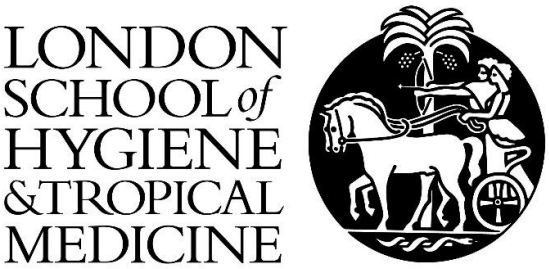


The epidemiology of infections among older people with diabetes mellitus and chronic kidney disease

HELEN ISABEL MCDONALD

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy of the

University of London

April 2015

Department of Non-Communicable Disease Epidemiology

Faculty of Epidemiology and Population Health

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by Kidney Research UK

Declaration

I, Helen McDonald, 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.

Signed

H McDonald

Helen McDonald

Date

2 April 2015

" Recognizing that we have the kind of internal environment we have because we have the kind of kidneys that we have, we must acknowledge that our kidneys constitute the major foundation of our physiological freedom... Superficially, it might be said that the function of the kidneys is to make urine; but in a more considered view one can say that the kidneys make the stuff of philosophy itself."

Homer W. Smith

From Fish to Philosopher (1953)

Plain language summary

People with kidney disease are at greater risk of dying from infection than people without kidney disease. This study investigated the relationship between kidney disease and infections among older people with diabetes. I used anonymised healthcare records for 219,145 patients, combined from primary care, admissions to hospital, and death certificate records.

Older people with diabetes had frequent chest and urinary tract infections. Chest and bloodstream infections were more common among people with kidney disease. The relationship between kidney disease and frequency of infection was stronger for bloodstream infection than for chest infection. Protein in the urine marked an increased risk of infection separately from the other standard marker of kidney disease, the estimated filtering rate of the kidneys.

In general, vaccines provide less protection for patients with kidney disease. A single 'flu or pneumococcal vaccine did not seem to offer effective protection against the burden of chest infections for older people with diabetes, whether or not they had kidney disease.

After being diagnosed with pneumonia or bloodstream infection, patients with severe kidney disease had a higher risk of dying than patients without kidney disease, but this did not seem to be true for patients with mild or moderate kidney disease.

Abstract

This thesis describes the epidemiology of community-acquired infections among older people with diabetes without a history of renal replacement therapy, according to markers of chronic kidney disease (CKD): proteinuria and reduced estimated glomerular filtration rate (eGFR). The thesis uses linked electronic health records from primary and secondary care, and mortality records.

Among a cohort of 219,145 patients with diabetes aged ≥65 years there was a high burden of community-acquired infection: lower respiratory tract infections (LRTIs) having the highest crude rate (152.7/1,000 years) followed by urinary tract infections (male 51.4, female 147.9/1,000 years). All-cause 28-day mortality was 32.1% for pneumonia (as a subset of LRTI) (3,115/9,697) and 31.7% for sepsis (780/2,461). Reduced eGFR was associated with a strong and graded increased risk of community-acquired LRTI, pneumonia and sepsis incidence, after adjustment for co-morbidities, smoking status and characteristics of diabetes mellitus. The effect sizes were larger for sepsis than pneumonia, and for pneumonia than LRTI. Proteinuria was a marker of increased risk of infection incidence independently of eGFR, for LRTI (rate ratio 1.07: 95%CI 1.05–1.09), pneumonia (1.26:1.19–1.33), and sepsis (1.33:1.20–1.47), after adjustment for co-morbidities, smoking status and characteristics of diabetes.

Advanced CKD (eGFR<30ml/min/1.73m^2) was associated with 28-day mortality following community-acquired pneumonia (risk ratio=1.27:95%CI 1.10–1.47) and sepsis (RR=1.42:1.10–1.84) compared to eGFR≥60 ml/min/1.73m^2), adjusted for age, sex, socio-economic status, smoking status and co-morbidities. Lesser reductions in eGFR and proteinuria were not associated with mortality.

The protective effects of pneumococcal vaccine against community-acquired pneumonia appeared to wane swiftly. There was scant evidence for any impact of influenza vaccination against the total burden of community-acquired LRTI.

This study allows patients, clinicians and public health planners to quantify infection risks among older people with diabetes according to CKD status. Further research could explore mechanisms and prevention strategies, including enhanced vaccination schedules.

Acknowledgements

I have been very lucky to be work under the guidance of my supervisors, Dr Dorothea Nitsch and Dr Sara Thomas. I am very grateful to them both for their advice, stimulating discussions, and support throughout this project.

I would like to thank all members of the Electronic Health Research group at LSTHM, for sharing their expertise and encouragement. I am particularly indebted to Elizabeth Millett, for so generously sharing her methods, skills and time.

I am grateful to Kidney Research UK, for supporting me to conduct this project: not just financially, but also for interesting discussions with patients and researchers at Kidney Research UK Fellows' Days. I hope this project will be a contribution towards the pursuit of health for people with, or at risk of, kidney disease.

I am grateful for the guidance I received from my other advisory committee members, Professor Simon de Lusignan and Professor Alan Sinclair, and from my upgrading examiners, Professor Liam Smeeth and Professor David Wheeler.

Finally, I would like to express my gratitude to my husband Donald, without whose patience and good humour during my studentship it would have been a lot less fun.

Funding

Helen McDonald was supported by a three-year research studentship from Kidney Research UK (ST2/2011) to undertake this study.

Table of Contents

List of tables

Tables in Appendices

List of figures

List of abbreviations

ACR	Albumin: creatinine ratio
AKI	Acute kidney injury
ARB	Angiotensin receptor II blocker
BMI	Body mass index
BNF	British National Formulary
CI	Confidence interval
CKD	Chronic kidney disease
COPD	Chronic obstructive pulmonary disease
CPRD	Clinical Practice Research Datalink
EHR	Electronic health records
ESRD	End-stage renal disease
eGFR	Estimated glomerular filtration rate
GFR	Glomerular filtration rate
GP	General practitioner
HES	Hospital Episode Statistics
ICD-9	International Classification of Diseases, Ninth Revision
ICD-10	International Classification of Diseases, Tenth Revision
IMD	Index of Multiple Deprivation
IQR	Interquartile range
K/DOQI	Kidney Disease Outcomes Quality Initiative
KDIGO	Kidney Disease: Improving Global Outcomes
LRTI	Lower respiratory tract infection

MI	Myocardial infarction
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NSAID	Non-steroidal anti-inflammatory drug
ONS	Office for National Statistics
OR	Odds ratio
PCR	Protein: creatinine ratio
PY	Person-years
RR	Rate ratio or risk ratio (specified in context)
RRT	Renal replacement therapy
SUS	Secondary Uses Service
TIA	Transient Ischaemic Attack
UK	United Kingdom
US	United States of America
UTI	Urinary tract infection
QOF	Quality and Outcomes Framework

BACKGROUND SECTION

This thesis uses routinely-collected electronic health records to investigate the epidemiology of acute, community-acquired infections according to markers of chronic kidney disease (CKD) prior to end-stage renal disease, among older people with diabetes mellitus.

This Background section introduces the study question and sets out the thesis aims and objectives. **Chapter 1** outlines the general background of acute, community-acquired infections as a cause of morbidity among older people, and the epidemiology of chronic kidney disease (CKD). The rationale for studying infections among older people with diabetes according to CKD status is presented and the aims and objectives of the thesis are described.

Chapter 2 presents a systematic review of the association between CKD and community-acquired infection incidence.

Chapter 1. General background

1.1 Community-acquired infection in older age

Community-acquired infections are common among older adults, causing a high burden of morbidity.[1-5] They are an important cause of mortality at older ages: pneumonia is the second commonest cause of death in people aged ≥75 years in England.[6]

The UK population is ageing. The proportion of the UK population aged ≥65 years is predicted to rise from 17% in 2010 to 23% in 2035. The fastest increase has been among the 'oldest old'. The number of people aged ≥85 years doubled from 0.7 million in 1985 to 1.4 million in 2010, and is predicted to reach 3.5 million (5% of the UK population) by 2035.[7]

Hospitalisations for infection are rising even faster than the population is ageing: age-standardised hospital admission rates for community-acquired pneumonia and urinary tract infections (UTIs) more than doubled between 2001/2001 and 2010/2011 in England.[8] The cost of hospitalisations was estimated at £235 million for pneumonia and £316 million for UTI in 2010/2011[8]. The increase in pneumonia hospitalisations has been most marked among older adults.[9] The driving factors behind this rise in admissions for community-acquired infections are not currently well understood but it does not appear to be purely due to lower thresholds for hospital admission, as the rising incidence of community-acquired LRTI remains when diagnoses in primary and secondary care are combined.[5] Suggested explanations include population ageing among the 'oldest old', together with higher prevalence of co-morbidities such as diabetes mellitus.[8, 9]

1.2 Chronic kidney disease

1.2.1 Definition of chronic kidney disease

Chronic kidney disease (CKD) is an impairment of kidney function or structure which persists for at least 3 months.[10] Kidney function is described by the glomerular filtration rate (GFR), the rate at which the glomerular capillaries in the kidney filter waste products such as creatinine. GFR is usually estimated from serum creatinine measurements adjusted for age, sex and ethnicity.[11, 12] Other evidence of kidney damage may include haematuria, structural abnormalities, or persistent protein in the urine (proteinuria).

Classification of CKD has since 2002 been based upon a 5-level staging of function (using two GFR estimations at least 3 months apart), and evidence of proteinuria.[13] Classification of CKD has evolved over the study period, and this is discussed in detail in **5.1.1**. The classification to which this thesis will refer most regularly is that recommended by the National Institute for Health and Care Excellence (NICE) in 2008, which was the dominant classification in UK clinical practice by the end of the study period (**Table 1.1**).[11]

CKD may progress to kidney failure, which is usually treated with renal replacement therapy: either kidney transplant or dialysis, in which waste products are filtered and removed from the blood (haemodialysis) or via the peritoneal cavity (peritoneal dialysis). In 2009, 2% of patients with CKD in England were receiving renal replacement therapy: of this group, approximately half were renal transplant recipients, with the majority of the rest treated with haemodialysis, and 8% with peritoneal dialysis.[14] Patients receiving renal replacement therapy are also referred to as having end-stage renal disease (ESRD).[13] In this thesis, the term end-stage renal disease will be used to identify patients receiving renal replacement therapy, and chronic kidney disease (CKD) will refer to patients with CKD not receiving renal replacement therapy, unless otherwise specified.

Table 1.1: NICE 2008 classification of chronic kidney disease (CKD)

CKD stage	GFR[1]	Evidence of kidney damage also required[2]
1	≥ 90	Yes
2	60-89	Yes
3A	45-59	
3B	30-44	
4	15-29	
5	< 15	

Based on the National Institute for Health and Care Excellence 2008 guidelines[11]

1. Glomerular filtration rate (ml/min/1.73m^2)

2. Persistent proteinuria, albuminuria or haematuria, or structural abnormalities

1.2.2 Chronic kidney disease as a public health problem

As CKD is usually asymptomatic until quite severe, it is often unrecognized, and estimates of prevalence vary. Estimates of the prevalence of CKD stages 3-5 among adults in England range from 4.3% to 8.5%.[15-18] This rises steeply with age: the 2009/2010 Health Survey for England identified stage 3–5 CKD among 29% of men and 35% of women aged ≥75 years.[16] Other risk factors for CKD include female sex, hypertension, diabetes mellitus, smoking and overweight at a younger age, many of which are modifiable.[12, 19] Among an ageing population with a rising prevalence of diabetes mellitus and hypertension, the

prevalence of CKD may be increasing but this is difficult to distinguish from increasing diagnosis and recording.[20, 21]

CKD is associated with a high burden of morbidity, mortality and health service use.[21, 22] The cost of CKD to the English NHS was recently estimated at £1.45 billion (1.3% of the total budget).[14] Even at early stages, CKD is associated with reduced quality of life, more frequent hospital admission and higher mortality compared to normal kidney function.[21, 23]

Older people with CKD are more likely to die of other causes than to develop end-stage renal disease.[22] Much of the burden of CKD is due to its association with non-renal adverse outcomes, such as cerebrovascular disease (causing stroke and cognitive impairment) and cardiovascular disease, which accounts for 58% of deaths among patients with CKD.[22-24]

1.2.3 CKD and infections

Infection is an important cause of morbidity and mortality among patients with ESRD. Patients with ESRD have higher rates of mortality caused by sepsis and pulmonary infections, and higher rates of infection-related hospitalisation, than the general population.[25-27] Among patients with ESRD in the US, the second commonest recorded cause of death after cardiac arrest is septicaemia.[28]

The association between ESRD and infection is partly driven by renal replacement therapy, which carries specific risks for infection. For example, patients who have received a kidney transplant require life-long immunosuppressive medication, while dialysis necessitates vascular or peritoneal access which disrupts the cutaneous barrier to infection.[27] However, the side effects of treatment do not fully account for the burden of infection in ESRD: the HEMO study found that only 23% of infection-related hospitalisations among haemodialysis patients in the US were related to vascular access.[29] It is possible that the association between CKD and infection is also present at earlier stages of CKD, prior to ESRD.

Patient characteristics which pre-dispose to infection are associated with all stages of CKD, including older age, high prevalence of co-morbidities and exposure to infectious agents from frequent healthcare attendance.[27]

Patients with ESRD and earlier stages of CKD are known to have a reduced response to some vaccinations. This is not only relevant as a risk factor for vaccine-preventable

infections, but also suggests that CKD may itself cause underlying impairment of the adaptive immune system which could increase incidence of infections in general even at early stages of CKD.[30, 31] A causal relationship between CKD itself and infection incidence or prognosis is plausible. There are multiple potential mechanisms for CKD to alter cell-mediated and humoral immune system function (such as malnutrition, hypoalbuminaemia, anaemia, complement loss, disrupted calcium regulation and vitamin D insufficiency, chronic renal inflammation, and immunosuppressive therapy for renal disease) which are not limited to patients receiving renal replacement therapy.[27, 32]

Even at early stages, CKD is associated with a higher mortality rate than among the general population, part of which is attributed to infection, and with more frequent infection-related admission to hospital.[33-35] Several expert narrative reviews have agreed that an association between CKD prior to ESRD and infection is plausible or even likely, but that the clinical epidemiology of such an association is insufficiently characterized to establish this at present.[27, 36-39]

1.3 Diabetes mellitus

Diabetes mellitus is a common endocrine disorder in which there is an insufficiency of, or resistance to, the hormone insulin, which regulates blood glucose levels. The estimated number of adults in England with diabetes mellitus was 3.1 million in 2010 and is predicted to rise to 4.6 million by 2030. Diabetes is more common among men, people with South Asian or Black ethnicity, and older adults. The estimated prevalence of diabetes among adults in England aged >75 years is 16.5% (95% CI 12.3–22.0%).[40]

Diabetes is associated with considerable morbidity and mortality: at 50 years old, a diagnosis of diabetes reduces life expectancy by 6 years.[41] Diabetes causes macrovascular complications such as cardiovascular disease and stroke, and microvascular complications such as retinopathy and nephropathy (diabetic kidney disease). Diabetes accounts for approximately 10% of UK health spending, and this is forecast to rise to 17% by 2035/2036. The cost of treating diabetic complications is £7.7 billion, and predicted to increase to £13.5 billion by 2035/6.[42]

1.3.1 Diabetes and CKD

Diabetic nephropathy is a major cause of CKD: diabetes is the commonest cause of CKD among patients requiring renal replacement therapy.[43] Patients with diabetes may also experience other causes of renal disease. In total estimates of the prevalence of stage 3–5

CKD among adults with diabetes in the UK range from 18% to 31%.[44-46] Death from renal disease is three times more common among patients with diabetes than patients without.[41]

1.3.2 Diabetes and infection

Diabetes has long been believed to increase susceptibility to infection (and certain rare infections occur almost exclusively among patients with diabetes),[47] but the epidemiology of infection among patients with diabetes was until recently surprisingly under-determined.[47, 48]

Diabetes is a risk factor for hospitalisation and mortality from infection.[41, 49, 50] This appears to be partly driven by an increased risk of hospitalisation and of death following infection onset.[50] However, an association between diabetes and infection diagnosed in primary care has also been observed among the general adult population, suggesting diabetes is likely to be a risk factor for infection incidence as well as severity.[50-52]

The association between diabetes and infection may be modified by age. A large population-based case-control study in Denmark found that the relative risk of hospitalisation with pneumonia for patients with diabetes compared to people without diabetes was considerably stronger among patients aged <40 years (adjusted RR 3.21: 95%CI 2.51–4.12) than those aged 65–79 (adjusted RR 1.22: 1.15–1.29) or ≥80 years (adjusted RR 1.11: 1.05–1.18).[49] Studies of risk factors for community-acquired infection among older adults have not been powered for precise estimates of infection incidence or risk ratios among the subgroup with diabetes.[53-57] Among older people with diabetes, data on the burden of infection from a community or primary care perspective, or risk factors for community-acquired infection, are scarce.

1.4 Thesis rationale, aims and objectives

1.4.1 Thesis rationale

Community-acquired infections are responsible for a large burden of morbidity and mortality among older people. This is a growing public health problem: not only is the UK population ageing, but hospitalisation rates for pneumonia and urinary tract infections are rising even after standardisation for age. One factor potentially driving the increasing incidence of infection-related hospitalisation could be the rising prevalence of co-morbidities such as diabetes and CKD.

CKD is common among older people, and in an ageing population the prevalence is expected to increase. Infection is an important cause of morbidity and mortality among patients with end-stage renal disease, and this is at least partly due to the immunosuppressive effects of renal replacement therapy. The majority of patients with CKD do not have, nor progress to, ESRD. Patients with earlier staged of CKD also have higher rates of infection-related hospitalisation and mortality than the general population. However, the precise relationship between CKD and infection is unclear.

Older people with diabetes are an important population in which to understand the epidemiology of CKD and infection. Diabetes is a risk factor for infection-related hospitalisation and mortality. The population of older people with diabetes is large and growing, with a high prevalence of CKD. Any role of CKD in increasing infection risk among the diabetic population would be of clinical and public health significance.

There are also epidemiological advantages to studying this population. Older age and diabetes are important *a priori* confounders of any association between CKD and infection: restricting the study population to older people with diabetes reduces confounding. Patients with diabetes are also regularly monitored for CKD, which should ensure reasonable ascertainment of CKD status from routinely-collected electronic health records.

An observational study of the epidemiology of infections according to CKD status is not well-suited to establishing the precise mechanisms which underlie any causal relationship between CKD and infection – but studying a focused question may still lead to a better understanding of a causal relationship and potential underlying mechanisms for further research. Studying patients prior to ESRD excludes any association resulting purely from the immunosuppressive effects of renal replacement therapy. Studying community-acquired infections allows identification of any inherent association of CKD with infection separately

from increased infection resulting from frequent hospital attendance. Thorough adjustment for co-morbidities may also clarify whether any association exists independently of co-morbidities.

Quantifying the risk of infection among older people with diabetes, overall and according to CKD status, would itself be valuable information for older people with diabetes and their clinicians, and also for planning health-care provision for this growing population, and health economics analyses of the impact of CKD.

Identifying whether and to what extent early stages of CKD are associated with increased risk of infection incidence, or greater severity of infections, could help ensure efforts at preventing excess infection-related mortality are targeted appropriately. Given the generally reduced response to vaccination among patients with ESRD, it is particularly important to identify whether vaccines are effective at preventing common infections among patients with earlier stages of CKD.

1.4.2 Aims and objectives

Among a cohort of people aged ≥65 years with diabetes mellitus, and using large, linked electronic health records, this thesis aims to describe:

- the incidence of, hospitalisation with and mortality from of acute community-acquired infections that are common (urinary tract and lower respiratory tract infections) or severe (pneumonia and sepsis); and
- the association between incidence of community-acquired infection and chronic kidney disease (excluding patients with a history of renal replacement therapy); and
- pneumococcal and influenza vaccine effectiveness according to stage of chronic kidney disease (excluding patients with a history of renal replacement therapy); and
- the association between short-term mortality following community-acquired infection and chronic kidney disease (excluding patients with a history of renal replacement therapy).

The objectives are detailed in **Table 1.2**.

Table 1.2: Thesis objectives

Objective	Study design	Population	Patients with a history of RRT	Infections	CKD markers	Effect measure
1. Describe the literature on the association of long term kidney disease with incidence of acute, community-acquired infection.	Systematic review	Adults in high-income countries.	Could be included.	Community-acquired UTI, LRTI, sepsis and CNS infections.	Any long-term kidney disease.	Any relative risk ratio.
2. Describe the burden of morbidity and mortality from acute community-acquired infections among older people with diabetes mellitus.	Cohort study	Patients aged ≥65 years with diabetes mellitus.	Not excluded.	Acute community-acquired UTI, LRTI (with pneumonia as a subset), and sepsis.	—	Incidence rates, 28-day and 90-day all-cause hospitalisation and mortality.
3. Describe the association between CKD and community-acquired infection incidence.	Cohort study	Patients aged ≥65 years with diabetes mellitus and a serum creatinine result available.	Excluded.	Acute community-acquired LRTI (with pneumonia as a subset), and sepsis.	eGFR and history of proteinuria.	Incidence rate ratios.
4. Describe pneumococcal and influenza vaccine effectiveness, according to stage of CKD.	Cohort study	Patients aged ≥65 years with diabetes mellitus and a serum creatinine result available.	Excluded.	Acute community-acquired LRTI.	eGFR and history of proteinuria.	Vaccine effectiveness.
5. Describe the association between CKD and short-term mortality following community-acquired infection.	Cohort study	Patients with community-acquired pneumonia or sepsis aged ≥65 years with diabetes mellitus, a serum creatinine result, and mortality data linkage available.	Excluded.	Acute community-acquired pneumonia and sepsis.	eGFR and history of proteinuria.	Risk ratios.

UTI, urinary tract infection; LRTI, lower respiratory tract infection; CNS infections, central nervous system infections; RRT, renal replacement therapy; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate

1.4.3 Organisation of the thesis

The thesis comprises 11 chapters, which are grouped into Background, Methods, Results and Discussion sections. **Chapter 1** introduced the study question, aims and objectives. Each thesis objective was addressed with a separate study, and **Table 1.2** which sets out the objectives, may be used to navigate the thesis. A systematic review of the association between kidney disease and acute, community-acquired infections (objective 1) is presented in **Chapter 2** to complete the Background section.

The Methods section (**Chapters 3–6**) presents the general materials and methods used in common across the thesis objectives. **Chapter 3** describes the data sources used in the study, and identification of the study population. **Chapter 4** describes the methods used to identify episodes of infection and calculate infection rates. **Chapter 5** describes the identification of chronic kidney disease. **Chapter 6** describes the definition of covariates.

The Results section (**Chapters 7–10**) contains research articles which present the study design and analysis, results and discussion specific to the particular study objective. **Chapter 7** describes the incidence of community-acquired LRTI (including pneumonia), pneumonia (as a subset of LRTI), UTI and sepsis among older people with diabetes mellitus (objective 2). **Chapter 8** presents estimates of the association between markers of CKD and incidence of LRTI (including pneumonia), pneumonia (as a subset of LRTI) and sepsis among older people with diabetes (objective 3). **Chapter 9** explores the effectiveness of pneumococcal vaccine to prevent community-acquired pneumonia, and influenza vaccine against community-acquired LRTI, and whether this varies according to CKD status (objective 4). **Chapter 10** describes the association of markers of CKD with all-cause short-term mortality following infection (objective 5).

Finally, the Discussion summarises the main results of each study, considers the overarching strengths and weaknesses of the thesis as a whole, and suggests implications of the findings for clinical practice and future research (**Chapter 11**).

Chapter 2. Systematic review of the association between chronic kidney disease and infection incidence

2.1 Introduction to Paper 1

This paper was published in *BMJ Open* and presents a systematic literature review of the association between chronic kidney disease and four acute, community-acquired infections: lower respiratory tract infection, urinary tract infection, central nervous system infection, and sepsis.

Fourteen studies were identified, all consistent with a positive association between CKD and infection risk. Considerable heterogeneity precluded meta-analysis, and most studies gave cause for concern about study quality. A few large, high-quality studies found a graded association between CKD and risk of hospitalisation with infection. Other than these, there was a scarcity of high-quality studies on this research topic, and in particular a lack of data on the relationship between proteinuria and infection incidence independently of glomerular filtration rate. There were few data available on the association of CKD with infection incidence using less severe outcome measures than hospitalisation, and thus it was not possible to identify an effect on susceptibility to infection separately from an effect on the severity of infection.

The study search terms and the detailed inclusion criteria and study quality assessment referred to in the article as supplementary material are available in this thesis as **Appendix A**.

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	Helen McDonald
Principal Supervisor	Dr Dorothea Nitsch
Thesis Title	The epidemiology of infections among older people with diabetes mellitus and chronic kidney disease

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?	2014		
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 led the design of the study, including search terms and inclusion criteria, with detailed advice from D. Nitsch and S. Thomas. I conducted the search, study selection, data extraction and quality assessment. D. Nitsch screened the sample of 100 abstracts. Final quality assessment decisions were agreed by discussion between all authors. I drafted the

	initial manuscript, and revised it in light of comments made by D. Nitsch and S. Thomas. The manuscript was peer-reviewed, and I incorporated suggestions from reviewers into the final manuscript.

Student Signature: HMcDonald **Date:** 10 April 2015

Supervisor Signature: Dorothea Nitsch **Date:** 14/4/15

BMJ Open

Chronic kidney disease as a risk factor for acute community-acquired infections in high-income countries: a systematic review

Helen I McDonald,[1] Sara L Thomas,[2] Dorothea Nitsch[1]

To cite: McDonald HI, Thomas SL, Nitsch D. Chronic kidney disease as a risk factor for acute community-acquired infections in high-income countries: a systematic review. *BMJ Open* 2014;**4**: e004100. doi:10.1136/bmjopen-2013-004100

► Prepublication history and additional material is available online. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2013-004100).

Received 23 September 2013
Revised 20 February 2014
Accepted 27 February 2014

[1]Department of Non-communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK
[2]Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK

Correspondence to
Dr Helen I McDonald;
helen.mcdonald@lshtm.ac.uk

ABSTRACT

Objective: A systematic review of the association of predialysis chronic kidney disease (CKD) with the incidence of acute, community-acquired infections.

Design: We searched the MEDLINE, EMBASE and Cochrane databases (inception to 16 January 2014) for studies analysing the association of predialysis kidney disease with the incidence of acute, community-acquired urinary tract infection (UTI), lower respiratory tract or central nervous system infections or sepsis. Studies were required to include at least 30 participants with and without kidney disease.

Setting and participants: Community-based populations of adults in high-income countries.

Outcome measures: Acute, community-acquired UTI, lower respiratory tract or central nervous system infections or sepsis.

Results: We identified 14 eligible studies. Estimates from two studies lacked 95% CIs and SEs. The remaining 12 studies yielded 17 independent effect estimates. Only three studies included infections managed in the community. Quality assessment revealed that probable misclassification of kidney disease status and poor adjustment for confounding were common. There was evidence from a few large high-quality studies of a graded association between predialysis CKD stage and hospitalisation for infection. One study found an interaction with age, with a declining effect of CKD on infection risk as age increased. There was evidence of between-studies heterogeneity (I^2=96.5%, p<0.001) which persisted in subgroup analysis, and thus meta-analysis was not performed.

Conclusions: Predialysis kidney disease appears to be associated with increased risk of severe infection. Whether predialysis kidney disease increases the susceptibility to infections and whether age modifies this association remains unclear.

Strengths and limitations of this study

- This study used a sensitive search strategy, with a broad definition of kidney disease, for a thorough and inclusive search.
- Study quality was assessed using a tool adapted to observational studies, providing a transparent assessment of the risk of a range of biases for each study.
- Between-study heterogeneity and the low quality of many of the studies limit the interpretation of results of the studies currently available.

INTRODUCTION

Chronic kidney disease (CKD) is common, and its prevalence is increasing.[1] Infection is a major cause of mortality in end-stage renal disease (ESRD) and hospitalisation at all stages of CKD. The second commonest cause of death among patients with ESRD in the USA is septicaemia, and patients with ESRD are at increased risk of death from infection compared to the general population.[2–4] Patients with ESRD and predialysis CKD in the USA are at higher risk of hospitalisation for infection than the general population.[2 5 6] Predialysis CKD has been found to increase mortality among patients hospitalised with infections.[7]

Increased mortality and hospitalisation from infection could be driven by increased severity of infection, that is, once an infection is present, the course of the associated illness is more severe, or increased incidence, that is, CKD may make people more susceptible to develop an infection. Patients with CKD display impaired host immunity: reduced vaccination responsiveness is observed at all stages of CKD.[8]

Among patients with ESRD, aspects of dialysis, such as vascular and peritoneal access for dialysis, may be a risk factor for infection incidence and severity. However, this does not tell the whole story, and only 23% of infection-related hospitalisations among patients undergoing haemodialysis in the USA were identified as related to vascular access in the

HEMO study.[9] Risk factors for infection identified among patients with ESRD which are not related to renal replacement therapy, and could apply at all stages of predialysis CKD, include: the causes and treatment of kidney disease; comorbidities; reduced vaccine effectiveness; and high levels of exposure to healthcare facilities.[10]

If there is an increased risk of infection incidence at early stages of CKD, this would affect a large and growing number of patients. Awareness and quantification of this risk could have benefits for patient management, more effective vaccination strategies and healthcare planning.

Narrative reviews have concluded that it is likely that CKD in itself increases infection incidence, but reported a lack of evidence.[10–12] We are not aware of any relevant systematic literature reviews of the effect of CKD on infection incidence.

This review sought to assess systematically whether predialysis CKD is a risk factor for the incidence of acute, community-acquired urinary tract infection (UTI), lower respiratory tract infection (LRTI), central nervous system (CNS) infection or sepsis, among community-based adults in high-income countries.

METHODS

Data sources and searches

One reviewer (HIM) searched the MEDLINE and EMBASE databases, and the Cochrane library, from inception to 16 January 2014. The search strategies combined text words and MeSH terms for three concepts: acute community-acquired infection (sepsis, UTI, LRTI or CNS infection), kidney disease and relative risk. We used search terms to identify studies among adult humans in high-income countries (according to the World Bank classification),[13] and limited the search to articles in English, French or German. The full strategies are available in online supplementary tables S1–S3.

We searched the reference lists of all included studies and any pertinent review articles to identify further eligible studies.

Study selection

One reviewer (HIM) screened titles and abstracts, reviewed the full text of identified studies and made initial decisions on eligibility according to prespecified inclusion criteria (see online supplementary table S4). Any borderline cases were discussed between HIM, DN and SLT. A second reviewer (DN) checked a sample of 100 abstracts, selected randomly after de-duplication of records, and a κ statistic was calculated to describe agreement in the selection of studies.

Eligible studies analysed the effect of predialysis kidney disease on the relative risk of at least one of the four specified acute, community-acquired infections among community-based adults in high-income countries. We excluded study populations managed in secondary care (unless for kidney disease), routinely treated with immunosuppressants, or exclusively of pregnant women, as these groups have a raised risk of infection, and the relationship of CKD to infection risk may be different among these groups compared to that in the general adult population in primary care. Ascertainment of CKD, as a silent disease, and, to a certain extent, ascertainment of acute community-acquired infections are dependent on high levels of monitoring and good access to healthcare, so we restricted our search to high-income countries. Chronic infections such as tuberculosis were not included, as the relationship between CKD and chronic infection is very likely to differ from that between CKD and acute infections, which was our focus in this review.

To maximise the sensitivity of our search strategy, we accepted a wide range of definitions of kidney disease, including: medical diagnosis of kidney disease, reduced estimated glomerular filtration rate or creatinine clearance, elevated creatinine, proteinuria, microalbuminuria or macroalbuminuria and renal structural abnormalities. We also accepted definitions which included some patients with ESRD among the patients with CKD, but excluded definitions which were exclusively patients receiving renal replacement therapy.

Outcomes of interest were relative risk estimates of acute community-acquired LRTIs, UTIs, CNS infections or sepsis. We accepted outcomes describing incidence of severe infections (such as hospitalisation with pneumonia).

We restricted our search to published studies which were sufficiently large to include at least 30 participants with and without kidney disease, to allow reasonable precision of the study estimate. Detailed eligibility criteria are listed in online supplementary table S4.

Data extraction and quality assessment

Data were extracted from relevant studies using a prespecified collection form. Study characteristics extracted included study design, data source, any participant exclusion criteria, number of participants, age, gender, baseline renal function, definition of renal impairment and definition of the outcome infection. An estimate of relative risk (rate ratio, risk ratio or OR) with any measures taken to address confounding was extracted from each eligible independent analysis in each study. Studies with no CIs and for which the SE was not calculable from the data presented were included in the review but not considered for meta-analysis.

When multiple estimates were available from a study but were not independent, a single estimate was identified for potential meta-analysis by selecting the estimate best adjusted for confounding, using the most recent data, comparing the level of CKD most common in the general population with no CKD.

Study quality was assessed using a prespecified tool adapted from Higgins *et al*[14] for observational studies. Studies were assigned a high, low or uncertain risk of each of the following: selection bias, non-differential measurement error for exposure and outcome, information bias in exposure and outcome, confounding and reverse causation. The minimum requirement for a low risk of bias

from confounding was appropriate management of confounding by age, sex and diabetes. The specific criteria used are detailed in online supplementary table S5.

Data synthesis and analysis

The relationship between CKD and UTIs was considered likely to differ from that of CKD to other infections, due to potential reverse causality. For example, repeat UTIs may cause kidney disease, or structural kidney disease may be identified through investigation of repeat UTIs. Therefore, in all quantitative analysis, UTIs were analysed separately from other infections.

Estimates were examined for heterogeneity using Cochran's Q statistic and the I^2 statistic as described by Higgins *et al*.[15] If I^2 was less than 50% and Cochran's Q statistic p≥0.1, fixed-effects meta-analysis was considered for each of the two categories (UTI and other infections). Funnel plots were constructed to look for publication bias. All analysis was conducted using STATAV.12.0.

RESULTS

The database searches identified 10 380 citations, of which 1204 were duplicates (figure 1). Both reviewers had 100% agreement on which studies to extract for full-text analysis from screening a random sample of 100 abstracts (Cohen's K=1).

We identified 14 eligible studies, with varying study characteristics (table 1). Four studies were case–control studies,[16–19] and 10 were cohort studies.[20–29] Seven studies investigated a range of risk factors for infection,[16–19 21 28 29] two studies reported the effect of CKD on infection as a confounder of the effect of interest[24 25] and five studies investigated the effect of CKD on infection risk as their primary research question.[5 20 22 26 27]

Seven studies were based among the general population.[5 16 19 21 23 28 29] Other study populations included: attendants at a specialist renal clinic,[22] patients with diabetes mellitus,[25] patients admitted to hospital for an acute cardiovascular event or an arterial revascularisation procedure,[24] and the Navajo Nation—a population which experiences 3–5 times higher rates of invasive pneumococcal disease than the general US population.[17] The population of the cohort studies in Calgary, Canada comprised adults with a serum creatinine test result available in their medical records.[26 27] There is some overlap in the study populations of these two cohort studies: residents aged over 65 years with a serum creatinine measurement between 1 July 2001 and 31 December 2001 and also between 1 July 2003 and 30 June 2004 would have been included in both studies for the period from the second creatinine measurement until 31 December 2004.[26 27]

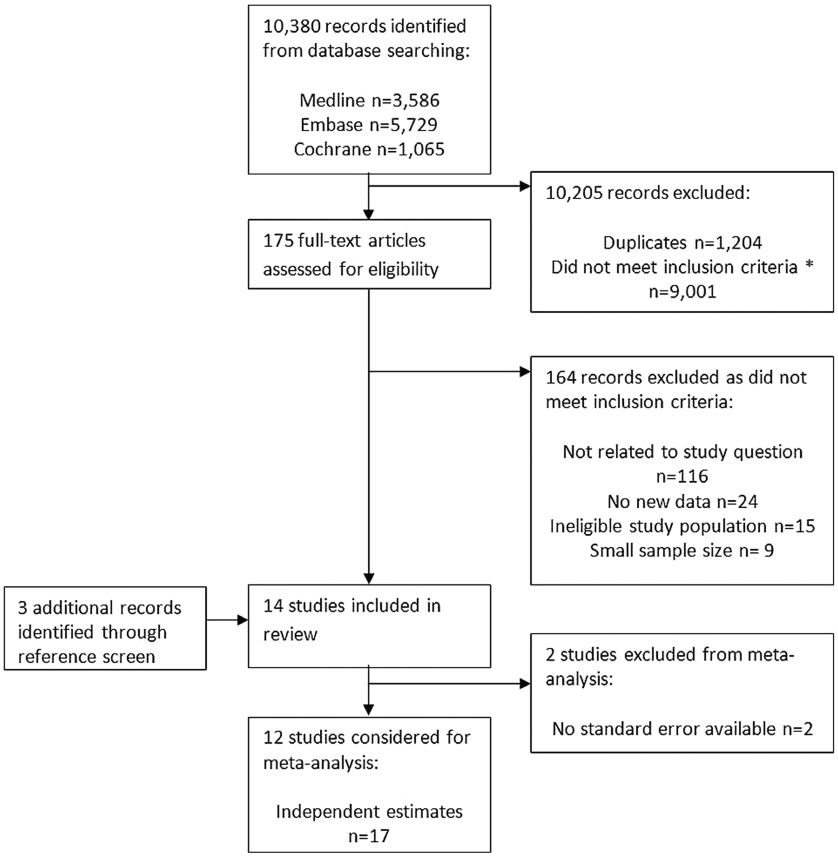


Figure 1 Flow chart of study selection. *Common examples of ineligible studies returned by the database searches included: studies in which renal failure and infection were both outcomes, studies in which renal failure and infection were both exclusion criteria, studies of acute renal failure resulting from sepsis or antibiotic use, studies of chronic infections (e.g. hepatitis C, BK viraemia, tuberculosis) following organ transplantation, descriptive studies of UTIs, descriptive studies of CKD, studies of predictors of prognosis among patients with infections, and review articles without any original data.

