct evidence that this occurred, but a clear understanding of

trends in clinical and coding practices is essential for interpretation of findings from both stand-alone and linked data.

5.4.3 Implications of the findings

5.4.3.1 Implications for CAP incidence estimates presented in Chapter 4

The CAP incidence estimates presented in Chapter 4 were produced using both stand-alone CPRD records (40.3% of patients over the entire study period) and CPRD HES-linked data, in order to get the best estimates from the whole CPRD population. In section 4.5.4, I hypothesised that if the discrepancy between linked and stand-alone data remained constant over time, it may be possible to use stand-alone data and somehow adjust estimates for the increase that would be expected if linked data had been used. The results presented here show that this is inadvisable, as the discrepancy between the record types varied so considerably over the study period.

Comparison with the results from the mixed data sources (overall incidence presented in Appendix D) shows the rates from the mixed data were closer to the estimates from the HES-linked data than the stand-alone data in these analyses. However, use of the results from this work to predict the level of underestimation reported in Chapter 4 is not straightforward. CPRD-HES linkage is currently only available for English CPRD practices, thus CPRD practices from Scotland, Wales and Northern Ireland (who comprised 20% of the patients who contributed to the incidence analyses in Chapter 4) are automatically ineligible for HES-linkage. These populations may have different age structures, underlying health problems and health behaviours (such as smoking status) to patients in England, and thus may differ slightly from the HES-linked patients included in this study population. Additionally, devolution of the NHS has led to each country controlling its own service provision, which may affect generalisability between the nations.

Secondly, the number of practices that consented to HES-linkage changed throughout the study period. This may have resulted in a varying amount of possible underestimation due to the stand-alone data over time.

Finally, even if results are limited to England, there is no clear gold standard between CPRD and HES without validation of diagnoses in both data sources. While CAP may be

under-recorded in CPRD, it is also possible that CAP was increasingly ascertained in HES, or that there is an increasing trend in labelling patients as having CAP. This will be discussed further in the next chapter.

5.4.3.2 Implications of accounting for clustering using random-effects versus population-averaged model

The analyses presented in this Chapter were repeated for publication using a population-averaged (PA) model (Appendix E). It is important to consider the different interpretation of the two measures when thinking about the implications of their use on the results shown. Random effects (RE) models produce subject-specific effects – the incidence estimates presented above are for the rate of CAP in the mean patient in the population under study. The PA models represent the average rate of CAP *across* the patients in the population under study. If there was no clustering of CAP at a patient level, the PA and RE models would produce the same estimates. When clustering is present, PA models have been shown to produce consistently lower results than those from random effects RE models.[114] However, the ratio between the parameter estimate and SE is similar in both models, and so p-values from the models are generally consistent. Thus, while the estimates presented in Appendix E and the percentage increase in incidence when using linked compared to stand-alone data are slightly lower than those presented here, both sets are equally valid (as long as they are interpreted correctly), and show the same general trend.

5.4.3.3 Implications for future research

This study provides evidence that use of primary or secondary care data in isolation may not give accurate incidence estimates for some infections in older populations. While the added value of using linked data has previously been shown for several non-communicable conditions,[139-142] there has been a paucity of evidence regarding this topic when investigating infectious conditions. In particular, illnesses for which patients may seek treatment in either a primary or secondary care setting are at risk of being underestimated. In addition, incomplete recording of hospitalised pneumonia episodes in stand-alone GP data limits its use in studies of pneumonia in older adults.

Neither the linked nor the stand-alone data were originally designed for research use, resulting in each having their own limitations. Although considerably easier to use,

stand-alone CPRD data seem to under-record pneumonia events which occurred in hospital, leading to underestimates of the burden of disease. HES-linked CPRD data require a very large amount of preparation before they can be analysed, but enable inclusion of both hospitalised and ambulatory events.

For conditions that can be treated both in the community and in hospital, single data sources such as stand-alone CPRD are also likely to be susceptible to changes to health services, patient and clinician behaviour. Using linked-data may protect against these external influences when attempting to accurately estimate the burden of disease over time, but they are still susceptible to changes to diagnostic thresholds and trends in coding practices. Validation of the accuracy of pneumonia diagnosis and coding in both CPRD and HES over time would aid better understanding of the results I have presented. Further research is needed to establish whether the results shown are repeated in other infections, populations and settings. It may be that due to the underlying health status of the older population and the severity of CAP in this group, the findings shown are more pronounced among those aged ≥65 years.

5.4.3.4 Implications for the subsequent work in this thesis

The second part of this thesis concerns severe outcomes after CAP episodes. In order to fully capture CAP events, only patients eligible for CPRD HES-linked data are included in the analyses presented in later chapters.

Chapter 6   Risk factors for hospitalisation after CAP in older adults, and their contribution to increasing hospitalisation rates

The second section of this thesis concentrates on CAP, which is both directly and indirectly responsible for a large number of hospitalisations and deaths among older adults. Due to the disparities in recording of CAP in the linked and stand-alone data (as outlined in the previous chapter), all subsequent analyses in Chapter 6 (objective 2) and Chapter 7 (objective 3) only include patients who were eligible for CPRD HES-linked data (Figure 6-1).

The focus of this chapter is on hospitalisations following CAP. The main analysis aims to identify risk factors for hospitalisation after a CAP episode, and to assess the effect of these factors on trends in hospitalisation after CAP over time (objective 2).

Firstly I provide a review of the literature on risk factors for hospitalisation after CAP. My main analyses for objective 2 are presented as a paper published in BMJ Open. This is followed by additional detail of the methods I used to identify the variables included as potential risk factors for hospitalisation post-CAP. Finally, a supplementary analysis is presented, which assesses the added benefit of using individual co-morbidities rather than the Charlson index when adjusting the regression model from Paper 2 for patients' co-morbidities.

Figure 6-1 Records included in the analyses in this Chapter
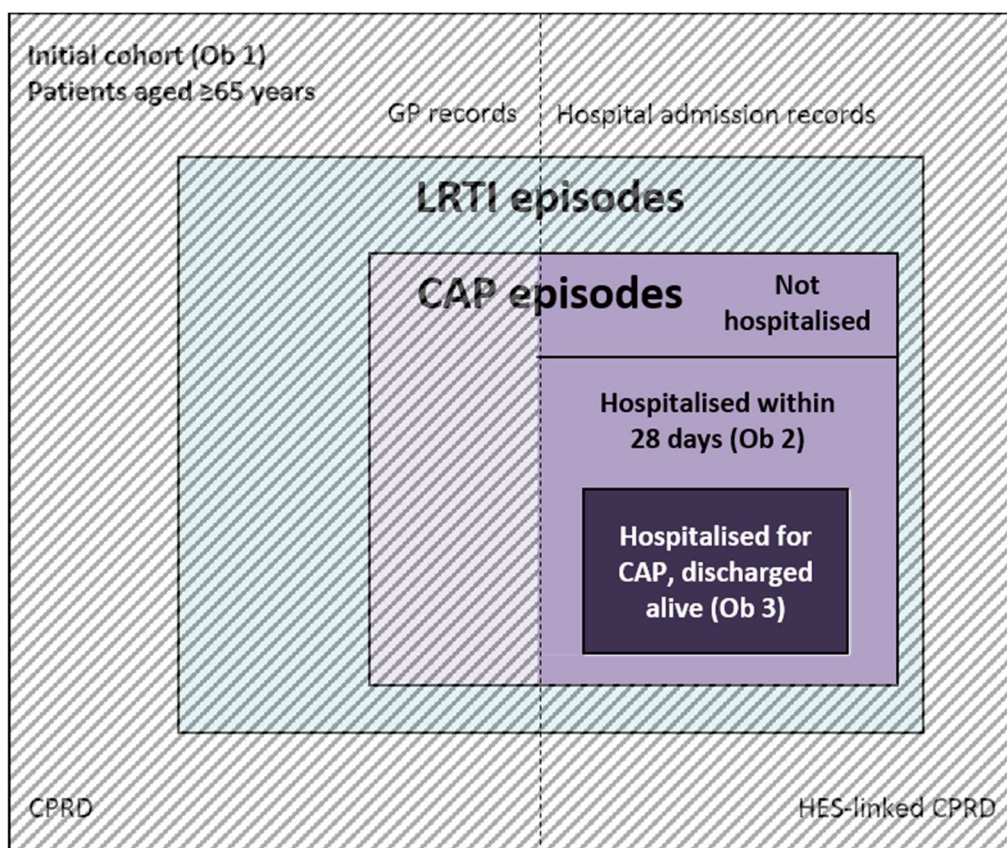
6.1 Background

6.1.1 Trends in hospitalisation for CAP over time

As highlighted in the Background (section 1.2.1), hospitalisations for pneumonia among older adults have risen in recent years both in the UK and other European countries. In England, increasing hospitalisations for CAP have been reported by two studies for periods spanning April 1997 and March 2011.[61, 62] Elsewhere in Europe, studies from Denmark,[64, 65] the Netherlands,[66] and Portugal,[67] have all shown rising admissions for pneumonia over the last 10 to 20 years. The extent of the increase varied by study, but consistently increased with age. However, as discussed in the Background, none of these studies were able to examine the observed trends in hospitalisation following pneumonia independent of underlying trends in pneumonia incidence. Additionally, only one English and one Dutch paper investigated the effect of increasing co-morbidity on hospitalisation trends over time. In both cases this was achieved using the Charlson co-morbidity index and not by investigating the role of individual co-morbidities.[61, 64] The effect of medications, vaccinations and frailty factors was not

considered in any of the studies, despite these factors being prevalent among the older population who make up a large proportion of pneumonia hospitalisations.

Could factors such as co-morbidities, increasing frailty and medication use lie behind the increasing levels of hospitalisation for pneumonia, and if so which of these factors might be driving the trend? In order to identify potential risk factors for hospitalisation for CAP for further investigation, I performed a literature review on the topic, which is described below.

6.2 Literature review of risk factors for hospitalisation for and after a CAP infection

6.2.1 Aim

The aim of this review was to identify factors among adults with CAP that have previously found to be associated with admission to hospital. The factors of interest included those potentially recorded in patients' electronic health records: co-morbidities, frailty factors, medications and vaccinations, and lifestyle factors.

6.2.2 Methods

6.2.2.1 Search strategy

A 2012 review paper which summarised risk factors among hospitalised CAP patients was taken as a starting point.[143] I searched Medline for articles published since the review (from 2008 (to overlap with the end of the review), until May 2015). The search strategy combined MeSH and free-text terms for community-acquired pneumonia, risk factors, hospitalisation and older adults (the full search strategy is provided in Appendix F).

6.2.2.2 Inclusion criteria

My primary interest was in the comparison between hospitalised and non-hospitalised CAP patients. Due to the small number of papers expected to meet this criterion, I also identified papers of secondary interest, in which hospitalised CAP patients were compared to the background population. These studies were considered separately due to their identifying risk factors for both developing CAP and being hospitalised for CAP. Studies comparing patients hospitalised for CAP to patients hospitalised for other

reasons were not included, as these were not thought to adequately identify risk factors for hospitalisation after CAP.

In addition to using one of the comparison groups above, papers were included if they: reported original research, reported a multivariable model including measures of effect (rate-/risk-/odds ratio) for risk factors for CAP hospitalisation, were set in a high-income country, were written in English, and included older adults in their study population. I did not specifically restrict the study population to older adults, as risk factors such as co-morbidities were considered likely to remain so across the adult age spectrum.

Articles were excluded if they were set within a specific sub-population which may have affected the generalisability of the results to the older population. This included those studying specific pathogens, CAP patients treated in ICU rather than all CAP hospitalisations (as these represent a specific more severe class of disease whose risk factors may not be generalisable to less severe hospitalised CAP), and CAP hospitalisations within a specific co-morbidity group (such as COPD) as risk factors within these groups may differ from those in the general population. Additionally, studies of the effect of a particular treatment (for example a specific type of antibiotic), or studies on risk factors for aspiration pneumonia were excluded as these cases were not included in this study's definition of CAP.

6.2.2.3   Article screening

I applied the inclusion criteria described in section 6.2.2.3 to the articles identified from the Welte review and from the Medline search. For articles I considered to be potentially eligible, I obtained the full-text and further checked the article against the inclusion criteria. I examined the reference lists of included articles in an attempt to identify additional papers of interest. Whenever I was unsure of a study's eligibility, I discussed it with my supervisor and a decision was reached by consensus.

6.2.2.4   Data extraction

I extracted relevant data into standardised data extraction forms in Excel. Separate forms were used for the articles of primary interest (hospitalised vs non-hospitalised CAP patients) and for those of secondary interest (hospitalised CAP vs general population). I extracted details of the study design and study population, the method

of ascertaining the outcome (hospitalisation after CAP), the hospitalisation rate, risk factors for hospitalisation that were identified including the ORs/RRs for these factors, and the statistical techniques used.

Studies were summarised narratively. I noted aspects of the studies that were particularly good, weak or that were unclear, but formal quality assessments were not performed.

6.2.3 Results

The Medline search identified 367 papers. Of these, only two compared CAP patients treated in a community setting to those hospitalised (the primary objective),[122, 125] and three compared hospitalised CAP patients to the background population (the secondary objective).[144-146] The Welte review provided one further study for the primary objective and one for the secondary objective.[130, 147]

6.2.3.1 Included studies

Primary objective – CAP treated in the community as a comparison group

The three studies which compared CAP patients treated in hospital with those treated in the community were all conducted in Southern Europe (40 provinces in Italy, the Barcelona region of Spain and a rural region of Crete, Greece, Table 6-1).[122, 125, 130] Two were prospective cohort studies in which data was collected from general practitioners,[130] and additionally pulmonary hospital clinics.[122] The third was a retrospective cohort study which utilised electronic health records from both primary and secondary care.[125] The studies were all relatively small, ranging from 124 to 699 patients with CAP. The Greek study was limited to older adults aged ≥50 years who had a hospitalisation rate of 32.3%.[130] The other papers included patients of all ages with pneumonia (Italy) and all adults aged ≥18 years (Spain), and reported hospitalisation rates of 22% and 42% respectively.[122, 125]

All three studies estimated effects using logistic regression and odds ratios. Age was the only risk factor included in all three models; the odds of hospitalisation were found to increase by 10% per year of age in the Spanish study,[125] while Greek patients aged ≥74 years had more than 7 times the odds of hospitalisation than those aged 50-73 years

(Table 6-1).[122] The Italian paper reported adjusting their model for age, sex and smoking status but did not report the measures of effect for these variables.[130]

Other factors reported by the three studies showed little overlap. The Greek study found that the presence of two or more co-morbidities (included as a co-morbidity count), being obese (defined as ≥30kg/m^2), and ≥40 pack years smoking were all associated with increased odds of hospitalisation (Table 6-1). Pneumococcal vaccination (over an undefined time period of vaccine receipt) was shown to be protective. Both of the other studies found associations between individual co-morbidities and odds of hospitalisation in multivariable analyses. The Spanish paper reported that presence of; stroke, dementia, liver disease, COPD and diabetes were associated with increased odds of hospitalisation among CAP patients (Table 6-1).[125] However, only odds ratios from the multivariable model with a p-value<0.002 were reported in the paper, indicating that other factors may also have been associated with hospitalisation. The Italian multivariable model included heart disease, neurological disease, asthma, and cancer, as well as previous corticosteroid therapy (Table 6-1).[130] All were shown to be associated with being hospitalised, however the results were only presented graphically and 95% confidence intervals for the last four factors appeared to cross one.

Table 6-1 Characteristics of studies comparing hospitalised and community treated CAP patients, and risk factors they identified

Author [ref]	Viegi [130]*	Sicras-Mainar [125]†	Bertsias [122]
Study population			
Region, Country	40 provinces in Italy	Barcelona, Spain	Crete, Greece
Study period	1999-2000	2008-2009	2011-2012
Study design	Prospective cohort	Retrospective cohort	Prospective cohort
Study population	Patients of all ages with 'suspected' and 'x-ray confirmed' pneumonia episodes as reported by GPs from 287 randomly selected general practices	Electronic health records of inpatients & outpatients aged ≥18y in 6 primary care centres & 2 hospitals with ICPC-2 (outpatient) or ICD-9-CM (inpatient) codes for pneumonia, & diagnosis confirmed by X-ray	Patients aged ≥50y presenting to 6 general practices & 2 pulmonary hospital clinics (including access to A&E) with CAP, defined as patients presenting with acute LRTI confirmed with lung infiltrate on X-ray, who had not been hospitalised <14d pre diagnosis.
Hospitalisation (case) ascertainment	Separate forms completed by GP when notified of hospitalised patients	ICD-9-CM code for pneumonia in hospitalisation records	Not explicitly defined.
Sample size	699	581	124
Age (years) Mean (SD)	1° care: 57.6 (19.2) 2° care: 66.7 (18.7)	57.5 (19.1)	Median 74 (range 50-95)
Male n (%)	328 (46.9)	323 (55.6)	64 (51.6)
Hospitalised n (%)	151 (21.6)	241 (41.5)	40 (32.3)
Risk factors identified: Odds Ratio (95% CI)			
Age	N/R	Per yr increase: 1.1	50-73 yrs: ref ≥74yrs : 7.13 (2.23-22.79)
Neurological disease	2.8 (1.5-5.5)~	Stroke: 3.6	
Dementia		3.5	
Liver disease		5.9	
Lung disease	Asthma: 2.0 (1.0-3.6)~	COPD: 2.9	
Diabetes		1.9	
Heart disease	4.0 (2.5-6.0)~		
Cancer	3.5 (0.95-1.75)~		
Co-morbidity count			Multi-morbidity (≥2 co-morbidities) 5.77 (1.81-18.42)
BMI			Obese : 3.36 (1.08-10.52)
Smoking	N/R		0 pack years: ref 1-39 pack yrs: 0.32 (0.06-1.73) 40+ pack yrs: 3.82 (1.07-13.68)
Pneumococcal vaccine			Vaccinated: 0.29 (0.09-0.95)
Previous steroid use	2.1 (1.0-3.6)~		
Area of residence	1.5 (0.9-2.1)~		

*model also included sex and area of residence but results not reported. †results only reported in paper if p<0.002, results reported to 1dp and no 95% CI provided. N/R: included in model but results not presented in paper.

~ORs estimated from Figure (no table provided) ICPC-2: International Primary Care Classification ICD-9-CM: International Classification of Diseases, ninth revision, Clinical Modification

Reporting of the statistical methods used was inadequate in all the studies. No detail was provided on their model building strategies, or the other factors investigated but not included in the final multivariable models. None of the studies were conducted for a sufficiently long period to assess the effect of the risk factors on hospitalisation rates over time.

Secondary objective – using the background population as a comparison group

Of the four papers that compared hospitalised CAP patients to the background population, three cohort studies were set in the USA,[144, 147, 146] and one case-control study was from Canada.[145] Hospitalisation rates for CAP in these study populations ranged from 6.2% to 45.2% (Table 6-2). One American study pooled data from two prospective cohorts including adults aged ≥45 years,[146] while the other three papers only included older adults. The second American paper used a subset of a larger cohort, restricting to patients with information on oral hygiene (who had fewer pneumonia outcomes than those who were excluded).[144] The final American study also used data from a previous prospective cohort (the Cardiovascular Health Study).[147] The Canadian case control study enrolled CAP cases at A&E, and recruited controls via random digit dialling.[145]

A variety of factors were considered by these studies. Age was consistently found to be associated with increased CAP hospitalisation,[144, 147, 146, 145] ranging from 7% increase per year of age[146] to a 28 fold increase in those aged 80-84 years compared to those aged <50 years.[147] Two articles reported men having around twice the rate of CAP hospitalisation than women.[147, 144]

The majority of co-morbidities were investigated by just two studies, with limited overlap between the factors they included in their multivariable models.[147, 145] Co-morbidities reported to increase the risk of CAP hospitalisation included diagnoses of COPD, congestive heart failure, coronary heart disease, diabetes, cerebrovascular accident, and renal disease (Table 6-2). Previous pneumonia was reported as tripling the rate of CAP hospitalisation by one study,[144] but having no clear effect by another.[147]

Table 6-2 Characteristics of studies comparing hospitalised CAP cases to the background population, and risk factors they identified

Author [ref]	Loeb [145]	Juthani-Mehta [144]	Yende [146]	O'Meara [147]
Study population				
Region, Country	Ontario & Alberta, Canada	Memphis & Pittsburgh, USA	Sites in six states, USA	Sites in four states, USA
Study period	2003 - 2005	1997-2008	1987-1990	1989 - 2001
Study design	Case-control	Prospective cohort	Prospective cohorts	Prospective cohort
Study population	Cases: patients presenting to A&E. Controls: selected contemporaneously using random digit dialling from same catchment area. All aged ≥65y.	Individuals aged 70-79y participating in the Health, Ageing and Body Composition study -subset with dental examination assessment.	Individuals aged ≥45y participating in the Cardiovascular Health Study or in the Atherosclerosis Risk in Communities study.	. Individuals aged ≥65y participating in the Cardiovascular Health Study.
Hospitalised CAP (case) ascertainment	Presented with ≥2 specified signs/symptoms & new opacity on chest x-ray consistent with pneumonia.	Adjudicated by a committee, using a combination of discharge summary, ICD-9 diagnosis code, admission history, physical exam & x-ray reports.	Hospitalisation for pneumonia within 10yr period identified using combination of ICD-9 codes, with chart review and physician adjudication.	Identified via patient self-report, corroborated by ICD-9 codes for pneumonia as 1st discharge diagnosis in Medicare data. Only included patient's 1st CAP hospitalisation
Sample Size	1584	1441	16260	5888
Age (mean (SD))	Cases:79.1 (7.6) Controls: 74.4 (6.7)	74.7 (2.9)	59.2 (10.1)	N/R (all ≥65)
Male n (%)	Cases: 429 (60.4)† Controls: 273 (31.5)	719 (49.9)	7172 (44.1)	2495 (42.4)
Hospitalised n (%)	716 (45.2)	193 (13.4)	1000 (6.2)	582 (9.9)
Risk Factors identified Hazard/Odds Ratio (95% CI)				
Age (years)	OR: 1.93 (1.64-2.27)	HR: 1.24 (0.93-1.65)	HR: <50:Ref 65: 4.84 (3.43-6.75) 70: 9.93 (7.00-14.08) 75: 14.72 (10.17-21.31) 80: 27.87 (18.67-41.58) 85: 16.13 (9.73-26.76)	HR:1.07 (1.05-1.09)
Male gender		2.07 (1.51-2.83)		1.96 (1.59-2.42)
Cerebrovascular accident				1.47 (1.06-2.05)
COPD	13.53 (7.8-23.48)			1.49 (1.17-1.89)
Diabetes				1.34 (1.05-1.70)
Coronary heart disease				1.51 (1.22-1.87)
Congestive heart failure	2.07 (1.22-3.49)			1.92 (1.38-2.77)
Renal disease	4.06 (1.98-8.35)			

Author [ref]	Loeb [145]	Juthani-Mehta [144]	Yende [146]	O'Meara [147]
Previous CAP*		3.09 (1.86-5.14)		1.07 (0.88-1.30)
Smoking*	Lifetime >100 cigarettes: 2.01 (1.26-3.36)	Pack years: 1.01 (1.00-1.01)	past: 1.26 (1.07-1.49) current: 2.06 (1.70-2.5)	current: 1.75 (1.35-2.26)
Alcohol	g/m prev yr (per 5g increase): 1.69 (1.08-2.61)			
Pneumococcal vaccine				1.14 (0.92-1.42)
Influenza vaccine in previous year				0.99 (0.81-1.21)
Living alone	0.48 (0.30-0.76)			
Immunosuppressants	15.13 (4.74-48.29)			
Other factors related to frailty	Dysphagia: 3.76 (1.60-8.88) Nutritional score (per unit reduction): 1.83 (1.19-2.80) Functional status (Barthel score 18+): 7.94 (3.77-16.69)	Incident mobility limitation 1.77 (1.32-2.38)		H/o claudication: 2.11 (1.4-3.18) Time walk 15ft: 1.02 (0.99-1.06) 3MSE (cognitive exam): 0.99 (0.97-1.01)

3MSE: Modified Mini-Mental State Examination, H/o history of, ICD-9-CM: International Classification of Diseases, ninth revision, *=different categorisation used in different models, † Missing data on sex for 6 cases,

Patients' smoking status was included in all four studies and consistently associated with increased CAP hospitalisation.[144-147]   In the Canadian study, alcohol consumption was found to increase the odds of hospitalisation by 69% with each 5g increase in alcohol.[145]

Neither pneumococcal nor influenza vaccine were associated with decreased odds of hospitalisation in the only study which included vaccinations in its multivariable model.[147]   The only medication included in any of the final models was immunosuppressants, which in the Canadian study was shown to have one of largest effects of all factors studied (OR: 15.13 (4.74-48.29)).[145]

Three of the four studies also investigated aspects of frailty. Decreased functional status and decreasing nutritional score were both found to increase odds of CAP hospitalisation, while living alone was found to be protective (Table 6-2).[145] The two American papers reported contrasting effects of limited mobility, one found it to be associated with increased CAP hospitalisation rates,[144] while the other found no effect.[147] Cognitive impairment was also found to have no effect on the rate of CAP hospitalisation in the latter study.[147]   However, all three of these studies were restricted by exclusion criteria which would have omitted more frail patients from the study, thus limiting investigation of aspects of frailty.[144, 145, 147]

In general, the reporting of statistical methods was more detailed than for the studies of primary interest, with two studies using backward stepwise selection of variables into their final models,[145, 147] and the third study using both forward and backward selection.[144]   The final study developed a prognostic model for risk of CAP hospitalisation over a 10-year period using a non-standard approach.[146]

6.2.4   Comment

In general, there is a paucity of information around risk factors for hospitalisation among older CAP patients, especially around frailty factors, medications and vaccinations. Only three relatively small studies examined risk factors for hospitalisation specifically among CAP patients. One of these studies was set within the older population, but used a co-morbidity count rather than investigating the association between individual co-morbidities and odds of hospitalisation. In the other two studies, seven co-morbidities

were reported as increasing the odds of CAP hospitalisation, but different categorisation of neurological disease and lung disease resulted in none of the same conditions being identified in both studies. In general, factors likely to influence decisions over site of care, such as frailty factors or place of residence (for example living alone or in residential accommodation) were not included in final multivariable models. A larger variety of conditions were included in the four studies included in the secondary objective, which compared hospitalised CAP patients to the background population. Six co-morbidities were shown to increase the risk of CAP hospitalisation in these studies, as was a history of CAP and a variety of frailty factors. However, the usefulness of these results is limited by the inability to separate whether the conditions identified were risk factors for acquiring CAP or for CAP patients' subsequent hospitalisation.

None of the seven studies investigated the effect of the risk factors for hospitalisation on trends in CAP hospitalisation over time, and so the importance of these factors on increasing levels of CAP hospitalisation remains unquantified.

Although my review highlights a lack of studies of risk factors for hospitalisation among CAP patients, some limitations need to be considered. It is possible that by restricting my search strategy to papers identified in Medline and written in English (due to time and resource constraints), I may not have included all papers which addressed the topic. I did attempt to identify further studies by detailed examination of the reference lists of all included papers, although none were identified by this method. The lack of a second reviewer to assess eligibility of studies and to extract data from selected studies may also have resulted in exclusion of eligible studies, as well as increasing the risk of errors in the reported results.

6.3 Objective 2: identification of risk factors for hospitalisation after CAP in older adults

6.3.1 Introduction

In this study I aimed to identify risk factors for hospitalisation after a CAP diagnosis among older adults, and to assess the contribution of these factors to increasing levels of hospitalisation with CAP over time. For this analysis I used all CAP episodes identified in the linked CPRD-HES data between 1st April 1998 and 31st March 2011. CAP records

from between 1st April 1997 and 31st March 1998 were not included, due to the unusually low numbers of CAP hospitalisations that were observed over this period (as reported in section 5.3.1). I compared CAP episodes which resulted in hospitalisation within 28 days to those that solely received care from a GP. All hospitalisations in the 28 days after a CAP episode were included, not just those for which CAP was the primary code of the first episode, in order to capture events that CAP may have precipitated, such as a fall or MI.

As discussed in section 2.8.2.2, when developing a causal model, the hierarchical relationship between the factors of interest and whether their effects are direct or mediated through other factors needs to be considered.[116] Figure 6-2 shows a hierarchical framework for the groups of variables included in the analyses of hospitalisation in the 28 days after a CAP diagnosis.

Figure 6-2 Hierarchical framework for hospitalisation after CAP

6.3.2 Research paper cover sheet

LONDON
SCHOOL *of*
HYGIENE
&TROPICAL
MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED <u>FOR EACH</u> RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Elizabeth Millett
Principal Supervisor	Sara Thomas
Thesis Title	The use of linked electronic health data to investigate the burden and outcomes of community-acquired pneumonia among older individuals in the United Kingdom.

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	BMJ Open			
When was the work published?	1st December 2015			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A			
Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes	

*If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	This study was conceived by Sara Thomas, who also obtained funding, ethical approval and the data for the study from CPRD. I developed the detailed study design, supervised by Sara Thomas and with statistical advice from Bianca De Stavola. The CAP episodes included in the study were

	identified using methods I developed for the incidence analyses (Chapter 4). Read and ICD-10 codelists for co-morbidities were developed by or adapted from the work of others at LSHTM by Sara Thomas. I developed Read and ICD-10 lists for the frailty factors, which were then agreed by Sara Thomas. I designed the methods to identify and categorise the co-variables of interest. Exceptions were as follows: inhaled corticosteroid use was adapted from the work of Jennifer Quint, categorisation of oral steroid use was adapted from the work of Harriet Forbes, and smoking status categorisation built upon work by Ian Douglas and Krishnan Bashkaran. Decisions around acute/chronic events and periods of exposure for medications were made based on current literature and after discussion with Sara Thomas. Throughout the work, additional advice on clinical aspects was provided by Jennifer Quint and Liam Smeeth. I conducted all data management, analysis, led the interpretation of results (with input from Sara Thomas and Bianca De Stavola) and I wrote the first draft of the paper. All co-authors contributed revisions, which I then incorporated. After peer-review, I further adapted the manuscript to include the reviewers' comments, with advice from all co-authors.

Student Signature: _____ **Date:** 04 / 12 /15

Supervisor Signature: _Sar Thms_ **Date:** 4/12/15

BMJ Open   Risk factors for hospital admission in the 28 days following a community-acquired pneumonia diagnosis in older adults, and their contribution to increasing hospitalisation rates over time: a cohort study

Elizabeth R C Millett, Bianca L De Stavola, Jennifer K Quint, Liam Smeeth, Sara L Thomas

CrossMark

Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, UK

Correspondence to
Elizabeth RC Millett;
Elizabeth.millett@lshtm.ac.uk

ABSTRACT

Objectives: To determine factors associated with hospitalisation after community-acquired pneumonia (CAP) among older adults in England, and to investigate how these factors have contributed to increasing hospitalisations over time.

Design: Cohort study.

Setting: Primary and secondary care in England.

Population: 39 211 individuals from the Clinical Practice Research Datalink, who were eligible for linkage to Hospital Episode Statistics and mortality data, were aged ≥65 and had at least 1 CAP episode between April 1998 and March 2011.

Main outcome measures: The association between hospitalisation within 28 days of CAP diagnosis (a 'post-CAP' hospitalisation) and a wide range of comorbidities, frailty factors, medications and vaccinations. We examined the role of these factors in post-CAP hospitalisation trends. We also looked at trends in post-CAP mortality and length of hospitalisation over the study period.

Results: 14 comorbidities, 5 frailty factors and 4 medications/vaccinations were associated with hospitalisation (of 18, 12 and 7 considered, respectively). Factors such as chronic lung disease, severe renal disease and diabetes were associated with increased likelihood of hospitalisation, whereas factors such as recent influenza vaccination and a recent antibiotic prescription decreased the odds of hospitalisation. Despite adjusting for these and other factors, the average predicted probability of hospitalisation after CAP rose markedly from 57% (1998–2000) to 86% (2009–2010). Duration of hospitalisation and 28-day mortality decreased over the study period.

Conclusions: The risk factors we describe enable identification of patients at increased likelihood of post-CAP hospitalisation and thus in need of proactive case management. Our analyses also provide evidence that while comorbidities and frailty factors contributed to

Strengths and limitations of this study

- Our novel use of large linked primary and secondary care data sets provided enriched patient medical and therapeutic histories, and allowed detailed identification of the determinants of hospitalisation after community-acquired pneumonia (CAP).
- The very large sample size, with more than 43 000 CAP episodes, enabled assessment of multiple potential risk factors for hospitalisation with precise estimation of relative risk.
- Using linked data also allowed us to distinguish trends in the tendency to hospitalise patients with CAP (the focus of this paper) from trends in CAP hospitalisations due simply to increasing CAP incidence.
- Our analyses suggested that frailty factors were suboptimally recorded by general practitioners, preventing full assessment of these factors and highlighting the need for better capture of frailty by practices.
- For similar reasons, analyses on the effect of smoking were performed on a subset of the data, due to previously described incomplete recording of smoking status pre-2004.

increasing post-CAP hospitalisations in recent years, the trend appears to be largely driven by changes in service provision and patient behaviour.

INTRODUCTION

Hospitalisations for ambulatory care sensitive conditions (ACSC, conditions which could be treated outside of hospital) have increased considerably over the past decade.[1 2] Pneumonia is one such condition, with >56 000 more pneumonia admissions in

2010/2011 compared with 2001/2002.[2] Most of this increased burden is found among patients aged 65 years and older who accounted for 70% of pneumonia admissions in 2012/2013.[3] In England, the recently introduced Unplanned Admissions Enhanced Service highlights the importance of proactive case management in primary care of at-risk patients, many of whom are expected to be older, to reduce ACSC hospitalisations.[4 5]

To date, risk factors for hospitalisation for community-acquired pneumonia (CAP) have not been quantified. Existing analyses based on stand-alone hospitalisation data are unable to compare hospitalised patients with CAP to those who were treated in the community. They therefore cannot distinguish between factors which affect a patient's likelihood of hospitalisation after a CAP diagnosis from risk factors for developing CAP. Furthermore, hospitalisation data have incomplete information on patients' medical histories, and contain little or no information on factors such as alcohol and smoking habits, frailty, or medications prescribed in the community. It is frequently hypothesised that changes in comorbidities and frailty factors have contributed to the increasing hospitalisation trends for older individuals with CAP in the UK. Use of a non-hospitalised comparison group would allow this hypothesis to be tested, and to distinguish an increasing tendency to hospitalise older patients with CAP from increasing incidence of CAP among older adults.[6]

In this study, we used large linked general practice, hospital admission and mortality data sets to assess a variety of potential risk factors (comorbidities, medications and other factors) for hospitalisation after CAP among older individuals in England. Use of the general practice data enabled more complete capture of patient histories than those derived from stand-alone hospital records. We also investigated the risk over time of hospitalisation after a CAP diagnosis. The choice of a study population who had been diagnosed with CAP allowed us to examine specifically trends in hospitalisation after CAP, independent of any trends in pneumonia incidence. We assessed to what extent the patient factors associated with hospitalisation explained these hospitalisation trends. The linked mortality data enabled further investigation into whether mortality rates in the 28 days after CAP had changed over the same period, including deaths occurring both in and outside hospital settings, as a marker of CAP severity.

MATERIALS AND METHODS
Data sources
The Clinical Practice Research Datalink (CPRD, formerly GPRD) is a large database of UK general practice records comprising a representative sample of around 8% of the UK population.[7 8] Anonymised data in CPRD include diagnoses (coded using Read codes), prescriptions, referrals, tests and patient demographics. Over 50% of CPRD patients living in England have their general practice records linked to Hospital Episode Statistics (HES) which includes all inpatient National Health Service (NHS) hospitalisations (coded using International Classification of Diseases (ICD)-10).[9] Each HES hospitalisation consists of one or more episode denoting the time a patient is under the care of one consultant. The data were also linked to Office for National Statistics (ONS) central mortality data to obtain vital status and, if relevant, date of death.

Study population
We included patients who were registered with a CPRD practice eligible for linkage to HES data, were aged 65 years or over between 1 April 1998 and 31 March 2011, and who had a CAP episode recorded during that period.

Defining CAP episodes
The methods for defining CAP illness episodes have been described in detail elsewhere.[6] In brief, lists of Read (CPRD) and ICD-10 (HES) codes for pneumonia and other lower respiratory tract infections (LRTIs) were agreed by three clinical epidemiologists. Pneumonia could be first diagnosed either in general practice or when presenting to hospital. In HES, in order to differentiate between illness present at hospital admission and subsequent hospital-acquired illness, only pneumonia coded as the reason for admission (ie, the primary code in the first episode of a hospitalisation) was included in the study. These HES pneumonia records were combined with CPRD pneumonia records to determine 'illness episodes' whereby records within 28 days of each other (or of an intermediate LRTI record) were deemed part of the same infection.[6] The earliest pneumonia record in the 'illness episode' was the incident (diagnosis) date of pneumonia.

To be defined as community-acquired, the CAP incident date needed to be ≥14 days after any HES inpatient hospital discharge. All CAP episodes in eligible patients during the study period were included in the study.

Defining hospitalisation after pneumonia
The outcome of interest was hospital admission (defined using HES) for any cause, on or up to 28 days after the CAP diagnosis date. Thus, a CAP diagnosed when a patient presented at hospital was automatically assigned as having the outcome; a CAP diagnosed in general practice had the outcome if the patient had a hospital admission in the next 28 days. We chose all-cause hospitalisation because we also wanted to capture hospitalisations for events which pneumonia could have precipitated in our older population, such as stroke, myocardial infarctions and falls, or worsening of underlying comorbidities such as chronic obstructive pulmonary disease or congestive heart failure.[10]

Other factors
Age was categorised in 5-year groups from 65 to 89 years, and ≥90 years.

Open Access

Time period

Year of hospitalisation used a financial year structure (1 April to 31 March) to ensure respiratory pathogens circulating throughout the winter were captured in the same year. Year was then grouped as 1998–2000, 2001–2003, 2004–2006, 2007–2008 and 2009–2010, to account for health service changes such as the introduction of payment-for-performance indicators in 2004 and 2009.[11] [12]

Comorbidities and frailty factors

Code lists for the 19 comorbidities in the Charlson Index and for additional cardiac, neurological and immune disorders that could affect pneumonia disease severity or a doctor's decision to hospitalise were devised by author SLT and at least one other clinical epidemiologist.

Prevalidated frailty scores such as the frailty phenotype or frailty index could not be utilised due to aspects of each not being recorded in the electronic health records used in this study (eg, grip strength and slow gait speed from the phenotype and sucking problems and poor muscle tone in neck from the index).[13] [14] Instead, a wide variety of factors identified in the frailty index as associated with frailty, for which information was potentially available in the databases, were considered. Authors ERCM and SLT devised Read and ICD-10 code lists and used other recording fields within the data to capture health deficits within the previous year (eg, history of falls, inability to self-care) which were likely to be recorded in patients' health records, as well as other factors that could increase the likelihood of hospitalisation (eg, living alone).

The presence of chronic conditions (such as diabetes or dementia) was determined using CPRD and HES records from any point up to and including the CAP incident date. For acute/potentially acute conditions (myocardial infarction, stroke, congestive heart failure, hemiplegia, falls, weight loss/undernutrition) which could have occurred as a result of the CAP, records from any point prior to but excluding the CAP incident date itself were used as evidence of a pre-existing condition.

Terminal illness was defined using Read and ICD-10 codes stating terminal illness, rather than specific conditions. In addition, primary care information on referrals to hospices was included.

Medications, vaccinations and health behaviours

Medications included oral steroids, inhaled steroids, immunosuppressive drugs, statins and antibiotics. We also considered influenza and pneumococcal vaccination status, and health behaviours such as smoking and excessive alcohol consumption.

As is common when using routinely collected health records, for all the factors aforementioned, the absence of a code for a condition was assumed to represent absence of the condition.

A full list of the factors considered, how they were categorised and timescales used to determine if they were present at CAP diagnosis, can be found in online supplementary file A.

Main analyses

Some patients had more than one CAP event during the study period. It is highly likely that decisions around whether to hospitalise a patient after CAP were affected by a previous history of CAP. Furthermore, decisions to hospitalise may have been similar for patients within a general practice, for example, due to local area service provision. To account for this clustering at patient and practice level, we used multilevel logistic regression models for the binary hospitalisation outcome.[15] The model had three levels: CAP episodes nested within patients who were nested within practices. The suitability of the three-level model was assessed by comparing it to simpler specifications (using likelihood ratio tests (lrt)) both before and after including explanatory factors in the model.

First, minimally adjusted ORs of hospitalisation following a CAP (adjusted for age, sex and year of CAP) were produced for each of the factors of interest. The size of these ORs and their 95% CIs were used to decide which variables to include in later models. These variables were grouped into (1) comorbidities; (2) frailty factors; and (3) medications, vaccinations and health behaviours. We added the three groups of variables sequentially to a model adjusted for age, sex and year of CAP, according to each group's hypothesised place on the causal pathway to hospitalisation. This enabled examination of the independent effect of each comorbidity (in the 'comorbidity' model), and how much of each comorbidity's effect was explained by resulting frailty and/or medications (in subsequent models). A possible interaction between age and sex on the odds of hospitalisation was investigated in the final (full) model, comparing full models with and without the interaction term using an lrt.

To investigate the extent to which patient risk factors for hospitalisation explained trends in the probability of post-CAP hospitalisations, ORs for hospitalisation for each temporal period relative to 2001–2003 were estimated, controlling for changes in comorbidities and other factors. Multilevel models produce cluster-specific ORs of hospitalisation (ie, effects measured within each cluster). When investigating the level of hospitalisation after CAP over time, results at a population level were deemed more useful. Thus, we used the predicted cluster-level odds of hospitalisation derived from the final multilevel model to calculate the population average of predicted percentages of CAPs hospitalised in each year group.[16]

Further analyses

Records for smoking enable the recording of a negative response (non-smoker), and so levels of missing data

3

were able to be established for this variable. Smoking status was more completely recorded over time, decreasing from 26% missing data in 1998 to <1% by 2010 (see online supplementary file B). Multiple imputation was not considered appropriate as data were unlikely to be missing at random, for example, with respect to comorbidity status. Analyses including smoking as a covariate were therefore restricted to a subset of the data, performing a complete case analysis of CAP episodes with a recorded smoking status from 2007 onwards (which included more than 97% of CAPs per year).

Trends in mortality in the 28 days after CAP were assessed using the linked mortality data and multilevel logistic regression modelling (as for hospitalisation). Odds of mortality over time were adjusted for age and sex, but not for comorbidities or other factors, to avoid overadjustment of CAP severity resulting from underlying health deficits.

Two other potential explanations for trends in hospitalisation were investigated. The length of hospital admission for hospitalised patients over time (a further potential proxy of severity of illness) was examined using HES data. The pathway of care for each CAP episode over time was assessed by examining method of admission data in HES, and whether there was a general practice consultation for CAP (or potential CAP) on the day of diagnosis. We widened the definition of CAP in general practice to any LRTI to allow for conservative coding by general practitioners (GPs) in the absence of radiographical confirmation of pneumonia.[17]

All analyses were performed using Stata MP V.11.2.

RESULTS

Of 917 859 potentially eligible patients, 39 211 had at least one recorded CAP and their 43 576 CAP illness episodes were included in the study. The median age at diagnosis was 81 years (lower-upper quartiles: 75–87 years) and 53% of CAPs were in females (table 1). Most patients (91%) experienced one CAP, 7% of patients had two episodes and 2% had 3–8 episodes. The reason for admission was coded with an ICD-10 Chapter X code (Diseases of the respiratory system) for 95% of admissions throughout the study period, with Chapter XVIII (Signs and symptoms not elsewhere classified) and Chapter IX codes (Circulatory disease) each contributing around 1% of admissions.

Risk factors for hospitalisation

After adjusting for age, sex and year, our study found little evidence that hemiplegia, mild renal disease, self-care problems, anxiety/depression, mobility issues, tiredness, history of falling or excessive alcohol consumption were risk factors for hospitalisation, and these factors were not included in subsequent analyses (see online supplementary file C).

Results for the remaining 16 comorbidities, 6 frailty factors and 7 medications/vaccinations are given in figure 1 and in the online supplementary file C. The figure reports the final model in which ORs are mutually adjusted; the table reports minimally and then consecutively adjusted ORs in successive models.

Comorbidities

In the final model, 12 comorbidities were associated with increased odds of hospitalisation after CAP (figure 1). Of the comorbidities common among this cohort, chronic lung disease, ischaemic heart disease, congestive heart failure, severe renal disease and diabetes (with and without complications) were associated with a 25–82% increased odds of hospitalisation. The greatest odds of hospitalisation were found for less common conditions such as metastatic cancer and other disorders of the immune mechanism (adjusted ORs 2.46 and 2.49, respectively). Only terminal illness and dementia remained clearly associated with decreased odds of hospitalisation. Comparison with earlier models indicated that adjustment for frailty factors and medications made little difference to effect estimates for individual comorbidities except for those for connective tissue disease (attenuated by medications) and dementia (attenuated by frailty factors, online supplementary file C).

Frailty factors

Visual problems was the only frailty factor associated with increased odds of hospitalisation in the final adjusted model, while the presence in the last year of bedsores, low weight/poor nutrition or incontinence had a negative effect on hospitalisation, as did residence in a nursing home (figure 1).

Medications/vaccinations

Patients with a prescription for antibiotics in the previous 8–28 days were less likely to be hospitalised after CAP than patients with no prescription in the previous 4 weeks, controlling for the other variables in the model. Oral steroid use was associated with increased odds of hospitalisation, but the strong effect of inhaled corticosteroids and other immunosuppressive medications disappeared after adjusting for comorbidities (see online supplementary file C). We did not observe a protective effect of statin use against hospitalisation. Influenza vaccination in the current influenza season lowered the odds of hospitalisation after CAP in the final adjusted model by 25% compared with those who had never been vaccinated. In contrast, receipt of pneumococcal vaccine showed no protective effect, with evidence of slightly increased odds among the group vaccinated ≥5 years ago (compared with unvaccinated).

Age/sex

In the final model, females remained at lower odds of hospitalisation than males, and hospitalisation odds increased with age up to 85–89 years. However, there was evidence that the effect of age on hospitalisation varied by sex ($p_{interaction} < 0.001$). In contrast to men, women

Table 1 Characteristics of the study population, factors of interest and hospitalisation within 28 days of CAP

	Hospitalised within 28 days n	Not hospitalised n	Total
All CAPs n (%)	33 321 (76.5)	10 255 (23.5)	43 576
Male n (%)	16 143 (79.5)	4151 (20.5)	20 294
Female n (%)	17 178 (73.8)	6104 (26.2)	23 282
Age (grouped) n (%)			
65–69	3469 (75.9)	1099 (24.1)	4568
70–74	4703 (78.8)	1262 (21.2)	5965
75–79	6039 (78.4)	1663 (21.6)	7702
80–84	7227 (79.5)	1865 (20.5)	9092
85–89	6666 (76.6)	2038 (23.4)	8704
90+	5217 (69.1)	2328 (30.9)	7545
Year of CAP (grouped) n (%)			
1998–2000	4008 (57.7)	2944 (42.3)	6952
2001–2003	6266 (69.9)	2701 (30.1)	8967
2004–2006	8269 (79.2)	2173 (20.8)	10 442
2007–2008	7039 (83.7)	1372 (16.3)	8411
2009–2010	7739 (87.9)	1065 (12.1)	8804
Individual comorbidities n (%)			
Ischaemic heart disease			
Pre-MI	7261 (81.5)	1644 (18.5)	8905
Post-MI	4914 (83.2)	994 (16.8)	5908
Congestive heart failure	8289 (79.6)	2124 (20.4)	10 413
Peripheral vascular disease	4661 (82.7)	976 (17.3)	5637
Dementia	4526 (66.8)	2248 (33.2)	6774
Chronic lung disease	14 571 (83.4)	2905 (16.6)	17 476
Connective tissue disease	3347 (81.9)	740 (18.1)	4087
Peptic ulcer	3343 (81.1)	778 (18.9)	4121
Liver disease			
Mild	241 (84.3)	45 (15.7)	286
Severe	165 (87.3)	24 (12.7)	189
Diabetes			
Diabetes	4678 (81.3)	1076 (18.7)	5754
With complications	1633 (86.6)	253 (13.4)	1886
Hemiplegia	1243 (76.4)	384 (23.6)	1627
Cancer			
Solid cancer	5208 (80)	1300 (20)	6508
Metastatic	1066 (83.9)	204 (16.1)	1270
Leukaemia/lymphoma	981 (85)	173 (15)	1154
Severe renal disease	7001 (88.6)	900 (11.4)	7901
Cerebrovascular disease	8338 (74.5)	2856 (25.5)	11 194
Neurological disease	2997 (73.1)	1103 (26.9)	4100
Disorders of the immune mechanism	243 (90)	27 (10)	270
Mild renal disease	401 (82.5)	85 (17.5)	486
Terminal illness	1190 (67.1)	584 (32.9)	1774
Frailty factors n (%)			
Recent carer	1418 (79.6)	364 (20.4)	1782
Living arrangements			
Not recorded	27 949 (77.6)	8070 (22.4)	36 019
Lives alone	1471 (81)	344 (19)	1815
Sheltered accommodation	477 (79)	127 (21)	604
Residential care	3424 (66.6)	1714 (33.4)	5138
Visual impairment	11 098 (78.8)	2984 (21.2)	14 082
Self-care	366 (79.7)	93 (20.3)	459
Anxious/depressed	2730 (76)	860 (24)	3590
Bedsore/ulcer	824 (59.7)	556 (40.3)	1380
Mobility issues	2072 (79)	552 (21)	2624
Tired	1957 (74.4)	672 (25.6)	2629
Low weight/poor nutrition	4460 (75.1)	1477 (24.9)	5937

Open Access

Table 1 Continued

	Hospitalised within 28 days n	Not hospitalised n	Total
Incontinence/catheter	3230 (71.7)	1274 (28.3)	4504
History of falling	4792 (76.4)	1484 (23.6)	6276
Excessive alcohol consumption			
Any excess alcohol code	1720 (80.3)	423 (19.7)	2143
Medications n (%)			
Immunosuppressants (other than steroids) in past 120 days	685 (85.3)	118 (14.7)	803
Inhaled corticosteroids			
None pre-CAP	22 414 (73.6)	8023 (26.4)	30 437
Within 60 days	6864 (84.7)	1239 (15.3)	8103
Within 61–180 days	1620 (81.6)	366 (18.4)	1986
Within 181–365 days	597 (81.4)	136 (18.6)	733
More than 365 days ago	1826 (78.8)	491 (21.2)	2317
Antibiotics			
None in previous 28 days	23 437 (77.2)	6926 (22.8)	30 363
In previous 1–7 days	5368 (76.9)	1610 (23.1)	6978
In previous 8–28 days	4516 (72.4)	1719 (27.6)	6235
Statins in previous 6 months	8829 (86.7)	1350 (13.3)	10 179
Oral steroids in previous 90 days	5242 (83)	1077 (17)	6319
Influenza vaccine receipt			
No vaccine pre-CAP	4940 (69.7)	2143 (30.3)	7083
14–365 days pre-CAP	20 554 (76.2)	6420 (23.8)	26 974
Last season	5990 (75.5)	1949 (24.5)	7939
2–5 years pre-CAP	1846 (71.7)	728 (28.3)	2574
>5 years pre-CAP	656 (76.4)	203 (23.6)	859
Pneumococcal vaccine			
No vaccine pre-CAP	13 126 (66.4)	6643 (33.6)	19 769
14–365 days pre-CAP	1872 (73.7)	669 (26.3)	2541
1–2 years pre-CAP	2095 (75.4)	682 (24.6)	2777
2–5 years pre-CAP	7260 (80.1)	1801 (19.9)	9061
>5 years pre-CAP	9633 (85.4)	1648 (14.6)	11 281

CAP, community-acquired pneumonia; MI, myocardial infarction.

aged ≥90 years were not at increased odds of hospitalisation compared with women aged 65–69 years after adjusting for comorbidities and other factors (table 2). The three-level model remained the most appropriate structure (compared with single or two-level) after the addition of all other factors to the model (all p<0.001).

The effect of comorbidities on trends in post-CAP hospitalisations

After adjusting for all factors and for clustering, a marked increase in the percentage of CAP cases hospitalised over time remained, rising from 57% to 86% hospitalised over the study period. The wide range of comorbidities and other factors identified as risk factors for hospitalisation contributed very little to this increase (table 3).

Smoking

In total, 17 008 CAP events between 2007 and 2010 were included in the complete-case smoking analyses. After adjusting for age and sex, smokers had nearly three times the odds of being hospitalised than non-smokers (OR=2.83, 95% CI 2.25 to 3.56) with ex-smokers at nearly twice the odds (OR=1.88, 95% CI 1.59 to 2.23). After

adjusting for comorbidities, smokers had 96% higher odds of hospitalisation than non-smokers, and ex-smokers 37% higher (see online supplementary file D).

Further analyses

The age-adjusted and sex-adjusted odds of dying in the 28 days post-CAP decreased progressively over the study period, with patients in 2009–2010 having a 38% reduction in the odds of dying within 28 days of CAP compared with those in 2001–2003 (table 4). Length of hospital admission decreased slightly over the study period, from 8 (IQR 4–16) days in 1998–2000 to 7 (IQR 3–13) days in 2009–2010 (table 4). The percentage of short-term admissions (<2 days) increased over time from 11.7% to 14.1%. The majority of admissions (95.6%) occurred on the date of the CAP diagnosis.

Emergency admissions recorded as being via Accident and Emergency (A&E) increased successively, from 50.6% of post-CAP admissions in 1998–2000 to 76.4% in 2009–2010. Conversely, emergency admissions coded as arriving via a GP fell from 41.6% to 18.1%, and there was a corresponding fall in the percentage of CAP events with a CAP or potential CAP record in the GP data, from 58% to 34%.