Table 1 Characteristics of eligible studies (n=14)

Case–control studies

	Study			Kidney disease			Infection			Kidney disease prevalence		
	Date	Setting	Population Age Percentage of female	Defined	ESRD included	Ascertained	Type	Defined	Ascertained	Cases	Controls	OR (95% CI)
Vinogradova *et al*[16]	1996–2005	UK	General population Any age Median age band 45–64 years Cases 49.3%, controls 49.1%	Chronic renal disease	Unclear	Primary care medical record diagnosis code in previous 5 years	Pneumonia	Medical diagnosis recorded in primary care records	READ code in primary care medical records	203/17 172 (1.2%)	386/71 299 (0.5%)	1.72 (1.3 to 2.07)[1]
Watt *et al*[17]	1999–2002	The Navajo Nation USA	Navajo adults ≥18 years Summary age and sex n/r	Chronic renal failure	17 participants receiving dialysis	Medical record abstraction	Invasive pneumococcal disease	*Streptococcus pneumoniae* isolated from a normally sterile body fluid during illness	Active laboratory surveillance system[2]	20/118 (16.9%)	12/353 (3.4%)	2.6 (0.87 to 7.7)[3] p=0.087
Loeb *et al*[18]	2002–2005	Ontario and Alberta Canada	General population ≥65 years Mean age: cases 79.1, controls 74.4 years. Cases 39.6%, controls 68.5%	Renal disease	Unclear	Cases: hospital interview. Controls: telephone interview at home	Pneumonia	Consistent chest X-ray and ≥2 of: chest pain, shortness of breath, productive cough, temperature >38° C, crackles on auscultation	Recruited patients attending emergency departments	127/690 (18.4%)	38/82 (4.4%)	4.06 (1.98 to 8.35)[4] p<0.001
Schnoor *et al*[19]	2002–2005	Germany	General population >18 years Mean age: cases 57, controls 57.5 years Cases 44.8%, controls 54.7%	Chronic renal disease	Unclear	Cases: reporting physician. Controls: self-reported questionnaire	Pneumonia	(1) Infiltrate on chest X-ray or (2) temperature ≥38.3°C with any of: cough, purulent sputum, positive auscultation Excluded if hospitalised within prior 4 weeks, or immunodeficient	Community-acquired pneumonia network registry reports (primary and secondary care)	49/1128 (4.3%)	27/1044 (2.6%)	1.7 (1.1 to 2.8) (unadjusted) p<0.05

Cohort studies

	Study			Kidney disease			Comparison group	Infection			
	Date	Setting Follow-up time	Population Number Age Sex	Defined number with kidney disease	ESRD	Ascertained	Defined	Type	Defined	Ascertained	Risk or rate ratio (95% CI)
Higgins[22]	1985	Oxford UK 1 year	Patients attending a renal unit with chronic renal failure n=211 17–77 years Mean 50.5 years Percentage of female n/r	Creatinine ≥250 µmol/L Number n/r	Excluded	Serum creatinine	Creatinine <250 µmol/L	UTI	>10^5 organism/mL and ≥10 leucocytes/hpf in clean catch urine specimen	Medical record review	Creatinine µmol/L <250 1 250–500 1.5[5] >500 2[5]
Dalrymple *et al*[23]	1989–2007	USA Mean 11.5 years	General community-dwelling population[6] n=5142	Baseline eGFR <90 mL/min/1.73 $m^{2\ 7}$ n=3863	Excluded	Baseline cystatin C	Baseline eGFR ≥90 mL/min/ 1.73 $m^{2\ 7}$	Pulmonary	Hospital admission with a principal discharge diagnosis of the relevant infection (ICD-9-CM codes)	Medical record review following patient report of hospital admission in cohort study	eGFR mL/min/1.73m^2 ≥90 1 60–89 1.22 (0.99 to 1.54)[8] 45–59 1.27 (0.94 to 1.71)[8]

Continued

Table 1 Continued

Cohort studies

	Study			Kidney disease			Comparison group	Infection			
	Date	Setting Follow-up time	Population Number Age Sex	Defined number with kidney disease	ESRD	Ascertained	Defined	Type	Defined	Ascertained	Risk or rate ratio (95% CI)
			>65 years Mean 72 years 61% female					Genitourinary			15–44 1.81 (1.25 to 2.63)[8] ≥90 1 60–89 1.08 (0.75 to 1.56)[8] 45–59 1.17 (0.67 to 2.05)[8] 15–44 2.63 (1.40 to 4.96)[8]
								Bacteraemia and sepsis			≥90 1 60–89 1.10 (0.77 to 1.58)[8] 45–59 1.55 (0.93 to 2.57)[8] 15–44 0.77 (0.29 to 2.03)[8]
Hackam *et al*[24]	1997–2002	Ontario Canada Mean 2.2 years	Patients with cardiovascular disease n=69 168 >65 years Mean 74.1 years 44% female	Chronic renal insufficiency n=7169	Unclear	Health record databases[9]	No chronic renal insufficiency	Sepsis	Hospital admission with a diagnosis of sepsis[10]	Health record database[11]	1.47 (1.27 to 1.72)[12]
Karunajeewa *et al*[25]	1999–2000	Western Australia Mean 2.9 years	Patients with diabetes n=496 >10 years Mean 66.1 years[13] 46.2% female	Albuminuria; serum urea; serum creatinine	Unclear	Baseline urinary ACR, serum urea, serum creatinine	HR per 2.72-fold increase in ACR or serum urea	Urinary sepsis and non-urinary sepsis	Hospitalisation diagnosis codes (principal diagnosis, or principal or secondary diagnosis)[14]	Health record database[15]	Urinary sepsis (principal code) Ln (ACR) 1.5 (1.1 to 1.9)[16] p=0.004 Urinary sepsis (principal or secondary code) Ln (ACR) 1.3 (1.1 to 1.6)[17] p=0.005 Non-urinary sepsis (principal) Ln (ACR) 1.4 (1.1 to 1.9)[16] Non-urinary sepsis (principal or secondary code) Ln (urea) 4.6 (2.3 to 9.4)[16] p<0.001
James *et al*[26]	2001–2004	Calgary Canada Mean 3.2 years	General population n=25 675 >65 years Mean by eGFR[18] 55.9% female	Baseline eGFR <60 mL/min/1.73 m^2 [19] n=6941	Excluded	Calgary Laboratory Services records	Baseline eGFR ≥60 mL/min/1.73 m^2 [19]	Bloodstream infection	Any pathogenic organism isolated from ≥1 blood cultures submitted from the community or ≤2 days of hospital admission	Calgary Laboratory Services records	eGFR mL/min/1.73 m^2 ≥60 1 45–59 1.17 (0.92 to 1.49)[20] 30–44 1.60 (1.20 to 2.13)[20] <30 2.95 (2.11 to 4.14)[20]
James *et al*[27]	2003–2006	Calgary Canada Median 2.5 years	General population n=252 516 ≥18 years Mean by eGFR[21] 42.3% female	Time updated eGFR <60 mL/min/1.73 m^2 [22] n=35 948	Excluded	Calgary Laboratory Services records	eGFR 60–104 mL/min/1.73 m^2 [22]	Pneumonia	ICD-10 code for pneumonia any position in hospital discharge report	Hospital discharge reports	eGFR mL/min/1.73m^2 18–54 years 60–104 1 45–59 3.23 (2.40 to 4.36)[23] 30–44 9.67 (6.36 to 14.69)[23] <30 15.04 (9.64 to 23.47)[23] Age 55–64 years

Continued

Table 1 Continued

Cohort studies

	Study			Kidney disease			Comparison group	Infection			
	Date	Setting Follow-up time	Population Number Age Sex	Defined number with kidney disease	ESRD	Ascertained	Defined	Type	Defined	Ascertained	Risk or rate ratio (95% CI)
											60–104 1 45–59 1.43 (1.11 to 1.84)[23] 30–44 1.94 (1.32 to 2.87)[23] <30 5.50 (3.83 to 7.92)[23] Age 65–74 years 60–104 1 45–59 1.18 (0.99 to 1.40)[23] 30–44 2.24 (1.84 to 2.73)[23] <30 3.23 (2.52 to 4.13)[23] Age ≥75 years 60–104 1 45–59 0.95 (0.85 to 1.05)[23] 30–44 1.03 (0.92 to 1.16)[23] <30 1.79 (1.55 to 2.06)[23]
Wang *et al*[28]	2003–2011	USA Mean .7 years	General population sample (weighted by age, geography and ethnicity)[24] n=30 239 ≥45 years 69%>60 years 55% female	Baseline eGFR <60 mL/min/1.73 m^{2} [25]	Unclear	Baseline serum creatinine	Baseline eGFR ≥60 mL/min/ 1.73m^{2} [25]	Sepsis	Among hospitalisations attributed by participants to serious infection, medical record review[26]	Initially reported by study participants, confirmed with medical record review	1.99 (1.73 to 2.29)[27]
Caljouw *et al*[29]	1998–2004	Leiden The Netherlands Mean 2.6 years	General population n=479 86–90 years All aged 86 years at entry 67.2% female	Creatinine clearance <30 mL/min[28] n=43	Unclear	Baseline serum creatinine	Creatinine clearance ≥30 mL/ min[28]	UTI	Diagnosed by treating physician based on signs, symptoms and urine analysis; or death records[29]	Physician interview and medical record review Statistics Netherlands for cause of death data	0.9 (0.5 to 1.7) (unadjusted) p=0.794
Campbell *et al*[21]	2009–2010	England UK 9 months	General population n=43.9 million 6 months—64 years Summary age and sex n/r	Chronic kidney disease n=182 000	Unclear	Cases: consultant microbiologist report Denominator: primary care population estimate[30]	No pre-existing conditions[30]	Pandemic influenza A (H1N1)	PCR test confirmation of pandemic influenza A (H1N1) from a hospital inpatient	Consultant microbiologist report to national surveillance system	17.5 (13.4 to 22.9)[31]

Continued

Table 1 Continued

Cohort studies											
	Study			Kidney disease			Comparison group	Infection			
	Date	Setting Follow-up time	Population Number Age Sex	Defined number with kidney disease	ESRD	Ascertained	Defined	Type	Defined	Ascertained	Risk or rate ratio (95% CI)
USRDS 2010[20]	2008	USA 1 year[32]	Medicare patients 66+ years	Chronic kidney disease	Excluded	Insurance database ICD-9-CM codes[33]	No CKD	Pneumonia UTI Bacteraemia/ septicaemia	Principal cause of hospital admission using hospital insurance claim records	ICD-9-CM codes 480–486 ICD-9-CM codes[34] ICD-9-CM codes 038.0–038.9	2.76 (unadjusted) 3.15 (unadjusted) 3.90 (unadjusted)

[1]Controls matched to cases on age at index data (within 1 year), sex, general practice and calendar time. Estimate adjusted for smoking status, Townsend deprivation score, use of influenza vaccine in previous 12 months, use of pneumococcal vaccine in previous 5 years, number of years of medical records data available in database, and comorbidities including: diabetes, chronic heart disease, chronic respiratory disease, asplenia, cerebrospinal shunt, chronic liver disease, sickle cell disease or coeliac disease, cochlear implant, HIV/AIDS, immunosuppression, stroke or transient ischaemic attack, rheumatoid arthritis, Parkinson's disease, multiple sclerosis, dementia, osteoporosis and any cancer.
[2]Center for American Indian Health surveillance system.
[3]Cases and controls matched by gender and age group. Adjusted for age, receipt of pneumococcal polysaccharide vaccine, congestive heart failure, alcohol use, body mass index and unemployment.
[4]Adjusted for age, non-English language spoken most at home, living in detached house, living alone, congestive heart failure, chronic obstructive pulmonary disease, dysphagia, functional status using Barthel Index, immunosuppressive medications, nutritional score, tobacco use (lifetime history and secondhand smoke), alcohol consumption and history of regular exposure to gases, fumes or chemicals at home or at work.
[5]Approximate numbers, read from bar graph in publication. No CIs available.
[6]Cohort selected for the Cardiovascular Health Study. Exclusion criteria included: inability to provide informed consent or communicate with the interviewer, institutionalisation, being homebound, receipt of hospice care, treatment with radiation or chemotherapy for cancer or plans to move out of the community within 3 years.
[7]Serum cystatin C measured by particle-enhanced immunonephelometric assay, and eGFR calculated using: $eGFR=6.7\times CysC^{-1.19}$.
[8]Adjusted for age, sex, race, tobacco use, body mass index, diabetes mellitus, coronary heart disease, stroke, heart failure, cancer, chronic obstructive pulmonary disease, serum albumin, C reactive protein, interleukin-6.
[9]Canadian Institute for Health Information Discharge Abstract database or Ontario Health Insurance Plan database.
[10]ICD-9 codes 003 1, 036 2 and 038 0–038 9.
[11]Canadian Institute for Health Information Discharge Abstract database.
[12]Adjusted for statins, age, sex, nature of index event, Charlson index, healthcare use, malignant disease, chemotherapy, neutropaenia, diabetes mellitus, oral steroids, antineoplastics, other immunosuppressants, history of aspiration, structural lung disease, infection (respiratory, GI, skin/soft tissue or other), recent trauma, transplant recipient, heart failure, stroke, chronic liver disease, chronic obstructive pulmonary disease, alcoholism, dementia, Parkinson's disease. Statin and non-statin users matched using propensity scoring for the above factors.
[13]Mean age among the 460 participants without asymptomatic bacteriuria, 66.1 years (SD 11.0): mean age among the 36 participants with asymptomatic bacteriuria, 67.7 years (SD 10.5).
[14]ICD-9 or ICD-10 codes for urinary sepsis were those encoding UTI, cystitis, pyelonephritis, orchitis, epididymitis and prostatitis; codes for non-urinary sepsis were those for sepsis, septicaemia and/or abscess.
[15]Western Australia Data Linkage System.
[16]Adjusted for presence of asymptomatic bacteriuria.
[17]Adjusted for presence of asymptomatic bacteriuria and age.
[18]Mean age±SD by eGFR. ≥60: 74.4±6.5 years. 45–59: 77.5±7.2 years. 30–44: 79.3±7.4 years. <30: 78.6±7.4 years.
[19]eGFR calculated using abbreviated Modification of Diet in Renal Disease Study equation (omitting ethnicity) from single outpatient serum creatinine result.
[20]Adjusted for age, sex, diabetes mellitus, comorbidity score and care in a CKD clinic.
[21]Mean age±SD by eGFR. ≥105: 38.7±14.6. 60–104: 50.9±15.4. 45–59: 67.0±14.1. 30–44: 74.5±12.9. <30: 73.3±15.2.
[22]eGFR calculated using abbreviated Modification of Diet in Renal Disease Study equation (omitting ethnicity) from most recent outpatient serum creatinine result.
[23]Adjusted for age, sex, socioeconomic status, ethnicity, diabetes mellitus, Charlson comorbidity score.
[24]Cohort selected for the Reasons for Geographic and Racial Differences in Stroke (REGARDS) study. Population weighted by age, ethnicity and geography according to local stroke incidence rates.
[25]eGFR calculated using the CKD-EPI equation.
[26]Medical record review confirming (1) serious infection as the major reason for admission and (2) ≥2 of heart rate >90 bpm, temperature >38.3°C or <36°C, tachypnoea >20 breaths/min or leucocytosis.
[27]Adjusted for age, sex, race, education, income, geographical region, alcohol use and smoking status.
[28]Creatinine clearance calculated from serum creatinine concentration and weight using the Cockcroft-Gault formula.
[29]Cause of death recorded as UTI (ICD-10 code N39.0)'.
[30]Department of Health-Health Protection Agency influenza vaccine uptake primary care monitoring system data.
[31]Adjusted for age.
[32]Smoothed estimate: models include data from the stated year and the 2 years preceding it, applying weights of 1, 1/4 and 1/8 with increasing distance in time.
[33]ICD-9-CM diagnosis codes recorded in insurance claims during the preceding year: 585.1-585.5 (chronic kidney disease stages 1–5); or 585.6 with no ESRD 2728 form or other indication of ESRD.
[34]Principal hospital admission ICD-9-CM codes: 590-590.9, 595-595.4, 597-597.89, 598, 599.0, 601-601.9, 604-604.9, 607.1-2, 608.0, 608.4, 616.1, 616.3-4 and 616.8.
ACR, albumin : creatinine ratio; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; n/r, not reported; USRDS, US Renal Data System; UTI, urinary tract infection.

Definitions of kidney disease included medical diagnoses of chronic renal disease, elevated creatinine levels, impaired creatinine clearance and structural abnormalities of the kidney. Five studies excluded patients with ESRD, and one specified the number included, but for the remaining eight studies it was unclear how many of the included patients received renal replacement therapy (table 1).

Three studies recorded infections diagnosed in primary care or outpatients,[16 19 29] two recorded infections identified from a positive culture result,[17 26] one included infections diagnosed in the emergency department,[18] seven required hospital admission for infection[5 21 23–25 27 28] and for one study the definition and severity of infection was unclear.[22]

For two studies, the results extracted had no CI or SE and these could not be calculated from the reported data. From the remaining 12 studies, 17 independent effect estimates with SEs were available for meta-analysis, among which UTI was the outcome in three estimates.

For all infections, there was strong evidence of considerable heterogeneity (Cochran's Q statistic p<0.001, I^2=96.5%). This persisted when estimates for UTIs were excluded (p<0.001, I^2=97.2%), when considering LRTIs alone (p<0.001, I^2=98.2%), when limited to cohort studies (p<0.001, I^2=97.3%), and when stratified by exclusion of patients with ESRD (ESRD excluded, p<0.001, I^2=88.9%: ESRD not excluded p<0.001, I^2=97.2%). Owing to this heterogeneity, meta-analysis was not performed.

All results are displayed in the Forest plot (figure 2). Despite the quantitative heterogeneity, the results were qualitatively similar: all estimates were compatible with a positive association between kidney disease and infection. The four studies which compared different stages of CKD found a graded association of increased risk of infection with more severe CKD. All four of these studies excluded patients with ESRD.[22 23 26 27] One study found that the effect of CKD on infection risk was modified by age, with a declining effect of CKD on infection risk as age increased.[27] This effect was consistent with the lower effect of CKD on UTI incidence found among 86–90 year-olds (0.90, 95% CI 0.50 to 1.77) compared with an adult study population with a mean age of 66 years (1.50, 95% CI 1.10 to 1.90).[25 29]

The funnel plot was sparsely populated, with widely scattered effect estimates, and provides no clear evidence for or against publication bias (see online supplementary figure S1).

Study quality was variable. Relying on routine medical diagnosis introduced a potential source of misclassification of kidney disease status for seven studies.[5 16–19 21 24] There was variable adjustment for confounding, from unadjusted crude estimates to estimates adjusted for a range of comorbidities, demographic and socioeconomic factors. Six studies did not meet this review's minimal requirements.[19 21 22 25 28 29] The summarised results are displayed in table 2, and the full quality assessment is in online supplementary table S5.

DISCUSSION

Our comprehensive search strategy identified 14 studies describing an association between kidney disease and acute community-acquired infection. Although between-study heterogeneity precluded meta-analysis, all studies were consistent with a positive direction of association. Four studies which reported estimates on more than one category of kidney disease found a graded association in which risk of infection increased with greater severity of CKD. These four studies excluded patients with ESRD, and three were at low risk of bias in all categories of quality assessment.[22 23 26 27]

To the best of our knowledge, this is the first review to address this research question systematically. We used a sensitive search strategy, with a broad definition of kidney disease, for a thorough and inclusive search. The results are consistent with the conclusion of previous narrative reviews: that an association between CKD and infection incidence is likely, but that there is a paucity of evidence.[10–12]

Heterogeneity between the studies precluded a meta-analysis of results. Variable study designs and biases may have contributed to the heterogeneity: for example, the four case–control studies calculated ORs, which may differ from equivalent rate ratios for common infections.[16–19] Failure to control the confounding effects of age, sex and diabetes would be likely to result in overestimation of the effect of CKD on infection. Non-differential misclassification of kidney disease status in studies which relied on routine medical diagnosis would be expected to underestimate the effect of CKD on infection risk. In general, the risk of ascertainment bias from increased monitoring for infection among patients with CKD is probably low, although one study assessed risk factors for hospitalisation with influenza during an influenza pandemic, in which context patients with influenza-like symptoms may have been more likely to be tested for influenza A(H1N1) if they also had CKD.[21]

The heterogeneity may reflect true differences in effect size between the studies.

First, the studies considered a range of outcomes. CKD may have a different effect on the incidence of different infections. For all but three studies, detection of infection required either hospital attendance for the infection or a positive blood culture. CKD may affect severity of infection, as an alternative, or in addition to any effect on infection incidence. CKD may also increase the probability of hospital admission for management of a moderately severe infection. Either would result in a larger effect of CKD on the risk of severe infectious outcomes (such as hospitalisation for sepsis) than on less severe infections (such as community-diagnosed LRTI), and could result in the graded association we observed, with increasing hospitalisation for patients with more severe stages of CKD.

Second, the studies included a variety of definitions of kidney disease. For example, proteinuria (and renal loss of complement) may represent a separate mechanism

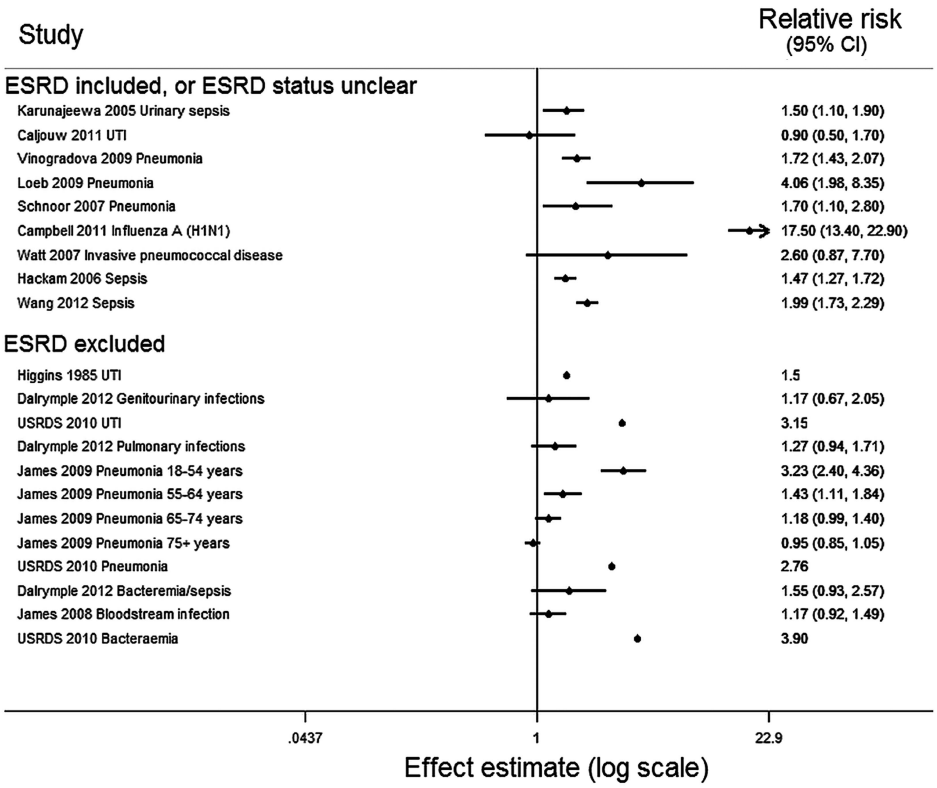


Figure 2 Forest plot of all estimates of the association of chronic kidney disease with infection (n=17) from all 14 studies identified. The estimates from Higgins 1985 and USRDS 2010 did not include SEs. Dalrymple 2012: presented estimates compare eGFR 45–59 with eGFR ≥90 mL/min/1.73m^2; James 2009: presented estimates compare eGFR 45–59 with eGFR 60–104 mL/min/1.73m^2; James 2008: presented estimates compare eGFR 45–59 with eGFR≥60 mL/min/1.73m^2. eGFR, estimated glomerular filtration rate; ESRD, end-stage renal disease; USRDS, US Renal Data System; UTI, urinary tract infection.

for risk of infection than uraemia. For the nine studies which did not exclude patients with ESRD, it is unclear to what extent the results reflect the effect of treatments associated with dialysis, such as vascular or peritoneal access for dialysis, on infection incidence.

Third, the association of CKD with infection may be modified by age. James *et al* observed a weaker association of CKD with hospitalisation for pneumonia as age increased. They suggested that such an observation could be explained by a lower baseline rate of hospitalisation for pneumonia among younger adults, the natural decline in renal function by age, and inaccuracy in the estimation of renal function using the Modification of Diet in Renal Disease (MDRD) study equation in older populations.[27] As their study population included only adults who had had a creatinine test result, reasons for testing creatinine could also be relevant confounders. As age increases, more comorbidities accrue which require creatinine tests to guide therapy. Hence, younger people who receive a creatinine test may be at an unusually high risk for infections and CKD due to the reasons associated with getting a creatinine test. A real age-dependency of the CKD-infection association would be consistent with the lower effect of CKD on UTI incidence found among 86–90-year-olds (0.90, 95% CI 0.50 to 1.77) compared with an adult study population with a mean age of 66 years (1.50, 95% CI 1.10 to 1.90). However, it may be that the study among the older adults measured a less severe outcome, and CKD may be associated with other factors that eventually lead to hospitalisation for UTI.[25 29]

CKD was not a component of the primary study question for nine of the 14 studies; thus, there is a risk that this association may have been reported and published only when CKD was found to be a risk factor for infection or an important confounder of another relationship. This would result in selective reporting bias, with a subsequent overestimation of the association of CKD with infection risk. This bias would be expected to affect smaller studies to a greater extent, and a funnel plot might show an asymmetry of relative risk estimates about the central pooled estimate among smaller studies. The sparsely populated funnel plot (see online supplementary figure S1) provides no clear evidence for or against selective reporting bias, but some evidence of selective reporting bias comes from within the individual studies. For example, the crude HR for the association of creatinine clearance with UTI incidence is reported in

Table 2 Summary of risk of bias within studies (quality assessment tool adapted from Higgins *et al*[14])

	Selection bias: control selection	Selection bias: participation	Selection bias: loss to follow-up	Non-differential misclassification: exposure	Information bias: exposure	Non-differential misclassification: outcome	Information bias: outcome	Confounding	Reverse causation
Case–control studies									
Vinogradova *et al*[16]			NA						
Watt *et al*[17]			NA						
Loeb *et al*[18]			NA						
Schnoor *et al*[19]			NA						
Cohort studies									
Higgins[22]	NA	NA							
Hackam *et al*[24]	NA	NA							
Dalrymple *et al*[23]	NA	NA							
Karunajeewa *et al*[25]	NA	NA							
James *et al*[26]	NA	NA							
James *et al*[27]	NA	NA							
Wang *et al*[28]	NA	NA							
Caljouw *et al*[29]	NA	NA							
Campbell *et al*[21]	NA	NA							
USRDS 2010[20]	NA	NA							

Key to table 2.
Low risk of bias
Uncertain risk of bias
High risk of bias

Caljouw *et al*[29] (0.9, 95% CI 0.5 to 1.7), but as creatinine clearance was not found to be significant in the multivariable model, the adjusted association is not reported.

The overlap in the study populations of the two large cohort studies based in Calgary, Canada could result in more similar estimates than if the study populations were independent.[26 27] Outcomes in the two studies are likely to be correlated with each other: hospitalisation with pneumonia could cause a positive blood culture, which would result in one infection being included as an outcome in both studies. This is unlikely to have a large effect, particularly in the qualitative assessment of the combined evidence, as the potential overlap of person-time is limited.

Although we excluded study populations routinely treated with specialist medication (unless for kidney disease), some study populations may have been at higher risk of infection than the general population, and this may have affected the relationship of CKD to infection. For example, the cohort of patients admitted for an acute cardiovascular event or an arterial revascularisation procedure will have had a higher prevalence of comorbidities (such as diabetes) than the general population and excluded patients with severe comorbidities who did not survive an acute cardiovascular event, or who were not fit enough to undergo the procedure.[24] Each of the selected study populations limits the generalisability of the individual study result, but the qualitatively similar findings across the variety of study populations, and their qualitative consistency with the studies based among the general population,[5 16 19 21 23 28 29] support a positive association between CKD and infection risk in a variety of study populations.

A few large, high-quality studies which excluded patients with ESRD have found a graded association between predialysis CKD and risk of hospitalisation with infection. All studies identified in this review were compatible with a positive association of CKD with increased infection risk. There are little data available on the association of CKD with infection incidence using less severe outcome measures than hospitalisation, and it is not possible in most studies to distinguish an effect on susceptibility to infection from an effect on the severity of infection.

The potential age-dependency of the relationship between CKD and infection is intriguing and needs further research. Also, there is currently no evidence on the relationship between proteinuria and infection incidence independent of the glomerular filtration rate. Future studies should identify infections in the community in addition to hospitalisations for infection, characterise the association of proteinuria adjusted for the glomerular filtration rate, explore the age-dependency of the association and assess vaccine efficacy among older people with CKD.

Contributors All authors designed the study strategy including the search terms, inclusion and exclusion criteria. HIM performed the search, study selection and data extraction. DN screened the randomly selected sample of 100 abstracts. All authors agreed on the quality assessment of included papers and interpretation of results by discussion. HIM drafted the article, which DN and SLT revised. All authors approved the final version of the manuscript.

Funding This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests This report is independent research arising from a Career Development Fellowship supported by the National Institute for Health Research, awarded to SLT (grant number CDF 2010-03-32). HIM is funded by a Kidney Research UK studentship, grant reference ST2/2011.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All data, including full search terms and eligibility criteria, are available either in the article or as online supplementary material submitted with this manuscript.

REFERENCES

1. Coresh J, Selvin E, Stevens LA, *et al*. Prevalence of chronic kidney disease in the United States. *JAMA* 2007;298:2038–47.
2. Collins AJ, Foley R, Herzog C, *et al*. United States Renal Data System 2007 Annual Data Report Abstract. *Am J Kidney Dis* 2008;51:A6–7.
3. Sarnak MJ, Jaber BL. Mortality caused by sepsis in patients with end-stage renal disease compared with the general population. *Kidney Int* 2000;58:1758–64.
4. Sarnak MJ, Jaber BL. Pulmonary infectious mortality among patients with end-stage renal disease. *Chest* 2001;120:1883–7.
5. US Renal Data System 2011 Annual Data Report. Morbidity and mortality in patients with chronic kidney disease. *Am J Kidney Dis* 2012;59:e59–68.
6. Naqvi SB, Collins AJ. Infectious complications in chronic kidney disease. *Adv Chronic Kidney Dis* 2006;13:199–204.
7. Viasus D, Garcia-Vidal C, Cruzado JM, *et al*. Epidemiology, clinical features and outcomes of pneumonia in patients with chronic kidney disease. *Nephrol Dial Transplant* 2011;26:2899–906.
8. Kausz AT, Gilbertson DT. Overview of vaccination in chronic kidney disease. *Adv Chronic Kidney Dis* 2006;13:209–14.
9. Allon M, Depner TA, Radeva M, *et al*. Impact of dialysis dose and membrane on infection-related hospitalization and death: results of the HEMO Study. *J Am Soc Nephrol* 2003;14:1863–70.
10. Dalrymple LS, Go AS. Epidemiology of acute infections among patients with chronic kidney disease. *Clin J Am Soc Nephrol* 2008;3:1487–93.
11. Foley RN. Infections in patients with chronic kidney disease. *Infect Dis Clin North Am* 2007;21:659–72, viii.
12. Foley RN. Infections and cardiovascular disease in patients with chronic kidney disease. *Adv Chronic Kidney Dis* 2006;13:205–8.
13. The World Bank. Country and lending groups. 2012 [6 Jun 2013]. http://data.worldbank.org/about/country-classifications/country-and-lending-groups
14. Higgins JP, Altman DG, Gotzsche PC, *et al*. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
15. Higgins JP, Thompson SG, Deeks JJ, *et al*. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–60.
16. 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:e329–38.
17. Watt JP, O'Brien KL, Benin AL, *et al*. Risk factors for invasive pneumococcal disease among Navajo adults. *Am J Epidemiol* 2007;166:1080–7.
18. Loeb M, Neupane B, Walter SD, *et al*. Environmental risk factors for community-acquired pneumonia hospitalization in older adults. *J Am Geriatr Soc* 2009;57:1036–40.
19. Schnoor M, Klante T, Beckmann M, *et al*. Risk factors for community-acquired pneumonia in German adults: the impact of children in the household. *Epidemiol Infect* 2007;135:1389–97.
20. Collins AJ, Foley RN, Herzog C, *et al*. US Renal Data System 2010 Annual Data Report. *Am J Kidney Dis* 2011;57(1 Suppl 1):A8, e1–526.
21. Campbell CNJ, Mytton OT, McLean EM, *et al*. Hospitalization in two waves of pandemic influenza A(H1N1) in England. *Epidemiol Infect* 2011;139:1560–9.
22. Higgins RM. Infections in a renal unit. *Q J Med* 1989;70:41–51.
23. Dalrymple LS, Katz R, Kestenbaum B, *et al*. The risk of infection-related hospitalization with decreased kidney function. *Am J Kidney Dis* 2012;59:356–63.
24. Hackam DG, Mamdani M, Li P, *et al*. Statins and sepsis in patients with cardiovascular disease: a population-based cohort analysis. *Lancet* 2006;367:413–18.
25. Karunajeewa H, McGechie D, Stuccio G, *et al*. Asymptomatic bacteriuria as a predictor of subsequent hospitalisation with urinary tract infection in diabetic adults: the Fremantle Diabetes Study. *Diabetologia* 2005;48:1288–91.
26. James MT, Laupland KB, Tonelli M, *et al*. Risk of bloodstream infection in patients with chronic kidney disease not treated with dialysis. *Arch Intern Med* 2008;168:2333–9.
27. James MT, Quan H, Tonelli M, *et al*. CKD and risk of hospitalization and death with pneumonia. *Am J Kidney Dis* 2009;54:24–32.
28. Wang HE, Shapiro NI, Griffin R, *et al*. Chronic medical conditions and risk of sepsis. *PLoS ONE* 2012;7:e48307.
29. Caljouw MA, den Elzen WP, Cools HJ, *et al*. Predictive factors of urinary tract infections among the oldest old in the general population. A population-based prospective follow-up study. *BMC Med* 2011;9:57.

METHODS SECTION

In this section, **Chapter 3** describes the data sources used in the study and identification of the study population, **Chapter 4** presents the methods used to identify episodes of infection and calculate infection rates, **Chapter 5** describes the identification of CKD, and **Chapter 6** provides detailed definition of all other variables used in analyses.

Chapter 3. Data sources and study population

3.1 Use of electronic health records for epidemiological research

Primary care in the United Kingdom (UK) is comprehensively computerised.[58] Clinicians record consultations and diagnoses directly into electronic clinical management systems, and prescribe electronically. Laboratory test results are increasingly reported electronically. The UK National Health Service (NHS) is free at the point of care for all UK residents, and provides near-universal coverage: 99% of the population are registered with a General Practitioner in primary care, and patients register with each practice for an average of 12 years.[59, 60] Once these electronic patient records have been collected, anonymised and aggregated into large administrative databases, they permit large observational studies with a dataset rich in detail, with long follow-up based among the general population.[61]

As with any secondary data analysis, care must be taken in interpretation.[62] It is important to assess data quality, including completeness and accuracy.[63] Appropriate data handling and interpretation require an understanding of the context within which the data were generated.[64] Secular trends in electronic health records may reflect changing epidemiology, patient health service use, clinical practice or recording patterns. In particular, the Quality and Outcomes Framework (QOF), introduced 1 April 2004, financially incentivised electronic recording of certain outcomes in primary care to demonstrate the achievement of 'pay for performance' indicators.[65]

Data linkage offers the opportunity to combine the advantages of different datasets: for example, to capture co-morbidities recorded in primary care, while identifying time in hospital from secondary care records.[63] This study used a large database of anonymised electronic health records from primary care in the UK, with linkages to secondary care, mortality and socio-economic status datasets.

3.2 Primary care data from the Clinical Practice Research Datalink

Primary care data were obtained from the Clinical Practice Research Datalink (CPRD). CPRD contains data continuously recorded from 1987 under its previous incarnations, Value Added Medical Products Ltd (VAMP) and the General Practice Research Database

(GPRD).[66, 67] Data were extracted in May 2011, when CPRD contained data for 12,808,177 patients at 627 practices across the UK.[68, 69]

The CPRD population has been found to be representative of the general UK population in terms of sex and age structure, with some under-representation of single-GP practices, and of practices in inner London.[67]

3.2.1 CPRD data collection

All participating practices use VISION Practice Management software. Events in primary care are recorded directly by the healthcare practitioner. Information recorded includes patient demographics and registration dates, consultation dates and diagnoses, lifestyle information (e.g. smoking status), clinical details (e.g. blood pressure), test results, prescriptions, and interactions with secondary care (e.g. referrals, discharge summaries).[69] Diagnoses are encoded contemporaneously by the healthcare practitioner using the Read coding system, the national standard.[70] General practices submit anonymised records for all registered patients to the central database.

3.2.2 CPRD data quality

CPRD checks data quality indicators by general practice and for individual patient records.[66, 67, 69] Specific requirements for data quality from each primary care practice have altered over time, but are based on markers of continuity of recording and a recorded mortality rate within predicted parameters. Each practice is considered 'up-to-standard' for the latest continuous period of time during which the practice records have met the quality and continuity standards.[66, 67, 69] Quality requirements for individual patient records include a valid gender, year of birth, reason supplied for any transfer out from the practice, and permanent registration. Only patients with acceptable records are included in the database made available to researchers. The commonest reason for rejecting a patient's record is temporary registration.[69] In May 2011, when the data for this thesis were extracted, all CPRD practices were 'up-to-standard' and 11,287,981 (88.1%) of the patients had acceptable records.[68]

One well-recognised quality issue is the over-estimation of incidence rates in the initial period following a patient's registration with the practice. Disease symptoms may lead some patients to register with a new practice, biasing observed incidence rates upwards for the time shortly after registration. In addition, pre-existing co-morbidities or past major medical events are often entered without distinction from new diagnoses, during the early patient visits in which previous medical history is established. Lewis *et al.* found that this

increased incidence rate following new patient registration returned to baseline within 6 months for most acute conditions (including pneumonia), and within a year for most chronic conditions (including diabetes mellitus).[71]

A 2010 systematic review of validation studies in CPRD found 357 validations for 183 diagnoses. Most studies estimated the positive predictive value of a CPRD diagnosis (the proportion of cases recorded in CPRD which represented true cases) by requesting confirmation from the patient's GP. Estimates of positive predictive value of a diagnosis in CPRD were generally good: the median proportion of cases confirmed by validation was 89% (range 24-100%). Fewer studies calculated sensitivity (the proportion of true cases identified as such in CPRD) or specificity (the proportion of patients without a disease correctly identified in CPRD as non-cases), but those that did found high validity for the diagnoses studied.[72] Disease rates in CPRD have been compared against other UK sources for 99 different diagnoses, including diabetes, pneumonia, chickenpox and asthma, providing some indirect support to claims of both good validity of recorded diagnoses and representativeness of the CPRD population to the UK.[72]

3.2.3 CPRD data structure

Data are recorded as a combination of free-text and coded data. For reasons of patient confidentiality, information entered as free text is not routinely available.[66] These may include comments on encoded material, or longer communications such as letters to or from secondary care. The essentials of longer free-text communications, such as conditions diagnosed in secondary care, may also be entered as encoded records.[69]

Data are available in a set of files, sorted by type of information (**Table 3.1**). Prescriptions are recorded in the therapy file. These are encoded using the Multilex product coding system, and also contain the BNF chapter for the prescription. CPRD provide a dictionary of medication product codes ("the CPRD Product Browser"), which may be searched by drug name and BNF code. Immunisations are recorded in the immunisation file using product codes and a CPRD variable labelled "immstype" which also records the vaccine type. A wide range of information including medical diagnoses, clinical symptoms and signs, and some lifestyle factors and test results are encoded using the Read code system. Read codes may be recorded in clinical, referral, immunisation and test files. CPRD provide a dictionary of all Read codes ("the CPRD Medical Browser") which occur in the database.