Risk_factors	OR (95% CI)
Sex	
Female	0.82 (0.76, 0.87)
Age (years)	
70-74	1.35 (1.18, 1.53)
75-79	1.44 (1.27, 1.63)
80-84	1.65 (1.45, 1.87)
85-89	1.56 (1.37, 1.77)
90+	1.18 (1.03, 1.34)
INDIVIDUAL CO-MORBIDITIES	
Congestive heart failure	1.25 (1.15, 1.36)
Chronic lung disease	1.62 (1.47, 1.78)
Peripheral vascular disease	1.27 (1.14, 1.41)
Cerebrovascular disease	0.98 (0.91, 1.06)
Neurological disease	1.08 (0.97, 1.21)
Dementia	0.71 (0.65, 0.78)
Severe renal disease	1.82 (1.64, 2.03)
Connective tissue disease	1.21 (1.07, 1.37)
Other immune mechanism disorders	2.49 (1.49, 4.15)
Peptic ulcer	1.17 (1.05, 1.32)
Leukaemia/lymphoma	1.94 (1.55, 2.43)
Terminal illness	0.30 (0.25, 0.35)
Ischaemic heart disease	
pre-MI	1.34 (1.23, 1.46)
post MI	1.31 (1.17, 1.46)
Liver disease	
Mild	1.71 (1.09, 2.66)
Severe	1.80 (1.03, 3.15)
Diabetes	
Diabetes, no complications	1.27 (1.14, 1.40)
With complications	1.31 (1.08, 1.59)
Solid cancer	
Non-metastatic	1.26 (1.15, 1.38)
Metastatic	2.46 (1.97, 3.07)
FRAILTY FACTORS	
Visual impairment	1.10 (1.02, 1.18)
Bedsore/ulcer	0.52 (0.44, 0.61)
Low weight/poor nutrition	0.84 (0.77, 0.93)
Incontinence/catheter	0.83 (0.75, 0.92)
Recent carer	1.17 (0.97, 1.41)
Living arrangements	
Lives alone	1.17 (0.97, 1.40)
Sheltered accommodation	1.11 (0.83, 1.48)
Residential Care	0.46 (0.41, 0.51)
MEDICATIONS/VACCINATIONS	
Steroids in previous 90 days	1.42 (1.27, 1.58)
Immunosuppressants	1.10 (0.84, 1.45)
Statins in previous 6 months	1.12 (1.01, 1.23)
Inhaled corticosteroids	
Within 60 days	0.99 (0.88, 1.12)
Within 61-180 days	0.83 (0.69, 0.98)
Within 181-365 days	0.93 (0.72, 1.21)
More than 365 days ago	0.82 (0.70, 0.95)
Antibiotics	
In previous 1-7 days	1.04 (0.95, 1.14)
In previous 8-28 days	0.69 (0.63, 0.76)
Influenza vaccine last received	
14-365 days pre CAP	0.75 (0.68, 0.83)
last season	0.76 (0.67, 0.85)
2-5 years pre CAP	0.82 (0.71, 0.96)
>5 years pre CAP	0.90 (0.70, 1.15)
Pneumococcal vaccine last received	
14-365 days pre CAP	1.02 (0.89, 1.18)
1-2 years pre CAP	1.02 (0.89, 1.17)
2-5 years pre CAP	1.07 (0.97, 1.18)
>5 years pre CAP	1.16 (1.04, 1.29)

X-axis: 0   .5   1   1.5   2   2.5   3   3.5   4   4.5

Figure 1 Mutually adjusted ORs (circles) with 95% CIs (lines) of hospitalisation in the 28 days after CAP for factors included in the final model*. The model also contained year of CAP diagnosis, but the results for year are not presented. *Baseline categories were age 65–69 years; condition or medication not present (for comorbidities, frailty factors, recent medications); unvaccinated/no record of vaccination (for influenza and pneumococcal vaccination) (CAP, community-acquired pneumonia; MI, myocardial infarction).

Open Access

Table 2 Results of the effect of age on post community-acquired pneumonia hospitalisation, in males and females

Age (years)	Male OR (95% CI)*	Female OR (95% CI)*
65–69	1	1
70–74	1.35 (1.13 to 1.61)	1.34 (1.10 to 1.61)
75–79	1.49 (1.26 to 1.76)	1.39 (1.16 to 1.66)
80–84	1.65 (1.40 to 1.96)	1.61 (1.35 to 1.93)
85–89	1.63 (1.36 to 1.94)	1.47 (1.23 to 1.75)
≥90	1.59 (1.31 to 1.94)	1.00 (0.84 to 1.19)

*Adjusted for: year, ischaemic heart disease, congestive heart failure, peripheral vascular disease, dementia, chronic lung disease, connective tissue disease, peptic ulcer, liver disease, diabetes, cancer, leukaemia/lymphoma, severe renal disease, cerebrovascular disease, neurological disease, disorders of the immune mechanism, terminal illness, recent carer, place of residence, vision problems, bed ulcer, underweight/nutritional replacement, incontinence/catheter, immunosuppressants (other than steroids), steroids, inhaled steroids, statins, antibiotics in previous 28 days, influenza vaccine.

DISCUSSION

This is the first UK study to use large linked data sets to explore the factors associated with hospitalisation among CAP cases, and thus help identify high-risk patients for proactive case management. The factors we investigated had varying effects on hospitalisation. We were able to identify a wide range of patient factors that increased the odds of hospitalisation, including conditions common in older populations such as chronic lung disease, ischaemic heart disease, congestive heart failure, severe renal disease and diabetes. Analysis of the subset of data with near-complete recording of smoking status illustrated that smoking is also a strong risk factor for hospitalisation, independent of comorbidity status.

Table 3 Average predicted probability (%) of hospitalisation within 28 days of community-acquired pneumonia diagnosis, by year

| Year | Average predicted probability of hospitalisation (%) | | |
	No adjustment	Adjusted for age, sex and comorbidities*	Full model†
1998–2000	58	57	57
2001–2003	70	67	68
2004–2006	80	76	78
2007–2008	84	80	81
2009–2010	89	85	86

*Comorbidities: ischaemic heart disease, congestive heart failure, peripheral vascular disease, dementia, chronic lung disease, connective tissue disease, peptic ulcer, liver disease, diabetes, cancer, leukaemia/lymphoma, severe renal disease, cerebrovascular disease, neurological disease, disorders of the immune mechanism, terminal illness.
†As for comorbidities, with addition of: recent carer, place of residence, vision problems, bed ulcer, underweight/nutritional replacement, incontinence/catheter, immunosuppressants (other than steroids), steroids, inhaled steroids, statins, antibiotics in previous 28 days, influenza vaccine.

Factors associated with decreased likelihood of hospitalisation included terminal illness, specific frailty factors and receipt of residential care. Individuals who had been recently vaccinated against influenza and those with recent antibiotic treatment were also less likely be hospitalised. An unexpected finding was that the oldest women (but not men) in our study were not at increased risk of hospitalisation compared with younger women in adjusted analyses; one possible explanation for this is a survivor effect among the oldest women.

Unlike previous studies that have reported pneumonia hospitalisation trends, we were able to demonstrate that hospitalisation after a CAP diagnosis is increasing independently of any trends in CAP incidence.[2 18] We additionally found that this increase does not appear to be driven by the underlying health and social issues of the older population. The average predicted probability of hospitalisation in the 28 days after CAP increased substantially in this population over the study period, from 57% to 86%, after extensive adjustment for changes in the prevalence of patient factors. All-cause mortality in the 28 days post-CAP and length of hospitalisation both decreased over the study period, with an increasing proportion of short-term (<2 day) admissions, suggesting that the increase in hospitalisation was not linked to increasing CAP severity. Owing to the lack of information on illness severity in these data, we cannot ascertain directly if less severely ill patients are being hospitalised over time, or whether hospital treatment has helped reduce mortality in the 28 days after admission.

We also found that over the study period progressively lower proportions of patients arrived in A&E after referral from their GP or with evidence that the GP had seen them for a LRTI on the day of CAP diagnosis, highlighting changes in patients' health-seeking behaviour.

Strengths

The use of large linked data sets meant that we could distinguish between community and hospital-acquired pneumonia and include non-hospitalised CAP episodes, which enabled assessment of risk factors specifically for hospitalisation. The very large linked data allowed thorough investigation of individual comorbidities and other variables, many of which are incompletely recorded or unrecorded in hospital admission data. The advantage of investigating individual comorbidities, compared with using a summary comorbidity score such as the Charlson score, is that we avoided masking of opposing associations of individual comorbidities on hospitalisation. For example, a Charlson score of 1 is given to a patient who has dementia, or to a patient with chronic lung disease. According to our analysis, a patient with dementia would have reduced odds of hospitalisation after CAP, whereas a patient with chronic lung disease would have increased odds. The linked data also allowed assessment of trends in hospitalisation independent of trends in CAP incidence, with detailed adjustment to account for any changes in the prevalence of patient

Open Access

Table 4 Post-CAP mortality, length of hospital admission and consultation behaviour on the day of CAP diagnosis, over time

	Year				
	1998–2000	2001–2003	2004–2006	2007–2008	2009–2010
OR for mortality within 28 days of CAP diagnosis*	1.01 (0.93 to 1.10)	1	0.84 (0.78 to 0.91)	0.73 (0.67 to 0.79)	0.62 (0.57 to 0.68)
Length of hospital admission median, (lower-upper quartile), days	8 (4–16)	8 (4–17)	8 (4–15)	7 (3–14)	7 (3–13)
0–2	468 (11.7)	660 (10.5)	981 (11.9)	999 (14.2)	1093 (14.1)
2–6	1138 (28.4)	1831 (29.2)	2595 (31.4)	2364 (33.6)	2731 (35.3)
7–13	1145 (28.6)	1737 (27.7)	2311 (27.9)	1827 (26)	1987 (25.7)
≥14	1256 (31.3)	2037 (32.5)	2383 (28.8)	1848 (26.3)	1930 (24.9)
Reason for admission, n (% of those hospitalised)					
Emergency: via A&E	2027 (50.6)	3760 (60)	5559 (67.2)	5073 (72.1)	5914 (76.4)
Emergency: via GP	1666 (41.6)	2016 (32.2)	2231 (27)	1522 (21.6)	1402 (18.1)
Emergency: via bed bureau	102 (2.5)	107 (1.7)	125 (1.5)	114 (1.6)	121 (1.6)
Emergency: via consultant outpatient clinic	26 (0.6)	42 (0.7)	43 (0.5)	41 (0.6)	42 (0.5)
Emergency: other means (including A&E from another place)	66 (1.6)	135 (2.2)	143 (1.7)	122 (1.7)	127 (1.6)
Transfer (non-emergency), elective, not known	120 (3)	205 (3.3)	169 (2)	166 (2.4)	135 (1.7)
Admitting diagnosis ICD10 Chapter X— diseases of the respiratory system	3798 (94.8)	5979 (95.4)	7939 (96)	6743 (95.8)	7426 (95.9)
Hospitalisations on CAP diagnosis date, n (% of those hospitalised)	3718 (92.8)	5893 (94.1)	7896 (95.5)	6804 (96.7)	7539 (97.4)
Relevant diagnosis on CAP date† (n, % all CAP)					
CPRD only	3234 (46.5)	3074 (34.3)	2546 (24.4)	1607 (19.1)	1265 (14.4)
HES only	2909 (41.8)	4644 (51.8)	5990 (57.4)	5266 (62.6)	5787 (65.7)
CPRD and HES	809 (11.6)	1249 (13.9)	1906 (18.3)	1538 (18.3)	1752 (19.9)

*Adjusted for age and sex using three-level model.
†General practice records included LRTI records as 'potential CAP', to allow for potentially conservative coding by GPs in the absence of radiographical confirmation of pneumonia (see Results section). HES records included any hospital admission record.
A&E, Accident and Emergency; CAP, community-acquired pneumonia; CPRD, Clinical Practice Research Datalink; GP, general practitioner; HES, Hospital Episode Statistics; ICD, International Classification of Diseases; LRTI, lower respiratory tract infection.

risk factors for hospitalisation over time. The linked CPRD population is representative of the population of England, making our findings generalisable to the population at large, and the linked hospital data enabled us to identify the outcome (hospitalisation) with minimal misclassification. Further linkage to central mortality records allowed us to estimate mortality without restricting analyses to the subset of patients who died in hospital, thus avoiding changes in mortality over time due simply to changes in the relative proportions of patients who died in and outside hospital.

In contrast to previous studies that used only the first CAP episode in a year, we included patients with repeated episodes of CAP.[18] [19] The association between specific comorbidities and hospitalisation could be particularly strong in this small but important subset of patients, and inclusion of their multiple episodes avoided potential underestimation of these associations.

Limitations
Validity of recorded diagnoses is generally high in CPRD, and comorbidities and other risk factors that were only recorded after the hospitalisation were not included; thus, any misclassification of these factors is likely to be relatively small and non-differential with respect to the outcome.[20] The linked data enriched our

9

overall comorbidity coding, but smoking histories were under-recorded in the earlier years of the study period. The introduction of the Quality Outcomes Framework (QOF) in 2004 has improved the recording of smoking status in GP records, and analysis of the subset of data with near-full recording of smoking indicated its importance as a risk factor for hospitalisation after CAP. Similarly, while the use of GP data enabled investigation of some factors associated with frailty, these factors were not frequently recorded by GPs which limited our ability to assess fully their association with hospitalisation. Owing to the frailty indicators included in the data, we were unable to use an established measure of frailty such as the frailty phenotype or frailty index.[13 14] However, the frailty index includes several of the comorbidities we included individually in our model, such as diabetes, myocardial infarction and lung disease, and so use of this score may have led to overadjustment for these other important conditions. We examined a wide variety of variables; thus, estimates in the final model with a 95% CI close to including the null value should be interpreted with caution.

The HES pneumonia diagnoses used in this study have not been validated. There have been small localised reports of overdiagnosis of pneumonia in English hospitals, but trends over time at a national level have not been reported.[21-23] As such we cannot exclude that overdiagnosis could have played a role in the increasing level of hospitalisation after CAP identified in this study. The forthcoming British Thoracic Society audit of CAP diagnoses will help to clarify this issue. Nevertheless, it is likely that the majority of these patients had a respiratory illness that was considered severe enough to merit hospitalisation, and these are of public health importance.

The data sources used in this analysis did not contain direct measures of pneumonia severity, such as those in the CURB score, and so severity of illness could not be measured directly.[24] However, our aim was to establish patients' pre-existing conditions which contributed to the increase in hospitalisation over time, not the mechanism by which this occurred (either by altering severity of CAP or by other means), and so we do not feel this detracts from our study findings. Furthermore, a recent systematic review highlighted suboptimal performance of CURB scores for oldest patients, and stressed the need to focus more on the presence of comorbidities and frailty in these patients.[25] Similarly, due to the nature of the coding used in these data, we were not able to examine the severity of many of the comorbidities we investigated, and so could not directly assess whether increasing severity of these comorbidities over time could have explained some of the increase in hospitalisations during the study period. However, where we could distinguish categories of severity (eg, for diabetes, liver disease, renal disease and ischaemic heart disease), the likelihood of hospitalisation was very similar for those with severe and milder manifestations of the condition, and adjustment for these factors did not materially affect increasing hospitalisations.

Findings in relation to other studies

We have previously shown that CAP incidence is rising among older individuals in the UK.[6] This study confirms that hospitalisation following a CAP diagnosis is also increasing, with a growing percentage of cases hospitalised within 28 days of diagnosis. Our findings enhance those from previous studies that used stand-alone hospitalisation data, which have shown increasing hospitalisations for pneumonia without distinguishing increasing CAP incidence from an increasing tendency to hospitalise patients with CAP.[2 18] In particular, our analyses of hospitalisation trends update and extend those of a previous English study, which reported increasing pneumonia hospitalisation rates between 1997 and 2004 after less extensive adjustment for comorbidities, using the Charlson Index.[18] Other studies that have investigated individual risk factors among older patients with CAP or LRTI have mostly been small and included fewer factors; some used hospitalisation or death as a composite outcome, which will obscure the opposing effects of conditions such as dementia or terminal care on these two outcomes.[19 26-29] Our finding that influenza vaccine receipt is associated with protection against hospitalisation after CAP is also consistent with the direction of effect shown in previous studies of influenza vaccine effectiveness against hospitalised CAP,[30 31] and studies showing a relative lack of long-term protection of pneumococcal vaccine, especially among those with underlying health conditions.[32 33] Our findings also add to those from a recent systematic review, which highlighted between-study heterogeneity in the association between statin use and outcomes of pneumonia.[34] The reduction in mortality seen over the study period echoes that from the earlier English study which focused on in-hospital mortality, as well as CAP mortality studies from Europe and the USA.[18 35 36]

Meaning, explanations and implications for future research

The risk factors identified in this study will be of benefit to clinicians managing patients in primary care settings, by helping to identify patients at high risk of unplanned admission to hospital who are in need of proactive case management. Our findings will further inform discussions with these patients about protecting against infection risk and seeking early treatment for symptoms.

Frailty is currently a health priority in the UK. The requirement in the 2014 general practice contract for increased identification of vulnerable older members of the practice population may result in better recording of frailty in general practice data and enable more thorough investigation of its effects on hospitalisation in future research.[37] The latter will be helped by a new

171

Open Access

primary care electronic Frailty Index, currently under development in England.[38]

Despite their importance in identifying high-risk patients, our adjusted analyses show that increasing prevalence of comorbidities and frailty are not driving the increase in hospitalisation rates. Declining mortality and length of hospital stay indicate that this is not due to increasing disease severity. What then explains the increasing hospitalisation trend? The guidelines for management of CAP issued by the British Thoracic Society have not changed significantly over the study period.[39 40] However, the diagnostic accuracy of pneumonia may have changed over time. An emergency department-based US study found that the accuracy of pneumonia diagnoses decreased after the change of a core quality measure (time to first antibiotic dose) from 8 to 4 h.[41] In England, the introduction of the 4 h A&E waiting time target in 2004 could have had a similar effect.

In addition, changes to service provision and utilisation have been highlighted as playing a role in the increase, with the change in out-of-hours access to GPs during the study period.[42] The effect of this is difficult to measure directly, but we found decreasing emergency admissions over time arriving via a GP, and a decreasing proportion of patients with a CAP or potential CAP recorded by their general practice on their CAP diagnosis date. Studies have also shown that the increase in emergency admissions among older individuals in England is not restricted to pneumonia but are seen for a range of other conditions, and that the percentage of patients who attended A&E and were then admitted rose by over a third between 2003 and 2012, with 75% of this rise attributed to increasing emergency admissions and 25% to an increase in A&E attendance.[43] Thus, an increasing tendency to hospitalise, coupled with an increasing inclination of patients to present to A&E rather than to their general practice, may be a main driver of the growth in hospitalisation after CAP. It would be interesting to compare our results with those from equally detailed studies that use linked data to investigate risk factors and hospitalisation trends for other conditions, such as COPD or cellulitis. Results from these studies would allow further interpretation of whether increasing hospitalisation and decreasing primary care consultation trends are not specific to CAP among older adults.

Our study, based on very large numbers (minimising random error) and with the ability to adjust hospitalisation rates for many factors, supports the argument that focusing on high-risk patients, while important for risk stratification, will not appreciably reduce emergency admissions.[42] If the incidence of CAP among those aged ≥65 years also continues to increase, these combined trends will place a joint expanding burden on the health service.

REFERENCES

1. Harrison MJ, Dusheiko M, Sutton M, et al. Effect of a national primary care pay for performance scheme on emergency hospital admissions for ambulatory care sensitive conditions: controlled longitudinal study. *BMJ* 2014;349:g6423.
2. Bardsley M, Blunt I, Davies S, et al. Is secondary preventive care improving? Observational study of 10-year trends in emergency admissions for conditions amenable to ambulatory care. *BMJ Open* 2013;3:pii: e002007.
3. Blunt I. *Focus on preventable admissions: trends in emergency admissions for ambulatory care sensitive conditions, 2001 to 2013*. London: The Health Foundation and The Nuffield Trust, 2013.
4. NHS England. *Avoiding unplanned admissions enhanced service: proactive case finding and care review for vulnerable people. Guidance and audit requirements*. London: NHS England, The BMA and NHS Employers, 2014.
5. BMA. Guide to avoiding unplanned admissions. http://bma.org.uk/practical-support-at-work/contracts/gp-contracts-and-funding/general-practice-funding/unplanned-admissions/contact-patient-following-discharge-from-hospital (accessed 04 Feb 2015).
6. Millett ER, Quint JK, Smeeth L, et al. Incidence of community-acquired lower respiratory tract infections and pneumonia among older adults in the United Kingdom: a population-based study. *PLoS ONE* 2013;8:e75131.
7. Campbell J, Dedman JD, Eaton SC, et al. Is the CPRD GOLD population comparable to the U.K. population? *29th International Conference on Pharmacoepidemiology & Therapeutic Risk Management*; Montreal: Pharmacoepidemiol Drug Saf, 2013:280.
8. CPRD. The Clinical Practice Research Datalink—CPRD. 2014. http://www.cprd.com/intro.asp (accessed 23 May 2014).
9. Health and Social Care Information Centre. HES Online. 2014. http://www.hscic.gov.uk/hes (accessed 23 May 2014).
10. Smeeth L, Thomas SL, Hall AJ, et al. Risk of myocardial infarction and stroke after acute infection or vaccination. *N Engl J Med* 2004;351:2611–18.
11. National Institute for Health and Clinical Excellence. About the Quality Outcomes Framework. http://www.nice.org.uk/aboutnice/qof/qof.jsp (accessed 23 May 2014).
12. Department of Health. Commissioning for Quality and Innovation (CQUIN) payment framework. http://www.institute.nhs.uk/world_class_commissioning/pct_portal/cquin.html (accessed 23 May 2014).
13. Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci* 2001;56:M146–57.

Open Access

14. Mitnitski AB, Mogilner AJ, Rockwood K. Accumulation of Deficits as a Proxy Measure of Aging. *TheScientificWorldJOURNAL* 2001;1:323–36.
15. Rabe-Hesketh S, Skrondal A. *Multilevel and longitudinal modeling using Stata*. Vols I and II. 3rd edn. Stata Press, 2012.
16. How can I compute predictive margins for xtmelogit with random effects? UCLA: Statistical Consulting Group. http://www.ats.ucla.edu/stat/stata/faq/predictive_margins_xtmelogit.htm (accessed 23 May 2014).
17. van Vugt SF, Verheij TJ, de Jong PA, *et al.* Diagnosing pneumonia in patients with acute cough: clinical judgment compared to chest radiography. *Eur Respir J* 2013;42:1076–82.
18. Trotter CL, Stuart JM, George R, *et al.* Increasing hospital admissions for pneumonia, England. *Emerg Infect Dis* 2008;14:727–33.
19. Winchester CC, Macfarlane TV, Thomas M, *et al.* Antibiotic prescribing and outcomes of lower respiratory tract infection in UK primary care. *Chest* 2009;135:1163–72.
20. Herrett E, Thomas SL, Schoonen WM, *et al.* Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *Br J Clin Pharmacol* 2010;69:4–14.
21. Pick H, Lacey J, Hodgson D, *et al.* S15 clinical characteristics of hospitalised patients misdiagnosed with community-acquired pneumonia. *Thorax* 2014;69(Suppl 2):A10.
22. Gildeh N, Errington T, Maxwell I, *et al.* The impact of miscoding of community-acquired pneumonia in a UK district general hospital. *Eur Respir J* 2014;44(Suppl 58):A204.
23. Pink K, Mitchell I, Davies H. P17 the accuracy of a diagnosis of pneumonia in a UK teaching hospital. *Thorax* 2012;67(Suppl 2):A71.
24. Lim WS, van der Eerden MM, Laing R, *et al.* Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* 2003;58:377–82.
25. Foss CH. Are the Pneumonia Severity Index (PSI) and CURB-65 suitable in very old patients hospitalized with pneumonia? *Eur Geriatr Med* 2013;4(Suppl 1):S115.
26. Hak E, Wei F, Nordin J, *et al.* Development and validation of a clinical prediction rule for hospitalization due to pneumonia or influenza or death during influenza epidemics among community-dwelling elderly persons. *J Infect Dis* 2004;189:450–8.
27. Bont J, Hak E, Hoes AW, *et al.* A prediction rule for elderly primary-care patients with lower respiratory tract infections. *Eur Respir J* 2007;29:969–75.
28. van de Nadort C, Smeets HM, Bont J, *et al.* Prognosis of primary care patients aged 80 years and older with lower respiratory tract infection. *Br J Gen Pract* 2009;59:e110–15.
29. Chan Carusone SB, Walter SD, Brazil K, *et al.* Pneumonia and lower respiratory infections in nursing home residents: predictors of hospitalization and mortality. *J Am Geriatr Soc* 2007;55:414–19.
30. Jefferson T, Di Pietrantonj C, Al-Ansary LA, *et al.* Vaccines for preventing influenza in the elderly. *Cochrane Database Syst Rev* 2010;(2):CD004876.
31. Baxter R, Ray GT, Fireman BH. Effect of influenza vaccination on hospitalizations in persons aged 50 years and older. *Vaccine* 2010;28:7267–72.
32. Andrews NJ, Waight PA, George RC, *et al.* Impact and effectiveness of 23-valent pneumococcal polysaccharide vaccine against invasive pneumococcal disease in the elderly in England and Wales. *Vaccine* 2012;30:6802–8. .
33. Huss A, Scott P, Stuck AE, *et al.* Efficacy of pneumococcal vaccination in adults: a meta-analysis. *CMAJ* 2009;180:48–58.
34. Khan AR, Riaz M, Bin Abdulhak AA, *et al.* The role of statins in prevention and treatment of community acquired pneumonia: a systematic review and meta-analysis. *PLoS ONE* 2013; 8:e52929.
35. van Gageldonk-Lafeber A, Bogaerts M, Verheij R, *et al.* Time trends in primary-care morbidity, hospitalization and mortality due to pneumonia. *Epidemiol Infect* 2009;137:1472–8.
36. Ruhnke GW, Coca-Perraillon M, Kitch BT, *et al.* Marked reduction in 30-day mortality among elderly patients with community-acquired pneumonia. *Am J Med* 2011;124:171–8.e1.
37. Changes to the General Medical Services contract for 2014/15 for England. http://www.nhsemployers.org/PayAndContracts/GeneralMedicalServicesContract/GMSContractChanges/Pages/Contractchanges201415.aspx#4 (accessed 27 May 2014).
38. New approach needed to frail elderly care. NHS England. 2014. http://www.england.nhs.uk/2014/05/07/elderly-care/ (accessed 27 Jun 2014).
39. British Thoracic Society Standards of Care Committee. BTS guidelines for the management of community acquired pneumonia in adults. *Thorax* 2001;56(Suppl 4):iv1–64.
40. Lim WS, Baudouin SV, George RC, *et al.* BTS guidelines for the management of community acquired pneumonia in adults: update 2009. *Thorax* 2009;64(Suppl 3):iii1–55. doi:10.1136/thx.2009.121434.
41. Welker JA, Huston M, McCue JD. Antibiotic timing and errors in diagnosing pneumonia. *Arch Intern Med* 2008;168:351–6.
42. Roland M, Abel G. Reducing emergency admissions: Are we on the right track? *BMJ* 2012;345:e6017.
43. Morse A. *Emergency admissions to hospital: managing the demand*. London: The Stationery Office, National Audit Office, 2013.

Supplementary files B-D can be found in Appendix G. A more detailed version of the online supplementary file A from the paper is presented below.

6.4 Supplementary Methods used to identify the presence of variables

This section expands upon the supplementary methods referred to in Paper 2 (online material A), including a more detailed account of how co-morbidities, frailty factors, medications vaccinations and lifestyle factors were identified.

6.4.1 Co-morbidities

The large size of the data included in this study, and detailed medical history contained within patients' linked records, enabled me to investigate the effects of individual co-morbidities rather than using a summary score such as the Charlson index.[68] I initially included all 19 co-morbidities in the Charlson index (as defined in the original 1987 paper),[68] as well as additional conditions which were thought to potentially affect the probability of being hospitalised following CAP (neurological disease, disorders of the immune mechanism, and terminal illness).

While it may be expected that GP records would contain the majority of a patient's historical and current diagnoses, there is added detail and value in also utilising the HES data.[148] Therefore, records for co-morbidities were sought from both CPRD and HES records in order to collect as complete a medical history for each patient as possible.

To date, there is no real consensus on how far back to look through a patient's medical records in order to identify the presence of underlying disease, and this will depend on the condition and the study question. 'Lookback' periods used in other papers (when specified) range from a year[149] to all available records.[64] If a patient has a long-standing chronic condition, this may not be recorded repeatedly in medical records, necessitating longer lookback periods to identify it. Additionally, when the outcome of interest can occur multiple times per patient (as with CAP), using a short lookback period could result in the patient fluctuating between disease states for different CAP episodes depending on when the last co-morbidity record occurred. For example, if a patient was diagnosed with COPD and nine months later had an episode of CAP, the patient would have COPD listed as a co-morbidity using a one-year lookback. If six months later the

174

patient had another CAP episode, but the COPD had not been re-recorded the patient would not be classified as having COPD during this illness. In order to avoid this scenario and to allow me to gather as much information on the cohort of older patients as possible, I did not place time restrictions on how long ago co-morbidities could be recorded in order to be considered in these analyses. Furthermore, I did not limit the records I used to those recorded after a certain time point (such as the date the practice was deemed 'up to standard' by CPRD), but included records for co-morbidities from any point in a patient's records prior to the CAP episode. This strategy was adopted as I was interested in whether the patient had ever experienced each disease rather than identifying incident cases, and so any recording of historical conditions potentially on the wrong date (pre-CAP) was not a concern.

Read (CPRD) and ICD-10 (HES) code lists were defined for each co-morbidity of interest by one or more senior clinical epidemiologist at LSHTM, taking into account previous published code lists.[150] These code lists were merged with the data, and the date of the earliest record recorded separately for CPRD and HES records. If either of these earliest co-morbidity dates were prior to or on the CAP incident date, the patient was classified as having the condition. For some conditions there were additional adaptations to this process, as described below.

6.4.1.1 Differentiation between acute and chronic co-morbid events

Conditions recorded for the first time on the same date as an incident CAP record were treated differently for acute and chronic illnesses. For conditions such as MI, stroke, and falls which may have been precipitated by a CAP episode, records up to the day before the CAP incident date were used.[151, 152] Records for chronic conditions (for example diabetes without complications), that were first recorded on the CAP incident date were included, as CAP is not known to lead to the onset of diabetes.

6.4.1.2 Co-morbidities with more than one level of severity

Specific co-morbidities were categorised with more than one level of severity, including ischaemic heart disease (categorised with no evidence or evidence of past MI), solid cancer (without/with metastases), liver disease (moderate and severe) and diabetes (without/with complications). Data were time-updated such that patients were classified as having the less severe category of disease until the date of their first code

at the higher level, when they permanently changed to the higher level. If patients only had codes for the higher level (e.g. metastatic cancer without a prior code for solid cancer) they were immediately categorised as having the higher level of disease.

6.4.1.3 Additional, non-Charlson broad co-morbidity terms that included several diseases

<u>Neurological disease</u> included Parkinsons disease and other extra-pyramidal/movement disorders, epilepsy, multiple sclerosis and other demyelinating diseases, systemic atrophies primarily affecting the central nervous system, hereditary and idiopathic neuropathy, cerebral palsy, Creutzfeldt-Jakob disease and atypical viral infections of the central nervous system.

<u>Disorders of the immune mechanism</u> included aplastic anaemia, immunodeficiency with predominantly antibody defects, combined immunodeficiencies, immunodeficiency associated with other major defects, common variable immunodeficiency, functional disorders of polymorphonuclear neutrophils, chronic myeloproliferative disease, HIV infection, and unspecified immunodeficiency. AIDS was also included when calculating the Charlson score (see section 6.5.2).

<u>Terminal illness</u> was defined using Read and ICD-10 codes that stated terminal illness, rather than specific conditions. In addition, GP information on referrals to hospices was included.

6.4.2 Frailty factors

Frailty is an important topic when considering the health of older adults. As discussed in section 1.2.1.1, the definition and classification of frailty is complex and the cause of much discussion, and currently there are several working definitions of frailty and associated ways to measure it. Two popular approaches (described in section 1.2.1.1) are the frailty phenotype, which recognises frailty from a set of five deficits, and the frailty index, which utilises a cumulative deficit approach (counting a number of factors for each patient across a range of systems, including co-morbidities, disabilities and clinical signs and symptoms).

It was not possible to categorise frailty using either the frailty index or frailty phenotype, as neither CPRD nor HES data include all the factors needed to calculate either score, such as grip strength, gait speed, or neck muscle strength.[73, 153] Instead, I developed a series of variables based initially on those factors that are included in the frailty index which could possibly have been recorded in a patient's records.[153] Read, and when possible ICD-10 code lists were developed (by myself and Sara Thomas), incorporating information from additional recording fields within the data such as the entity type, and consultation type (when the type of consultation was recorded, e.g. residential home visit). Unlike co-morbidities, only codes used within the previous year were counted as evidence of a frailty factor due to the ability of the frailty state to fluctuate. The exception to this was visual impairment, for which more complex rules were developed as explained in detail later in this section. The specific methodology used for each factor is laid out below.

6.4.2.1   Frailty factors included if they were recorded in the year pre-CAP

Anxiety/depression utilised Read and ICD-10 codes for anxiety and depression.

Bedsores/ulcers included Read and ICD-10 codes for bedsores and pressure sores/ulcers.

History of falls used information from both CPRD and HES to define whether a patient had fallen within the last year (excluding records on the CAP incident date). Codes for hip fractures were also included in this category as it has been estimated that up to 95% of hip fractures are caused by falls.[154] Codes for falls were included if the reason for the fall could have been due to a deficit in underlying health. For example, codes pertaining to an accidental fall not associated with potential loss of function (such as falling from a bicycle) were not included.

Incontinence included Read, ICD-10 and entity codes for bladder or bowel incontinence, and codes and prescriptions for catheters, leg bags and associated products.

Mobility problems included Read codes and entity codes for problems with general mobility and with stairs, walking, being house- or chair-bound, immobile, in a wheelchair, as well as use of a zimmer/walking frame and difficulties getting in/out of bed.

Body mass index (BMI) was initially investigated, but the data were not included as a potential risk factor due to high levels of missing data and a lack of timeliness in the records available. A quarter of patients included in the analysis had neither a recorded BMI value in their records nor information on their height and weight to enable BMI to be derived. Furthermore, for those patients with a BMI record, timing of the record relative to the next CAP episode was suboptimal for much of the study period. Less than 40% of CAP episodes between 1997 and 2002 had a BMI value recorded in the previous two years, reaching 60-70% by the end of the study period. This difficulty is not specific to this CAP cohort, as these values are similar to those presented over time for patients aged ≥65 years in a recent data resource profile of CPRD.[99] In this older population with relatively high prevalence of co-morbidities, weight recorded several years prior to a CAP episode may not be accurate by the CAP incident date. For example, patients may have gained weight due to mobility problems and inactivity, or have lost weight due to ill health or depression. Thus, for weight records to accurately reflect a patient's current BMI status they need to have been recorded relatively recently prior to the CAP episode.

Previous studies have shown that compared to patients with a normal weight/BMI (and after taking other factors such as co-morbidity into account), patients who were overweight or obese were not at higher risk of developing CAP,[31, 155] and there is conflicting evidence whether higher BMI is a risk factors for being hospitalised for CAP.[122, 156] However, there is stronger evidence that patients who were underweight were at increased risk of developing CAP,[31, 155] and being underweight is also an aspect of frailty. Due to the high level of missing data and lack of timely recording for BMI, it was decided not to include BMI in this analysis. Rather, I specifically looked at whether patients were recorded as underweight or needing nutritional supplementation, as described below.

Low weight/poor nutrition included Read, ICD-10, prescription and entity type codes for weight loss, malnourishment, lack of appetite, anorexia, and prescriptions for nutritional supplementation. As described above, due to the long period between BMI measurements and CAP episodes, low BMI values were not included.

Recent support from a carer was defined in CPRD using Read codes and entity types in the year before the CAP episode. Records were restricted to the prior year in order to

strike a balance between excluding historical records for patients who were no longer in receipt of care while including those for patients with long-term carers which may not be recorded frequently.

It was hypothesized that a change in housing might result in modified use of a carer. Thus, patients could be classified as having a carer irrespective of which residence category they were in (defined below), but the carer variable was modified accordingly. Patients whose most recent residence code was within a year of the CAP index date were only classified as having a carer if the carer code was more recent than the residence code. If there were no records relating to the patient's place of residence or the residence code was recorded more than a year before the CAP index date, then the patient was coded as having a carer if it was coded in the year prior to the CAP.

Self care problems included Read and entity codes for an inability to; wash/clean oneself, maintain personal hygiene, get dressed without assistance, use the toilet without assistance, perform housekeeping activities, buy and/or prepare food.

Tiredness included Read codes regarding tiredness, fatigue, malaise, lethargy and chronic fatigue syndrome.

6.4.2.2   Frailty factors included if recorded at any point pre-CAP

Place of residence was defined using CPRD records only, as over 85% of records in HES were coded as admitted from 'usual place of residence' which could relate to independent housing or residential care. Within CPRD the most recent record (using Read codes or the consultation type field) was used, as patients could move in and out of residential care, for example while recuperating after an illness or fall. The categories living alone, sheltered accommodation and residential care were treated as being mutually exclusive, and when more than one category was used on a date the less independent place of residence was assigned. The category residential care also included nursing homes – the Read codes used did not differentiate well between residential and nursing homes, so it was necessary to use a single category for both types of care.

Visual impairment was defined using records from both CPRD and HES (using Read, ICD-10 codes and entity codes). To aid data management, visual codes were categorised as

long-term conditions (such as blindness), and those which were potentially treatable. The treatable conditions were further categorised as cataracts (diagnosis or surgery) and other eye conditions. Within HES, codes within the primary code of the first episode (the main condition treated) were assumed to be admissions for cataract removal surgery.

Patients with any long-term condition, other visual impairment at any point before CAP, or with a cataract diagnosis code but no cataract surgery code before the CAP, were coded as having a vision problem.

6.4.2.3 How well did my approach really capture frailty?

I chose not to use a cumulative deficit approach in my risk factors for hospitalisation analyses, as inclusion of both a frailty index and individual co-morbidities in the regression models may have led to over-adjustment of co-morbidities. However, the use of the individual aspects of the frailty index to adjust for frailty (as I did) is not a validated approach, and may not have captured patients' frailty status adequately. The frailty index weights each included factor equally when estimating the cumulative deficit, resulting for example in patients with four different co-morbidities and two signs/symptoms having the same score as a patient with one co-morbidity and five signs/symptoms (both being classified as 'mild' frailty). The model I built would not have had the same result, as the majority of co-morbidities were associated with increased odds of hospitalisation, whereas the other frailty factors I considered were not. Thus, in my model the deficits would not have accumulated in a comparable way, and patients with a high number of non-co-morbidity frailty factors would most likely have been assigned lower odds of hospitalisation than patients with multiple morbidities but no other frailty factors. While my approach allowed me to investigate a wide range of components of frailty, it did not enable me to explore any potentially cumulative effect these factors may have had, and thus may not have fully captured patients' frailty status.

6.4.2.4 A new approach - the Electronic Frailty Index

Since I completed the work on this thesis, a study which developed and validated an electronic frailty index for use in English primary care records has been published.[157] Clegg et al developed and internally validated (using the split sample approach) their electronic frailty index (eFI) using a UK primary care data source (ResearchOne), and

then externally validated the index using a separate primary care data source (The Health Improvement Network, THIN).

Patients aged 65-95 years who were registered with a practice contributing to either data source on 14 October 2008 were included, and these patients were followed-up for up to five years. Thirty-six deficits were chosen for inclusion in the index across a range of categorises including disease states, signs/symptoms, and disabilities. Using the cumulative deficit approach, each patient had their eFI calculated as the number of deficits present divided by 36 (the total number of possible deficits). Four categories of frailty were defined, based on the quartiles of the frailty score: 1) Fit (those with scores of 0-0.12, i.e. 0 to 4 deficits present); 2) Mild frailty (scores of >0.12-0.24, 5 to 8 deficits); 3) Moderate frailty (scores of >0.24-0.36, 9 to 12 deficits), and 4) Severe frailty (scores of >0.36, 13 or more deficits). The derivation and internal validation cohorts assigned 50% of patients as fit, 35% mildly frail, 12% moderately frail and 3% severely frail. In the external validation THIN cohort, fewer patients were assigned as fit (43%), with slightly higher levels of moderately frail (16%) and similar levels of mildly frail (37%), and severely frail (4%).

These four categories were then assessed against a range of outcomes; mortality, emergency hospitalisation and nursing home admission at one, three and five years. The authors described the discrimination of the model as good for the outcomes of mortality and nursing home admission (internal validation c-statistics ranged from 0.72 at 1 year to 0.69 at 5 years for mortality) and 'moderate' for emergency hospitalisation (c-statistic 0.66 at 1 year to 0.63 at 5 years). External validation showed improved discrimination for mortality (0.76 and 0.75) and emergency hospitalisation (0.71 and 0.69) at one and five years, although the hospitalisation values in particular still have room for improvement.

6.4.2.5 Limitations of assessing frailty using EHR

The new eFI appears to have a moderate ability to identify older adults at risk of emergency hospitalisation within one year. In general, frailty indices include a range of factors, as some of the physiological changes that bring about frailty do not result in specific diseases. Consequently, an index which only includes co-morbidities (for example, the Charlson score) will not identify all frail patients. Many of the factors

included in the eFI are co-morbidities, and thus the score relies on the completeness of coding of the other frailty factors in order to differentiate it from a simpler 'co-morbidity count' approach.

As highlighted in the paper above and discussed in section 6.5.4, one potential disadvantage of using a score approach is that individual co-morbidities may have opposing effects on the risk of hospitalisation, and these opposing effects may not be identified when a score is used.

No matter whether a frailty index (such as the eFI) or an individual factor approach (as in my analysis) is chosen, the most important limitation of assessing patients' frailty using EHR is the lack of information on validity of the recording of non-disease factors. For example, when codes are used to denote that a patient has mobility problems (for example, an inability to climb stairs), these are likely to have been correctly applied to patients who have these signs/symptoms, and the codes are therefore likely have a high positive predictive value. However, it is possible that not all patients with mobility problems are assigned such a code, either because they have not reported this issue to their GP, or because the GP has noted the issue (for example in a free text field) but not formally coded it. This would result in the coding of these factors having low sensitivity and thus a patient's frailty score would be underestimated.

The increasing interest in frailty and its current priority status in the UK may result in more complete recording of the non-disease frailty factors in general practice records.[158] Updates of the eFI in future years may benefit from this, resulting in an improved discriminative ability for emergency hospitalisation, and it will be interesting to see how the updated eFI changes over time.

6.4.3 Medications

Information for all medication variables was obtained from the therapy files in CPRD. Linked-HES data do not currently include information on medication prescribed during a patient's hospital admission, and as a result drugs given during any hospital admission could not be included in the analysis.

Prescription recording in CPRD can include information on the date of prescription, drug name, quantity of tablets prescribed, the numeric daily dose and length of treatment. Unfortunately the information on daily dose and length of treatment are sometimes poorly recorded. For medications such as oral steroids it can thus be difficult to work out whether some patients were prescribed a short high-dose course or a low long-term maintenance dose. To work around this difficultly, I used both the time between the CAP incident date and the closest preceding prescription, as well as the length of a prescription (when recorded) in order to define exposure to the drug of interest, using different time windows depending on the medication.

Inhaled corticosteroids: Timing of inhaled corticosteroids (ICS) prescribing was based upon recent work by Suissa et al.[33] CAP episodes without an ICS prescription in the previous year were classed as 'non-users'. Those who did have a prescription were categorised by their most recent prescription pre- CAP; up to 60 days prior, 61-180 days prior, and 181-365 days prior to the CAP diagnosis.

Immunosuppressants other than steroids: Patients were categorised as being recent/current immunosuppressant users if they had a prescription record within 120 days of their CAP event date (assuming a 90 day prescription, and 30 day washout period).

Oral steroids: As highlighted above, the categorisation of oral steroid use was challenging due to incomplete information in the numeric daily dose (36.9% missing) and length of treatment course (93.6% missing) fields. This was a particular issue when patients had prescriptions for multiple tablet strengths (e.g. 1mg and 5mg) on the same day, and I was unable to differentiate regimens in which a patient was tapering their dose from those in which the patient was simply on a dose non-divisible by 5 (e.g. 8mg). To combat this problem, the timing of oral steroid use was based upon the strategy adopted by Dixon et al.[159] This study explored several models for categorising exposure to oral steroids using routinely collected GP data, and the 90 day model had the best fit of the conventional 'binary' models presented. Patients were categorised as being recent/current oral steroid users if there was evidence of a prescription for oral steroids within the 90 days pre-CAP, or if they had a prescription that would last into this period. When prescriptions were issued up to 14 days pre-CAP (but not in the 15 to

90 days before the CAP), patients were not classified as being 'on steroids', as the treatment may have been given for an LRTI preceding the CAP (e.g. for an infection-related COPD exacerbation), and because the immunosuppressive effects of the medication may not have taken effect before the CAP episode.

Statin use was defined as a prescription within the previous six months (183 days). A longer period was used than that for steroids or other immunosuppressants, as statins are prescribed as a long term therapy and not frequently used as an acute treatment.

Antibiotic prescriptions were classified as given in the 1-7 or 8-28 days before the incident CAP diagnosis.

6.4.4 Vaccines

Vaccination status was determined using GP records alone. I decided to prioritise records from the CPRD 'immunisation' and 'therapy' files (rather than the 'clinical' files), as these were thought more likely to record a current vaccination event rather than an historical one. To account for the time taken for a vaccine to elicit an immune response, patients were considered 'unvaccinated' for the 14 days after the vaccination date.[53, 48] Timings were categorised in similar manner to Vindogrova et al, as outlined below.[160]

Influenza vaccine status was categorised to reflect that influenza vaccine is offered to patients yearly. The vaccination season was defined as 1st September to 31st August the next year.

Individuals were classed as unvaccinated if they had no influenza vaccination record prior to their CAP diagnosis, or if they were first vaccinated less than 14 days pre-CAP. Those vaccinated ≥2 weeks prior to the CAP episode were classified according to their most recent vaccination date as being vaccinated in: the current season, the previous season, 2-5 years ago or >5 years ago.

Pneumococcal vaccine status was categorised similarly to influenza vaccine, but without the use of a vaccination season; categories were unvaccinated, or vaccinated: this year, last year, 2-5 years and >5 years before the CAP diagnosis.

For both vaccines, patients with only a 'history of vaccination' code were classified conservatively as being vaccinated more than five years previously, as there was no way of identifying the correct date.

6.4.5 Lifestyle/social factors

Patients' smoking status and excess alcohol consumption were considered as potential risk factors. An episode of pneumonia may modify a patient's behaviour, potentially making them more health conscious. For this reason, only the most recent record for smoking and excess alcohol status prior to the CAP episode was used to classify a patient's status for each CAP; records occurring after their CAP diagnosis were not used.

Excess alcohol consumption: Previous studies have found a link between excessive alcohol consumption and risk of CAP.[31, 161] I used markers for excess alcohol consumption in addition to specific measures of consumption as only 20% of CAP episodes had a record regarding alcohol consumption in the previous year (rising to 31.3% in the previous two years). Excess alcohol consumption was therefore identified using a Read/ICD code for high/excess alcohol consumption, for alcoholism, or a code specifying harm due to alcohol (for example, alcohol-induced hepatitis) recorded at any point before the CAP episode. Patients with prescriptions for medications to treat alcohol dependency at any time before/on the date of the CAP episode, and those recorded as consuming ≥6 units of alcohol a day, or ≥42 units a week were also included.

Smoking status was derived using Read/ICD-10 codes, prescriptions for nicotine replacement therapy and appropriate entity codes. Patients with a code for 'non-smoker' were amended to 'ex-smoker' if they had previous current or ex-smoker records. Patients on nicotine replacement therapy were coded as current-smokers, as relapse rates are high.[162]

Exclusion of socioeconomic status: I decided not to include a measure of socioeconomic status (i.e. IMD quintile) in this analysis, as IMD quintile was not available at the individual level for more than 4000 patients (>10% of patients included in this study population). While practice-level deprivation quintiles were available for these patients, this should have been accounted for in part by the practice-level clustering included in the multilevel model. Furthermore, SES is a distal determinant of hospitalisation,

mediated by the other variables of interest such as co-morbidities and lifestyle factors. Therefore the inclusion of these factors in the multivariable model will have at least partially accounted for the effect of SES.

6.5   Additional analysis comparing co-morbidity adjustment using individual co-morbidities versus the Charlson index

6.5.1   Background

The range of co-morbidities included in the preceding analysis was based upon those in the Charlson index, supplemented by additional potentially important disease groups. The diseases in the Charlson index were used as the starting point due to the index's widespread use in epidemiological studies (the original paper has been cited >15,000 times),[163] in particular its use in the analysis of CAP,[61, 148, 164, 165] and studies using electronic health records.[61, 148, 164, 165]   By supplementing the diseases included by Charlson with other cardiac, neurological and immune disorders it was felt that a wide and comprehensive range of potentially important co-morbidities was included.

More specifically, the study of increasing hospitalisations for pneumonia in England by Trotter et al which originally inspired the analyses above found that using the Charlson co-morbidity index did not explain the increasing hospitalisation trend over time.[61] While my intention was never to use the Charlson index to adjust for co-morbidities in the main hospitalisation analysis, I thought it would be useful to compare its use to that of individual co-morbidities in investigating the underlying reasons for the hospitalisation trends.

6.5.1.1   Potential issues with using the Charlson index to adjust for co-morbidity

In the thirty years since the Charlson paper was originally published, the methods for developing prognostic models, and the prevalence and treatment of the diseases contained within the index, have changed considerably. This has resulted in heightened awareness of the limitations in its use.

Methodical limitations

Methodologically there are several problems with the approach used to develop the Charlson index. The score was developed by assigning the rounded HRs for each predictor, rather than their Beta coefficients (the log HRs). This is important, as the Beta coefficients are developed on an additive scale and therefore summing them to provide an overall score is appropriate, whereas the HRs are developed on a multiplicative scale and thus should not be summed but multiplied. This incorrect method of totalling patients' scores may decrease the calibration and predictive ability of the model.[166] Since the analyses for this thesis were completed, the effect of scoring Charlson co-morbidities using HRs rather than the Beta coefficients has been investigated (in addition to several other methods) using CPRD records for older adults, to predict their one-year mortality.[167] To compare the performance of these different scores, each was compared to a logistic regression model adjusted for age and sex only. The score that used the individual factors' Beta-coefficients was found to fit the data better than the original Charlson score. When compared to the classification provided by the simpler model (adjusted only for age and sex), the Beta coefficient model was found to reclassify patients into the correct risk stratum a higher percentage of the time (9.43%) compared to the original Charlson score (6.95%).[167]

Current advice is that use of Beta coefficients that have not been rounded to the nearest integer may produce models that better fit the data and have better performance. This was also assessed by Mehta et al, and again the Beta coefficient model was a better fit to the data and correctly reclassified patients into the correct risk stratum to a greater extent (9.43%) than the Beta model that was rounded to the nearest integer (7.79%).[167] This rounding can result in diseases that have very different associations with increased mortality being assigned the same score. For example, when scored using HRs (as Charlson did), conditions assigned a score of two may have been associated with anywhere between 50% and 150% increased mortality. Rounding of scores for individual conditions does make the index easier to use in a clinical setting, where time pressures may make scores with more complex sums inappropriate. However, in a research setting the inclusion of this extra detail should not be problematic.

The exclusion of predictive factors below a certain threshold (in the case of Charlson, diseases with HRs<1.2 were not included) is also now considered poor practice. In

addition to excluding diseases that have a comparatively small but positive association with the outcome, this threshold also removes all diseases that decrease the risk of the outcome (as these have a HR of less than 1.0). In order to fully understand a patient's risk, all factors that may positively or negatively influence their score should be included.

Additional limitations

An important consideration when using the Charlson score in epidemiological studies is the loss of information on the effects of individual diseases on the outcome of interest. In order to calculate the score, the presence of each co-morbidity must be established and therefore it would require little or no extra work to include these individual coefficients in the final adjusted model. This would enable researchers and clinicians alike to assess the importance of each disease on a variety of outcomes, and discover in which scenarios certain co-morbidities were protective and others of no importance. While use of a co-morbidity score is simpler in that it only necessitates the inclusion of one categorical variable in a model rather than several additional variables, it also results in a potential wealth of information being lost.

Treatments for the diseases that are included in the score and their survival probabilities will have changed over the last thirty years, and thus some of the HRs that informed the score (and their Beta-coefficients) may be out of date. Therefore, the predictive performance of the model may have decreased over time.

Furthermore, use of a summary index may not be appropriate when assessing trends over time. As the co-morbidities included in the index may have opposing effects on the risk of the outcome, trends in the prevalence of individual conditions may affect the outcome rate in ways that are not identified when using a summary measure.

Attempts have been made at updating the Charlson index, but as yet they have not been as popular or widely used as the original index.[168, 169] Despite its limitations, the Charlson index is still frequently used to adjust for patients' underlying co-morbidity status. The following analysis assessed the added benefit of using individual co-morbidities rather than the Charlson index to investigate the increasing trend in hospitalisation after CAP.

6.5.2 Methods

Information on the 19 variables included in the Charlson index had already been gathered for use in the analyses presented in Paper 2. The presence of co-morbidities was assessed as outlined in the paper and in section 6.4.1. I also calculated Charlson scores for individuals at each CAP episode using the scores in Table 1-1 and then categorised as none (0), mild (1-2), moderate (3-4), severe (≥5).[68]

The OR for hospitalisation in the 28 days after CAP by time period was then re-calculated using three-level unadjusted and minimally adjusted models (including age and sex), with the addition of the categorised Charlson score.

6.5.3 Results

Over 40% of patients were assigned a Charlson score of 1-2, and 27% a score of 3-4. The unadjusted odds of hospitalisation after CAP increased steeply with increasing Charlson score, although this was somewhat attenuated when additionally adjusted for age, sex and year (Table 6-3).

Table 6-3 Number of CAP episodes that resulted in hospitalisation within 28 days by Charlson score, and unadjusted and minimally adjusted ORs.