Structured data are recorded using CPRD 'entity types'. These are used to record test results and clinical data entered using templates (such as blood pressure recordings).The

entity type variable encodes the record type, e.g. 'Blood pressure recording'. Each entity type has up to seven data fields attached which contain the test results or template contents. CPRD provide a key for each entity type of what it records, how the data are structured, and look-up tables to translate each data field. Entity types with all data fields are recorded in additional clinical details files and test files. Entity codes may also be recorded in clinical files, but without accompanying data fields.

Thus: diagnostic records are available as Read codes in clinical, referral, immunisation and test files; prescription records are available as product codes in therapy files; and test results are available as entity codes in test and additional clinical details files and may also be entered using Read codes in clinical, referral, immunisation and test files.

Table 3.1: Data files in the Clinical Practice Research Datalink (CPRD)

File	Selected contents
Patient	Patient demographics (gender, year of birth, death date) and registration dates
Practice	Practice details (region, date of last data collection, date since which practice has met data quality standards)
Staff	Practice staff details (gender and role)
Consultation	Details of the type of consultation (e.g. home visit)
Clinical	Medical history details
Additional Clinical Details	Information recorded using templates, e.g. blood pressure recordings
Referral	Patient referrals to external centres (specialty, urgency, type e.g. day case)
Immunisation	Immunisation records
Test	Test results
Therapy	Prescriptions issued in primary care (Multilex product code, BNF code, quantity)

BNF, British National Formulary

3.2.4 Principles of data management in CPRD

To identify each diagnosis in CPRD, a Read codelist was compiled using a combination of text-based and hierarchical searching. The CPRD Medical Browser was first searched using text terms, to identify an initial codelist. Relevant Read code headings were identified from this codelist, and used to search the CPRD Medical Browser hierarchically, and the two codelists were then combined.

To identify prescriptions, a medication code lists was compiled using the CPRD Product Browser using search terms identified from the British National Formulary (BNF).[73]

To identify test results or other structured data records, the relevant entity codes were identified from CPRD meta-data.

Records were then extracted from the relevant data files, and merged with a patient list to identify events of interest among the study population. Records were cleaned and any required algorithms applied, to identify the event of interest. The detailed methods used to define particular diagnoses are described in **Chapter 4** for infections, **Chapter 5** for CKD, and **Chapter 6** for other variables.

3.3 Data linkages with CPRD

CPRD patient records are linked to other routine datasets for English general practices, subject to practice-level consent.[68] This thesis used linkages to inpatient secondary care records from Hospital Episode Statistics (HES), mortality data from death registrations and socio-economic data from the national census, the latter both held by the Office for National Statistics (ONS).

All linkages were performed by a third party (the Health and Social Care Information Centre), so that patient anonymity to the database users was maintained.[74] Linkage between CPRD and HES was based on the patient's NHS number, gender, and partial date of birth. Linkage between CPRD and mortality data was based on the patient's NHS number, supplemented with date of birth and postcode. Socio-economic status estimates were linked to CPRD records by the postcode of patient residence.

3.4 Secondary care data from Hospital Episode Statistics

Hospital Episode Statistics (HES) is a database of all attendances at NHS hospitals in England. This study used anonymised HES inpatient admissions data, which have been collected from 1989 onwards.[75] The seventh version of the CPRD-linked HES inpatient admissions dataset was used.

3.4.1 HES data collection

Hospital care providers submit coded records of all hospital inpatient admissions to the Secondary Uses Service (SUS). Data are encoded retrospectively, typically by administrative coders, using inpatient records and hospital discharge summary. Since the phased introduction of Payment by Results in the mid-1990s, payment for hospital care providers has been based on activity levels calculated from these reported data, which has thus

provided a strong financial incentive for full recording of patient diagnoses and treatments, and admission and discharge dates, since before the start of our study period.[76] Data are extracted monthly and annually from SUS and cleaned to form Hospital Episode Statistics.[77]

3.4.2 HES data quality

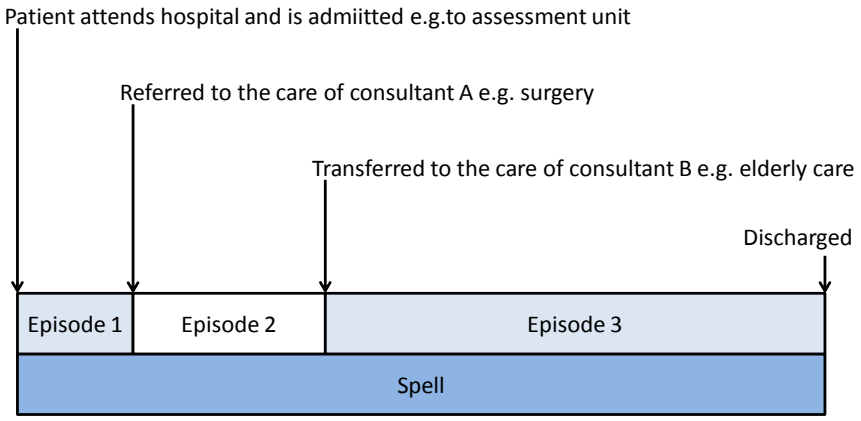
Inpatient data collection includes a system of checks designed to assure data quality. Tools to check the validity of reported data are made available to hospital care providers, to encourage good quality data coding at source. SUS data submissions are audited for completeness and invalid data formats, with results fed back to hospital care providers. HES data extracted from SUS are also cleaned and checked for internal validity.[77]

The effectiveness of these data quality checks is unclear. The Audit Commission audits the payment claims against the cost of care provided annually. Errors in clinical coding vary from 0–20% by hospital care provider when considered in terms of whether an appropriate payment was claimed.[78] The accuracy of clinical coding considered in terms of whether the diagnoses recorded provide an accurate assessment of the health status of each patient from a clinical perspective is less clear. The Commission noted that co-morbidities were inconsistently recorded, and that coding was more accurate when there was clinical overview of the diagnoses recorded.[78]

3.4.3 HES data structure

HES data are structured in episodes, which together form spells. Any period of time a patient spends continuously in hospital is referred to as a spell, which spans the time from hospital admission to discharge. When the clinical team responsible for a patient's care changes, a new episode is recorded within that spell **(Figure 3.1)**. A set of up to 20 diagnoses are recorded for each episode, of which the first (primary) diagnosis for the first episode encodes the main reason for admission.[75] The coding system is based on the 10th revision of the International Classification of Diseases (ICD-10).[79]

Figure 3.1: Hospital Episode Statistics (HES) data structure



Reason for admission: primary diagnosis of episode 1

The HES dataset is provided as a set of text files. The files used in this thesis are summarised in **Table 3.2**.

Table 3.2: Selected data files in Hospital Episode Statistics (HES)

File	Selected contents
Patient file	Patient identifier, ethnicity
Episode file: one record per episode	Spell number (constant for all episodes which together form a single spell in hospital for a patient) Episode key (identifier for each episode of care) Admission date (Date of admission for the spell) Discharge date (Date of discharge for the spell) E-order (field which identifies the order of episodes within a spell)
Episode diagnosis file	Episode key (identifier for each episode of care) ICD-10 codes for the episode D-order (field which identifies the order of diagnoses within an episode)

3.4.4 Principles of data management in CPRD-HES linked data

ICD-10 codes for a diagnosis of interest were selected using a hierarchical approach from the ICD-10 classification.[75] These were extracted from the *'episode diagnosis'* files, and merged with the *'episode'* files using the episode key to locate the diagnosis within a hospital spell. The primary reason for admission was identified as the diagnosis recorded as the primary diagnosis (d-order value of 1 in the episode diagnosis file) for the first episode of an admission (e-order of 1 in the episode file).

3.5 Mortality data from the Office for National Statistics

This study used anonymised Office for National Statistics (ONS) mortality data derived from death registration records. Updated mortality data were obtained during this study. The

original version of ONS mortality data was used to identify exit from the study cohort for objectives 2–4, and to describe mortality following infection in the analysis of burden of infection among patients ≥65 with diabetes (objective 2). The updated version was used for the analysis of the association of CKD with post-infection mortality (objective 5).

3.5.1 ONS data collection

Date and cause of death for all deaths in England and Wales are collated from civil registration records. Information on cause of death is recorded on a death certificate by a doctor in approximately 80% of cases, and these are coded automatically by a software programme. Causes of death are recorded in free-text by a coroner (following post-mortem and/or inquest) in approximately 20% of cases, and these are coded manually.[80]

3.5.2 ONS data quality

Mortality data derived from death registration records in the UK are high quality, with good completeness, systematic coding rules applied consistently to clinically-led diagnosis selection, and low use of ill-defined codes.[81]

3.5.3 ONS data structure

Text files contained the date of death and up to 15 causes of death. The updated dataset also included meta-data describing the quality of the linkage to CPRD.

Causes of death were coded using the Ninth Revision of the International Classification of Disease (ICD-9) until 2001, and the Tenth Revision (ICD-10) subsequently. This not only altered the format and number of codes available, but also the rule governing the selection of the underlying cause of death (Rule 3). This rule states that if the condition which would otherwise be selected as the underlying cause of death "is obviously a direct consequence of another reported condition... select this primary condition". In particular, this has reduced the number of deaths with underlying cause of death assigned to pneumonia. The conditions considered underlying causes if they are co-reported differ according to the precise code used to record pneumonia, but may include any conditions that impair the immune system, wasting diseases (such as cancer), diseases causing paralysis (such as stroke), and chronic lung disease.[80]

A bridging study used both ICD-9 and ICD-10 coding to assign underlying cause of death to deaths which occurred in 1999, and compared the results. For nearly half of all deaths with pneumonia assigned as the underlying cause of death using the ICD-9 coding system, the underlying cause of death was reassigned using the ICD-10 coding system. Most of the affected deaths were reassigned to Chapter IX (Diseases of the circulatory system, which

includes stroke and ischaemic heart disease), and a substantial proportion were reassigned to dementia, Parkinson's disease, and neoplasms.[82] Hence, recording of infection and particularly pneumonia as a specific cause of death is vulnerable to artefactual change, and full capture of infection-related mortality necessitates use of both underlying and contributory cause of death ICD codes.

3.6 ONS socio-economic status data

The study used linked ONS socio-economic status data for individual patients, based on residence.

Socio-economic status was described using the index of multiple deprivation (IMD), a composite area-level marker of deprivation.[83] Deprivation is identified in seven dimensions: housing, income, employment, education and training, health and disability, living environment, and crime. These are combined to form a single estimate of multiple deprivation, which is calculated at the level of lower super output area, a geographical unit with a population of approximately 1,000–3,000. The lower super output areas are then ranked, and divided into quintiles. ONS IMD estimates from 2007 were used.

3.7 Software

Data management and analysis were conducted in Stata unless otherwise specified: version 12.0 for objective 2, and version 13.0 for objectives 3–5.

3.8 Ethics

This study formed part of a project led by Dr Sara Thomas on the burden, determinants and outcomes of severe infections in older people in the UK, which received ethics approval from the LSHTM ethics committee (LSHTM ethics reference 6116) and GPRD/CPRD data use approval by the CPRD ethics committee (ISAC reference 11_033).

An amendment which provided further detail of the definition of chronic kidney disease as part of aim 4 of the original protocol, was submitted to and approved by the CPRD ethics committee (ISAC reference 11_033A). This amendment made no changes to the original protocol.

3.9 Identifying the study cohort

The primary study population comprised patients registered in CPRD aged ≥65 years with diabetes mellitus (study population A, **3.9.1**). The rationale for studying older people with diabetes mellitus is discussed in **1.4.1**.

For analyses of the association of CKD with infection incidence or vaccine effectiveness among patients with CKD (objectives 3 and 4), patients were required to have a valid serum creatinine result, and patients with a history of renal replacement therapy were excluded (study population B, **3.9.2**).

To study the association between CKD and short-term post-infectious mortality (objective 5), the study population was additionally restricted to patients with available linkage to ONS mortality data, and the cohort was identified anew using the updated linked ONS mortality data (study population C, **3.9.3**).

3.9.1 Study cohort A: patients aged ≥65 years with diabetes mellitus

The primary study population comprised all 219,145 patients registered in CPRD during the study period (1 April 1997 – 31 March 2011) aged ≥65 years with diabetes mellitus.

To ensure data quality, patients were only eligible if both patient and practice data met CPRD quality standards (**3.2.2**). To allow good assessment of co-morbidities at baseline, and prevent over-estimation of infection incidence (**3.2.2**), patients were required to be registered with the practice for at least a year before study entry.[71]

Patients entered study cohort A at the last date of: diabetes diagnosis date, 65th birthday, one year anniversary of current registration at the practice, date practice data reached quality control requirements 'up-to-standard', or 1 April 1997. Patients exited study cohort A at the first date of: date of death, leaving the practice, last date of data collection from the practice, or 31 March 2011.

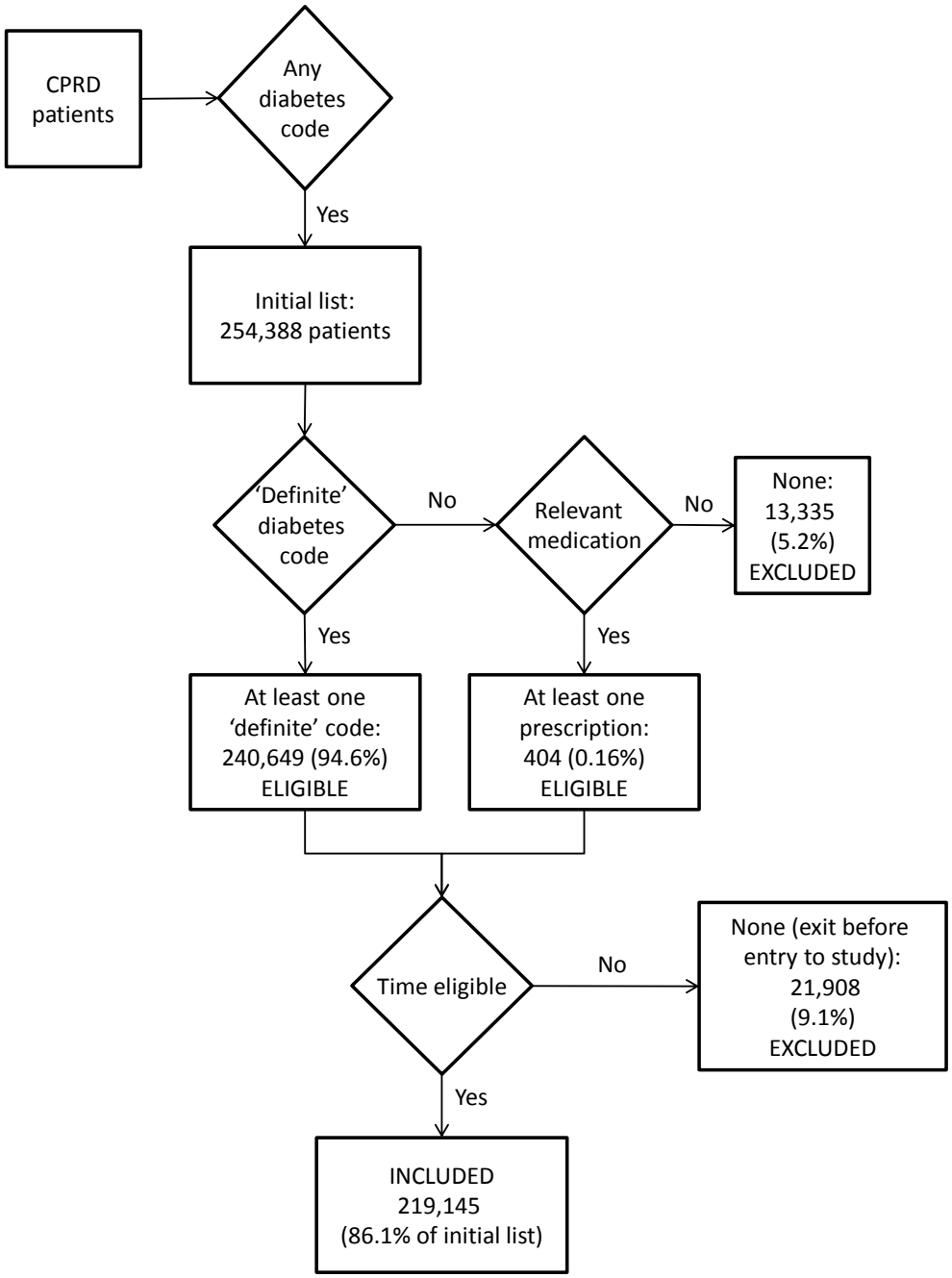
Diabetes mellitus was identified from diagnoses recorded in primary care. Two Read codelists were used, both compiled by Dr Sara Thomas, listed in **Appendix B**. The first codelist contained Read codes which were considered to encode a definite diagnosis of diabetes. Any one of these 'definite' codes identified a patient with diabetes mellitus. The second codelist contained Read codes which suggested a possible diagnosis of diabetes. These codes identified a patient with diabetes only if the patient had also received a prescription for an antidiabetes medication. I compiled product code lists for antidiabetes

medications, which are discussed in **6.6.3** and listed in **Appendix B**. The 'diabetes diagnosis date' was the earliest date of any qualifying diagnosis code or prescription.

To maintain patient anonymity, CPRD provides year but not day or month or birth. To calculate age from year of birth, all patients were assigned a nominal birthday of 1 July. Date of death was identified from ONS mortality records if the patient had a death recorded in linked ONS data, and from CPRD date of death records if not.

Identification of the cohort is described in **Figure 3.2**. In total, 13.8% (35,186/254,388) of all patients were excluded: 13,335 patients had only 'possible' codes for diabetes, with no relevant medications to confirm diabetes; and 21,850 patients met a criterion for exit before becoming eligible for entry, and thus contributed no time to the study. Among the 219,203 eligible patients, the median time eligible was 3.9 years (range 29 days to 14.0 years).

Figure 3.2: Defining study cohort A: patients aged ≥65 years with diabetes mellitus



CPRD, Clinical Practice Research Datalink

3.9.2 Study cohort B: patients with a serum creatinine test and no history of renal replacement therapy

Study cohort B comprised the subset of study population A who had a valid serum creatinine result recorded, after patients with a history of renal replacement therapy had been excluded.

Excluding patients with a history of renal replacement therapy

Renal replacement therapy was identified in CPRD by a Read code for renal transplant or dialysis. The Read code list is described in **5.3** and listed in **Appendix D**.

For objective 2, patients with a history of renal replacement therapy in CPRD prior to the study were excluded, and patients with a first incidence of renal replacement therapy in CPRD during the study were censored from this date.

Subsequently, identification of renal replacement therapy was improved by additionally excluding patients with an ICD-10 code for renal transplant or dialysis in HES (listed in **Appendix D**). A sensitivity analysis was conducted for objective 2 in which this was also applied to study cohort B: an additional 62 patients were excluded, and the study exit date was altered for 440 patients. Results of this are presented in **8.4**.

Identifying patients with a valid serum creatinine result

A valid serum creatinine result was required to define the patient's CKD status.

Clinical reasons for testing serum creatinine include acute illness, particularly infection, and patient characteristics which may be related to infection. There is therefore a risk of ascertainment bias in estimates of the association between CKD and infection from greater creatinine recording among patients with higher incidence of, or potential mortality from, infection. This risk may be mitigated by routine monitoring and complete recording. Several of the entry criteria to the study cohort (such as a diagnosis of diabetes mellitus, age ≥65 years, or practice recording reaching required CPRD data standards) may be associated with completeness of serum creatinine monitoring or recording. We examined completeness of serum creatinine recording in the year prior to entry to study cohort A compared to the first year after entry to study cohort A among patients without a history of renal replacement therapy. Serum creatinine recording was more complete after entry to study cohort A (**5.4.2**).

Classifying patients' CKD using serum creatinine results prior to fulfilling the eligibility criteria for study population A would therefore have increased the risk of ascertainment

bias in estimation of the association between CKD and infection incidence. It was therefore decided that patients would only enter study cohort B from the date of the first valid serum creatinine test after fulfilling other eligibility criteria. Serum creatinine results were identified and cleaned as described fully in **5.4.1**.

Cohort B entry and exit dates

Patients entered study cohort B on the date of the first valid creatinine result recorded after the last time-point of: diabetes diagnosis date, 65th birthday, one year anniversary of current registration at the practice, date practice data reached quality control requirements 'up-to-standard', or 1 April 1997.

Patients exited study cohort B at the first date of: date of death, leaving the practice, last date of data collection from the practice, renal replacement therapy (dialysis or renal transplant) or 31 March 2011.

3.9.3 Study population C: patients with ONS mortality data linkage

Analysis of the association of CKD with short-term post-infection mortality (objective 5) was restricted to patients with an available linkage to ONS mortality data, to optimize timely ascertainment of mortality. An updated linkage to ONS mortality data was available, and so the study cohort were defined anew using the date of death recorded in the updated ONS dataset.

The updated mortality dataset recorded deaths registered from 1 January 1998 to 10 January 2012, and so the study period was restricted to start on 1 January 1998. Dates of death in CPRD were not used to identify patient death for the cohort exit date. As death would not be identified after the date on which the datasets linkage was established, the date on which record linkage was checked (linkage date) was additionally used to define cohort exit.

Study cohort C was otherwise defined using the same criteria as cohort B.

Cohort C study entry and exit dates

Patients entered study cohort C on the date of the first valid creatinine result recorded after the last time-point of: diabetes diagnosis date, 65th birthday, one year anniversary of current registration at the practice, date practice data reached quality control 'up-to-standard', or 1 January 1998. Patients exited study cohort C at the first date of: date of death (recorded in the updated ONS dataset), leaving the practice, last date of data

collection from the practice, renal replacement therapy (dialysis or renal transplant in CPRD or HES), ONS data linkage date, or 31 March 2011.

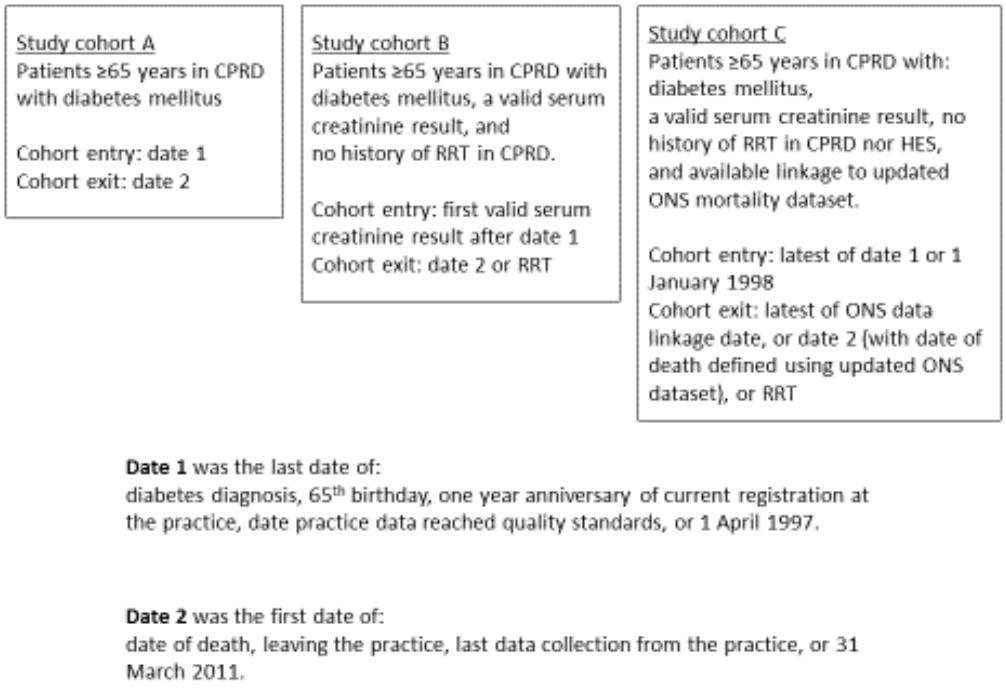
3.9.4 Relationship between study cohorts

Study cohort A comprised all patients in CPRD with diabetes mellitus aged ≥65 years. Study cohort B represented the subset of cohort A who had valid serum creatinine result and no history of renal replacement therapy. These restrictions were introduced to permit analysis of infection incidence and vaccine effectiveness according to CKD status, excluding patients with end-stage renal disease (Objectives 3 and 4).

Study cohort C would have been the subset of cohort B who had ONS mortality data linkage available. However, an updated mortality dataset linkage was available. Defining the cohort anew with this dataset offered the opportunity to study a larger cohort for this objective. As date of death was defined using an updated mortality dataset for study cohort C, which could alter the date of study exit to be earlier or later, study cohort C was not a direct subset of study cohort A.

The relationship between study cohorts A, B and C is illustrated in **Figure 3.3**. Any additional study exclusion criteria (such as exclusion of patients with missing data) are described in the chapter presenting the relevant data analysis.

Figure 3.3: Overview of eligibility criteria for study cohorts A, B and C



CPRD, Clinical Practice Research Datalink; RRT, renal replacement therapy; HES, Hospital Episode Statistics; ONS, Office for National Statistics

Chapter 4. Definition of infections

Four acute, community-acquired infections were studied: urinary tract infection (UTI), lower respiratory tract infection (LRTI), pneumonia (as a subset of LRTI) and sepsis.

The strategy for defining episodes of infection was based on previous work by Elizabeth Millett and Sara Thomas in defining LRTIs and pneumonia.[5] The general principles were that: infections were identified from clinical diagnoses in primary or secondary care; diagnoses recorded within 28 days of one another were treated as a continuation of the same infection; hospital-acquired infections were excluded; time in hospital was excluded from time at risk of that infection; and combined CPRD and HES data were used whenever HES-linked data were available for a patient.

I used the data management files created by Elizabeth Millett and Dr Sara Thomas to define episodes of LRTI (including pneumonia), and adapted these to define episodes of other infections **(Elizabeth Millett, PhD thesis, LSHTM, unpublished)**. Both Elizabeth Millett and Dr Sara Thomas provided extensive practical advice and support towards applying their methods to this study's dataset, and adapting them for UTIs and sepsis. Elizabeth Millett's methods for defining LRTI and pneumonia are described in **4.1**, followed by adaptations for this thesis to sepsis and UTI.

4.1 Lower respiratory tract infections (LRTIs), including pneumonia

4.1.1 LRTI diagnosis codes

The Read code list used to identify clinical diagnoses of LRTI in CPRD, and the ICD-10 code list used to identify clinical diagnoses of LRTI in HES, were developed by three clinical epidemiologists (Dr Sara Thomas, Dr Jennifer Quint, a respiratory physician, and Prof. Liam Smeeth, GP). The codes are listed in **Appendix C**. Exacerbations of chronic obstructive pulmonary disease (COPD) were not included unless they specified an infectious aetiology. Codes identified as pneumonia codes defined both an LRTI and a pneumonia. Codes such as acute bronchitis and influenza identified LRTI but not pneumonia. The Read code list included one diagnosis for a post-operative infection, H262.00 *'Postoperative pneumonia'* which was used to identify hospital-acquired infections (**4.1.4**).

4.1.2 CPRD data extraction

CPRD records which included a Read code for LRTI were extracted from clinical, test and referral files. There were no instances of the listed Read codes in the immunisation files.

4.1.3 Definition of LRTI episodes in CPRD

An LRTI diagnosis code defined the incident date for an LRTI if there were no LRTI diagnoses recorded within the previous 28 days. A diagnosis recorded within 28 days of a previous LRTI diagnosis continued the same episode of LRTI. The duration of each episode of LRTI was from the incident date until 28 days after the last diagnosis of LRTI within the episode (**Figure 4.1: A**).

4.1.4 Identification of hospital-acquired LRTI episodes using unlinked CPRD data

If the episode of LRTI included the diagnostic code H262.00 *'Postoperative pneumonia'*, the entire episode was considered an episode of hospital-acquired pneumonia.

If the incident date of an episode was within 14 days after a CPRD record of hospitalisation (including inpatient admissions, hospice stays, day cases and non-urgent admissions), the episode was judged to be hospital-acquired (**Figure 4.1: B**). Records which defined inpatient admission were: Read codes for inpatient admission or hospice stay in a clinical, test or referral file; consultation type field recorded as 23 *'Hospital admission'* or 47 *'Hospital Inpat Rept'* in a patient's consultation file; or inpatient field recorded as 1 *'Inpatient'* in a referral file. Records which defined day case or non-urgent admissions were: Read codes for day cases or non-urgent admissions in a clinical, test or referral file; consultation type field recorded as 25 *'Day Case Report'* in a consultation file; or inpatient field recorded as 2 *'Day Case'* in a referral file. Emergency department attendances were not used to identify hospitalisations.

Hospital-acquired LRTIs were not included as outcomes, and the duration of a hospital-acquired LRTI was removed from time at risk of a community-acquired LRTI.

Episodes of LRTI which did not contain a code for post-operative infection, and did not occur within 14 days of a CPRD record of hospitalisation were considered community-acquired for patients without CPRD-HES linkage available.

4.1.5 HES data extraction and cleaning

All hospital episodes for the study population were extracted from the HES *'episode'* file. Episodes which included an LRTI were identified by extracting records which included an

LRTI ICD-10 code from the HES *'episode diagnosis'* file. LRTI diagnoses were merged into the *'episode'* files by matching records on patient identifier, spell number and the episode key.

The data were collapsed to form one record for each hospital spell, with an indicator for whether an LRTI occurred as the primary diagnosis for the first episode of the spell, or at any point during the spell. The first episode was identified using the e-order field of the *'episode'* file, and the primary diagnosis of the episode was identified using the d-order field, of the *'episode diagnosis'* file (**3.3.3**).

Hospital spells with a missing admission or discharge date, and hospital spells with a discharge date earlier than the admission date, were discarded. When a patient had two spells with the same admission date, at least one had a discharge date on the same day for all cases, and so these are likely to represent same day readmissions. Hospital spells for a single patient which overlapped in time or were nested one within another were also identified. For patients with multiple spells on one date, nested spells, or overlapping spells, these spells were combined to form one record of continuous hospitalisation from the earliest admission date to the latest discharge date. The spell with the earliest admission date (and lowest spell number for same day admissions) was used to determine whether LRTI was the primary reason for admission at the start of the combined period of continuous hospitalisation. All spells were used to identify whether there was any LRTI diagnosis at any point during the combined period of continuous hospitalisation.

4.1.6 Definition of LRTI episodes in CPRD-HES linked data

Combined CPRD and HES data were used to define LRTIs for all patients with HES-linkage available.

When a patient is admitted to hospital for a community-acquired infection, LRTI should be recorded as the main reason for hospital admission. Hospital spells with LRTI recorded as the primary diagnosis for the first episode of the spell are referred to in this thesis as "hospitalisations for LRTI", and defined an incident community-acquired LRTI, with duration from hospital admission until 28 days after hospital discharge (**Figure 4.1: C**).

Patients may also be admitted to hospital with a community-acquired infection which is not their main reason for admission (for example, if a myocardial infarction occurs secondary to a community-acquired LRTI). The LRTI which is recorded cannot be distinguished in the HES records from a hospital-acquired LRTI. Hospital spells with an LRTI diagnosis which was not recorded as the primary diagnosis for the first episode are referred to in this thesis as

"hospitalisations including an LRTI". In the absence of other evidence, these were assumed to represent hospital-acquired LRTIs, with duration from hospital admission to 28 days after hospital discharge (**Figure 4.1: D**).

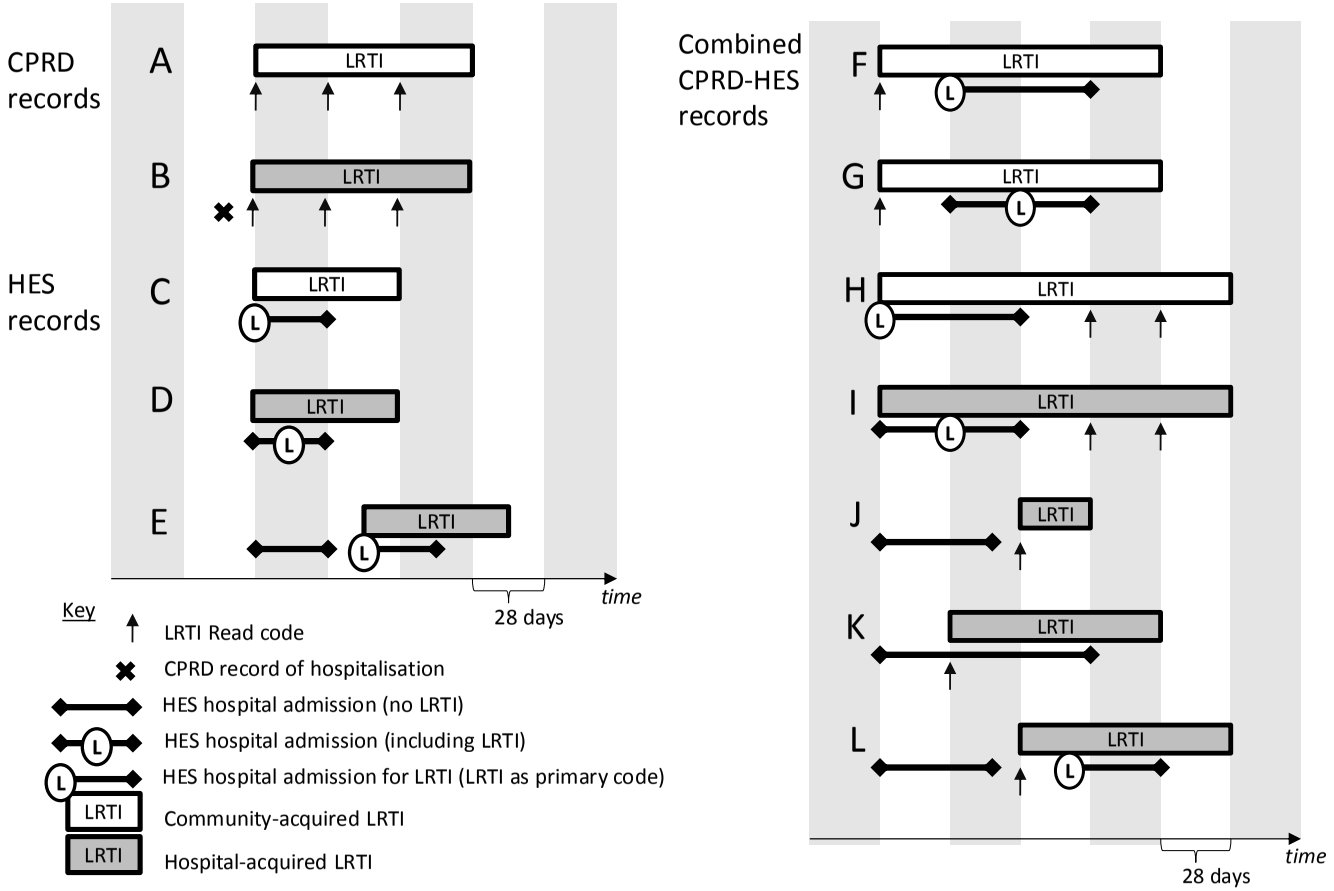
Either HES or CPRD diagnoses could continue an ongoing LRTI, without changing its status as community- or hospital-acquired. If a hospital admission for or including an LRTI occurred during a community- or hospital-acquired LRTI identified from CPRD diagnosis, the LRTI was continued until 28 days after hospital discharge (**Figure 4.1: F and G**). A CPRD diagnosis of LRTI recorded during the 28 days after discharge from a hospital spell for or including an LRTI continued the LRTI until 28 days after the last diagnosis (**Figure 4.1: H and I**).

Hospital records with no LRTI were also used to define LRTIs as hospital-acquired. Any incident LRTI occurring within 14 days after any hospital discharge recorded in HES was defined as hospital-acquired (**Figure 4.1: E and J**). A diagnosis of LRTI in CPRD which was during a hospital spell recorded in the linked HES data was also considered a hospital-acquired LRTI, with duration until 28 days after hospital discharge (**Figure 4.1: K**).

An exception was made when a CPRD diagnosis of LRTI was recorded on the same day as a HES admission for a spell which did not have any record of LRTI in the spell. It was considered possible for these patients that the patient had been diagnosed as having possible pneumonia in primary care, and referred to hospital for investigation and admission. However, as LRTI was absent from all records of that hospital spell it appeared likely that investigation in secondary care had provided an alternative diagnosis. In this case, the increased investigation possible in secondary care (for example, the easy availability of chest X-rays) were considered to cast doubt upon the primary care diagnosis. Thus, if a CPRD diagnosis of LRTI occurred on the same day as a hospital admission for a spell with no record of LRTI in the spell, this was not used to define an LRTI.

Any one episode of LRTI could combine multiple records of LRTI in CPRD and HES in any order; the same principles were applied (**Figure 4.1: L**).

Figure 4.1: Definition of hospital-acquired and community-acquired lower respiratory tract infections (LRTIs) using HES-linked records (summary of method developed by E.Millett)[1]



LRTI, Lower Respiratory Tract Infection; CPRD, Clinical Practice Research Datalink; HES, Hospital Episode Statistics

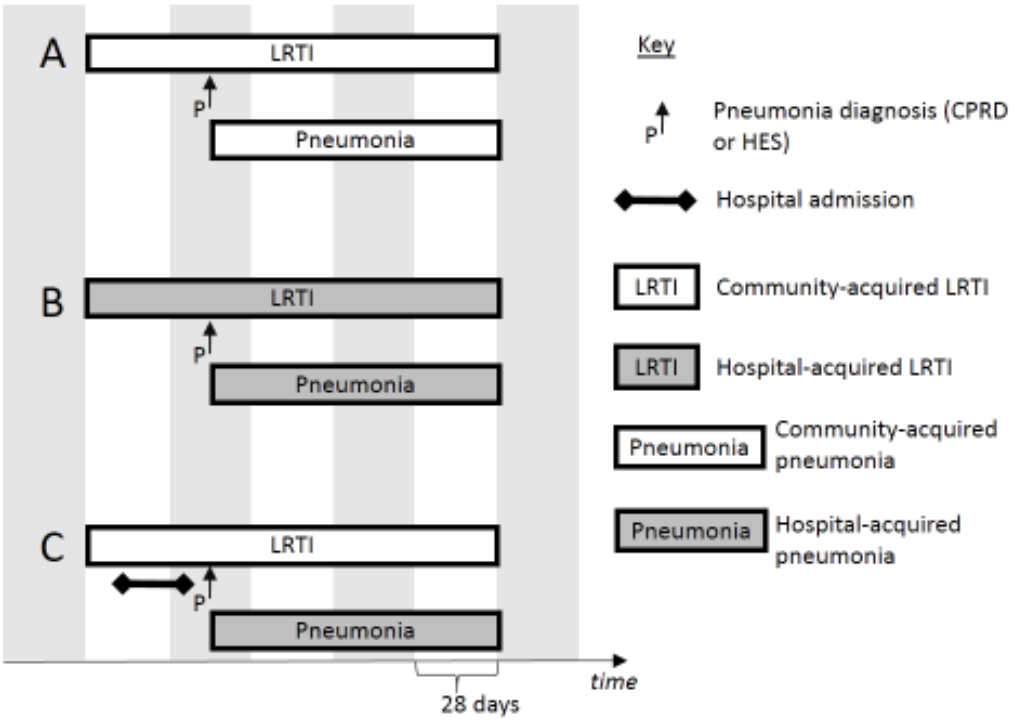
1. Millett E., Ongoing PhD thesis London School of Hygiene & Tropical Medicine, UK

4.1.7 Pneumonia as a subset of LRTI

Pneumonia was identified as a subset of LRTI. Codes which defined pneumonia diagnoses are listed in **Appendix C**. The first pneumonia diagnosis within an LRTI defined the pneumonia incident date. The duration of the pneumonia episode was from the pneumonia incident date to the end of the LRTI within which it occurred.

A major difference between community-acquired and hospital-acquired pneumonia is that each is associated with a different profile of pathogens.[84] If a pneumonia episode developed from a hospital-acquired LRTI, the causative pathogen was likely to be hospital-acquired. If a patient was admitted to hospital with a community-acquired LRTI, any subsequent progression to pneumonia might be caused by hospital-acquired pathogens. To ensure we classified as community-acquired only those episodes of pneumonia which were likely to be caused by pathogens acquired in the community, pneumonia was defined as community-acquired only if it was within a community-acquired LRTI and there was no hospitalisation between the LRTI incident date and the pneumonia incident date (**Figure 4.2**).

Figure 4.2: Defining community-acquired and hospital-acquired pneumonia episodes



LRTI, Lower Respiratory Tract Infection; CPRD, Clinical Practice Research Datalink; HES, Hospital Episode Statistics

4.2 Sepsis

Clinical diagnoses were used to identify community-acquired and hospital-acquired episodes of sepsis, applying the same strategy as described above for defining episodes of LRTI.

4.2.1 Identification of diagnostic codes for sepsis

The outcome of interest was clinically diagnosed acute, community-acquired sepsis. Sepsis is defined as a systemic inflammatory response in the presence of infection.[85] Bacteraemia, or disseminated infections which commonly arise through haematogenous spread, may occur in the absence of a systemic inflammatory response, and so these codes were not considered sufficient evidence to identify sepsis, and were not included in the diagnosis code lists. Likewise, positive blood culture tests may occur without a systemic inflammatory response, and were not considered sufficient evidence of a clinically relevant episode of sepsis. Systemic infections which are commonly subacute, or chronic, or with a long latent period, such as disseminated blastomycosis or systemic cryptococcosis, were not considered acute, and were not included.

I identified diagnostic Read codes for sepsis using the search strategy in **Table 4.1**, which started with a text-based search and then also used the identified codes to search hierarchically. I combined this list with a Read code list compiled separately by Dr Sara Thomas and discussed this with Dr Sara Thomas and Dr Dorothea Nitsch to compile a final set of Read codes, listed in **Appendix C**. Codes specifying that sepsis was post-operative or a result of a blood transfusion were identified, but none were recorded for the study population, and so these were not used to identify hospital-acquired episodes of sepsis.

Table 4.1: Search strategy for sepsis Read codes

Search details	CPRD Medical browser version 1.3.2 Database build 'ever' Search date 10 April 2012
Stage 1	Text-based search
Read terms	*sepsis*, *septic*, *blood*infect*, *blood*poison*, *endocard*, *pyaem*, pyem*, *viraem*, *virem*, *fungaem*, *fungem*, *system*mycos*, *system*inf*, *SIR*, *bacteraem*, *bacterem* or *culture*
Stage 2	Identifying relevant Read code hierarchies
	Sorted results of text-based search by Read code to identify relevant Read code headings.
Stage 3	Hierarchical search
Read codes	14O*, A*, G51*, G54*, H5*, J65*, J66*, J67*, L04*, L09*, L29*, L40*, L43*, Q40*, Qyu*, R05*, R10*, SL4*, SP2*, SP3*

CPRD, Clinical Practice Research Datalink

ICD-10 codes were searched hierarchically to identify diagnostic codes for sepsis. Two codes were identified which were not sufficient to identify a sepsis episode, but could have been consistent with a diagnosis of sepsis. These were: B349 *'Viral infection'*, unspecified and A499 *'Bacterial infection'*, unspecified. These would have been used to confirm a diagnosis of sepsis if they occurred during a hospital spell with a CPRD diagnosis on the day of admission, but there were no instances of this combination. The ICD-10 code list for sepsis diagnoses is listed in **Appendix C**.

4.3 Urinary tract infections

Acute urinary tract infections (UTIs) were identified from clinical diagnoses using the same strategy as used to define LRTIs, modified to incorporate diagnostic codes for suspected UTI, and with a difference in how a primary care diagnosis on the same day as hospital admission was classified.

4.3.1 Identification of diagnostic codes for UTIs

The outcomes of interest were acute, community-acquired infections of the bladder and upper urinary tract. Urethritis and male accessory gland infections such as prostatitis have different clinical presentations from bladder and upper urinary tract infections, with different risk profiles.[86] We considered these to be separate clinical entities from bladder and upper urinary tract infections. As less severe conditions, they may also be more vulnerable to ascertainment bias from health-seeking behaviour, diagnosis and recording bias. We thus did not include urethritis and male accessory gland infections as UTIs. While UTIs may cause positive urine cultures, these may also occur in the context of asymptomatic bacteruria. To ensure we only identified clinically-relevant infections, we used clinical diagnoses to identify UTIs, and did not use urine bacterial culture test results.

Diagnostic codes for UTI were identified by Dr Sara Thomas and code lists were finalised in discussion with Dr Dorothea Nitsch and myself. The Read and ICD-10 codes selected to define UTIs are listed in **Appendix C**. Codes for 'recurrent UTI' were assumed to be recorded in the context of a current infection. Codes which specified that a UTI had occurred in pregnancy were assumed to be historical in this age group and were not used.

If Read code K190200 *'Post operative urinary infection'* occurred during a UTI episode, the entire episode was classified as hospital-acquired.

Codes for pyelonephritis were vulnerable to cross-use to describe chronic kidney disease. For example, acute renal infections are recorded in HES records using the ICD-10 codes N10

'Acute tubulo-interstitial nephritis' or N12 *'Tubulo-interstitial nephritis, not specified as acute or chronic'*, both with an additional code B95-99 to identify the infectious agent. If an infectious agent is not recorded, code N12 could represent a chronic tubulo-interstitial nephritis (a cause of CKD), or an episode of pyelonephritis where the infectious agent was not identified or not recorded. To avoid misclassifying patients with CKD as having episodes of pyelonephritis, a cautious approach was taken in selecting codes to identify pyelonephritis. Codes which were ambiguous or required another code to identify an infectious aetiology were not used to identify pyelonephritis.

4.3.2 Resolving conflict between primary care diagnosis and hospital records

For patients admitted to hospital on the same day as a diagnosis of UTI in primary care, the hospital records did not always include a diagnosis of UTI for the admission.

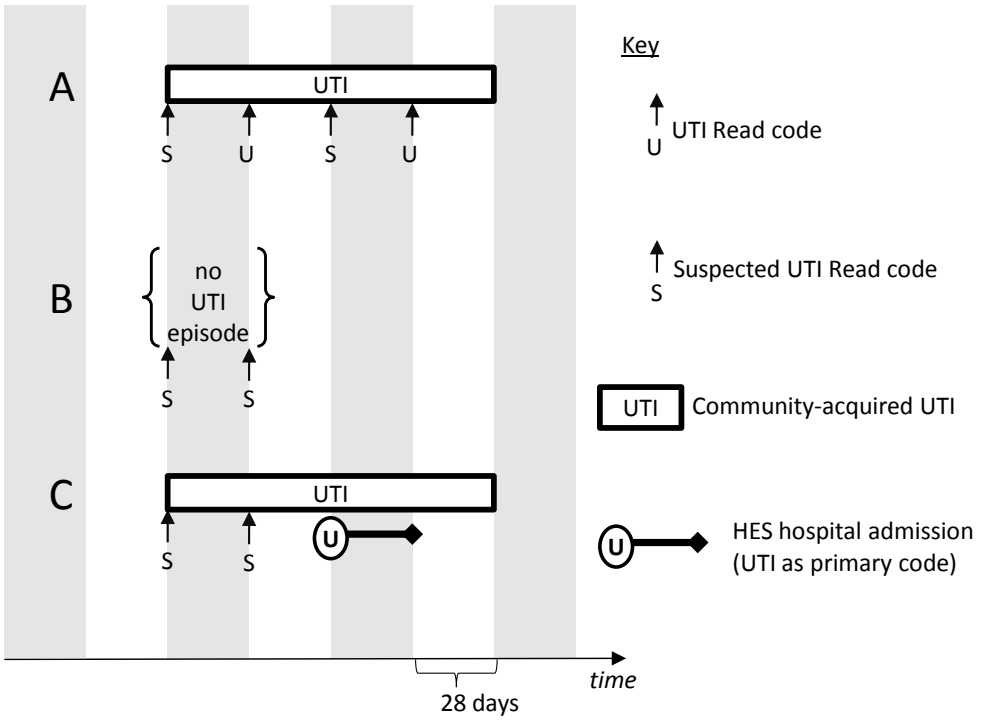
Urinary tract infection is usually a clinical diagnosis in primary care, and the diagnostic gold standard is urine microscopy and culture, which is available in primary care.[87] Primary care diagnoses were therefore treated with some confidence. It was thought plausible that recording of UTIs in hospital discharge records could be incomplete for patients admitted to hospital who had a community-acquired UTI at the time.

If a UTI was recorded in CPRD on the same day as a hospital admission, this was defined as a community-acquired UTI even if the hospital spell did not include an ICD-10 code for UTI. Otherwise, CPRD and HES records were combined to identify community-acquired and hospital-acquired UTI using the same principles as those used to define LRTI.

4.3.3 Codes for suspected UTI

In clinical practice, a patient with symptoms of UTI may be recorded as having a suspected UTI: pending the results of urine microbiology, for example. We included Read code IJ4..00 *'Suspected UTI'* in the UTI code list used to define episodes of UTI. 'Suspected UTI' codes were treated in the same way as other UTI diagnostic codes for defining the start, end or duration of any group of UTI codes within 28 days of one another to form a UTI episode (**Figure 4.3: A**). However, a group of diagnoses which comprised only codes for suspected UTI without any other code for UTI did not define an episode of UTI (**Figure 4.3:B**). Other than this requirement, suspected UTI codes were treated identically to other UTI diagnostic codes in defining UTIs in combined CPRD-HES records (**Figure 4.3: C**).

Figure 4.3: Use of Read code IJ4..00 'Suspected UTI' in defining episodes of UTI



UTI, Urinary Tract Infection; HES, Hospital Episode Statistics

4.4 Identifying infectious aetiology

Data sources identifying infectious aetiology were available within both primary and secondary care datasets. Aetiology may be recorded in diagnostic codes for infection, either as Read codes in primary care records (e.g. H21..00 *Lobar (pneumococcal) pneumonia*) or as ICD-10 codes in secondary care records (either directly in the diagnostic ICD-10 code, such as J13 *Pneumonia due to Streptococcus pneumonia*, or by attachment of a separate code recording the causative pathogen).[76] Microbiological test results sent from primary care were available in CPRD, although tests conducted in secondary care were not available in HES.

Infectious aetiology is incompletely identified and recorded for community-acquired infections: fewer than half of patients with community-acquired pneumonia have a pathogen identified.[84, 88] In part this is due to the low diagnostic yield of many microbiological tests, especially among patients who have commenced antibiotic treatment.[89] However, it is also due to selective investigation, particularly in primary care. Guidelines on the management of community-acquired UTI and LRTI in primary care

recommend that microbiological tests should be sent only for selected patients.[88, 90, 91] Impaired renal function is an indication for microbiological investigation for UTI and if associated with poorer prognosis of infection would also influence microbiological investigation more generally.[90] Thus, information on infectious aetiology was incomplete among the study population, and was likely to be differential by CKD status. Restricting any analysis to infections with identified aetiology would have included only a selected subset of patients, and would have been likely to introduce selection bias into estimates of the association of CKD with infection incidence or mortality. Description of infectious aetiology was therefore not attempted for this thesis.

4.5 Time at risk of community-acquired infection

Time at risk of community-acquired infection was identified separately for each type of infection. Patients were considered at risk of a community-acquired infection unless they were experiencing an ongoing episode of infection, or were a current hospital inpatient, or were within 14 days of discharge from hospital. Community-acquired or hospital-acquired infections were removed from person-time at risk for each patient. When HES-linkage was available, time in hospital was also removed from time at risk for each patient.

4.6 Calculating infection incidence

Infection rates were calculated using Poisson regression models. Poisson regression was selected in preference to Cox regression because it allowed me to look at how infection rates changed over several dimensions of time (for example age and calendar year). Fitting Poisson regression models is also computationally less intensive than fitting Cox regression models, which is relevant with this large dataset. Lexis expansions were used to allow adjustment for age and calendar year. The Poisson regression model assumes that baseline hazard is constant within a period of time which does not cross boundaries of age category or calendar year.

A random effects model was used to adjust for clustering of multiple infections among individual patients. Random effects models specify explicitly the between-cluster variation, and include it in the model. This adjusts both the parameter estimate and standard errors for clustering, with less weight given to infections which are part of a cluster of infections occurring to one patient. This approach also allows likelihood ratio tests to be used.

Chapter 5. Defining chronic kidney disease in electronic health records

This chapter describes the identification of chronic kidney disease (CKD) from electronic health records among older people with diabetes. CKD was classified using two separate markers of kidney disease: estimated glomerular filtration rate, and a history of proteinuria. The general principles were that: CKD status was identified from primary care records; eGFR was calculated from serum creatinine test results using the CKD-EPI equation, and classified according to clinical categories; patients without a serum creatinine test result were excluded from studies of the association of CKD with infection-related outcomes; proteinuria was identified from test results and Read code records; a history of proteinuria was a binary non-reversible exposure in which the first valid proteinuria record changed the patient's status from negative to positive for the rest of the patient's follow-up; a cautious approach was taken to identifying positive proteinuria status (for example, a 'trace' of proteinuria was not counted as positive); proteinuria records coincident with a urinary tract infection were removed; and the absence of a positive proteinuria record was assumed to indicate a negative status.

This definition of CKD was developed in the light of how CKD is classified, monitored and recorded in electronic health records, and following an exploration of data quality in CPRD. This chapter first describes the classification, identification and monitoring of CKD in clinical practice, and discusses the likely impact of changes in clinical practice over the study period (1997–2011) on CKD recording. The chapter then discusses what was already known about the validity of CKD recording in CPRD. The data cleaning and categorisation used to define CKD in this thesis are presented in detail, and alternative approaches which were explored are briefly reviewed.

5.1 Chronic kidney disease

Chronic kidney disease (CKD) is an impairment of kidney structure or function that persists for at least three months. CKD is almost always asymptomatic until advanced, and diagnosis depends on screening or incidental findings of impaired kidney structure or function. Although the kidney has many functions, the rate at which the glomerular capillaries in the kidney filter waste products, known as the glomerular filtration rate (GFR), is considered the best overall indicator of kidney function. Impaired kidney structure may be evidenced by a broad range of signs, including structural abnormalities observed on imaging, histological abnormalities, and protein in the urine (proteinuria).[10]

5.1.1 Classification of CKD

Classification of CKD in the UK has evolved over the last decade. Changes in CKD classification have been driven by accumulating evidence from large epidemiological studies of the prognostic importance of both proteinuria and even moderate reductions of estimated GFR (eGFR), particularly in combination.[10] Albuminuria and reduced eGFR are associated with increased cardiovascular events, and all-cause and cardiovascular mortality, even at early stages of CKD.[23, 92-94]

The first widely adopted classification of CKD was developed by the US National Kidney Foundation as part of the Kidney Disease Outcomes Quality Initiative (K/DOQI) in 2002 **(Table 5.1)**.[13] This described five stages of CKD. Stages 3-5 were identified based on two GFR estimations at least three months apart. Stages 1-2 additionally required other evidence of kidney damage, such as proteinuria. This classification was recommended for use in the UK by the Department of Health in 2005, and will have been used clinically in the UK for the majority of the study period.[95]

In 2008, the UK National Institute for Health and Care Excellence (NICE) recommended the division of stage 3 into 3a (eGFR 45–59 ml/min/1.73m^2) and 3b (eGFR 30–44 ml/min/1.73m^2), and acknowledgement of persistent proteinuria within each stage.[11]

Table 5.1: K/DOQI 2002 classification of chronic kidney disease

CKD stage	GFR[1]	Evidence of kidney damage required[2]	Description
1	≥ 90	Yes	Kidney damage with normal/↑eGFR
2	60-89	Yes	Kidney damage with mild ↓eGFR
3	30-59		Moderate ↓eGFR
4	15-29		Severe ↓eGFR
5	< 15 or dialysis		Kidney failure

From the Kidney Disease Outcomes Quality Initiative (K/DOQI) 2002 classification of CKD [13]
1. Glomerular filtration rate (ml/min/1.73m^2) on at least two occasions three months apart
2. Persistent proteinuria, albuminuria or haematuria, or structural abnormalities

In 2012, the Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group in the US recommended that CKD be classified by GFR status, albuminuria status, and the cause of CKD.[10] In this classification, GFR is categorised as previously, including the distinction between 3a and 3b. The ratio of albumin to creatinine in the urine (ACR) is categorised as normal/ mildly elevated (<3 mg/mmol), moderately elevated, or severely elevated (>30 mg/mmol). These categories are combined to describe the severity of CKD: for example, a patient with eGFR 50ml/min/1.73m^2 on at least two occasions three months apart, and persistent albuminuria of 20 mg/mmol has CKD status G3a A2 **(Table 5.2).** This

classification is based on evidence of the prognosis associated with the combination of GFR and albuminuria status, and represents recommended practice in the UK at the time of writing but was published after the end of the thesis study period.[96]

Table 5.2: KDIGO 2012 classification of CKD prognosis by GFR and albuminuria status

				Persistent ACR categories Description and range		
				A1	A2	A3
				<3 mg/mmol	3–30 mg/mmol	>30 mg/mmol
				Normal to mildly increased[1]	Moderately increased	Severely increased
GFR categories (ml/min/1.73m^2) Description and range	G1	>90	Normal	Low risk[2]	Moderately increased risk	High risk
	G2	60–89	Mildly decreased[1]	Low risk[2]	Moderately increased risk	High risk
	G3a	45–59	Mildly to moderately decreased	Moderately increased risk	High risk	Very high risk
	G3b	30–44	Moderately to severely decreased	High risk	Very high risk	Very high risk
	G4	15–29	Severely decreased	Very high risk	Very high risk	Very high risk
	G5	<15	Kidney failure	Very high risk	Very high risk	Very high risk

KDIGO, Kidney Disease: Improving Global Outcomes; CKD, chronic kidney disease; GFR, glomerular filtration rate; ACR, albumin:creatinine ratio

1. Relative to young adult level
2. No chronic kidney disease if no other markers of kidney disease

5.1.2 Estimating glomerular filtration rate

GFR cannot be measured directly. It can be measured indirectly by measuring the rate at which an exogenous marker of filtration, such as inulin or iothalamate, is cleared from plasma into urine. This is expensive, requires close monitoring over an interval of time, may involve exposure to radiation, and is not used in routine clinical practice.[97]

GFR is more conveniently estimated using endogenous markers of filtration, such as serum creatinine, which is simple and inexpensive to measure. Creatinine is a waste product of muscle metabolism which is eliminated renally, with free glomerular filtration and only a small component of tubular excretion. The level of creatinine in the serum reflects the equilibrium between the rate at which it is produced and the rate at which it is filtered. Creatinine production is influenced by a range of factors including age, sex, ethnicity, muscle mass and dietary protein. Among patients with low creatinine production, reduced GFR may therefore co-exist with a 'normal' level of serum creatinine.[97]

Equations such as the Cockcroft-Gault equation, which address this problem by adjusting the serum creatinine result for age, sex and weight to estimate GFR, have existed since the

1970s.[97] However, at the start of the study period, the commonest approach to identifying impaired renal function in primary care was to classify an unadjusted serum creatinine level as normal (<120 μmol/l), raised (120-150 μmol/l) or requiring nephrology referral (>150 μmol/l).[44] Thus patients with impaired eGFR but low creatinine production were not identified. As age is a strong predictor of creatinine production, this particularly affected older people. In a Canadian study, 47.3% (316/668) of primary care patients aged ≥70 years with a 'normal' serum creatinine (≤130 μmol/l) had a reduced eGFR ≤50 ml/min using the Cockcroft-Gault equation.[98]

The Modification of Diet in Renal Disease (MDRD) Study equation was published in 1999,[99] and widely adopted following its recommendation in Royal College of General Practitioner (RCGP) 2005 guidelines,[100] and automatic reporting of calculated eGFR by UK laboratories from April 2006.[20, 101] The MDRD equation tends to underestimate eGFR at higher levels of GFR, and a new estimating equation, the CKD Epidemiology Collaboration (CKD-EPI) creatinine equation was published in 2009.[102] The CKD-EPI equation has been validated among diverse populations, including older people, and provides a more accurate estimation of GFR, which is more closely correlated with prognosis than estimates using the MDRD equation.[97]

A more accurate estimation of GFR is also possible by combining creatinine with another endogenous marker, cystatin, using the 2012 CKD-EPI creatinine-cystatin C equation.[97] However, serum cystatin is infrequently measured in clinical practice, and so this is not available in routinely collected historic electronic health records.

Laboratory estimation of serum creatinine has seen improvements since 1997. Laboratory variation in estimation of serum creatinine was addressed with the phased introduction of laboratory-specific standardisation from 2006.[103] Creatinine assays are now calibrated to a reference assay using isotope-dilution mass spectrometry. The CX3 assay method used for the majority of the study period tended to a slight positive bias in creatinine estimation: serum creatinine results are 5% lower after IDMS-standardisation.[97, 104]

5.1.3 Measurement of proteinuria

A variety of tests exist to identify proteinuria. These range from urine reagent strips ('dipsticks') which provide results at the bedside within minutes, to more accurate laboratory measurement of protein: creatinine ratio or albumin: creatinine ratio.

Proteinuria, although most commonly caused by kidney damage, may result from other aetiologies (for example, multiple myeloma causes overproduction of immunoglobulin light

chains, which circulate in the plasma and are excreted by the kidneys into the urine). Albumin is a plasma protein normally found at low levels in the urine, but which at increased quantities is pathognomonic of kidney disease.[10] The urinary albumin: creatinine ratio (ACR) is more sensitive to low levels of proteinuria than the protein: creatinine ratio (PCR).[10] It is albuminuria which has been studied and found to have a graded association with all-cause and cardiovascular mortality.[92] For these reasons, urinary ACR was recommended by NICE in 2008 as the test of choice for screening for diabetic nephropathy and is considered best practice for CKD screening in general.[10, 11]

Bedside urine reagent strips are used in clinical practice for a variety of reasons, such as investigating suspected UTIs. Although ACR has been the recognised test of choice for CKD screening among patients with diabetes since at least 2008, the less precise and less renally-specific test of 'dipstick' for total proteinuria will also occur in primary care records, both for CKD screening during the early years of the study, and for other clinical purposes.

5.1.4 The effect of changes in primary care on CKD diagnosis

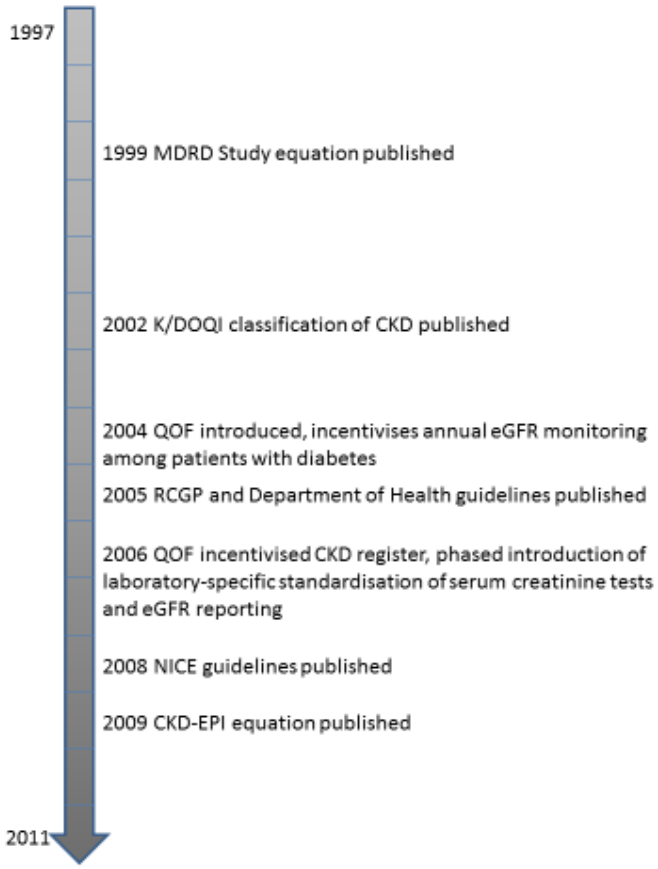
Over the study period 1997–2011, CKD diagnosis in primary care has changed considerably. Changes in CKD testing and classification have been driven by a growing evidence base for the prognostic significance of CKD, which may itself have influenced clinical practice. In particular, uncertainty as to whether early stages of CKD held any prognostic implications for older people may have reduced enthusiasm for formally diagnosing CKD among older people at the start of the study period.[105] Any hesitance is likely to have reduced as evidence has accumulated that reduced eGFR and proteinuria among older people identify individuals at higher risk of hospitalisation, and reduced eGFR is associated with a higher risk of all-cause and cardiovascular mortality, although the significance of eGFR 45–60 ml/min/1.73m^2 remains less clear among this age group than among younger adults.[105-108] The clinical value of monitoring for proteinuria has also been enhanced by finding that treatment of proteinuria with angiotensin converting enzyme inhitors can slow progression of CKD.[109]

Adoption of changes in CKD testing and classification has been reinforced by other external factors and events (**Figure 5.1**). A number of guidelines for the identification and monitoring of CKD in primary care were published during 2005-8.[11, 95, 100] The Quality and Outcomes Framework (QOF) has incentivised GPs to monitor serum creatinine or creatinine clearance of diabetic patients annually since 2004, and to keep a register of

patients with CKD using a defined set of Read codes since 2006, which are likely to have increased monitoring and standardised recording of CKD, respectively.[65]

The effect of these changes is difficult to quantify precisely, as increased CKD diagnosis could reflect a combination of increased ascertainment and increasing CKD prevalence. Many of the changes clustered around the years 2005-6 (**Figure 5.1**). Hobbs *et al.* identified a 47.5% increase in referrals to renal clinics in Kent in 2006–2008 compared with 2004–2006, and this steep and sudden increase is likely to be substantially explained by changes in clinical practice rather than a step change in CKD prevalence.[20] The increasingly routine use of eGFR rather than absolute serum creatinine level over the study period will have directly resulted in increased recognition and earlier diagnosis of CKD in primary care, particularly among older people.[110, 111] This change is likely to have been substantial: among 5,072 patients in primary care with diabetes, serum creatinine of >120 µmol/l identified only 33% of the 1,588 patients diagnosed with CKD stages 3-5 when eGFR was estimated using the MDRD equation.[46] Other changes are likely to have contributed incrementally to earlier recognition of CKD. For example, automated laboratory reporting of eGFR in Australasia was temporally associated with a 4% decrease in late referral for renal replacement therapy.[112]

Figure 5.1: Timeline of changes to identification of CKD in primary care during the study period 1997-2011



MDRD, Modification of Diet in Renal Disease; K/DOQI, Kidney Disease Outcomes Quality Initiative; QOF, Quality and Outcomes Framework; RCGP, Royal College of General Practitioners; NICE, National Institute for Health and Care Excellence; CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration

Thus CKD identification in primary care has changed enormously during the study period, and records of CKD diagnosis in CPRD will reflect this. If CKD were ascertained in our study exclusively from clinical diagnoses, there would be a risk of ascertainment bias. Rates of infection have increased over time in older individuals, and if ascertainment of CKD also increased over time due to changes in primary care practice, this could result in severe over-estimation of an association between CKD and infection rates.[5]

Throughout the many changes in clinical practice over the study period which have affected ascertainment and recording of CKD, serum creatinine testing has remained the constant underlying test for CKD in primary care.

5.1.5 Classifying renal disease by cause

Clinical guidelines recommend that clinicians consider the cause of disease when classifying CKD.[10] Investigation and diagnosis of the cause of CKD generally occurs in secondary care outpatient clinics, but would be communicated to primary care via letter, which may be recorded in free-text (and hence unavailable in the CPRD dataset) or encoded (and available).

Not all patients with CKD will routinely have underlying renal pathology investigated, and this may particularly be true among older patients with multiple co-morbidities. Renal impairment among young patients with no co-morbidities is more likely to have an underlying cause which would benefit from investigation and specialist management than among older patients with diabetes.[100]

A wide range of pathologies may cause CKD, and within each condition there may be a range of severity of CKD. Assessing the prognostic significance of the underlying cause of renal disease for a patient with CKD is informed at an individual patient level by knowledge of the presentation and progression of the renal pathology, and the patient's medical history, and relies on clinical judgment rather than a systematic classification of aetiologies.

For these reasons, the cause of disease is likely to be unavailable for most older patients with diabetes in CPRD, and population-level classification of CKD by aetiology is less practicable than classification by eGFR and proteinuria status.

5.2 CKD recording in electronic health records

Most patients with CKD are identified, monitored and managed entirely in primary care.[100] CPRD is therefore the most relevant database for identifying CKD status in this thesis.

Electronic health records are secondary data, which were not generated for the purpose of this research.[62] The primary purpose for which these data were recorded is clinical care of the patient. Data generation is driven by clinical practice: for example, screening for CKD is a response to patient risk factors. Secondary reasons for data recording include local audit and quality improvement, external pressures such as financial incentives, and awareness that the practice participates in research by submitting records to the CPRD. Aspects of data recording, information technology, and database curation may also influence how CKD is recorded in CRPD, separately from clinical changes. The valid identification of CKD in CPRD is thus dependent on a wide range of factors.

5.2.1 Data recording affects completeness and validity of CKD records

Data recording in primary care, and curation of the research database may affect the quality and availability to researchers of different data types encoding CKD status. How clinical data are recorded in CPRD in general was described in **3.2**.

The variety of tests for proteinuria will be reflected in less uniform recording of proteinuria test results than serum creatinine tests. Proteinuria tests may be recorded in a variety of formats in the database, and may frequently be unrecorded, especially if proteinuria is absent. In particular, bedside urine reagent strip tests may be recorded in free-text rather than coded, particularly if negative. In a validation study of primary care records, among 432 patients with a recorded diagnosis of CKD stage 3–5, proteinuria testing was only recorded using a Read code or structured test result record if positive. The Read code 4672 *'Urine protein test negative'* was not used at all, among 93 records of proteinuria testing recorded in structured data. Manual test searches identified tests for proteinuria recorded in free text without a Read code. Nearly half of these were negative (69/142).[113] To maintain anonymisation of the database, free-text is not available in CPRD without the considerable expense of checking by a third party for any identifying information, and thus a high proportion of negative proteinuria test results may be absent in the dataset available for this thesis.

In contrast, the shift of laboratory test result reporting in primary care from fax or letter to automated electronic reporting directly into patient records, which was almost universal by

2007,[101] is likely to have enhanced consistency of recording, reduced data entry error, and reduced the risk of under-recording of 'normal' results for serum creatinine test results.

5.2.2 What do we know about completeness of CKD recording among older people with diabetes mellitus?

Older people with diabetes are a highly-monitored population for CKD. Diabetic nephropathy has been well-recognised as a potential complication of diabetes mellitus for some time, and standard care for patients with diabetes should have included regular monitoring for CKD throughout the study period. Guidelines published during the study have consistently recommended annual screening for, and at least annual monitoring of, impaired renal function and proteinuria in primary care.[11, 95, 100] Among 5,072 patients with diabetes registered at 17 English practices in 2003–4, 92% had a serum creatinine measurement recorded in the previous 2 years.[46] There was sufficient information to estimate GFR for 4,139 (81.6%) using the simplified MDRD equation. Completeness of creatinine recording is likely to have improved since, as this study pre-dated QOF incentivisation of creatinine monitoring. Albuminuria was poorly recorded: only 36% of patients with diabetes had any record of albuminuria having been measured.[46]

5.2.3 What do we know about the validity of CKD recorded in CPRD?

The validity of classification of a variable is generally considered from two perspectives: sensitivity and specificity; or positive and negative predictive values.[62] These describe the relationship of the classification to the true underlying state **(Table 5.3)**.

Table 5.3: Describing validity: classification of binary variables compared to the true underlying state

		Recorded CKD status		
		Positive	**Negative**	**Total**
Gold standard (true CKD status)	**True positive**	a	b	a + b
	True negative	c	d	c+d
Total		a + c	b + d	a + b + c + d

Specificity is the probability of correctly classifying as unexposed a person who is truly unexposed. This can be represented as d/(c+d). Sensitivity is the probability of correctly classifying as exposed a person who is truly exposed. This can be represented as a/(a+b).

The positive predictive value is the probability that a person who is classified as exposed is truly exposed, which can be represented as a/(a+c). The negative predictive value is the probability that a person classified as unexposed is truly unexposed, which can be represented as d/(b+d).

The positive predictive value of diagnoses recorded in CPRD are generally good.[72] However, as CKD is 'silent', or asymptomatic, until quite advanced, under-ascertainment is a particularly high risk for CKD compared to many diseases. This would reduce sensitivity of CKD diagnosis recording.

Most studies of the validity of CKD coding in administrative database coding and existing datasets have investigated inpatient hospital claims databases in the US. These datasets differ considerably from CPRD, in that coding uses ICD classifications, the purpose of coding is for cost claims not case management, and the study population is usually inpatient.[114, 115]

Van Staa *et al*. conducted an external validation of a range of recorded diagnoses in CPRD against a reference standard of diagnoses on hospital discharge summaries for 500 patients. For diabetic nephropathy there was 96.1% agreement: this high agreement may be misleading, as for 413 patients with concordant results the diagnosis was absent in both sources, which may simply reflect incomplete recording in both CPRD and hospital discharge summaries.[116]

Two studies have assessed the performance of CKD Read codes in primary care records against eGFR calculated from serum creatinine. In a general practice of approximately 11,000 patients, a quarter of patients had a creatinine recorded, of whom 492 had CKD stage 3-5 according to the calculated eGFR. CKD Read code searches identified 36/492 (7.3% sensitivity compared to eGFR). Full-text manual searching of patient records only identified a further 4 cases: 452/492 (92%) were undiagnosed cases.[113] A list of 45 CKD Read codes were validated against eGFR for all subjects with at least one creatinine measurement between 2002–8 in The Health Improvement Network, a large electronic primary care dataset. Read codes had 48.8% sensitivity and 98.2% specificity to identify CKD stages 3–5 compared to recorded eGFR, with a positive predictive value of 88.9% (95% CI 88.7–89.1) and negative predictive value of 86.5%.[117]

GFR estimated from routinely monitored serum creatinine test is not a true reference standard, and may itself lack validity. For example, CKD may be over-estimated when using a single serum creatinine result.[118] It would therefore be a simplification to see these results purely as demonstrating the sensitivity lost downstream of creatinine testing by the requirement for clinical interpretation, diagnosis and Read code recording. These internal validations may also not be generalisable to patients without a recorded serum creatinine test.

However, the results do suggest that CKD Read codes are considerably less sensitive for identifying CKD in primary care than eGFR calculated from serum creatinine results. This inference is supported by the NEOERICA study, which studied the prevalence of CKD using primary care records for the population of Kent, Manchester and Surrey. The prevalence of stage 3–5 CKD identified using serum creatinine results (8.5%) was among the highest UK estimates to date, suggesting that this method of establishing CKD status is not unduly limited by missing creatinine results, while only 1.6% of the cohort had a diagnosis of renal disease in their primary care records.[17]

If CKD were ascertained in our study from diagnosed CKD only, there would be a greater risk of under-ascertainment of CKD than if serum creatinine results are used to estimate GFR. Non-differential under-ascertainment of CKD could result in under-estimation of the associations between CKD and infections, as the observed associations would be diluted versions of the true underlying associations. If diagnosis were influenced by clinical factors which were risk factors for infection as well as risk factors for CKD – such as smoking – use of CKD diagnosis to identify CKD in CPRD could cause ascertainment bias, in which the association between CKD and infection would be over-estimated.