Charlson score	Hospitalised within 28 days	Not hospitalised	Total	Unadjusted OR (95% CI)	Minimally adjusted OR (age, sex, year)
0 (n (%))	3980 (67.6)	1906 (32.4)	5886	1	1
1-2 (n (%))	13105 (73.1)	4830 (26.9)	17935	1.57 (1.42 - 1.73)	1.37 (1.25 - 1.5)
3-4 (n (%))	9380 (79.4)	2432 (20.6)	11812	2.61 (2.33 - 2.92)	1.9 (1.71 - 2.11)
≥5 (n (%))	6856 (86.3)	1087 (13.7)	7943	4.98 (4.32 - 5.73)	2.8 (2.46 - 3.18)

Interestingly, adjusting the average predicted probability of hospitalisation after CAP over time using the Charlson score resulted in very similar results to those which were adjusted using individual co-morbidities (presented in Paper 2). In the two earlier time periods use of the Charlson model resulted in probabilities 2% and 1% lower than the individual co-morbidity model, but from 2004 onwards both models provided the same estimates (Table 7.5).

Table 6-4 Average predicted probability (%) of hospitalisation within 28 days of CAP diagnosis, by time period, including adjustment for Charlson score

Calendar period	Average predicted probability of hospitalisation (%)			
	No adjustment[†]	Charlson score (categorised)*	Individual co-morbidities included in final model[†]	Full model[†]
1998-2000	58	55	57	57
2001-2003	70	66	67	68
2004-2006	80	76	76	78
2007-2008	84	80	80	81
2009-2010	89	85	85	86

[†]Originally presented in Table 3 of Paper 2, presented here for ease of comparison
*Charlson score categorised as 0, 1-2, 3-4 and ≥5

6.5.4 Discussion

There was very little difference between the average predicted probability of hospitalisation after CAP when adjusted for co-morbidities using the Charlson score or when adjusted for individual conditions. Multivariable analyses indicated that increasing prevalence of co-morbidities accounted for only 3-4% of the increase in hospitalisation post-CAP over the study period, irrespective of whether the co-morbidities were included individually or as the Charlson score. It may be that any difference between the two methods of adjustment was not noticeable when co-morbidities appeared to account for such a small percentage of increasing levels of hospitalisation over time.

Nevertheless, this analysis demonstrates that investigation of the effect of individual co-morbidities (rather than use of a summary score) provided valuable information. As outlined in Paper 2, the use of individual co-morbidities instead of a summary score in the models developed in this study showed that the diseases included in the Charlson score have contrasting directions of effect on the odds of hospitalisation. For example, dementia and chronic lung disease both have a Charlson score of one, but dementia was found to decrease the odds of hospitalisation post-CAP while the presence of chronic lung disease increased the odds. While both models provided similar population averaged results, the use of individual co-morbidities in this work has provided greater insight into the characteristics of patients who were hospitalised post-CAP than would have been possible if this work had been limited to using the Charlson score.

My intention in this analysis was to highlight the added value of investigating the effect of individual co-morbidities on the risk of hospitalisation after CAP, as opposed to using a summary co-morbidity index. Had I been aiming for a more in-depth investigation of the utility of the Charlson index compared to the utility of models that included individual co-morbidities, additional analyses could have been attempted. For example, I could update the index, following the suggestions of Quan et al and reassigning weights to each co-morbidity based on more recent mortality data,[169] or using the version by Mehta et al that corrects the erroneous use of HRs by deriving the score using the Beta coefficients.[167] Addressing these limitations of the existing Charlson index could enable better assessment of the relative benefits of using a combined co-morbidity measure and individual co-morbidities. However, this would not deal with the differences in the direction of risk for hospitalisation and death for some of the co-morbidities of interest. Development of a new co-morbidity index for the risk of hospitalisation after infection, although outside the scope of the present work, could enable further examination of the combined effects of co-morbidities on hospitalisation.

6.6   Implications of the work presented in this Chapter

In this Chapter I have shown that hospitalisations after CAP among older adults have risen independently of any increase in CAP incidence, and that changes in CAP patients' underlying health status does not appear to explain this trend. Nevertheless, the risk factors I identified will help GPs identify patients at risk of unplanned hospital admission, should they become ill with CAP. I have also highlighted the poor level of recording of frailty factors in primary care, which is particularly timely given the 2014 general practice contract requirement to identify older vulnerable members of the practice population.[158] Further work is needed to better understand the real impact of changes in service provision and utilisation on hospitalisation trends over time.

Increasing hospitalisations and decreasing mortality post-CAP over time have also resulted in a growing population of older adults with complex underlying health status who survive pneumonia hospitalisations. These patients then return to the community, where they are cared for by their GP. It seems that an increasing number of these patients present straight to hospital when ill, resulting in their GP being unaware of their illness until after the event. As discussed in section 1.3.2.4, older adults hospitalised for

pneumonia have a higher mortality risk for at least a year post-discharge compared with patients hospitalised for other reasons. In the next Chapter I present the final analyses of this thesis, which aimed to assist GP decision making by identifying patients at higher risk of dying in the post-discharge period.

Chapter 7 Development of prognostic models for long-term mortality risk for patients after CAP hospital discharge, in order to assist GP decision making

In Chapter 6 I showed that hospitalisations within 28 days of a CAP diagnosis among older adults increased between 1998 and 2011, but over the same period mortality (either directly or indirectly due to CAP) in the 28 days after a CAP diagnosis declined. Thus there is a growing population of older patients who survive hospitalisation after CAP and are released back into their GPs' care. In this Chapter I examine predictors of subsequent mortality for these individuals, in order to try to aid GP decision making about plans for future care and support.

I first provide a brief rationale for this study, and review the literature on risk factors associated with increased mortality after adults with CAP are discharged from hospital. This is followed by an outline of the methodology behind prognostic modelling, a statistical technique that enables the development of risk scores to aid and inform clinical decision making. I then apply these methods to the patients in the linked CPRD-HES cohort who were hospitalised for CAP and survived the hospitalisation. The mortality rate and cause of death of these hospitalised CAP patients in the year after their discharge is examined, and I create a series of prognostic models to try to help GPs identify patients with a high predicted mortality risk over this period. Finally, I discuss the limitations of my approach and alternative strategies which could be used to tackle this aim.

7.1 Rationale – the need for prognostic models to predict mortality post-CAP in older adults

7.1.1 Increasing GP interaction with older adults and those at high risk of hospital admission

As outlined in section 1.3.2.4, patients hospitalised for CAP have a higher mortality risk for at least a year post-discharge compared with both the general population and with patients hospitalised for other conditions.[88, 92] The combination of several trends has resulted in an increasing number of older adults belonging to this higher-risk group

of individuals who survive a CAP hospitalisation: an expanding older population, the rising incidence of CAP, the rising number of hospitalisations for CAP and decreasing in-hospital mortality following CAP.

These patients are returned to the care of their GP, who may learn about the CAP episode for the first time via a hospital discharge summary and who then need to decide what future care the patient requires. This has recently become a particular focus for a specific group of patients, as changes to the GP contract (2014/15) in England have resulted in 'at-risk' patients being placed on the Enhanced Service (ES) register. Patients placed on this list (a minimum of 2% of the practice list) have been identified as at risk of an unplanned admission to hospital, and their general practice is required to contact them within three days of discharge should admission occur.[170] Given the increased mortality risk after a CAP hospitalisation, a prognostic model to assess a patient's risk of death in the year after CAP hospital discharge would be useful for GPs at the time of this post-discharge contact, in order to inform decisions on the kind of support to be offered to the patient by the practice. While not all patients on the ES register will be older adults, patients are chosen using a risk stratification tool or clinical judgement, making it highly likely that many will be aged ≥65 years. Additional changes have resulted in all patients aged ≥75 years being assigned a named GP who is responsible for general oversight of their care.[171] A simple to use prognostic model for mortality that requires minimal clinical input could also prove useful in health care planning for this group after a CAP hospitalisation.

7.1.2 Limitations of currently available models

Prognostic models are available to predict mortality risk post-CAP over short risk periods, however commonly used tools such as CURB-65 and the PSI were designed to be used at the point of CAP diagnosis, which may make them less suitable for use post-discharge.[81, 80] Patients who die during a CAP hospitalisation are likely to have different health profiles to those who survive, and therefore these populations need to be considered separately. For example those with cardiovascular disease may have high in-hospital mortality (where it would thus be an important predictor of mortality), and so these patients would be less well represented in the post-discharge population (where it may be a less important predictor). Fitting prognostic models to CAP survivors'

characteristics would result in more accurate future risk predictions than those obtained from a model used at the point of hospital admission. Additionally, models such as CURB-65 and PSI include clinical signs and symptoms present at the point of diagnosis such as respiratory rate, which are unlikely to be available in a primary care setting when the GP takes over a patient's care post-discharge.

As described in section 1.3.2.3, only three scores have been specifically developed for use in older patients in an outpatient or primary care setting.[85-87] Unfortunately these three models were not limited to patients with pneumonia but also included patients with other LRTI, and all used a combined end point of hospitalisation or death. This combined outcome is problematical, as some of the risk factors for hospitalisation are likely to differ from those for death, or the direction of effect may not be the same. None of the previous primary care models were optimised for patients who survive a CAP hospitalisation, or to predict mortality over a period longer than 30 days after diagnosis.

A wealth of information regarding patients' medical histories, such as their co-morbidity profile, vaccination status and lifestyle factors is available to GPs. A model including readily accessible factors such as these would therefore be appropriate in the CAP post-discharge setting.

7.1.3   Aim

The objective of this study was to develop easy to use prognostic models to predict the risk of death in the year after a patient's discharge from a CAP hospitalisation. The results generated by the score could be used in addition to the GP's own knowledge of the patient to aid in planning the patient's future care. An important feature of a good prognostic model is ease of use. Some of the software used for clinical management by GPs already has clinical risk scores built in, such as QRISK, a score to assess cardiovascular risk.[172] The prognostic models developed in this study were designed to similarly utilise patients pre-existing electronic health records, and so require minimal clinical input.

Before discussing the methodology behind developing a prognostic model, I looked at the existing literature around risk factors for mortality post CAP-discharge, to see what

models were currently available, and to inform my choice of potential prognostic factors.

7.2 Literature review of factors associated with long-term mortality after a CAP hospitalisation

7.2.1 Aim

The aim of this literature review was to identify studies of risk factors and prognostic models for mortality after a CAP hospitalisation in older adults. Factors of interest were those which would be readily recorded in general practice records, or provided in a hospital discharge summary (for example co-morbidities, history of vaccinations and lifestyle factors). Results from physical or laboratory examinations performed at point of hospital admission or during a hospitalisation (and therefore not immediately available to GPs) were not specifically of interest but were noted for completeness.

7.2.2 Methods

7.2.2.1 Search strategy

A review of research about longer-term prognosis after an episode of community-acquired pneumonia was published in 2013 by Restrepo et al.[93] I updated this review by searching the Medline database for additional studies published since the review. The Medline search was performed using a combination of MeSH and free text terms for community-acquired pneumonia, mortality, older adults and risk factors/prognosis (see Appendix H).

7.2.2.2 Inclusion criteria

The Medline search was restricted to original research articles written in English and added to the database between 2011 (to overlap with the end of the Restrepo review) and May 2015. Case series were excluded, as were animal studies. In order to restrict included research to that broadly generalisable to my older English study population, only papers from high-income countries were included, and the study population had to include (but not necessarily be restricted to) patients aged 65 years and over.

As the interest of this work was in mortality after hospital discharge, studies which included in-hospital mortality (alone or in combination with post-discharge mortality) were excluded from the review, as predictors of death during admission are likely to differ from those post-discharge. Additionally, studies which included patients treated for CAP as outpatients in addition to inpatients (and did not analyse the two groups separately) were excluded. The rationale for this decision was that patients hospitalised for pneumonia will differ from those who are not hospitalised, either by severity of CAP or prevalence of severe co-morbidities including terminal illness (for which CAP may be an expected end of life event). Thus patients who are not hospitalised are likely to have a different set of risk factors for mortality than hospitalised patients. Similarly, studies set in an intensive care unit (ICU) were excluded, as ICU patients represent a more severe subset of CAP cases, and survivors of an ICU admission may differ from the general CAP hospitalisation cohort.

Studies were also excluded if they; reassessed or compared established severity scores (such as PSI/CURB-65) without incorporating any new factors, focussed on CAP due to specific pathogens (e.g. *Legionella* pneumonia, as patients who have pathogens identified may differ from the wider older population), assessed specific treatments such as a type of antibiotic on patients' mortality risk, focussed on aspiration pneumonia (as this was not included in my study definition of CAP), built models comparing mortality in CAP patients to that of those with HAP or HCAP, or used combined endpoints such as ICU admission/mortality.

7.2.2.3 Article screening

I initially screened all articles cited in the Restrepo review and those identified in the Medline search using their title and abstract, applying the inclusion/exclusion criteria outlined above. For papers which appeared to meet the criteria, I obtained the full paper and again checked against the criteria above. I also checked the reference lists of included articles for further possible papers of interest. As in the previous reviews, if I was uncertain of a papers eligibility I discussed it with my supervisor and a decision was reached by consensus.

7.2.2.4 Data extraction

I extracted data from the eligible papers into a standardised form in Excel. Study characteristics of interest included the study population, year and duration of the study, method of ascertaining the outcome and mortality rate, demographics, and risk factors included in the final fully adjusted analyses.

Study findings were summarised narratively. Aspects of the study that were considered good or poor were noted, but a formal quality assessment was not performed.

7.2.3 Results

The Medline search identified 276 papers, of which only two were eligible for inclusion in this review. Of the 85 articles cited by Restrepo, three were found to be eligible for this review, providing a total of five included studies. The key findings of these studies are discussed below, and summarised in Table 7-1.

7.2.3.1 Included studies

The post-hospitalisation mortality risk period investigated by the studies varied: two reported follow-up of one year,[173, 174] one ended after 18 months [175] and two lasted several years.[176, 92] Four studies were analyses of risk factors for mortality, and one developed a new prognostic model.[175] None of the papers were restricted to older adults (all included patients aged ≥18 years), but of those that summarised participants' age, only one had a mean age <60 years (Waterer, mean age=58.1).[176]

Further reporting of the factors found to be associated with longer-term mortality are discussed below, categorised by the length of follow-up used in the model.

Table 7-1 Characteristics of included studies, and the risk factors identified as associated with longer-term mortality among CAP patients post-discharge

Length of follow-up	365 days		>365 days		
Author Year (REF)	Carriere 2004 [173]	Adamuz 2014 [174]	Guertler 2011 [175]	Waterer 2004 [176]	Bruns 2011 [92]
Region, Country	Alberta, Canada	Barcelona, Spain	6 centres, Switzerland	Memphis, USA	Multiple centres, Netherlands
Study period	1994-2000	2007-2011	2006-2008	1998-2001	2000-2003
Study design	Retrospective cohort	Prospective cohort	Prospective cohort	Prospective cohort	Prospective cohort
Study population	Patients (age ≥18y) in two administrative health databases for whom diagnosis of CAP was considered most responsible for hospital admission	All adults with CAP admitted to hospital	All CAP patients (age ≥18y) previously enrolled in multicentre trial (ProHOSP) who survived initial CAP hospitalisation.	All CAP patients (age ≥18y) admitted to Methodist Healthcare Memphis Hospitals for whom written consent could be obtained.	Hospitalised patients (age ≥18y) from two, multicentre RCTs of antibiotic treatment strategies for CAP
Case ascertainment (mortality)	Vital statistics declaration of death certificates merged with the Alberta Health Care Premium Registry databank	Patient healthcare database of Catalan Health Service (routinely collects data on hospitalisations and in-hospital or home mortality).	Vital status ascertained in telephone interviews on days 30, 180 & 540 Relatives or treating primary care physician interviewed if patient could not be contacted	Social security number-linked death records, review of hospital and outpatient pharmacy records, contact with all known treating physicians, and postal contact at the last known home address.	Ascertained post discharge using the Dutch Municipal Public Health Records Database
Sample size	43642	1284	877	366	356
Age (years)	Median (IQR) age-group: 65-75y (45-64, 75-84)	Age ≥65y: 779 (60.7%)	Median 73 (59-82)	Mean 58.1y (range, 18–99)	Mean 66y SD 16.1
Male, n (%)	NR	844 (65.7)	512 (58)	165 (45.1)	131 (6.8)
Post CAP mortality rate	26%	7.2%	18 months: 17.3%	34% (mean follow-up 1058 days)	1 yr: 17% 5 yrs: 43% 7 yrs: 53%

Length of follow-up	365 days		>365 days		
Author **Year (REF)**	**Carriere** **2004 [173]**	**Adamuz** **2014 [174]**	**Guertler** **2011 [175]**	**Waterer** **2004 [176]**	**Bruns** **2011 [92]**
FACTORS INVESTIGATED	(ORs)		(HRs)		
Age (years)	18-44: ref 45-64: 2.54 65-74: 3.05 75-84: 3.90 ≥85: 6.18	≥65: 1.59	<59: ref 59-73: 1.5 73-82: 2.0 >82: 3.0		≥65: 1.96
Male gender	1.39	x	1.7		
Chronic respiratory disease/COPD*		1.76	1.5		1.72 (SS)
Diabetes		2.13	x		
Chronic heart disease/ failure		1.71	0.8		1.72 (SS)
Cardiovascular disease				1.72	1.72 (SS)
Cancer*		2.62	2.5	x	2.07
Renal disease*		1.33	1	x	1.72 (SS)
Cerebrovascular disease*		1.01		2.52	
Dementia		3.86			
Liver disease				x	1.72 (SS)
In nursing home	3.43	2.03			2.04
Prognostic score:	Co-morbidity count: 1: 2.3 2: 3.77 >2: 4.79	PSI >90: 1.35	PSI x		PSI >90: 2.13
Other factors investigated	Readmitted <30d: 7.49 Race Aboriginal: 0.48 Export from home region: 1.45 Respiratory failure/arrest: 1.52 Per capita hospital beds: 0.90 Hypotension/ shock: 1.51 Urban residence: x Rural hospitals: x Regional hospitals: x	Steroids: 1.31 Gastric acid suppressants: 0.94 Aspiration pneumonia: 1.94 Non-stable on discharge: 1.34 Smoking: x Chronic liver disease x Previous CAP x Alcohol x Influenza vaccine x Pneumococcal vaccine x Statins x	Chills: 0.6 Temp: 37-37.9: 0.8 37.9-38.7:0.9 >38.7: 0.4 Procalcitonin x WBC x Albumin x ProADM & CRP x	Altered mental state: 3.13 Haemocrit <35%: 1.61 Increasing blood glucose: x Other variables in PSI (individually): x	

*different categories used in different models, x = factor investigated but not selected into final model, SS= summary measure used for the presence of any of these variables. RCT: randomised controlled trial, PSI: pneumonia severity index, WBC: white blood cell count, ProADM: ProAdrenomedullin,

Mortality within one year of hospital discharge

Two studies examined risk factors for mortality within a year of CAP hospital discharge, using logistic regression.[173, 174] One was a large retrospective cohort study using administrative data from Alberta, Canada which found a high one-year mortality of 26.1% among patients who were discharged from hospital after an episode of CAP.[173] The second was a prospective cohort study which recruited all patients who were admitted to a hospital in Barcelona, Spain with CAP. They reported a lower one-year mortality rate of 7.2%.[174]

Both studies identified increasing age as a risk factor for mortality, although only the Canadian study was large enough to show this over several age groups among older adults (Table 7-1). The study also showed that women had 28% lower odds of mortality than men after adjusting for all other factors in the model.[173] The two papers reported different approaches to including co-morbidities in their models. The Canadian study used a co-morbidity count, and showed increasing odds of mortality with increasing levels of multimorbidity (Table 8-1).[173] In contrast the Spanish study investigated the effect of individual co-morbidities, and found that diagnoses of dementia, cancer and diabetes had the largest odds of mortality; patients with COPD, chronic renal disease, chronic heart disease and cerebrovascular disease also had increased odds of death and were included in the final model (Table 8-1).[174] Surprisingly, a high PSI score (≥90) was included in the model in addition to many of the individual factors used to calculate this score, thus over-adjusting for these variables. Smoking status, alcohol use, influenza and pneumococcal vaccinations were investigated by the study but not included in the final model (Table 7-1). Both studies found that patients in nursing homes had higher odds of death than patients who lived in the community (OR=3.43 and OR=2.03 respectively).

Mortality in periods longer than a year after hospital discharge

Three studies used prospectively collected data to examine risk factors for mortality over periods of more than one year post-CAP discharge, using Cox regression with one study developing a prognostic model (Table 7-1). The studies were all smaller than those that examined the mortality up to a year post-discharge. The Swiss study included patients previously hospitalised for CAP who had been enrolled in a multicentre trial and

had the shortest duration with follow-up of up to 18 months.[175] The US study recruited CAP patients admitted to hospitals in Memphis, and followed patients for up to four years post-discharge. Finally, a study in the Netherlands created a cohort of hospitalised CAP patients from two randomised placebo-controlled trials of antibiotic treatment strategies, and collected information on mortality for up to seven years.[176, 92] Mortality rates in these studies ranged from 17.3% at 18 months to 53% at seven years.[92, 175]

All three studies included age in their final risk models for mortality. In the Dutch study, those aged ≥65 years were found to have almost twice the mortality rate of those aged <65 years (Table 7-1).[92] The Swiss study used smaller age groups and showed that mortality rates increased with rising age amongst the older population (Table 7-1).[175] The American study did not report the HRs for age from its final model.[176] The effect of sex was only examined by the Swiss paper, which found that men had 70% higher mortality rates than women.[175]

Again, the method of adjusting for co-morbidity varied by paper, with the Swiss and American studies investigating the effects of individual conditions,[175, 176] while the Dutch study combined several co-morbidities into one group.[92] Cancer was the only co-morbidity investigated by all three studies and was included in the multivariable models of two, where it was shown to more than double the risk of death (Table 7-1).[175, 92] The Swiss final prognostic model also included COPD, chronic heart disease, and chronic renal disease, although the two latter factors both had wide 95% CIs which spanned the null. A range of additional signs/symptoms and blood markers were also investigated, as was inclusion of the PSI, although this was not retained in the final prognostic score (Table 7-1). In contrast, the American study did not include COPD, but did include cardiovascular disease, cerebrovascular disease, altered mental status and blood test values in its multivariable model (Table 7-1). Liver and renal diseases were investigated but not included in the final model.[176]

In the Dutch study, the combined co-morbidity variable (any of malignancy, COPD, congestive heart failure, cerebrovascular disease, liver disease, and renal disease) was associated with a 71% higher mortality rate over seven years (Table 7-1).[92] However, in addition to malignancy being included in the co-morbidity variable, it was also

reported as having been included in a separate factor. The final model also included PSI score>90, in addition to several constituent parts of the PSI score as individual factors (living in a nursing home, age and three co-morbidities).

7.2.4   Discussion

Only five studies were identified which examined risk factors for long-term mortality after discharge from a CAP hospitalisation. None of the studies were set specifically among the older population, and none investigated risk factors for mortality in periods less than one year after hospital discharge. Only one study developed a prognostic model. The studies investigated a range of factors; increasing age, chronic respiratory disease, chronic heart disease, cancer, and living in a nursing home were found to be risk factors for increased longer-term mortality rates in at least three of the five studies, while male sex and cardiovascular disease were shown to increase mortality rates in two studies. All of the studies included laboratory or physical examination findings (either individually, or via inclusion of the PSI) which might not be routinely available in a patient's hospital discharge letter received by GPs.

All five studies used robust methods to ascertain deaths. However, the quality of the papers presented was highly variable with respect to the analysis. For example, the double inclusion of malignancy in the multivariable model in the Dutch study made the findings difficult to decipher.[92] Similarly, two studies included both the PSI in their final model and several of the constituent factors of PSI as individual covariates (age, cancer, liver disease, congestive heart failure, cerebrovascular disease and renal disease).[174, 92] This will have resulted in an over-adjustment for these variables, and may have led to an underestimation of their individual associations with mortality in the final models.

The American study did not present measures of effect for all co-variables included in the model and so the size and direction of their effects could not be discussed.[176] The studies investigated many potential risk factors and some of the studies had small sample sizes, but only the Swiss prognostic model discussed the power of the sample size to investigate so many factors.[175]

The Swiss study was the only one to develop a new prognostic model to predict long-term mortality risk, but the methods used to do so were incorrect.[175] In prognostic models, the log HRs from the multivariable model should be used to calculate the scores for each factor, but the Swiss study used exponentiated (i.e. non-logged) HRs, rounded to the nearest integer values. Also, the model was not validated either externally in a new dataset, or internally using boot-strapping or a split-sample approach. As such, it is not possible to say whether the model would be appropriate for use in another population. The methodological aspects of prognostic modelling are discussed in detail in section 7.3, below.

The majority of the burden of CAP, hospitalisations for CAP and mortality directly or indirectly due to CAP is among those aged 65 years and older. However, none of the studies identified were specifically set among this high risk group and the inclusion of younger patients who are likely to have differing underlying health and risk profiles will have made the studies' results less valid for those aged ≥65. There is currently a real need for specific models for use in the high-risk older population.

My literature search did not identify any studies which had looked at increased risk of mortality for periods of less than a year after discharge from a CAP hospitalisation. The cohort of patients who survive a CAP hospitalisation is likely to change over these longer time periods; in a similar manner to patients who die during hospitalisation differing from those who survive it, the risk factors and strength of effect of included factors may change over a period of a year post-discharge. Thus it seems sensible to model several different time periods within a year, in order to capture these changes and produce models which better fit the underlying trends.

Again, limitations of the review process itself need consideration. As with other reviews I conducted for this thesis, a systematic approach was used for the search strategy, eligibility assessment and data extraction. However, restriction of my search strategy to only one database (Medline) and the exclusion of papers not written in English could have resulted in some papers on the topic being excluded from the literature review. I assessed eligibility of studies single-handedly (except when I was unsure about eligibility, when I consulted with my supervisor), and extracted the data from selected

studies myself, and this will have increased the possibility of excluding eligible studies and introducing errors in the data extraction.

Despite the potential limitations of the review, it appears that there is a real paucity of information for GPs on their patients' mortality risk post-discharge after an episode of CAP. Currently, there are no simple models which utilise readily available data to aid clinical decision making post CAP-discharge. The increasing number of patients hospitalised and subsequently discharged due to CAP, coupled with the new GP service requirements mean there is a need for such a tool to be developed. In the next section I outline how these tools can be developed and assessed.

7.3   Overview of the methods used to develop and validate a prognostic model

Creating a prognostic model involves several stages. First the model itself must be developed. Secondly its performance is assessed to check that it discriminates between patients with and without the outcome, and how well it works in a population different to the one it was developed in (model validation). Finally its clinical impact is assessed. These stages are outlined below.

7.3.1   Developing the model

It is crucial that an appropriate dataset is used when developing the model - a key feature of the process is that the data used in model development should be representative of the population the model will be applied to. If this is not the case, the model is unlikely to make accurate predictions and therefore not be clinically useful. Models can be built using a range of statistical techniques. As I use Cox regression in the analyses presented below, I will concentrate in this section on the methodology appropriate to Cox models.

7.3.1.1   Candidate predictors

The model itself must be defined from a set of candidate predictor factors. To improve the generalisability of the model, these factors should be clearly defined, standardised and should be reproducible in the data the model will be applied to in a real world situation.[177] Factors considered are typically demographic, clinical or other characteristics of a patient. Factors known to be associated with or to be predictors of

the outcome under study should be included, as should those hypothesized to have an effect.[178] While causal factors ought to be included as prognostic factors where possible, not all prognostic factors will cause the outcome.[177] As with other modelling processes, predictors that are highly correlated with each other should either be combined (if both are important predictors), or only one should be included in the final model.[115]

Caution has been advised when using medications as predictors in data from observational studies, as indications for their prescription may not be completely standardised across the study population.[177] Prognostic models developed using clinical trial data frequently include medications as predictors, as the studies prescribe medications according to a strict protocol. However for studies developing models using observational data, prescribing behaviour can vary between clinicians, and thus the inclusion of medications in the model would be modelling clinician behaviour rather than the true risk associated with the medication.

In order for the study to have suitable power, it is commonly stated that the number of candidate predictors should not exceed one per ten outcomes in the study population.[179] This calculation should include all candidate predictors assessed, not just those included in the final model.

7.3.1.2 Model specification

Commonly, development of a prognostic model uses a predictive modelling strategy (described in section 2.8.2.1). Backward stepwise selection is frequently used to decide which variables should be included in the model. For prognostic modelling, it is important that the criterion for inclusion is not too stringent, as use of a $p\text{-value} < 0.05$ may exclude strong but uncommon predictors.[115]

7.3.1.3 Calculating scores and predicted risk

A patient's score is calculated by multiplying the regression coefficients for each factor in the multivariable model by the level of the factor that applies to the patient (for example the coefficient for cancer would be multiplied by 0 for no, 1 for yes), and summing these together. When using Cox regression, it is also necessary to consider the risk over time in order to calculate the expected probability of the outcome (here,

mortality). In order to obtain the predicted survival at time *t,* the total score is multiplied by the baseline survival at time *t* (as shown in Equation 7-1).[115]

Equation 7-1 Predicted survival

$$S(t|X) = S(t)^{exp(\beta X)}$$

Where S(t|X) is the predicted survival at time t given a set of predictors X,

S(t) is the baseline survival (the Kaplan-Meier estimate) at time t and

βX is the linear predictor, i.e. the sum of the logHR for each factor present in a patient.

Consequently, the predicted mortality at time *t* is:

Equation 7-2 Predicted mortality at time t

$$1 - S(t)^{exp(\beta X)}$$

After the model has been built and the prognostic scores and predicted risk calculated, the model's performance must be assessed.

7.3.1.4 Model performance

The performance of the development model is assessed by looking at the discrimination and calibration of the model.

Discrimination is the model's ability to distinguish between patients who do and do not have the event of interest. This is calculated using Harrell's concordance-statistic (c-statistic), which estimates the probability of concordance between predicted and observed responses. A c-statistic of 0.5 represents no predictive discrimination, and a value of 1 represents perfect separation of patients with different outcomes. The c-statistic of prognostic models is usually found to be between 0.6-0.85.[178]

Additionally, discrimination can be informally assessed graphically. Patients' scores are split into risk groups; if there are no clear clinical criteria on which to base these groupings, cut points can be chosen to give a good spread of risk. Ideally, these groups should be of different sizes to enable identification of patients at very high or low risk, grouping together the other patients with similar prognoses.[180] One method, developed by Cox, cuts the data into risk groups at the 16th, 50th and 84th centiles. On a

normal scale, these points are roughly equivalent to 0 and ±1, i.e. the mean value ±1 standard deviation, and they were specifically chosen to keep the loss of information (that can occur when data is split into groups) to a minimum.[180] Kaplan-Meier curves are plotted of these risk groups over time; the better the separation between the curves, the better the discrimination of the model.[180]

After discrimination, the model must be calibrated. Calibration refers to the level of agreement between predicted and observed outcomes. This can be assessed graphically by splitting the predicted mortality (calculated using Equation 8-2) into deciles, and plotting the mean proportion predicted and observed with the outcome for each decile group, with the aim that the predicted and observed risks for each decile will be similar. Greater spread between the observed deciles suggests a model with better discrimination.[181]

7.3.2   Validation of the model

The objective of model validation is to assess the model's predictive performance in a different population to that in the development dataset. While the model may describe the dataset it was developed in well, creation of the model can lead to it being too closely fitted to the data (called over-fitting), resulting in predictions that are not valid when the model is applied to a new population. It is important that any optimism in the model's performance is quantified and assessed. Optimism is the difference between the model's performance in a new population (its "true" performance), minus its performance in the development dataset (the "apparent" performance).

Validation of the model can be performed in the re-sampled original data (internal validation), or using a new data source (external validation). External validation is preferred as it truly tests the model by applying it to a completely new population, but is frequently not possible due to a lack of a second data source. Other methods have therefore been developed to enable a model to be developed and validated in the same data. The efficiency of the three common internal validation procedures (bootstrap validation, split-sample and cross-sample validation) have been compared using logistic regression models containing the same potential predictors on the same dataset to assess the best internal validation technique.[182] This analysis found that bootstrap

validation provided more stable estimates with lower levels of bias than the other techniques. Due its superiority, I used bootstrap validation in the analyses in this study.

7.3.2.1 Bootstrap validation

The steps involved in the bootstrap method are outlined below [115] (format adapted from Harrell et al [183]).

1. Develop the prognostic model (for example, using a Cox model and backwards stepwise selection to identify predictors) using all subjects (n) available in the data. Calculate the c-statistic for this model, $C_{apparent}$

2. Randomly generate a bootstrap sample of size n with replacement (meaning patients can be sampled multiple times) from the original dataset.

3. Develop a new prognostic model, using the same technique as used in the brackets in step 1.

4. Calculate the c-statistic for this new model, C_{boot}.

5. Retain the model fitted in step 3, and fit it to the original dataset. Calculate the c-statistic C_{orig}

6. Calculate the optimism of the c-statistic from the bootstrap sample, which is C_{boot} - C_{orig}.

7. Repeat steps 2 to 6, 200 times.

8. Average the 200 optimism estimates calculated in step 6 to get O.

9. Calculate the optimism-corrected c-statistic using $C_{apparent}$ minus O.

This optimism-corrected c-statistic represents the model's real ability to distinguish between patients who do and do not have the event of interest, (having taken into account the optimism of the prognostic model).

The number of times each factor is selected into the bootstrap models can be used to assess the stability of the model.[183]

7.4 Methods used in this study to develop prognostic models for long-term mortality risk after CAP hospital discharge

7.4.1 Time period of study

The study period started at 1st April 2004, when financial incentives such as QOF resulted in GP recording becoming more standardised. This resulted in more complete recording for important incentivised factors such as smoking status and diabetes. While this decision led to a reduced study population, recording habits post-2004 are closer to those used currently, and so better represent the data the model will be applied to.[99] CAP events up to 31st March 2011 were included.

7.4.2 Inclusion criteria and time 'at risk'

Patients were included in the study if they met all of the following criteria (Figure 7-1);

1) They contributed to CPRD and were eligible for hospital- and mortality-linked data at any time between 01/04/2004 and 31/03/2011.

2) They were hospitalised for CAP (defined using HES) during this period.

3) They did not die during the CAP hospital admission (or on the discharge date).

4) They were aged ≥65 years at the point of CAP diagnosis.

Patients must have survived until the day after hospital discharge in order to be included in the study, as it was not possible to distinguish patients who had died in hospital from those who died at home on the discharge date.

Figure 7-1 Data sources and population included in this study
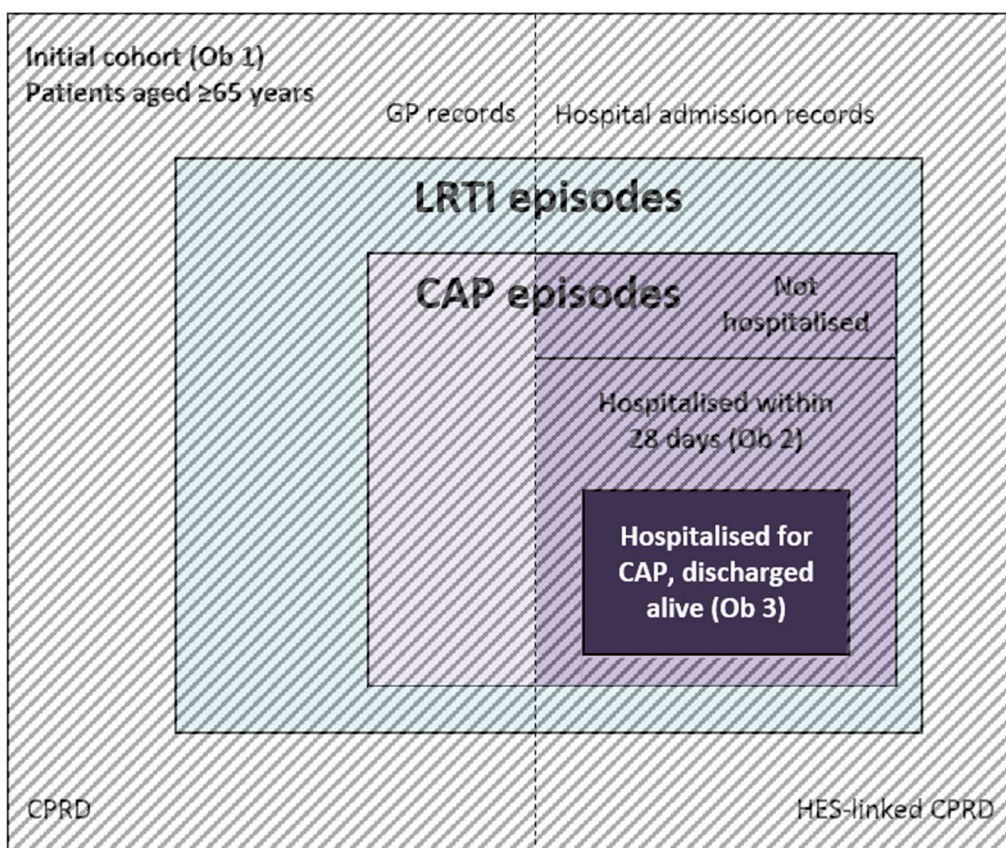
Only the first CAP episode that resulted in hospitalisation in a 365 day period was included in the study. The patient was included as 'at risk' from the date of discharge until the earliest of 365 days post-discharge, death, the study end date (31st March 2011) or the date they stopped contributing to CPRD.

Patients with no recorded smoking status prior to their CAP hospitalisation were excluded from the study, as there were very few such patients (<3%) and smoking status was considered an important potential predictor.[183] As explained in Paper 2 (Chapter 6) smoking status is unusual in CPRD as it is possible to code all three available states (non, ex, current smoker). For all other factors considered, the absence of a code was assumed to be due to the absence of the factor in question.

Patients with a terminal care code or with metastatic cancer were excluded as the death may not have been unexpected and these patients were not representative of the population the model would be applied to in future use.

7.4.3 Outcome definition

The outcome of interest was death, defined using the date of death provided in the linked ONS data. I additionally used the information on underlying and contributory causes of death from this data when investigating how causes of death changed over the year post-discharge.

7.4.4 Eligibility of potential prognostic factors

The model was designed to require very little manual input or calculation by clinical staff. Theoretically it could be built into the GP patient software and programmed to run automatically, checking through a patient's electronic records to identify codes for the factors included in the model and then calculate a patient's predicted risk of death in each time period. In order to make this possible, the presence of co-morbidities and other factors of interest were only determined from CPRD data (the data available in the GP's system), and while hospitalisations for CAP were defined using HES, the data used to develop the model did not utilise HES records.

Variables included as potential prognostic factors were limited to those whose recording was considered generally standardised and reproducible in the data the model would be applied to (i.e. general practice records outside CPRD). When there was evidence that factors were not well recorded (for example certain frailty measures such as 'immobile'), they were not considered for inclusion in the model.

Similarly, medications were not included as potential prognostic factors as it was difficult to assess if they were prescribed in a uniform fashion (as outlined in section 7.3.1.1, above). Additionally, it could not be known whether medications had been newly prescribed or altered during the hospitalisation (as this information is not provided in HES), which would have led to misclassification of medication status. Pneumococcal and influenza vaccination status were included as both are offered uniformly to all older adults.

7.4.5 Factors selected as 'candidate predictors'

Sex and age were included in all models as a priori prognostic factors. Age was categorised into five-year groups from 65 to 89 years, and then as ≥90 years.

7.4.6   Co-morbidities

Co-morbidities were coded as present if there was a Read code for the condition in a patient's CPRD record before the date of their hospital admission for CAP. Factors investigated were based on those identified in the literature review in section 7.2 as well as those in the Charlson co-morbidity index and included; chronic heart failure, peripheral vascular disease, dementia, chronic lung disease, connective tissue disease, peptic ulcer, hemiplegia, leukaemia/lymphoma, other neurological disease, other immunological disease, cerebrovascular disease, solid cancer and previous/history of pneumonia. These were binary variables, indicating presence of disease or no code for the disease. Several other variables contained an additional level of disease severity. These were diabetes (without complications, with complications) liver disease (mild/moderate, severe), ischaemic heart disease (pre-MI, post-MI), and renal disease (mild/moderate, severe).

Frailty and lifestyle factors included were: smoking status (non-smoker, ex-smoker, current smoker), living arrangements (lives alone, in sheltered housing, in residential care) underweight/poor nutrition, history of falls within the last year, recent use of a carer, and excessive alcohol consumption.

Receipt of influenza and pneumococcal vaccinations were recorded from the most recent record of vaccination as: no record/not vaccinated, vaccinated in the current year, vaccinated in the previous year, vaccinated 2-5 years previously, vaccinated >5 years previously.

7.4.7   Cause of death

Patients' cause of death was investigated using the causes recorded in the ONS-linked data. The number of deaths with pneumonia recorded as a) the underlying, or b) a contributory cause of death were calculated. Other underlying causes of death were categorised in the same way as co-morbidities used in the analysis (for example, MI, solid cancer), and underlying causes which contributed ≥2.5% of deaths in each time period were plotted graphically.

7.4.8 Statistical analyses

7.4.8.1 Models developed

The primary aim of this study was to predict longer-term post-discharge mortality risk and to identify patients who may benefit from additional attention after a CAP hospitalisation. However, the predictors of mortality are likely to vary in the immediate period post-discharge. After discussion with the clinical members of my advisory panel I decided to build three models, for risk periods 1-7 days, 8-30 days and 31-365 days after hospital discharge.

The one to seven day model covers the immediate period post-discharge, and will capture many of the subset of patients who have been discharged from hospital to enable them to die at home.

The eight to thirty day model was decided upon to transition between the immediate post-discharge risk-period, and the longer period of follow-up. GPs are likely to receive a patient's discharge summary shortly after the date of discharge, and should then contact ES patients within three days. This model would enable GPs to identify patients who may require attention relatively soon.

The final model (the model of primary interest) was the 31-365 day model. This identifies patients' longer-term mortality risk, after the initial post-CAP period is over. The timing was limited to a year, as older adults' health status may change considerably over any period longer than that, which would be likely to lead to inaccurate predictions.

7.4.8.2 Model development

All models were built using Cox regression with robust standard errors to account for clustering of CAP hospital discharges within patients. The Cox model was deemed most appropriate as the risk of death was likely to vary over time within the three risk periods, and there was a degree of censoring in the data making a logistic model inappropriate.

The total numbers of CAP episodes which resulted in hospital discharge included in each model and the number of events with the outcome (death) were calculated for each factor of interest. Univariable and minimally adjusted (age and sex) Cox regression analyses were performed for each factor to enable assessment of the influence of other

factors when examining the full model (as recommended when reporting results of prognostic model research).[184] Variables with more than two levels were included as a group of indicator variables, rather than a single factor variable in order for the bootstrapping command including robust standard errors to be able to run. This did not affect the model, but did result in the logHR for each indicator variable being multiplied by one when the patients score was calculated (rather than increasing to 2, 3, 4 as the categories increased, as illustrated using smoking status in Table 7-2).

Table 7-2 Different coding used for factor and indicator variables, using smoking status as an example

	Factor variable	Indicator variables values		
Smoking status	Values	Smoking_0	Smoking_1	Smoking_2
Never	0	1	0	0
Ex	1	0	1	0
Current	2	0	0	1

The complete set of variables was included in a backwards stepwise elimination process, with the cut-off for inclusion being a p-value<0.2 (defined using the Wald-test).[113] The p-value of 0.2 was chosen in an attempt to avoid over-fitting of the model to the data.[115] Age and sex were forced into the model, and then each condition was considered in turn. The condition with the highest p-value was removed from the model if p≥0.2. The process was then repeated until all conditions had a p-value<0.2, resulting in the final model. The c-statistic for the final model was calculated.

7.4.8.3 Calculation of risk scores and predicted risks

Risk scores and predicted risk of mortality for each subject were then calculated. The predicted risk was spilt into deciles and plotted against the mean observed percentage that died in each group.

The older patients included in the study population were very diverse in their underlying health status. Rather than create pre-defined risk groups (such as low, medium, high etc.) which may not be clinically useful, I decided that GPs should be able to compare the mortality risk of their patient to the general (overall) mortality risk of a patient of the same sex and age group in the general population.

However, purely as a technical device to allow the separation of patients with different scores to be considered, the score was split into four groups (at the 16th, 50th and 84th centiles) as per the Cox method described in section 7.3.1.4. Kaplan-Meier estimates for each model stratified by these risk groups were produced.

As a comparison group, mortality rates for CPRD-ONS linked patients aged ≥65 years, who were active in CPRD between 1st April 2004 and 31st March 2011 were calculated by age group. These rates were converted to a risk for each of the three time periods using Equation 7-3.[113] Risk ratios of predicted risks of mortality to those of the linked CPRD population were calculated.

Equation 7-3 Conversion of a rate to a risk

$$Risk\ (up\ to\ time\ t) = 1 - e^{-\lambda t}$$

Where λt= hazard (rate) at time t.

7.4.8.4 Model validation

Internal validation was assessed using bootstrap re-sampling with 200 repetitions. As described in section 7.3.2.1, Cox regression models were developed in the same way as in the original data, c-statistics calculated and the model then applied to the original data, where the c-statistic was calculated again. This process was repeated 200 times. The means of these two sets of c-statistics were calculated, and the difference between them provided the optimism of the model.[115]

7.4.9 Presentation of risk scores to a clinical audience

When used in a clinical setting, it is important that predicted risks are presented in a way that is easy to interpret as a population level statistic. Ideally this would be done using text and/or a graphic explaining (for example) "Out of 1000 patients with this combination of health factors, on average we would expect 470 to die within the next year. This level of risk is 4 times higher than the general population of the same age group and sex". This provides the clinician with the patient's level of risk, as well as appropriate contextualisation. I therefore developed tables to demonstrate this approach.

7.5   Results

7.5.1   Study population and baseline characteristics

There were 13,589 CAP hospital discharges recorded in 12,983 patients included in the study cohort; the derivation of the original discharge cohort, and the numbers included in each model are presented in Figure 7-2. Characteristics of the population in each time period are shown in Table 7-3. The median age of patients at CAP diagnosis was 80 years, and 51% of episodes were in women. The majority of episodes were among patients who had at least one existing co-morbidity of interest also recorded (the median number of co-morbidities in each model period was two (lower to upper quartiles, 1 to 3)). The most common co-morbidity was chronic lung disease (41%) while in 30% of episodes the patients had a history of ischaemic heart disease and 20% had cerebrovascular disease. In the first week after discharge, 313 (2.3%) patients in the initial cohort died. Over 8-30 days 13,051 episodes were included of which 520 (4.0%) ended in death. In the longer 31-365 day model, 12,204 episodes were included of which 2,415 (19.8%) resulted in the patient's death.

Figure 7-2 Flow chart for inclusion in the study

15282 hospitalisations for CAP where the patient survived until at least the day after discharge

EXCLUDED: Hospitalisation within a year of a previous CAP discharge n=810

N=14472

EXCLUDED: Terminally ill n=451
Metastatic cancer n=78

N=13943

EXCLUDED: Smoking status not recorded n=354

1-7 days:
13589 discharges
(12983 patients)

Died: 313, Censored: 225

8-30 days:
13051 discharges
(12467 patients)

Died: 520, Censored: 327

31-365 days:
12204 discharges
(11667 patients)
Follow-up:
Median 334 days
(IQR: 172-334)

Died: 2415, Censored: 2419

Table 7-3 Number of CAP hospital discharges and deaths per time period, and by potential prognostic factors

Potential prognostic factor	Value	Model: 1-7 days		8-30 days		31-365 days	
		CAPs	Died	CAPs	Died	CAPs	Died
All CAP discharges		13589	313	13051	520	12204	2415
Sex	Male	6660	143	6411	258	6004	1274
	Female	6929	170	6640	262	6200	1141
Age group (years)	65-69	1765	24	1722	33	1658	195
	70-74	2169	24	2128	55	2038	308
	75-79	2595	38	2519	86	2392	401
	80-84	2926	68	2808	116	2612	555
	85-89	2481	77	2355	116	2164	546
	90+	1653	82	1519	114	1340	410
Number of co-morbidities	0	1685	31	1631	42	1554	181
	1	3468	64	3341	106	3143	560
	2	3436	98	3287	135	3066	613
	3	2585	58	2483	133	2295	499
	≥4	2415	62	2309	104	2146	562
Congestive heart failure		2264	60	2159	80	2030	571
Peripheral vascular disease		1423	43	1362	64	1268	301
Dementia		1051	51	963	90	824	270
Chronic lung disease		5658	115	5470	200	5150	1066
Connective tissue disease		1341	23	1298	42	1228	265
Peptic ulcer		1128	28	1084	50	1013	199
Hemiplegia		117	1	115	8	105	30
Solid Cancer		2059	63	1964	100	1816	433
Leukaemia/Lymphoma		334	8	321	18	296	89
Cerebrovascular disease		2835	94	2686	152	2472	597
Other neurological disease		896	28	841	59	769	189
Other immunological disease		52	0	52	1	51	13
Diabetes	Diabetes	1962	41	1893	66	1783	365
	+ complications	778	6	760	19	718	138
Liver Disease	Mild/moderate	61	1	59	2	56	10
	Severe	20	2	18	1	17	3
Ischaemic heart disease	pre-MI	2395	47	2313	108	2154	444
	post-MI	1760	39	1696	64	1591	349
Renal Disease	Mild/moderate	255	4	248	10	231	42
	Severe	2827	63	2694	120	2474	548
Smoking status	Non	2018	72	1915	114	1741	388
	Current	2637	47	2548	104	2380	478
	Ex	8934	194	8588	302	8083	1549
Previous pneumonia		2060	41	1989	87	1864	447
Excessive alcohol consumption		739	11	714	22	677	145
Weight loss		960	51	893	93	776	287

Table 7-3 Number of CAP hospital discharges and deaths per time period, and by potential prognostic factors (continued)

Potential prognostic factor	Value	Model: 1-7 days CAPs	1-7 days Died	8-30 days CAPs	8-30 days Died	31-365 days CAPs	31-365 days Died
Fall in previous year		1412	43	1324	72	1193	289
Carer in previous year		478	12	456	17	420	92
Living arrangements	No records	11368	225	10969	356	10353	1872
	Lives alone	550	12	530	25	487	109
	Sheltered accommodation	160	6	151	6	139	35
	Residential Care	1511	70	1401	133	1225	399
Influenza vaccination	No records	1267	32	1218	46	1141	207
	14-365 days pre CAP	8991	189	8658	320	8125	1587
	Last season	2478	59	2371	121	2198	476
	2-5 years pre CAP	623	27	582	28	532	106
	>5 years pre CAP	230	6	222	5	208	39
Pneumococcal vaccination	No records	3041	90	2898	126	2711	538
	14-365 days pre CAP	751	11	729	29	683	138
	1-2 years pre CAP	988	19	958	42	904	185
	2-5 years pre CAP	3525	75	3407	115	3242	642
	>5 years pre CAP	5284	118	5059	208	4664	912

Kaplan-Meier survival curves for the complete year after discharge and for each model period are shown in Figure 7-3. The mortality rate was highest in the week after discharge, then slowed and was approximately linear by day 150 post discharge (day 121 in the third (31-365 day) model).

Figure 7-3 Kaplan Meier survival plots for the full year after discharge and each model period
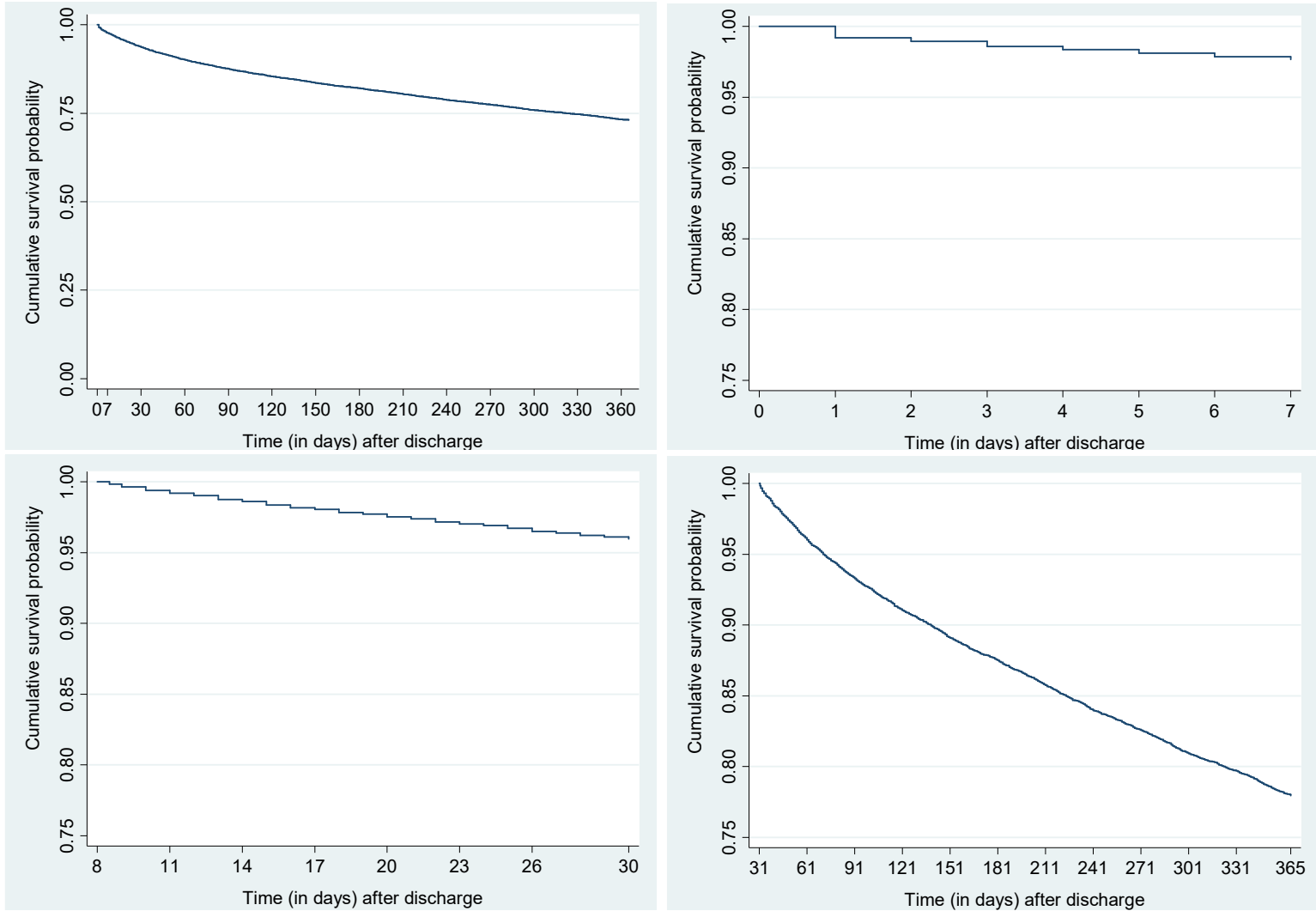
NOTE: differing scales on y-axis

7.5.2 Cause of death

Pneumonia was named as the underlying cause for 20% of deaths in the 1-7 day post-discharge period, 13% in 8-30 days and 9% in 31-365 days (Figure 7-4). The percentage of death certificates with pneumonia named as a contributory cause of death also decreased over the three time periods, and was named on 58%, 41% and 34% respectively (data not shown). The percentage of deaths attributed to dementia decreased with increasing period after discharge, while those for cardiovascular conditions (in particular cerebrovascular disease and ischaemic heart disease (pre-MI)) made up an increasing percentage of deaths as time since discharge increased (Figure 7-4). The contribution of chronic lung disease did not differ substantially over the three time periods, and nor did that of solid cancers.