5.3 Identification of patients with CKD in CPRD

As discussed in **5.1.1**, current best clinical practice is to classify CKD according to eGFR, albuminuria and cause.[10] There is good evidence that eGFR and albuminuria are both independently associated with CKD prognosis.[92] Classifying CKD status according to eGFR and proteinuria independently and according to the categories in current classification schemes would allow comparison of the relationships between CKD and infection with those between CKD and cardiovascular disease, and would facilitate translation of results into clinical practice.

This study identified patients with CKD in CPRD by first excluding patients with a history of renal replacement (**5.3.1**), estimating glomerular filtration rates from serum creatinine tests (**5.4**) and, separately, identifying patients with a history of proteinuria (**5.5**).

5.3.1 Exclusion of patients with a history of renal replacement therapy

A history of renal replacement therapy was used to exclude patients from study populations B and C (**3.9.2 and 3.9.3**).

I identified Read codes for renal replacement therapy using the search strategy in **Table 5.10 (5.5.2)**. These were sorted into codes specifying renal transplant, dialysis, and

acquired arteriovenous fistula **(Table 5.4)**. For most dialysis codes it was not specified whether dialysis was historical or ongoing, nor whether the dialysis was short-term (for AKI) or longer-term for CKD. Therefore, no distinctions were made between short-term and long-term dialysis, nor between historical and ongoing dialysis.

Codes recording the creation of an arterio-venous fistula were not used to exclude patients from the study population, as patients may be prepared for haemodialysis in advance of renal failure, and so this would have risked introducing selection bias to analyses of the association between CKD and infection. For example, if patients with fewer co-morbidities or receiving more active preventive care received arterio-venous fistulae at an earlier stage of CKD, this could create a selection bias in which patients remaining in the study at CKD stages 4-5 had higher risk of infection than those excluded, resulting in a spurious association between CKD stages 4-5 and infection risk.

Table 5.4: Summary of Read codes identifying renal replacement therapy in CPRD

Subgroup	Number of codes	Number of events in CPRD	Commonest code in CPRD
Renal transplant	24	11,718	7B00.00 Transplantation of kidney
Dialysis (any)	34	13,114	7L1A200 Haemodialysis NEC
Acquired arteriovenous fistula	16	3,008	7A60100 Creation of arteriovenous fistula NEC
Total	74	29,840	

CPRD, Clinical Practice Research Datalink

Patients with a history of renal replacement therapy were identified from CPRD using the Read codes listed in **Appendix D**.

Patients with a history of renal replacement therapy recorded in Hospital Episode Statistics were identified using a list of ICD-10 codes developed by Dr Sara Thomas (**Appendix D**).

5.4 Estimating glomerular filtration rate from serum creatinine tests in CPRD

CKD status was described among the 218,688 patients aged ≥65 years with diabetes mellitus and no history of renal replacement therapy in CPRD.

GFR status was assigned by calculating eGFR from serum creatinine tests (using the CKD-EPI equation and adjusting results to the current assay standardisation), and categorised according to eGFR thresholds used in current clinical CKD classification.[10]

5.4.1 Data extraction and cleaning of serum creatinine test results

Test results in CRPD are curated as "entity codes" (**3.2.3**). Entity code 165 records serum creatinine test results. This contains seven data fields: operator (e.g. >), value (e.g. 78), unit of measure (e.g. µmol/l), qualifier (e.g. [8] High), normal range from, normal range to (e.g. 120), normal range basis (e.g. [3] Age and sex based). An entity code record may also have an attached Read code.

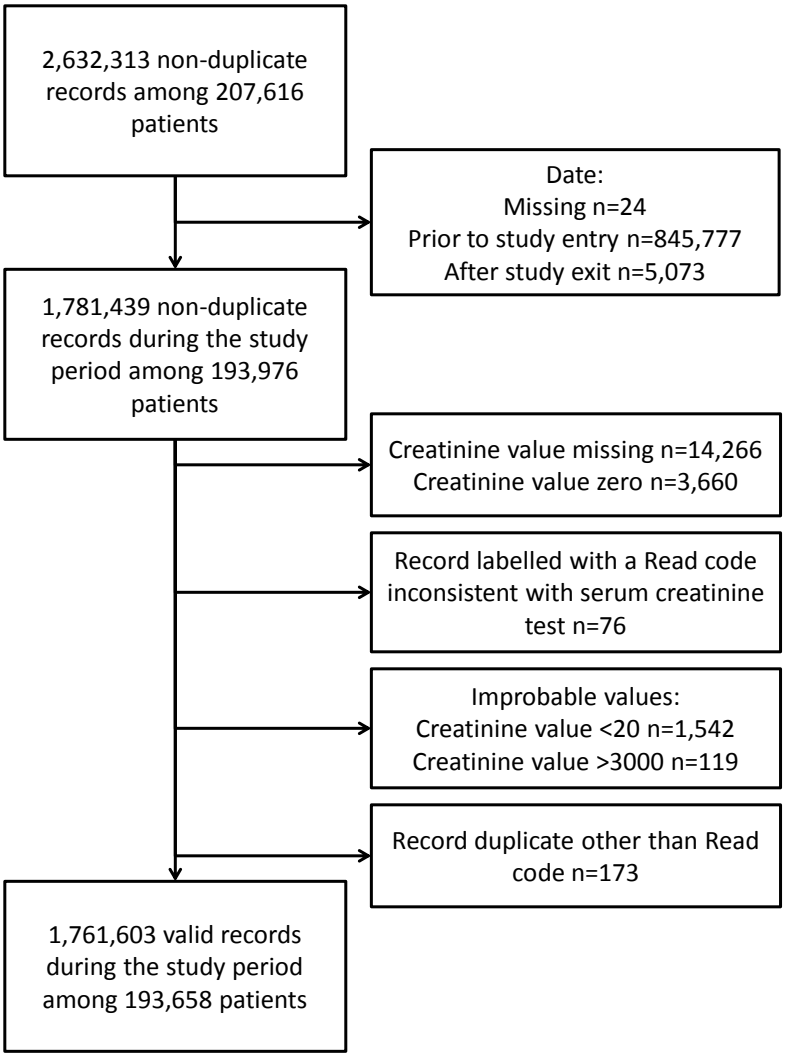
All records with entity code 165 were extracted from the test files. Entity codes may also be recorded in clinical files but these files do not include the data fields containing the test result value, and so were not extracted. Entity codes may be stored in additional clinical files with all relevant data fields, but there were none in this location for entity code 165.

Duplicate records were dropped, and the 2,632,313 non-duplicate records with entity code 165 among 207,616 patients were cleaned as **Figure 5.2**.

Of the 76 records excluded for an inconsistent Read code, 66 had Read code 44HG.00 *'Serum creatine kinase level'*. Other Read codes judged inconsistent were 4679 *'Urine dipstick for protein'*, 46W..00 *'Urine microalbumin'*, C04..13 *'Hypothyroidism'* and G581.00 *'Left ventricular failure'*.

Acceptable serum creatinine values (20–3000 µmol/l) were selected by consulting the improbable values excluded by the National Diabetes Audit[119] and discussion with Dr Dorothea Nitsch, nephrologist. Non-duplicate records on the same day were kept in this dataset at this point.

Figure 5.2: Data cleaning of serum creatinine tests (entity code 165)



Additional data cleaning could have been provided by checking whether the Read code was consistent with the serum creatinine value, for example, looking for patients with a high serum creatinine value and a Read code for low serum creatinine. The potential for this was limited, as 96.8% (1,705,552/1,761,603) of the included serum creatinine tests were labelled with Read code 44J3.00 *'Serum creatinine'*. In addition, classification of an absolute serum creatinine level as normal is consistent with impaired GFR, which could make interpretation of Read codes for normal renal function unreliable. This data cleaning option was therefore not used.

The potential for using the unit of measure to identify records in alternative units (such as mg/dl) or units raising suspicion of a GFR estimate recorded as a serum creatinine result was explored by tabulating creatinine values by the unit recorded **(Table 5.5)**. Again, this appeared to have limited potential for data cleaning: 93.6% of valid tests had the expected

unit for serum creatinine in the UK (umol/L), and median test result values did not differ markedly across units of measure.

Table 5.5: Serum creatinine test values by unit of measure among valid tests (n=1,762,411)

Number of valid tests		Unit of measure	Median value (interquartile range)
n	%		
1,649,647	93.6	umol/L	96 (81 – 119)
79,598	4.5	missing	97 (83 – 118)
23,768	1.3	mmol/L	98 (83 – 118)
8,564	0.5	mol/L	93 (76 – 118)
299	0.02	/uL	92 (74 – 124)
289	0.02	microU/L	112 (94 – 149)
61	0.003	Mmol	90 (75 – 113)
57	0.003	Umol	99 (88 – 115)
<20		mL/min, 1, mg/mmol, mmol/mmol (creat)	
<10		ng, nmol/L, h, mmol/mol, mg/dL, u/L, umol/min, %, mg/L, pmol/L, L, mg/mmol (creat), iu/L, m,ug.L, umol/g (creat), umol/mmol	
1		1(tot), 1/mL, mL, mmol/d, ng/mL, rad, u, ug, um, umol/mL/h, umol/mol	

Total exceeds 1,761,603 (the number of valid serum creatinine tests), as records which are duplicates for serum creatinine value and date but with different units of measure are included in this table.

5.4.2 The decision not to use serum creatinine tests prior to study entry

Several study entry criteria (such as diagnosis of diabetes mellitus) may be associated with subsequent increased serum creatinine screening and monitoring. Valid serum creatinine test results prior to study entry were also identified and cleaned identically to the process in **Figure 5.2**, other than selection by study entry date, to see if serum creatinine results were less complete in the year prior to study entry. Completeness of serum creatinine recording was compared between the year prior to study entry and the first year in the study, stratified according to whether study entry was pre- or post-QOF (1 April 2004), as this was expected to alter screening for CKD. Completeness was 12.6% higher after study entry among patients who entered the study pre-QOF **(Table 5.6)**. Including historical serum creatinine tests prior to study entry therefore carried a risk of introducing ascertainment bias (as discussed in **3.9.2**), hence it was decided to use serum creatinine results only after study entry.

Table 5.6: Are serum creatinine results less complete in the year prior to study entry?

Year of creatinine record	Number of creatinine results	Number of patients with a creatinine result (%)
Study entry before 1 April 2004 (111,591 patients)		
Year prior to study entry	69,873	44,623 (40.0%)
First eligible year	105,575	58,701 (52.6%)
Study entry on or after 1 April 2004 (107,097 patients)		
Year prior to study entry	185,868	92,769 (86.6%)
First eligible year	195,919	90,716 (84.7%)

5.4.3 Estimating GFR from serum creatinine tests

Serum creatinine values were multiplied by 0.95 for assay adjustment to IDMS-standardised serum creatinine,[104] and divided by 88.4 to convert units from µmol/l to mg/dl. GFR was estimated using the CKD-EPI equation:[102]

$$\mathbf{GFR = 141 \times \min(Scr/\kappa, 1)^{\alpha} \times \max(Scr/\kappa, 1)^{-1.209} \times 0.993^{Age} \times 1.018\ [if\ female]\ \_\ 1.159\ [if\ Black\ ethnicity]}$$

where: Scr is serum creatinine (mg/dL)

κ is 0.7 for females and 0.9 for males,

α is -0.329 for females and -0.411 for males,

min indicates the minimum of (Scr/κ) or 1, and

max indicates the maximum of (Scr/κ) or 1.

For example, eGFR for a 65-year old woman with black ethnicity and serum creatinine 0.8mg/dL:

$$GFR = 141 \times 1 \times (0.8/0.7)^{-1.209} \times (0.993)^{65} \times 1.018 \times 1.159 = 90\ ml/min/1.73\ m^2$$

Methods for identifying patients with Black ethnicity are described in **6.1.2**.

Individual serum creatinine records were classified corresponding to clinical categories used during the study period (eGFR <15, 15-29, 30-44, 45-59, or ≥60 ml/min/1.73m^2).[10] To allow analysis to compare with epidemiological studies, a longer classification was also used, which classified eGFR results over 60 as 60-74, 75-89, 90-104, and ≥105 ml/min/1.73m^2).[92] There were 1,501 pairs of records which produced different estimates of eGFR on the same day according to the longer classification. As omitting these results was likely to have a minimal effect on data completeness, and an intuitive way to identify the estimate most likely to be correct was to consult the previous result for that patient, it was decided to omit records on the same day where the estimated GFR categories clashed.

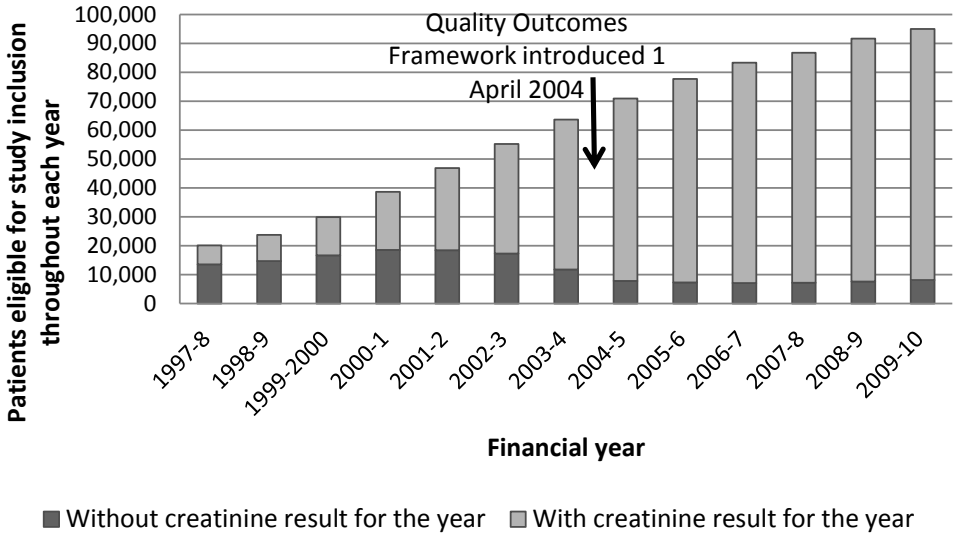
5.4.4 Completeness of serum creatinine test results

Serum creatinine tests provided good completeness with regular updating during the study period. Among the 218,688 patients in the study population, 88.5% (193,646) had 1,760,094 valid serum creatinine tests during the study period. **Figure 5.3** shows the percentage of patients who had a valid serum creatinine test available in each financial year for which they were eligible for inclusion throughout the full year. Completeness of

test result availability improved steadily from 32.8% in 1997-8 to 89.0% in 2004-5, and remained steady at over 90% for each subsequent year.

Serum creatinine was frequently monitored. Among the 193,646 patients with at least one result, the median number of serum creatinine results per patient during the study period was 15 (range 1-272, IQR 8-21). For 89% of records, the next test result was within a year. The time from each GFR estimate to the next test result or study exit was a median of 137 days (IQR 56-242).

Figure 5.3: Completeness of serum creatinine recording among patients eligible for each full financial year (among the 193,646 patients with at least one creatinine result during follow up)



5.4.5 Identifying chronicity of impaired eGFR

The clinical classification of CKD requires that impaired renal function persist for at least 3 months for CKD diagnosis. In epidemiological studies of the effect of CKD on mortality, the definition of CKD is often based on a single measurement of serum creatinine at baseline.[24] However, serum creatinine fluctuates, and identifying CKD based on a single estimate of GFR will tend to over-estimate CKD prevalence.[118] A single impaired GFR may also be caused by acute kidney injury, AKI. AKI describes a rapid reduction in renal function, identified by a sudden decline in eGFR.[120] Both impaired eGFR and albuminuria, as well as being markers of CKD, are strongly and independently associated with an increased risk of AKI.[121] The commonest trigger of AKI among hospital inpatients is sepsis.[120] Misclassifying AKI as CKD could therefore result in over-estimation of the association between CKD and infection.

Two tests at least three months apart could be used at baseline to identify CKD, and this would be more robust to misclassification from creatinine fluctuation or AKI. However, creatinine monitoring may be more frequent among patients at higher risk of infection. In particular, acute or severe infections will prompt monitoring of eGFR, and if patients with more frequently monitored creatinine are more likely to be identified as having CKD (for example, as they are more likely to be eligible for a definition requiring two test results rather than one) this could also introduce ascertainment bias which would result in over-estimation of the association of CKD with infection incidence.

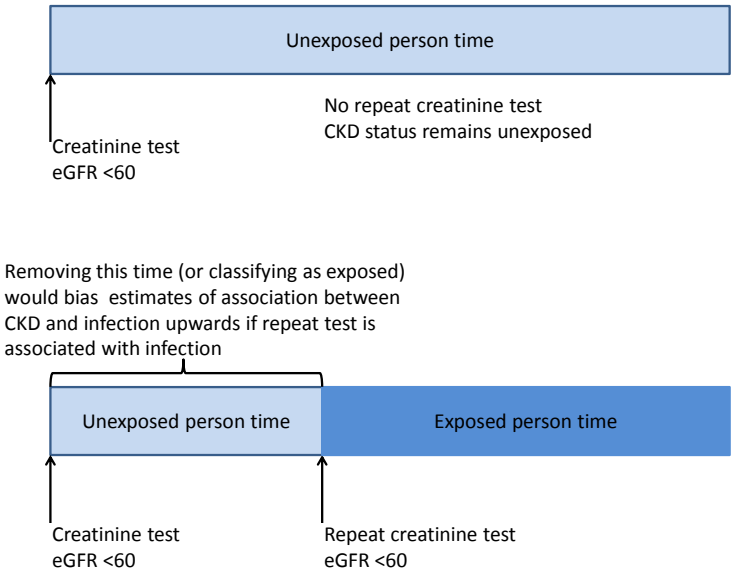
Thus, if using serum creatinine test results to assign CKD status based on eGFR, a balance must be found between minimising misclassification of a single reduced eGFR as CKD by requiring two results, and minimising ascertainment bias that would result from differential testing by using a single eGFR result. How best to do this may depend on the length of follow-up time. Over a short period of follow-up, a *'best-of-two'* method could be used. CKD status could be ascertained at baseline using the highest eGFR result of the latest two serum creatinine tests recorded at baseline that were at least three months apart. This would approximate the clinical requirement that eGFR be impaired at a certain level for at least three months before that stage of CKD be diagnosed. In the absence of two suitable results, the single most recent creatinine result could be accepted, to reduce ascertainment bias.

Measuring CKD at baseline without updating status over time might not be suitable for a study with a long follow-up time, over which patients' CKD may progress. If patients with more rapid progression of CKD also experience higher rates of infection, the association of CKD with infection could be severely under-estimated by identifying CKD only at baseline. Risk factors for progression of CKD are not fully understood but poor glycaemic control may be a common risk factor for CKD progression and increased risk of infection incidence.[122]

If two results are required to define time-updated CKD status in epidemiological studies, the time between the first and second result must be handled with care to avoid immortal time bias.[123] If this time is considered 'exposed' to CKD (or removed from analysis), infections following the first test are attributed to the exposed group (or excluded) for patients who have a second test but to the unexposed group for patients who have a single test. If infections are associated with increased creatinine testing this would bias estimates the association between CKD and infection upwards. The approach of choice is to classify this time as unexposed (**Figure 5.4**). Events attributed to the unexposed group may reflect

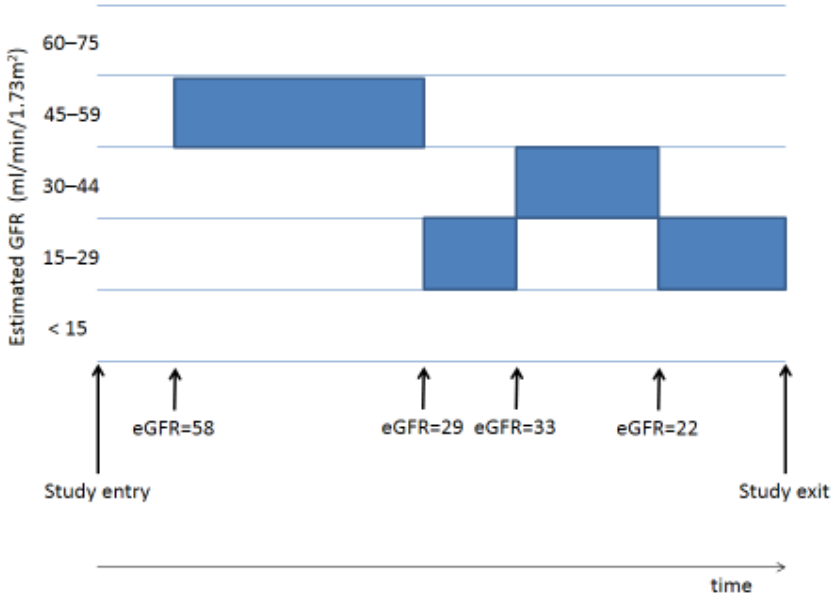
the effect of early CKD, overestimating the infection rate among the unexposed: this results in a conservative estimate of the effect of CKD on infection incidence.

Figure 5.4: The risk of immortal time bias if two creatinine test results are required to define CKD status



The 'last-carried-forward' method used by James *et al.* may be more suitable for studies with a sufficiently long follow-up time for CKD progression to be an issue.[124, 125] In this approach, CKD stage is defined at any given time using the GFR estimate produced by the single most recent creatinine result (**Figure 5.5**). This method allows updating of the patient's status as CKD progresses. Although creatinine fluctuation and AKI will result in over-estimation of CKD stage, the patient's status will be updated at the next test result, minimising misclassified person-time. If AKI is evident among serum creatinine results being monitored in primary care, frequent testing would be expected until GFR has stabilised, and so misclassified person-time from AKI should be minimal.

Figure 5.5: The last-carried-forward method for establishing GFR status



eGFR, estimated glomerular filtration rate

Another option could be to use Read codes to try to identify CKD rather than temporarily reduced GFR. However, encoding of CKD may not be sufficiently complete to permit this, and the negative predictive value of Read codes for CKD is unclear. In a validation of Read codes in CPRD, 10% of potential cases of acute renal failure were found to have CKD on manual review of the patient records, despite all patients with identified CKD Read codes having been removed from the sample.[126]

5.5 Identification of proteinuria status in CPRD

Proteinuria tests are less consistently and completely recorded than serum creatinine tests, and are particularly likely to be missing from the database if negative. Test records and Read codes for proteinuria were combined to identify a proteinuria variable. This was a binary non-reversible exposure recording a history of proteinuria. The first valid positive proteinuria record changed the patient's status from negative to positive, which it then remained until study exit. The absence of a positive proteinuria record was assumed to indicate negative status. As this approach allows a single result to define ongoing proteinuria status, caution was indicated in interpreting test results as positive. A 'trace' of proteinuria was not counted as positive, and the results were extensively cleaned to ensure internal consistency. This approach involved tolerating likely under-ascertainment of proteinuria, but should result in conservative estimates for the association between proteinuria and infection.

The possibilities of quantifying proteinuria, and of identifying albuminuria separately from total proteinuria, to allow classification consistent with current clinical practice, were also explored.

5.5.1 Proteinuria entity code data extraction

Three entity codes recorded proteinuria test results (**Table 5.7**). Albumin: creatinine ratio test results were not available as a separate entity code in the May 2011 CPRD database I used to select my study population. All test records with entity code 287, 431 or 435 were extracted from the test files. None were recorded in the additional clinical files. Entity codes were cleaned as described in **Figure 5.6**.

Table 5.7: Entity codes available in CPRD for identifying proteinuria status

Entity code		Data format
287	Urinalysis-Protein test	4 data fields: qualifier (e.g. [8] High), normal range from (e.g. 3), normal range to (e.g. 10), normal range basis (e.g. [3] Age and sex based). May also have Read code label.
431	Urine dipstick for protein test	
435	Urine microalbumin test	

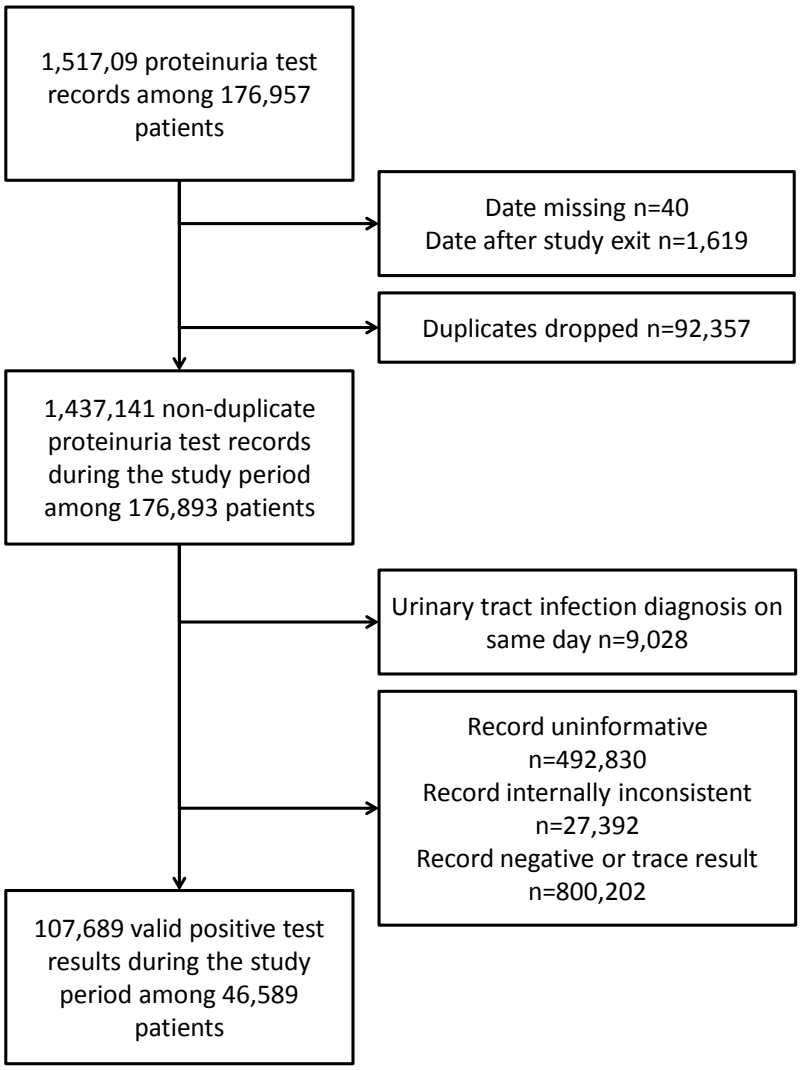
Removal of records coincident with urinary tract infection

Persistent albuminuria is a sign of renal damage used to classify CKD: transient proteinuria has a variety of causes including fever and urinary tract infection (UTI).[127] Misclassifying transient proteinuria resulting from fever as persistent proteinuria indicating CKD could result in ascertainment bias which could overestimate the association between CKD and

infection. Ideally, the definition of proteinuria for identifying CKD would avoid this, for example by using only Read codes which clearly indicated persistent proteinuria, or by requiring repeated positive proteinuria tests. This was likely to be impractical without severely reducing ascertainment of persistent proteinuria. As a compromise, proteinuria records coincident with a urinary tract infection (UTI) were removed.

Records with a Read code for UTI on the same day were identified using Read codes listed in **Appendix C**. This list was an expanded version of the list of Read codes used to identify diagnoses of UTI for the purposes of defining episodes of infection (**4.3.1**). Infectious urethritis was included, as urethritis may cause transient proteinuria. Of the 9,028 proteinuria records on the same day as a UTI, 3,264 (36.2%) recorded a positive proteinuria result.

Figure 5.6: Data cleaning of proteinuria tests (entity codes 287, 431 and 435)



Removal of uninformative records

A test result could not be obtained from 492,830 records where both the test value (recorded in the qualifier field) and the Read code were missing or uninformative (e.g. Read term 467..00 *'Urine protein test'* with test value [0] missing).

Checking for internal consistency of records

Internal consistency was assessed by comparing the test value with the attached Read code as shown in **Table 5.8**. Records were judged internally inconsistent if: (1) the test value was [20] *'Not examined'*; (2) the Read code was inconsistent with the label for a proteinuria result (of the 5,874 records judged to have an inconsistent Read code, 98.9% had Read code 4691 *'Urine protein test not done'*); or (3) one of the test value and Read code recorded a positive result while the other recorded a negative result. The distribution of these internally inconsistent tests is described in **Table 5.9**. Trace results were considered consistent with either positive or negative interpretation, depending on clinical context, and so records containing a trace result and either a positive or negative value were not considered internally inconsistent. The internally inconsistent test records which contained both a positive and a negative result formed 0.2% (2537/1,437,141) of the non-duplicate proteinuria test records, and probably represent data entry error.

Table 5.8: Comparison of test value and Read code in proteinuria test records

		Test value (qualifier field)					
		Missing	Uninformative	Inconsistent	Positive	Trace	Negative
Read code	Missing				Use		
	Uninformative				Use		
	Inconsistent						
	Positive	Use	Use		Use		
	Trace						
	Negative						

	Uninformative record	n=492,830
	Internally inconsistent record	n=27,392
	Negative or trace result	n=800,202
Use	Positive result	n=107,689

Table 5.9: The internal consistency of proteinuria test records

		Test value						
		Missing	Uninformative	Inconsistent	Positive	Trace	Negative	Total
Read code	Missing	1	486,421	4,472	16,951	4,190	99,834	611,869
	Uninformative	0	6,408	0	14	0	0	6,422
	Inconsistent	0	18,932	1,052	19	1	29	20,033
	Positive	0	85,553	4	5,171	86	137	90,951
	Trace	0	42,303	3	862	2,486	342	45,996
	Negative	0	604,621	343	2,400	66	45,412	652,842
Total		1	1,244,238	5,874	25,417	6,829	145,754	1,428,113

5.5.2 Identifying a Read code list for proteinuria

I extracted 2,174 Read codes identifying CKD status using the strategy in **Table 5.10**, which starts by identifying text-based search terms, and uses these to identify relevant chapters for a hierarchical search. I then identified the subset describing proteinuria or proteinuric disease manually.

Table 5.10: Search strategy for Read codes identifying chronic kidney disease

Search details	CPRD Medical browser version 1.3.2 Database build 'ever' Search date 17 January 2012
Stage 1	Text-based search
Read terms	*rena* OR *nephr* OR *kidn* OR *glomerul* OR *creatinin* OR *GF*
Stage 2	Identifying relevant Read code hierarchies
	Sorted results of text-based search by Read code to identify relevant Read code headings.
Stage 3	Hierarchical search
Read codes	ZV* TB11* TB00* TA22*TA02*SP15*SP14*SP08*SK0* SB24*S76*R14* R13* R08*Q48* Q20* Pyu* PD*P76* L393*L162*L121* Kyu*K1*K0*G76*G72*G71*G70*G23*G22*D31*D21* Cyu*C35*C31* C10*B9* B8*B7*B5*B4*9O*9N*9h* 9b*8L*8H*8A*7P1* 7P0*7N5* 7L* 7B1* 7B0* 7A6*6A* 68*66* 585*57*557*53B* 4Q*4I*46*45*262*1Z,, 1J*, 1A5*, 14V*, 14*

As the aim was to identify proteinuria separately from eGFR status, codes which did not unambiguously record proteinuria were not included. Although diabetic nephropathy causes albuminuria, a record of diabetic nephropathy in primary care for an older patient with diabetes mellitus might also be used less formally to identify CKD in the presence of diabetes. Codes for diabetic nephropathy which did not specify proteinuria were therefore not included. For example, neither C109012 *'Type 2 diabetes mellitus with renal complications'* nor C10FC00 *'Type 2 diabetes mellitus with nephropathy'* were included, whereas C10FL00 *'Type 2 diabetes mellitus with persistent proteinuria'* was included. Codes which identified proteinuria in pregnancy (such as pre-eclampsia) were not included, as they would not imply ongoing persistent proteinuria among a population aged ≥65 years.

This codelist was also compared with a Read code list for proteinuria compiled separately by Dr Catriona Shaw (Clinical Research Fellow, UK Renal Registry) and Dr Anoop Shah (Clinical Research Association, University College London) with Dr Dorothea Nitsch. The comparison identified 13 additional codes which were added to this codelist.

The final proteinuria Read code list comprised four subgroups (**Table 5.11**) and is listed in full in **Appendix D**. Codes which could represent a single positive test result or transient

proteinuria were identified within this codelist. Several of these codes (such as R110.00 *'[D]Proteinuria'* and R110z00 *'[D]Proteinuria'*) were codes incentivised for use to record persistent proteinuria among patients with diabetes by QOF.[65] As these codes were ambiguous, a cautious approach was taken and these were treated as potentially recording transient proteinuria, although they were in practice likely to indicate persistent proteinuria (these codes are flagged in **Appendix D**).

Table 5.11: Read codes identifying proteinuria

Subgroup	Sample Read codes	Number of codes	May be transient
Proteinuria test result or status	R110.00 [D]Proteinuria	8	Yes
	4675 Urine protein test = ++		
Albuminuria test result or status	R110300 [D]Microalbuminuria	3	Yes
	46W0.00 Urine microalbumin positive		
Proteinuria status that does not appear to reflect a single test result	C10FL00 Type 2 diabetes mellitus with persistent proteinuria	33	No
	1Z1B.00 Chronic kidney disease stage 3 with proteinuria		
	K190X00 Persistent proteinuria, unspecified		
Proteinuric diseases	K011.00 Nephrotic syndrome with membranous glomerulonephritis	68	No
	K020.00 Chronic proliferative glomerulonephritis		

5.5.3 Data extraction and cleaning of Read codes for proteinuria

All Read codes for proteinuria were extracted from test, clinical and referral files. There were 84,627 Read codes for proteinuria extracted for 41,281 patients. The 26,349 records which had already been cleaned as they were proteinuria test records were dropped. The 61 records with a missing date and 215 with a date after study exit were dropped. Records were de-duplicated, and 56,828 records for 31,845 patients remained.

A small number of Read codes (6.6%, 3,762/56,828) were attached to an entity code record which was not an entity code for a proteinuria analysis. The entity codes to which Read codes were attached were all considered consistent with a proteinuria result being indicated by the Read code (**Table 5.12**). These entity code test values were not used to identify proteinuria status, but the Read code for proteinuria remained in the analysis.

Table 5.12: Entity codes for tests other than proteinuria analysis to which proteinuria Read codes were attached

Entity code	N	%
286 *'Urinalysis – Glucose'*	1,739	46.2
240 *'Urine test'*	1,356	36.0
288 *'Other laboratory tests'*	576	15.3
340 *' Urine biochemistry'*	59	1.6
430 *'Urine dipstick for glucose'*	32	0.9
Total	3,762	

There were 48,443 records with Read codes which could identify transient proteinuria. They were kept in the analysis unless there was a UTI recorded on the same day, using the same Read codes for UTI as were used to clean the proteinuria test results in **5.5.1**. Once the 429 proteinuria records which were recorded on the same day as a UTI diagnosis were removed, this left 56,399 valid, non-duplicate records among 31,684 patients.

5.5.4 Records defining onset of a history of proteinuria

Using the combined proteinuria test results and Read codes, there were 165,247 records of proteinuria before study exit among 62,367 (28.5%) of the 218,688 eligible patients. For 48% of these patients, this status relied upon a single record of proteinuria before or during the study (29,822/62,367). **Table 5.13** shows the record types defining onset of a history of proteinuria. For 90% of patients with a history of proteinuria identified before study exit, this was defined by a proteinuria or albuminuria test result or status which could have been transient (56,324/62,367).

Table 5.13: Description of records defining onset of proteinuria (n=62,367)

	Read term or test type	n (%)	Total n (%)
Proteinuria test result or status	'Urinalysis-Protein test' (entity code 287)*	22,773 (36.5)	34,113 (54.7)
	[D] Proteinuria	3,326 (5.3)	
	'Urine dipstick for protein test' (entity code 431)*	3,157 (5.1)	
	Proteinuria	2,480 (4.0)	
	Urine protein test = +	1,613 (2.6)	
	Urine protein test = ++	491 (0.8)	
	Urine protein test = +++	194 (0.3)	
	Urine protein test = ++++	48 (0.1)	
	[D]Proteinuria NOS	31 (0.1)	
Albuminuria test result or status	'Urine microalbumin test' (entity code 435)*	13,191 (21.2))	22,211 (35.6)
	[D] Microalbuminuria	6,823 (10.9)	
	[D] Albuminuria	1,858 (3.0)	
	Urine microalbumin positive	339 (0.5)	
Proteinuria status that does not appear to reflect a single test result	Type 2 diabetes mellitus with persistent proteinuria	1,355 (2.2)	2,277 (3.7)
	Type 2 diabetes mellitus with persistent proteinuria	452 (0.7)	
	Chronic kidney disease stage 3 with proteinuria	138 (0.2)	
	Persistent proteinuria, unspecified	77 (0.1)	
	Chronic kidney disease stage 3A with proteinuria	60 (0.1)	
	Chronic kidney disease stage 3B with proteinuria	45 (0.1)	
	Type 1 diabetes mellitus with persistent proteinuria	33 (0.1)	
	Chronic kidney disease stage 2 with proteinuria	30 (0.1)	
	Chronic kidney disease stage 4 with proteinuria	27 (<0.1)	
	Type 1 diabetes mellitus with persistent proteinuria	24 (<0.1)	
	Benign postural proteinuria	11 (<0.1)	
	Other records with <10 cases each	25 (<0.1)	
Proteinuric diseases	Nephrotic syndrome	131 (0.2)	473 (0.8)
	Acute glomerulonephritis	60 (0.1)	
	Unspecified glomerulonephritis NOS	50 (0.1)	
	Nephrotic syndrome in diabetes mellitus	45 (0.1)	
	Acute interstitial nephritis	39 (0.1)	
	Berger's IgA or IgG nephropathy	20 (<0.1)	
	Glomerulosclerosis	16 (<0.1)	
	Other records with <10 cases each	112 (0.2)	
Multiple records			3,293 (5.3)
Total			62,367

* with a positive result indicated within the entity type record

5.5.5 Exploration of quantification of proteinuria, and identification of albuminuria in test results

Positive proteinuria test results were explored to see whether albuminuria could be distinguished from total proteinuria, and whether albuminuria could be quantified. Read codes specifying that the analysis was albuminuria were almost exclusively limited to records with entity code 435 *'Urine microalbumin'*. These Read codes therefore added little

information, although they do provide evidence of internal consistency for these records. Records with entity code 435 formed 38.0% of the proteinuria test records, and all but two had a Read code specifying albuminuria measurement, and so it is reasonable to assume that at least 38% of proteinuria tests recorded albuminuria rather than total proteinuria **(Table 5.14)**.

It was not possible to quantify albuminuria in these data. The only quantification available for any of these entity types was in terms of +/++/+++/++++, a system which is used for total protein urine reagent strips. This result was also occasionally recorded in Read codes, e.g. 4674 *'Urine protein test = +'*. This information was only available for 63.7% of positive proteinuria tests, and so precise quantification was not attempted.

Table 5.14: Description of positive proteinuria test results

	Result quantified n (row %)	**Conflict in quantification*** n (row %)	**Total**
Entity code 287 *'Urinalysis-Protein test'*			
Microalbuminuria positive	0	0	3
Urinary albumin measurement	2 (16.7)	0	12
Albumin not specified	53,764 (94.3)	216 (0.4)	57,005
Entity code 431 *'Urine dipstick for protein'*			
Urinary albumin measurement	4 (100)	0	4
Albumin not specified	9,647 (99.1)	0	9,736
Entity code 435 *'Urine microalbumin'*			
Albumin:creatinine ratio measurement	40 (97.6)	0	41
Microalbuminuria positive	726 (8.9)	0	8,156
Microalbuminuria measurement	4,397 (13.4)	0	32,730
Albumin not specified	2 (100)	0	2
Total	68,582 (63.7)	216 (0.2)	107,689

* conflict between proteinuria quantity recorded in test value and quantity specified in Read code

5.6 Summary of definition of CKD in CPRD

Two markers of CKD, estimated glomerular filtration rate (eGFR) and proteinuria, were identified separately, to allow analyses to treat these as potential independent markers of infection risk.

5.6.1 Defining eGFR in CPRD

Serum creatinine tests offered a relatively complete and frequently updated source of GFR estimates for a high proportion of the study population. They offered good granularity of GFR categorisation, both distinguishing stage 3a and 3b, and permitting categorisation of GFR at levels above 60 ml/min/1.73m^2. As the most consistently recorded data available for assessing GFR status, they were the data source least vulnerable to ascertainment bias arising from changes in the clinical context of CKD recording over the study period. The completeness of recording among this study population minimised the risk of under-ascertainment and ascertainment bias from CKD identification according to patient characteristics (such as smoking status), changes in clinical practice and data recording patterns.

For these reasons, eGFR was calculated from serum creatinine test results using the CKD-EPI equation, including adjustment for ethnicity, and classified according to NICE 2008 guidelines.[11, 102]

The choice of the last-carried-forward method or the best-of-two method for each cohort study was based upon the length of individual patient follow-up, to allow the tension between misclassification of CKD with the risk of ascertainment bias to be balanced according to whether CKD progression was likely during the follow-up period.

5.6.2 Defining a history of proteinuria in CPRD

Proteinuria was less consistently recorded. The quality of data recording did not permit adequate ascertainment of a negative status: thus a pragmatic approach was taken, in which it was assumed that absence of a positive test result implied a negative proteinuria status. As a single positive result defined a positive status for the rest of the study, caution was taken in accepting potential positive records. A conservative approach to identifying positive results was taken which comprised excluding proteinuria records on the same day as a UTI, not counting trace results as positive, extensive data cleaning to check internal consistency of records, and not including diagnostic codes for diabetic nephropathy in the codelist for proteinuria identification.

A binary variable of ‘history of proteinuria’ was defined in which absence of a positive record was assumed to indicate no history of proteinuria. Data quality did not permit confident identification of albuminuria, or quantification of proteinuria.

5.7 Potential alternative approaches to identifying CKD

CPRD contains several other sources of data which could inform CKD status.

5.7.1 Unspecified tests in CPRD

Entity codes may be recorded together with a Read code: thus, a serum creatinine result could be inferred from the unspecified test of entity code 288 *'Other laboratory tests'* if this were labelled with a Read code identifying a serum creatinine test or test result, such as 44J3300 *'Serum creatinine raised'*. Non-specific entity code test results could be interpreted as proteinuria tests if labelled with a relevant Read code, such as 46TC. *'Protein:Creatinine Ratio'*. This inference depends on a Read code having been used accurately and for the purpose of identifying the accompanying test result. Potential for data entry error may be high among these records, as entry of a test result under 'Other laboratory test' suggests an abnormal data entry context. Processing a GFR estimate as a crude serum creatinine level, for example, would result in considerable misclassification of CKD status. For these reasons, entity code 288 *'Other laboratory tests'* was not used to identify GFR status nor proteinuria.

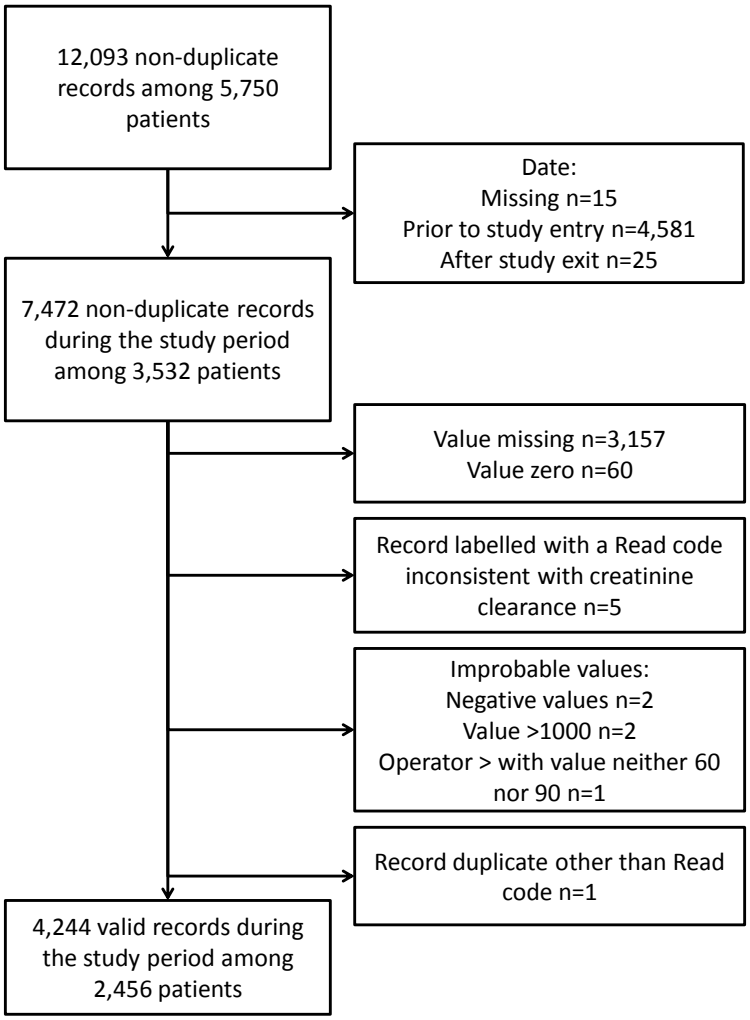
5.7.2 Creatinine clearance test results in CPRD

Entity code 166 *'Creatinine clearance'* records creatinine clearance test results, with the same seven data fields as entity code 165 '*Serum creatinine'* (**5.4.1**). All records with entity code 166 were extracted from the test files: there were none in the additional clinical files.

Duplicate records were dropped, and the 12,093 non-duplicate records with entity code 166 among 5,750 patients were cleaned as **Figure 5.7**. Read codes judged inconsistent with creatinine clearance were G20..00 *'Essential hypertension'* (n=1), 46TC.00 *'Urine albumin:creatinine ratio'* (n=2), and 44J7.00 *'Albumin/creatinine ratio'* (n=2). Non-duplicate records on the same day were kept in this dataset at this point.

Creatinine clearance tests added little to completeness of data. There were only 4,244 valid creatinine clearance tests among 2,456 patients during the study period. For 895 creatinine clearance tests, there was a serum creatinine test recorded on the same day. They also restricted categorisation of eGFR: 375 records recorded only that creatinine clearance was >60 ml/min. They were not found to be a useful addition to serum creatinine test results, and were therefore not used.

Figure 5.7: Data cleaning of creatinine clearance tests (entity codes 166)



5.7.3 Read codes recording GFR status

From the search strategy described in **Table 5.10,** I identified Read codes which could be used to identify staged CKD. These comprised: CKD stages 1–2; CKD stages 3–5; and end-stage renal failure (assumed to indicate CKD stage 5). These codes are listed in **Appendix D** and summarised in **Table 5.15**.

Table 5.15: Summary of Read codes identifying staged CKD status

Subgroup	Number of codes	Codes indicating proteinuria	Commonest example in CPRD
CKD stage 1-2	9	4	1Z11.00 Chronic kidney disease stage 2
CKD stage 3-5	25	10	1Z12.00 Chronic kidney disease stage 3
End-stage renal failure	3	0	K050.00 End stage renal failure
Total	37	14	

These Read codes were extracted from clinical and referral files (there were none present in test files). Six records with missing dates were dropped, records were de-duplicated and restricted to those between study entry and study exit.

Staged Read codes added only slightly to completeness: there were 75,731 staged CKD Read codes among 48,012 patients during the study period, of which 18,680 (24.7%) were on the same day as a serum creatinine result. They permitted a less fine categorisation of GFR than serum creatinine results. Among 58,795 read codes for CKD stage 3 during the study period, 55,990 (95.2%) did not specify between stage 3a or 3b, which has been found to be an important distinction in terms of prognosis among older people.[107, 108] They were not found a useful addition to serum creatinine tests among this population, and were not used.

This approach is supported by a recently published study of CKD prevalence in the total population of CPRD, which identified CKD from either two eGFR measurements <90 ml/min/1.73m^2, or a Read code for CKD, or both. Only 0.5% of the patients with CKD were identified from a Read code alone. [18]

5.7.4 Renal disease aetiology

The cause of renal disease may not be fully investigated among this population, and classification of CKD by cause of disease involves clinical interpretation of an individual patient's health, rather than application of a systematic classification of aetiologies (**5.1.5**). As infection itself and the presence of risk factors for infection may influence clinical investigation of the cause of CKD, classification of CKD according to aetiology risks introducing ascertainment bias to the association between CKD and infection. However, validity of diagnostic codes may differ among the older people with diabetes from the general population, and the potential use of diagnostic codes to supplement CKD classification by eGFR and proteinuria was explored in this population.[114]

From the codes extracted using the search strategy in **Table 5.10**, I identified Read codes for renal pathology likely to cause CKD. These were disparate and difficult to classify according to CKD status. In particular, many aetiologies could have varying relationships with eGFR, proteinuria and prognosis, according to presentation and severity. Within each subgroup, the most commonly occurring codes were usually those that would be 'high-level' codes in a hierarchical coding system. These codes provide the least detailed information and are thus least useful when seeking granularity of data. These Read codes could be used as exclusion characteristics, to obtain a patient population unlikely to have

CKD, but are less useful for classifying CKD status for an individual patient, as many codes would have poor specificity for CKD.

5.7.5 Could we have used secondary care records to identify CKD?

The majority of CKD is identified, managed and monitored in primary care.[100] Patients may be referred to secondary care for specialist investigation and management of underlying renal disease, management of complications such as uncontrolled hypertension, or initiation of renal replacement therapy. Patients referred to nephrologists from primary care, for example for management of advanced CKD, or investigation of underlying renal disease, are mostly treated in outpatient clinics: only about 5% of patients under the care of a nephrologist at any one time are inpatients.[43] Outpatient records which provided details of renal disease among these patients could be a useful supplement to primary care records, but were not available to this study.

Linked hospital admission records were available, and co-morbidities are recorded in these. However, there is a high risk of ascertainment bias if CKD status is established from inpatient admission records. Infection is a major cause of hospital admission among older people.[9] Infection is the leading trigger of AKI among hospital inpatients, and admission with infection would therefore prompt monitoring of renal function.[120, 128] Pre-existing CKD is also a recognised risk factor for development of AKI during infection.[121] CKD may therefore be more likely to be recorded for patients admitted with infection, and so using hospital acute admission records to supplement identification of CKD status would risk increasing CKD ascertainment differentially among patients with a history of hospital admission for infection. This could introduce ascertainment bias, with over-estimation of the association between CKD and infection. For this reason, secondary care records were not used as a data source for CKD status in this thesis.

Even if this were not the case, HES admission records would still not be ideal for ascertaining CKD status. The HES inpatient database does not include secondary care test results. CKD diagnoses may be under-recorded in hospital admission records: the sensitivity of a diagnostic code of kidney disease in a Canadian administrative database of hospital admissions was 38% compared to eGFR from serum creatinine results.[129] Co-morbidities in hospital admission records are encoded using the 10th revision of the International Classification of Diseases (ICD-10).[75] This coding system is not ideal for establishing CKD status, as kidney disease is classified according to aetiology rather than function, which would not permit ascertainment of staged CKD status.[79]

Primary care of patients with CKD continues after referral to secondary care, and the GP retains responsibility for co-ordination of the patient's overall care. It is therefore to be hoped that CKD identified in secondary care would be recorded in primary care records.

Chapter 6. Describing the cohort

The cohort was described in terms of demographics, health behaviours, co-morbidities, and characteristics of diabetes. Not all characteristics were relevant to every analysis: the rationale for how covariates were included in each analysis is presented in **Chapters 7-11**.

6.1 Demographics

6.1.1 Age, gender and region

Year of birth, gender and region of residence were obtained from patient files in CPRD. These records were complete for all patients. To calculate age from year of birth, patients were assigned a nominal birthday of 1 July. Unless otherwise specified, age was categorised in five-year bands up to age ≥85 years.

6.1.2 Ethnicity

Ethnicity is recorded in CPRD and HES. A study by Mathur *et al.* of usability of CPRD and HES ethnicity records found that the consistency of ethnic category was high: within each dataset, for patients with multiple ethnicity records within either dataset; and between datasets, for patients with an ethnicity recorded in both datasets. In CPRD, ethnicity recording increased considerably after recording was financially incentivised by the Quality and Outcomes Framework (QOF) in 2006. Patients in CPRD with a recorded ethnicity in 2011 had a very similar ethnic breakdown to that recorded for the UK population in the 2011 census.[130]

The single ethnicity provided in the HES patient file was used as a descriptive variable for objective 2. This variable is based on all ethnicity recordings in HES inpatient records for each patient with HES linkage. Where multiple ethnicities were recorded for one patient, the most frequently recorded code was provided. Where there was a tie, the more specific ethnicity code was provided.

A more complete identification of Black ethnicity was required for calculation of estimated glomerular filtration rate (eGFR) to identify CKD status for objectives 3–5. For this purpose, CPRD and HES records were combined. A Read code list developed by Rohini Mathur was used to extract ethnicity records from clinical and referral files in CPRD (there were none in other files). Patients with a single usable ethnicity record in CPRD were assigned this ethnicity. Patients with multiple ethnicity records in CPRD were assigned the most frequently recorded code for the patient: in the event of a tie, the most recently recorded

of the frequently recorded codes for the patient was used. For patients with HES linkage and no usable record of ethnicity in CPRD, the ethnicity variable in the HES patient file was assigned. Where Mixed Black ethnicity was recorded, the patient was assigned Black ethnicity for the purposes of adjusting for Black ethnicity in calculation of eGFR.

6.1.3 Socio-economic status

Socio-economic status was described using the index of multiple deprivation.[83] For all patients, practice-level socio-economic status was available. This was the index of multiple deprivation quintile for the postcode of the patient's primary care practice. For patients with data linkage, patient-level socio-economic status was available. This was the index of multiple deprivation quintile for the postcode of the patient's home address. The type of socio-economic status used was specified in each analysis.

6.2 Nursing or residential home

A history of residence in a nursing or residential home was identified using a combination of CPRD records, including a Read code list compiled by Elizabeth Millett. Many of the commonly used Read codes were not sufficiently granular to distinguish between nursing or residential homes (e.g. 13F6.00 *'Nursing/other home'* and 13F7.00 *'Residential institution'*). Warden-controlled and sheltered home residency were not included. A patient was classified as having a history of nursing or residential home residence if there was any eligible positive record without evidence of internal inconsistency (**Table 6.1)**.

Table 6.1: Records defining nursing or residential home status

Record type	Positive record	Evidence of internal inconsistency (records with these characteristics were discarded)
Consultation type field in the consultation file.	Values 30, *'nursing home visit'*, or 31, *'residential home visit'*.	
Read code in clinical or referral files (none in immunisation or test files).	Read code for nursing or residential home status, e.g. 9N1G.00 *'Seen in nursing home'.*	Eligible Read code was labelling an entity type 132 record which also recorded that residence was warden-supported or sheltered accommodation, or that the patient lived alone.
Entity type 132 *'Residence'* in additional clinical details file.	Any entity type 132 record with residence field values 1 'nursing home' or 2 'residential home'.	Record labelled with a Read code for warden-controlled or sheltered accommodation, or with a Read code recording that the patient lived alone.

6.3 Smoking

Smoking status was identified as a descriptive variable for objective 2 using an algorithm and data management file developed by Dr Ian Douglas. Entity type code 4 *'Smoking'* in CPRD records smoking status as 'Non', 'Current', or 'Ex'. The latest smoking record prior to the relevant index date for each study objective was used to identify ex- or current smokers. Patients with their most recent smoking status recorded as non-smoker were only identified as non-smokers if all previous records were consistent with this: if they had a prior record of current or ex-smoking but their most recent status at the index date was non-smoker, they were classified as an ex-smoker. If there was no smoking record prior to the index date, the first smoking record after the index date was used. To be identified as a non-smoker from the first record after the index date, all later values must also have been consistent with never smoking: otherwise, the patient was classified as an ex-smoker.

For all other objectives, a fuller ascertainment of smoking status was developed which combined entity type 4 *'Smoking'* records, Read codes for smoking status (using a list compiled by Dr Sara Thomas and Elizabeth Millett), prescriptions for smoking cessation therapy (using a medication code list compiled by Dr Sara Thomas) from CPRD, and ICD-10 codes Z716 *'Tobacco abuse counselling'* and Z720 *'Tobacco use'* from HES. Smoking cessation at baseline was identified at the timepoint of the latest relevant record prior to study entry or (if there were no relevant records prior to study entry) the first record after study entry.

As the success of smoking cessation attempts is generally low, success of smoking cessation was not presumed, and a prescription for smoking cessation therapy was taken as evidence of current smoking.[131] Similarly, both ICD-10 codes were taken as evidence of current smoking. Read codes were categorised as: current smoker (e.g. 1374 *'Moderate smoker - 10-19 cigs/d'*), non-smoker (e.g. 1371.11 *'Non-smoker'*), ex-smoker (e.g. 1379 *'Ex-moderate smoker (10-19/day)'*), or ever-smoker (where it was unclear whether the patient was a current or ex-smoker, e.g. H310100 *'Smokers' cough'*).

A recorded non-smoker status may be consistent with never having smoked, or with being an ex-smoker. If a patient had a record with ex-smoker status on the same day as a record with non-smoker status (e.g. an entity type 4 record with value 'non-smoker' labelled with a Read code 137K.00 *'Stopped smoking'*), the patient was assumed to be an ex-smoker, and was assigned ex-smoker status. A record of current or ex-smoker status re-categorised any subsequent non-smoker record to ex-smoker status for that patient.

For 0.1% of smoking status records (2,183/ 2,068,167) the patient was recorded as both a current smoker and a non-smoker on the same day. These records could not be assigned a smoking status, and were discarded.

For 3.3% of smoking status records (68,016/ 2,068,167) it was unclear whether the patient was a current or ex-smoker, either because the only record was an 'ever-smoker' Read code, or because there was a records of both current and ex-smoker status on the same day. This was too small a group to be viable as a separate category in analyses. It was considered better to classify these patients as either current or ex-smokers, to avoid any risk of another record classifying them as non-smokers. These patients were classified as current smokers. In the context of inability to quantify cigarette pack-years, the effect of potentially misclassifying a small number of ex-smokers as current smokers was thought likely to be minor.

6.4 Body mass index

Body mass index (BMI) was calculated from entity type codes for height and weight, using an algorithm and data management file developed by Dr Krishnan Bhaskaran. Records of height were cleaned, including conversion of values >100 from centimetres to metres, and applying cut-offs at range extremes of <1.37m or >2.3m (4' and 7'6''). Records of weight were cleaned, which included discarding records at range extremes of <25.4kg or >222 kg (4 and 35 stone). BMI was calculated using the weight recorded closest to the index date.

6.5 Co-morbidities

Co-morbidities were identified using diagnoses in primary care using Read code lists developed by Dr Sara Thomas and Elizabeth Millett. Co-morbidities identified, for different purposes according to individual objectives, included: hypertension, congestive heart failure, ischaemic heart disease (including myocardial infarction, angina pectoris, and other ischaemic heart disease), peripheral vascular disease, cerebrovascular disease (including stroke, transient ischaemic attack, and cerebrovascular dementia), other dementia, chronic lung disease (including chronic obstructive pulmonary disease and other chronic lung conditions, but not asthma), cancer (including haematological and metastatic cancers), chronic liver disease (not including hepatorenal disease), connective tissue disorders (including rheumatoid arthritis and systemic lupus erythematosus), human

immunodeficiency virus infection, hyposplenia (including asplenia, coeliac disease and sickle cell disease), and history of cochlear implant.

Read codes were extracted from clinical, referral and test files. All co-morbidities were modelled as binary variables. A patient was defined as having a history of the co-morbidity from the first instance of a Read code for the relevant co-morbidity, until study exit. Absence of a Read code was assumed to indicate a negative status, as the absence of a co-morbidity would not be recorded in a patient's health care records. The diagnosis of hypertension was supplemented with entity code type 15 'Hypertension register' for the analysis of the association of CKD with post-infection mortality (objective 5).

6.6 Medications

6.6.1 Steroid use

Oral steroid prescriptions were extracted from the CPRD therapy files, using a medication code list compiled by Dr Sara Thomas, to identify steroid prescriptions within the three months before study entry for objective 5, the association of CKD with short-term mortality following infection.

6.6.2 Influenza and pneumococcal vaccinations

A history of pneumococcal vaccination or influenza vaccination was identified by extracting pneumococcal and influenza vaccine prescriptions from therapy files in CPRD (using a medication code list compiled by Elizabeth Millett) and Read codes for pneumococcal or influenza vaccination status from clinical, test, referral and immunisation files in CPRD. Pneumococcal and influenza vaccination status were determined separately. Any relevant prescription or Read code defined a positive vaccinated status for the relevant vaccine: the absence of a record was assumed to indicate a negative vaccination status. This classification of influenza and pneumococcal vaccine status was used as a descriptive variable for objective 2.

A more detailed definition of influenza and pneumococcal vaccination status was developed for the estimation of vaccine effectiveness according to CKD status (objective 4) and this is described in **Chapter 9**.

6.6.3 Anti-diabetes medications

The CPRD Product browser was searched using text terms and a hierarchical search using the strategy in **Table 6.2**.

Table 6.2: Search strategy for anti-diabetes medications

Search details	CPRD Product browser version 1.3.2 Database build 'ever' Search date 2 February 2012
Stage 1	BNF search to identify text-terms and BNF headings
BNF	BNF version 62 online
Stage 2	Text search of product names and drug substance names
Text terms	*insulin* and *insu*, *hypurin*, *bovine*, *porcine*, *actrapid*, *humulin*, *aspart*, *novo*, *glulisine*, *apidra*, *lispro*, *huma*, *detemir*, *levemir*, *glargine*, *lantus*, *lente*, *isophane*, *basal*, *protamine*, *biphasic* *sulfonyl* , sulphonyl*, *gliben*, *gliclaz*,*zicron*, *diamicro*, *vitil*, *glime*, *amaryl*, *glipiz*, *minod*, *tolbu* *biguan*, *metfor*, *metsol*, *glucoph*, *bolamy*, *metabe* *acarbo*, *glucob*, *exena*, *byet*, *liragl*, *victoz*, *nategl*, *starl*, *piogl*, *actos* (and did not upload lactose-related codes), *competac*, *repagl*, *prandi*, *saxagl*, *ongl*, *sitagl*, *janu*, *vildagl*, *galv*, *eucrea*
Stage 3	Hierarchical search
BNF headings	06010101*, 06010102*, 06010103*, 06010201*, 06010202*, 06010203*

BNF, British National Formulary

The product codes identified for insulin therapy are listed in **Appendix B**. A prescription of insulin itself was required, rather than products which may be prescribed as adjuncts to insulin treatment. Thus glucagon prescriptions, lancets, blood sugar monitoring devices and delivery devices (unless pre-filled with insulin) were not considered sufficient evidence on their own of insulin use, and were not included.

The product codes identified for oral anti-diabetes medications are listed in **Appendix B**. Medications no longer licensed at the time of the search, such as first generation sulphonylureas and rosiglitazone, were included, to capture historical prescriptions. Metformin was included, as it is a commonly prescribed first-line oral anti-diabetes medication. Metformin also has other indications, including treatment of polycystic ovarian syndrome and metabolic syndrome. All patients included in the study cohort were required to have a code for diabetes mellitus, and so prescription of metformin for indications other than diabetes is likely to be rare among this cohort. Guar gum, a dietary fibre which can be prescribed as an anti-diabetes medication was not included, as this is not its main indication, and it is not commonly used for this purpose. Acarbose, which may be prescribed for either type 1 or type 2 diabetes mellitus, was not included as (unlike other oral anti-diabetes medications) it is not an indicator of type 2 diabetes mellitus. If would be

unusual to be prescribed this as a solitary anti-diabetes medication, and so its exclusion was unlikely to have affected the ascertainment of anti-diabetes medications for many patients.

6.7 HbA1C

Glycated haemoglobin, or HbA1C, reflects mean blood glucose levels over the preceding three months. It should be monitored at least every six months among patients with diabetes, and more frequently if blood glucose control is not stable.[132]

HbA1C test results in CPRD may be reported in two forms, each with a different reference range. HbA1C assays in the UK were predominantly aligned to the Diabetes Control and Complications Trial (DCCT) assay and reported as a percentage during the study period. From 1 June 2009, HbA1C assays were International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) standardised and reported in mmol/mol, although for a period of dual reporting, the equivalent percentage was also provided, labelled "DCCT-aligned".[132]

I aimed to identify the latest HbA1C test result at baseline (study entry). As regular monitoring of HbA1C is an important component of care for patients with diabetes, I included a 'none recorded' category. Rather than being a missing data category, this was an indicator for diabetes which may have been poorly monitored.

6.7.1 Data extraction and cleaning

HbA1C could be recorded in CPRD using a template, available in the database as entity type 275 'HbA1c - diabetic control', as a Read code containing the HbA1C result, or as a more general test result (such as entity type 288 'Other laboratory tests' with a Read code label indicating that this was an HbA1C result.

Read codes were identified from the CPRD Medical Browser searching for Read terms (*HbA*; *Hb*; *diab*control*; *glyc*haem*; *glyc*Hb*; *A1*) and Read codes (42c*; 42V*; 42W*; 44T*; 66A*; C108*; C109*; C10E*;C10F*). This identified six codes which provided a value for an HbA1C test, and 11 terms which were labels for an HbA1C test but without the result (**Table 6.3**).

From clinical, test and referral files I extracted all records with an entity code of 275, or any of the eligible Read codes in **Table 6.3**. There were no eligible records in the additional clinical files.

Table 6.3: Read codes for HbA1C tests and results

	Read code	Read term
Read code containing HbA1C test result	42c0.00	HbA1 < 7% - good control
	42c1.00	HbA1 7 - 10% - borderline control
	42c2.00	HbA1 > 10% - bad control
	42W1.00	Hb. A1C < 7% - good control
	42W2.00	Hb. A1C 7-10% - borderline
	42W3.00	Hb. A1C > 10% - bad control
Read code indicating HbA1C test	42c..00	HbA1 - diabetic control
	42c3.00	HbA1 level (DCCT aligned)
	42W..00	Hb. A1C - diabetic control
	42W..11	Glycosylated Hb
	42W..12	Glycated haemoglobin
	42W4.00	HbA1c level (DCCT aligned)
	42W5.00	Haemoglobin A1c level - IFCC standardised
	42WZ.00	Hb. A1C - diabetic control NOS
	44TB.00	Haemoglobin A1c level
	44TC.00	Haemoglobin A1 level
	44TL.00	Total glycosylated haemoglobin level

HbA1C, glycated haemoglobin

Records with no result available (Read code indicating a test performed but no entity type attached to contain the results) were discarded (n=16,762). Two records with entity type 363 'Lipoprotein electrophoresis' were discarded, because these were unlikely to encode HbA1C results, and because this entity type had no numerical data field attached for an appropriate result. Records with a missing data (n=70) or a date after study exit (n=5,983) were discarded. This left 2,850,694 records, 97.7% of which were entity code 275 *'HbA1c – diabetic control'* with an attached Read code for an HbA1C test (**Table 6.4**).

Table 6.4: Origin of records of HbA1C tests

		Entity code					
		213 Blood glucose	**274 Fasting glucose**	**275 HbA1c – diabetic control**	**288 Other laboratory tests**	**none**	**Total**
Read code	**Read code including result**	0	0	44,369	40	11,487	55,896
	Read code for HbA1C test	8,047	1	2,784,487	2,191	–	2,794,726
	None	–	–	72	–	–	72
Total		8,047	1	2,828,928	2,231	11,487	2,850,694

HbA1C, glycated haemoglobin

6.7.2 Categorising HbA1C results

HbA1C results were categorised as good (<7% or <53mmol/mol), borderline (7–10%, or 53–86 mmol/mol), poor (>10% or >86 mmol/mol), or none recorded (no result prior to study entry).

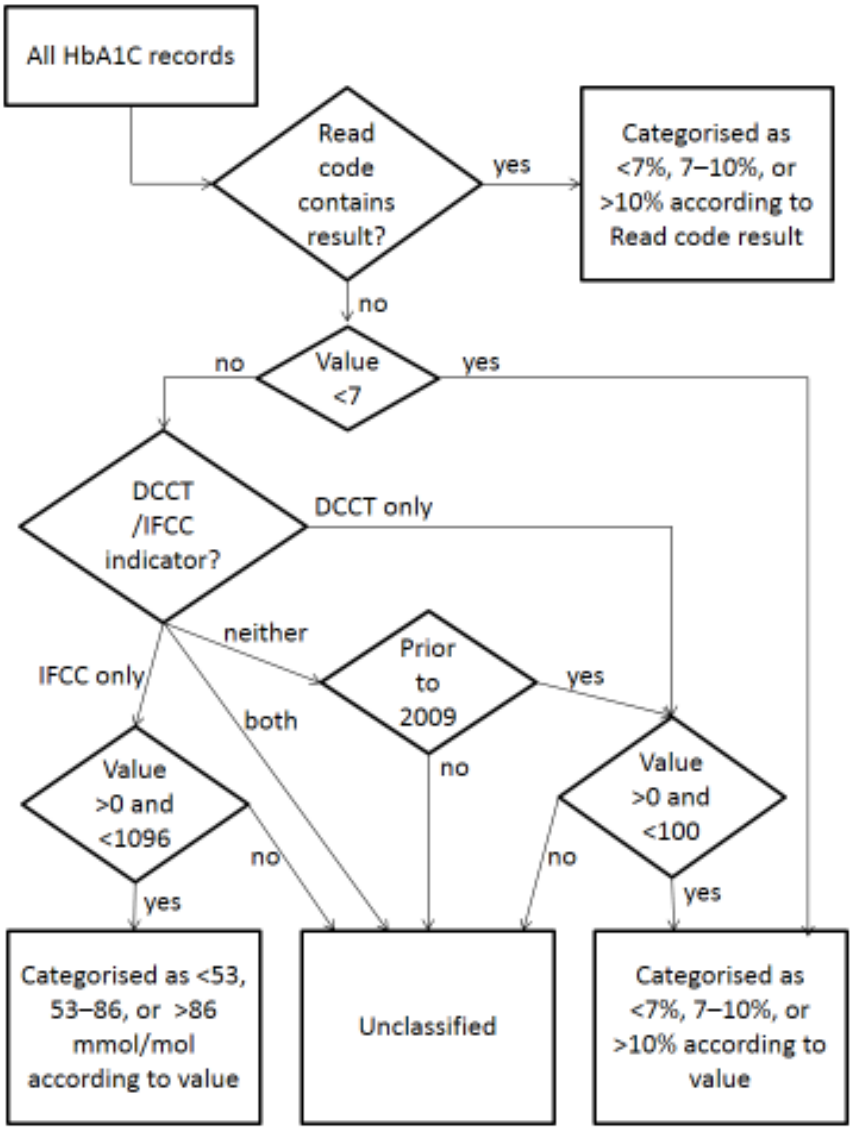
For numerical results recorded as an entity code, there was potentially some ambiguity as to whether these should be treated as DCCT-aligned (%) or IFCC-standardised (mmol/mol). Each entity code had seven data fields attached. The second data field contained a numerical value, which provided the test result. Other data fields of interest were the first, which contained an 'operator' (e.g. <, >, =), the third, which contained a 'unit' (e.g. %), and the fifth and sixth, which contained the lower and upper limits of the normal range. These data fields were used to identify DCCT-aligned from IFCC-standardised results, for appropriate categorisation.

Indicators of DCCT-alignment were: Read codes 42c3.00 *'HbA1 level (DCCT aligned)'* or 42W4.00 *'HbA1c level (DCCT aligned)'*, value 1–6, value 7 with operator recorded as <, unit recorded as %, or normal range recorded as 4–6. Indicators of IFCC-standardisation were: Read code 42W5.00 *'Haemoglobin A1c level - IFCC standardised'*, unit recorded as mmol/mol, or normal range recorded as 20–42.

HbA1C records were categorised as shown in **Figure 6.1**. This assumed that results prior to 2009 were DCCT-aligned unless otherwise specified. This left 10,692 (0.4%) of the entity records with a potentially usable result as unclassified. These records either had an indicator of both IFCC and DCCT with value ≥7, or no indicator of IFCC nor DCCT in or after 2009 with value ≥7.

Where multiple records occurred on the same date, I prioritised: (1) results from Read codes (as these had less potential for misclassification); (2) entity test results with an uncontradicted DCCT or IFCC indicator; and (3) other entity test results. After prioritisation, 5528 pairs and 3 triplets of clashing results on the same day remained, and these were discarded.

Figure 6.1: Categorisation of HbA1C records



6.8 Alternative approaches considered

6.8.1 Why were secondary care records not used to identify co-morbidities?

The impossibility of distinguishing a patient who truly does not have a co-morbidity from a patient with missing co-morbidity data in routinely-collected health records means that co-morbidity status is particularly vulnerable to ascertainment bias. Diagnoses were not supplemented with HES diagnoses, as this might have resulted in greater ascertainment of co-morbidities among patients with more hospital attendances. Patients with CKD have higher rates of hospital attendance than patients without CKD, and infection is a common cause of hospital admission among older people.[9, 107] Differential ascertainment of co-morbidities among patients with CKD or among patients with frequent infection was relevant to each analysis in this thesis. For example, differential ascertainment of co-morbidity status would be likely to bias estimates of the association between CKD and infection after adjustment for co-morbidities (objective 3). The direction of bias could differ for each co-morbidity, and would be difficult to predict, as it would depend on the relationship of the co-morbidity with CKD prevalence, all-cause hospital admission rates, infection-related hospitalisation rates, and whether the co-morbidity coding in secondary care was differential according to the reason for hospitalisation.

6.8.2 Why were medications not used to identify co-morbidities?

Medication status may vary systematically depending on CKD status. A large number of commonly prescribed medications are contraindicated by impaired renal function.[73] In addition, medications used to treat CKD may overlap with those prescribed for other co-morbidities. For example, an angiotensin converting enzyme inhibitor (ACE inhibitor) should be prescribed to patients with diabetes and proteinuria: but is also commonly prescribed as an anti-hypertensive medication.[96] Adjusting estimates of the association between CKD and infection for medications rather than diagnosis could result in bias, the direction of which would be difficult to predict as it would depend upon the relationship between the medication and CKD, the medication and the other co-morbidities for which it is prescribed, and the relationship of these other co-morbidities with infection. As a general approach, therefore, we did not use medications as proxies for co-morbidities, and preferred to use clinical diagnoses to identify co-morbidities.

6.8.3 Could hypertension have been identified from blood pressure recordings?

As hypertension is 'silent', or asymptomatic, until quite advanced, and so may be under-diagnosed, I investigated using blood pressure recordings to identify hypertension in

addition to diagnoses. I identified patients as hypertensive from the first date of: a diagnostic code for hypertension (as described above); inclusion on the hypertension register (entity type code 15); or the third recording of raised blood pressure if all 3 records were within a year of each other (not including records on the same day). Recordings of raised blood pressure could be either a Read code recording raised blood pressure, or an entity type code 1 ('Blood pressure') record with systolic blood pressure recorded as ≥130 mm Hg (only acceptable if ≤350 mm) or diastolic blood pressure >80 mmHg (only acceptable if ≤200). Entity type 1 records were only included if they were also labelled with an appropriate Read code indicating a blood pressure measurement. The thresholds for raised blood pressure were obtained from NICE guidelines for medication of hypertension for patients with diabetes.[133]

Applying this definition to study cohort A obtained a prevalence of hypertension of 76.0% (166,626/219,145) at baseline and 89.3% (195,621/219,145) by study exit. This was not useful for discriminating between patients. It was therefore decided to use clinical diagnosis to define hypertension, without blood pressure recordings. The process of diagnosis and recording suggests that clinically relevant hypertension has been identified, which may be more discriminatory in terms of risk. It is likely to capture patients with hypertension controlled by medication, and risks missing patients with undiagnosed (and uncontrolled) hypertension.

6.8.4 Could we have distinguished type 1 from type 2 diabetes mellitus?

Both infection risk and CKD prevalence may vary by the type of diabetes mellitus. Ideally, patients with type 1 and type 2 diabetes mellitus would be distinguished to allow consideration of this. However, diagnostic coding of type of diabetes is known to be problematic, with most patients not having the type of diabetes recorded in electronic health records, and considerable misclassification of those who do.[134] Completeness and validity can be improved by assigning diabetes type according to an algorithm developed by de Lusignan *et al.* which includes prescriptions, age at diagnosis and BMI at diagnosis.[135] Unfortunately, when patients register with a primary care practice, prevalent and incident diagnoses of diabetes cannot reliably be distinguished.[71] Hence, as many patients will have changed primary care practice between first diagnosis of diabetes and eligibility for our study population at age 65 years, the date of first diagnosis cannot be identified using these data for this study population. Neither duration of diabetes, nor age and BMI at diagnosis are therefore identifiable. We described the patients' prescription histories of

insulin and oral anti-diabetes medications, but were unable otherwise to identify type 1 from type 2 diabetes mellitus.