7.5.3 Crude associations between potential predictors and mortality

Unadjusted and minimally adjusted hazard ratios for each predictor in each time period are presented in Table 7-4. In univariable analyses, age groups ≥80 years, a diagnosis of dementia, solid cancer, cerebrovascular disease, being underweight/requiring nutritional supplementation in the last year, having fallen in the last year and living in residential care were all associated with an increased mortality rate across all three time periods under study. There were a larger number of factors associated with mortality in the 31-365 day period than either the 1-7 or 8-30 day periods. These included relatively common conditions among older adults such as chronic lung disease and congestive heart failure.

Figure 7-4 Distribution of underlying causes of death by time period post-CAP hospital discharge
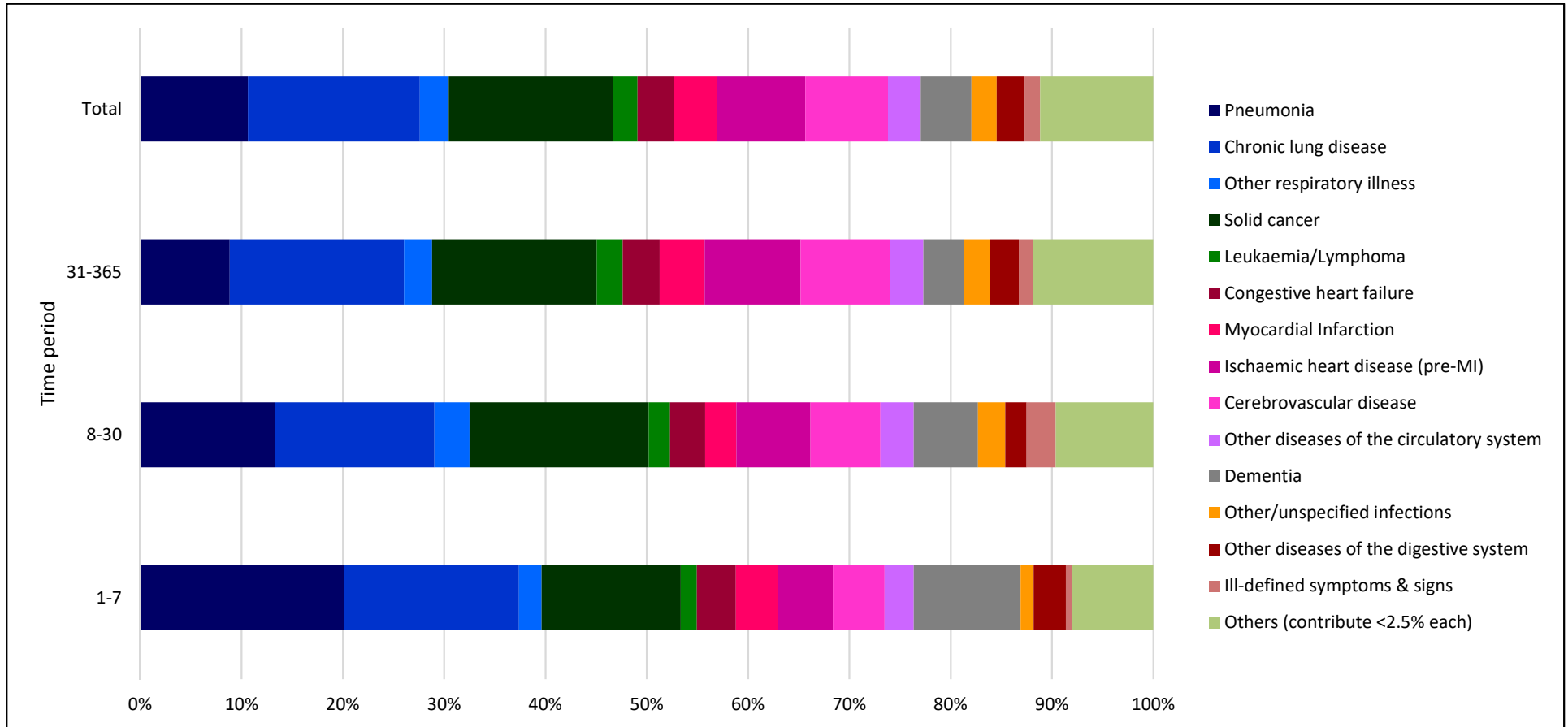
Table 7-4 Univariable & minimally adjusted mortality hazard ratios (HR) over each time period, for potential prognostic factors of interest (cont…)

Potential prognostic factor	Value	1-7 days		8-30 days		31-365 days	
		Unadjusted HR	Minimally adjusted HR	Unadjusted HR	Minimally adjusted HR	Unadjusted HR	Minimally adjusted HR
Sex	Female	1.15 (0.92 - 1.43)	1.00 (0.80 - 1.24)	0.98 (0.83 - 1.17)	0.88 (0.74 - 1.05)	0.86 (0.79 - 0.93)	0.78 (0.72 - 0.84)
Age	70-74	0.81 (0.46 - 1.43)	0.81 (0.46 - 1.43)	1.35 (0.88 - 2.08)	1.36 (0.88 - 2.09)	1.30 (1.09 - 1.55)	1.30 (1.09 - 1.56)
	75-79	1.08 (0.65 - 1.79)	1.08 (0.65 - 1.79)	1.80 (1.20 - 2.68)	1.80 (1.21 - 2.69)	1.46 (1.23 - 1.73)	1.47 (1.24 - 1.74)
	80-84	1.72 (1.08 - 2.74)	1.72 (1.08 - 2.74)	2.20 (1.49 - 3.24)	2.21 (1.50 - 3.26)	1.94 (1.65 - 2.28)	1.98 (1.68 - 2.33)
	85-89	2.30 (1.46 - 3.64)	2.31 (1.46 - 3.64)	2.63 (1.79 - 3.88)	2.67 (1.81 - 3.94)	2.44 (2.08 - 2.88)	2.52 (2.14 - 2.97)
	90+	3.73 (2.37 - 5.87)	3.73 (2.37 - 5.87)	4.11 (2.79 - 6.05)	4.22 (2.86 - 6.25)	3.15 (2.66 - 3.73)	3.33 (2.81 - 3.95)
Congestive heart failure		1.19 (0.90 - 1.58)	1.03 (0.77 - 1.37)	0.92 (0.72 - 1.16)	0.80 (0.63 - 1.02)	1.65 (1.50 - 1.81)	1.45 (1.32 - 1.60)
Peripheral vascular disease		1.37 (0.99 - 1.88)	1.41 (1.02 - 1.96)	1.21 (0.93 - 1.57)	1.19 (0.92 - 1.55)	1.26 (1.12 - 1.42)	1.22 (1.08 - 1.38)
Dementia		2.36 (1.75 - 3.18)	1.88 (1.38 - 2.55)	2.75 (2.19 - 3.44)	2.30 (1.81 - 2.91)	1.98 (1.74 - 2.25)	1.65 (1.45 - 1.88)
Chronic lung disease		0.81 (0.64 - 1.02)	0.96 (0.77 - 1.21)	0.86 (0.72 - 1.03)	0.97 (0.81 - 1.16)	1.09 (1.01 - 1.18)	1.21 (1.12 - 1.32)
Connective tissue disease		0.72 (0.47 - 1.10)	0.72 (0.47 - 1.10)	0.79 (0.58 - 1.08)	0.80 (0.58 - 1.10)	1.12 (0.99 - 1.27)	1.16 (1.02 - 1.31)
Peptic ulcer		1.08 (0.74 - 1.60)	1.12 (0.76 - 1.64)	1.18 (0.88 - 1.58)	1.17 (0.87 - 1.57)	1.00 (0.86 - 1.15)	0.97 (0.84 - 1.13)
Solid Cancer		1.41 (1.07 - 1.86)	1.38 (1.04 - 1.83)	1.35 (1.09 - 1.68)	1.30 (1.04 - 1.62)	1.31 (1.18 - 1.45)	1.23 (1.10 - 1.36)
Leukaemia/Lymphoma		1.04 (0.52 - 2.10)	1.16 (0.57 - 2.34)	1.43 (0.89 - 2.28)	1.54 (0.96 - 2.47)	1.62 (1.32 – 2.00)	1.72 (1.40 - 2.12)
Cerebrovascular disease		1.64 (1.29 - 2.08)	1.48 (1.15 - 1.89)	1.61 (1.33 - 1.94)	1.45 (1.20 - 1.75)	1.35 (1.23 - 1.48)	1.21 (1.11 - 1.33)
Other neurological disease		1.40 (0.95 - 2.07)	1.45 (0.99 - 2.14)	1.88 (1.43 - 2.46)	1.89 (1.44 - 2.48)	1.32 (1.14 - 1.53)	1.31 (1.13 - 1.52)
Other immunological disease		0 (0 - 0)	0 (0 - 0)	0.47 (0.07 - 3.43)	0.59 (0.08 - 4.25)	1.41 (0.84 - 2.36)	1.71 (1.03 - 2.83)
Diabetes	Diabetes	0.85 (0.61 - 1.18)	0.92 (0.66 - 1.27)	0.83 (0.64 - 1.08)	0.87 (0.67 - 1.13)	1.05 (0.94 - 1.17)	1.08 (0.97 - 1.21)
	+complications	0.31 (0.14 - 0.70)	0.35 (0.16 - 0.80)	0.59 (0.38 - 0.94)	0.64 (0.40 - 1.01)	0.99 (0.84 - 1.18)	1.04 (0.88 - 1.23)
Liver Disease	Mild/moderate	0.71 (0.10 - 4.99)	0.90 (0.13 - 6.39)	0.84 (0.21 - 3.33)	1.00 (0.25 - 3.99)	0.94 (0.51 - 1.73)	1.12 (0.60 - 2.06)
	Severe	4.44 (1.12 - 17.51)	5.34 (1.28 - 22.23)	1.39 (0.19 - 10.02)	1.61 (0.22 - 11.93)	0.93 (0.30 - 2.82)	0.98 (0.31 - 3.12)

Potential prognostic factor	Value	1-7 days		8-30 days		31-365 days	
		Unadjusted HR	Minimally adjusted HR	Unadjusted HR	Minimally adjusted HR	Unadjusted HR	Minimally adjusted HR
Ischaemic heart disease	Pre-MI	0.81 (0.59 - 1.11)	0.79 (0.58 - 1.08)	1.22 (0.98 - 1.51)	1.17 (0.94 - 1.45)	1.07 (0.97 - 1.19)	1.02 (0.92 - 1.14)
	Post-MI	0.92 (0.65 - 1.29)	0.91 (0.65 - 1.29)	0.98 (0.75 - 1.27)	0.93 (0.71 - 1.22)	1.17 (1.04 - 1.31)	1.09 (0.97 - 1.22)
Renal Disease	Mild/moderate	0.67 (0.25 - 1.79)	0.72 (0.27 - 1.92)	1.04 (0.56 - 1.95)	1.10 (0.59 - 2.05)	0.98 (0.73 - 1.34)	1.02 (0.76 - 1.39)
	Severe	0.95 (0.72 - 1.26)	0.83 (0.62 - 1.09)	1.17 (0.95 - 1.44)	1.04 (0.85 - 1.28)	1.24 (1.12 - 1.36)	1.10 (1.00 - 1.22)
Hemiplegia		0.36 (0.05 - 2.60)	0.39 (0.05 - 2.74)	1.79 (0.89 - 3.61)	1.87 (0.93 - 3.78)	1.46 (1.02 - 2.09)	1.44 (0.99 - 2.09)
Weight loss		2.59 (1.93 - 3.49)	2.40 (1.78 - 3.23)	3.07 (2.45 - 3.83)	2.91 (2.33 - 3.65)	2.29 (2.03 - 2.59)	2.22 (1.97 - 2.50)
Fall in last year		1.39 (1.01 - 1.91)	1.09 (0.78 - 1.50)	1.46 (1.14 - 1.87)	1.20 (0.93 - 1.55)	1.35 (1.19 - 1.53)	1.14 (1.01 - 1.30)
Carer in last year		1.10 (0.62 - 1.95)	0.78 (0.44 - 1.40)	0.94 (0.58 - 1.53)	0.72 (0.44 - 1.17)	1.14 (0.93 - 1.40)	0.89 (0.72 - 1.09)
Previous pneumonia		0.84 (0.61 - 1.17)	0.83 (0.60 - 1.15)	1.11 (0.88 - 1.40)	1.07 (0.85 - 1.35)	1.32 (1.19 - 1.46)	1.26 (1.13 - 1.39)
Excessive alcohol consumption		0.63 (0.35 - 1.15)	0.85 (0.46 - 1.56)	0.76 (0.49 - 1.16)	0.94 (0.61 - 1.46)	1.11 (0.94 - 1.31)	1.30 (1.10 - 1.55)
Smoking status	Current	0.50 (0.34 - 0.72)	0.72 (0.50 - 1.05)	0.67 (0.52 - 0.88)	0.90 (0.68 - 1.19)	0.88 (0.77 – 1.00)	1.10 (0.96 - 1.27)
	Ex	0.61 (0.46 - 0.79)	0.71 (0.54 - 0.94)	0.58 (0.47 - 0.72)	0.64 (0.51 - 0.80)	0.83 (0.74 - 0.93)	0.88 (0.79 - 0.99)
Living arrangements	Lives alone	1.11 (0.62 - 1.98)	0.76 (0.42 - 1.37)	1.48 (0.99 - 2.22)	1.16 (0.76 - 1.76)	1.28 (1.06 - 1.56)	1.01 (0.82 - 1.23)
	Sheltered accommodation	1.89 (0.85 - 4.22)	1.28 (0.57 - 2.90)	1.23 (0.55 - 2.73)	0.94 (0.42 - 2.10)	1.45 (1.04 - 2.04)	1.14 (0.81 - 1.59)
	Residential Care	2.37 (1.82 - 3.1)	1.78 (1.34 - 2.36)	3.03 (2.49 - 3.70)	2.52 (2.04 - 3.11)	2.04 (1.83 - 2.28)	1.72 (1.53 - 1.92)
Influenza vaccination	14-365	0.40 (0.17 - 0.97)	0.53 (0.22 - 1.28)	0.82 (0.50 - 1.33)	0.98 (0.60 - 1.61)	1.14 (0.94 - 1.38)	1.29 (1.06 - 1.57)
	Previous season	1.26 (0.52 - 3.05)	1.85 (0.76 - 4.52)	0.46 (0.15 - 1.44)	0.64 (0.20 - 2.01)	1.13 (0.81 - 1.57)	1.49 (1.06 - 2.09)
	2-5 years	0.91 (0.13 - 6.38)	1.24 (0.19 - 8.15)	1.12 (0.28 - 4.45)	1.48 (0.37 - 5.93)	0.72 (0.32 - 1.61)	0.89 (0.40 - 2.01)
(timing pre-CAP)	>5 years	0.83 (0.57 - 1.21)	0.80 (0.55 - 1.17)	0.98 (0.72 - 1.33)	0.92 (0.67 - 1.25)	1.08 (0.94 - 1.25)	1.00 (0.86 - 1.15)
Pneumococcal vaccination	14-365	0.94 (0.61 - 1.45)	0.88 (0.57 - 1.35)	1.36 (0.97 - 1.90)	1.24 (0.88 - 1.74)	1.27 (1.08 - 1.49)	1.15 (0.98 - 1.35)
	1-2 years	1.74 (1.04 - 2.89)	1.61 (0.96 - 2.68)	1.29 (0.80 - 2.06)	1.18 (0.74 - 1.89)	1.21 (0.95 - 1.52)	1.11 (0.87 - 1.40)
	2-5 years	1.03 (0.43 - 2.46)	0.96 (0.40 - 2.30)	0.60 (0.24 - 1.50)	0.55 (0.22 - 1.39)	1.13 (0.81 - 1.57)	1.05 (0.75 - 1.46)
(timing pre-CAP)	>5 years	0.49 (0.26 - 0.92)	0.53 (0.28 - 0.99)	0.91 (0.61 - 1.37)	0.95 (0.63 - 1.41)	0.98 (0.82 - 1.19)	1.01 (0.84 - 1.22)

7.5.4 Model selection and performance

When all the levels of the variables identified as candidate predictors were taken into account, the number of potential factors was 44. The 31-365 day model had 54 outcomes per potential factor, and the 8-30 day model had 11 outcomes per factor, but the 1-7 day model had only 7 outcomes per factor (and thus was under the '10 events per potential prognostic factor' rule). Rather than fit a different list of factors to each model, I decided to use the same list and have slightly lower power for the 1-7 day mortality model, as there were no candidate predictors which were an obvious choice for omission.

Backwards stepwise regression selected 12 factors (with 17 levels) into the 1-7 day model, 17 factors (with 23 levels) into the 8-30 day model and 17 predictive factors (with 21 levels) into the 31-365 day model, in addition to the a priori factors age (5 levels) and sex (1 level). Seven factors were selected into all three models (the 'consistent' factors: peripheral vascular disease, dementia, connective tissue disease, cerebrovascular disease, solid cancer, living arrangements and low weight), while five variables were not included in any of the final models (hemiplegia, peptic ulcer, renal disease, other immune mechanism disorders, and falls in the last year) (Table 7-5).

Each of the sets of data was consistent with the proportional hazards assumption.

7.5.5 1-7 day model

The model for the shortest period included diabetes, influenza vaccination and liver disease, in addition to age, sex and the seven consistent factors. These 12 predictors had the widest range of scores of any of the models, ranging from -1.11 for diabetes with complications (negative coefficients represent a decreased mortality risk) to 1.81 for severe liver disease, although most scores were within the range -0.3 to 0.4 (Table 7-5). The uncorrected c-statistic was 0.699, suggesting the model was reasonably good at discriminating between patients who lived and died over the seven day period.

The calibration was the weakest of the three models, with mortality over-predicted by the model in the lowest five deciles and observed mortality higher in decile six than in decile seven (Figure 7-5). The Kaplan-Meier survival curves stratified by Cox-centile

showed little separation between the two lowest Cox-centile groups, but better separation between those at the two higher levels of risk (Figure 7-6).

The optimism of the model was 0.031, and the corrected c-statistic 0.669. The stability (in terms of repeated selection of factors into the model) was lower than the two longer-duration models, with only low weight and peripheral vascular disease being selected into >70% of bootstraps (Table 7-6).

7.5.6 8-30 day model

The 8-31 day model contained 17 factors; diabetes and influenza vaccine were again selected as predictors, whereas liver disease was no longer a predictive factor. Six additional factors were included when compared to the 1-7 day model; two were assigned negative coefficients (congestive heart failure and having a recent carer), three were found to be positive predictors (ischaemic heart disease, leukaemia/lymphoma and other neurological disease), and smoking status had a mixed effect depending on whether patients were current or ex-smokers. Of the factors included in the model, low weight was assigned the largest score followed by residential care and the higher age groups (Table 7-5). Among the factors assigned negative coefficients (and thus seen as 'protective' against mortality) influenza vaccination >5 years ago and diabetes with complications had the largest negative effects, reducing patients' scores by 0.49 and 0.46 respectively. Again, the model showed near-reasonable discrimination (unadjusted c-statistic=0.696)

The 8-30 day model slightly over-predicted mortality risk in the first three risk deciles, and slightly under-predicted risk in deciles seven to nine (Figure 7-5). The Kaplan-Meier survival curves stratified by Cox-centile showed reasonable separation between the groups, with good separation for the highest risk group (Figure 7-6).

The optimism was the highest of all the models (0.038), thus the optimism corrected c-statistic was sub-optimal at 0.658 (Table 7-5). Living arrangements, low weight/poor nutrition, and dementia (kept in 200, 200, and 191 bootstraps respectively) showed high stability across the bootstrapping process (Table 7-6).

7.5.7 31-365 day model

Of the 17 factors retained in the 31-365 day model, only leukaemia/lymphoma, other neurological disease, and smoking status remained from those selected into the 8-30 day model. Additionally, chronic lung disease, a history of pneumonia and excessive alcohol consumption were all predictors of mortality, and pneumococcal vaccination had a mixed effect depending on the timing of the vaccination. Increasing age, low weight, leukaemia/lymphoma, residential care, dementia and congestive heart failure, were the strongest predictors of mortality (Table 7-5). Being female, an ex-smoker and having received pneumococcal vaccine over a year ago each had negative score coefficients. The discrimination of the model was lower than for the two shorter time periods (unadjusted c-statistic=0.647).

The model showed fair calibration (Figure 7-5), with observed risk increasing as the predicted risk increased. However, mortality risk was slightly over-predicted in the two lowest deciles compared to the observed risk (Figure 7-5). The Kaplan-Meier curves stratified by Cox-centiles showed good separation between the groups (Figure 7-6).

The model displayed relatively poor discrimination and little optimism (0.018) with an optimism-corrected c-statistic of 0.630 (Table 7-5). The six strongest predictors also showed high stability, and were selected into ≥99% of bootstrap models (Table 7-6).

Table 7-5 Scores for variables included in each final model, and their C-statistics.

		Score for factor (logHR) Model period (days)			Hazard Ratios Model period (days)		
		1-7	8-30	31-365	1-7	8-30	31-365
Sex	Female	0.02	-0.15	-0.24	1.02	0.86	0.79
Age	70-74	-0.23	0.28	0.24	0.79	1.32	1.28
	75-79	0.01	0.53	0.36	1.01	1.70	1.44
	80-84	0.43	0.67	0.63	1.54	1.95	1.87
	85-89	0.67	0.79	0.84	1.95	2.21	2.32
	90+	1.10	1.20	1.13	3.01	3.33	3.08
Diabetes	Diabetes	-0.14	-0.15		0.87	1.13	
	& complications	-1.11	-0.46		0.33	0.86	
Peripheral vascular disease		0.40	0.23	0.16	1.49	0.63	1.18
Dementia		0.42	0.45	0.37	1.52	1.56	1.44
Living arrangements	Lives alone	-0.28	0.33	0.01	0.76	1.39	1.01
	Sheltered housing	0.24	0.04	0.07	1.27	1.04	1.07
	Residential Care	0.45	0.70	0.42	1.56	2.02	1.53
Connective tissue disease		-0.31	-0.21	0.14	0.73	0.81	1.15
Influenza vaccination*	14-365 days	-0.25	-0.11		0.78	0.90	
	Previous season	-0.21	0.14		0.81	1.16	
	2-5 years	0.44	0.12		1.55	1.13	
	>5 years	0.00	-0.49		1.00	0.61	
Solid cancer		0.35	0.28	0.19	1.42	1.33	1.21
Liver Disease	Mild/moderate	-0.10			0.91		
	Severe	1.81			6.11		
Low weight/poor nutrition		0.64	0.94	0.66	1.90	2.56	1.93
Congestive heart failure			-0.19	0.36		0.83	1.44
Cerebrovascular disease		0.29	0.18	0.09	1.34	1.20	1.09
Recent carer			-0.36			0.69	
Ischaemic heart disease	pre-MI		0.23			1.26	
	post-MI		0.02			1.02	
Leukaemia/lymphoma			0.41	0.56		1.50	1.75
Other neurological disease			0.42	0.19		1.52	1.21
Smoking status	Current		0.06	0.11		1.06	1.12
	Ex		-0.27	-0.10		0.76	0.91
Chronic lung disease				0.20			1.23
Previous pneumonia				0.14			1.16
Excess alcohol consumption (Yes)				0.19			1.21
Pneumococcal vaccination*	14-365 days			0.01			1.01
	1-2 years			-0.08			0.92
	2-5 years			-0.12			0.89
	>5 years			-0.14			0.87
Uncorrected C-stat		0.699	0.696	0.647			
Optimism		0.031	0.038	0.018			
Corrected C-stat		0.669	0.658	0.630			

Empty cells denote the factor not being selected into the model. Scores are rounded to 2 decimal places *Timing of most recent vaccine pre-CAP

Table 7-6 Stability of the models: Number of times each variable was selected into each model made in the bootstrap samples of data (200 repetitions).

Variable	1-7 day model	8-30 day model	31-365 day model
Sex*	200 (100%)	200 (100%)	200 (100%)
Age*	200 (100%)	200 (100%)	200 (100%)
Chronic heart failure	29 (14.5%)	88 (44%)	200 (100%)
Peripheral vascular disease	149 (74.5%)	80 (40%)	148 (74%)
Dementia	137 (68.5%)	191 (95.5%)	200 (100%)
Chronic lung disease	27 (13.5%)	25 (12.5%)	198 (99%)
Connective tissue disease	66 (33%)	71 (35.5%)	115 (57.5%)
Peptic ulcer	23 (11.5%)	48 (24%)	42 (21%)
Hemiplegia	8 (4%)	62 (31%)	69 (34.5%)
Leukaemia/lymphoma	16 (8%)	87 (43.5%)	198 (99%)
Other neurological disease	49 (24.5%)	167 (83.5%)	144 (72%)
Other immunological disease	7 (3.5%)	3 (1.5%)	56 (28%)
Cerebrovascular disease	123 (61.5%)	101 (50.5%)	96 (48%)
Previous pneumonia	55 (27.5%)	27 (13.5%)	156 (78%)
Solid cancer	138 (69%)	152 (76%)	189 (94.5%)
Smoking status	69 (34.5%)	167 (83.5%)	191 (95.5%)
Diabetes	119 (59.5%)	87 (43.5%)	22 (11%)
Liver Disease	134 (67%)	30 (15%)	10 (5%)
Ischaemic heart disease	47 (23.5%)	79 (39.5%)	9 (4.5%)
Renal disease	38 (19%)	30 (15%)	36 (18%)
Living arrangements	123 (61.5%)	200 (100%)	200 (100%)
Influenza vaccination (most recent)	124 (62%)	81 (40.5%)	68 (34%)
Pneumococcal vaccination (most recent)	52 (26%)	43 (21.5%)	78 (39%)
Low weight/poor nutrition	159 (79.5%)	200 (100%)	200 (100%)
History of falling	18 (9%)	20 (10%)	25 (12.5%)
Recent carer	30 (15%)	61 (30.5%)	48 (24%)
Excessive alcohol consumption	36 (18%)	28 (14%)	127 (63.5%)

Cells in green show variables included in the final model.

*variables were forced into the model.

Figure 7-5 Calibration of the three models, split into deciles of predicted mortality
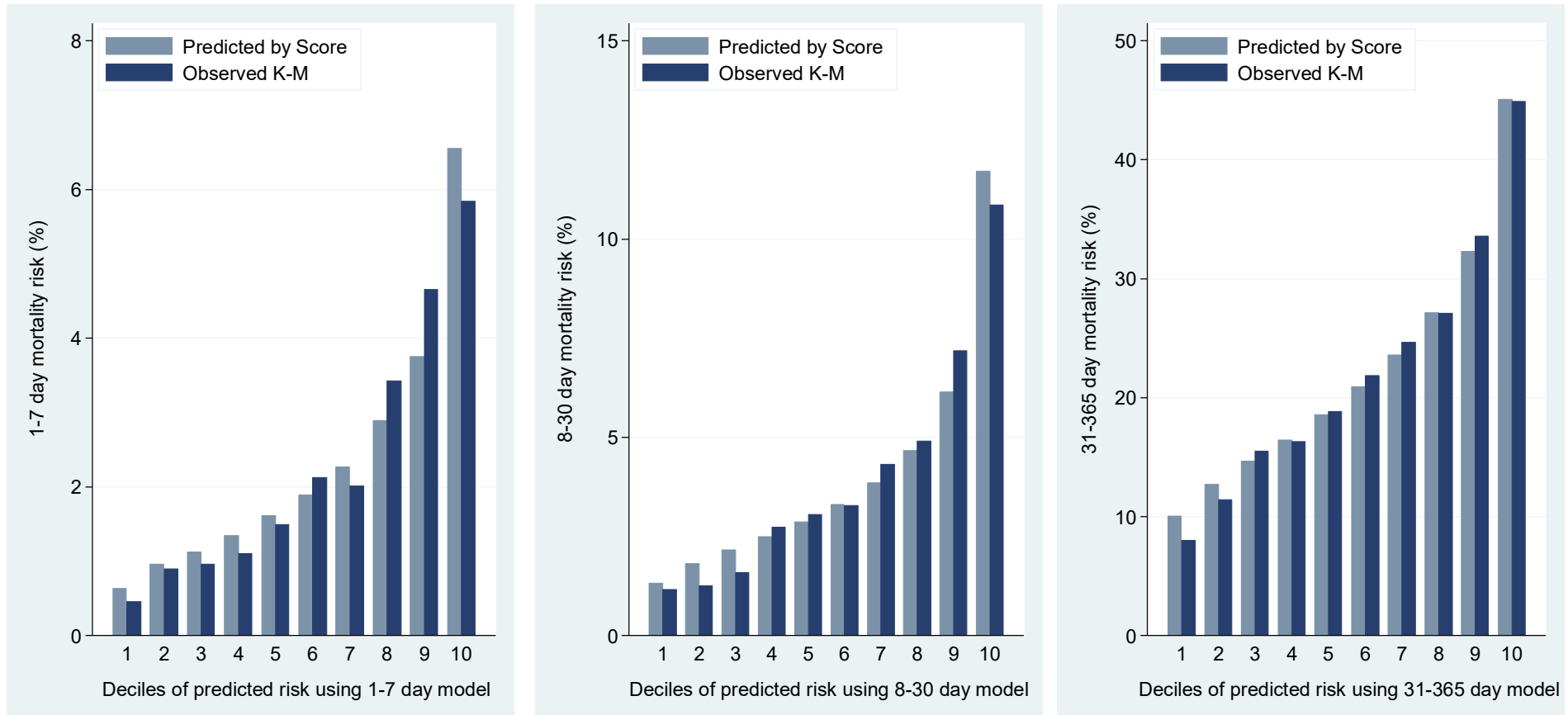
NOTE: different y-axis scales used on each panel of the figure.

Figure 7-6 Kaplan-Meier curves stratified by Cox-centiles for each model
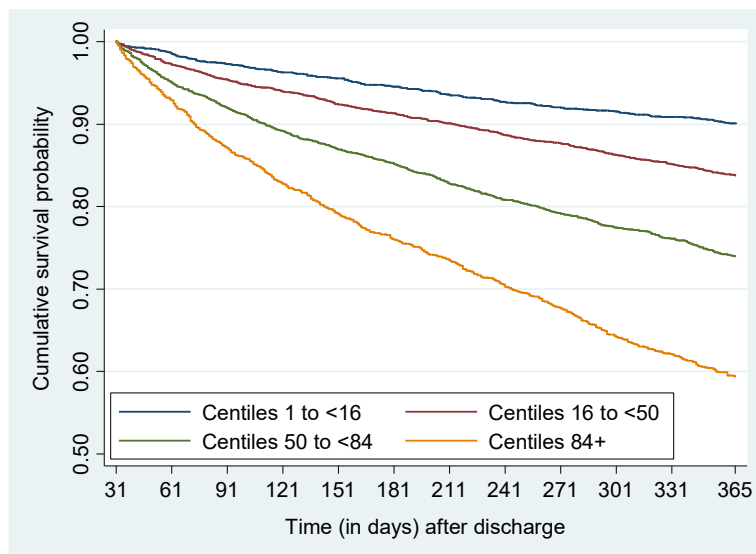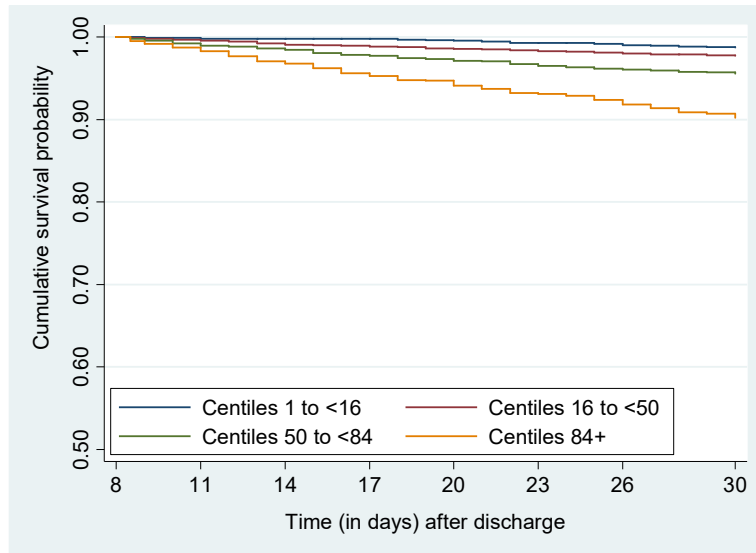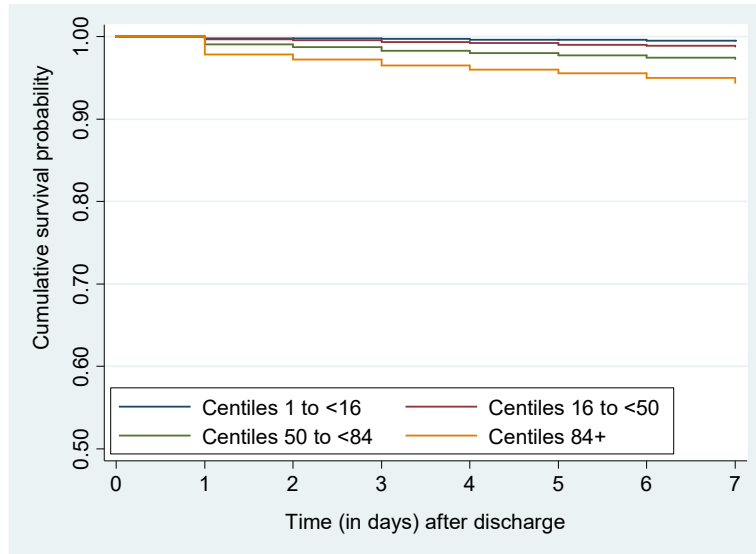
7.5.8 Predicted risks and risk ratios for (hypothetical) patients

Predicted mortality risks for a range of hypothetical scenarios and the risk ratios comparing the predicted risk for each scenario to the risk in the CPRD-ONS linked population are shown separately for women and men in Table 7-7 and Table 7-8.

Predicted risks (on the left side of Table 7-7 and Table 7-8) are presented as the number of patients per 1000 that would be expected to die in each risk period, given a specified combination of factors. For example, out of 1000 women aged 80-84 with a previous cancer diagnosis and of low-weight, on average 343 would be expected to die within 31-365 days post-CAP discharge (Table 7-7 – left-hand side); these women have 7.2 times the mortality risk of an 80-84 year old woman in the general CPRD-linked cohort (as shown in right-hand side of the table). Similarly, of 1000 65-69 year-old men who were ex-smokers with COPD and who received an influenza vaccine this season, 126 would be expected to die within 31-365 days of a CAP discharge, which is 7.7 times the risk compared to men of the same age in the linked-CPRD population (left and right sides of Table 7-8). The ratios were calculated using mortality rates (converted to risks) for the CPRD-linked HES cohort, which can be found in Appendix I.

Table 7-7 Examples of predicted risk of mortality among women, and of risk ratios comparing predicted mortality of female patients post-CAP hospital discharge to the mortality risk of female patients in the general CPRD population for a range of factors in each model

Factors included when calculating mortality risk among WOMEN	Risk presented as number of patients per 1000 predicted to die in each time period given the set of factors stated						Relative risk of dying post-CAP discharge compared to patients in the general CPRD population) given the set of factors stated						KEY (women)
Age	65-69	70-74	75-79	80-84	85-89	90+	65-69	70-74	75-79	80-84	85-89	90+	**Risk/1000**
Age alone	14	11	14	22	28	42	61.0	35.1	25.4	21.4	15.0	11.2	1-7 day
	17	23	29	33	38	56	23.4	22.4	16.3	10.4	6.5	4.8	8-30
	91	115	128	164	199	255	8.2	7.5	4.8	3.5	2.4	1.5	31-365
Diabetes & Peripheral vascular disease	19	15	19	28	36	55	79.1	45.6	33.0	27.8	19.4	14.4	**RR**
	19	24	31	36	41	61	25.2	24.1	17.6	11.2	7.0	5.1	1-7 day
	106	134	149	190	230	293	9.6	8.7	5.6	4.0	2.7	1.8	8-30
Flu vaccine this season, COPD & Ex-smoker	11	9	11	17	22	33	47.5	27.3	19.8	16.7	11.7	8.7	31-365
	12	16	20	23	26	39	16.1	15.4	11.3	7.2	4.5	3.3	
	101	126	141	180	218	279	9.0	8.3	5.3	3.8	2.6	1.7	
Cancer & low weight	38	30	39	58	73	110	162.4	93.7	67.7	56.8	39.4	29.1	
	57	75	96	109	123	179	78.1	74.0	53.7	34.1	21.1	15.1	
	201	249	275	343	406	499	18.1	16.2	10.3	7.2	4.8	3.0	
Dementia, Residential care	34	27	34	51	65	98	143.6	82.9	59.9	50.3	34.9	25.8	
	53	70	89	102	114	167	72.6	68.8	50.0	31.8	19.7	14.1	
	190	236	261	326	387	477	17.1	15.4	9.7	6.9	4.6	2.9	
Pre-MI Ischaemic heart disease, Flu vaccine this season & PPV >5 years	11	9	11	17	22	33	47.5	27.3	19.8	16.7	11.7	8.7	
	20	26	33	38	43	64	26.5	25.3	18.5	11.8	7.3	5.4	
	79	100	112	144	175	225	7.1	6.5	4.2	3.0	2.1	1.4	

Table 7-8 Examples of predicted risk of mortality among men, and of risk ratios comparing predicted mortality of male patients post-CAP hospital discharge to the mortality risk of male patients in the general CPRD population for a range of factors in each model

Factors included when calculating mortality risk among MEN	Risk presented as number of patients per 1000 predicted to die in each time period given the set of factors stated						Relative risk of dying post-CAP discharge compared to patients in the general CPRD population) given the set of factors stated							KEY (men)
Age	65-69	70-74	75-79	80-84	85-89	90+	65-69	70-74	75-79	80-84	85-89	90+		Risk/1000
Age alone	14	11	14	22	27	42	40.7	22.1	17.0	14.7	11.2	9.8		1-7 day
	20	26	34	39	44	65	18.4	16.6	12.9	8.4	5.7	4.9		8-30
	114	143	160	203	245	311	7.0	6.0	4.1	3.0	2.2	1.7		31-365
Diabetes & Peripheral vascular disease	18	15	18	28	35	54	52.7	28.7	22.0	19.0	14.5	12.6		RR
	22	28	36	42	47	70	19.8	17.9	13.8	9.0	6.1	5.2		1-7 day
	133	166	185	234	282	355	8.1	7.0	4.7	3.5	2.6	1.9		8-30
Flu vaccine this season, COPD & Ex-smoker	11	9	11	17	21	33	31.7	17.2	13.2	11.4	8.7	7.6		31-365
	14	18	23	27	30	45	12.7	11.4	8.9	5.8	3.9	3.4		
	126	157	175	223	268	339	7.7	6.6	4.5	3.3	2.4	1.8		
Cancer & low weight	37	30	38	57	72	108	108.3	59.2	45.3	38.9	29.5	25.4		
	66	87	110	126	141	204	61.1	54.7	42.0	27.3	18.4	15.3		
	247	304	335	413	483	583	15.1	12.8	8.6	6.1	4.4	3.2		
Dementia, Residential care	33	26	34	50	64	96	95.8	52.3	40.0	34.4	26.1	22.6		
	62	81	103	117	131	191	56.8	50.9	39.1	25.4	17.2	14.3		
	234	288	318	394	462	560	14.3	12.1	8.1	5.8	4.2	3.0		
Pre-MI Ischaemic heart disease, Flu vaccine this season & PPV >5 years	11	9	11	17	21	33	31.7	17.2	13.2	11.4	8.7	7.6		
	23	30	38	44	49	73	20.8	18.8	14.5	9.5	6.5	5.5		
	99	125	140	178	216	276	6.1	5.3	3.6	2.6	2.0	1.5		

7.6   Discussion

In this study I have developed a series of risk scores to try to aid clinical decision-making around care in the year after a patient is discharged from hospital after a CAP episode. All three models have been designed to require minimal clinical input, and include important and common conditions such as cancer, cerebrovascular disease, ischaemic heart disease and dementia. These scores would be usable by GPs whether or not the patient had consulted them for the CAP episode prior to their hospital admission, which is valuable given that GP consultations by patients prior to a CAP-related hospitalisation are decreasing (as described in Chapter 6). Many existing prognostic models include signs and symptoms present at the point of diagnosis, which GPs are unlikely to have uniformly recorded, making these models unusable in a primary care setting.

Seven predictors other than age and sex were included in all three models, predicting risk of death for varying lengths of time after patients with CAP are discharged from hospital. These predictors were peripheral vascular disease, dementia, solid cancer, cerebrovascular disease, connective tissue disease, living arrangements and low weight/nutritional supplementation. Of these, living alone (1-7 days) and connective tissue disease (1-7 and 8-30 days) had an initial protective effect which later became predictive of mortality in the longer-term models. Six additional factors were present in two of the models, and seven were predictors in only one time period.

In the main (31-365 day) model, 17 predictors were identified. The model showed relatively poor ability to distinguish between patients who died and survived (discrimination) but good agreement between predicted and observed outcomes (calibration) in the mid-range of predicted risk, where it would be of most use to clinicians.

It is important to consider the factors found to be predictive of mortality in this analysis in context of the study population. Only patients who survived the initial hospitalisation for CAP were included in the study population, whereas those with the highest mortality risk after CAP were likely to have died in hospital. Thus several factors which one may expect to play a role in subsequent mortality may not be prominent risk factors for surviving patients. A good example of such a situation is ischaemic heart disease, where

the 'pre-MI' (less severe) category had a higher score than 'post-MI' in the 8-30 day model (0.23 and 0.02 respectively). Two possible reasons for this are; firstly, that patients with severe ischaemic heart disease are more likely to die in hospital, and thus not be included in the study population. Secondly, those who survive the CAP episode may have been hospitalised for longer than patients without severe ischaemic heart disease to ensure they were sufficiently stable to be discharged. Death in hospital, or a more cautious approach to discharge of specific CAP patients may also explain why chronic lung disease was a predictor of mortality only in the 31-365 day model, and why those with conditions associated with immune dysregulation and thus more severe infection (such as diabetes and connective tissue disorders), those with congestive heart failure and those living alone had a negative association with mortality in the 1-7 and/or 8-30 day models. Conversely, severe liver disease (with a coefficient of 1.81 in the 1-7 day model) had the highest score of any factor in any model, including age. This could be due to patients with severe liver disease being at higher risk of short-term death from conditions such as variceal bleeds or sudden unexplained death.[185]

As previously stated, the exact mechanisms behind the increased mortality risk after a CAP hospital discharge are unclear, but the increased mortality is likely to be due to a combination of CAP being a marker for underlying ill-health and CAP having longer-term effects on the body (as outlined in sections 1.1.2 and 1.3.2.4). For example, the increased association of dementia and solid cancer with death in all three models, and leukaemia/lymphoma and other neurological disease in the two longer periods may be more representative of CAP as a marker for underlying ill-health. In all cases, the HRs remained high into the third period, with a minimum 21% increase. Infections can also worsen existing comorbid conditions, most notably chronic lung disease.[186]

As highlighted in section 1.3.2.4, changes to endothelial activity and platelet activation increase the risk of acute ischaemic events after systemic infection, which explains the inclusion of cerebrovascular disease in all three models, and pre-MI ischaemic heart disease in the 8-30 day model, as well as other co-morbidities such as diabetes that are associated with an increased cardiovascular risk after infection.[187] The same mechanism could also explain the inclusion of peripheral vascular disease, although this was associated with increased mortality in the 1-7 and 31-365 models and decreased mortality in 8-30 days. Similarly, myocardial injury and volume overload (due to the

mechanisms described in section 1.3.2.4), which lead to worsening heart failure, could explain its inclusion and increased risk of mortality in the longer model. The decreased risk of congestive heart failure shown in the 8-30 day model is less obvious, but could perhaps be explained by heart failure being more acutely monitored in the month post-CAP discharge.

7.6.1 Findings in relation to other studies

Direct comparison between the predictors identified in this study and the risk factors identified in the literature review is problematical, due to the many differences between those papers and this work. Neither of the two papers which examined risk factors in up to one year post-CAP discharge separated their follow-up into more than one risk period. This resulted in their capturing in a single model early as well as late post-discharge deaths, which I have demonstrated have some differing prognostic factors.[174, 173] However, as in my analyses, both papers found that mortality risk increased with increasing age, and that residence in a nursing home was associated with higher mortality.[174, 173] To investigate the role of co-morbidities, the Canadian study used a simple count of the number of co-morbidities present,[173] while the Spanish paper adjusted for both co-morbidities and high PSI score.[174] Several of the individual factors the Spanish study identified as risk factors were also included in my 31-365 day prognostic model (chronic lung disease, congestive heart failure, cancer, cerebrovascular disease and dementia), albeit with generally less pronounced measures of effect, possibly due to exclusion of earlier deaths from my model.[174]

As well as providing the most accurate predictions possible, a prognostic model also needs to be fit for use in the appropriate setting. Due to the intended use of this score by GPs, factors relating to the severity of the CAP and physical/laboratory findings such as those found in the PSI or hypotension/shock were not included in the model, despite being identified as risk factors for mortality in periods of a year or more after discharge in other studies.[92, 173-176] Similarly, despite being identified by the Spanish paper as being associated with an increased mortality risk, re-hospitalisation was not included as a candidate predictor in this study.[174] While it is possible to measure re-hospitalisation in a retrospective cohort study, 'future hospitalisation' would not be

known at the point this score would be calculated (post-CAP hospital discharge), and so cannot be used to inform predictions.

Two papers reported causes of death at one year of follow-up, with contrasting results;[174, 92] in general the results of this study sit between those presented by the two papers. I found that over the full year of follow-up post-discharge, the majority of underlying causes of deaths were due to circulatory illnesses (28% ICD chapter I), followed by chronic lung disease and cancer with just under 20% each. Pneumonia accounted for 10.7% of deaths. The Spanish study found that 25.8% of 93 deaths were due to pneumonia, 22.6% to other infectious diseases, 20.4% to cardiovascular causes and 8.6% to cancer.[174] In contrast, the Dutch paper depicted in graphical form that only approximately 3% of 198 deaths were due to pneumonia, 10% due to circulatory diseases, 27% to cancer and 7% to chronic lung disease.[92]

The pattern of underlying causes found across the three model periods I present supports a theory by Mortensen et al that deaths within 45-days of a CAP diagnosis are CAP-related, and those after 45 days less likely to be so.[188] I found that pneumonia was named either as the underlying or contributory cause of death in 57.8% of patients who died within 7 days of discharge, and was the underlying cause for 20.1%. This is consistent with pneumonia being a part of the terminal care pathway, with some patients possibly being discharged from hospital to enable them to die at home. These percentages of CAP-related deaths decreased with each model period, and were named in 34.4% of underlying or contributory causes and 8.9% of underlying causes of deaths in the 31-365 day model.

7.6.2 Strengths and limitations

This study included a sizable number of older patients who had been hospitalised with CAP and subsequently discharged. A large number of potential prognostic factors were investigated, and these were used to develop models for three separate risk periods (for which patients had differing underlying causes of death and for which different prognostic factors for mortality were identified). The models were designed to be used specifically in a primary care setting, and were developed to enable them to be incorporated into pre-existing software to require minimal input from staff. Pre-existing

prognostic scores (such as CURB-65, PSI or Charlson) were not included as candidate predictors, and thus age and co-morbidities were not double-counted.

The work I present here may also inform preventative approaches to CAP. There has been considerable research around risk factors for developing CAP, and on risk factors for dying which can be identified at the point of CAP diagnosis. This study adds to that body of work by identifying those at increased mortality risk in the year following CAP hospital discharge, thus further informing who might particularly benefit from strategies to prevent CAP (such as increased uptake of pneumococcal and influenza vaccines).

I validated the models using bootstrapping, which is the recommended method of internal validation and has been shown to be robust.

7.6.2.1   Use of stand-alone CPRD data in analyses

A key decision was that, in order to create models that could be built into the GP patient software and programmed to run automatically, I only included co-morbidity diagnoses from within patients' primary care records. Recent work has shown that the addition of HES data may capture more co-morbidity diagnoses than stand-alone CPRD data,[148] but the inclusion of these HES data would have violated one of my primary aims – ease of use. Thus, the models need to be interpreted taking into account that the included factors are only those known to and recorded by the practice in the patient's electronic health record. It seems likely that important pre-existing conditions a patient regularly receives treatment for are included in their general practice records, and validity of CPRD diagnoses has been shown to generally be high.[106]   However, the analyses excluded factors not recorded in the primary care record that were present at or arose during the CAP admission, and the potential limitations of this are discussed below.

An important consideration is that the information made available to GPs about recent hospitalisations is different to hospitalisation information made available to researchers. General practitioners now receive an electronic version of the hospital discharge summary soon after the patient is discharged. However, much of the information from this summary is typically not coded into a patient's electronic health record, and thus will not be in the anonymised CPRD data. For example, if practice staff code only the CAP diagnosis and that the patient was recently hospitalised, then aspects

of the presenting features and the course of the CAP episode that are potential predictors of subsequent mortality (such as hypotension or high temperature at admission, or the patient being non-stable at discharge, Table 7.1), will be known to the GP from the discharge letter but will not be available in the primary care electronic record.

Incorporation of HES data into the model is unlikely to have addressed this problem, as the coded information provided in HES is not the same as that contained in the discharge summary received by GPs. Both sources of information are derived from the patient's records during their hospitalisation, but the discharge summary is written by a member of clinical staff (primarily a doctor), usually at the point of discharge. In contrast, HES records are derived from the information provided by a team of trained clinical coders, who extract information from the patient's hospital records on the conditions present at admission or arising during hospitalisation, after the patient has been discharged. Presenting features of an illness and its clinical course are rarely captured in the ICD-coded HES data, and thus I would not have been able to investigate these features in the prognostic models by using the HES data.

It is also important to note that GPs do not receive HES data, and that hospital discharge letters may not include all information on the hospitalisation. There are a limited number of studies which have evaluated the completeness of English discharge summaries, but reports suggest the information on co-morbid illnesses is not complete. An audit of discharge summaries received by a primary care trust in Eastern England between January and March 2011 found that only 50.3% of the summaries written by doctors contained complete information on patients' co-morbidities, although information on the presenting diagnosis was well recorded (approximately 94% complete (data read from the figure provided in the report)).[177] A separate 2009 survey by the Care Quality Commission of 280 general practices across England found that 56% of the practices reported receiving inaccurate or incomplete discharge summaries over the last year 'some of the time', and 16% reported that this occurred 'most of the time', compared to 27% reporting 'not very often' or 'never' receiving inaccurate/incomplete summaries.

Assuming the findings from these reports were broadly applicable to all discharge summaries over the time period of this study, then the information on diagnoses provided in HES might not be sufficiently similar to that provided to GPs in discharge summaries to be useful for developing prognostic models for use in primary care. As previously mentioned (section 7.3.1), a key feature of the data used to develop a prognostic model is that these data are as similar as possible to the population the model will be applied to. Inclusion of HES data could have increased the discriminative ability and internal calibration of the model. However, if the model was then applied to an external sample of primary care data from practices for which HES records were not available (but instead potentially incomplete recording from discharge summaries), the calibration of the model would suffer due to the differences in recording, and the predictions made by the model would be less reliable (i.e. the external calibration would be poor). The addition of HES information and its potential effect on the models is further discussed in section 7.6.4.1.

A final potential limitation relates to the acceptability of the score to general practitioners. Although I discussed the development of the prognostic models with two general practitioners, I did not consult more widely with a sample of GPs to determine whether excluding information about a patient's most recent hospitalisation would be clinically acceptable to them. If this exclusion was not acceptable, GPs would be unlikely to use the score.

7.6.2.2 Other considerations

The low, non-uniform recording of frailty factors revealed in Chapter 6 resulted in the majority of the frailty factors included in that work being unsuitable for inclusion as potential predictors of mortality. This was unfortunate, as such measures of underlying deficit may have strengthened the discrimination of these scores. Recent attention on the effects and importance of frailty among older adults will hopefully lead to more complete recording of these factors over time, so they can be included in future, updated versions of these models.[158]

Despite the absence of these frailty-specific factors in the models, it is possible that the models were simply identifying frail older patients. In order to assess whether the models were identifying factors specific to mortality after a CAP hospitalisation, I would

have needed to apply the models to an alternative study population (such as patients hospitalised with an unrelated condition to CAP) and assess their calibration and predictive accuracy among this group. I also did not compare my models to other more generic models such as Charlson, or the eFI (which was not published at the time of this work), and so cannot be certain that my models are better at identifying patients at higher mortality risk compared to those already available.

Patients with a code for terminal illness or metastatic cancer were excluded from the analyses, as the aim of the study was to predict unexpected deaths. Directives such as 'do not resuscitate' orders or living wills are very poorly recorded in CPRD, and so patients who had expressed this wish (and thus would receive little/no active treatment for their CAP) but did not have terminal illness codes could not be excluded as 'expected' deaths. This may have affected the shorter 1-7 and 8-30 day models, but should not have been such an issue for the primary (31-365 day) model.

While more than one CAP hospital discharge per patient was included in the dataset, only the first discharge in one 365 day period was used. This may mean that predictions from the model are less clinically useful for predicting mortality after a patient's second or third CAP hospitalisation in a year. Multiple CAP hospitalisations in such a period could be considered themselves a marker for underlying health deficit, and GPs could additionally use this as a marker of increased risk.

The longer models (31-365 and 8-30 days) both had considerably more than 10 outcomes per factor investigated, but the shortest (1-7 day) model had slightly too few outcomes per factor. However, this model may be the least useful of the three, due to the timing of receipt of discharge summaries by GPs.

While I validated the prognostic models using the well-regarded internal bootstrapping technique, I was unable to perform external validations. Were these models to be considered for clinical use, future work would need to include validating the models developed in the CPRD data in other large, UK GP databases to assess the accuracy of their predictions in other settings.

7.6.3 Clinical relevance of the model

Since this work was completed, GRADE guidelines have been published on the evaluation of tests used for clinical decision-making.[189] The full framework has mainly been described in its role as an aid for panels who are making decisions and recommendations about tests and their usage, but it can also be applied to individual studies. The GRADE guidelines include several items which have already been discussed in this thesis; consideration of whether the problem is a priority and the importance of the outcome are covered in section 7.1, the accuracy of the tests (in this case the models themselves) is covered in section 7.5.4. Below, I discuss other aspects of the framework.

7.6.3.1 Evidence of the suitability of an intervention to implement post risk stratification

I developed these models with the aim of helping GPs to identify older patients at increased risk of mortality after a CAP hospitalisation. However, as with any test, when developing a new prognostic model it is important not only to assess whether the model correctly identifies patients at increased risk of the outcome, but also to consider its clinical applicability, and whether any proposed intervention has any effect on the outcome.[189]

Prognostic models are designed to be used in addition to (not instead of) clinical judgement. As such, it's unlikely that the scores developed in this study would be used to aid assessment of the most obviously high risk patients (for example those aged ≥90, residing in a nursing home and with dementia). The aim is to aid GPs when they are not sure of the level of mortality risk, or when the risk is less obvious (i.e. in the mid older-age groups, and the middle/upper centiles on the calibration graphs, which is where these models performed most reasonably). While there is likely to be a greater level of uncertainty in the model among patients with a high number of conditions (as combining many coefficients' standard errors will lead to wider confidence intervals), GPs will be aware of these patients due to their worse health.

Unlike previously developed scores for other conditions, to the best of my knowledge there are currently no specific preventative interventions for patients at high-risk of mortality post-CAP (given the varied causes of death after CAP), and there is no

definitive evidence that identifying high-risk patients will affect mortality rates among this group. The Swiss study identified in the earlier literature review (section 7.2.3) suggested that closer patient contact should be assessed in future prospective trials.[175] My original idea when developing these models was that GPs could implement a form of case management, consisting of the assessment of patients with subsequent formulation of care plans specific to their pre-existing co-morbidities, and potentially their social care needs.[190] Based on previous research, possible items for inclusion in the care management plans include a patient's clinical history, current health status, medication review, ADL ability and needs, mobility and cognitive function, formal and informal care arrangements, physical and social care needs.[190]

Studies have assessed the impact of case management in primary care on preventing adverse outcomes in a range of patient populations. Within primary care, case management has most commonly been used to try to prevent emergency hospitalisations, but mortality has also been considered as an outcome. A 2015 systematic review of case management for 'at risk' patients in primary care found that the 12 studies which assessed short-term mortality (0-12 months), and the 13 studies that assessed longer-term mortality (13+ months), both showed heterogeneous results. None of these studies were in an older population recently discharged from hospital after CAP. When results were pooled there was no evidence of an overall effect of case management on decreasing mortality. However, subgroup analyses indicated that there might be small benefits of delivering case management via a multidisciplinary team and of including a social worker, and the authors suggested that these findings might be further explored in future studies.[191]

7.6.3.2 Benefits and harms of the models

Due to the nature of the models, there are no associated side-effects which could harm a patient. As discussed above, there is currently little/no evidence that my proposed intervention of enhanced case management would lead to reduced mortality rates among the target population for the model, and as such the models may have no/limited benefits.

7.6.3.3 Balance between desirable and undesirable effects

The unclear benefit of the case management approach also impacts on the balance between desirable and undesirable effects. While the undesirable effects of using the model should be minimal, the predicted benefits of my proposed intervention could also be low.

7.6.3.4 Acceptability of the score and intervention to patients and GPs

Should the use of the score and case management be found to help reduce one-year mortality in older patients, then acceptability and adoptability of both the score and the intervention in the real world would need serious consideration. The score was designed to be able to be inserted into software used by GPs, and thus to be easy to use. However, an ever-increasing number of prognostic scores are available in primary care and it is possible that GPs would not be willing or have time to adopt them all. Additionally (as outlined in section 7.6.2.1) it is not clear whether GPs would find it acceptable to adopt a score that did not incorporate the most recent patient data from the hospitalisation, and this would need to be assessed.

From a patient perspective, the score itself would likely be acceptable as it requires no input or action. After being identified as at increased risk, some patients may be reticent to spend more time with medical professionals after hospital discharge, and may not feel that case management is necessary. The reasons for the case management approach would need to be carefully explained to this group in order to let them make a fully informed decision about how to proceed.

7.6.3.5 Equity

The integration of the score into GPs software would result in it running automatically if a patient had been coded as being hospitalised with CAP. This automated approach should result in the score being applied to patients equitably.

7.6.3.6 Feasibility and resource use

Case management is resource intensive due to its potential multidisciplinary nature. My intention was to aid GPs in identifying patients with increased risk of mortality but who might not otherwise have been identified (i.e. not the most frail patients, with whom

the GP would likely to already have frequent contact). The relatively high cost of this intervention may preclude it from being used in a population not necessarily considered classically 'high-risk'. If effective, case management could prevent some decline in health and further hospital admissions, and thus avert the associated healthcare costs. However, without conducting a cost-effectiveness analysis, the viability of this approach is uncertain.

7.6.3.7 Conclusion of GRADE assessment

Due to population ageing and increasing hospitalisation for CAP, increased mortality after CAP hospitalisation should be considered a priority. However, due to the lack of evidence that my proposed intervention would have an impact on patients' mortality risk, the potentially high resource use and the moderate/weak predictive ability of the models I developed, they would not currently meet the GRADE criteria to be introduced into clinical practice.

7.6.4 Reflection on my approach to aspects of this work

Below I reflect further on the lessons learned from this work, discuss aspects that could have been carried out differently and the potential impact this would have had on the models and their performance.

7.6.4.1 Exclusion of information regarding the most recent hospitalisation

As discussed in section 7.6.2, I chose not to include information available in HES about patients' CAP hospitalisation in the calculation of their risk score. If the recording of co-morbidities on discharge summaries becomes more consistent with that recorded in HES, the addition of co-morbidities recorded in the HES data during the CAP admission could enhance the predictive ability of my models by more accurately reflecting patients' most recent health status. This could better delineate those who are starting to decline from those who are generally well. These co-morbidities could be included as separate 'recent' variables if not previously recorded in the patient's primary care record, to signal their treatment during the hospitalisation. Knowledge of the more recent occurrence of a co-morbidity may increase its prognostic score value, for example an MI occurring during a hospitalisation for CAP may have a stronger association with one-year mortality than an MI recorded five years ago.

Although HES data are suboptimal for recording factors that indicated the severity of CAP at presentation or during the hospitalisation, one factor I could have considered from the HES data (also included in hospital discharge summaries) is the patient's length of stay in hospital. This could have provided useful information in the severity of illness and been used as a proxy for the patient's underlying health, as those who are frailer may have longer admissions than patients who are generally fit and well. Were I to repeat these analyses, I could investigate length of stay as a potential predictor to ascertain whether it could improve the prognostic ability of the models.

An alternative strategy would be to include patients' self-reported details of their hospitalisation. Information such as conditions arising in hospital could be ascertained, although information regarding any change in medications should strictly be excluded (due to the potential of modelling prescribing behaviour rather than the true risk associated with the medication, as described in section 7.3.1.1). The accuracy of self-report would need to be validated prior to its inclusion in the model, in particular among the older, more infirm patients who may recall these details less well than their younger or healthier counterparts.

7.6.4.2 Assessing whether the inclusion of information from the most recent hospitalisation improves the prognostic ability of the models

There are formal methods to investigate whether the addition of new candidate predictors (here, using predictors from the hospitalisation data), improve a model's ability to classify patients into the correct outcome group, i.e. its sensitivity and specificity. Net reclassification improvement (NRI), is a method of measuring whether patients with the outcome move up a classification category, while those without the outcome move down. It is simplest to calculate and interpret when only one cut-point is being evaluated (i.e. the score only categorises patients into two groups). In this case, the event NRI and non-event NRI are first obtained. The event NRI is calculated among those with the outcome, as the proportion who moved up a category minus the proportion who moved down a category (the improvement in sensitivity). The non-event NRI is calculated among those without the outcome, as the proportion who moved down a category minus those who moved up a category (the improvement in specificity).[181, 192] The overall NRI is then calculated as the sum of the event NRI and

non-event NRI. Large overall NRI values may have been driven by an increase in correct reclassification of subjects either up or down (an increase in sensitivity or specificity), and thus without knowing the event and non-event NRI the direction of any improvement (and the new model's clinical usefulness) cannot be ascertained. The denominators for the event and non-event groups are likely to differ, resulting in the NRI being weighted by the event rate and thus complicating its interpretation. Therefore, several factors must be considered when using and interpreting the NRI.

Most importantly for the models I present, NRI is only recommended for use where clinically meaningful cut points (or risk categories) can be chosen a priori, [192] and therefore the continuous NRI would be a better choice in this setting. The continuous NRI considers the change in risk for all events and non-events without the need to create categories. It is interpreted differently to the NRI, as not all changes in predicted risk would lead to a change in clinical management of a condition or implementation of an intervention, for example if a patient's risk only changed minimally. It has also been reported that the continuous NRI is often positive for weak markers, and that it can be strongly affected by miscalibration of the model.[192] Correct calibration of models and ensuring that the two models to be compared were developed from the same data can minimise these issues.[192]

Both NRI and continuous NRI are predominantly used to assess the addition of one or a small number of predictors to a pre-existing model. When the classification abilities of two different models are to be compared and they are not well correlated, then both the discrimination and calibration of the models should be considered instead of the reclassification.[192] This is due to the actual classification of patients being of greater interest than the movement of patients between groups.

To assess whether the addition of discharge summary information improved the classification of the model, I would therefore need to first examine how well the two models were correlated. Given the relatively small amount of additional data, I would expect to be able to use the continuous NRI. In the case of the models I present, an increase in specificity (identifying more patients at a lower one-year mortality risk, and thus not assigning them to case management) would save clinicians time and health care resources, while not causing problems for the patient. Increases in sensitivity (correctly

re-classifying patients at a higher mortality risk) would enable these patients to receive case management, and thus increases in sensitivity could be considered of higher importance than increases in specificity.

7.6.4.3 Changing study population

When developing these prognostic models, the study population comprised older patients who had been hospitalised with CAP between 2004 and 2011. As shown in Chapter 6, hospitalisations with CAP increased over this period, but the increase did not appear to be explained by changes to the population's co-morbidity profile, and short-term mortality decreased. The inclusion of data collected over a period during which there may have been an increasing tendency to hospitalise older patients with CAP could have created a cohort that had changing risks for longer-term mortality. For example, the one-year mortality risk among those with specific co-morbidities may have decreased over time if those hospitalised with CAP more recently had less severe underlying disease compared to CAP patients hospitalised in 2004. Had the prognostic models been calculated separately for each year, this scenario could have resulted in HRs for specific co-morbidities that were closer to null in later years than in 2004. Had I restricted the study population to patients hospitalised with CAP in a later period, such as from 1st April 2008 (when CAP incidence first peaked in the linked data, as shown in Chapter 5) the cohort might have been less heterogeneous, and it is possible that the discrimination and calibration of the models could have been improved. The Beta-coefficients of the parameters in the model could then be updated annually (as happens with other prognostic scores such as QRisk,)[193] to take into account changes in disease severity in hospitalised patients that could result from changing health service patterns.

A second consideration is that trends in disease ascertainment and recording are likely to have changed over the study period for some conditions, due to enhanced case finding and recording (for example, as a result of QOF). This could have resulted in an already heterogeneous population becoming more diverse, and further diversified the group of patients categorised as having the specific co-morbidity in the prognostic models. If disease severity was similar in those with ascertained and unascertained disease, the misclassification of disease status could have underestimated the effect of that co-morbidity on mortality risk in earlier years. However, if (as seems more likely)

co-morbidities not ascertained until later in the study period were less severe that those ascertained earlier, the more recently diagnosed, less severely affected, patients could have had a lower one-year mortality risk than those diagnosed earlier. Again, this may have resulted in the HRs associated with the co-morbidity over the entire study period being further from the null compared to those that would have been estimated had the study only included more recent data.

For four of the co-morbidities of interest (diabetes, ischaemic heart disease, liver disease and renal disease) it was possible to broadly categorise patients into those with mild/moderate and severe disease, due to the codes used for these conditions. For other illnesses such as chronic lung disease, it was difficult to differentiate between those with mild and severe disease. For these latter groups in particular, the broad case-mix is likely to have resulted in the assignment of a score for subsequent mortality that was too low for the more severely ill patients in the group, and too high for those with mild illness. Similarly, the increased risk associated with specific age groups may not have captured the range of underlying health states, including frailty, within each age categorisation, and how this may have varied over the study period.

7.6.4.4 Alternative models or strategies that could have been investigated

It can seem attractive to create a single prognostic model for use in a wide range of patients; however, this may not always be the most appropriate approach. If I were to repeat these analyses, one alternative could have been to use the now-available eFI to stratify the study population into those who were fit (0-4 frailty deficits), mildly frail (5-8 deficits) or moderately/severely frail (≥9 deficits), and then develop separate prognostic models for each group. The case-mix of patients included in each of these models may be less heterogeneous, which could increase the prognostic ability of the model. In particular, the Beta-coefficients for variables such as age could be more appropriate for a larger proportion of the patients in each age group. This approach would not tackle directly the lack of depth of coding for some co-morbidities. However, in many cases, severe illness results in additional co-morbidity or increasing frailty and thus using the eFI to pre-stratify patients could also aid in creating more homogenous co-morbidity groups within each model.

Another potential approach would be to develop several disease-specific-models to assess predictors of mortality among patients with CVD or chronic lung disease. However, this approach could be complicated by patients with multi-morbidity potentially being assigned a different mortality risk in each disease-specific model, limiting its utility. Due to this major limitation, the eFI stratification approach may be more appropriate.

As highlighted above, restricting the study population to patients who were hospitalised with CAP from 2008 onwards could have been attempted to decrease heterogeneity in the case-mix over time. The drawback of developing models over a shorter time period, and possibly further stratifying by eFI to address differences in underlying frailty, is that the study populations would be considerably smaller than the one I used. However, CPRD have recently gained access to anonymised GP data from practices that use EMIS general practice software, supplementing the existing data collected from practices that use Vision software. Future provision of these EMIS data for research use will greatly increase the size of CPRD, and thus may help address sample size issues. If the study populations of interest are still not large enough to meet the ten outcomes per potential predictor rule, then it may be necessary to attempt to limit the number of candidate predictors.

7.7   Conclusions

Information widely recorded in patients' GP records has the potential to be used to inform GPs about patients' one-year mortality risk after they have been discharged from hospital due to CAP. The predictive value of the models was suboptimal, but could be increased in the future by using only recently collected data to develop the models, stratifying study populations by the extent of frailty, and including additional co-morbidity information from the most recent hospitalisation (once recording on hospital discharge summaries is more consistent with that found in HES). Before further models are developed and implemented, either the evidence for case management as an intervention would need to be strengthened, or an alternative intervention which could be implemented among higher-risk patients would need to be identified. Then models such as those presented here could be useful for GPs who are contacting patients on the

ES register, and enable GPs to assess the level of post-hospital discharge support that older individuals may require in the longer-term after a CAP hospitalisation.

Chapter 8   Discussion

The overall aim of this work was to use linked electronic health records to better understand the burden of community-acquired pneumonia among older adults in the UK, and to enable thorough assessment of risk factors for hospitalisation after CAP and for longer-term mortality after a CAP hospitalisation. This chapter summarises the key findings of the main objectives, considers the strengths and limitations of using linked electronic health records for their investigation, and highlights the implications of the findings for research and clinical practice.

8.1   Objective 1: Use of linked data to estimate the incidence of CAP and all LRTI

8.1.1   What was known

CAP and LRTI more broadly are common infections among older adults. Given the increasing size of the UK's older population, clarifying the burden of disease in this age group is important for health service planning. However, there were relatively few studies of the incidence of CAP (or of LRTI in general) in the UK, and none specifically set among the older population. Estimates of CAP incidence from the UK were generally lower than those from the rest of Europe. This could be because the UK studies only utilised primary care data; CAP can be diagnosed and treated in either primary or secondary care, thus use of only one of these data sources may underestimate the total burden of infection. Furthermore, there was a paucity of information about the burden of these infections in the UK over time, particularly age-specific estimates of incidence among those aged ≥65 years.

8.1.2   What this study adds

Novel methods were developed to identify episodes of pneumonia (and LRTI as a whole) from linked electronic health records (as described in Chapter 2), and to classify these as community- or hospital-acquired. Within the primary care data, I identified an appropriate start of follow-up which excluded historical episodes of CAP reported when a patient registered with their general practice (Chapter 3). The creation of illness-episodes prevented repeat consultations for an ongoing infection from being counted as new events, whilst allowing the inclusion of subsequent incident episodes. Use of

linked general practice and hospital data enabled capture of both moderate and severe cases of disease, more accurate identification (and subsequent exclusion) of hospital-acquired infections, and more accurate estimation of person-time at risk of a community-acquired infection.

The incidence analyses (Chapter 4) presented rates of CAP (and of all LRTI) among more than 1.5 million older patients over a 14 year period. Crude incidence of CAP was estimated as 7.99 episodes/1000 person-years (and 122.9/1000 for all LRTI). When stratified by age and sex, CAP estimates were almost 20% higher than those from a UK study using stand-alone primary care over a similar period. The large study population permitted incidence estimates to be finely categorised by age, which revealed that between the ages of 65-69 years and 85-89 years, rates of CAP increased more than seven-fold and rates of LRTI as a whole doubled. When examined over time, incidence of both CAP and of all LRTI generally increased (with fluctuations) between 1997/98 and 2010/11; the increasing CAP trend was somewhat attenuated after adjusting for age, suggesting that the rising trend could be due to population ageing.

Further assessment of the value of using linked data was performed by comparing CAP incidence using stand-alone CPRD data with that derived from linked CPRD and HES records in a single group of patients (Chapter 5). This revealed increasing divergence between estimates from the two sources over time. Rates derived from CPRD records did not increase over the study period, whereas those from the linked-data rose considerably. Most of this increase was due to a higher number of CAP events recorded in the HES data over time. Examination of CPRD records revealed that consultation with a GP on the day of the CAP diagnosis decreased over the study period from 82% to 43% of CAP episodes, and that consultation for any LRTI in the 28 days prior to the CAP diagnosis decreased from 15% to 10%. This suggests that in addition to patients increasingly presenting to hospital when suffering from CAP, these hospitalised episodes are not completely recorded in primary care data. This finding reinforces the value of using linked primary and secondary care data, and thus capturing both ambulatory and hospitalised episodes of disease.

8.2   Objective 2: Time trends and risk factors for hospitalisation after CAP

8.2.1   What was known

Hospitalisations for pneumonia in England have been shown to be increasing since 1997, and the majority of pneumonia hospitalisations are known to be among older adults. Use of secondary care data in isolation prevented these studies from distinguishing increasing pneumonia incidence from an increasing tendency to hospitalise older individuals with pneumonia. Prior to this work, no UK studies had attempted to examine risk factors for hospital admission among patients with CAP, and thus estimate the contribution of factors such as individual co-morbidities, medication use or frailty to the increasing hospitalisation levels.

8.2.2   What this study adds

The cohort study of CAP patients in Chapter 6 showed that the population-averaged percentage of CAP episodes which resulted in hospital admission rose from 57% to 86% of episodes between 1998/2000 and 2009/10 after adjustment for a wide range of factors. My use of linked primary and secondary care data revealed that hospitalisation post-CAP increased independently of any increases in CAP incidence over the study period. The linked data also permitted thorough investigation of potential explanatory factors, some of which (such as vaccination status) were not routinely recorded in HES data.

Fourteen co-morbidities, five frailty factors, and four medications/vaccinations were found to be associated with hospitalisation within 28 days of a CAP diagnosis. Use of individual factors rather than a summary co-morbidity score identified very different associations between individual factors and odds of hospitalisation. These ranged from terminal illness and dementia (which decreased patients' odds of hospitalisation) to metastatic cancer, chronic lung disease and severe renal disease which all increased the odds of hospitalisation. Replacing the individual co-morbidities included in the model with the Charlson co-morbidity index had almost no impact on the predicted probability of hospitalisation over time, but did conceal the direction and magnitude of individual risk factors' association with hospitalisation.

This study also showed that over the study period, the odds of dying in the 28 days post-CAP decreased progressively, and there was a slight decrease in length of hospital admission. Together, these factors suggest that increasing CAP severity may not lie behind the rising trend in hospitalisation.

While the factors identified explained little of the increase in hospitalisation post-CAP over the study period, investigation of other elements of the data provided possible explanation. Emergency admissions coded as originating from A&E increased, while those arriving via a GP declined over the study period. Records from primary care confirmed this trend; patients with a GP record of CAP (or LRTI more broadly) on the date of hospital admission also decreased over the study period. This finding is in line with that from the data source comparison analysis (Chapter 5) which showed decreasing consultations with GPs on the date of CAP diagnosis over time. Together, these analyses suggest an increasing tendency for older patients with CAP to present directly to hospital rather than via their GP.

8.3 Objective 3: Prognostic models to predict longer-term mortality post-CAP hospitalisation

8.3.1 What was known

Older patients hospitalised for CAP have an increased mortality risk compared with both patients hospitalised for other reasons and the population at large. Previous studies have shown that this risk persists after hospital discharge for periods ranging from one to more than ten years. A tool to aid clinicians in identifying patients at increased risk of death after hospital discharge would enable GPs to monitor such patients' underlying health conditions more closely, and potentially intervene to prevent some of these deaths. Several prognostic scores are available to assist clinicians in decision making around whether to hospitalise a CAP patient at the point of diagnosis, based on predicted risk of death within 30 days. However, studies of risk factors for longer-term mortality after discharge from a CAP hospitalisation are uncommon, and none have been undertaken specifically in the high-risk older population. The risk factors identified in these papers were not restricted to those that are readily available to GPs, and thus would be of limited use after hospital discharge. As described in the previous Chapter, only one existing study developed a prognostic model to be used post-hospital

discharge, but did so using incorrect methods and without validating the final model either internally or externally, and so its ability to perform in real-world settings is unknown.

8.3.2 What this study adds

Prognostic models for mortality post-CAP discharge were developed for three time periods: 1-7 days, 8-30 days and 31-365 days after hospital discharge, and each was internally validated using bootstrapping. The prognostic scores were designed to need as little input from GPs as possible, making use of pre-existing information held within electronic general practice records. In addition to the a priori variables age and sex, the candidate predictors included in the models varied by time period. Of the 12 factors (1-7 days), 17 factors (8-30 days) and 17 factors (31-365 days) included in each model, seven (peripheral vascular disease, dementia, connective tissue disease, cerebrovascular disease, solid cancer, living arrangements and low weight) were included in all three models. The shorter scores showed reasonable discrimination, while the longest was sub-optimal. All showed fair calibration.

8.4 Strengths of the studies in this thesis

Databases of electronic health records are increasingly commonly used in epidemiological research, due to their many strengths. The linkage of these data to additional data sources is also becoming more common, providing researchers with a wealth of detailed data. The use of linked data has resulted in the studies in this thesis having many strengths, as outlined below.

8.4.1 Use of linked data

Linkage of CPRD to HES and ONS death records provided an additional level of detail not available in stand-alone CPRD that benefitted the studies in a number of ways.

8.4.1.1 Identification of CAP

Linkage of CPRD to HES was important for many aspects of determining CAP and its incidence. Firstly, inclusion of hospital admissions for pneumonia enabled more complete capture of these events than would have been possible using CPRD alone, and resulted in a dataset of pneumonia episodes treated in hospital as well as episodes

managed in the community. As shown in Chapters 4 and 5, this resulted in the reporting of more comprehensive and detailed CAP incidence estimates among older adults than had been achieved previously in the UK. Secondly, use of the linked data informed the categorisation of pneumonia episodes into CAP or HAP (and thus the study populations of Chapter 6 and Chapter 7). Although CPRD contains some information on patients' hospitalisations such as the admission diagnosis, admission and discharge dates are not included. Furthermore, the date a hospitalisation is recorded in CPRD may refer to the date of admission, date of discharge or the date the discharge summary was received by the general practice. The accurate dates of admission and discharge provided by HES-linkage were crucial to these analyses for several reasons, outlined below.

8.4.1.2 Identifying episodes of infection, person-time at risk and CAP incidence over time

Identification of all pneumonia (and other LRTI) records in both data sources enabled me to differentiate those that were related to an ongoing illness from those that were a new episode, by using a 28 day illness-episode structure. I was then able to include the incident episodes experienced by each patient in my analyses, unlike many previous studies which have either included all consultations, or restricted analyses to the first diagnosis of CAP.

The CAP illness-episode structure, together with hospital admission and discharge dates were used to define person-time not at risk of community-acquired infections in the incidence analyses presented in Chapter 4. This improved the accuracy of the estimated incidence rates by only including person-time at risk of community-acquired infection.

The linked data also facilitated better estimation of trends in CAP incidence. Use of stand-alone primary or secondary care data may show a change in illness trends attributable simply to patients moving from one care setting to another over time. This was illustrated in Chapter 5, in which CAP incidence estimated using stand-alone CPRD data suggested a slightly decreasing rate over time. However, this was likely to have resulted (at least in part) from an increasing number of patients presenting and being first diagnosed with CAP in hospital, rather than a real decrease in disease incidence. Linkage of data sources provided protection against mis-characterisation of disease burden due to changing service provision or utilisation.

8.4.1.3 Better identification of risk factors for severe outcomes of CAP

The date of CAP diagnosis and the hospital admission date provided by the linked data were used to define the outcome (hospitalised within 28 days of a CAP diagnosis) in the analysis of risk factors for hospitalisation. Risk factors for hospitalisation post-CAP diagnosis are likely to differ from those for contracting CAP, but it is difficult to separate these using stand-alone data. Use of linked data allowed restriction of the study population to those who had been diagnosed with CAP and thus examination of characteristics associated with their subsequent hospitalisation risk.

The linked data were also essential for developing the prognostic models for mortality. Firstly, date of hospital discharge (from HES) provided an accurate start date for the mortality risk period. The ONS-linked mortality data provided an accurate date of death, which was of utmost importance when examining time to death. While CPRD records include a date of death, it is estimated from up to three dates that can be recorded within the general practice records, and so cannot be considered definitive. When analysing short periods of follow-up (for example 1 to 7 or 8 to 31 days in the prognostic modelling work), errors in the date of death by even a few days may result in patients being included in the wrong mortality model, and inaccurate estimation of predictors of mortality.

8.4.1.4 Identification of potential risk factors for severe outcomes

The linked data provided more complete information on patients' co-morbidity status. The CPRD data provided a record of patients' longer-term medical history, while the use of HES records ensured that diagnoses made in hospital were fully captured. Research comparing CPRD and/or HES to disease registries for MI and cancer show that linking CPRD to HES captures 23% and 33% more events respectively than use of stand-alone CPRD.[194, 139] Additionally, each data source provided information not available in the other. For example, CPRD provided detailed information on factors such as prior vaccination, therapies, smoking status and frailty, while admission diagnoses in HES enabled identification of hospitalisations for events such as cataract surgery.

8.4.2 Large sample size

One of the primary strengths of both CPRD and HES is their considerable size, containing data collected over many years. The study population for the incidence analyses in Chapter 4 contained over 1.5 million patients, over 900,000 of whom were eligible for HES linkage. The large sample size enabled thorough examination of rates by age and sex over time, in a level of detail not previously reported in studies of CAP among older UK adults.[128, 121] The >900,000 patients in the linked data provided more than 45,000 CAP episodes in England over 13 years for the analysis of risk factors for hospitalisation. Having a sizable study population facilitated the risk factor analysis, obtaining precise estimates of relative risk for a wide variety of individual co-morbidities and other factors which could influence hospitalisation decisions. The availability of patient records over several years enabled me to then assess the contribution of these factors to increasing hospitalisation trends over time, which previous smaller studies comparing ambulatory and hospitalised CAP had not been able to do.[122, 125, 130] In Chapter 7, patients' CAP episodes were restricted to those that required hospitalisation and were discharged between 2004/5 and 2010/11, and >12,000 CAP episodes were eligible for inclusion in the development of prognostic models for long-term post-CAP mortality. This provided a large population in which to investigate the contribution of the 44 candidate predictors of mortality, and (with the exception of the 1-7 day model) the analyses contained more than the recommended 10 outcomes per potential predictor. The studies presented in this thesis would have been either impossible or severely limited in their scope if the data available had been smaller in size.

8.4.3 Prospectively collected data

Both CPRD and HES contain routinely recorded information, collected prospectively. CPRD captures all consultations between a patient and their GP (which are typically coded during the consultation), as well as patients' medical history, prescriptions and vaccinations. Diagnoses in HES capture the illnesses patients are treated for during a hospitalisation (including the main condition treated, any underlying conditions that require treatment and medical events that arise during the hospitalisation), and these diagnoses are translated into their equivalent codes at the end of the hospitalisation by specialised teams of medical coders. The prospective nature of the recording of

exposures such as co-morbidities, vaccinations and medications eliminated the possibility of recall bias (in which patients' reporting of their exposure status is influenced by their outcome status). Observer bias (when knowledge of exposure status influences the classification of the outcome) should not have been an issue, as the outcomes were hospitalisation or death, defined using dates from the HES- and ONS-linked data.

8.4.4 Assessment of hospitalisation for any cause after CAP

I did not restrict the analysis in Chapter 6 to hospitalisations specifically for CAP, but also included hospitalisations for any reason in the 28 days after CAP. This will have captured the fuller effect of CAP via its role in worsening of patients' co-morbidities, in addition to any acute events it may have precipitated such as MI, stroke or falls. This will be of particular use when forecasting healthcare use among the older population.

8.5 Potential limitations

Despite their many strengths, it is important to remember that these records are primarily intended for clinical use, and the use of linked data thus provides some challenges and some limitations. Those of specific studies were discussed in the relevant chapters; these are also summarised here along with overarching potential limitations.

8.5.1 Misclassification

8.5.1.1 Misclassification of CAP due to imperfect diagnostic validity

As mentioned in previous chapters, pneumonia diagnoses have not yet been validated in either CPRD or HES, and thus there is no clear gold-standard. As highlighted in section 1.1.4, clinical diagnosis of pneumonia in adults has been shown to have low sensitivity compared to diagnoses made with chest radiographs and the diagnosis can be particularly difficult in older adults, who can present with fewer, non-specific symptoms.[18-20]

A previous multi-centre European study of adults has indicated that clinical diagnoses of pneumonia in general practice had sensitivity as low as 29% when compared to subsequent chest x-ray in the week following diagnosis.[20] It is possible that a similar level of under-diagnosis occurred in the CPRD data, although the older study population

included in these analyses may have been treated more cautiously by GPs than the somewhat younger patients that were included in the European paper.[20] If GPs in the UK did regularly under-diagnose less severe cases of CAP (for example due to the less specific presentation of CAP among some older patients), this would have consequences for objectives 1 and 2 in this thesis. In the incidence study CAP would have been under-ascertained in primary care, and thus the incidence estimates presented in Chapter 4 are likely to be underestimates. Any under-ascertainment of ambulatory CAP episodes would have also affected the analysis of trends in post-CAP hospitalisation in Chapter 6, leading to an overestimate of the percentage of cases hospitalised. As discussed in Chapter 5, the higher estimates of CAP incidence obtained from linked hospitalisation data compared to stand-alone general practice data are likely to be due to CAP diagnoses made in hospital that were either undiagnosed when seen earlier by GPs, or that were not seen by GPs (because patients presented directly to hospital), with the hospital diagnosis being incompletely recorded by general practices. If the under-ascertainment in general practice increased (i.e. sensitivity decreased) over time, while the sensitivity of pneumonia diagnoses in hospital remained stable or even increased (for example due to more frequent use of CT-scans), this would also have contributed to the greater increase over the study period in diagnoses of pneumonia in HES than CPRD reported in Chapter 5, as well as the growing hospitalisation trend shown in Chapter 6.

When thinking about misclassification of disease status, it is important to consider not just the sensitivity and specificity, but also whether the recorded diagnoses are correct (their positive predictive value). Again, the lack of access to chest x-rays in primary care may have resulted in pneumonia diagnoses in HES having a higher positive predictive value than those in CPRD. In the multi-centre European study of adults with cough, clinical diagnoses of pneumonia by GPs had a PPV of 57%.[20] However, prevalence of pneumonia in the European study was only 5%, and this limits generalisability of these PPV findings to older study populations with cough, who are likely to have higher prevalence of pneumonia. On its own, imperfect positive predictive value of recorded pneumonia diagnoses would have resulted in an overestimation of CAP incidence. However, if the number of non-CAP cases incorrectly included was lower than the number of cases missed due to the imperfect sensitivity of the diagnoses, this would

have resulted in an overall underestimate of incidence. In the hospitalisation risk factor analyses, inclusion of some ambulatory cases that were not CAP may also have affected effect estimates for some covariates.

8.5.1.2 Misclassification of CAP due to diagnostic coding practices

Over the period of study, the method by which hospitals were assigned funding changed, and this may have influenced the coding of diagnosis in the hospitalisation dataset used for this thesis. Prior to 2003/4, hospitals were assigned funds according to a block grant based on their previous year's costs and activity.[195] Between 2003/4 and 2007/8, a system called Payment by Results (PbR) was gradually introduced in England, which was developed to better link payment to the case-mix of patients treated. The aims of PbR were to provide more transparent funding, reward efficiency, and encourage hospitals to work toward reducing waiting times.[196] The system is similar to those used in the US, Canada, Australia and several European countries, whereby hospitals are paid for the services they have provided. Within the English NHS, this payment is derived from the diagnoses made during a period of care, for example a hospital admission, and these diagnoses are translated post-discharge into ICD-10 codes by clinical coders. The coded diagnoses are collected centrally, and these are the codes provided to researchers to be used in HES. For the purposes of PbR, the codes are grouped into Healthcare Resource Groups, from which a payment for the care is assigned.[196] National tariffs are published every year, along with adjustments which may be made for short/long stays, best practice and specialised care.[196]

This type of payment system is susceptible to a practice called 'upcoding' or 'gaming', whereby patients are coded as having a condition that is more severe or more expensive to treat than the condition they actually have, thus increasing the income of the healthcare provider. Instances of upcoding have been reported in the US and Europe, while preliminary studies of upcoding in England have provided conflicting evidence.[195] A systematic review of discharge coding in the UK compared discharge coding accuracy before and after the introduction of PbR in 2004, and found no differences in coding accuracy overall (pre-PbR 77.0% (IQR: 66.2-89.0), vs. post-PbR 86.1% (IQR: 73.1-96.1%)). Accuracy of the primary diagnosis improved after PbRs

implementation (pre-PbR 73.8% (IQR: 59.3-92.1%), vs. post-PbR 96.0% (IQR: 89.3-96.2%), p=0.020).[197]

However, absence of evidence is not evidence of absence, and it is possible that upcoding occurs within the NHS. There are several ways in which upcoding could have affected this work. One is if LRTIs less severe than pneumonia were coded as pneumonia in order to attract a higher payment tariff. This could explain at least some of the increase in hospitalisation after CAP over time that was described in Chapter 6, and would be compatible with the slight decrease in length of hospitalisation over that period. Alternatively, some patients who had pneumonia, but who were admitted and treated primarily for a different condition that would have resulted in a lower tariff, could have been wrongly assigned a primary diagnostic code of pneumonia. If these cases had pneumonia that was hospital-acquired, this would have resulted in over-estimation of the incidence of CAP, and of hospitalisation after CAP over time. It is possible that patients without pneumonia or any other LRTI were also assigned a pneumonia code, but given the increasing validity of HES diagnoses over time, there is little evidence to support this.

8.5.1.3 Misclassification of CAP due to methods used to manage the data

I categorised pneumonia episodes which started within 14 days of a hospitalisation as hospital-acquired. As previously discussed, in stand-alone CPRD data hospitalisation was identified using information recorded by the practice from discharge summaries, which may have been recorded on one of several dates. If the date of the record was before the real date of discharge, or if the hospitalisation was not recorded in CPRD, some pneumonia episodes would have been incorrectly categorised as CAP rather than HAP. Conversely, if the recorded date was some time after the real discharge date, some episodes of CAP may have been incorrectly categorised as HAP. More generally across both stand-alone and linked data, the use of a 14 day rule could have led to some misclassification between CAP and HAP. There is no consistent choice across the literature of the timing of the exclusion window for HAP; previously used exclusion periods vary from between 7 to 30 days after hospital discharge.[81, 176, 19] A two week exclusion period is commonly used,[80, 85, 198, 86] but as with any rule, exceptions are possible. For example, a small number of CAP episodes may have been

excluded for patients who were infected soon after discharge from hospital. Misclassification of CAP and HAP may thus have occurred to some small extent due to incorrect recording of hospitalisation timing in stand-alone CPRD or to the use of the 14 day hospitalisation rule. Nevertheless, the analyses in this thesis approach the issue of separating hospital and community-acquired infections in a more detailed and thorough manner than has previously been attempted when examining CAP in the UK.

Re-consultation for an ongoing LRTI within 28 days has been shown to be common (25% to 33% of patients),[35, 57-60] so I used a 28 day illness episode structure in order to combine multiple consultations for one illness into a single CAP illness episode. It is possible, albeit very unlikely, that a small number of patients did experience two CAP episodes within 28 days, of which only one would have been included in the study. Over-recording of CAP was potentially more of an issue, although this too would have been at a low level due to the vast majority of patients only experiencing one CAP event over the study period. Over-recording may have occurred if one episode of illness was recorded twice more than 28 days apart, either separately in CPRD and HES, or repeatedly in CPRD (for example once as the patient's initial diagnosis and then subsequent recording of a hospital diagnosis). Any over-counting of the number of episodes would have resulted in overestimation of CAP incidence. If the recording occurred in both CPRD and HES, these single episodes would also have been included twice in the risk factors for hospitalisation analysis, and thus the importance of their underlying co-morbidities on odds of hospitalisation may have been slightly over-estimated. The restriction of the mortality prognostic models to the first CAP in a year will have prevented this from being an issue in these models. Electronic discharge summaries were increasingly used across the study period, with the addition in 2008 of a 72-hour target for their receipt by GPs (decreasing further to 24 hours in 2010).[199] If very late recording did occur, it would have been in low and in decreasing numbers, and should not have significantly changed the results presented.

As outlined in section 2.1.2, HES data can include up to 20 recorded diagnoses across an episode (a period of consultant care), along with the dates each episode began and ended. Unfortunately, information which would have enriched this work, such as the specific date each diagnosis was made, the diagnosis at presentation to the hospital or the main reason the patient was admitted to hospital is not provided. I used the primary

code of the first episode as a proxy for the reason for admission, as it seemed reasonable to assume that the majority of patients will have been admitted to hospital for the main condition that was subsequently treated or investigated over the episode of care. This approach was also taken to minimise inclusion of episodes of pneumonia arising in hospital (HAP), assuming that HAP would usually be coded with a secondary diagnostic code. However, there will have been occasions when patients admitted to hospital with CAP did not have pneumonia as the primary diagnostic code. For example, if a patient with CAP was admitted for a more severe condition or one which required a larger amount of care than pneumonia (such as a myocardial infarction, or sepsis following pneumonia), then this condition would have been the primary code of the first episode, and the pneumonia recorded as a secondary code. I would not have included these cases (unless they were also coded as pneumonia in CPRD on or before the day of admission, see section 2.4.3), thus underestimating the incidence of CAP and level of hospitalisation.

Conversely, some of the hospitalisations in the 28 days following a pneumonia diagnosis made in the community, and some of the hospitalisations that had a pneumonia code as the primary diagnosis for the first episode, may not have been a result of CAP. For example, if a patient developed a mild case of CAP in the month before a scheduled elective hospitalisation and the CAP was treated promptly, it is possible that the elective surgery subsequently went ahead. However, it is unlikely that this would apply to many of the CAP episodes in the study, as most of the hospitalisations after a CAP diagnosis were on or very soon after the CAP diagnostic date. Alternatively, if a patient was admitted with a less severe condition than pneumonia and subsequently acquired pneumonia in hospital during the first episode of care, it is possible that the pneumonia would have been chosen as the primary code of the first episode and a case of HAP incorrectly included in my analyses. In retrospect, there were additional ways I could have utilised the data to further tease apart community- and hospital-acquired infections.

One possibility would have been to limit pneumonia diagnoses to those classified as emergency admissions. As shown in Table 4 of the hospitalisation paper (section 6.3.3), there were very few cases in which HES pneumonia codes were included and the admission type was recorded as 'elective' (defined as "when the decision to admit could

be separated in time from the actual admission").[200] These cases made up 2% of total CAP hospitalisations included in the analyses in Chapter 6, and this fluctuated a little by year (from 2.7% in 2001 to 1.5% in 2009).   While this was a small proportion of the CAP episodes included my analyses, I would exclude them if I was repeating this work, as they are more likely to have been HAP than CAP. I would also exclude any elective admissions in the 28 days after a CAP diagnosis made in the community.

I could also have examined the treatment speciality of the consultant to which the patient was assigned during the initial period of care. This may have enabled me to identify cases with pneumonia who were treated on a surgical ward, and these cases could also have been excluded as probable HAP. A more wide-reaching method would have been to flag all first episodes in HES which included both a primary code for pneumonia (i.e. when pneumonia was the main condition treated during the episode) and a code for any unrelated surgical procedure, as being potential HAP episodes. This would have helped to identify patients who had surgery and subsequently developed HAP, which was then treated whilst under the care of the same consultant surgeon. Pneumonias diagnosed after surgery and treated by a medical rather than surgical team would have been recorded in a subsequent episode of the hospitalisation and thus not included in my analyses. However, given the large number of surgical codes that this approach would have required to be considered, and the lower incidence of HAP than CAP, this would have been a labour intensive method to identify a small number of cases in addition to those I could have identified using the 'elective' admission type.

8.5.1.4   Misclassification of other covariates of interest (co-morbidities, etc.)

The positive predictive value of diagnoses in CPRD has generally been found to be high, and the accuracy of diagnoses in HES is improving.[106, 112] When ascertaining the presence of co-morbidities in the analyses in Chapters 6 and 7, I included any diagnosis prior to the CAP incident date as evidence of a patient having the condition. For some conditions such as cancer or stroke, a code may represent either an ongoing or a resolved event. Utilising diagnoses from the complete patient record prior to the CAP considers ongoing and resolved comorbid events to have an equal strength of association with the outcome, which may not be the case. However, the inclusion of all

prior diagnoses also ensures the inclusion of current co-morbidities such as chronic lung disease, that GPs may not repeatedly code (as they are not required to do so).

Fewer validation studies have examined the sensitivity of diagnoses in these data.[106] In the analysis of risk factors for hospitalisation, the use of linked CPRD-HES data should have enhanced the recording (i.e. sensitivity) of co-morbidity status for patients who had been hospitalised for any reason prior to their CAP episode. If these patients were also more likely to be hospitalised within 28 days of CAP this could have resulted in differential misclassification of exposure status, leading to possible over-estimation of effects of these co-morbidities on the odds of hospitalisation. The vast majority of the older patients included in this analysis had hospital admission records at some point in time (94.4%), and so any bias is likely to have been very small. Conversely, it is also possible that some factors, such as those pertaining to frailty, BMI, and smoking (in the earlier period of the study) were under recorded (had low sensitivity). The effect of any bias from this misclassification would depend on whether these variables were better recorded in less well patients (who were more likely to be hospitalised or die) than in patients without the outcome of interest. An important outcome of this thesis is that the low level of frailty recording in these data has now been highlighted.

Non-acute co-morbidities which were recorded for the first time during a CAP hospitalisation, but only after the first episode of the hospitalisation, were not coded as 'present' in my analyses in Chapter 6. The rationale for this decision was to exclude conditions that had not yet occurred at the time patients with CAP were assessed by their GP or at hospital, as well as conditions that were present but only diagnosed after the patient was hospitalised. For some of the longer-term conditions such as cancer or diabetes, new diagnoses may have been made later in the hospitalisation after the patient had presented with CAP. Alternatively, it is possible that the diagnosis may have been known to the GP or admitting physician but not previously coded in the patient's electronic health record. Thus, exclusion of diagnoses recorded only after the first episode of the CAP hospitalisation could have resulted in some misclassification of co-morbidity status. In order for this to have substantially affected the ORs reported, a substantial number of newly recorded diagnoses would have needed to be made for patients hospitalised with CAP, during their second (or later) episode of care. If I were

to repeat the analyses, it would be interesting to assess how much difference the addition of newly recorded chronic conditions would have made.

The analyses to develop prognostic models for longer term mortality were limited to CPRD co-morbidity diagnoses only, to reflect the intended setting for these models. The co-morbidity statuses used in these models thus reflect GPs' recording practices rather than patients' full underlying health status, and will have resulted in some under-recording of patients' medical history. Again, the effect of any under-recording of co-morbidity status on the results from this analysis depend on whether the under-recording differed with respect to patients' subsequent mortality. Many of the conditions of interest require regular treatment and are thus likely to have been known to the GP. Inclusion of HES diagnoses may have improved the sensitivity somewhat, but would have resulted in a model unsuited to automated use in primary care.

8.5.2 Completeness of recording of pneumonia diagnoses across the data sources

Over the time period covered by this study, there have been important changes to the provision of primary care, which had the potential to affect the recording of pneumonia diagnoses in both CPRD and HES.

In 2004, GPs became able to opt-out of providing out-of-hours (OOH) care (broadly defined as care between 6.30pm and 8.00am, on weekends and bank holidays) which could instead be provided by a third party. The majority of general practices have since opted out with only 10% of practices providing their own OOH care in 2013/14. Over the same period, the number of cases being handled by OOH care declined from 8.6 million in 2007-8 to 5.8 million in 2013-14. The majority of this decrease was attributed to increased use of NHS Direct, a 24 hour telephone advice service staffed by nurses who provided basic health advice or directed patients with more serious illnesses to the appropriate part of the NHS, such as their GP or A&E.[201]

Of the 5.8 million OOH cases in 2013-14, 3.3 million were face-to-face consultations including 800,000 GP home-visits. The number of in-hours consultations provided by GPs that year was estimated at over 300 million.[202] While the OOH consultations make up a small proportion of the total cared for in primary care, it is important that their information is still available in primary care records. To ensure that GPs are aware

of OOH consultations, since 2005 OOH service providers have been required to meet a set of national quality requirements (NQR). The second of these is that:

> "Providers must send details of all OOH consultations (including appropriate clinical information) to the practice where the patient is registered by 8.00 a.m. the next working day. Where more than one organisation is involved in the provision of OOH services, there must be clearly agreed responsibilities in respect of the transmission of patient data."[202]

Despite the requirement of the service provider to share the information of patients treated out-of-hours, some CAP diagnoses made during OOH consultations may not have been coded into patients' GP records (and thus would not have been captured in CPRD).

Over a longer time period, trends in A&E attendance rose considerably from 14.12 million attendances in 1996/7 to 21.38 million in 2010/11.[203] The majority of this increase happened from 2004 onwards (as OOH care changed providers), and was largely due to increasing visits to minor injury units and urgent care centres rather than major A&E departments.[204] While A&E records were available for the HES-linked data from 2007 onwards, previous research has shown that the records were not complete until after 2010/11.[204, 205] Additionally, the recorded diagnoses in the A&E data were lacking in detail and would not have provided sufficient depth of coding to have been useful in this work (for example, general headings such as 'respiratory disease' are widely used rather than 'pneumonia' or 'COPD exacerbation').[Rachel Williams, CPRD, personal communication] Therefore the A&E data were not considered to be of high enough quality to include in my analyses. The quality of this coding is now starting to improve, and it may be possible to include A&E diagnoses in future work.

Within CPRD there is provision of a 'constype' field within the consultation file to record when and where the consultation (and the resulting diagnoses) occurred, such as a surgery consultation, night visit at home, A&E, or out-of-hours visit. As the location of the consultation was not initially of interest, this was not something I investigated thoroughly. In retrospect, an analysis of the 'constype' of CAP records in CPRD over time

could have provided useful detail when evaluating trends in recording of OOH and A&E care.

As a result of the changes to OOH care and trends in A&E attendance, the analyses of incidence (Chapters 4 and 5) and of risk-factors for hospitalisation (Chapter 6) may not have fully captured patients who required out-of-hours treatment or presented to A&E with pneumonia, especially if these patients were not subsequently admitted to hospital or seen by their GP. While it is not possible to know the extent of any underestimation in incidence, there are two reasons to believe it may not have been sufficiently large to change appreciably the results shown.

Firstly, the NQR for providers of OOH care to send details of the OOH care to GPs may have reduced the potential for underascertainment of the relatively small proportion of consultations via this method, provided they were coded by the GP once reported.

Secondly, while A&E and affiliated services such as minor injury units and urgent care centres accounted for a larger number of attendances than OOH care, an increasing percentage of hospitalisations for/after CAP were found to arise from A&E consultations (from 50.6% in 1998-2000, to 76.4% in 2009-10, see Chapter 6). This suggests that many of the increased A&E consultations translated into admissions.

Excluding some CAP events that were diagnosed in OOH services or in A&E may also have affected the hospitalisation risk factor analysis in Chapter 6. It is possible that patients with CAP who were seen in these settings would have been more likely to have been captured in the CPRD or HES data (and thus included in the hospitalisation analyses) if they had specific characteristics. For example, given the potential severity of CAP with increasing age and among those with co-morbid diseases, the oldest patients and those with co-morbidities could have had a higher probability of referral to their own GP for follow-up, or for hospitalisation after A&E assessment. In contrast, relatively young patients and those with fewer co-morbidities could have been simply prescribed antibiotics by the clinician they consulted. Exclusion of the latter patients from the 'non-hospitalised' group in the hospitalisation analyses may have resulted in an overestimation of the ORs for hospitalisation in younger age-groups, and an underestimation of the effects of some co-morbidities.

8.5.3 Accuracy of linkages to HES and to mortality data

The analyses in this thesis assume that the linkages between CPRD, HES and ONS are accurate. This is likely to be true for the majority of cases given the deterministic algorithm used by the trusted third party to link the datasets, which includes NHS number and other patient-identifiable data. However, information regarding the accuracy of individual patient linkage between CPRD and HES is not readily available, and it is possible that linkage errors occurred at a low level.

There are two types of linkage error which may have occurred. False matches involve a patient's CPRD records being incorrectly linked to a different patient's HES records. This would have resulted in an incorrect medical history being assigned to a CPRD patient via the HES records, in addition to incorrect dates of hospital admission and pneumonia diagnoses (if any occurred in the HES data). Any false matched HES pneumonia diagnoses may have resulted in a small overestimation of CAP incidence and levels of hospitalisation post-CAP. They may also have slightly biased the association between risk factors recorded in the patient's CPRD record and hospitalisation post-CAP, or mortality post-CAP hospital discharge. Should the CPRD patient have additionally been matched to the correct HES data, then their HES records will have included both correct and incorrect information (although this double-linkage seems extremely unlikely).

Secondly, missed matches may have occurred where a patient's CPRD records are not linked to their HES records. This would result in their having missing data on hospital admissions (including the associated person-time not at risk, and any CAPs diagnosed in hospital) and thus potential under-ascertainment of the outcome in the hospitalisation analysis, and their exclusion from the mortality analysis.

The lack of patient identifiers available to researchers makes recognition of these possible linkage errors between CPRD and HES extremely difficult. A recent analysis attempted to identify possible false-matches for children and adolescent records within stand-alone HES data in the absence of patient-level identifiers, using simultaneous admission at multiple hospitals and readmission after death to identify false matches.[206] Theoretically, this may also be possible to assess matching of CPRD to HES, although it is outside of the scope of this thesis. However, readmission after death in HES data could equally be due to incorrect coding of death in the discharge

destination or discharge method. A better understanding of the validity of these data fields may be advisable before using them as a gold standard to assess matching accuracy.

Both of these types of linkage error may also have occurred when linking CPRD to ONS records. However CPRD have recently started to provide information on this linkage including the strength of the match (as described in section 2.1.4) and the number of CPRD patients with more than one ONS date of death match. Multiple matching occurred in an extremely small minority of patients in this study (87 of the >900,000 eligible, <0.01%) and I was able to identify the most appropriate linked record in each of these cases.

Considering the very small level of false-matching found between CPRD and ONS data, it seems likely that it occurred at a very low level between CPRD and HES, and thus any false matches should not have significantly affected the results I present. While missed matches between the linked data sources may have occurred, again, it is unlikely this will have been at a high enough level to change my findings.

8.5.4   Confounding

If variables are misclassified, this can result in residual confounding. In the majority of the analyses included in this work, covariates were included as factors of interest in their own right. However, it was possible for variables to confound the effect of other factors in the model. For example, the analyses in Chapter 6 showed that the strong effects shown in minimally adjusted models for some co-morbidities were diminished after the model was additionally adjusted for other co-morbidities. It is therefore conceivable that in some cases in the analyses in Chapter 6 and Chapter 7, misclassification of covariates may have resulted in some low-level residual confounding.