RESULTS SECTION

Chapter 7 describes the burden of acute community-acquired infection among older people with diabetes (objective 2). **Chapter 8** presents an investigation into the association between CKD and incidence of community-acquired LRTI (including pneumonia as a subset) and sepsis (objective 3). **Chapter 9** presents a study of the extent to which community-acquired LRTI may be preventable with pneumococcal and influenza vaccination among older people with diabetes, according to CKD status (objective 4). **Chapter 10** describes a study of the association between CKD and short-term mortality following community-acquired pneumonia and sepsis (objective 5).

Results are presented in journal article format, each with a brief introduction, methods, results, and discussion of the individual study. Any additional results or further discussion relevant to the individual study is presented in the relevant chapter following the journal article.

Chapter 7. The burden of community-acquired infection among older people with diabetes mellitus

This chapter presents and discusses the results of a study to investigate the burden of acute, community-acquired infections among older people with diabetes mellitus (objective 2). The results are presented and discussed in Paper 2. The Read code lists used to identify patients with diabetes mellitus (referred to in the article as additional supporting information) are available in **Appendix B**. The chapter concludes with a supplementary exploration of the impact of incomplete data linkages.

7.1 Introduction to Paper 2

This study used a retrospective cohort design to assess the infection incidence rates among patients aged ≥65 years with diabetes mellitus in CPRD (study cohort A, **3.9.1**). This was a descriptive study. Key characteristics of the study population (including demographics, smoking status and common co-morbidities) were described at baseline, using definitions described in **Chapter 6**.

The primary outcomes were incidence rates of community-acquired lower respiratory tract infection (LRTI), pneumonia (as a subset of LRTI), urinary tract infection (UTI), and sepsis. These infections were expected to be responsible for a high burden of morbidity or mortality among the study population. LRTI and UTI are common infections, responsible for a high burden of morbidity among the general population, and particularly among older people.[5, 136] Pneumonia and sepsis are severe infections, and the commonest infectious causes of death among patients receiving dialysis.[28] Methods used to identify community-acquired infection incidence were detailed in **Chapter 4**. Infection incidence rates were summarised by age, sex, region and year.

Secondary outcomes were all-cause hospital admission within 28 days of infection onset and all-cause mortality within 28 and 90 days of infection onset. Infections may cause hospital admissions and death directly or indirectly through complications, such as myocardial infarction or stroke.[137] All-cause hospitalisation and mortality were therefore described within a time-period following infection, rather than including only hospitalisations and deaths attributed to infection. The time period was defined with respect to infection onset date, even for infections with longer duration, to avoid biasing estimates of hospitalisation and mortality upwards for patient groups prone to longer infection episodes.

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	Helen McDonald
Principal Supervisor	Dr Dorothea Nitsch
Thesis Title	The epidemiology of infections among older people with diabetes mellitus and chronic kidney disease

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?	Diabetic Medicine		
When was the work published?	2014		
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)	The original idea for this study was developed by Dorothea Nitsch and Sara Thomas, who wrote the study proposal and obtained ethics approval and funding for the study. Sara Thomas obtained the study data from CPRD. I developed a detailed study design from the original study proposal, with supervision from Dorothea Nitsch and Sara

	Thomas, and advice from Alan Sinclair. Identification of the study population was discussed in detail with Dorothea Nitsch, Sara Thomas, Alan Sinclair and Simon de Lusignan and used a codelist of Read codes for diabetes mellitus which was compiled by Sara Thomas. Lower respiratory tract infections (LRTI) and pneumonia were identified using methods designed by Elizabeth Millett and Sara Thomas and previously implemented by Elizabeth Millett to define LRTI and pneumonia among older people using linked CPRD-HES records. These methods (including do-files) were shared with me, and I applied them to the study population of older people with diabetes mellitus. Elizabeth Millett and Sara Thomas provided advice and practical support with this time-consuming task. I adapted Elizabeth Millett's methods to identify urinary tract infections (UTI) and sepsis, with advice from Dorothea Nitsch and Sara Thomas. Codelists to identify UTI and sepsis were compiled by Sara Thomas and myself, respectively. Individual codelists developed and shared by colleagues at LSHTM and used to identify other patient characteristics are attributed in Chapter 6. In particular, body mass index and smoking status data were extracted, cleaned and classified using do-files written by Krishnan Bhaskaran and Ian Douglas, respectively. I conducted all data management and analysis, and drafted the manuscript, which was commented on by all co-authors. The article was peer-reviewed, and I edited the manuscript in response to feedback from reviewers, with advice from all co-authors.

Student Signature: HMcDonald **Date:** 10 April 2015

Supervisor Signature: Dorothea Nitsch **Date:** 14/4/15

DIABETICMedicine

DOI: 10.1111/dme.12384

Research: Complications

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

H. I. McDonald[1], D. Nitsch[1], E. R. C. Millett[1], A. Sinclair[2] and S. L. Thomas[1]

[1]London School of Hygiene and Tropical Medicine, London and [2]Institute of Diabetes for Older People at the University of Bedfordshire, Luton, Bedfordshire, UK

Accepted 29 November 2013

Abstract

Aim To describe the incidence of acute community-acquired infections (lower respiratory tract infections, urinary tract infections and sepsis) among the UK population aged ≥65 years with diabetes mellitus, and all-cause 28-day hospital admission rates and mortality.

Methods We used electronic primary care records from the Clinical Practice Research Datalink, linked to death certificates and Hospital Episode Statistics admission data, to conduct a retrospective cohort study from 1997 to 2011.

Results Among the 218 805 older people with diabetes there was a high burden of community-acquired infection, lower respiratory tract infections having the highest incidence (crude rate: 152.7/1000 person-years) followed by urinary tract infections (crude rates 51.4 and 147.9/1000 person-years for men and women, respectively). The incidence of all infections increased over time, which appeared to be driven by the population's changing age structure. Most patients diagnosed with pneumonia and sepsis were hospitalized on the same day (77.8 and 75.1%, respectively). For lower respiratory tract infections and urinary tract infections, a large proportion of 28-day hospitalizations were after the day of diagnosis (39.1 and 44.3%, respectively), and a notable proportion of patients (7.1 and 5.1%, respectively) were admitted for a cardiovascular condition. In the 4 weeks after onset, all-cause mortality was 32.1% for pneumonia (3115/9697), 31.7% for sepsis (780/2461), 4.1% for lower respiratory tract infections (5685/139 301) and 1.6% for urinary tract infections (1472/91 574).

Conclusions The present large cohort study provides up-to-date detailed infection incidence estimates among older people with diabetes in the community, with variation by age, sex and region and over time. This should be of use for patient communication of risk and future healthcare planning.

Diabet. Med. 31, 606–614 (2014)

Introduction

Community-acquired infections are common among older adults, with a high burden of morbidity and mortality [1–4]. Pneumonia is the second most common cause of death in people aged ≥75 years in England [5]. Hospitalizations for infection are rising: age-standardized hospital admission rates for community-acquired pneumonia and urinary tract infections more than doubled between 2000/2001 and 2010/2011 in England [6]. The increase in pneumonia hospitalizations has been most marked among older adults [7]. The cost of hospitalizations was estimated at £235 m for pneumonia and £316 m for urinary tract infection in 2010/2011 [6]. Diabetes is a risk factor for hospitalization with and mortality from infection [8–13]. A higher prevalence of comorbidities such as diabetes mellitus has been suggested as a driving factor for the rising burden of infection-related hospitalizations [6,7].

The number of adults in England with diabetes mellitus is predicted to rise from 3.1 million in 2010 to 4.6 million by 2030 [14]. Data on the burden of infection among older adults with diabetes from a community or primary care perspective, and their relationship with hospitalization and mortality, are scarce. A large cohort study in Canada

Correspondence to: Helen I. McDonald. E-mail: helen.mcdonald@lshtm.ac.uk

What's new?

- The present large cohort study is the first to describe the burden of acute infections, including infections managed in primary care, among older people with diabetes for use in healthcare planning and communication of risk with patients.
- On average per year among 1000 patients there were 152.7 lower respiratory tract infections (95% CI 151.3–154.1) and 99.6 urinary tract infections (95% CI 98.4–100.8).
- All infection rates were found to be increasing over time.
- Within 28 days of pneumonia, 81.4% of patients were hospitalized and 32.1% had died.
- An appreciable proportion of 28-day hospitalizations after lower respiratory or urinary tract infection were for cardiovascular conditions.

described the incidence of a range of infections among people with diabetes of all ages, but not by age group, and noted that the association between diabetes and infection changed with age [15]. Studies of risk factors for community-acquired infection among older adults have not been designed for precise estimates of infection incidence among the subgroup with diabetes [16,17].

Understanding the burden of infections in the community among older adults with diabetes, and the short-term risk of hospitalization and mortality after infection, is essential to communicating risk of infections to patients, for designing effective preventive care strategies, and for future healthcare service planning.

The aim of the present study was to describe the incidence of acute community-acquired infections (urinary tract infections, lower respiratory tract infections, including pneumonia as a subset, and sepsis) among the UK population aged ≥ 65 years with diabetes mellitus, and the short-term hospitalization and mortality rates following these infections.

Materials and methods

Data sources

The Clinical Practice Research Datalink (CPRD, formerly GPRD) is a large UK database of anonymized primary care medical records [18]. We used the May 2011 dataset, comprising 12.8 million patient records at 627 practices. Diagnoses are entered directly by healthcare workers during the patient consultation, in the form of Read codes. The records also include patient demographics, prescriptions, health behaviours, test results and interactions with secondary care such as referrals. The CPRD population has been found to be representative of the general UK population [19]. The CPRD asserts a range of data quality checks, and the validity of recorded diagnoses in the CPRD is generally high [19,20].

Another strength of the CPRD is the availability of data linkage for the subset of patients (> 50% of the total patients included in the CPRD) registered at practices in England which participate in CPRD data linkage [18]. The present study used linked data on dates and diagnoses for all hospital inpatient admissions to NHS hospitals in England from Hospital Episodes Statistics, and mortality and socio-economic status from the Office for National Statistics [21,22].

Study population and follow-up

Patients in CPRD aged ≥65 years with a diagnostic code for diabetes mellitus were eligible for inclusion in the study. Two lists of Read codes were used, 'defining' codes (sufficient evidence of diabetes) and 'possible' codes (requiring confirmation). All patients with a 'defining' code were included (e.g. C10F.11 *'Type II diabetes mellitus'*). Patients with a 'possible' code (e.g. 9N1Q.00 *'Seen in diabetic clinic')* were only included if there was a history of prescription of insulin or oral antidiabetes medication (Tables S1 and S2).

Patients entered the study at whichever was the latest of the following timepoints: the diabetes diagnosis date, 65th birthday, 1 year after practice registration date, the date the practice reached CPRD quality control standards or 1 April 1997. The 1-year delay from registration was to prevent overestimation of incidence from recording of historical events at new patient registration and early consultations [23]. Patients exited the study at whichever was the earliest of the following timepoints: date of death (recorded in CPRD or Office for National Statistics data), patient transferring out from the practice; last data collection from the practice, or 31 March 2011.

Definition of infections

We studied urinary tract infection, lower respiratory tract infection, pneumonia and sepsis. Urinary tract infections and lower respiratory tract infections are common and are responsible for a high burden of morbidity and mortality among the older population, while pneumonia and sepsis are rare but serious events which we would expect to be well-ascertained in primary care records. Urinary tract infections, lower respiratory tract infections and sepsis were defined and analysed separately, while pneumonia was a subset of lower respiratory tract infections.

Each infection was defined by a clinical diagnosis recorded in primary care or hospital discharge records. To avoid overestimation from repeat attendances for the same infection, diagnostic codes recorded within 28 days of one another were attributed to a single episode of infection, with the index date defined by the first diagnostic code, and duration until 28 days after the last diagnostic code. Three clinical epidemiologists agreed each list of diagnostic codes

to be used to define each infection before data analysis. Code lists are available on request.

Pneumonia codes were a subset of lower respiratory tract infection codes. If any lower respiratory tract infection included a pneumonia code, the pneumonia index date was the date of the first diagnosis of pneumonia, and the pneumonia episode ended on the end-date of the lower respiratory tract infection episode within which it occurred.

For conservative estimates, Read code IJ4.00 '*Suspected urinary tract infection*' alone did not define an infection but did continue an ongoing episode of urinary tract infection if it occurred within 28 days of another urinary tract infection diagnostic code. This was designed to avoid over-recording of urinary tract infections from non-confirmed diagnoses or from repeat attendances with ongoing infection.

An infection was designated as hospital-acquired: if the index diagnosis occurred during or within 14 days of a hospital admission (recorded in Hospital Episodes Statistics records for patients with linked data, or recorded in the CPRD for patients with unlinked data); if the index diagnosis was recorded in a hospital discharge record, but was not the primary reason for hospital admission; or if any of the diagnoses in the episode recorded that the infection was postoperative or otherwise hospital-acquired (e.g. K190299 '*Postoperative urinary infection*'). Other infections were classified as community-acquired. These were necessarily either first recorded in primary care records, or represented the primary reason for hospital admission. The results presented are for community-acquired infections only: hospital-acquired infections were not included as outcomes.

Patients were at risk of a community-acquired infection while eligible for study inclusion except during an infection episode (community- or hospital-acquired) or during a Hospital Episodes Statistics hospitalization spell or the 14 days after hospital discharge. Time at risk was calculated separately for each type of infection; a patient could be at risk of a lower respiratory tract infection despite an ongoing urinary tract infection, for example.

The methods described above were based on or adapted from previous work by Millett *et al.* [24] defining lower respiratory tract infections and pneumonia.

Covariates

Socio-economic status was described using the index of multiple deprivation, a composite area-level marker of deprivation [22]. The Office for National Statistics index of multiple deprivation estimates from 2007 were linked to individual patient records by the postcode of patient residence. Smoking status and BMI were described using CPRD data at the start of follow-up (baseline). Comorbidities, medications and vaccination status were described using CPRD records at baseline and at study exit. No influenza vaccinations recorded > 3 years before study entry were included at either timepoint. For baseline comorbidities, medications and pneumococcal vaccination, any record of positive status in the patient's records from their registration at the practice up to or including the study entry date was eligible. Comorbidities were based on diagnostic Read codes and included cardiovascular disease (myocardial infarction, other ischaemic heart disease, congestive heart failure, stroke and transient ischaemic attacks), chronic lung disease (chronic obstructive pulmonary disease and chronic interstitial lung diseases but not asthma) and peripheral vascular disease. Code lists are available on request.

Data analysis

All data were used to calculate incidence and mortality estimates. Analyses were restricted to patients with Hospital Episodes Statistics-linked data for description of hospitalizations. Incidence rates were calculated for each infection using Poisson regression with lexis expansions for age and calendar year, and a random-effects model to accommodate multiple episodes. We conducted likelihood ratio tests for the association of sex with incidence of each infection type, and for the sex-specific linear association of age group with incidence of each infection type. Age standardization used the Office for National Statistics mid-year population estimate for the UK in 2004 [22]. Age-standardized regional rates were presented only within England, as Hospital Episodes Statistics-linked data are not available for Scotland, Wales and Northern Ireland, and the inability to exclude time in hospital for these regions meant their rates were not comparable. For each infection, the proportion of infections among patients with Hospital Episodes Statistics-linked data who were admitted to hospital within 4 weeks of diagnosis, and the subset of these admitted to hospital on the day of admission, were calculated. The cause of admissions for patients admitted within 4 weeks was described using the primary diagnosis of the first episode of the first admission after infection onset. For each infection, the 4-week case-fatality rate using all-cause mortality was calculated, using deaths recorded in Office for National Statistics-linked mortality data or CPRD records.

Data analysis was conducted using Stata version 12.0, except age standardization, for which Microsoft Excel 2007 was used.

Ethics approval

The study was approved by the Independent Scientific Advisory Group of the CPRD (ISAC reference 11_033) and the London School of Hygiene and Tropical Medicine Ethics Committee (LSHTM reference 6116).

Results

The study population comprised 218 805 patients, with up to 7 years follow-up (Table 1). Hospital Episodes Statistics

Table 1 Characteristics of study population at baseline and before end of follow-up (N = 218 805)

	At Baseline		Before end of follow-up	
	Men	Women	Men	Women
Median (interquartile range) time in study, years*	3.9 (1.7–6.1)	4.0 (1.7–7.1)		
Median (interquartile range) age, years	70 (65–76)	72 (66–79)		
Ethnicity, ***n*** **(%)**				
Bangladeshi	85 (0.8)	60 (0.1)		
Black African	125 (0.1)	127 (0.1)		
Black Caribbean	477 (0.4)	519 (0.5)		
Black Other	88 (0.1)	83 (0.1)		
Chinese	86 (0.1)	85 (0.1)		
Indian	836 (0.8)	725 (0.7)		
Mixed	89 (0.1)	99 (0.1)		
Other Asian	216 (0.2)	169 (0.2)		
Pakistani	270 (0.2)	219 (0.2)		
White	44 400 (40.1)	43 524 (0.3)		
Other	446 (0.4)	433 (0.4)		
Missing or unknown	63 617 (57.4)	62 027 (57.4)		
Socio-economic status: index of multiple deprivation quintile, ***n*** **(%)**				
1 (least deprived)	12 105 (10.9)	10 485 (9.7)		
2	14 037 (12.7)	12 831 (11.9)		
3	12 290 (11.1)	11 893 (11.0)		
4	11 595 (10.5)	12 273 (11.4)		
5 (most deprived)	8626 (7.8)	9494 (8.8)		
Not available	52 082 (47.0)	51 094 (47.3)		
BMI, ***n*** **(%)**				
<18.5 kg/m^2	619 (0.6)	1741 (1.6)		
18.5–24.9 kg/m^2	22 140 (20.0)	22 691 (21.2)		
25–29.9 kg/m^2	47 111 (42.5)	34 086 (31.5)		
30–34.9 kg/m^2	26 076 (23.5)	24 279 (22.5)		
≥35 kg/m^2	9628 (8.7)	16 505 (15.3)		
Missing	5161 (4.7)	8768 (8.1)		
Smoking status, *n* (%)				
Non-smoker	28 844 (26.1)	54 759 (50.7)		
Current	16 380 (14.8)	12 812 (11.9)		
Previous	63 764 (57.6)	37 650 (34.8)		
Missing	1747 (1.6)	2849 (2.6)		
Comorbidities, *n* (%)				
Cardiovascular disease	43 927 (39.7)	34 979 (32.4)	55 539 (50.2)	46 439 (43.0)
Chronic lung disease	9258 (8.4)	7112 (6.6)	14 524 (13.1)	10 939 (10.1)
Peripheral vascular disease	9963 (9.0)	5609 (5.2)	16 195 (14.6)	9488 (8.8)
Vaccinations, *n* (%)				
Pneumococcal: ever	55 641 (50.2)	51 103 (47.3)	87 418 (78.9)	80 532 (74.5)
Influenza: < 3 years before study entry	83 479 (75.4)	80 288 (74.3)	100 296 (90.6)	95 688 (88.5)
Antidiabetes medications, *n* (%)				
Insulin only	4260 (3.8)	4049 (3.7)	4242 (3.8)	3897 (3.6)
Oral medication only	41 268 (37.3)	37 806 (35.0)	59 249 (53.5)	55 255 (51.1)
Both insulin and oral medication	6529 (5.9)	6082 (5.6)	15 924 (14.4)	14 491 (13.4)
None recorded	58 678 (53.0)	60 133 (55.6)	31 320 (28.3)	34 427 (31.9)
Total	110 735	108 070		

*Time in study includes person-time as a hospital inpatient or during an infection, which was excluded from time at risk of community-acquired infection.

data linkage was available for 128 373 patients (58.7%). The median (interquartile range) age of the cohort at baseline was 71 (65–77 years). Few patients (n = 8137; 3.9%) were medicated with insulin alone (with no history of oral antidiabetes medications), consistent with a preponderance of Type 2 diabetes mellitus in this age group. At baseline, over a third of the cohort had a history of cardiovascular comorbidity and 74.8% had received an influenza vaccination within the previous 3 years. Before the end of follow-up, 76.7% of patients had received a pneumococcal vaccine: 28.7% of unvaccinated patients (14 579/50 855) exited the study before 2003, when the pneumococcal vaccine was phased in for people > 60 years old [25].

Incidence by age and sex

For all infections, incidence increased sharply with increasing age (Table 2). For example, pneumonia incidence was 6–8 times higher among patients aged ≥85 years than patients aged 65–69 years.

Women were more likely to experience urinary tract infections than men in every age group, but this difference

Table 2 Infection incidence by age and sex among older people with diabetes (N = 218 805)

		Incidence rate /1000 person-years (95% CI)					
Age group		65–69 years	70–74 years	75 –79 years	80–84 years	≥85 years	*P**
Lower respiratory tract infection	Men	135.2 (129.6–140.8)	145.2 (139.8–150.7)	159.9 (153.8–166.0)	161.8 (155.9–167.8)	217.9 (209.2–226.7)	<0.001
	Women	137.1 (134.1–140.1)	145.2 (142.2–148.1)	151.9 (148.7–155.1)	167.4 (163.5–171.4)	208.9 (203.2–214.6)	<0.001
Pneumonia	Men	4.51 (4.15–4.87)	6.91 (6.45–7.36)	11.03 (10.36–11.71)	18.91 (17.69–20.14)	36.88 (34.19–39.56)	<0.001
	Women	4.19 (3.79–4.59)	5.81 (5.35–6.26)	7.83 (7.26 –8.39)	13.67 (12.75–14.60)	27.24 (25.51–28.96)	<0.001
Urinary tract infection	Men	29.4 (28.3–30.5)	40.8 (39.4–42.1)	59.1 (57.1–61.1)	83.4 (80.0–86.7)	141.2 (134.0–148.5)	<0.001
	Women	110.2 (107.5–112.9)	123.4 (120.6–126.2)	146.1 (142.8–149.3)	175.0 (170.8–179.3)	223.0 (216.6–229.3)	<0.001
Sepsis	Men	1.53 (1.32–1.74)	1.97 (1.73–2.21)	2.92 (2.58–3.27)	3.92 (3.38–4.45)	5.77 (4.83–6.71)	0.75
	Women	1.87 (1.60–2.13)	1.92 (1.66–2.17)	2.37 (2.08–2.66)	2.81 (2.43–3.18)	4.31 (3.77–4.86)	0.02

*Sex-specific likelihood ratio test for trend by age group.

reduced with increasing age. For lower respiratory tract infections, the incidence was similar for men and women within each age group. For pneumonia, the incidence was higher among men than women for all age groups > 70 years.

Trends over time

The crude incidence of all infections increased over time (Fig. 1A). For example, the crude incidence of lower respiratory tract infections among men was 117.1/1000 person-years in 1997 (95% CI 110.9–123.4), and 154.6 /1000 person-years in 2010 (95% CI 150.5–158.8). Much of this trend disappeared after standardization for age (Fig. 1B).

Differences by region

For lower respiratory tract infections, the highest incidence rates were in the North West (men 139.9/1000 person-years [95% CI 140.2–167.6]; women 177.5/1000 person-years [163.3–191.7]), Yorkshire and the Humber and the West Midlands regions, while London experienced the lowest incidence (men 116.8/1000 person-years [95% CI 112.1–121.6]; women 114.6/1000 person-years [110.3–119.0] (Fig. 2). For urinary tract infection, the two regions with the highest incidence were the West Midlands (men

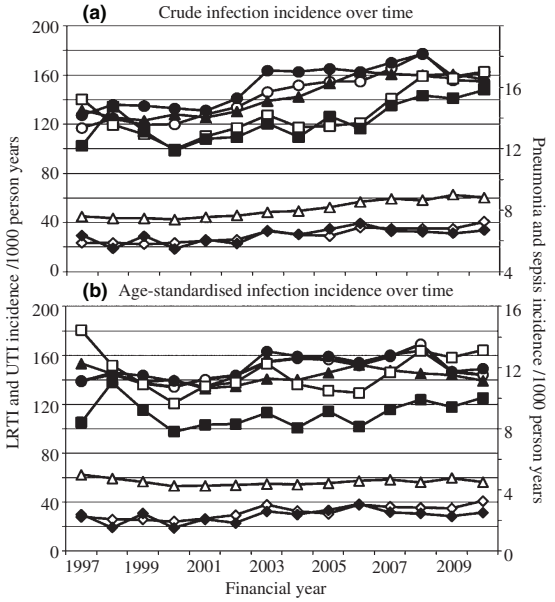


FIGURE 1 Crude and age-standardized infection incidence rates by sex over time among older people with diabetes (N = 218 805). UTI, urinary tract infection; men, white symbols; women, black symbols; circles, lower respiratory tract infection; squares, pneumonia; triangles, UTI; diamonds, sepsis.

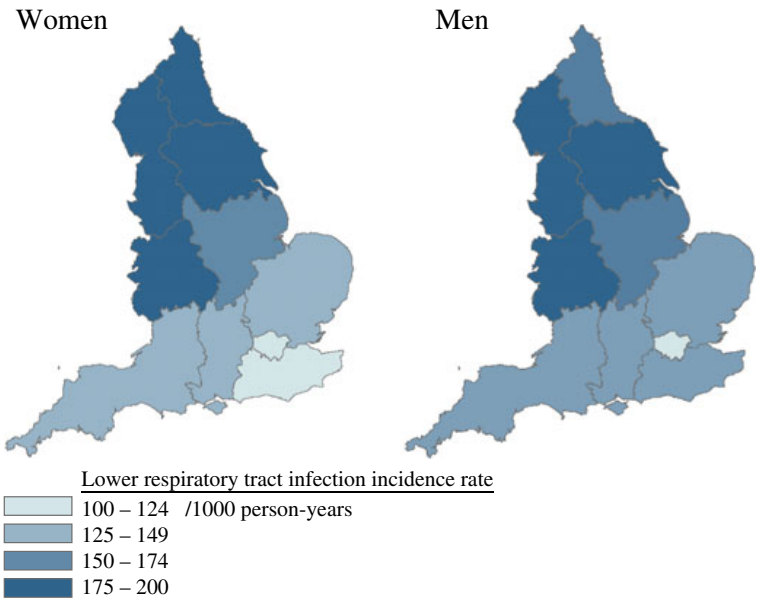


FIGURE 2 Age-standardized lower respiratory tract infection rates by region among older people with diabetes (patients eligible for Hospital Episode Statistics data linkage, N = 128 373). Boundary data provided through EDINA UKBORDERS with the support of the Economic and Social Research Council (ESRC) and Jisc and is copyright of the Crown.

74.5/1000 person-years [95% CI 69.6–79.4]; women 176.9/1000 person-years [95% CI 169.7–184.1]) and the North East for both sexes (Table S3). Incidence rates by region are not presented for sepsis and pneumonia because of small numbers of events. Regional variation in lower respiratory tract infection incidence was similar for men and women.

Hospitalizations

The proportion of infections that resulted in hospitalization within the subsequent 4 weeks were highest for pneumonia and sepsis and lowest for lower respiratory tract infections as a whole (including pneumonia) and urinary tract infections (Table 3); however, the number of hospitalizations after an urinary tract infections diagnosis exceeded the number after pneumonia and sepsis combined, for both same-day and 28-day hospitalizations. Most patients who were hospitalized for pneumonia and sepsis were admitted on the day of diagnosis. For lower respiratory tract infections and urinary tract infections, a large proportion of hospitalizations within 4 weeks were not on the day of infection diagnosis (6588/16 835; 39.1%, and 5159/11 651; 44.3%, respectively), and a notable proportion (1198/16 835; 7.1%, and 596/11 651; 5.1%, respectively) were admitted for a cardiovascular condition (International Classification of Diseases codes 10, chapter I *'Diseases of the circulatory system'*).

Mortality

The case-fatality rate after pneumonia (32.1%) was similar to that after sepsis [31.7% (Table 3)]. Although the 28-day case-fatality rate after urinary tract infection was lower (1.6%), the absolute number of deaths in the 28-days after diagnosis of urinary tract infections (n = 1472) was still high compared with those after sepsis (n = 780) because of the higher incidence rate of urinary tract infections. The case-fatality rate was similar for men and women for each infection.

Discussion

To our knowledge this is the first large cohort study to give detailed estimates of community-acquired infection rates among older people with diabetes mellitus, including infections managed in primary care. There is a high burden of community-acquired infection among older people with diabetes, lower respiratory tract infections having the highest incidence, followed by urinary tract infections. The incidence of all infections increased with age (particularly pneumonia) and increased over the study period; our age-standardization analyses suggest that this increasing trend was driven by the changing age structure of the population. Regional variation in age-standardized rates could be attributable to the prevalence of risk factors for infection such as socio-economic status, smoking, overweight and obesity, or diabetes control.

Table 3 Infection incidence and 28-day all-cause hospitalization and mortality after infection onset among older people with diabetes (N = 218 805)

		Number of infections	Crude incidence rate /1000 person-years (95% CI)	*P***	28-day all-cause mortality		All-cause hospitalization*					
							Eligible infections	On day of onset		≤28 days after onset		
					n	%	*n*	*n*	%	*n*	%	
Lower respiratory tract infection	Men	68 183	148.6 (146.7–150.5)	<0.001	2831	4.2						
	Women	71 118	156.8 (154.8–158.9)		2854	4.0						
	Total	139 301	152.7 (151.3–154.1)		5685	4.1	83 501	10 247	12.3	16 835	20.1	
Pneumonia	Men	5095	10.7 (10.4–11.1)	0.001	1558	30.6						
	Women	4602	9.9 (9.5–10.2)		1557	33.8						
	Total	9697	10.3 (10.0–10.5)		3115	32.1	6802	5297	77.8	5534	81.4	
Urinary tract infection	Men	24 180	51.4 (50.49–52.40)	<0.001	582	2.4						
	Women	67 394	147.9 (145.7–150.2)		890	1.3						
	Total	91 574	99.6 (98.4–100.8)		1472	1.6	57 683	6492	11.3	11 651	20.2	
Sepsis	Men	1246	2.53 (2.38–2.68)	0.75	374	30.0						
	Women	1215	2.49 (2.34–2.64)		406	33.4						
	Total	2461	2.51 (2.40–2.62)		780	31.7	1407	1057	75.1	1141	81.1	

*Among patients eligible for Hospital Episode Statistics data linkage (N = 128 373).
**Likelihood ratio test for difference in crude incidence rate by gender.

For lower respiratory tract infections and urinary tract infections, it is interesting that a high proportion of hospitalizations within 28 days were not on the day of diagnosis. This could reflect high underlying hospitalization rates in this cohort, or could be attributable to exacerbation of underlying comorbidities by acute infections. The considerable number of hospitalizations within 4 weeks of lower respiratory tract infections and urinary tract infections with a cardiovascular cause of admission (International Classification of Diseases codes 10 chapter I '*Diseases of the circulatory system*') are particularly intriguing, as both infections have been found to exacerbate underlying cardiovascular comorbidity [26].

Estimates of infection incidence rates specifically among patients with diabetes are scarce. A cohort study of patients with diabetes of all ages reported higher rates of pneumonia and sepsis than the present study, but included both hospital-acquired and community-acquired infections together [15].

A cohort study in the CPRD of the general population aged ≥65 years (from which we drew our population of patients with diabetes), using the same methodology as the present study, found an incidence of 122.9/1000 person-years for lower respiratory tract infection and 8.0/1000 person-years for pneumonia among the general population aged ≥65 person-years [24]. The crude incidence of lower respiratory tract infection and pneumonia observed in the present study is ~50% higher. This could be consistent with a direct effect of diabetes mellitus on infection incidence, or a higher prevalence of risk factors for infection among patients with diabetes, such as cardiovascular comorbidity or obesity. The rates of pneumonia and urinary tract infections observed in the present study are similarly raised compared with other cohort studies of the general population of older people in the UK/Europe [16,27].

Our sepsis estimate is lower than the rate of community-acquired bloodstream infections in a large cohort study in Canada, which found a rate of 4.5/1000 person-years (95% CI 3.7–5.6) among a selected subset of the general population aged ≥65 years with normal kidney function [28]; however, the Canadian study measured laboratory-defined bacteraemia, while our outcome of interest was clinically diagnosed sepsis.

The regional variation of lower respiratory tract infection incidence within England has a similar pattern to that observed in the general population [29].

The main strengths of the present study are: the large, nationally representative cohort of an older population with diabetes followed up over a prolonged period; the use of primary care medical records to ascertain medically diagnosed community-acquired infections more fully than previous hospital-based studies; detailed definitions of community-acquired infections including distinguishing repeated infection-related consultations within 28 days from recurrent infections; use of CPRD hospitalization codes and

linked hospital data to exclude infections within 14 days of discharge from hospital and to remove hospitalization periods from time at risk for a more accurate estimation of time at risk than previous studies; and the range of infections considered.

The study methods were designed to produce conservative estimates of incidence rates, as follows. We used strict criteria to identify and exclude possible hospital-associated infections; we did not count diagnoses as new episodes of infection if the record fell within 28 days of a previous code for the same infection; patients without Hospital Episodes Statistics data linkage had infections excluded from incidence rates using CPRD indicators of possible hospital-acquired provenance (such as postoperative infection codes) without removal of hospitalized time from time at risk.

Potential limitations include secular changes in management and diagnosis, but in the present study we saw no evidence of this. We could not remove hospitalizations from person-time at risk for patients with no Hospital Episodes Statistics data linkage, which will have led to a slight underestimate of infection rates.

Conclusions

The present study quantifies the high risk of community-acquired infection among older people with diabetes, and the proportion of patients who are admitted to hospital or die within 4 weeks of infection onset. This will facilitate discussions about risk of infections among older patients with diabetes. Knowledge of regional variations and the steep increase in risk with age among older people may assist with designing effective preventive care strategies. Healthcare planners should consider the high infection incidence in primary care and the proportion and pattern of 28-day hospital admission in planning future healthcare provision for this large and growing section of the UK population and of healthcare users.

Future research should clarify the risk factors for infection incidence, hospitalization and mortality, in particular any modifiable risk or protective factors, among this growing population.

Funding sources

The present work was supported by the National Institute for Health Research [CDF 2010-03-32 to SLT]; and Kidney Research UK [ST2/2011 to HIM]. The views expressed in this publication are those of the authors and not necessarily those of the UK National Health Service, the National Institute for Health Research or the Department of Health. The financial supporters of the study had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

None declared.

Acknowledgements

We thank Dr Jennifer Quint, who developed the code lists used for lower respiratory tract infection and pneumonia together with S.L.T.

References

1 Ruben FL, Dearwater SR, Norden CW, Kuller LH, Gartner K, Shalley A *et al.* Clinical infections in the noninstitutionalized geriatric age group: methods utilized and incidence of infections. The Pittsburgh Good Health Study. *Am J Epidemiol* 1995; **141**: 145–157.
2 Laupland KB. Gregson DB, Flemons WW, Hawkins D, Ross T, Church DL. Burden of community-onset bloodstream infection: a population-based assessment. *Epidemiol Infect* 2007; **135**: 1037–1042.
3 Laupland KB, Ross T, Pitout JD, Church DL, Gregson DB. Community-onset urinary tract infections: a population-based assessment. *Infection* 2007; **35**: 150–153.
4 Kaplan V. Angus DC, Griffin MF, Clermont G, Scott Watson R, Linde-Zwirble WT. Hospitalized community-acquired pneumonia in the elderly: age- and sex-related patterns of care and outcome in the United States. *Am J Respir Crit Care Med* 2002; **165**: 766–772.
5 Ruth K, Verne J. *Deaths in older adults in England*. 2010, National End of Life Care Intelligence Network, Bristol, UK. Available at http://www.endoflifecare-intelligence.org.uk/resources/publications/deaths_in_older_adults Last accessed 6 January 2014.
6 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.
7 Trotter CL, Stuart JM, George R, Miller E. Increasing hospital admissions for pneumonia. *England. Emerg Infect Dis* 2008; **14**: 727–733.
8 Kornum JB, Thomsen RW, Riis A, Lervang HH, Schonheyder HC, Sorensen HT. Diabetes, glycemic control, and risk of hospitalization with pneumonia: a population-based case-control study. *Diabetes Care* 2008; **31**: 1541–1545.
9 Joshi N, Caputo GM, Weitekamp MR, Karchmer AW. Infections in patients with diabetes mellitus. *N Engl J Med* 1999; **341**: 1906–1912.
10 Seshasai SR, Kaptoge S, Thompson A, Di Angelantonio E, Gao P, Sarwar N *et al.* Diabetes mellitus, fasting glucose, and risk of cause-specific death. *N Engl J Med* 2011; **364**: 829–841.
11 Seminog OO, Goldacre MJ. Risk of pneumonia and pneumococcal disease in people hospitalized with diabetes mellitus: English record-linkage studies. *Diabet Med* 2013; **30**: 112–119.
12 Vardakas KZ, Siempos II, Falagas ME. Diabetes mellitus as a risk factor for nosocomial pneumonia and associated mortality. *Diabet Med* 2007; **24**: 1168–1171.
13 Knapp S. Diabetes and Infection: Is There a Link? - A Mini-Review. *Gerontology* 2012; **59**: 99–104.
14 Holman N, Forouhi NG, Goyder E, Wild SH. The Association of Public Health Observatories (APHO) Diabetes Prevalence Model: estimates of total diabetes prevalence for England, 2010-2030. *Diabet Med* 2011; **28**: 575–582.
15 Shah BR, Hux JE. Quantifying the risk of infectious diseases for people with diabetes. *Diabetes Care* 2003; **26**: 510–513.