8.5.5   Multiple testing

When large numbers of tests are carried out in one analysis, some of the associations which are revealed may have occurred by chance (a Type 1 error). In these situations results should be interpreted cautiously.[113] In this thesis this applies in particular to the risk factor analysis presented in Chapter 6, in which the final model included

fourteen co-morbidities, five frailty factors, and four medications/vaccinations in addition to age and sex. P-values were intentionally excluded from the results presented in this paper, because even small differences are likely to have small p-values in analyses of very large data such as these. The potential effect of multiple testing should be borne in mind when assessing the adjusted ORs for each factor in the model, and the size of effect estimates and examination of the 95% CIs were used instead to guide interpretation of the data. Multiple testing also occurred during the development of the prognostic models in Chapter 7. The p-value for selection of variables into the models was intentionally large (p<0.2) in order to include strong but uncommon predictors, as recommended when developing these scores.[115] Unlike causal models, the inclusion of variables in prognostic models due to chance (known as 'noise variables') has been shown to have only a limited effect on their predictive ability, and is not thought to be a cause for concern.[115]

8.5.6 Generalisability of the results

Patients contributing to CPRD as a whole have been shown to be broadly similar to the UK population, most recently with regards to age and sex when compared to UK Census data from 2011.[99] However, HES-linked data were only available for England, due to the different versions of HES used in Scotland, Wales and Northern Ireland. The analyses of risk factors for hospitalisation and the prognostic models for mortality were thus both restricted to using English CPRD HES-linked data. Linkage only occurs when practices consent to it, however patients with HES-linked data have been shown to be similar to those without these data.[104] Differences in health policy since the devolution of UK health services in 1999 may have led to a lack of generalisability between the results presented here and for patients in the rest of the United Kingdom. The results from the hospitalisation analysis in Chapter 7 may also have limited generalisability to older populations in other European nations, due to changes in service provision and usage which seem to have driven so much of the rising hospitalisation trend in England.

When investigating CAP hospitalisation trends, I included all hospitalisations within 28 days of a CAP diagnosis, rather than those specifically for CAP. This enabled me to additionally capture hospitalisations for other conditions which were precipitated by CAP, but as a result the risk factors identified are not necessarily specific to

hospitalisation for CAP itself (although 95% of admitting diagnoses were coded as ICD-10 Chapter X, diseases of the respiratory system).

8.6   Implications of the study findings for clinical practice and future research

8.6.1   Implications for clinical practice/health policy

As the UK population ages and progresses through the older age groups, the overall incidence of CAP in the UK will increase, necessitating greater healthcare provision. Thus, CAP will continue to present an increasing burden to the health system. Previous research on CAP from the UK has been among adults of all ages, and has lacked detailed stratification of the older population by age. The incidence analyses I present show that older individuals comprise a highly heterogeneous population, with wide ranging disease burdens that are concealed by the summarised age grouping used in many studies. These enhanced estimates of the burden of CAP in the older population provide important information for health planners as they prepare for the consequences of the ageing of the UK population.

Rising numbers of CAP events will result in further increases in hospitalisations. My thorough characterisation of the population of CAP patients at risk of hospitalisation, and the contribution of risk factors to hospitalisation trends will also help inform health provision in the future. As a consequence of increasing hospitalisations, GPs will be responsible for ever larger numbers of older patients at increased risk of mortality. The prognostic scores developed in Chapter 7 could assist GPs in identifying specific patients at heightened risk, and in future may enable targeted follow-up to ensure their co-morbidities are well managed and their underlying health is not worsening. Should external validation further confirm the scores to be an accurate clinical tool, it may be possible to build them directly into GP software systems.

Frailty and older patients' health are currently being highlighted as priority areas for primary care. As discussed above, I have shown that the present level of frailty recording in CPRD precludes its use for extensive research on this topic, which is a missed opportunity. Changes to the GP contract, including longer appointment times, may go some way toward supporting fuller recording of these aspects of patients' lives.[158] The benefits and results of enhanced recording should also be fed back to the GP

community at large. Time with patients will and should always take precedence over time spent coding, but without thorough and accurate coding the knowledge gained from these records is limited, and an opportunity lost.

8.6.2 Implications for future research

The findings presented in this thesis provide several areas for further research. One of the most important of these is the validation of the pneumonia (and LRTI) codes used in this study. A better understanding of the positive predictive value and sensitivity of these codes in both primary and secondary care data over time would allow more thorough interpretation of the results I have presented. In particular it may inform the marked divergence in CAP rates estimated from stand-alone CPRD compared to linked CPRD-HES. Currently, there is no 'gold standard' between HES and CPRD for pneumonia records, but validation of diagnoses in both data sources may help to identify one.

Combining data from several sources can be complex and time consuming, and detailed, well-designed plans for merging the data must be made. A thorough understanding of the individual data sources can result in their combined use greatly enhancing epidemiological analyses. The methods I developed to use linked-data to identify episodes of disease, person-time at risk of infection and differentiate community- from hospital-acquired infection are transferable to many other infections across the wider population. So far, my methods have been used in subsequent analyses of the incidence of community-acquired LRTI, urinary tract infections and sepsis among older adults with diabetes,[207] and LRTI, CAP and sepsis among older adults with chronic kidney disease and diabetes.[208, 209]

My detailed analysis of the risk factors for hospitalisation post-CAP revealed that increasing hospitalisations cannot be attributed to changes to patients' underlying health. What then is the reason behind the considerable increase in hospitalisations after CAP, from 57% of cases in 1998-2000 to 87% in 2009-2011? The forthcoming BTS audit of CAP diagnoses and a large ongoing validation study of CAP diagnosis in CPRD may provide some information on this topic. A more comprehensive analysis of patterns of patient care within the older population may also be warranted in order to attempt to answer this question, and to stem the increasing tendency to treat CAP in hospital.

Information on frailty factors, and more detail on patients' residential status (for example delineation between residential and nursing homes) is particularly lacking within CPRD. Supplementary linkage to social care records would further enrich the CPRD data and facilitate more thorough investigation of the contribution of these factors to older patients' health. The 'oldest old' will make up 5% of the UK population by 2035,[44] and a better understanding of the effects of frailty by this point would greatly help in caring for this population.

Further linkage to laboratory data would provide information on the causative pathogen behind these infections, enabling assessment of the role of specific pathogens in trends over time. Additionally, the effectiveness of the pneumococcal vaccine and influenza vaccine against CAP among the older population could be estimated more accurately.

It is widely agreed that prognostic models should be externally validated before being introduced into routine use. There are several other primary care databases available, such as QResearch or ResearchOne which would be suitable for external validation of the models developed in Chapter 7.[210, 211] If the risk scores were ever introduced into general use, future analyses would need to be undertaken to determine whether identification by GPs of high-risk patients after a CAP hospitalisation improved longer-term patient mortality.

8.7 Overall conclusions

Use of linked primary, secondary and mortality records provided a large and detailed study population over a 14 year period, enabling a thorough investigation of the burden and outcomes of CAP among older individuals in the UK that was not previously possible.

This thesis exploited linked data to meet several objectives. CAP episodes from both primary and secondary care were combined to produce more complete incidence estimates than those currently available in the UK. Increasing CAP hospitalisation was shown to be occurring separately to increasing CAP rates, and the linked-data enabled identification of a range of risk factors for hospitalisation, although patients' worsening underlying health status was not found to explain the hospitalisation trends. Finally, prognostic models to predict longer-term mortality post-CAP discharge were developed

to aid clinical decision making. All of these analyses of this common and serious public health problem were made possible using data from multiple linked data sources.

References

1. Lim WS, Baudouin SV, George RC, Hill AT, Jamieson C, Le Jeune I et al. BTS guidelines for the management of community acquired pneumonia in adults: update 2009. Thorax. 2009;64(Suppl 3):iii1-iii55. doi:10.1136/thx.2009.121434.

2. Society ER. European Lung White Book: Chapter 18. Acute lower respriatory infections. Sheffield: European Respiratory Society Journals; 2013.

3. Walker BR, Colledge NR, Ralston S, Penman ID, Britton R. Davidson's principles and practice of medicine. Twenty-second ed. Principles and practice of medicine. Edinburgh New York: Churchill Livingstone/Elsevier; 2014.

4. Dickson RP, Huffnagle GB. The Lung Microbiome: New Principles for Respiratory Bacteriology in Health and Disease. PLOS Pathog. 2015;11(7):e1004923. doi:10.1371/journal.ppat.1004923.

5. Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB. The Microbiome and the Respiratory Tract. Annual review of physiology. 2016;78:481-504. doi:10.1146/annurev-physiol-021115-105238.

6. Kumar P, Clark M. Kumar & Clark's clinical medicine. Eighth ed. Kumar and Clark's clinical medicine. Edinburgh, [Scotland]: Saunders; 2012.

7. Moldoveanu B, Otmishi P, Jani P, Walker J, Sarmiento X, Guardiola J et al. Inflammatory mechanisms in the lung. J Inflamm Res. 2009;2:1-11.

8. Torres A, Blasi F, Peetermans WE, Viegi G, Welte T. The aetiology and antibiotic management of community-acquired pneumonia in adults in Europe: a literature review. Eur J Clin Microbiol Infect Dis. 2014;33(7):1065-79. doi:10.1007/s10096-014-2067-1.

9. Woodhead M, Blasi F, Ewig S, Garau J, Huchon G, Ieven M et al. Guidelines for the management of adult lower respiratory tract infections - Summary. Clin Microbiol Infect. 2011;17:1-24. doi:10.1111/j.1469-0691.2011.03602.x.

10. Enne VI, Personne Y, Grgic L, Gant V, Zumla A. Aetiology of hospital-acquired pneumonia and trends in antimicrobial resistance. Curr Opin Pulm Med. 2014;20(3):252-8. doi:10.1097/mcp.0000000000000042.

11. Guidelines for the Management of Adults with Hospital-acquired, Ventilator-associated, and Healthcare-associated Pneumonia. Am J Respir Crit Care Med. 2005;171(4):388-416. doi:10.1164/rccm.200405-644ST.

12. Kollef MH, Shorr A, Tabak YP, Gupta V, Liu LZ, Johannes RS. Epidemiology and outcomes of health-care–associated pneumonia*: Results from a large us database of culture-positive pneumonia. Chest. 2005;128(6):3854-62. doi:10.1378/chest.128.6.3854.

13. Chalmers JD, Rother C, Salih W, Ewig S. Healthcare-Associated Pneumonia Does Not Accurately Identify Potentially Resistant Pathogens: A Systematic Review and Meta-Analysis. Clin Infect Dis. 2014;58(3):330-9. doi:10.1093/cid/cit734.

14. Office for National Statistics. Changes in the Older Resident Care Home Population between 2001 and 2011. London: ONS, 2014.

15. Woodhead M. New guidelines for the management of adult lower respiratory tract infections. European Respiratory Journal. 38 (6) (pp 1250-1251), 2011. Date of Publication: 01 Dec 2011.; 2011.

16. Ewig Sa, Welte Tb, Torres Ac. Is healthcare-associated pneumonia a distinct entity needing specific therapy? Curr Opin Infect Dis. 2012;25(2):166-75.

17. Metlay JP, Kapoor WN, Fine MJ. Does this patient have community-acquired pneumonia? Diagnosing pneumonia by history and physical examination. JAMA. 1997;278(17):1440-5. doi:10.1001/jama.1997.03550170070035.

18. Metlay JP, Schulz R, Li Y, et al. Influence of age on symptoms at presentation in patients with community-acquired pneumonia. Arch Intern Med. 1997;157(13):1453-9. doi:10.1001/archinte.1997.00440340089009.

19. Riquelme R, Torres A, El-Ebiary M, Mensa J, Estruch R, Ruiz M et al. Community-acquired Pneumonia in the Elderly. Am J Respir Crit Care Med. 1997;156(6):1908-14. doi:10.1164/ajrccm.156.6.9702005.

20. van Vugt SF, Verheij TJM, de Jong PA, Butler CC, Hood K, Coenen S et al. Diagnosing pneumonia in patients with acute cough: clinical judgment compared to chest radiography. Eur Respir J. 2013;42(4):1076-82. doi:10.1183/09031936.00111012.

21. Campbell SG, Murray DD, Hawass A, Urquhart D, Ackroyd-Stolarz S, Maxwell D. Agreement between emergency physician diagnosis and radiologist reports in patients discharged from an emergency department with community-acquired pneumonia. Emerg Radiol. 2005;11(4):242-6. doi:10.1007/s10140-005-0413-4.

22. Basi SK, Marrie TJ, Huang JQ, Majumdar SR. Patients admitted to hospital with suspected pneumonia and normal chest radiographs: Epidemiology, microbiology, and outcomes. Am J Med 2004;117(5):305-11. doi:http://dx.doi.org/10.1016/j.amjmed.2004.03.029.

23. Novack V, Avnon LS, Smolyakov A, Barnea R, Jotkowitz A, Schlaeffer F. Disagreement in the interpretation of chest radiographs among specialists and clinical outcomes of patients hospitalized with suspected pneumonia. Eur J Intern Med. 2006;17(1):43-7. doi:http://dx.doi.org/10.1016/j.ejim.2005.07.008.

24. Harrington J, Tariq K, Collins A, Burhan H, McDermott P. Disagreement in radiological diagnosis of community acquired pneumonia and its impact on management and mortality: Is the term 'pneumonia' outdated? Eur Respir J. 2015;46(suppl 59). doi:10.1183/13993003.congress-2015.PA2593.

25. Albaum MN, Hill LC, Murphy M, Li Y-H, Fuhrman CR, Britton CA et al. Interobserver Reliability of the Chest Radiograph in Community-Acquired Pneumonia. Chest. 1996;110(2):343-50. doi:http://dx.doi.org/10.1378/chest.110.2.343.

26. Hopstaken RM, Witbraad T, van Engelshoven JMA, Dinant GJ. Inter-observer variation in the interpretation of chest radiographs for pneumonia in community-acquired lower respiratory tract infections. Clin Radiol. 2004;59(8):743-52. doi:http://dx.doi.org/10.1016/j.crad.2004.01.011.

27. NICE Short Clinical Guidelines Technical Team. Respiratory tract infections – antibiotic prescribing. Prescribing of antibiotics for self-limiting respiratory tract infections in adults and children in primary care. In: National Institute for Health and Clinical Excellence, editor. London2008.

28. Petersen I, Johnson AM, Islam A, Duckworth G, Livermore DM, Hayward AC. Protective effect of antibiotics against serious complications of common respiratory tract infections: retrospective cohort study with the UK General Practice Research Database. BMJ. 2007;335(7627):982. doi:10.1136/bmj.39345.405243.BE.

29. Welte T, Torres A, Nathwani D. Clinical and economic burden of community-acquired pneumonia among adults in Europe. Thorax. 2012;67(1):71-9. doi:10.1136/thx.2009.129502.

30. Torres A, Blasi F, Dartois N, Akova M. Which individuals are at increased risk of pneumococcal disease and why? Impact of COPD, asthma, smoking, diabetes, and/or

chronic heart disease on community-acquired pneumonia and invasive pneumococcal disease. Thorax. 2015. doi:10.1136/thoraxjnl-2015-206780.

31. Torres A, Peetermans WE, Viegi G, Blasi F. Risk factors for community-acquired pneumonia in adults in Europe: a literature review. Thorax. 2013;68(11):1057-65. doi:10.1136/thoraxjnl-2013-204282.

32. Khan AR, Riaz M, Bin Abdulhak AA, Al-Tannir MA, Garbati MA, Erwin PJ et al. The Role of Statins in Prevention and Treatment of Community Acquired Pneumonia: A Systematic Review and Meta-Analysis. PLoS ONE. 2013;8(1):e52929. doi:10.1371/journal.pone.0052929.

33. Suissa S, Patenaude V, Lapi F, Ernst P. Inhaled corticosteroids in COPD and the risk of serious pneumonia. Thorax. 2013;68(11):1029-36. doi:10.1136/thoraxjnl-2012-202872.

34. Macfarlane J, Holmes W, Gard P, Macfarlane R, Rose D, Weston V et al. Prospective study of the incidence, aetiology and outcome of adult lower respiratory tract illness in the community. Thorax. 2001;56(2):109-14. doi:10.1136/thorax.56.2.109.

35. Macfarlane JT, Colville A, Guion A, Macfarlane RM, Rose DH. Prospective study of aetiology and outcome of adult lower-respiratory-tract infections in the community. Lancet. 1993;341(8844):511-4.

36. Castle SC. Clinical Relevance of Age-Related Immune Dysfunction. Clin Infect Dis. 2000;31(2):578-85. doi:10.1086/313947.

37. Office for National Statistics. General lifestyle survey overview: A report on the 2010 general lifestyle survey. London: ONS, 2012.

38. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet. 2012;380(9836):37-43. doi:10.1016/S0140-6736(12)60240-2.

39. Bouillon K, Kivimaki M, Hamer M, Sabia S, Fransson E, Singh-Manoux A et al. Measures of frailty in population-based studies: an overview. BMC Geriatr. 2013;13(1):64.

40. Gill TM, Gahbauer EA, Allore HG, Han L. Transitions between frailty states among community-living older persons. Arch Intern Med. 2006;166(4):418-23. doi:10.1001/archinte.166.4.418.

41. Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. Lancet. 2013;381(9868):752-62. doi:10.1016/S0140-6736(12)62167-9.

42. Collard RM, Boter H, Schoevers RA, Oude Voshaar RC. Prevalence of Frailty in Community-Dwelling Older Persons: A Systematic Review. J Am Geriatr Soc. 2012;60(8):1487-92. doi:10.1111/j.1532-5415.2012.04054.x.

43. Bennett JE, Li G, Foreman K, Best N, Kontis V, Pearson C et al. The future of life expectancy and life expectancy inequalities in England and Wales: Bayesian spatiotemporal forecasting. Lancet. 2015;386(9989):163-70. doi:10.1016/S0140-6736(15)60296-3.

44. Office for National Statistics. Statistical Bulletin - Older People's Day 2011. London: ONS, 2011.

45. Office for National Statistics. Summary: UK Population Projected to Reach 70 Milllion by Mid-2027. London: ONS, 2011.

46. Muller-Pebody B, Crowcroft NS, Zambon MC, Edmunds WJ. Modelling hospital admissions for lower respiratory tract infections in the elderly in England. Epidemiol Infect. 2006;134(6):1150-7. doi:10.1017/S0950268806006376.

47. Health Protection Agency. Invasive pneumonococcal infection, England and Wales: 2000. CDR Weekly. 2003;13(21):7-10.

48. Public Health England, Department of Health. Pneumococcal (Chapter 25). Immunisation Against Infectious Disease. (The Green Book) 2013.

49. Public Health England. Pneumococcal Polysaccharide Vaccination (PPV) Uptake. http://webarchive.nationalarchives.gov.uk/20140714084352/http:/www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/Pneumococcal/PPVUptake/. Accessed 19/03/2015.

50. Begum F, Pebody R. Pneumococcal Polysaccharide Vaccine (PPV) uptake report Department of Health2012.

51. Andrews NJ, Waight PA, George RC, Slack MPE, Miller E. Impact and effectiveness of 23-valent pneumococcal polysaccharide vaccine against invasive pneumococcal disease in the elderly in England and Wales. Vaccine. 2012;30(48):6802-8. doi:http://dx.doi.org/10.1016/j.vaccine.2012.09.019.

52. Moberley SA, Holden J, Tatham DP, Andrews RM. Vaccines for preventing pneumococcal infection in adults. Cochrane Database Syst Rev. 2008(1):Cd000422. doi:10.1002/14651858.CD000422.pub2.

53. Public Health England, Department of Health. Influenza (Chapter 19) Immunisation Against Infectious Disease. (The Green Book) 2013.

54. Fleming DM, Andrews NJ, Ellis JS, Bermingham A, Sebastianpillai P, Elliot AJ et al. Estimating influenza vaccine effectiveness using routinely collected laboratory data. Journal of epidemiology and community health. 2010;64(12):1062-7. doi:10.1136/jech.2009.093450.

55. Jefferson T, Di Pietrantonj C, Al-Ansary LA, Ferroni E, Thorning S, Thomas RE. Vaccines for preventing influenza in the elderly. Cochrane Database Syst Rev. 2010(2):Cd004876. doi:10.1002/14651858.CD004876.pub3.

56. Pneumonia: Diagnosis and management of community- and hospital-acquired pneumonia in adults, (2014).

57. Creer DD, Dilworth JP, Gillespie SH, Johnston AR, Johnston SL, Ling C et al. Aetiological role of viral and bacterial infections in acute adult lower respiratory tract infection (LRTI) in primary care. Thorax. 2006;61(1):75-9. doi:10.1136/thx.2004.027441.

58. Cals JW, Hood K, Aaftink N, Hopstaken RM, Francis NA, Dinant G-J et al. Predictors of patient-initiated reconsultation for lower respiratory tract infections in general practice. Br J Gen Pract. 2009;59(567):761-4. doi:10.3399/bjgp09X472656.

59. Holmes WF, Macfarlane JT, Macfarlane RM, Lewis S. The influence of antibiotics and other factors on reconsultation for acute lower respiratory tract illness in primary care. Br J Gen Pract. 1997;47(425):815-8.

60. Monte SV, Paolini NM, Slazak EM, Schentag JJ, Paladino JA. Costs of treating lower respiratory tract infections. Am J Manage Care. 2008;14(4):190-6.

61. Trotter CL, Stuart JM, George R, Miller E. Increasing hospital admissions for pneumonia, England. Emerg Infect Dis. 2008;14(5):727-33. doi:10.3201/eid1405.071011.

62. Bardsley M, Blunt I, Davies S, Dixon J. Is secondary preventive care improving? Observational study of 10-year trends in emergency admissions for conditions amenable to ambulatory care. BMJ Open. 2013;3(1). doi:10.1136/bmjopen-2012-002007.

63. Blunt I. Focus on preventable admissions: Trends in emergency admission for ambulatory care sensitive conditions, 2001 to 2013. London: Nuffield Trust October 2013.

64. Thomsen RW, Riis A, NØRgaard M, Jacobsen J, Christensen S, McDonald CJ et al. Rising incidence and persistently high mortality of hospitalized pneumonia: a 10-year population-based study in Denmark. J Intern Med. 2006;259(4):410-7. doi:10.1111/j.1365-2796.2006.01629.x.

65. Søgaard M, Nielsen RB, Schønheyder HC, Nørgaard M, Thomsen RW. Nationwide trends in pneumonia hospitalization rates and mortality, Denmark 1997–2011. Respir Med. 2014;108(8):1214–22. doi:http://dx.doi.org/10.1016/j.rmed.2014.05.004.

66. Van Gageldonk-Lafeber A, Bogaerts M, Verheij R, Van der Sande M. Time trends in primary-care morbidity, hospitalization and mortality due to pneumonia. Epidemiol Infect. 2009;137(10):1472-8. doi:doi:10.1017/S0950268809002258.

67. Froes F, Diniz A, Mesquita M, Serrado M, Nunes B. Hospital admissions of adults with community-acquired pneumonia in Portugal between 2000 and 2009. Eur Respir J. 2013;41(5):1141-6. doi:10.1183/09031936.00216711.

68. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-83.

69. Grijalva CG, Zhu Y, Williams DJ, et al. Association between hospitalization with community-acquired laboratory-confirmed influenza pneumonia and prior receipt of influenza vaccination. JAMA. 2015;314(14):1488-97. doi:10.1001/jama.2015.12160.

70. Gross PA, Hermogenes AW, Sacks HS, Lau J, Levandowski RA. The Efficacy of Influenza Vaccine in Elderly PersonsA Meta-Analysis and Review of the Literature. Ann Intern Med. 1995;123(7):518-27. doi:10.7326/0003-4819-123-7-199510010-00008.

71. Puig-Barberà J, Díez-Domingo J, Varea ÁB, Chavarri GS, Rodrigo JAL, Hoyos SP et al. Effectiveness of MF59™-adjuvanted subunit influenza vaccine in preventing hospitalisations for cardiovascular disease, cerebrovascular disease and pneumonia in the elderly. Vaccine. 2007;25(42):7313-21. doi:http://dx.doi.org/10.1016/j.vaccine.2007.08.039.

72. Jackson LA, Neuzil KM, Yu O, Benson P, Barlow WE, Adams AL et al. Effectiveness of Pneumococcal Polysaccharide Vaccine in Older Adults. New Engl J Med. 2003;348(18):1747-55. doi:doi:10.1056/NEJMoa022678.

73. Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J et al. Frailty in Older Adults: Evidence for a Phenotype. J Gerontol A Biol Sci Med Sci. 2001;56(3):M146-M57. doi:10.1093/gerona/56.3.M146.

74. Searle S, Mitnitski A, Gahbauer E, Gill T, Rockwood K. A standard procedure for creating a frailty index. BMC Geriatr. 2008;8(1):24.

75. Rockwood K, Mitnitski A. Limits to deficit accumulation in elderly people. Mechanisms of ageing and development. 2006;127(5):494-6. doi:10.1016/j.mad.2006.01.002.

76. Heron M. Deaths: Leading causes for 2012. National vital statistics reports. 2015;64(10).

77. Office for National Statistics. Mortality Statistics: Deaths Registered in England and Wales (Series DR), 2013 Release2013.

78. Brock A, Griffiths C, Rooney C. The impact of introducing ICD-10 on analysis of respiratory mortality trends in England and Wales. Health Statistics Quarterly. 2006(29):9-17.

79. Office for National Statistics. Guidance for doctors completing Medical Certificates of Cause of Death in England and Wales. In: Group DCA, editor. London2010.

80. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax. 2003;58(5):377-82.

81. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE et al. A Prediction Rule to Identify Low-Risk Patients with Community-Acquired Pneumonia. N Engl J Med. 1997;336(4):243-50. doi:doi:10.1056/NEJM199701233360402.

82. Neill AM, Martin IR, Weir R, Anderson R, Chereshsky A, Epton MJ et al. Community acquired pneumonia: aetiology and usefulness of severity criteria on admission. Thorax. 1996;51(10):1010-6. doi:10.1136/thx.51.10.1010.

83. McNally M, Curtain J, O'Brien KK, Dimitrov BD, Fahey T. Validity of British Thoracic Society guidance (the CRB-65 rule) for predicting the severity of pneumonia in general practice: systematic review and meta-analysis. Br J Gen Pract. 2010;60(579):e423-33. doi:10.3399/bjgp10X532422.

84. Bont J, Hak E, Hoes AW, Macfarlane JT, Verheij TJ. Predicting death in elderly patients with community-acquired pneumonia: a prospective validation study reevaluating the CRB-65 severity assessment tool. Arch Intern Med. 2008;168(13):1465-8. doi:10.1001/archinte.168.13.1465.

85. Bont J, Hak E, Hoes AW, Schipper M, Schellevis FG, Verheij TJ. A prediction rule for elderly primary-care patients with lower respiratory tract infections. Eur Respir J. 2007;29(5):969-75. doi:10.1183/09031936.00129706.

86. van de Nadort C, Smeets HM, Bont J, Zuithoff NP, Hak E, Verheij TJ. Prognosis of primary care patients aged 80 years and older with lower respiratory tract infection. Br J Gen Pract. 2009;59(561):e110-5. doi:10.3399/bjgp09X420239.

87. Hak E, Wei F, Nordin J, Mullooly J, Poblete S, Nichol KL. Development and validation of a clinical prediction rule for hospitalization due to pneumonia or influenza or death during influenza epidemics among community-dwelling elderly persons. J Infect Dis. 2004;189(3):450-8. doi:JID30157 [pii] 10.1086/381165.

88. Kaplan V, Clermont G, Griffin MF, Kasal J, Watson RS, Linde-Zwirble WT et al. Pneumonia: still the old man's friend? Arch Intern Med. 2003;163(3):317-23. doi:10.1001/archinte.163.3.317.

89. Rozzini R, Trabucchi M. Pneumonia and mortality beyond hospital discharge in elderly patients. Chest. 2011;139(2):473-4; author reply 4-5.

90. Yende S, Angus DC, Ali IS, Somes G, Newman AB, Bauer D et al. Influence of comorbid conditions on long-term mortality after pneumonia in older people. J Am Geriatr Soc. 2007;55(4):518-25. doi:10.1111/j.1532-5415.2007.01100.x.

91. Koivula I, Stén M, Mäkelä P. Prognosis after community-acquired pneumonia in the elderly: A population-based 12-year follow-up study. Arch Intern Med. 1999;159(14):1550-5. doi:10.1001/archinte.159.14.1550.

92. Bruns AHW, Oosterheert JJ, Cucciolillo MC, El Moussaoui R, Groenwold RHH, Prins JM et al. Cause-specific long-term mortality rates in patients recovered from community-acquired pneumonia as compared with the general Dutch population. Clin Microbiol Infect. 2011;17(5):763-8. doi:10.1111/j.1469-0691.2010.03296.x.

93. Restrepo MI, Faverio P, Anzueto A. Long-term prognosis in community-acquired pneumonia. Curr Opin Infect Dis. 2013;26(2):151-8. doi:10.1097/QCO.0b013e32835ebc6d.

94. Yende S, D'Angelo G, Kellum JA, Weissfeld L, Fine J, Welch RD et al. Inflammatory Markers at Hospital Discharge Predict Subsequent Mortality after Pneumonia and Sepsis. Am J Respir Crit Care Med. 2008;177(11):1242-7. doi:10.1164/rccm.200712-1777OC.

95. Singanayagam A, Singanayagam A, Elder DHJ, Chalmers JD. Is community-acquired pneumonia an independent risk factor for cardiovascular disease? Eur Respir J. 2012;39(1):187-96. doi:10.1183/09031936.00049111.

96. Corrales-Medina VF, Musher DM, Shachkina S, Chirinos JA. Acute pneumonia and the cardiovascular system. Lancet. 2013;381(9865):496-505. doi:10.1016/S0140-6736(12)61266-5.

97. Millett ERC, Quint JK, Smeeth L, Daniel RM, Thomas SL. Incidence of Community-Acquired Lower Respiratory Tract Infections and Pneumonia among Older Adults in the United Kingdom: A Population-Based Study. PLoS ONE. 2013;8(9):e75131. doi:10.1371/journal.pone.0075131.

98. The Clinical Practice Research Datalink. http://www.cprd.com/home/. Accessed 16/02/2015.

99. Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol. 2015. doi:10.1093/ije/dyv098.

100. Analysis: Making GP data count. GP magazine. 2013. http://www.gponline.com/analysis-making-gp-data-count/article/1215294. Accessed 27/12/2015.

101. HES data dictionary, clinical coding. http://www.datadictionary.nhs.uk/version2/web_site_content/pages/codes/clinical_coding/clinical_coding_introduction.asp?shownav=0.

102. Noble M, McLennan D, Wilkinson K, Whitworth A, Dibben C, Barnes H. English Indices of Deprivation 2007. Communities and Local Government: London2008.

103. Office for National Statistics. Super Output Area (SOA). http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html. Accessed 16/02/2015.

104. Gallagher AM, Puri S, Van Staa TP. Linkage of the general practice research database (GPRD) with other data sources. Pharmacoepidemiol Drug Saf. 2011;20:S230-S1.

105. García Rodríguez LA, Pérez Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. Br J Clin Pharmacol. 1998;45(5):419-25. doi:10.1046/j.1365-2125.1998.00701.x.

106. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol. 2010;69(1):4-14. doi:10.1111/j.1365-2125.2009.03537.x.

107. National Institute for Health and Clinical Excellence. About the Quality Outcomes Framework. http://www.nice.org.uk/aboutnice/qof/qof.jsp. Accessed 23/05/2014.

108. Roland M. Linking Physicians' Pay to the Quality of Care — A Major Experiment in the United Kingdom. New Engl J Med. 2004;351(14):1448-54. doi:doi:10.1056/NEJMhpr041294.

109. Doran T, Kontopantelis E, Valderas JM, Campbell S, Roland M, Salisbury C et al. Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK Quality and Outcomes Framework. BMJ. 2011;342. doi:10.1136/bmj.d3590.

110. Lewis JD, Brensinger C. Agreement between GPRD smoking data: a survey of general practitioners and a population-based survey. Pharmacoepidemiol Drug Saf. 2004;13(7):437-41. doi:10.1002/pds.902.

111. Booth HP, Prevost AT, Gulliford MC. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. Pharmacoepidemiol Drug Saf. 2013;22(12):1357-61.

112. Commission A. Improving data quality in the NHS. Annual report on the PbR assurance programme. London: Audit Commission2010.

113. Kirkwood B, Sterne J. Essential Medical Statistics. Second ed. Blackwell.

114. Szmaragd C, Clarke P, Steele F. Subject specific and population average models for binary longitudinal data: a tutorial. 2013. 2013;4(2):19. doi:10.14301/llcs.v4i2.249.

115. Steyerberg EW. Clinical prediction models. A practical approach to development, validation and updating. Statistics for Biology and Health. Springer; 2009.

116. Victora CG, Huttly SR, Fuchs SC, Olinto MT. The role of conceptual frameworks in epidemiological analysis: a hierarchical approach. Int J Epidemiol. 1997;26(1):224-7.

117. Lewis JD, Bilker WB, Weinstein RB, Strom BL. The relationship between time since registration and measured incidence rates in the General Practice Research Database. Pharmacoepidemiol Drug Saf. 2005;14(7):443-51. doi:10.1002/pds.1115.

118. Wang Y, Hunt K, Nazareth I, Freemantle N, Petersen I. Do men consult less than women? An analysis of routinely collected UK general practice data. BMJ Open. 2013;3(8). doi:10.1136/bmjopen-2013-003320.

119. Fleming D. Weekly Returns Service of the Royal College of General Practitioners. Commun Dis Public Health. 1999;2(2):96-100.

120. Vila-Corcoles A, Ochoa-Gondar O, Rodriguez-Blanco T, Raga-Luria X, Gomez-Bertomeu F. Epidemiology of community-acquired pneumonia in older adults: A population-based study. Respir Med. 2009;103(2):309-16. doi:http://dx.doi.org/10.1016/j.rmed.2008.08.006.

121. RCGP. Annual Report of the RCGP Research & Surveillance Centre2005 - 2010.

122. Bertsias A, Tsiligianni IG, Duijker G, Siafakas N, Lionis C, Cretan CAPRG. Studying the burden of community-acquired pneumonia in adults aged >50 years in primary health care: an observational study in rural Crete, Greece. NPJ Prim Care Respir Med. 2014;24:14017.

123. Patrzalek M, Gorynski P, Albrecht P. Indirect population impact of universal PCV7 vaccination of children in a 2 + 1 schedule on the incidence of pneumonia morbidity in Kielce, Poland. Eur J Clin Microbiol Infect Dis. 2012;31(11):3023-8.

124. Capelastegui A, Espana PP, Bilbao A, Gamazo J, Medel F, Salgado J et al. Study of community-acquired pneumonia: incidence, patterns of care, and outcomes in primary and hospital care. J Infect. 2010;61(5):364-71.

125. Sicras-Mainar A, Ibanez-Nolla J, Cifuentes I, Guijarro P, Navarro-Artieda R, Aguilar L. Retrospective epidemiological study for the characterization of community- acquired pneumonia and pneumococcal pneumonia in adults in a well-defined area of Badalona (Barcelona, Spain). BMC Infect Dis. 2012;12:283.

126. Almirall J, Bolibar I, Vidal J, Sauca G, Coll P, Niklasson B et al. Epidemiology of community-acquired pneumonia in adults: a population-based study. Eur Respir J. 2000;15(4):757-63.

127. Jokinen C, Heiskanen L, Juvonen H, Kallinen S, Karkola K, Korppi M et al. Incidence of Community-Acquired Pneumonia in the Population of Four Municipalities in Eastern Finland. Am J Epidemiol. 1993;137(9):977-88.

128. Myles PR, McKeever TM, Pogson Z, Smith CJ, Hubbard RB. The incidence of pneumonia using data from a computerized general practice database. Epidemiol Infect. 2009;137(5):709-16. doi:10.1017/S0950268808001428.

129. Tichopad A, Roberts C, Gembula I, Hajek P, Skoczynska A, Hryniewicz W et al. Clinical and economic burden of community-acquired pneumonia among adults in the Czech Republic, Hungary, Poland and Slovakia. PLoS ONE. 2013;8(8):e71375.

130. Viegi G, Pistelli R, Cazzola M, Falcone F, Cerveri I, Rossi A et al. Epidemiological survey on incidence and treatment of community acquired pneumonia in Italy. Respir Med. 2006;100(1):46-55. doi:http://dx.doi.org/10.1016/j.rmed.2005.04.013.

131. Hak E, Rovers M, Kuyvenhoven M, Schellevis F, Verheij T. Incidence of GP-diagnosed respiratory tract infections according to age, gender and high-risk co-morbidity: the Second Dutch National Survey of General Practice. Fam Pract. 2006;23(3):291-4. doi:10.1093/fampra/cmi121.

132. Voordouw BC, Sturkenboom MC, Dieleman JP, Stijnen T, van der Lei J, Stricker BH. Annual influenza vaccination in community-dwelling elderly individuals and the risk of lower respiratory tract infections or pneumonia. Arch Intern Med. 2006;166(18):1980-5.

133. Sliedrecht A, den Elzen WPJ, Verheij TJM, Westendorp RGJ, Gussekloo J. Incidence and predictive factors of lower respiratory tract infections among the very elderly in the general population. The Leiden 85-plus Study. Thorax. 2008;63(9):817-22. doi:10.1136/thx.2007.093013.

134. Ochoa-Gondar O, Vila-Corcoles A, de Diego C, Arija V, Maxenchs M, Grive M et al. The burden of community-acquired pneumonia in the elderly: the Spanish EVAN-65 study. BMC Public Health. 2008;8:222. doi:10.1186/1471-2458-8-222.

135. Vila-Corcoles A, Ochoa-Gondar O, Llor C, Hospital I, Rodriguez T, Gomez A. Protective effect of pneumococcal vaccine against death by pneumonia in elderly subjects. Eur Respir J. 2005;26(6):1086-91.

136. Vila-Corcoles A, Ochoa-Gondar O, Hospital I, Ansa X, Vilanova A, Rodriguez T et al. Protective effects of the 23-valent pneumococcal polysaccharide vaccine in the elderly population: the EVAN-65 study. Clin Infect Dis. 2006;43(7):860-8.

137. Syrjala H, Broas M, Suramo I, Ojala A, Lahde S. High-resolution computed tomography for the diagnosis of community-acquired pneumonia. Clin Infect Dis. 1998;27(2):358-63.

138. Self WH, Courtney DM, McNaughton CD, Wunderink RG, Kline JA. High Discordance of Chest X-ray and CT for Detection of Pulmonary Opacities in ED Patients: Implications for Diagnosing Pneumonia. Am J Emerg Med. 2013;31(2):401-5. doi:10.1016/j.ajem.2012.08.041.

139. Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. BMJ. 2013;346. doi:10.1136/bmj.f2350.

140. Crooks C, Card T, West J. Defining upper gastrointestinal bleeding from linked primary and secondary care data and the effect on occurrence and 28 day mortality. BMC Health Serv Res. 2012;12(1):392.

141. Morgan CL, Currie CJ, Stott NCH, Smithers M, Butler CC, Peters JR. Estimating the prevalence of diagnosed diabetes in a health district of Wales: the importance of using primary and secondary care sources of ascertainment with adjustment for death and migration. Diabet Med. 2000;17(2):141-5. doi:10.1046/j.1464-5491.2000.00221.x.

142. Vollmer WM, O'Connor EA, Heumann M, Frazier EA, Breen V, Villnave J et al. Searching multiple clinical information systems for longer time periods found more prevalent cases of asthma. J Clin Epidemiol. 2004;57(4):392-7. doi:http://dx.doi.org/10.1016/j.jclinepi.2003.08.014.

143. Welte T. Risk factors and severity scores in hospitalized patients with community-acquired pneumonia: prediction of severity and mortality. Eur J Clin Microbiol Infect Dis. 2012;31(1):33-47. doi:10.1007/s10096-011-1272-4.

144. Juthani-Mehta M, De Rekeneire N, Allore H, Chen S, O'Leary JR, Bauer DC et al. Modifiable risk factors for pneumonia requiring hospitalization of community-dwelling older adults: the Health, Aging, and Body Composition Study. J Am Geriatr Soc. 2013;61(7):1111-8.

145. Loeb M, Neupane B, Walter SD, Hanning R, Carusone SC, Lewis D et al. Environmental risk factors for community-acquired pneumonia hospitalization in older adults. J Am Geriatr Soc. 2009;57(6):1036-40.

146. Yende S, Alvarez K, Loehr L, Folsom AR, Newman AB, Weissfeld LA et al. Epidemiology and long-term clinical and biologic risk factors for pneumonia in community-dwelling older americans: Analysis of three cohorts. Chest. 2013;144(3):1008-17. doi:10.1378/chest.12-2818.

147. O'Meara ES, White M, Siscovick DS, Lyles MF, Kuller LH. Hospitalization for pneumonia in the Cardiovascular Health Study: incidence, mortality, and influence on longer-term survival. J Am Geriatr Soc. 2005;53(7):1108-16. doi:10.1111/j.1532-5415.2005.53352.x.

148. Crooks CJ, West J, Card TR. A comparison of the recording of comorbidity in primary and secondary care by using the Charlson Index to predict short-term and long-term survival in a routine linked data cohort. BMJ Open. 2015;5(6). doi:10.1136/bmjopen-2015-007974.

149. Ruhnke GW, Coca-Perraillon M, Kitch BT, Cutler DM. Marked Reduction in 30-Day Mortality Among Elderly Patients with Community-acquired Pneumonia. Am J Med. 2011;124(2):171-8.e1. doi:10.1016/j.amjmed.2010.08.019.

150. Sundararajan V, Quan H, Halfon P, Fushimi K, Luthi JC, Burnand B et al. Cross-national comparative performance of three versions of the ICD-10 Charlson index. Med Care. 2007;45(12):1210-5. doi:10.1097/MLR.0b013e3181484347.

151. Smeeth L, Thomas SL, Hall AJ, Hubbard R, Farrington P, Vallance P. Risk of myocardial infarction and stroke after acute infection or vaccination. N Engl J Med. 2004;351(25):2611-8.

152. Rubenstein LZ. MSD Manual: Falls in the Elderly. Kenilworth, NJ, USA 2013.

153. Rockwood K, Song X, MacKnight C, Bergman H, Hogan DB, McDowell I et al. A global clinical measure of fitness and frailty in elderly people. Can Med Assoc J. 2005;173(5):489-95. doi:10.1503/cmaj.050051.

154. Nyberg L, Gustafson Y, Berggren D, Brannstrom B, Bucht G. Falls leading to femoral neck fractures in lucid older people. J Am Geriatr Soc. 1996;44(2):156-60.

155. Jackson ML, Nelson JC, Jackson LA. Risk factors for community-acquired pneumonia in immunocompetent seniors. J Am Geriatr Soc. 2009;57(5):882-8. doi:10.1111/j.1532-5415.2009.02223.x.

156. Kornum JB, Nørgaard M, Dethlefsen C, Due KM, Thomsen RW, Tjønneland A et al. Obesity and risk of subsequent hospitalisation with pneumonia. Eur Respir J. 2010;36(6):1330-6. doi:10.1183/09031936.00184209.

157. Clegg A, Bates C, Young J, Ryan R, Nichols L, Ann Teale E et al. Development and validation of an electronic frailty index using routine primary care electronic health record data. Age Ageing. 2016;45(3):353-60. doi:10.1093/ageing/afw039.

158. Changes to the General Medical Services contract for 2014/15 for England. http://www.nhsemployers.org/PayAndContracts/GeneralMedicalServicesContract/GMSContractChanges/Pages/Contractchanges201415.aspx#4. Accessed 27/05/2014.

159. Dixon WG, Abrahamowicz M, Beauchamp M-E, Ray DW, Bernatsky S, Suissa S et al. Immediate and delayed impact of oral glucocorticoid therapy on risk of serious infection

in older patients with rheumatoid arthritis: a nested case–control analysis. Ann Rheum Dis. 2012. doi:10.1136/annrheumdis-2011-200702.

160. Vinogradova Y, Hippisley-Cox J, Coupland C. Identification of new risk factors for pneumonia: population-based case-control study. Br J Gen Pract. 2009;59(567):e329-38. doi:10.3399/bjgp09X472629.

161. Koivula I, Sten M, Makela PH. Risk factors for pneumonia in the elderly. Am J Med.96(4):313-20. doi:10.1016/0002-9343(94)90060-4.

162. Ferguson J, Bauld L, Chesterman J, Judge K. The English smoking treatment services: one-year outcomes. Addiction. 2005;100:59-69. doi:10.1111/j.1360-0443.2005.01028.x.

163. Web of Science citation database entry for Charlson score paper. https://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=4&SID=T1yv1gsHXBTzbyCsc32&page=1&doc=1.

164. Chollet J, Saragoussi D, Clay E, François C. A Clinical Research Practice Datalink Analysis of Antidepressant Treatment Patterns and Health Care Costs in Generalized Anxiety Disorder. Value in Health. 2013;16(8):1133-9. doi:10.1016/j.jval.2013.09.001.

165. Walker AJ, West J, Card TR, Humes DJ, Grainge MJ. Variation in the risk of venous thromboembolism in people with colorectal cancer: a population-based cohort study from England. J Thromb Haemost. 2014;12(5):641-9. doi:10.1111/jth.12533.

166. Moons KGM, Harrell FE, Steyerberg EW. Should scoring rules be based on odds ratios or regression coefficients? J Clin Epidemiol. 2002;55(10):1054-5. doi:http://dx.doi.org/10.1016/S0895-4356(02)00453-5.

167. Mehta HB, Mehta V, Girman CJ, Adhikari D, Johnson ML. Regression coefficient–based scoring system should be used to assign weights to the risk index. J Clin Epidemiol. 2016;79:22-8. doi:http://dx.doi.org/10.1016/j.jclinepi.2016.03.031.

168. Bottle A, Aylin P. Comorbidity scores for administrative data benefited from adaptation to local coding and diagnostic practices. Journal of Clinical Epidemiology. 2011;64(12):1426-33. doi:https://doi.org/10.1016/j.jclinepi.2011.04.004.

169. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P et al. Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. American Journal of Epidemiology. 2011;173(6):676-82. doi:10.1093/aje/kwq433.

170. BMA. Guide to avoiding unplanned admissions. http://bma.org.uk/practical-support-at-work/contracts/gp-contracts-and-funding/general-practice-funding/unplanned-admissions/contact-patient-following-discharge-from-hospital. Accessed 04/02/2015.

171. BMA. Named GP for patients aged 75 years and older. http://bma.org.uk/practical-support-at-work/contracts/gp-contracts-and-funding/general-practice-funding/introduction/gp-contract-faqs/gp-contract-faqs-named-gp. Accessed 04/02/2015.

172. QRISK score availability in EMIS system. https://www.emishealth.com/products/enterprise-search-and-reports/. Accessed 16/07/2015.

173. Carriere KC, Jin Y, Marrie TJ, Predy G, Johnson DH. Outcomes and Costs Among Seniors Requiring Hospitalization for Community-Acquired Pneumonia in Alberta. J Am Geriatr Soc. 2004;52(1):31-8. doi:10.1111/j.1532-5415.2004.52007.x.

174. Adamuz J, Viasus D, Jimenez-Martinez E, Isla P, Garcia-Vidal C, Dorca J et al. Incidence, timing and risk factors associated with 1-year mortality after hospitalization for community-acquired pneumonia. J Infect. 2014;68(6):534-41.

175. Guertler C, Wirz B, Christ-Crain M, Zimmerli W, Mueller B, Schuetz P. Inflammatory responses predict long-term mortality risk in community-acquired pneumonia. Eur Respir J. 2011;37(6):1439-46. doi:10.1183/09031936.00121510.

176. Waterer GW, Kessler LA, Wunderink RG. Medium-Term Survival after Hospitalization with Community-Acquired Pneumonia. Am J Respir Crit Care Med. 2004;169(8):910-4. doi:10.1164/rccm.200310-1448OC.

177. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009;338. doi:10.1136/bmj.b375.

178. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: Developing a prognostic model. BMJ. 2009;338. doi:10.1136/bmj.b604.

179. Harrell FE, Jr., Lee KL, Matchar DB, Reichert TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. Cancer treatment reports. 1985;69(10):1071-77.

180. Royston P, Altman D. External validation of a Cox prognostic model: principles and methods. BMC Med Res Methodol. 2013;13(1):33.

181. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128-38. doi:10.1097/EDE.0b013e3181c30fb2

00001648-201001000-00022 [pii].

182. Steyerberg EW, Harrell Jr FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774-81. doi:http://dx.doi.org/10.1016/S0895-4356(01)00341-9.

183. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15(4):361-87. doi:10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

184. Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW et al. Reporting and Methods in Clinical Prediction Research: A Systematic Review. PLoS Med. 2012;9(5):e1001221. doi:10.1371/journal.pmed.1001221.

185. Lee SS, Liu H. Cardiovascular determinants of survival in cirrhosis. Gut. 2007;56(6):746-8. doi:10.1136/gut.2006.112169.

186. Fitzpatrick ME, Sethi S, Daley CL, Ray P, Beck JM, Gingo MR. Infections in "Noninfectious" Lung Diseases. Ann Am Thorac Soc. 2014;11(Suppl 4):S221-S6. doi:10.1513/AnnalsATS.201401-041PL.

187. Bertoni AG, Saydah S, Brancati FL. Diabetes and the Risk of Infection-Related Mortality in the U.S. Diabetes Care. 2001;24(6):1044-9. doi:10.2337/diacare.24.6.1044.

188. Mortensen EM, Coley CM, Singer DE, Marrie TJ, Obrosky DS, Kapoor WN et al. Causes of death for patients with community-acquired pneumonia: results from the Pneumonia Patient Outcomes Research Team cohort study. Arch Intern Med. 2002;162(9):1059-64. doi:ioi10325 [pii].

189. Schünemann HJ, Mustafa R, Brozek J, Santesso N, Alonso-Coello P, Guyatt G et al. GRADE Guidelines: 16. GRADE evidence to decision frameworks for tests in clinical practice and public health. Journal of Clinical Epidemiology. 2016;76:89-98. doi:http://dx.doi.org/10.1016/j.jclinepi.2016.01.032.

190. Ross S, Curry N, Goodwin N. Case management What it is and how it can best be implemented. London: The King's Fund2011.

191. Stokes J, Panagioti M, Alam R, Checkland K, Cheraghi-Sohi S, Bower P. Effectiveness of Case Management for 'At Risk' Patients in Primary Care: A Systematic Review and Meta-Analysis. PLoS ONE. 2015;10(7):e0132340. doi:10.1371/journal.pone.0132340.

192. Leening MG, Vedder MM, Witteman JM, Pencina MJ, Steyerberg EW. Net reclassification improvement: Computation, interpretation, and controversies. A literature review and clinician's guide. Ann Intern Med. 2014;160(2):122-31. doi:10.7326/M13-1522.

193. QRisk2 - Explaining annual updates. https://www.qrisk.org/. Accessed 01/04/2017.

194. Fortuny J, Kaye J, Margulis A, Plana E, Calingaert B, Perez-Gutthann S et al. 899: Validity of Cancer Diagnoses in General Practitioner Medical Records. Pharmacoepidemiol Drug Saf. 2015;24:1-587. doi:10.1002/pds.3838.

195. Dixon A, Mays N, Jones L. Understanding New Labour's market reforms of the English NHS. London: The Kings Fund; 2011.

196. Department of Health Payment by Results team. A simple guide to Paymeny by Results. Leeds2012.

197. Burns EM, Rigby E, Mamidanna R, Bottle A, Aylin P, Ziprin P et al. Systematic review of discharge coding accuracy. J Public Health. 2012;34(1):138-48. doi:10.1093/pubmed/fdr054.

198. Metersky ML, Waterer G, Nsa W, Bratzler DW. Predictors of in-hospital vs postdischarge mortality in pneumonia. Chest. 2012;142(2):476-81. doi:10.1378/chest.11-2393.

199. Today PC. New targets for hospital discharge letters. http://primarycaretoday.co.uk/news/new-targets-for-hospital-discharge-letters. Accessed 12/10/2015.

200. HES data dictionary, admission method. http://www.datadictionary.nhs.uk/data_dictionary/attributes/a/add/admission_method_de.asp.

201. Archived NHS Direct website. http://webarchive.nationalarchives.gov.uk/20140220132333/http://www.nhsdirect.nhs.uk/About/WhatIsNHSDirect/History.

202. Out-of-hours GP services in England. National Audit Office2014 9 July 2014.

203. Nuffield Trust. A&E attendences in England. 2013. https://www.nuffieldtrust.org.uk/chart/a-e-attendances-in-england.

204. Blunt I. Focus on: A&E attendances. London: Quality Watch2014.

205. Hospital Episode Statistics: Accident and Emergency Attendances in England (Experimental Statistics), 2010-11. Health and Social Care Information Centre2012.

206. Hagger-Johnson G, Harron K, Gonzalez-Izquierdo A, Cortina-Borja M, Dattani N, Muller-Pebody B et al. Identifying Possible False Matches in Anonymized Hospital Administrative Data without Patient Identifiers. Health Services Research. 2015;50(4):1162-78. doi:10.1111/1475-6773.12272.

207. McDonald HI, Nitsch D, Millett ERC, Sinclair A, Thomas SL. New estimates of the burden of acute community-acquired infections among older people with diabetes mellitus: a retrospective cohort study using linked electronic health records. Diabet Med. 2014;31(5):606-14. doi:10.1111/dme.12384.

208. McDonald HI, Nitsch D, Millett ERC, Sinclair A, Thomas SL. Are pre-existing markers of chronic kidney disease associated with short-term mortality following acute community-acquired pneumonia and sepsis? A cohort study among older people with diabetes using electronic health records. Nephrol Dial Transplant. 2015;30(6):1002-9. doi:10.1093/ndt/gfu401.

209. McDonald HI, Thomas SL, Millett ERC, Nitsch D. CKD and the Risk of Acute, Community-Acquired Infections Among Older People With Diabetes Mellitus: A Retrospective Cohort Study Using Electronic Health Records. Am J Kidney Dis. 2015;66(1):60-8. doi:http://dx.doi.org/10.1053/j.ajkd.2014.11.027.

210. QResearch homepage. http://www.qresearch.org/SitePages/Home.aspx. Accessed 12/10/2015.

211. The Health Improvement Network (THIN). http://www.thin-uk.net/. Accessed 12/10/2015.

212. Snijders B, van der Hoek W, Stirbu I, van der Sande MAB, van Gageldonk-Lafeber AB. General practitioners' contribution to the management of community-acquired pneumonia in the Netherlands: a retrospective analysis of primary care, hospital, and national mortality databases with individual data linkage. Prim Care Respir J. 2013;22(4):400-5. doi:10.4104/pcrj.2013.00085.

213. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. Ther Adv Drug Saf. 2012;3(2):89-99. doi:10.1177/2042098611435911.

Appendix A Read and ICD-10 codes for all LRTI and pneumonia

LRTI and pneumonia Read codes used in CPRD

Readcode	Readterm	Flag
65VA.00	Notification of whooping cough	
A022200	Salmonella pneumonia	pneumonia
A203.00	Primary pneumonic plague	pneumonia
A205.00	Pneumonic plague, unspecified	pneumonia
A33..00	Whooping cough	
A330.00	Bordetella pertussis	
A331.00	Bordetella parapertussis	
A33y.00	Whooping cough - other specified organism	
A33yz00	Other whooping cough NOS	
A33z.00	Whooping cough NOS	
A3BXA00	Mycoplasma pneumoniae [PPLO] cause/dis classifd/oth chaptr	
A3By100	Eaton's agent infection	
A3By400	Pleuropneumonia-like organism (PPLO) infection	
A521.00	Varicella pneumonitis	pneumonia
A54x400	Herpes simplex pneumonia	pneumonia
A551.00	Postmeasles pneumonia	pneumonia
A730.00	Ornithosis with pneumonia	pneumonia
A785000	Cytomegaloviral pneumonitis	pneumonia
A789300	HIV disease resulting in Pneumocystis carinii pneumonia	pneumonia
AB24.11	Pneumonia - candidal	pneumonia
AB40500	Histoplasma capsulatum with pneumonia	pneumonia
AB40600	Acute pulmonary histoplasmosis capsulati	
AB41500	Histoplasma duboisii with pneumonia	pneumonia
AD04.00	Toxoplasma pneumonitis	pneumonia
AD63.00	Pneumocystosis	pneumonia
Ayu3A00	[X]Whooping cough, unspecified	
G520300	Acute myocarditis - influenzal	
H04..00	Acute laryngitis and tracheitis	
H041.00	Acute tracheitis	
H041000	Acute tracheitis without obstruction	
H041100	Acute tracheitis with obstruction	
H041z00	Acute tracheitis NOS	
H042.00	Acute laryngotracheitis	
H042000	Acute laryngotracheitis without obstruction	
H042100	Acute laryngotracheitis with obstruction	
H042z00	Acute laryngotracheitis NOS	
H04z.00	Acute laryngitis and tracheitis NOS	
H06..00	Acute bronchitis and bronchiolitis	
H060.00	Acute bronchitis	
H060.11	Acute wheezy bronchitis	
H060300	Acute purulent bronchitis	
H060400	Acute croupous bronchitis	
H060500	Acute tracheobronchitis	
H060600	Acute pneumococcal bronchitis	
H060700	Acute streptococcal bronchitis	

H060800	Acute haemophilus influenzae bronchitis	
H060900	Acute neisseria catarrhalis bronchitis	
H060A00	Acute bronchitis due to mycoplasma pneumoniae	
H060B00	Acute bronchitis due to coxsackievirus	
H060C00	Acute bronchitis due to parainfluenza virus	
H060D00	Acute bronchitis due to respiratory syncytial virus	
H060E00	Acute bronchitis due to rhinovirus	
H060F00	Acute bronchitis due to echovirus	
H060w00	Acute viral bronchitis unspecified	
H060x00	Acute bacterial bronchitis unspecified	
H060z00	Acute bronchitis NOS	
H061.00	Acute bronchiolitis	
H061000	Acute capillary bronchiolitis	
H061200	Acute bronchiolitis with bronchospasm	
H061300	Acute exudative bronchiolitis	
H061500	Acute bronchiolitis due to respiratory syncytial virus	
H061600	Acute bronchiolitis due to other specified organisms	
H061z00	Acute bronchiolitis NOS	
H062.00	Acute lower respiratory tract infection	
H06z.00	Acute bronchitis or bronchiolitis NOS	
H06z000	Chest infection NOS	
H06z011	Chest infection	
H06z100	Lower resp tract infection	
H06z112	Acute lower respiratory tract infection	
H06z200	Recurrent chest infection	
H07..00	Chest cold	
H2...00	Pneumonia and influenza	pneumonia
H20..00	Viral pneumonia	pneumonia
H20..11	Chest infection - viral pneumonia	pneumonia
H200.00	Pneumonia due to adenovirus	pneumonia
H201.00	Pneumonia due to respiratory syncytial virus	pneumonia
H202.00	Pneumonia due to parainfluenza virus	pneumonia
H20y.00	Viral pneumonia NEC	pneumonia
H20y000	Severe acute respiratory syndrome	pneumonia
H20z.00	Viral pneumonia NOS	pneumonia
H21..00	Lobar (pneumococcal) pneumonia	pneumonia
H21..11	Chest infection - pneumococcal pneumonia	pneumonia
H22..00	Other bacterial pneumonia	pneumonia
H22..11	Chest infection - other bacterial pneumonia	pneumonia
H220.00	Pneumonia due to klebsiella pneumoniae	pneumonia
H221.00	Pneumonia due to pseudomonas	pneumonia
H222.00	Pneumonia due to haemophilus influenzae	pneumonia
H222.11	Pneumonia due to haemophilus influenzae	pneumonia
H223.00	Pneumonia due to streptococcus	pneumonia
H223000	Pneumonia due to streptococcus, group B	pneumonia
H224.00	Pneumonia due to staphylococcus	pneumonia
H22y.00	Pneumonia due to other specified bacteria	pneumonia
H22y000	Pneumonia due to escherichia coli	pneumonia
H22y011	E.coli pneumonia	pneumonia
H22y100	Pneumonia due to proteus	pneumonia

H22y200	Pneumonia - Legionella	pneumonia
H22yX00	Pneumonia due to other aerobic gram-negative bacteria	pneumonia
H22yz00	Pneumonia due to bacteria NOS	pneumonia
H22z.00	Bacterial pneumonia NOS	pneumonia
H23..00	Pneumonia due to other specified organisms	pneumonia
H23..11	Chest infection - pneumonia organism OS	pneumonia
H230.00	Pneumonia due to Eaton's agent	pneumonia
H231.00	Pneumonia due to mycoplasma pneumoniae	pneumonia
H232.00	Pneumonia due to pleuropneumonia like organisms	pneumonia
H233.00	Chlamydial pneumonia	pneumonia
H23z.00	Pneumonia due to specified organism NOS	pneumonia
H24..00	Pneumonia with infectious diseases EC	pneumonia
H24..11	Chest infection with infectious disease EC	
H240.00	Pneumonia with measles	pneumonia
H241.00	Pneumonia with cytomegalic inclusion disease	pneumonia
H242.00	Pneumonia with ornithosis	pneumonia
H243.00	Pneumonia with whooping cough	pneumonia
H243.11	Pneumonia with pertussis	pneumonia
H246.00	Pneumonia with aspergillosis	pneumonia
H247000	Pneumonia with candidiasis	pneumonia
H247z00	Pneumonia with systemic mycosis NOS	pneumonia
H24y.00	Pneumonia with other infectious diseases EC	pneumonia
H24y000	Pneumonia with actinomycosis	pneumonia
H24y100	Pneumonia with nocardiasis	pneumonia
H24y200	Pneumonia with pneumocystis carinii	pneumonia
H24y300	Pneumonia with Q-fever	pneumonia
H24y400	Pneumonia with salmonellosis	pneumonia
H24y500	Pneumonia with toxoplasmosis	pneumonia
H24y600	Pneumonia with typhoid fever	pneumonia
H24y700	Pneumonia with varicella	pneumonia
H24yz00	Pneumonia with other infectious diseases EC NOS	pneumonia
H24z.00	Pneumonia with infectious diseases EC NOS	pneumonia
H25..00	Bronchopneumonia due to unspecified organism	pneumonia
H25..11	Chest infection - unspecified bronchopneumonia	pneumonia
H26..00	Pneumonia due to unspecified organism	pneumonia
H26..11	Chest infection - pnemonia due to unspecified organism	pneumonia
H260.00	Lobar pneumonia due to unspecified organism	pneumonia
H260000	Lung consolidation	pneumonia
H261.00	Basal pneumonia due to unspecified organism	pneumonia
H262.00	Postoperative pneumonia	postoperative
H27..00	Influenza	
H270.00	Influenza with pneumonia	pneumonia
H270.11	Chest infection - influenza with pneumonia	pneumonia
H270000	Influenza with bronchopneumonia	pneumonia
H270100	Influenza with pneumonia, influenza virus identified	pneumonia
H270z00	Influenza with pneumonia NOS	pneumonia
H271.00	Influenza with other respiratory manifestation	
H271z00	Influenza with respiratory manifestations NOS	
H27y.00	Influenza with other manifestations	
H27y000	Influenza with encephalopathy	

H27y100	Influenza with gastrointestinal tract involvement	
H27yz00	Influenza with other manifestations NOS	
H27z.00	Influenza NOS	
H28..00	Atypical pneumonia	pneumonia
H29..00	Avian influenza	
H2A..00	Influenza due to Influenza A virus subtype H1N1	
H2A..11	Influenza A (H1N1) swine flu	
H2y..00	Other specified pneumonia or influenza	
H2z..00	Pneumonia or influenza NOS	
H30..11	Chest infection - unspecified bronchitis	
H3y0.00	Chronic obstruct pulmonary dis with acute lower resp infectn	
H510900	Pneumococcal pleurisy	
H510A00	Staphylococcal pleurisy	
H510B00	Streptococcal pleurisy	
H511.00	Bacterial pleurisy with effusion	
H511000	Pneumococcal pleurisy with effusion	
H511100	Staphylococcal pleurisy with effusion	
H511z00	Bacterial pleurisy with effusion NOS	
H530200	Gangrenous pneumonia	pneumonia
H530300	Abscess of lung with pneumonia	pneumonia
H540000	Hypostatic pneumonia	pneumonia
H540100	Hypostatic bronchopneumonia	pneumonia
H564.00	Bronchiolitis obliterans organising pneumonia	pneumonia
Hyu0400	[X]Flu+oth respiratory manifestations,'flu virus identified	
Hyu0500	[X]Influenza+other manifestations,influenza virus identified	
Hyu0600	[X]Influenza+oth respiratory manifestatns,virus not identifd	
Hyu0700	[X]Influenza+other manifestations, virus not identified	
Hyu0800	[X]Other viral pneumonia	pneumonia
Hyu0A00	[X]Other bacterial pneumonia	pneumonia
Hyu0B00	[X]Pneumonia due to other specified infectious organisms	pneumonia
Hyu0D00	[X]Pneumonia in viral diseases classified elsewhere	pneumonia
Hyu0H00	[X]Other pneumonia, organism unspecified	pneumonia
Hyu1.00	[X]Other acute lower respiratory infections	
Hyu1000	[X]Acute bronchitis due to other specified organisms	
Hyu1100	[X]Acute bronchiolitis due to other specified organisms	
16L..00	Influenza-like symptoms	
1J72.00	Suspected influenza A virus subtype H1N1 infection	
1J72.11	Suspected swine influenza	
1W0..00	Possible influenza A virus H1N1 subtype	
43jQ.00	Avian influenza virus nucleic acid detection	
43jx.00	Parainfluenza type 1 nucleic acid detection	
43jy.00	Parainfluenza type 2 nucleic acid detection	
43jz.00	Parainfluenza type 3 nucleic acid detection	
4J3L.00	Influenza A virus H1N1 subtype detected	
4JU0.00	Influenza H1 virus detected	
4JU2.00	Influenza H3 virus detected	
4JU3.00	Influenza H5 virus detected	
4JU4.00	Influenza A virus, other or untyped strain detected	
4JU5.00	Influenza B virus detected	
4JUF.00	Human parainfluenza virus detected	

4JUK.00	Mycoplasma pneumoniae detected	
A39y000	Pulmonary nocardiosis	
AB42.00	Pulmonary histoplasmosis	
H042.11	Laryngotracheitis	
H052.00	Pharyngotracheitis	
H053.00	Tracheopharyngitis	
H060v00	Subacute bronchitis unspecified	
H271000	Influenza with laryngitis	
H271100	Influenza with pharyngitis	
H27z.11	Flu like illness	
H27z.12	Influenza like illness	
H50..00	Empyema	
H500.00	Empyema with fistula	
H500000	Empyema with bronchocutaneous fistula	
H500100	Empyema with bronchopleural fistula	
H500400	Empyema with pleural fistula NOS	
H501.00	Empyema with no fistula	
H501000	Pleural abscess	
H501200	Pleural empyema	
H501300	Lung empyema NOS	
H501400	Purulent pleurisy	
H501500	Pyopneumothorax	
H501600	Pyothorax	
H50z.00	Empyema NOS	
1419.00	H/O: pertussis	History of
1419.11	H/O: whooping cough	History of
14B2.00	H/O: pneumonia	History of
14B3.11	H/O: bronchitis	History of
H341.00	Post-infective bronchiectasis	History of

LRTI and pneumonia ICD10 codes used in HES

ICD10 code	Diagnostic name	Flag
A37	Whooping cough	
A370	Whooping cough due to Bordetella pertussis	
A371	Whooping cough due to Bordetella parapertussis	
A378	Whooping cough due to other Bordetella species	
A379	Whooping cough, unspecified	
B012	Varicella pneumonia	pneumonia
B052	Measles complicated by pneumonia	pneumonia
B206	HIV disease resulting in Pneumocystis carinii pneumonia	pneumonia
B250	Cytomegaloviral pneumonitis	pneumonia
B960	Mycoplasma pneumoniae as cause dis class oth chaps	
J041	Acute tracheitis	
J042	Acute laryngotracheitis	
J09	Influenza due to other identified influenza virus	
J10	Influenza due to identified influenza virus	
J100	Influenza with pneumonia, influenza virus identified	pneumonia
J101	Influenza with oth resp manifest influenza virus identified	
J108	Influenza with other manifest influenza virus identified	
J11	Influenza, virus not identified	
J110	Influenza with pneumonia, virus not identified	pneumonia
J111	Influenza with oth resp manifestation virus not identified	
J118	Influenza with other manifestations, virus not identified	
J12	Viral pneumonia, not elsewhere classified	pneumonia
J120	Adenoviral pneumonia	pneumonia
J121	Respiratory syncytial virus pneumonia	pneumonia
J122	Parainfluenza virus pneumonia	pneumonia
J128	Other viral pneumonia	pneumonia
J129	Viral pneumonia, unspecified	pneumonia
J13	Pneumonia due to Streptococcus pneumoniae	pneumonia
J13X	Pneumonia due to Streptococcus pneumoniae	pneumonia
J14	Pneumonia due to Haemophilus influenzae	pneumonia
J14X	Pneumonia due to Haemophilus influenzae	pneumonia
J15	Bacterial pneumonia, not elsewhere classified	pneumonia
J150	Pneumonia due to Klebsiella pneumoniae	pneumonia
J151	Pneumonia due to Pseudomonas	pneumonia
J152	Pneumonia due to staphylococcus	pneumonia
J153	Pneumonia due to streptococcus, group B	pneumonia
J154	Pneumonia due to other streptococci	pneumonia
J155	Pneumonia due to Escherichia coli	pneumonia
J156	Pneumonia due to other aerobic Gram-negative bacteria	pneumonia
J157	Pneumonia due to Mycoplasma pneumoniae	pneumonia
J158	Other bacterial pneumonia	pneumonia
J159	Bacterial pneumonia, unspecified	pneumonia
J16	Pneumonia due to other infectious organisms NEC	pneumonia

J160	Chlamydial pneumonia	pneumonia
J168	Pneumonia due to other specified infectious organisms	pneumonia
J17	Pneumonia in diseases classified elsewhere	pneumonia
J170	Pneumonia in bacterial diseases classified elsewhere	pneumonia
J171	Pneumonia in viral diseases classified elsewhere	pneumonia
J172	Pneumonia in mycoses	pneumonia
J173	Pneumonia in parasitic diseases	pneumonia
J178	Pneumonia in other diseases classified elsewhere	pneumonia
J18	Pneumonia, organism unspecified	pneumonia
J180	Bronchopneumonia, unspecified	pneumonia
J181	Lobar pneumonia, unspecified	pneumonia
J182	Hypostatic pneumonia, unspecified	pneumonia
J188	Other pneumonia, organism unspecified	pneumonia
J189	Pneumonia, unspecified	pneumonia
J20	Acute bronchitis	
J200	Acute bronchitis due to Mycoplasma pneumoniae	
J201	Acute bronchitis due to Haemophilus influenzae	
J202	Acute bronchitis due to streptococcus	
J203	Acute bronchitis due to coxsackievirus	
J204	Acute bronchitis due to parainfluenza virus	
J205	Acute bronchitis due to respiratory syncytial virus	
J206	Acute bronchitis due to rhinovirus	
J207	Acute bronchitis due to echovirus	
J208	Acute bronchitis due to other specified organisms	
J209	Acute bronchitis, unspecified	
J21	Acute bronchiolitis	
J210	Acute bronchiolitis due to respiratory syncytial virus	
J218	Acute bronchiolitis due to other specified organisms	
J219	Acute bronchiolitis, unspecified	
J22	Unspecified acute lower respiratory infection	
J22X	Unspecified acute lower respiratory infection	
J440	Chronic obstructive pulmonary disease with acute lower respiratory infection	
J851	Abscess of lung with pneumonia	pneumonia
J86	Pyothorax	
J860	Pyothorax with fistula	
J869	Pyothorax without fistula	
U04	Severe acute respiratory syndrome [SARS]	pneumonia
U049	Severe acute respiratory syndrome, unspecified	pneumonia

Appendix B Additional results for comparison of LRTI incidence in those who registered pre and post-UTS over the first three years of follow-up (Chapter 3)

Overall comparison
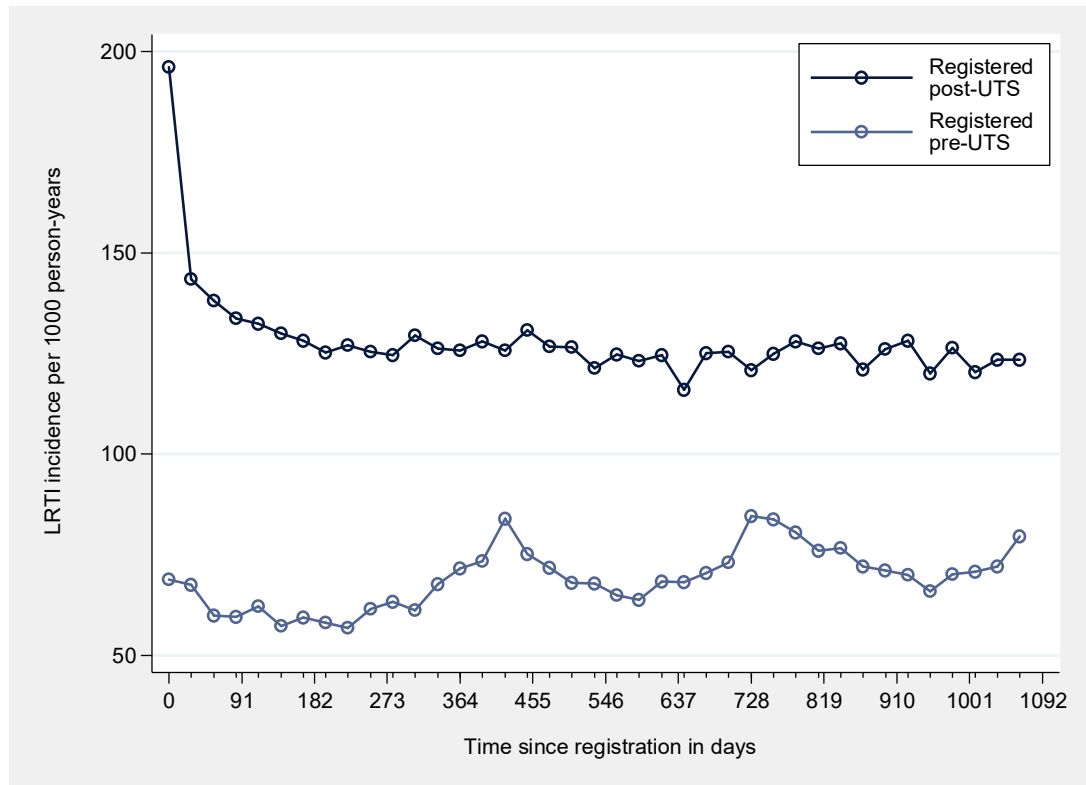
Comparison of age-stratified LRTI incidence in those who registered pre and post-UTS over the first three years of follow-up
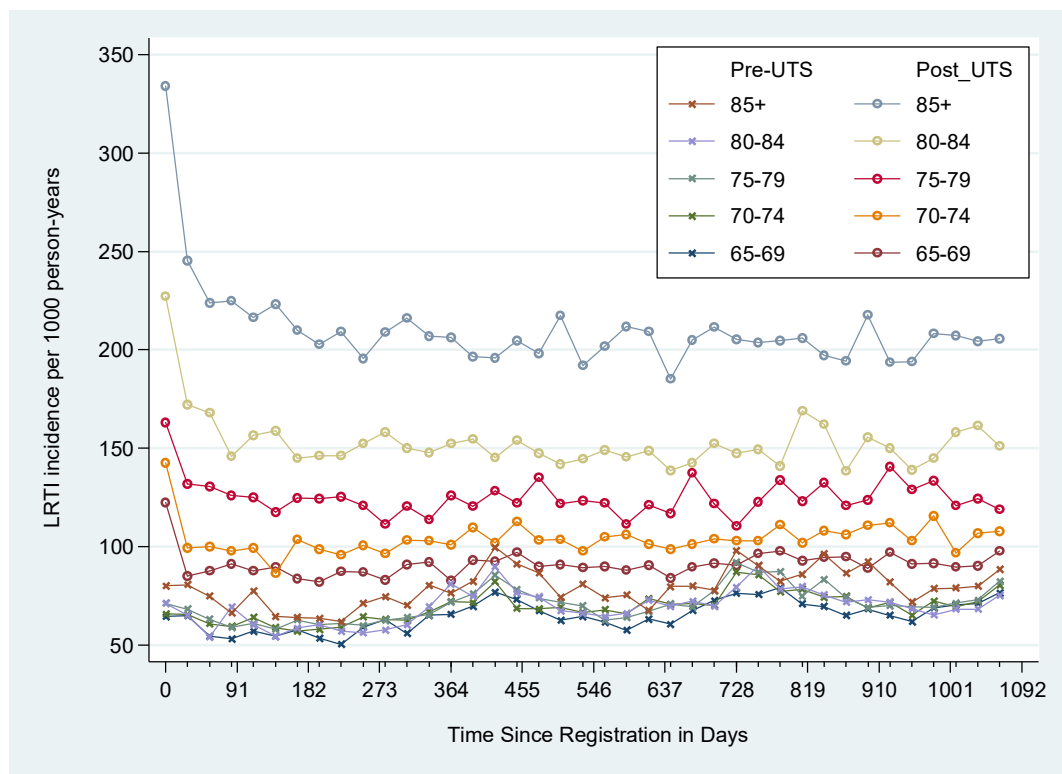
Comparison of LRTI incidence stratified by year of start of follow-up in those who registered pre and post-UTS over the first three years of follow-up
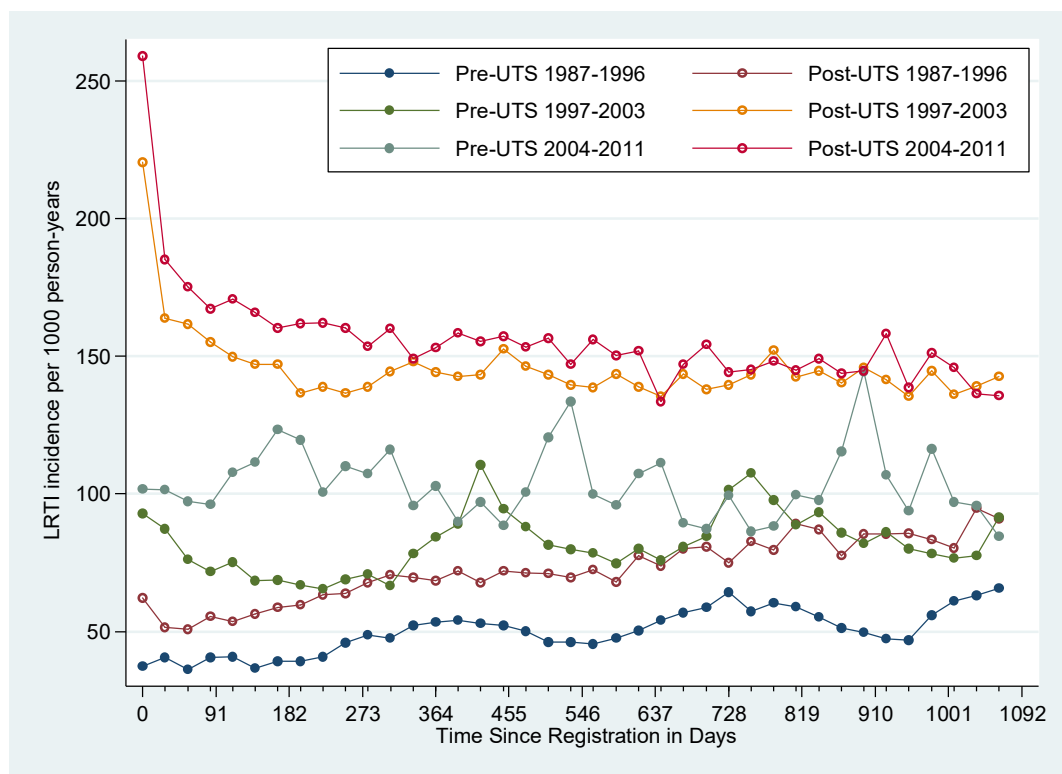
Comparison of LRTI incidence over the first year of follow-up in the post-UTS group stratified by health check exclusion
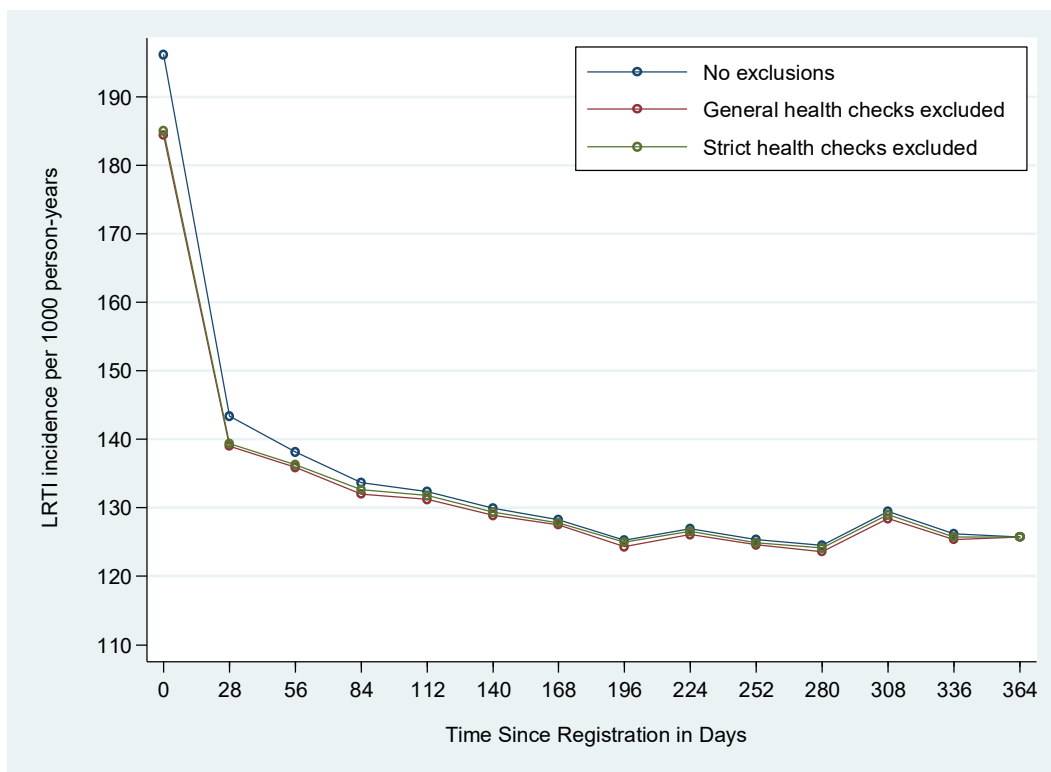
Comparison of LRTI incidence stratified by sex in those who registered pre and post-UTS over the first three years of follow-up
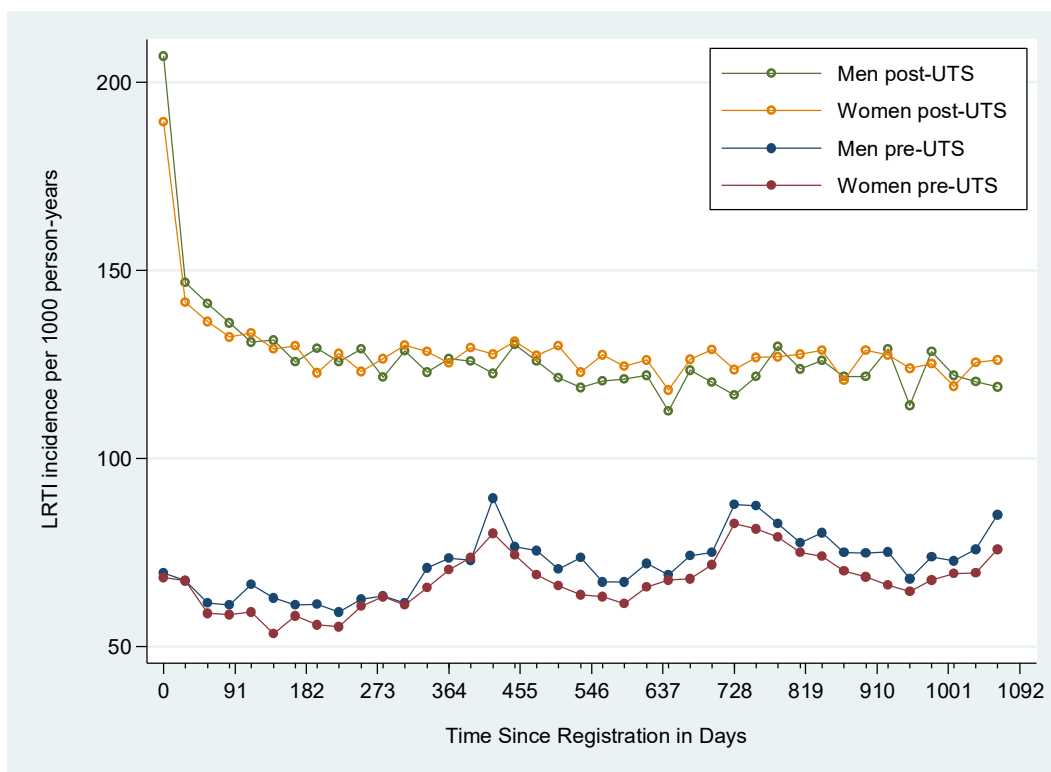
Appendix C Medline search strategy for the European incidence of CAP and all LRTI in literature review (Chapter 4)

1	exp pneumonia/
2	(pneumonit* or pneumonia).ti,ab.
3	bronchopneumonia.ti,ab.
4	pleuropneumonia.ti,ab.
5	((lung or lobar or pulmonary) adj2 inflamm*).ti,ab.
6	(lower respiratory adj3 (infection* or inflamm*)).ti,ab.
7	LRTI.ti,ab.
8	Bronchitis/ep, mo, sn [Epidemiology, Mortality, Statistics & Numerical Data]
9	exp Bronchopneumonia/ep, mo [Epidemiology, Mortality]
10	exp Bronchiolitis/ep, mo [Epidemiology, Mortality]
11	exp Trachietis/ep, mo, sn
12	(bronchitis or bronchiolitis).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
13	pneumonias.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
14	(chest adj infection*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
15	LRTI.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
16	or/1-15
17	exp incidence/
18	exp Epidemiology/sn, td [Statistics & Numerical Data, Trends]
19	exp Population Surveillance/
20	inciden*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
21	epidemiol*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
22	surveillance.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
23	occur*.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
24	frequency.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
25	or/17-24
26	exp Aged, 80/ and over.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
27	exp Aged/

28	((old adj age*) or elderly or (senior adj citizen)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
29	26 or 27 or 28
30	exp Albania/
31	(Albania or Albanian or Albanians).ti,ab.
32	exp Andorra/
33	(Andorra or Andorran or Andorrans).ti,ab.
34	exp Armenia/
35	(Armenia or Armenian or Armenians).ti,ab.
36	exp Austria/
37	(Austria or Austrian or Austrians).ti,ab.
38	exp Azerbaijan/
39	Azerbaijan*.ti,ab.
40	exp Belgium/
41	(Belgium or Belgian*).ti,ab.
42	exp "Bosnia and Herzegovina"/
43	(Bosnia*-Her#egovin* or BOSNIA* or HER#EGOVIN*).ti,ab.
44	exp Bulgaria/
45	(Bulgaria or Bulgarian or Bulgarians).ti,ab.
46	exp Croatia/
47	(Croatia or Croatian or Croatians).ti,ab.
48	exp Cyprus/
49	(Cyprus or Cypriot or Cypriots).ti,ab.
50	exp Czechoslovakia/
51	exp Czech Republic/
52	(Czech Republic or Czechoslovakia or Czech or Czechs).ti,ab.
53	exp Denmark/
54	(denmark or faeroe islands or Danish).ti,ab.
55	exp Estonia/
56	(Estonia or Estonian or Estonians).ti,ab.
57	exp Europe/
58	EUROPE*.ti,ab.
59	exp Finland/
60	(Finland or Finnish).ti,ab.
61	exp France/
62	(France or French).ti,ab.
63	exp "Georgia (republic)"/
64	(Georgian or Georgians).ti,ab.
65	exp Germany/
66	(Germany or German or Germans).ti,ab.
67	exp United Kingdom/
68	(great britain or GBR or united kingdom or UK or northern ireland or scotland or channel islands or (isle adj2 man) or British or Scottish or (wales not new south wales) or (england not new england)).ti,ab.
69	exp Greece/
70	(Greece or Greek or Greeks).ti,ab.

71	exp Hungary/
72	(Hungary or Hungarian or Hungarians).ti,ab.
73	exp Iceland/
74	(Iceland or Icelandic).ti,ab.
75	exp Ireland/
76	(eire or ireland or Irish).ti,ab.
77	exp Israel/
78	(Israel or Israeli or Israelis).ti,ab.
79	exp Italy/
80	(Italy or Italian or Italians).ti,ab.
81	exp Kazakhstan/
82	(kazakh or kazakhs or kazakhstan or kazakhstani).ti,ab.
83	exp Kyrgyzstan/
84	(kirgizstan or kyrgyz or kirghizia or kirghiz or kyrgyzstan or Kyrgyzstani).ti,ab.
85	exp Latvia/
86	(Latvia or Latvian or Latvians).ti,ab.
87	exp Liechtenstein/
88	(liechtenstein or leichtenstein).ti,ab.
89	exp Lithuania/
90	(Lithuania or Lithuanian or Lithuanians).ti,ab.
91	exp Luxembourg/
92	(luxembourg* or luxemburg* or luxemborg).ti,ab.
93	exp "Macedonia (republic)"/
94	(Macedonia or Macedonian or Macedonians).ti,ab.
95	exp Malta/
96	(Malta or Maltese).ti,ab.
97	exp Moldova/
98	(Moldavia or Moldavian or Moldova or Moldovan or Moldovans).ti,ab.
99	exp Monaco/
100	(Monaco or Monegasque).ti,ab.
101	(Montenegro or Montenegrin or Montenegrins).ti,ab.
102	exp Netherlands/
103	(netherlands or holland or Dutch).ti,ab.
104	exp Norway/
105	(Norway or Norwegian or Norwegians).ti,ab.
106	exp Poland/
107	(Poland or (Polish adj3 (population or patient* or people))).ti,ab.
108	exp Portugal/
109	(Portugal or Portuguese).ti,ab.
110	exp Belarus/
111	(belarus or byelarus or belorussia or Belarusian or Belarusians).ti,ab.
112	exp Romania/
113	(Romania or Romanian or Romanians).ti,ab.
114	exp USSR/
115	(Russia or Russian or Russians).ti,ab.
116	exp San Marino/

117	exp Russia/
118	(San Marino or Sammarinese).ti,ab.
119	exp Scandinavia/
120	(Scandinavia or Scandinavian).ti,ab.
121	exp Serbia/
122	(Serbia or Serbian or Serbians).ti,ab.
123	exp Slovakia/
124	(slovakia or slovak or Slovakian or Slovakians or Slovak or Slovaks).ti,ab.
125	exp Slovenia/
126	(Slovenia or Slovenian or Slovenians).ti,ab.
127	exp Spain/
128	(spain or balearic islands or canary islands or Spanish).ti,ab.
129	exp Sweden/
130	(Sweden or Swedish).ti,ab.
131	exp Switzerland/
132	(Switzerland or Swiss).ti,ab.
133	exp Tajikistan/
134	(tadjikistan or tadzhik or tadzhikistan or tajikistan).ti,ab.
135	exp Turkey/
136	(turkey or Turkish).ti,ab.
137	exp Turkmenistan/
138	(turkmen or turkmenistan or Turkmens).ti,ab.
139	exp Ukraine/
140	(Ukraine or Ukrainian or Ukrainians).ti,ab.
141	exp Uzbekistan/
142	(uzbekistan or uzbek or Uzbeks).ti,ab.
143	exp "Yugoslavia (pre-1992)"/ or exp Yugoslavia/
144	(Yugoslavia or Yugoslav or Yugoslavs or Yugoslavian or Yugoslavians).ti,ab.
145	30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 or 80 or 81 or 82 or 83 or 84 or 85 or 86 or 87 or 88 or 89 or 90 or 91 or 92 or 93 or 94 or 95 or 96 or 97 or 98 or 99 or 100 or 101 or 102 or 103 or 104 or 105 or 106 or 107 or 108 or 109 or 110 or 111 or 112 or 113 or 114 or 115 or 116 or 117 or 118 or 119 or 120 or 121 or 122 or 123 or 124 or 125 or 126 or 127 or 128 or 129 or 130 or 131 or 132 or 133 or 134 or 135 or 136 or 137 or 138 or 139 or 140 or 141 or 142 or 143 or 144
146	16 and 25 and 29 and 145
147	exp Case Reports/
148	Animals/
149	Humans/
150	148 not (148 and 149)
151	146 not 150
152	151 not 147
153	152
154	limit 153 to yr="1980 -Current"
155	154
156	limit 155 to english language

Appendix D Supplementary material for Paper 1 (Chapter 4)

Community-acquired LRTI incidence rates overall and over time by sex, age, region of England and IMD quintile

| | Sex | | | Age (years) | | | | | |
	Male	Female	65-69	70-74	75-79	80-84	85-89	≥90
Overall	122.93	121.76	123.83	92.21	107.40	126.07	151.37	187.91	262.87
95% CI	122.49-123.37	121.1-122.42	123.24-124.42	91.7-92.71	106.8-108	125.33-126.81	150.36-152.39	186.32-189.49	259.79-265.94
1997	100.96	100.87	100.89	94.10	104.72	113.58	126.88	144.38	170.43
	99.91-102.01	99.27-102.48	99.51-102.28	92.01-96.19	102.38-107.06	110.91-116.26	123.37-130.38	139.49-149.27	163.18-177.69
1998	104.37	101.35	106.54	93.24	102.54	117.40	133.64	158.26	177.00
	103.35-105.38	99.82-102.88	105.18-107.9	91.29-95.19	100.37-104.72	114.88-119.92	130.16-137.13	153.4-163.12	170.03-183.97
1999	105.40	102.74	107.32	92.49	103.44	113.37	129.50	152.10	186.11
	104.44-106.36	101.3-104.19	106.04-108.61	90.7-94.29	101.42-105.46	111.1-115.65	126.32-132.69	147.72-156.47	179.44-192.79
2000	102.13	100.69	103.14	86.91	98.31	108.03	121.03	144.89	173.51
	101.22-103.03	99.32-102.05	101.94-104.34	85.29-88.54	96.46-100.15	105.92-110.13	118.23-123.82	140.87-148.92	167.46-179.56
2001	105.90	104.07	107.23	85.74	99.73	111.24	123.05	144.29	183.92
	105-106.81	102.7-105.44	106.03-108.44	84.18-87.3	97.92-101.54	109.14-113.34	120.37-125.72	140.38-148.2	177.86-189.98
2002	112.22	110.64	113.39	91.00	101.98	115.03	128.12	148.88	185.77
	111.3-113.15	109.23-112.04	112.16-114.63	89.42-92.57	100.2-103.77	112.92-117.14	125.49-130.75	144.9-152.85	179.79-191.74
2003	127.30	124.11	129.72	98.61	113.08	129.55	143.43	167.98	211.54
	126.3-128.3	122.61-125.61	128.38-131.05	96.99-100.23	111.19-114.97	127.29-131.8	140.69-146.18	163.65-172.32	205.14-217.94
2004	131.97	130.66	133.00	99.20	113.45	131.45	146.52	172.24	223.38
	130.95-132.99	129.11-132.21	131.64-134.35	97.6-100.8	111.57-115.32	129.2-133.7	143.78-149.26	167.88-176.6	216.8-229.96
2005	131.76	131.06	132.34	97.69	111.24	125.93	144.46	166.92	222.45
	130.74-132.78	129.51-132.62	130.99-133.69	96.12-99.26	109.4-113.08	123.76-128.1	141.74-147.19	162.84-171	215.89-229
2006	134.09	134.70	133.72	97.57	111.26	126.59	145.70	162.82	223.53
	133.05-135.12	133.11-136.3	132.35-135.08	96-99.14	109.43-113.1	124.41-128.77	142.94-148.46	158.93-166.7	216.92-230.14
2007	139.88	142.07	138.36	99.66	113.43	130.54	148.68	171.04	234.79
	138.81-140.95	140.42-143.73	136.96-139.77	98.08-101.25	111.58-115.28	128.32-132.76	145.86-151.49	167.11-174.97	227.84-241.73
2008	148.04	149.65	147.02	101.86	117.61	134.62	158.30	183.45	267.05
	146.93-149.16	147.92-151.37	145.56-148.49	100.26-103.47	115.71-119.51	132.33-136.91	155.34-161.26	179.34-187.55	259.28-274.82
2009	131.10	132.49	130.24	90.26	104.38	118.38	134.93	159.58	229.95
	130.06-132.14	130.88-134.1	128.87-131.61	88.77-91.74	102.6-106.16	116.24-120.52	132.23-137.63	155.8-163.35	222.78-237.11
2010	137.30	135.46	138.90	93.95	107.69	119.46	140.93	167.30	240.21
	136.21-138.38	133.81-137.1	137.46-140.35	92.42-95.48	105.85-109.53	117.27-121.64	138.13-143.74	163.33-171.27	233.03-247.38

					Region					
	North East	North West	Yorkshire & The Humber	East Midlands	West Midlands	East of England	South West	South Central	London	South East Coast
Overall	145.47	158.81	156.37	126.43	154.29	115.72	110.27	114.11	96.84	104.30
95% CI	141.85-149.09	157.23-160.39	153.87-158.88	124.2-128.66	152.48-156.11	114.38-117.05	108.95-111.59	112.84-115.37	95.68-98	103.03-105.58
1997	104.47	134.85	110.13	93.82	111.37	91.68	94.63	96.80	90.03	78.03
	97.94-111	131.47-138.23	105.65-114.61	89.45-98.18	107.77-114.96	88.69-94.68	91.15-98.1	93.12-100.47	86.8-93.26	74.85-81.21
1998	112.42	125.24	124.57	108.18	130.84	101.35	94.42	94.97	93.54	79.21
	105.52-119.31	122.13-128.35	119.87-129.27	103.61-112.76	127.14-134.54	98.35-104.36	91.22-97.62	91.67-98.26	90.35-96.73	76.07-82.34
1999	135.29	125.99	125.62	98.18	146.35	97.61	98.25	98.01	85.66	84.53
	127.49-143.09	122.97-129.01	120.96-130.28	94.05-102.31	142.49-150.21	94.8-100.41	95.28-101.22	95.04-100.98	82.8-88.51	81.52-87.53
2000	122.24	127.19	131.39	106.53	125.68	94.40	86.60	93.83	91.12	81.63
	115.32-129.16	124.2-130.18	126.66-136.12	102.17-110.9	122.33-129.03	91.72-97.08	83.92-89.29	91.2-96.47	88.32-93.92	78.88-84.39
2001	136.47	126.64	150.14	112.37	132.35	104.34	97.19	103.33	96.77	90.76
	128.95-143.99	123.73-129.55	145.08-155.19	107.93-116.8	128.89-135.81	101.52-107.15	94.31-100.08	100.63-106.04	93.95-99.59	87.89-93.64
2002	152.07	142.05	160.25	116.39	146.89	111.47	102.17	102.96	91.00	95.52
	143.91-160.23	138.99-145.12	155.07-165.42	111.86-120.92	143.19-150.58	108.58-114.37	99.21-105.12	100.24-105.67	88.32-93.68	92.64-98.4
2003	148.22	165.30	182.43	134.71	165.18	121.73	115.81	115.33	98.48	104.90
	140.12-156.33	161.91-168.69	176.7-188.16	129.73-139.69	161.21-169.14	118.66-124.8	112.64-118.99	112.43-118.23	95.65-101.31	101.9-107.9
2004	177.09	173.26	178.00	139.49	168.54	126.02	116.06	121.68	100.24	106.26
	167.98-186.2	169.76-176.76	172.31-183.68	134.26-144.71	164.49-172.59	122.89-129.15	112.91-119.22	118.68-124.69	97.35-103.13	103.25-109.26
2005	159.09	174.18	173.41	144.65	167.85	133.99	116.01	120.20	97.75	107.35
	150.54-167.64	170.66-177.7	167.79-179.02	139.26-150.05	163.76-171.93	130.71-137.26	112.87-119.16	117.2-123.2	94.92-100.59	104.34-110.37
2006	152.34	176.98	175.66	156.95	168.53	125.31	118.64	124.05	98.54	114.06
	143.96-160.72	173.39-180.58	169.79-181.52	151.2-162.71	164.43-172.63	122.08-128.53	115.47-121.8	120.98-127.12	95.67-101.41	110.97-117.15
2007	162.59	186.16	180.05	154.15	170.63	129.19	126.68	126.05	106.05	117.67
	154.12-171.06	182.4-189.91	173.95-186.16	148.31-160	166.43-174.83	125.79-132.58	123.38-129.99	123.03-129.07	103.09-109.01	114.53-120.8
2008	176.49	193.87	192.55	162.69	181.75	138.13	132.73	137.58	108.18	130.18
	167.46-185.52	190-197.74	185.6-199.5	156.41-168.97	177.38-186.12	134.49-141.78	129.32-136.14	134.42-140.74	105.19-111.18	126.83-133.54
2009	162.08	177.40	170.31	135.91	162.97	123.46	113.54	116.89	96.91	117.51
	153.38-170.79	173.71-181.09	163.04-177.57	129.98-141.85	158.89-167.06	119.99-126.94	110.44-116.64	114.04-119.74	94.13-99.7	114.4-120.62
2010	165.09	192.69	169.65	136.60	175.98	133.71	121.80	125.60	99.86	123.34
	156.22-173.95	188.76-196.62	162.24-177.06	129.35-143.86	171.68-180.29	129.93-137.49	118.52-125.08	122.64-128.56	97.06-102.66	120.11-126.57

	0 (least deprived)	1	2	3	4 (most deprived)
			IMD Quintile		
Overall	106.48	113.98	124.96	137.94	182.12
95% CI	105.39-107.57	112.86-115.1	123.63-126.28	136.4-139.47	179.79-184.46
1997	85.30	90.83	96.71	111.58	147.93
	82.48-88.13	88.11-93.55	93.65-99.77	108.14-115.02	143.27-152.6
1998	92.70	97.59	103.73	113.91	141.04
	89.93-95.47	94.88-100.3	100.71-106.75	110.58-117.25	136.68-145.4
1999	98.55	97.65	102.77	116.35	155.08
	95.93-101.18	95.09-100.21	99.95-105.59	113.12-119.57	150.58-159.58
2000	89.87	94.66	101.23	114.68	150.27
	87.52-92.23	92.29-97.03	98.56-103.9	111.6-117.77	145.97-154.56
2001	96.82	103.00	114.62	118.80	153.41
	94.39-99.24	100.56-105.45	111.8-117.45	115.73-121.87	149.12-157.71
2002	97.81	107.35	116.83	129.44	175.75
	95.39-100.22	104.86-109.83	113.97-119.69	126.23-132.65	171.08-180.42
2003	110.06	117.13	129.80	147.04	194.89
	107.48-112.64	114.52-119.74	126.75-132.85	143.55-150.53	189.86-199.92
2004	111.16	122.91	136.51	149.93	201.40
	108.58-113.74	120.23-125.6	133.35-139.68	146.37-153.49	196.21-206.6
2005	112.51	120.96	135.56	151.76	198.54
	109.91-115.1	118.31-123.61	132.38-138.73	148.14-155.38	193.34-203.75
2006	114.92	125.28	139.07	151.50	195.91
	112.29-117.54	122.58-127.98	135.85-142.29	147.87-155.14	190.7-201.12
2007	114.45	125.76	142.85	155.35	204.71
	111.86-117.04	123.06-128.46	139.59-146.1	151.66-159.05	199.34-210.08
2008	125.74	135.11	151.18	167.15	217.22
	123.01-128.46	132.3-137.93	147.8-154.56	163.26-171.04	211.58-222.85
2009	110.04	118.71	134.39	150.72	205.75
	107.55-112.53	116.12-121.29	131.25-137.53	147.04-154.4	200.2-211.31
2010	114.76	126.32	140.00	155.29	219.00
	112.17-117.36	123.61-129.03	136.72-143.27	151.47-159.1	213.2-224.81

Age-standardised* incidence of LRTI and CAP by year, region of England and index of multiple deprivation quintile.

	LRTI Age-standardised* rate/1000 person years				CAP Age-standardised* rate/1000 person years			
	Men	95% CI	Women	95% CI	Men	95% CI	Women	95% CI
Year								
1997	122.40	120.31-124.5	107.72	106.18-109.27	10.27	9.68-10.87	6.82	6.48-7.17
1998	121.34	119.38-121.34	112.11	110.62-113.6	9.91	9.36-10.46	7.51	7.17-7.85
1999	120.24	118.44-122.05	110.64	109.27-112.01	9.89	9.39-10.4	6.79	6.49-7.1
2000	114.01	112.38-115.64	103.87	102.63-105.11	8.31	7.88-8.74	5.56	5.3-5.81
2001	115.00	113.42-116.59	105.80	104.59-107.02	8.99	8.56-9.41	5.77	5.52-6.02
2002	118.49	116.92-120.07	110.16	108.94-111.37	8.78	8.37-9.19	5.83	5.58-6.08
2003	129.80	128.16-131.44	123.90	122.6-125.19	9.76	9.32-10.19	6.54	6.28-6.8
2004	133.16	131.52-134.81	124.75	123.47-126.03	9.47	9.05-9.89	6.62	6.36-6.88
2005	129.94	128.34-131.55	122.30	121.04-123.56	9.04	8.64-9.44	6.43	6.18-6.69
2006	130.10	128.5-131.69	122.13	120.87-123.39	9.36	8.96-9.76	6.43	6.17-6.68
2007	134.88	133.25-136.51	124.82	123.54-126.1	9.74	9.33-10.15	6.53	6.27-6.79
2008	140.55	138.85-142.25	131.36	130.04-132.69	11.09	10.65-11.53	7.65	7.37-7.93
2009	122.79	121.22-124.36	115.42	114.18-116.65	10.82	10.39-11.26	7.17	6.89-7.44
2010	124.18	122.6-125.77	121.24	119.95-122.52	11.00	10.57-11.44	7.64	7.36-7.92
Region								
North East	138.39	133.56-143.21	143.34	139.49-147.2	13.10	11.95-14.24	8.73	8.05-9.4
North West	162.52	160.3-164.74	153.74	152.06-155.42	11.67	11.23-12.1	7.91	7.65-8.17
Yorkshire & The Humber	160.95	157.44-164.47	146.74	144.12-149.36	11.02	10.34-11.69	7.23	6.83-7.63

East Midlands	129.12	126.01-132.23	121.78	119.37-124.18	10.41	9.73-11.08	6.58	6.18-6.97
West Midlands	157.17	154.67-159.68	145.14	143.26-147.02	10.87	10.39-11.35	7.43	7.14-7.72
East of England	120.29	118.41-122.17	108.10	106.7-109.5	10.20	9.78-10.63	6.65	6.39-6.9
South West	111.64	109.85-113.43	102.41	101.02-103.81	9.92	9.5-10.33	7.01	6.74-7.27
South Central	118.30	116.51-120.08	104.20	102.89-105.52	12.00	11.55-12.45	8.25	7.97-8.53
London	97.38	95.76-98.99	94.41	93.11-95.71	9.23	8.82-9.63	6.31	6.06-6.56
South East Coast	107.87	106.08-109.67	96.57	95.22-97.91	8.44	8.05-8.83	5.85	5.61-6.09
IMD Quintile								
0 (least deprived)	116.31	114.75-117.86	99.67	98.51-100.83	9.51	9.15-9.86	6.35	6.13-6.56
1	123.75	122.17-125.33	105.82	104.65-106.99	10.19	9.83-10.55	6.89	6.67-7.11
2	134.28	132.43-136.13	116.33	114.95-117.71	11.85	11.42-12.29	8.13	7.86-8.39
3	147.27	145.13-149.41	131.15	129.53-132.77	12.65	12.16-13.13	8.11	7.83-8.39
4 (most deprived)	188.05	184.91-191.19	177.79	175.26-180.31	15.35	14.68-16.02	10.77	10.36-11.19

*Age-standardised-UK mid-year population estimates, 2004.

LRTI: Lower respiratory tract infection

CAP: community-acquired pneumonia

IMD – index of multiple deprivation

Community-acquired LRTI incidence rates including COPD exacerbation codes overall and over time by sex, age, region of England and IMD quintile

	Sex			Age (years)					
	Overall	Male	Female	65-69	70-74	75-79	80-84	85-89	≥90
Overall	132.06	132.25	131.90	96.98	115.93	138.56	167.02	205.94	283.39
95% CI	131.57-132.54	131.52-132.99	131.25-132.54	96.45-97.51	115.29-116.57	137.75-139.37	165.9-168.14	204.19-207.69	280.01-286.77
1997	102.99	103.13	102.67	97.44	108.89	117.79	131.31	147.66	172.24
	101.93-104.05	101.51-104.75	101.27-104.07	95.29-99.6	106.47-111.31	115.03-120.55	127.7-134.92	142.68-152.64	164.92-179.55
1998	107.20	104.37	109.17	96.76	108.03	122.73	138.61	162.71	178.74
	106.17-108.24	102.82-105.92	107.79-110.55	94.75-98.77	105.77-110.3	120.12-125.35	135.02-142.2	157.73-167.68	171.71-185.77
1999	108.66	106.25	110.33	96.58	108.78	118.80	134.70	156.84	187.88
	107.68-109.64	104.77-107.72	109.02-111.64	94.72-98.44	106.67-110.88	116.44-121.16	131.41-137.98	152.35-161.32	181.16-194.61
2000	106.13	105.18	106.72	90.83	104.15	114.84	126.44	149.35	176.59
	105.2-107.05	103.78-106.59	105.48-107.95	89.15-92.52	102.22-106.08	112.63-117.05	123.54-129.33	145.22-153.48	170.45-182.72
2001	110.88	109.64	111.72	90.17	106.02	118.29	130.46	149.88	187.60
	109.94-111.82	108.22-111.06	110.48-112.96	88.54-91.79	104.12-107.91	116.09-120.49	127.66-133.25	145.85-153.91	181.45-193.74
2002	118.37	117.62	118.87	96.00	109.20	123.67	135.77	154.99	188.69
	117.41-119.33	116.15-119.09	117.59-120.14	94.35-97.64	107.32-111.09	121.43-125.9	133.01-138.52	150.88-159.1	182.64-194.74
2003	135.09	133.35	136.38	104.36	121.73	139.58	152.51	175.13	216.31
	134.04-136.13	131.77-134.94	134.99-137.77	102.67-106.06	119.73-123.73	137.19-141.97	149.62-155.4	170.65-179.61	209.79-222.82
2004	141.46	141.89	141.16	105.96	122.82	143.08	157.15	180.19	229.10
	140.38-142.54	140.23-143.54	139.75-142.58	104.28-107.65	120.83-124.81	140.68-145.49	154.25-160.04	175.67-184.71	222.39-235.81
2005	142.54	143.99	141.54	104.71	121.01	138.39	156.13	176.86	228.82
	141.46-143.63	142.31-145.67	140.11-142.96	103.05-106.37	119.05-122.97	136.05-140.72	153.24-159.02	172.59-181.12	222.12-235.52
2006	146.43	149.37	144.35	105.11	122.53	139.74	157.93	173.59	230.53
	145.31-147.54	147.64-151.11	142.9-145.81	103.45-106.78	120.56-124.51	137.39-142.09	155-160.87	169.5-177.67	223.75-237.31
2007	153.88	159.00	150.26	107.82	125.19	145.17	162.54	182.49	243.09
	152.72-155.04	157.18-160.81	148.77-151.76	106.13-109.5	123.19-127.19	142.76-147.58	159.52-165.55	178.35-186.63	235.95-250.24
2008	164.29	169.30	160.84	111.09	131.21	150.71	174.46	195.47	275.88
	163.08-165.51	167.38-171.21	159.26-162.42	109.37-112.81	129.14-133.29	148.2-153.21	171.26-177.65	191.14-199.79	267.89-283.87
2009	147.23	151.77	144.13	99.12	117.72	134.10	150.20	172.31	238.24
	146.09-148.38	149.97-153.57	142.64-145.62	97.52-100.72	115.77-119.68	131.74-136.46	147.26-153.14	168.3-176.32	230.87-245.62
2010	156.15	157.56	155.42	104.10	123.21	138.42	158.71	180.61	248.54
	154.95-157.36	155.7-159.43	153.84-157	102.44-105.76	121.17-125.25	135.97-140.87	155.63-161.79	176.39-184.84	241.15-255.93

	Region									
	North East	North West	Yorkshire & The Humber	East Midlands	West Midlands	East of England	South West	South Central	London	South East Coast
Overall	162.65	174.90	167.21	137.18	166.00	123.46	117.59	121.05	102.81	110.23
95% CI	158.54-166.76	173.13-176.68	164.47-169.94	134.72-139.65	164.01-168	122.01-124.91	116.15-119.02	119.68-122.42	101.56-104.06	108.86-111.6
1997	109.65	140.62	112.76	96.63	113.58	92.99	95.77	98.12	91.64	79.07
	102.93-116.38	137.14-144.09	108.22-117.3	92.18-101.07	109.95-117.22	89.97-96.01	92.28-99.25	94.42-101.82	88.38-94.9	75.86-82.28
1998	118.58	133.60	128.21	111.55	133.46	104.20	96.52	96.31	95.02	80.62
	111.46-125.71	130.36-136.85	123.42-133	106.88-116.23	129.72-137.21	101.14-107.26	93.28-99.76	92.99-99.63	91.8-98.25	77.45-83.79
1999	143.51	135.32	129.80	100.77	149.58	100.68	100.70	99.92	87.92	86.44
	135.39-151.62	132.15-138.49	125.04-134.57	96.56-104.97	145.66-153.49	97.82-103.54	97.69-103.72	96.91-102.92	85.02-90.82	83.39-89.49
2000	130.13	135.65	138.19	111.04	131.86	97.90	90.03	95.86	93.92	83.63
	122.9-137.36	132.52-138.78	133.29-143.08	106.54-115.54	128.4-135.33	95.16-100.65	87.28-92.78	93.18-98.53	91.06-96.78	80.83-86.43
2001	149.33	136.90	156.42	118.35	139.05	108.12	100.71	106.57	100.91	92.94
	141.3-157.37	133.83-139.98	151.21-161.64	113.73-122.97	135.46-142.63	105.23-111.01	97.75-103.67	103.8-109.34	98-103.82	90.01-95.86
2002	167.10	155.44	168.93	123.41	155.76	115.93	106.64	106.84	95.61	98.32
	158.32-175.89	152.15-158.72	163.53-174.32	118.67-128.16	151.89-159.62	112.94-118.93	103.59-109.69	104.05-109.64	92.84-98.39	95.37-101.27
2003	165.73	181.23	192.46	145.16	175.62	128.86	121.47	120.09	104.48	108.93
	156.9-174.56	177.58-184.89	186.47-198.46	139.87-150.46	171.46-179.78	125.65-132.07	118.18-124.76	117.09-123.08	101.52-107.43	105.84-112.01
2004	195.80	190.95	191.22	151.80	182.43	135.40	122.06	128.23	106.87	111.97
	185.9-205.71	187.16-194.74	185.2-197.24	146.2-157.4	178.12-186.74	132.08-138.71	118.78-125.34	125.09-131.36	103.84-109.9	108.84-115.1
2005	175.79	192.56	189.46	160.46	183.45	144.96	122.55	128.88	105.34	113.05
	166.5-185.09	188.73-196.39	183.43-195.48	154.6-166.33	179.07-187.83	141.47-148.45	119.27-125.83	125.71-132.04	102.35-108.33	109.91-116.18
2006	173.20	197.78	192.41	174.70	185.01	135.87	127.37	134.38	106.81	121.85
	163.9-182.5	193.83-201.73	186.09-198.73	168.42-180.98	180.59-189.44	132.43-139.32	124.02-130.71	131.12-137.65	103.76-109.86	118.6-125.1
2007	188.32	210.21	198.00	176.02	189.19	142.20	137.13	136.88	114.70	126.26
	178.73-197.9	206.04-214.37	191.39-204.61	169.53-182.52	184.62-193.75	138.54-145.86	133.6-140.65	133.65-140.1	111.55-117.84	122.94-129.59
2008	206.01	221.58	215.92	185.92	202.87	152.24	145.86	149.19	116.67	141.00
	195.69-216.33	217.24-225.93	208.29-223.55	178.92-192.93	198.08-207.66	148.3-156.18	142.18-149.54	145.8-152.58	113.49-119.86	137.43-144.58
2009	191.12	203.62	192.21	157.98	182.45	139.07	128.36	128.44	106.22	127.67
	181.11-201.14	199.47-207.77	184.23-200.19	151.3-164.66	177.96-186.93	135.25-142.88	124.95-131.77	125.37-131.52	103.23-109.2	124.34-131
2010	217.42	224.76	191.29	164.59	200.78	150.10	138.89	139.18	109.42	136.14
	206.39-228.44	220.26-229.25	183.14-199.45	156.3-172.87	195.98-205.58	145.95-154.24	135.26-142.52	135.96-142.4	106.4-112.43	132.65-139.64

	IMD Quintile				
	0 (least deprived)	1	2	3	4 (most deprived)
Overall	111.42	121.05	133.93	151.46	206.08
95% CI	110.26-112.59	119.84-122.26	132.49-135.37	149.75-153.17	203.4-208.76
1997	86.67	92.23	98.29	114.58	156.35
	83.82-89.52	89.49-94.96	95.21-101.38	111.09-118.07	151.51-161.19
1998	94.28	100.39	105.80	119.08	151.99
	91.49-97.08	97.64-103.14	102.75-108.85	115.65-122.51	147.41-156.56
1999	100.27	100.28	105.90	122.08	166.95
	97.62-102.92	97.68-102.88	103.03-108.78	118.75-125.41	162.21-171.7
2000	92.40	97.73	105.28	121.50	161.79
	90-94.79	95.3-100.15	102.54-108.02	118.29-124.72	157.25-166.32
2001	99.28	106.67	119.49	127.18	167.42
	96.81-101.75	104.16-109.18	116.57-122.41	123.95-130.41	162.83-172.01
2002	100.73	112.34	122.96	139.18	196.06
	98.26-103.19	109.77-114.91	120-125.93	135.78-142.58	190.97-201.14
2003	114.15	122.98	137.64	159.96	218.27
	111.49-116.81	120.27-125.69	134.45-140.83	156.22-163.69	212.76-223.78
2004	115.90	129.77	146.40	165.13	228.02
	113.23-118.56	126.97-132.57	143.05-149.74	161.28-168.98	222.28-233.77
2005	118.15	128.83	146.47	167.33	227.71
	115.45-120.84	126.05-131.61	143.1-149.85	163.41-171.25	221.89-233.52
2006	120.76	134.16	151.23	170.58	228.84
	118.04-123.49	131.31-137.02	147.79-154.67	166.58-174.58	222.93-234.74
2007	121.56	136.08	157.38	175.22	242.34
	118.84-124.28	133.2-138.95	153.86-160.9	171.14-179.3	236.17-248.51
2008	133.98	147.48	167.09	191.22	257.30
	131.11-136.86	144.45-150.5	163.42-170.77	186.87-195.58	250.8-263.8
2009	117.72	130.41	150.12	174.76	248.70
	115.09-120.34	127.62-133.2	146.68-153.56	170.6-178.91	242.21-255.2
2010	123.33	140.50	158.68	184.22	272.37
	120.58-126.07	137.55-143.46	155.06-162.3	179.85-188.6	265.41-279.32

Community-acquired pneumonia incidence rates overall and over time by sex, age, region of England and IMD quintile

CAP	Overall	Sex		Age (years)					
		Male	Female	65-69	70-74	75-79	80-84	85-89	≥90
Overall	7.99	8.60	7.53	2.81	4.31	6.94	12.05	21.84	41.94
95% CI	7.92-8.07	8.49-8.72	7.44-7.62	2.74-2.87	4.22-4.4	6.81-7.08	11.82-12.27	21.39-22.29	40.87-43.01
1997	7.34	7.87	6.95	2.59	4.48	6.85	12.38	19.98	34.65
	7.07-7.61	7.44-8.3	6.61-7.3	2.27-2.9	4.04-4.91	6.25-7.44	11.38-13.38	18.32-21.65	31.57-37.74
1998	7.69	7.67	7.71	2.88	4.30	7.31	12.98	20.98	35.54
	7.43-7.95	7.26-8.07	7.36-8.05	2.56-3.19	3.9-4.71	6.74-7.89	11.98-13.98	19.36-22.6	32.6-38.48
1999	7.36	7.71	7.10	2.89	4.36	6.72	11.59	19.80	33.25
	7.12-7.6	7.33-8.08	6.79-7.41	2.6-3.18	3.98-4.74	6.21-7.23	10.71-12.48	18.35-21.26	30.64-35.87
2000	6.30	6.70	6.00	2.16	3.40	6.11	9.37	17.70	27.31
	6.09-6.51	6.37-7.03	5.73-6.27	1.92-2.39	3.09-3.71	5.65-6.58	8.65-10.09	16.4-19	25.09-29.52
2001	6.83	7.45	6.38	2.40	4.16	5.96	9.47	18.27	29.51
	6.62-7.05	7.11-7.79	6.1-6.65	2.16-2.64	3.83-4.5	5.51-6.4	8.79-10.15	16.99-19.55	27.3-31.73
2002	6.92	7.48	6.50	2.43	3.88	6.43	9.90	17.48	27.95
	6.71-7.14	7.14-7.81	6.23-6.78	2.2-2.66	3.56-4.2	5.97-6.89	9.23-10.57	16.24-18.73	25.85-30.04
2003	7.81	8.34	7.41	2.71	3.85	7.01	11.03	20.31	32.86
	7.59-8.04	7.98-8.69	7.12-7.71	2.47-2.96	3.53-4.16	6.53-7.48	10.34-11.72	18.94-21.68	30.61-35.12
2004	7.93	8.33	7.62	2.68	4.23	7.05	11.10	18.42	33.16
	7.7-8.16	7.98-8.69	7.32-7.91	2.45-2.92	3.9-4.56	6.58-7.52	10.42-11.78	17.13-19.71	30.91-35.41
2005	7.84	8.22	7.54	2.90	4.38	6.36	10.75	17.45	30.04
	7.61-8.06	7.87-8.57	7.24-7.83	2.65-3.14	4.05-4.71	5.92-6.81	10.08-11.42	16.27-18.64	27.93-32.16
2006	8.16	8.75	7.70	2.87	4.38	6.73	11.35	16.57	30.57
	7.93-8.39	8.38-9.11	7.4-8	2.63-3.12	4.05-4.71	6.27-7.19	10.65-12.04	15.46-17.68	28.44-32.69
2007	8.49	9.21	7.92	2.88	4.54	7.00	11.07	18.37	30.74
	8.26-8.73	8.84-9.59	7.62-8.23	2.64-3.12	4.21-4.87	6.54-7.47	10.38-11.75	17.22-19.52	28.59-32.89
2008	9.91	10.66	9.30	3.57	5.03	7.77	13.41	20.92	36.12
	9.64-10.17	10.25-11.07	8.96-9.64	3.3-3.85	4.67-5.38	7.28-8.27	12.64-14.18	19.7-22.15	33.73-38.52
2009	9.52	10.44	8.77	3.18	5.00	7.44	12.81	20.28	34.84
	9.26-9.78	10.03-10.84	8.44-9.1	2.93-3.44	4.64-5.35	6.95-7.92	12.05-13.57	19.08-21.48	32.48-37.19
2010	10.06	10.82	9.44	3.61	4.87	7.84	13.37	21.49	34.97
	9.79-10.33	10.4-11.24	9.09-9.79	3.34-3.88	4.52-5.23	7.33-8.35	12.59-14.16	20.23-22.76	32.69-37.25

					Region					
CAP	North East	North West	Yorkshire & The Humber	East Midlands	West Midlands	East of England	South West	South Central	London	South East Coast
Overall	10.21	9.11	8.73	7.93	8.71	8.25	8.77	10.27	7.76	7.18
	9.61-10.81	8.88-9.33	8.37-9.09	7.58-8.28	8.45-8.96	8.01-8.48	8.51-9.02	10.01-10.54	7.53-8	6.95-7.4
1997	7.35	7.27	7.64	6.87	6.69	7.41	7.65	10.99	7.73	6.72
	5.8-8.9	6.55-7.99	6.53-8.74	5.73-8	5.87-7.5	6.6-8.22	6.7-8.6	9.79-12.18	6.83-8.64	5.83-7.62
1998	7.38	7.50	7.69	7.31	7.78	8.90	7.57	9.79	7.75	7.57
	5.79-8.96	6.79-8.21	6.61-8.77	6.19-8.42	6.95-8.61	8.05-9.74	6.7-8.43	8.77-10.8	6.87-8.64	6.63-8.51
1999	8.95	6.65	7.33	6.89	7.46	7.54	8.89	9.71	6.56	6.72
	7.18-10.72	6-7.29	6.29-8.38	5.87-7.9	6.67-8.24	6.81-8.28	8.05-9.73	8.82-10.6	5.81-7.31	5.91-7.54
2000	8.15	6.21	6.81	5.72	6.98	6.59	5.98	7.77	7.29	4.61
	6.52-9.78	5.6-6.81	5.82-7.8	4.78-6.65	6.27-7.7	5.92-7.25	5.31-6.64	7.06-8.49	6.55-8.04	3.98-5.24
2001	8.05	7.31	7.79	6.24	7.47	7.63	7.23	8.28	7.78	5.02
	6.4-9.7	6.67-7.95	6.74-8.83	5.28-7.19	6.72-8.21	6.93-8.34	6.49-7.96	7.57-8.99	7.03-8.53	4.38-5.66
2002	8.68	8.24	7.63	6.93	8.09	6.94	7.13	8.68	6.78	5.07
	6.94-10.43	7.57-8.9	6.62-8.64	5.92-7.94	7.31-8.87	6.28-7.6	6.41-7.86	7.94-9.41	6.09-7.46	4.45-5.7
2003	10.89	8.83	8.56	7.66	9.10	7.63	7.64	11.57	7.03	6.04
	8.89-12.88	8.13-9.52	7.46-9.67	6.59-8.73	8.27-9.93	6.94-8.33	6.9-8.38	10.71-12.44	6.32-7.74	5.37-6.71
2004	9.81	9.59	9.55	8.59	9.25	7.17	8.56	10.52	7.79	5.93
	7.91-11.71	8.86-10.32	8.37-10.74	7.42-9.76	8.41-10.1	6.5-7.84	7.78-9.34	9.69-11.34	7.03-8.55	5.28-6.59
2005	13.10	9.44	8.95	7.61	9.45	7.87	8.41	9.73	5.98	7.63
	10.87-15.34	8.71-10.16	7.79-10.1	6.5-8.72	8.58-10.32	7.16-8.58	7.64-9.18	8.94-10.53	5.33-6.64	6.88-8.38
2006	12.13	9.76	8.87	8.81	8.74	8.03	8.84	10.88	6.68	7.75
	9.96-14.29	9-10.51	7.67-10.06	7.59-10.03	7.91-9.58	7.29-8.77	8.06-9.63	10.03-11.73	5.97-7.38	7-8.49
2007	10.73	10.58	11.14	8.91	9.35	8.12	10.16	10.72	7.72	7.64
	8.74-12.72	9.78-11.37	9.73-12.54	7.64-10.17	8.47-10.23	7.35-8.9	9.31-11.01	9.9-11.54	6.97-8.47	6.89-8.38
2008	14.17	12.12	11.19	10.36	11.18	10.29	11.23	13.36	9.39	10.18
	11.81-16.53	11.26-12.99	9.64-12.74	8.92-11.8	10.21-12.15	9.38-11.19	10.33-12.13	12.44-14.28	8.56-10.22	9.31-11.06
2009	12.14	11.68	12.99	11.71	10.02	11.23	10.26	10.85	10.11	10.42
	9.93-14.35	10.82-12.53	11.11-14.86	10.1-13.33	9.1-10.94	10.26-12.2	9.4-11.11	10.03-11.67	9.25-10.97	9.54-11.3
2010	14.53	12.84	12.88	12.74	10.94	12.26	11.00	11.46	10.06	10.68
	12.06-17	11.93-13.76	10.96-14.79	10.7-14.78	9.96-11.91	11.21-13.32	10.1-11.91	10.62-12.3	9.2-10.91	9.77-11.59

317

CAP	IMD Quintile				
	0 (least deprived)	1	2	3	4 (most deprived)
Overall	7.58 7.38-7.77	8.35 8.15-8.55	9.93 9.68-10.17	10.08 9.82-10.35	12.78 12.4-13.15
1997	7.27 6.47-8.06	6.63 5.93-7.33	7.72 6.9-8.54	7.71 6.86-8.56	8.46 7.45-9.47
1998	6.89 6.17-7.61	7.49 6.78-8.2	8.80 7.96-9.63	8.54 7.68-9.4	10.47 9.37-11.57
1999	7.48 6.8-8.17	6.84 6.2-7.48	8.23 7.48-8.98	8.55 7.74-9.37	9.45 8.45-10.46
2000	5.61 5.06-6.17	6.10 5.54-6.67	7.32 6.65-7.99	7.53 6.81-8.26	8.87 7.93-9.81
2001	6.45 5.86-7.04	6.88 6.3-7.47	8.66 7.94-9.38	7.69 6.98-8.4	9.94 8.96-10.92
2002	6.61 6.03-7.2	7.15 6.56-7.74	8.28 7.58-8.98	8.43 7.69-9.17	10.82 9.8-11.85
2003	7.59 6.96-8.22	7.65 7.05-8.26	10.15 9.37-10.94	9.08 8.31-9.86	11.91 10.82-13.01
2004	7.71 7.08-8.34	8.41 7.77-9.04	10.63 9.81-11.44	8.92 8.14-9.69	12.79 11.64-13.95
2005	7.20 6.59-7.81	8.58 7.94-9.23	9.76 8.98-10.54	10.20 9.36-11.04	13.61 12.39-14.84
2006	7.99 7.35-8.63	8.48 7.84-9.11	10.14 9.34-10.93	10.81 9.93-11.68	13.50 12.26-14.74
2007	8.14 7.49-8.78	9.13 8.46-9.79	10.83 10-11.65	12.01 11.08-12.94	15.83 14.47-17.2
2008	9.65 8.94-10.35	10.62 9.9-11.34	13.28 12.35-14.2	14.35 13.32-15.38	17.82 16.34-19.31
2009	9.67 8.97-10.37	10.62 9.91-11.34	12.61 11.71-13.51	13.24 12.24-14.24	18.48 16.92-20.04
2010	9.29 8.59-9.99	11.33 10.57-12.08	13.31 12.37-14.26	14.11 13.05-15.17	20.35 18.68-22.01

Appendix E Alternative analysis of Chapter 5, presented as a paper published in the Journal of Clinical Epidemiology

London School of Hygiene & Tropical Medicine
Keppel Street, London WC1E 7HT
www.lshtm.ac.uk

Registry
T: +44(0)20 7299 4646
F: +44(0)20 7299 4656
E: registry@lshtm.ac.uk

LONDON
SCHOOL of
HYGIENE
&TROPICAL
MEDICINE

RESEARCH PAPER COVER SHEET

PLEASE NOTE THAT A COVER SHEET MUST BE COMPLETED FOR EACH RESEARCH PAPER INCLUDED IN A THESIS.

SECTION A – Student Details

Student	Elizabeth Millett
Principal Supervisor	Sara Thomas
Thesis Title	Improved incidence estimates from linked vs. stand-alone electronic health records

If the Research Paper has previously been published please complete Section B, if not please move to Section C

SECTION B – Paper already published

Where was the work published?	Journal of Clinical Epidemiology
When was the work published?	9 January 2016
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A

Have you retained the copyright for the work?*	Yes	Was the work subject to academic peer review?	Yes

If yes, please attach evidence of retention. If no, or if the work is being included in its published format, please attach evidence of permission from the copyright holder (publisher or other author) to include this work.

SECTION C – Prepared for publication, but not yet published

Where is the work intended to be published?	
Please list the paper's authors in the intended authorship order:	
Stage of publication	Choose an item.

SECTION D – Multi-authored work

For multi-authored work, give full details of your role in the research included in the paper and in the preparation of the paper. (Attach a further sheet if necessary)	I conceived the idea for this study. Sara Thomas obtained the funding, ethical approval and the data used in the analyses from CPRD. I developed the study design, supervised by Sara Thomas and with statistical advice from Bianca De Stavola. As for the Incidence paper, the Read and ICD-10 codelists for LRTI and pneumonia

were devised by Sara Thomas and two other clinical epidemiologists. Sara Thomas derived the hospitalisation Read codelist, which we each independently categorised into the eight hospitalisation groups. We compared categorisations and discussed those which did not match until we reached a consensus.

I designed the methods to derive illness-episodes, differentiate community- from hospital-acquired infections and person-time at risk, with input from Sara Thomas and contributions from Helen McDonald. Clinical advice was provided by Jennifer Quint and Liam Smeeth.

I conducted all data management, analysis, led the interpretation of results (with input from Sara Thomas and Bianca De Stavola) and I wrote the first draft of the paper. All co-authors contributed revisions, which I then incorporated. After peer-review, I further adapted the manuscript to include the reviewers' comments, with advice from all co-authors.

Student Signature: _____ Date: 24 /10 /2017

Supervisor Signature: _____ Date: 24/10/2017

320

Improved incidence estimates from linked vs. stand-alone electronic health records

Elizabeth R.C. Millett[a,*], Jennifer K. Quint[b], Bianca L. De Stavola[c], Liam Smeeth[b], Sara L. Thomas[a]

[a]Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
[b]Department of Non-communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
[c]Department of Medical Statistics, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK

Accepted 4 January 2016; Published online 9 January 2016

Abstract

Objective: Electronic health records are widely used for public health research, and linked data sources are increasingly available. The added value of using linked records over stand-alone data has not been quantified for common conditions such as community-acquired pneumonia (CAP).

Study Design and Setting: Our cohort comprised English patients aged ≥65 years from the Clinical Practice Research Datalink, eligible for record linkage to Hospital Episode Statistics. Stand-alone general practice (GP) records were used to calculate CAP incidence over time using population-averaged Poisson regression. Incidence was then recalculated for the same patients using their linked GP-hospital admission data. Results of the two analyses were compared.

Results: Over 900,000 patients were included in each analysis. Population-averaged CAP incidence was 39% higher using the linked data than stand-alone data. This difference grew over time from 7% in 1997 to 83% by 2010. An increasingly larger number of pneumonia events were recorded in the hospital admission data compared to the GP data over time.

Conclusion: Use of primary or secondary care data in isolation may not give accurate incidence estimates for important infections in older populations. Further work is needed to establish the extent of this finding in other diseases, age groups, and populations. © 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords: Pneumonia; Electronic health records; Data linkage; Aged; England/epidemiology; Cohort

1. Introduction

Electronic health records are extensively used in epidemiological research, because of their wide and detailed population coverage. It is increasingly possible to link electronic data sources to enhance available data. For example, linked primary and secondary care data provide more complete information on outcomes, enriched data on covariates such as patients' medical and therapeutic histories, and accurate timing of events such as hospitalizations. The value of linked over stand-alone data has been investigated for conditions such as cardiovascular events, asthma, diabetes, and upper gastrointestinal bleeding [1−4]. However, the potential benefits of linked data for examining the burden of important infectious diseases are unclear.

Community-acquired pneumonia (CAP) causes considerable morbidity among older individuals and can be treated in either primary or secondary care. Large-scale

321

What is new?

Key findings

- Use of linked primary-secondary care health data provided markedly higher incidence estimates of community-acquired pneumonia compared to stand-alone general practice (GP) records for the same group of English older adults.

- Comparison of the data sources revealed diverging incidence estimates over time, rising from 7% higher in 1997/98 to 83% higher in 2010/11 when using the linked data compared to the stand-alone GP data.

What this adds to what was known?

- The benefits of the use of linked electronic health records (compared to single data sources) have been demonstrated for conditions such as cardiovascular diseases; this is the first article to demonstrate the benefits for an important, common infection.

What is the implication and what should change now?

- Use of primary or secondary care data in isolation may not give accurate estimates of burden of disease for important infections in older populations.

- Further work is needed to establish if this trend is seen in other infections and diseases.

studies of CAP incidence trends have commonly used either stand-alone general practice (GP) records, potentially excluding patients who present to hospital if practices record hospitalized events suboptimally, or stand-alone hospital records which exclude cases treated in the community. Two recent studies used large linked GP and hospital data sets to assess disease burden of CAP but did not assess the added value of using the linked data [5,6].

We thus investigated the utility of linked primary/secondary care data in better determining trends in CAP disease burden in England among those aged ≥65 years by comparing incidence of CAP derived from stand-alone primary care data with that from linked primary-secondary care data. Each analysis used essentially the same cohort of patients over the same time period, using the same analytical approach.

2. Methods

The Clinical Practice Research Datalink (CPRD) is a nationally representative UK primary care dataset, containing a range of information including Read-coded diagnoses [1]. Hospital Episode Statistics (HES) contain inpatient records with ICD10-coded diagnoses, including admission and discharge dates. CPRD and HES records are linked at a patient-level for consenting English practices. By March 2011, CPRD contained >12 million patient records, with HES-linkage available for 65% of English CPRD practices (around 5% of the English population) [7].

Practices and patients joined CPRD throughout the study period, providing dynamic cohorts of patients. To ensure comparability of the two data sources, a near-identical group of patients were used in both analyses. Patients included in the study were eligible for record linkage, were aged ≥65 years, and contributed ≥1 day of follow-up. Follow-up started at the latest of the study start date (April 1, 1997), the patient's 65th birthday, the date the practice met CPRD quality standards or 28 weeks after patient registration (to exclude historical illnesses retrospectively reported) [6]. Follow-up ended at the earliest of the study end date (March 31, 2011), death, the practice's last data collection date, or the date the patient left the practice.

We have previously described in detail definitions for pneumonia illness episodes in CPRD and HES, using pneumonia and other lower respiratory tract infection records [6]. In brief, records for which pneumonia was recorded in CPRD (stand-alone and linked data) or as the admitting diagnosis (primary code of the first episode) in HES (linked data only) within 28 days of each other or of a record for lower respiratory tract infection were considered to be part of the same episode. The incident date of the episode was the date of the first of these pneumonia codes.

In both analyses, pneumonia illness episodes which started ≤14 days after a hospitalization were assumed to be hospital-acquired (HAP) and were excluded; episodes with no such hospitalization record were classed as community acquired. The method for defining hospitalizations, and thus distinguishing between CAP and HAP, differed between the two analyses. In the stand-alone CPRD data, hospitalization records were identified using Read codes and other relevant fields in the GP files. In the linked cohort, the 14-day period started at the discharge date of any hospital admission.

Patients were not considered "at-risk" of pneumonia during any pneumonia episode (CAP or HAP) or for 28 days after the last record in the episode, and this time was excluded from the denominator in both cohorts. A key difference in the linked data analysis was the capacity to also exclude the duration of any hospital admission and the subsequent 14 days from person-time at risk of a community-acquired infection and thus obtain more accurate denominator data. This was not possible in the stand-alone data as hospital admission, and discharge dates were not available.

Population-averaged Poisson models were used to calculate the incidence of CAP across clusters of CAP episodes per patient. Rates were calculated stratified by year, age group, and sex.

The financial year structure (April 1−March 31) was used to assign respiratory pathogens circulating over winter months to the same year.

In the linked data, whether patients had consulted with a GP (either face to face or by telephone) on the CAP incident date was examined using the "constype" field in the consultation file.

3. Results

The study population included 917,852 patients in the stand-alone data from 351 practices across England. The linked analysis included 916,128 (>99.8%) of these patients who had ≥1 day of follow-up after additionally excluding person-time in hospital. In both analyses 53% of patients were aged 65−69 years at start of follow-up and 56% were female. Using only GP records, we identified 31,575 CAP episodes during the study period. Using linked GP/hospital admission data identified 45,285 CAP episodes. In both analyses, >80% of patients had only one CAP episode during follow-up.

Incidence estimates using linked data were higher than those using stand-alone data. Overall, incidence was 39% higher using the linked data, and the difference increased markedly over time from 7% (6.18 vs. 5.77/1,000 person-years) in 1997/98 to 83% higher (10.13 vs. 5.54/1,000 person-years) in 2010/11 (Fig. 1). Although rates of CAP rose with age in both data sources, the relative increase in CAP estimates using the linked compared to GP stand-alone data was comparable for each age group, and so, the disparity was not attributable to a specific age group (data not shown). Incidence was higher in men than women in both analyses, but the divergence between estimates was observed in both sexes.

Because of the dynamic nature of the cohort, the number of patients contributing to each analysis increased over the
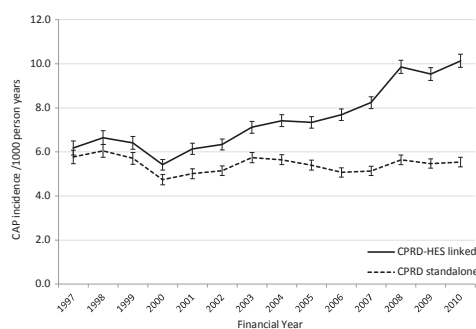

Fig. 1. Population-averaged incidence of CAP among older adults by data source over time. *Abbreviations:* CAP, community-acquired pneumonia; CPRD, Clinical Practice Research Datalink; HES, hospital episode statistics.

study period, increasing the person-time included. However, the increase in person-time within each analysis was similar (91% increase in linked vs. 93% in stand-alone data), whereas the increase in CAP episodes was substantially larger in the linked data (147% vs. 52% in stand-alone).

Between 1997 and 2010, the percentage of patients who had consulted with their GP on the day of the CAP diagnosis decreased from 82% to 43%. Over the same period, consultation with a GP for an LRTI in the 28 days before the CAP diagnosis decreased from 15% to 10%.

4. Discussion

Our investigation of incidence trends for a major infectious disease shows the benefits of using linked data. Use of primary care data alone yielded CAP incidence estimates that were 28% lower than estimates from linked primary/secondary care data. The divergence between estimates increased appreciably over the 14-year study period, and linked data estimates were 83% higher than those from stand-alone GP records by March 2011.

In the linked data analysis, we could refine estimated person-time at risk of community-acquired infection, by discounting the person-time patients were in hospital. However, it seems that the diverging estimates were attributable largely to the higher number of CAP episodes in the linked data. All pneumonias recorded in GP records are included in linked GP/hospital data, but pneumonias from hospital admissions are only included in stand-alone GP data if patients consulted their GP prehospitalization, or hospital diagnoses were retrospectively recorded by the patients' GP. Our analyses demonstrate that CAP identified in hospital is incompletely recorded by GPs, and this underrecording, coupled with the known increase in CAP hospitalizations in England over the study period, may explain the divergence we report [8]. Patients with CAP may have increasingly presented directly to Accident and Emergency Departments because of changes in GP service provision or perceived severity of illness, and the threshold for admission for these older patients may also have decreased. Both these scenarios are consistent with the larger increase in CAP episodes in the HES records and with decreasing consultations with a GP on the day of a CAP diagnosis. They also highlight that for conditions that can be treated both in the community and in hospital, changes to health services, patient, and clinician behavior could all result in marked underestimation of disease burden if single data sources are used.

Our analyses used large, nationally representative data sets containing ≥900,000 patients [9]. Overall validity of diagnoses in CPRD data has been shown to be high, although few studies have assessed the sensitivity of recording [10]. Over 99.8% of the same patients were included in both analyses, enabling examination of the

differences in CAP estimates due to the data source and methodology used. We are unaware of other studies that have assessed the added value of using linked vs. stand-alone data within the same population for estimating the burden of any infectious disease.

The two data sources use different coding systems, and changes to coding practices over time within each source are a further consideration. For example, "tentative" pneumonia codes such as "Influenza or pneumonia" (available in the Read but not ICD10 coding system) were not included in this study. Patients assigned a tentative pneumonia code by their GP and subsequently hospitalized with CAP would have been included in the linked data but not in the stand-alone data. However, to have contributed to the disparity, GPs would have needed to use these tentative diagnoses increasingly over time. Alternatively, if hospital physicians increasingly diagnosed or labeled older patients as having pneumonia, this would contribute to the divergent trends. We have no evidence that this occurred, but a clear understanding of trends in coding practices is essential for interpreting findings from both stand-alone and linked data.

In conclusion, use of primary or secondary care data in isolation may underestimate disease incidence for certain conditions, particularly those that can be treated in either care setting. Additionally, incomplete recording of events in UK stand-alone GP data limits its use in studies of the burden of pneumonia in older adults. Further work is needed to establish if this trend is seen in other diseases and age groups.

References

[1] Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. BMJ 2013; 346.

[2] Crooks C, Card T, West J. Defining upper gastrointestinal bleeding from linked primary and secondary care data and the effect on occurrence and 28 day mortality. BMC Health Serv Res 2012; 12:392.

[3] Morgan CL, Currie CJ, Stott NCH, Smithers M, Butler CC, Peters JR. Estimating the prevalence of diagnosed diabetes in a health district of Wales: the importance of using primary and secondary care sources of ascertainment with adjustment for death and migration. Diabet Med 2000;17:141—5.

[4] Vollmer WM, O'Connor EA, Heumann M, Frazier EA, Breen V, Villnave J, et al. Searching multiple clinical information systems for longer time periods found more prevalent cases of asthma. J Clin Epidemiol 2004;57:392—7.

[5] Snijders B, van der Hoek W, Stirbu I, van der Sande MAB, van Gageldonk-Lafeber AB. General practitioners' contribution to the management of community-acquired pneumonia in the Netherlands: a retrospective analysis of primary care, hospital, and national mortality databases with individual data linkage. Prim Care Respir J 2013; 22(4):400—5.

[6] Millett ERC, Quint JK, Smeeth L, Daniel RM, Thomas SL. Incidence of community-acquired lower respiratory tract infections and pneumonia among older adults in the United Kingdom: a population-based study. PLoS One 2013;8:e75131.

[7] Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. Ther Adv Drug Saf 2012;3:89—99.

[8] Bardsley M, Blunt I, Davies S, Dixon J. Is secondary preventive care improving? Observational study of 10-year trends in emergency admissions for conditions amenable to ambulatory care. BMJ Open 2013;3:e002007.

[9] Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data resource profile: clinical practice research datalink (CPRD). Int J Epidemiol 2015;44:827—36.

[10] Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010;69:4—14.

Appendix F Medline search strategy for the hospitalisation post-CAP literature review (Chapter 6)

1	exp Community-Acquired Infections/
2	(community adj1 acqui*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
3	exp Hospitalization/
4	hospitali$*.mp.
5	(hospital adj2 discharge).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
6	3 or 4 or 5
7	exp Prognosis/
8	(prognostic adj (score or rule or model)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
9	(predicti* adj (score or rule or model)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
10	(risk adj (score or rule or model)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
11	(predictive adj1 factor).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
12	exp Risk Assessment/
13	(decision adj (score or rule or model)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
14	exp Decision Support Techniques/
15	7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
16	exp Risk Factors/
17	(risk adj1 factor*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
18	exp pneumonia/
19	(pneumonit* or pneumonia).ti,ab.
20	bronchopneumonia.ti,ab.
21	pleuropneumonia.ti,ab.
22	exp Bronchopneumonia/ep, mo [Epidemiology, Mortality]
23	pneumonias.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
24	exp Aged, 80/ and over.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
25	exp Aged/

26	((old adj age*) or elderly or (senior adj citizen)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
27	community-acquired.mp.
28	exp Case Reports/
29	Animals/
30	Humans/
31	determinant*.mp.
32	16 or 17 or 31
33	18 or 19 or 20 or 21 or 22 or 23
34	1 or 2 or 27
35	24 or 25 or 26
36	33 and 34
37	15 or 32
38	35 and 36 and 37
39	6 and 38
40	29 not (29 and 30)
41	39 not 40
42	41 not 28
43	limit 42 to english language
44	limit 43 to yr="2008 -Current"

Appendix G Supplementary material from Paper 2 (Chapter 6).

Supplementary File B. Completeness of recording of smoking status over time

Year	Non smoker	Current smoker	Ex-smoker	Missing	Total
	Smoking status n (%)				
1998	62 (2.8)	350 (16)	1194 (54.6)	582 (26.6)	2188
1999	124 (5.1)	400 (16.4)	1319 (54.1)	595 (24.4)	2438
2000	116 (5)	378 (16.3)	1228 (52.8)	604 (26)	2326
2001	144 (5.2)	502 (18.1)	1499 (53.9)	634 (22.8)	2779
2002	157 (5.4)	561 (19.3)	1591 (54.8)	592 (20.4)	2901
2003	256 (7.8)	569 (17.3)	1887 (57.4)	575 (17.5)	3287
2004	472 (13.7)	663 (19.3)	1902 (55.4)	396 (11.5)	3433
2005	471 (13.7)	682 (19.9)	2044 (59.7)	229 (6.7)	3426
2006	582 (16.2)	685 (19.1)	2152 (60.1)	164 (4.6)	3583
2007	624 (16.2)	722 (18.7)	2435 (63.1)	80 (2.1)	3861
2008	783 (17.2)	850 (18.7)	2853 (62.7)	64 (1.4)	4550
2009	714 (16.4)	840 (19.3)	2757 (63.5)	31 (0.7)	4342
2010	710 (15.9)	909 (20.4)	2811 (63)	32 (0.7)	4462

Supplementary File C Multivariable models adjusted for additional groups of explanatory factors (continued over four pages)

		OR (95% CI) Adjusted for: (cumulative left to right)			
		Age, sex, year	& Individual co-morbidities	& frailty factors [†]	& medications/ vaccinations [¥]
Male		1	1	1	1
Female		0.71 (0.66 - 0.76)	0.8 (0.75 - 0.85)	0.81 (0.76 - 0.87)	0.82 (0.76 - 0.87)
Age (grouped)	65-69	1	1	1	1
	70-74	1.41 (1.24 - 1.61)	1.31 (1.16 - 1.49)	1.34 (1.17 - 1.52)	1.35 (1.18 - 1.53)
	75-79	1.45 (1.28 - 1.65)	1.39 (1.23 - 1.57)	1.43 (1.26 - 1.62)	1.44 (1.27 - 1.63)
	80-84	1.6 (1.41 - 1.81)	1.58 (1.39 - 1.78)	1.62 (1.43 - 1.84)	1.65 (1.45 - 1.87)
	85-89	1.33 (1.18 - 1.51)	1.44 (1.28 - 1.63)	1.52 (1.34 - 1.73)	1.56 (1.37 - 1.77)
	90+	0.88 (0.78 - 1)	1.02 (0.9 - 1.16)	1.13 (0.99 - 1.28)	1.18 (1.03 - 1.34)
Year of CAP (grouped)	1998-2000	0.52 (0.47 - 0.58)	0.56 (0.51 - 0.62)	0.55 (0.5 - 0.6)	0.54 (0.48 - 0.59)
	2001-2003	1	1	1	1
	2004-2006	1.91 (1.73 - 2.1)	1.81 (1.65 - 1.98)	1.99 (1.81 - 2.19)	1.94 (1.76 - 2.14)
	2007-2008	2.77 (2.48 - 3.1)	2.34 (2.1 - 2.6)	2.64 (2.36 - 2.95)	2.51 (2.23 - 2.82)
	2009-2010	4.38 (3.87 - 4.96)	3.63 (3.22 - 4.09)	4.16 (3.67 - 4.71)	3.86 (3.38 - 4.4)
Individual co-morbidities					
Ischaemic heart disease	pre-MI	1.64 (1.5 - 1.78)	1.38 (1.26 - 1.5)	1.35 (1.24 - 1.47)	1.34 (1.23 - 1.46)
	post MI	1.73 (1.55 - 1.92)	1.37 (1.23 - 1.52)	1.35 (1.21 - 1.5)	1.31 (1.17 - 1.46)
Congestive heart failure		1.68 (1.55 - 1.82)	1.22 (1.13 - 1.33)	1.25 (1.15 - 1.36)	1.25 (1.15 - 1.36)
Peripheral vascular disease		1.59 (1.43 - 1.77)	1.26 (1.13 - 1.4)	1.27 (1.14 - 1.41)	1.27 (1.14 - 1.41)
Dementia		0.45 (0.41 - 0.5)	0.56 (0.51 - 0.62)	0.69 (0.62 - 0.75)	0.71 (0.65 - 0.78)
Chronic lung disease		1.85 (1.71 - 1.99)	1.62 (1.51 - 1.74)	1.59 (1.48 - 1.71)	1.62 (1.47 - 1.78)
Connective tissue disease		1.53 (1.36 - 1.72)	1.34 (1.19 - 1.5)	1.35 (1.2 - 1.52)	1.21 (1.07 - 1.37)
Peptic ulcer		1.32 (1.18 - 1.49)	1.17 (1.04 - 1.31)	1.17 (1.05 - 1.32)	1.17 (1.05 - 1.32)

		OR (95% CI) Adjusted for: (cumulative left to right)			
		Age, sex, year	& Individual co-morbidities	& frailty factors [†]	& medications/ vaccinations [¥]
Liver disease	Mild	1.92 (1.22 - 3.03)	1.68 (1.08 - 2.61)	1.7 (1.09 - 2.65)	1.71 (1.09 - 2.66)
	Severe	2.12 (1.2 - 3.76)	1.81 (1.03 - 3.16)	1.78 (1.01 - 3.11)	1.8 (1.03 - 3.15)
Diabetes	Diabetes	1.4 (1.27 - 1.55)	1.26 (1.14 - 1.39)	1.28 (1.16 - 1.42)	1.27 (1.14 - 1.4)
	With complications	1.81 (1.5 - 2.19)	1.41 (1.17 - 1.7)	1.34 (1.1 - 1.62)	1.31 (1.08 - 1.59)
Cancer	Solid cancer	1.24 (1.13 - 1.36)	1.24 (1.13 - 1.37)	1.26 (1.14 - 1.38)	1.26 (1.15 - 1.38)
	Metastatic	1.6 (1.3 - 1.98)	2.55 (2.05 - 3.18)	2.55 (2.04 - 3.18)	2.46 (1.97 - 3.07)
Leukaemia/lymphoma		1.93 (1.53 - 2.42)	1.95 (1.56 - 2.44)	1.91 (1.52 - 2.39)	1.94 (1.55 - 2.43)
Severe renal disease		2.24 (2.01 - 2.49)	1.84 (1.66 - 2.05)	1.85 (1.66 - 2.05)	1.82 (1.64 - 2.03)
Cerebrovascular disease		0.85 (0.79 - 0.92)	0.91 (0.84 - 0.98)	0.98 (0.91 - 1.06)	0.98 (0.91 - 1.06)
Neurological disease		0.74 (0.66 - 0.83)	0.95 (0.85 - 1.05)	1.05 (0.95 - 1.18)	1.08 (0.97 - 1.21)
Disorders of the immune mechanism		3.4 (2.01 - 5.74)	2.35 (1.41 - 3.92)	2.49 (1.49 - 4.15)	2.49 (1.49 - 4.15)
Terminal illness		0.34 (0.29 - 0.4)	0.27 (0.23 - 0.32)	0.3 (0.25 - 0.36)	0.3 (0.25 - 0.35)
Frailty factors					
Recent carer		1.37 (1.15 - 1.63)		1.18 (0.97 - 1.42)	1.17 (0.97 - 1.41)
Living arrangements	Not recorded	1		1	1
	Lives alone	1.32 (1.11 - 1.58)		1.17 (0.97 - 1.41)	1.17 (0.97 - 1.4)
	Sheltered accommodation	1.21 (0.91 - 1.61)		1.09 (0.82 - 1.45)	1.11 (0.83 - 1.48)
	Residential Care	0.33 (0.3 - 0.37)		0.45 (0.4 - 0.5)	0.46 (0.41 - 0.51)
Visual impairment		1.3 (1.21 - 1.4)		1.1 (1.02 - 1.18)	1.1 (1.02 - 1.18)
Bedsore/ulcer		0.38 (0.32 - 0.45)		0.51 (0.43 - 0.6)	0.52 (0.44 - 0.61)
Low weight/poor nutrition		0.72 (0.65 - 0.79)		0.84 (0.77 - 0.93)	0.84 (0.77 - 0.93)
Incontinence/catheter		0.66 (0.59 - 0.73)		0.81 (0.73 - 0.9)	0.83 (0.75 - 0.92)

		OR (95% CI) Adjusted for: (cumulative left to right)			
		Age, sex, year	& Individual co-morbidities	& frailty factors [†]	& medications/ vaccinations [¥]
Medications					
Immunosuppressants (other than steroids) in last 120 days		1.65 (1.26 - 2.17)			1.1 (0.84 - 1.45)
Inhaled corticosteroids	None pre CAP	1			1
	Within 60 days	1.73 (1.57 - 1.9)			0.99 (0.88 - 1.12)
	Within 61-180 days	1.46 (1.24 - 1.71)			0.83 (0.69 - 0.98)
	Within 181-365 days	1.56 (1.21 - 2.03)			0.93 (0.72 - 1.21)
	More than 365 days ago	1.29 (1.12 - 1.5)			0.82 (0.7 - 0.95)
Antibiotics	None in previous 28 days	1			1
	In previous 1-7 days	1.02 (0.93 - 1.11)			1.04 (0.95 - 1.14)
	In previous 8-28 days	0.69 (0.63 - 0.76)			0.69 (0.63 - 0.76)
Statins in previous 6 months		1.57 (1.43 - 1.72)			1.12 (1.01 - 1.23)
Steroids in previous 90 days		1.8 (1.62 - 2)			1.42 (1.27 - 1.58)
Influenza vaccine	No vaccine pre CAP	1			1
	Vaccinated 14-365 days pre CAP	0.87 (0.8 - 0.96)			0.75 (0.68 - 0.83)
	Vaccinated last season	0.84 (0.75 - 0.94)			0.76 (0.67 - 0.85)
	Vaccinated 2-5 years pre CAP	0.88 (0.75 - 1.03)			0.82 (0.71 - 0.96)
	Vaccinated >5 years pre CAP	1.05 (0.81 - 1.34)			0.9 (0.7 - 1.15)
Pneumococcal vaccine	No vaccine pre CAP	1			1
	Vaccinated 14-365 days pre CAP	1.01 (0.88 - 1.17)			1.02 (0.89 - 1.18)
	Vaccinated 1-2 years pre CAP	1.07 (0.93 - 1.22)			1.02 (0.89 - 1.17)
	Vaccinated 2-5 years pre CAP	1.24 (1.13 - 1.36)			1.07 (0.97 - 1.18)
	Vaccinated >5 years pre CAP	1.57 (1.42 - 1.74)			1.16 (1.04 - 1.29)

		OR (95% CI) Adjusted for: (cumulative left to right)			
		Age, sex, year	& Individual co-morbidities	& frailty factors [†]	& medications/ vaccinations [¥]
Factors not included in later models					
Hemiplegia		0.85 (0.72 - 1.01)			
Mild renal disease		0.97 (0.7 - 1.34)			
Self-care problems		0.79 (0.57 - 1.08)			
Anxious/depressed		1.07 (0.95 - 1.2)			
Mobility issues		1.00 (0.87 - 1.15)			
Tired		0.94 (0.83 - 1.07)			
History of falling		1.06 (0.97 - 1.16)			
Excessive alcohol consumption	Any excess alcohol code	1.10 (0.93 - 1.29)			

† frailty factors: recent carer, place of residence, vision problems, bed ulcer, underweight/nutritional replacement, incontinence/catheter.

¥ medication/vaccination: Immunosuppressants (not steroids), steroids, inhaled steroids, statins, antibiotics in previous 28 days, influenza vaccine

All categorised as in (main text) Table 1.

Supplementary File D. Distribution of smoking status (where not missing) 2007-2010 and univariable, minimally adjusted, and adjusted ORs for hospitalisation. n=17008

	Hospitalised n (%)	Not hospitalised n (%)	Total	Unadjusted OR	Minimally adjusted (age, sex, year)	Adjusted co-morbidities
Non smoker	2260 (79.8)	571 (20.2)	2831	1	1	1
Current smoker	2969 (89.4)	352 (10.6)	3321	2.93 (2.34 - 3.67)	2.83 (2.25 - 3.56)	1.96 (1.58 - 2.44)
Ex-smoker	9386 (86.5)	1470 (13.5)	10856	2.05 (1.74 - 2.43)	1.88 (1.59 - 2.23)	1.37 (1.17 - 1.61)

Appendix H Medline search strategy for the longer-term post-CAP mortality literature review (Chapter 7)

1	exp Community-Acquired Infections/
2	(community adj1 acqui*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
3	exp Mortality/
4	mortality.mp.
5	death.mp.
6	3 or 4 or 5
7	exp Prognosis/
8	(prognostic adj (score or rule or model)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
9	(predicti* adj (score or rule or model)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
10	(risk adj (score or rule or model)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
11	(predictive adj1 factor).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
12	exp Risk Assessment/
13	(decision adj (score or rule or model)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
14	exp Decision Support Techniques/
15	7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
16	exp Risk Factors/
17	(risk adj1 factor*).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
18	exp pneumonia/
19	(pneumonit* or pneumonia).ti,ab.
20	bronchopneumonia.ti,ab.
21	pleuropneumonia.ti,ab.
22	exp Bronchopneumonia/ep, mo [Epidemiology, Mortality]
23	pneumonias.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
24	exp Aged, 80/ and over.mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
25	exp Aged/

26	((old adj age*) or elderly or (senior adj citizen)).mp. [mp=title, abstract, original title, name of substance word, subject heading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier]
27	community-acquired.mp.
28	exp Case Reports/
29	Animals/
30	Humans/
31	determinant*.mp.
32	16 or 17 or 31
33	18 or 19 or 20 or 21 or 22 or 23
34	1 or 2 or 27
35	24 or 25 or 26
36	33 and 34
37	15 or 32
38	35 and 36 and 37
39	6 and 38
40	29 not (29 and 30)
41	39 not 40
42	41 not 28
43	limit 42 to english language
44	limit 43 to yr="2011 -Current"

Appendix I Mortality rates (from 1st April 2004 to 31st March 2011) in patients aged ≥65 years who were eligible and

contributed to CPRD-ONS linked data (Chapter 7)

Age group	Women				Men			
	Mortality rate/1000 person-years	Mortality risk in CPRD-ONS linked population over time period (%)			Mortality rate/1000 person-years	Mortality risk in CPRD-ONS linked population over time period (%)		
		1-7 days	8-30 days	31-365 days		1-7 days	8-30 days	31-365 days
65-69	12.2 (11.7-12.8)	0.02	0.07	1.11	18 (17.3-18.8)	0.03	0.11	1.64
70-74	16.9 (16.5-17.3)	0.03	0.10	1.53	26.3 (25.8-26.9)	0.05	0.16	2.38
75-79	29.7 (29.1-30.2)	0.06	0.18	2.68	43.7 (42.9-44.5)	0.08	0.26	3.92
80-84	53.2 (52.4-54.1)	0.10	0.32	4.75	76.5 (75.2-77.8)	0.15	0.46	6.76
85-89	96.7 (95.3-98.2)	0.19	0.58	8.47	127.1 (124.8-129.4)	0.24	0.76	10.98
90+	197.9 (195.1-200.6)	0.38	1.19	16.56	222.9 (218-228)	0.43	1.33	18.45