16 Caljouw MA, den Elzen WP, Cools HJ, Gussekloo J. Predictive factors of urinary tract infections among the oldest old in the general population. A population-based prospective follow-up study. *BMC Med* 2011; **9**: 57.
17 Sliedrecht A, den Elzen WP, Verheij TJ, Westendorp RG, 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**: 817–822.
18 CPRD. *The Clinical Practice Research Datalink - CPRD* 2012, 30 August 2012; Available from: http://www.cprd.com/intro.asp Last accessed 6 January 2014.
19 Walley T, Mantgani A. The UK General Practice Research Database. *Lancet* 1997; **350**: 1097–1099.
20 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.
21 The Health and Social Care Information Centre. *HESonline* 2012, 14 September 2012; Available from: www.hesonline.nhs.uk Last accessed 6 January 2014.
22 Office for National Statistics. *Office for National Statistics* 2012, 14 September 2012; Available from: www.ons.gov.uk Last accessed 6 January 2014.
23 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**: 443–451.
24 Millett ER, 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.
25 Pebody RG, Begum F, Gates P, Noakes K, Salisbury D. Influenza vaccination coverage in England, 2000-2008. *Euro Surveill* 2008; **13**: 19074.
26 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**: 2611–2618.
27 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**: 709–716.
28 James MT, Laupland KB, Tonelli M, Manns BJ, Culleton BF, Hemmelgarn BR. Risk of bloodstream infection in patients with chronic kidney disease not treated with dialysis. *Arch Intern Med* 2008; **168**: 2333–2339.
29 Davies S. *Annual Report of the Chief Medical Officer, Volume One, 2011, On the State of the Public's Health*. Department of Health: London, 2012.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Diabetes mellitus 'defining' Read codes used alone to define eligibility for study.
Table S2. Diabetes mellitus 'possible eligibility' Read codes used to define eligibility for study only in combination with a history of an antidiabetes medication.

7.3 Discussion of the likely impact of incomplete data linkages on infection incidence estimates

This section presents a more detailed discussion of the likely impact of incomplete data linkage on infection incidence estimates. Additional results referred to in this discussion are presented in **7.3.3**.

7.3.1 What is the likely impact of incomplete CPRD-HES linkage availability?

HES linkage was available for 128,373/218,805 patients (58.7%). Practices in Northern Ireland, Scotland and Wales, which were not eligible for HES-linkage, contributed 50,910 patients to the study population (23.3%). HES-linkage availability varied by region within England: only 57.5% of patients in CPRD within the East Midlands region had HES-linkage available, compared to 89.3% of patients in CPRD in the South West. Other than regional differences, patient characteristics were similar for patients with and without HES-linkage (**Table 7.1**).

Time in hospital was removed from time at risk of community-acquired infection for patients with HES linkage, but this was not possible for patients without HES linkage. Thus, for patients without HES linkage, infection incidence rates are an under-estimation. The extent of this under-estimation will have been greater for patient groups who spent more time in hospital, such as older patients. Thus, the relationship between infection incidence and age may be even steeper than we observed.

The difference in HES-linkage by region could result in greater under-estimation of infection incidence for regions with lower availability of HES-linkage. **Table 7.2** presents region-specific estimates of infection incidence with 95% confidence intervals, to allow better assessment of the relationship between infection incidence rates by region than possible from the journal figure. There is no clear relationship between availability of HES linkage and region-specific infection incidence. For example, the East Midlands region has the lowest availability of HES-linkage, but is ranked 5th of 10 regions in terms of LRTI infection incidence for men and women. In addition, the regional variation in LRTI incidence we observed among older people with diabetes mirrors regional variation in primary care consultations for LRTI among the general population using Royal College of General Practitioners data.[138] This suggests that bias from varying HES-linkage availability is not sufficiently large to grossly distort estimates of regional variation in

infection incidence: the overall pattern is likely to be real, but region-specific incidence rates should be interpreted with caution.

The proportion of patients admitted to hospital in the 28 days following infection onset was described only among patients with HES linkage available. Any differences in hospital admission thresholds between patients with and without HES linkage available could limit generalisability for patients without HES linkage. There may be differences in clinical practice resulting in different hospital admission thresholds between England and Scotland, Wales, and Northern Ireland for example, due to different provision of care in the community. However, hospital admission thresholds are unlikely to differ for patients with and without HES linkage within England, so these results are likely to be generalisable across England, rather than only to patients with HES linkage.

7.3.2 What is the likely impact of incomplete CPRD-ONS linkage?

A date of death was available from linked ONS records for 38,430 patients, of whom 23,707 (61.7%) had a death recorded on the same day in CPRD (**Table 7.3**). Dates of death recorded in both CPRD and ONS showed good agreement, suggesting that CPRD death dates are reasonably accurate. Among the 7,600 patients with different dates of death recorded in CPRD and ONS datasets, 1934 (25.4%) were one day apart, and 6779 (89.2%) were within 28 days of each other. As CPRD death dates were used for patients without a death recorded in ONS, incomplete CPRD-ONS linkage may have resulted in under-estimation of post-infection mortality. However, this appears to be a limited concern, as a high proportion of patients with a death recorded in ONS also had a death recorded in CPRD.

7.3.3 Additional results comparing the population with and without available data linkage

The additional results presented here are discussed in **7.3.1** and **7.3.2**.

Table 7.1: Baseline characteristics of patients with and without HES linkage

		Percentage of patients with HES linkage	Characteristics of patients with HES linkage median (IQR)	Characteristics of patients without HES linkage median (IQR)
Age (years)		N/A	71 (65–77)	71 (65–77)
Time in study (years)		N/A	4.0 (1.8–7.3)	3.8 (1.8– 6.8)
		row %	n (column %)	n (column %)
Gender	Female	58.6	63,295 (49.3)	44,775 (49.5)
Socio-economic status [1]	1 (least deprived)	50.5	19,236 (15.0)	18,853 (20.9)
	2	65.6	25,716 (20.0)	13,482 (14.9)
	3	58.8	26,664 (20.8)	18,670 (20.7)
	4	58.8	29,391 (22.9)	20,632 (22.8)
	5 (most deprived)	59.3	27,366 (21.3)	18,795 (20.8)
Region	North East	73.0	2,800 (2.2)	1,037 (1.2)
	North West	83.5	22,712 (17.7)	4,479 (5.0)
	Yorkshire & The Humber	68.6	6,793 (5.3)	3,111 (3.4)
	East Midlands	57.7	5,168 (4.0)	3,785 (4.2)
	West Midlands	80.5	15,403 (12.0)	3,740 (4.1)
	East of England	79.3	15,455 (12.0)	4,042 (4.5)
	South West	89.3	17,075 (13.3)	2,046 (2.3)
	South Central	66.9	13,566 (10.6)	6,713 (7.4)
	London	71.4	15,694 (12.2)	6,284 (7.0)
	South East Coast	76.2	13,707 (10.7)	4,285 (4.7)
	Northern Ireland	–	–	7,601 (8.4)
	Scotland	–	–	19,248 (21.3)
	Wales	–	–	24,061(26.6)
Body mass index (BMI)	<18.5	57.4	1,354 (1.1)	1,006 (1.1)
	18.5–24.9	59.3	26,565 (20.7)	18,266 (20.2)
	25–29.9	58.9	47,822 (37.3)	33,375 (36.9)
	30–34.9	58.1	29,245 (22.8)	21,110 (23.3)
	35+	58.5	15,298 (11.9)	10,835 (12.0)
	Missing	58.1	8,089 (6.3)	5,840 (6.5)
Smoking status	Non-smoker	58.0	48,470 (37.8)	35,133 (38.9)
	Current	57.0	16,644 (13.0)	12,548 (13.9)
	Ex-smoker	59.8	60,635 (47.2)	40,779 (45.1)
	Unknown	57.1	2,624 (2.0)	1,972 (2.2)
Co-morbidities	Cardiovascular disease	57.3	45,185 (35.2)	33,721 (37.3)
	Chronic lung disease	57.1	9,351 (7.3)	7,019 (7.8)
	Peripheral vascular disease	56.4	8,787 (6.8)	6,785 (7.5)
Total patients		58.7	128,373	90,432

HES, Hospital Episode Statistics

1. practice-level index of multiple deprivation quintile

Table 7.2: Age-standardised infection incidence rates by region (n=218,805)

Region	n	LRTI /1,000 years (95% CI)		UTI /1,000 years (95% CI)	
		Female	Male	Female	Male
North East	3,837	177.5 (163.3-191.7)	153.9 (140.2-167.6)	182.4 (166.1–198.8)	66.9 (57.0–76.7)
North West	27,191	199.5 (193.6-205.4)	185.1 (179.1-191.1)	151.9 (146.6–157.2)	61.0 (57.6–64.5)
Yorkshire & the Humber	9,904	190.8 (180.9-200.7)	183.6 (173.6-193.6)	131.7 (123.5–139.9)	53.5 (48.3–58.7)
East Midlands	8,953	152.7 (144.4-161.0)	153.8 (144.6-163.0)	143.2 (143.4–152.0)	55.8 (50.2–61.4)
West Midlands	19,143	190.4 (183.6-197.3)	194.5 (187.1-201.8)	176.9 (169.7–184.1)	74.5 (69.6–79.4)
East of England	19,497	143.0 (137.5-148.5)	149.9 (144.0-155.7)	143.7 (137.7–149.8)	59.8 (55.9–63.7)
South West	19,121	133.7 (128.3-139.0)	136.6 (131.2-142.0)	149.8 (143.5–156.2)	56.9 (53.2–60.7)
South Central	20,279	134.1 (129.0-139.2)	138.0 (132.6-143.3)	161.2 (154.7–167.8)	62.2 (58.3–66.2)
London	21,978	114.6 (110.3-119.0)	116.8 (112. 1-121.6)	131.2 (125.9–136.6)	56.8 (53.0–60.6)
South East Coast	17,992	122.6 (117.5-127.7)	129.3 (123.9-134.7)	134.3 (128.3–140.3)	51.9 (48.2–55.6)

LRTI, lower respiratory tract infection: UTI, urinary tract infection: 95% CI, 95% confidence interval

Table 7.3: Deaths recorded in CPRD and ONS for patients aged ≥65 with diabetes mellitus (n=218,805)

Source of date of death record	Number (%)
CPRD only	31,613 (45.1)
ONS only	7,123 (10.2)
CPRD and ONS coincident dates	23,707 (33.8)
CPRD and ONS discrepant dates	7,600 (10.9)
Total	70,043

CPRD, Clinical Practice Research Datalink; ONS, Office for National Statistics

Chapter 8. The association of chronic kidney disease with incidence of acute, community-acquired infection

This chapter presents a study of the association of markers of chronic kidney disease (CKD) with incidence of acute, community-acquired infections (objective 3). The results are presented and discussed in the journal article, and additional discussion follows in **8.3** and **8.4**. The full results of all sensitivity analyses referred to are available in **Appendix E.**

8.1 Introduction to Paper 3

This paper presents a retrospective cohort study to identify incidence rate ratios of infection according to markers of CKD using primary care records from the Clinical Practice Research Datalink linked to Hospital Episode Statistics admissions data. The study population comprised 191,672 patients aged ≥65 years with diabetes mellitus and a valid serum creatinine result during follow-up, with no history of renal replacement therapy (study population B), identified as described in **3.9.2**.

The outcomes were acute, community-acquired lower respiratory tract infection (LRTI), pneumonia (as a subset of LRTI) and sepsis. Methods used to identify community-acquired infection incidence were detailed in **Chapter 4**.

The exposure of interest was CKD, identified by a history of proteinuria or reduced estimated glomerular filtration rate (eGFR) as described in **Chapter 5**.

Incidence rate ratios were adjusted for *a priori* potential confounders of the association between markers of CKD and infection as described in the article. Detailed definitions of these confounders are described in **Chapter 6**. The interpretation of adjusted analyses are discussed in the article, and further detailed discussion follows in **8.3.4**.

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	Helen McDonald
Principal Supervisor	Dr Dorothea Nitsch
Thesis Title	The epidemiology of infections among older people with diabetes mellitus and chronic kidney disease

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?	American Journal of Kidney Diseases		
When was the work published?	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)	The original idea for this study was conceived by Dorothea Nitsch and Sara Thomas, who wrote the study proposal and obtained ethics approval and funding for the study. Sara Thomas obtained the study data from CPRD. I developed a detailed study design from the original study proposal, with advice and supervision from Dorothea Nitsch

	and Sara Thomas. This study used the dataset of community-acquired infections among study population B which I had constructed for objective 2. Episodes of community-acquired infection were defined using methods developed by Elizabeth Millett (Chapter 4). I developed an approach to identifying markers of chronic kidney disease in CPRD (described in Chapter 5), with advice and supervision from Dorothea Nitsch and Sara Thomas. I extracted, cleaned and classified the data to define other variables as described in Chapter 6, and was fortunate to be able to use codelists developed by colleagues at LSHTM (each attributed individually in Chapter 6). I extracted, cleaned and classified body mass index data using a do-file written by Krishnan Bhaskaran. I conducted all data analysis, and drafted the manuscript, which was then commented on by all co-authors. The manuscript was peer-reviewed, and I incorporated suggestions from reviewers into the final manuscript.

Student Signature: K McDonald **Date:** 10 April 2015

Supervisor Signature: Dorothea **Date:** 14/4/15

AJKD
Original Investigation

CKD and the Risk of Acute, Community-Acquired Infections Among Older People With Diabetes Mellitus: A Retrospective Cohort Study Using Electronic Health Records

Helen I. McDonald, MSc,[1] Sara L. Thomas, PhD,[2] Elizabeth R.C. Millett, MSc,[2] and Dorothea Nitsch, MD[1]

Background: Hospital admissions for community-acquired infection are increasing rapidly in the United Kingdom, particularly among older individuals, possibly reflecting an increasing prevalence of comorbid conditions such as chronic kidney disease (CKD). This study describes associations between CKD (excluding patients treated by dialysis or transplantation) and community-acquired infection incidence among older people with diabetes mellitus.

Study Design: Retrospective cohort study using primary care records from the Clinical Practice Research Datalink linked to Hospital Episode Statistics admissions data.

Setting & Participants: 191,709 patients 65 years or older with diabetes mellitus and no history of renal replacement therapy, United Kingdom, 1997 to 2011.

Predictor: Estimated glomerular filtration rate (eGFR) and history of proteinuria.

Outcomes: Incidence of community-acquired lower respiratory tract infections (LRTIs, with pneumonia as a subset) and sepsis, diagnosed in primary or secondary care, excluding hospital admissions from time at risk.

Measurements: Poisson regression was used to calculate incidence rate ratios (IRRs) adjusted for age, sex, smoking status, comorbid conditions, and characteristics of diabetes. Estimates for associations of eGFR with infection were adjusted for proteinuria, and vice versa.

Results: Strong graded associations between lower eGFRs and infection were observed. Compared with patients with eGFRs $\geq$ 60 mL/min/1.73 m^2, fully adjusted IRRs for pneumonia among those with eGFRs $<$ 15, 15 to 29, 30 to 44, and 45 to 59 mL/min/1.73 m^2 were 3.04 (95% CI, 2.42-3.83), 1.73 (95% CI, 1.57-1.92), 1.19 (95% CI, 1.11-1.28), and 0.95 (95% CI, 0.89-1.01), respectively. Associations between lower eGFRs and sepsis were stronger, with fully adjusted IRRs up to 5.56 (95% CI, 3.90-7.94). Those associations with LRTI were weaker but still clinically relevant at up to 1.47 (95% CI, 1.34-1.62). In fully adjusted models, a history of proteinuria remained an independent marker of increased infection risk for LRTI, pneumonia, and sepsis (IRRs of 1.07 [95% CI, 1.05-1.09], 1.26 [95% CI, 1.19-1.33], and 1.33 [95% CI, 1.20-1.47]).

Limitations: Patients without creatinine results were excluded.

Conclusions: Strategies to prevent infection among people with CKD are needed.

INDEX WORDS: Community-acquired infections; lower respiratory tract infections (LRTIs); pneumonia; sepsis; non–dialysis-dependent chronic kidney disease (CKD); decreased renal function; estimated glomerular filtration rate (eGFR); proteinuria; diabetes mellitus; aged; elderly; electronic health records.

Hospitalization rates for infections in the United Kingdom are increasing rapidly, particularly among older individuals: age-standardized hospital admission rates for community-acquired pneumonia more than doubled between 2000 and 2010.[1,2] The driving factors behind this increase are unclear, but a higher prevalence of comorbid conditions in the aging population has been suggested.[1,2]

One comorbid condition associated with hospitalization for infection is chronic kidney disease (CKD). Patients receiving renal replacement therapy may be at increased infection risk due to their treatment. This study focuses on patients with CKD not treated by dialysis or transplantation, which will be referred to as CKD. A graded association between increasing severity of CKD and higher risk of hospitalization with pneumonia and sepsis has been reported, even at early stages of CKD.[3-5] These studies identified CKD by reduced estimated glomerular filtration rate (eGFR). Proteinuria also has been found to indicate an

From the Departments of [1]Non-communicable Disease Epidemiology and [2]Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, United Kingdom.

Received July 7, 2014. Accepted in revised form December 2, 2014.

Address correspondence to Helen I. McDonald, MSc, Department of Non-communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, Keppel Street, London, WC1E 7HT, United Kingdom. E-mail: helen.mcdonald@lshtm.ac.uk

0272-6386
http://dx.doi.org/10.1053/j.ajkd.2014.11.027

increased risk of infection-related hospitalization among patients with diabetes.[6,7] CKD is a risk factor for poor prognosis from infection, so this could be driven by a higher chance of hospital admission for patients with community-acquired infection if they have CKD.[8] It is unclear whether CKD is a risk factor for higher incidence of infection in the community. One large case-control study identified CKD as a risk factor for incidence of community-acquired pneumonia in primary care, but relied on routine diagnosis of CKD in the general population and did not exclude patients receiving renal replacement therapy.[9]

There is a high prevalence of CKD among people with diabetes, particularly older people.[10-12] Patients with diabetes are monitored regularly in primary care for CKD, and this has been financially incentivized in the United Kingdom since 2004.[13,14] Thus, studying people with diabetes minimizes the potential for ascertainment bias in estimating the association between CKD and infection from routinely collected electronic health records. The subset of the UK population with diabetes mellitus and aged 65 years or older is large and growing and experiences a high burden of infection.[15,16] This population also is at higher risk of infectious complications such as acute kidney injury.[17] If CKD is a risk factor for infection incidence among older people with diabetes, this could be important to health service planning, as well as to patients and their clinicians.

We aimed to describe, among older people with diabetes, the associations between CKD (excluding patients with a history of renal replacement therapy) and community-acquired lower respiratory tract infection (LRTI), pneumonia (as a subset of LRTI), and sepsis. We used linked health records to identify infections managed in primary or secondary care.

METHODS

Data Sources

We used the May 2011 data set of the Clinical Practice Research Datalink (CPRD), a database of anonymized primary care medical records comprising 12.8 million patient records at 627 practices in the United Kingdom.[18] Data include patient demographics, health behaviors, test results, diagnoses, and prescriptions. The CPRD population is representative of the general UK population and the validity of recorded diagnoses generally is high.[19,20] Linked data are available for patients registered at consenting English practices. For linked patients, this study used linked data for all hospital inpatient admissions to National Health Service hospitals in England from Hospital Episodes Statistics (HES) and socioeconomic status from the Office for National Statistics.[21,22]

Study Population and Follow-up

All patients in CPRD at any point between April 1997 and March 2011 with diabetes mellitus, aged 65 years or older, with at least one valid serum creatinine result during the study period, and with no history of renal replacement therapy were eligible. The definition of diabetes was based on diagnostic Read codes. "Definite" codes, for example, C10F.00 Type 2 diabetes mellitus, were sufficient evidence of diabetes. "Possible" codes, for example, 90LA.11 Diabetes monitored, required an antidiabetes medication prescription for confirmation. Full code lists were published previously.[15] Patients met eligibility criteria at the latest date of diabetes diagnosis, 65th birthday, 1 year after practice registration, practice fulfilling CPRD quality control standards, or April 1, 1997. Their study entry date was their first valid serum creatinine result after the eligibility criteria were met. Patients left the study at the earliest date of death, leaving the practice, last data collection from the practice, renal replacement therapy (dialysis or kidney transplantation), or March 31, 2011.

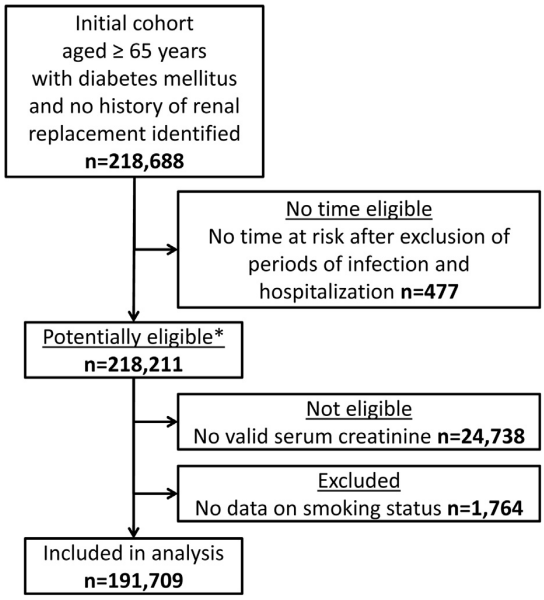


Figure 1. Flowchart of study eligibility and participation. *Baseline characteristics described in Table 1.

Definition of CKD

We estimated glomerular filtration rate from primary care serum creatinine test results, multiplied by 0.95 to correct for lack of isotope-dilution mass spectrometry standardization, using the CKD-EPI (CKD Epidemiology Collaboration) creatinine equation.[23-25] We included adjustment for black ethnicity.[26] To reduce misclassification of eGFR from variability in serum creatinine results or acute illness, we used a last-carried-forward method, with eGFR initially defined using the creatinine result that marked entry to the study and updated at each subsequent creatinine result so that eGFR was always defined by the single most recent creatinine result, as previously performed by James et al.[3,27] For our main analyses, we used eGFR categories corresponding to those used in diagnosis (<15, 15-29, 30-44, 45-59, and ≥60 mL/min/1.73 m^2).[25] During the study period, many UK laboratories did not report the specific value of eGFR results if they were ≥60 mL/min/1.73 m^2 and so we did not distinguish eGFR categories > 60 mL/min/1.73 m^2 in the main analysis. We repeated the final model including additional separate categories for eGFR of 60 to 74, 75 to 89, and ≥90 mL/min/1.73 m^2, with patients with eGFRs of 75 to 89 mL/min/1.73 m^2 as a reference group.

Either a positive urine protein test result or a diagnosis of proteinuric kidney disease in CRPD defined onset of a history of proteinuria. We excluded urine protein test results that occurred on the same day as a diagnostic Read code for urinary tract infection.

Table 1. Baseline Characteristics of Potentially Eligible Study Population by Baseline eGFR

	No Scr (n = 24,738)[a]	eGFR ≥ 60 (n = 124,521)[b]	eGFR < 60 (n = 68,952)[b]
Female sex	12,662 (51.2)	54,246 (43.6)	40,907 (59.3)
Age category			
65-69 y	7,356 (29.7)	63,364 (50.9)	15,849 (23.0)
70-74 y	4,852 (19.6)	28,549 (22.9)	13,668 (19.8)
75-79 y	4,570 (18.5)	18,689 (15.0)	15,610 (22.6)
80-84 y	3,843 (15.5)	9,297 (7.5)	12,816 (18.6)
≥85 y	4,117 (16.6)	4,622 (3.7)	11,009 (16.0)
SES[c] by practice			
1: least deprived	3,839 (15.5)	21,818 (17.5)	12,341 (17.9)
2	4,269 (17.3)	22,312 (17.9)	12,515 (18.2)
3	5,101 (20.6)	26,136 (21.0)	13,962 (20.3)
4	6,037 (24.4)	28,355 (22.8)	15,500 (22.5)
5: most deprived	5,492 (22.2)	25,900 (20.8)	14,634 (21.2)
Smoking status			
Current smoker	3,976 (16.1)	21,398 (17.2)	9,127 (13.2)
Ex-smoker	7,034 (28.4)	51,901 (41.7)	25,991 (37.7)
Nonsmoker	10,968 (44.3)	50,551 (40.6)	32,741 (47.5)
Missing	2,760 (11.2)	671 (0.5)	1,093 (1.6)
Comorbid conditions			
Ischemic heart disease	6,381 (25.8)	30,743 (24.7)	23,308 (33.8)
Congestive heart failure	3,108 (12.6)	6,221 (5.0)	10,122 (14.7)
Hypertension	12,229 (49.4)	73,263 (58.8)	45,915 (66.6)
Cerebrovascular disease	4,169 (16.9)	13,157 (10.6)	11,500 (16.7)
Other dementia	1,350 (5.5)	1,651 (1.3)	1,889 (2.7)
Chronic lung disease	2,344 (9.5)	9,266 (7.4)	5,847 (8.5)
Antidiabetes medications			
Insulin only	1,407 (5.7)	3,936 (3.2)	3,346 (4.9)
Oral medication only	9,811 (39.7)	54,635 (43.9)	27,169 (39.4)
Both insulin and oral	1,473 (5.6)	8,078 (6.5)	5,764 (8.4)
None recorded	12,047 (48.7)	57,872 (46.5)	32,673 (47.4)
Hemoglobin A_{1c}[d]			
Good	7,256 (29.3)	58,177 (46.7)	31,026 (45.0)
Borderline	5,185 (21.0)	46,122 (37.0)	24,157 (35.0)
Poor	1,431 (5.8)	7,433 (6.0)	4,142 (6.0)
None recorded	10,866 (43.9)	12,789 (10.3)	9,627 (14.0)

Note: N = 218,211. Values are given as number (percentage). Baseline is date of study entry for study participants (n = 191,709) and date of eligibility for study entry for patients not included in the study due to having no available Scr result or no available smoking status. eGFRs expressed in mL/min/1.73 m^2.

Abbreviations: eGFR, estimated glomerular filtration rate; Scr, serum creatinine; SES, socioeconomic status.

[a]These patients had no Scr result available and hence were not included in study.

[b]These patients had an Scr result available and were included in study unless smoking status was missing.

[c]Index of multiple deprivation.

[d]Good, <53 mmol/mol (<7%); borderline, 53-86 mmol/mol (7%-10%); poor, >86 mmol/mol (>10%).

We did not count trace results as positive and checked records for internal consistency.

Definition of Infections

We studied 3 acute community-acquired infections: LRTI (which included diagnoses such as influenza and acute bronchitis), pneumonia (as a subset of LRTI), and sepsis. Either diagnostic Read codes in CPRD or any *International Classification of Diseases, Tenth Revision* code that formed the primary diagnostic code on hospital admission in HES could define an infection. To avoid overestimation from repeat attendances for the same infection, diagnostic codes recorded within 28 days of one another were attributed to a single episode of infection, with index date defined by the first diagnostic code and duration until 28 days after the last diagnostic code. If any LRTI included a pneumonia code, the pneumonia episode was considered to start from the first instance of the pneumonia code and end on the end date of the LRTI within which it occurred. Any infection with onset date during an HES hospitalization spell, within 14 days after hospital discharge, or that included a code for postoperative infection was identified as a hospital-acquired infection and excluded. These methods were described in detail previously.[15,28]

Time at Risk

Patients were not at risk of a community-acquired infection during an infection (community or hospital acquired), during an HES hospitalization spell, or within 14 days following hospital discharge, and these periods were removed from time at risk. Time

at risk was calculated separately for each type of infection; a patient could be at risk of sepsis despite an ongoing LRTI, for example.

Definition of Covariates

Age was defined in 5-year age bands up to a final category of 85 years or older. Socioeconomic status was assigned at practice level using 2007 Office for National Statistics estimates of the Index of Multiple Deprivation, a composite area-level marker of deprivation.[22] Smoking status was identified as current, ex-smoker, or nonsmoker from both HES and CPRD data. Smoking cessation products were considered to indicate current smoking because cessation success rates are low.[29] The most recent smoking status record by the study entry date defined smoking status at baseline when available; if not recorded, the first subsequent record defined smoking status at baseline. Comorbid conditions (ischemic heart disease, congestive cardiac failure, hypertension, cerebrovascular disease, other dementia, and chronic lung disease) were defined using diagnostic CPRD Read codes. The first diagnostic record at any point in the patient's records defined onset of the condition. Baseline hemoglobin A_{1c} level was defined by the most recent hemoglobin A_{1c} test result in CPRD prior to (or on) the study entry date. Baseline diabetic medication history was defined using CPRD prescription records.

Data Analysis

Incidence rates were calculated for each infection using Poisson regression with lexis expansions for age and a random-effects model to adjust for multiple infection episodes. Analysis was conducted separately for each type of infection (LRTI, pneumonia, and sepsis) using 3 main regression models.

Negative proteinuria test results tend to be under-recorded in primary care records.[30] For comorbid conditions and proteinuria status, absence of a positive record was treated as absence of disease, and for hemoglobin A_{1c}, absence of a recorded result was included as a category of hemoglobin A_{1c} status. We excluded patients with no smoking status available.

Our first model adjusted for age, sex, socioeconomic status at practice level, and date prior to or post April 1, 2004, when Quality Outcomes Framework guidelines introduced financial incentives for recording CKD status among people with diabetes in primary care that are suggested to have improved ascertainment of CKD in primary care.[31] Our second model additionally adjusted for confounding by smoking status and comorbid conditions. This second model was run both with all variables assessed at baseline and separately with new onset of comorbid condition time updated during the study (but not smoking status because changes in smoking status are particularly vulnerable to reverse causation). Our final model additionally adjusted for hemoglobin A_{1c} level and diabetic medication history at baseline. All nonbinary covariates were modeled as categorical variables.

Sensitivity analyses repeated the final model with the following adjustments: limiting follow-up to post April 1, 2004; restricting the data set to patients with HES linkage available; and using only the first infection as an outcome.

All estimates for the associations of eGFR with infection were adjusted for proteinuria, and vice versa, so that the effect estimates for eGFR and proteinuria are independent.

We looked for evidence of interaction between eGFR and proteinuria and between eGFR and age (65-74 and ≥75 years) in the final models for LRTI and pneumonia using likelihood ratio tests to assess nested models with and without interaction terms. We did not look for interaction in the sepsis regression model due to the smaller number of events.

Stata, version 13.1 (StataCorp LP), was used for data analyses. All code lists are available on request.

Ethics Approval

The study was approved by the Independent Scientific Advisory Committee of the CPRD (ISAC reference 11_033A) and the London School of Hygiene & Tropical Medicine Ethics Committee (LSHTM reference 6116).

RESULTS

Of 218,211 patients potentially eligible for inclusion, 191,709 (87.9%) had a valid serum creatinine result and complete data available (Fig 1). Study participants were followed up for a median of 4.6 (interquartile range [IQR], 2.3-7.6) years. Median age at study entry was 71 (IQR, 66-78) years. For 113,106 study participants (59.0%), HES linkage was available. The population with no available serum creatinine result had a high prevalence of missing data for both smoking status (2,760 of 24,738 [11.2%]) and hemoglobin A_{1c} results (10,866 of 24,739 [43.9%]), suggesting that this population may not attend primary care services frequently. The population with no available creatinine result had a comorbid condition profile similar to patients with CKD stages 3 to 5 in terms of prevalence of congestive heart failure and cerebrovascular disease (Table 1).

We found good completeness of serum creatinine result recording: only 11% of potentially eligible patients lacked a valid serum creatinine result (Fig 1). The median time for which each creatinine result was carried forward was 137 (IQR, 56-242) days. At study entry, 67,859 (35.4%) participants had CKD stages 3 to 5, defined as eGFR $<$ 60 mL/min/1.73 m^2, and 25,433 (13.3%) had a history of proteinuria (Table 2).[25]

We observed 115,080 LRTIs among 56,076 patients, 7,870 episodes of pneumonia among 7,095 patients, and 1,980 episodes of sepsis among 1,902 patients. Crude incidence rates were as follows: LRTI, 155.8 (95% confidence interval [CI], 154.3-157.4)/1,000 person-years; pneumonia, 10.3 (95% CI,

Table 2. Prevalence of Markers of CKD at Baseline for Study Participants

eGFR	Proteinuria Absent	History of Proteinuria	Total
<15	307	234 (43.3)	541
15-29	3,373	1,205 (26.3)	4,578
30-44	14,857	3,417 (18.7)	18,274
45-59	38,672	5,794 (13.0)	44,466
60-74	52,168	6,726 (11.4)	58,894
75-89	41,446	5,379 (11.5)	46,825
≥90	15,453	2,678 (14.8)	18,131
Total	166,276	25,433 (13.3)	191,709

Note: n = 191,709. Markers of CKD are eGFR and history of proteinuria. Values are given as number or number (row percentage). eGFR categories expressed in mL/min/1.73 m^2.

Abbreviations: CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate.

Table 3. Infection Rates With Corresponding Rate Ratios by eGFR

Infection	No. of Events	Time at Risk (person-y)	Crude Rate/1,000 Person-y (95% CI)	Minimally Adjusted[a,b]	Adjusted for Comorbid Conditions at Baseline[b,c]	Adjusted for Time-Updated Comorbid Conditions[b,d]	Adjusted for Characteristics of Diabetes[b,e]
LRTI	115,080	808,194	155.8 (154.3-157.4)				
eGFR < 15	607	2,532	295.3 (265.8-324.8)	1.78 (1.61-1.96)	1.67 (1.51-1.85)	1.52 (1.38-1.68)	1.47 (1.34-1.62)
eGFR 15-29	5,153	25,016	228.0 (219.8-236.2)	1.38 (1.33-1.43)	1.28 (1.23-1.33)	1.20 (1.15-1.24)	1.17 (1.13-1.22)
eGFR 30-44	16,557	96,214	188.7 (184.8-192.6)	1.19 (1.16-1.22)	1.13 (1.10-1.15)	1.09 (1.07-1.12)	1.08 (1.05-1.10)
eGFR 45-59	29,783	204,866	159.9 (157.4-162.4)	1.06 (1.05-1.08)	1.04 (1.03-1.06)	1.03 (1.02-1.05)	1.03 (1.01-1.04)
eGFR ≥ 60	62,980	479,565	143.0 (141.3-144.7)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Pneumonia[f]	7,870	816,517	10.3 (10.1-10.6)				
eGFR < 15	99	2,570	52.8 (40.7-65.0)	4.26 (3.37-5.38)	3.69 (2.92-4.65)	3.25 (2.58-4.10)	3.04 (2.42-3.83)
eGFR 15-29	650	25,362	30.6 (27.8-33.3)	2.29 (2.07-2.53)	2.01 (1.82-2.23)	1.82 (1.65-2.01)	1.73 (1.57-1.92)
eGFR 30-44	1,523	97,360	17.4 (16.4-18.4)	1.42 (1.33-1.53)	1.31 (1.22-1.40)	1.23 (1.15-1.32)	1.19 (1.11-1.28)
eGFR 45-59	1,980	207,025	10.2 (9.7-10.7)	1.01 (0.95-1.07)	0.98 (0.93-1.04)	0.96 (0.91-1.02)	0.95 (0.89-1.01)
eGFR ≥ 60	3,618	484,200	7.8 (7.6-8.1)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sepsis	1,980	816,826	2.5 (2.4-2.6)				
eGFR < 15	41	2,572	17.8 (11.8-23.9)	7.40 (5.19-10.55)	6.93 (4.86-9.90)	6.19 (4.34-8.82)	5.56 (3.90-7.94)
eGFR 15-29	186	25,382	8.1 (6.8-9.3)	3.29 (2.75-3.94)	3.01 (2.52-3.61)	2.69 (2.25-3.23)	2.50 (2.08-3.00)
eGFR 30-44	387	97,416	4.2 (3.7-4.6)	1.80 (1.58-2.06)	1.70 (1.49-1.95)	1.60 (1.39-1.83)	1.51 (1.32-1.73)
eGFR 45-59	499	207,103	2.5 (2.2-2.7)	1.19 (1.06-1.34)	1.17 (1.04-1.32)	1.14 (1.02-1.29)	1.11 (0.99-1.25)
eGFR ≥ 60	867	484,352	1.8 (1.7-2.0)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)

Note: Unless otherwise indicated, values are given as rate ratio (95% CI). eGFR categories expressed in mL/min/1.73 m^2.

Abbreviations: CI, confidence interval; eGFR, estimated glomerular filtration rate; LRTI, lower respiratory tract infection.

[a]Adjusted for proteinuria (updated), age (updated), sex, socioeconomic status by practice, and financial year prior to or post 2004.

[b]$P < 0.001$ for all rate ratios. Likelihood ratio test for inclusion of eGFR as a categorical variable in the model.

[c]Adjusted for proteinuria (updated), age (updated), sex, socioeconomic status by practice, financial year prior to or post 2004, ischemic heart disease, congestive cardiac failure, hypertension, cerebrovascular disease, other dementia, chronic lung disease, and smoking at baseline.

[d]Adjusted for proteinuria (updated), age (updated), sex, socioeconomic status by practice, financial year prior to or post 2004, ischemic heart disease (updated), congestive cardiac failure (updated), hypertension (updated), cerebrovascular disease (updated), other dementia (updated), chronic lung disease (updated), and smoking (baseline).

[e]Adjusted for proteinuria (updated), age (updated), sex, socioeconomic status by practice, financial year prior to or post 2004, ischemic heart disease (updated), congestive cardiac failure (updated), hypertension (updated), cerebrovascular disease (updated), other dementia (updated), chronic lung disease (updated), smoking (baseline), hemoglobin A_{1c} level (baseline), and antidiabetes medication history (baseline).

[f]Subset of LRTI.

10.1-10.6)/1,000 person-years; and sepsis, 2.5 (95% CI, 2.4-2.6)/1,000 person-years (Table 3).

Both eGFR and proteinuria were independent risk markers for incidence of LRTI, pneumonia, and sepsis ($P < 0.001$ for each analysis). A high incidence of infection was observed among patients with CKD. For example, crude LRTI rates were 228.0 (95% CI, 219.8-236.2)/1,000 person-years among patients with eGFRs of 15 to 29 mL/min/1.73 m^2 compared to 143.0 (95% CI, 141.3-144.7)/1,000 person-years among patients with eGFRs ≥ 60 mL/min/1.73 m^2. The association between eGFR and infection incidence was graded, with increased infection incidence even at early stages of CKD. Strong and graded associations between reduced eGFR and infection remained after adjustment for age, sex, smoking status, comorbid conditions, and characteristics of diabetes. Compared to eGFR ≥ 60 mL/min/1.73 m^2, fully adjusted incidence rate ratios (IRRs) for pneumonia were 3.04 (95% CI, 2.42-3.83), 1.73 (95% CI, 1.57-1.92), 1.19 (95% CI, 1.11-1.28), and 0.95 (95% CI, 0.89-1.01) for eGFRs < 15, 15 to 29, 30 to 44, and 45 to 59 mL/min/1.73 m^2, respectively. The associations between reduced eGFR and sepsis were stronger, with fully adjusted IRRs up to 5.56 (95% CI, 3.90-7.94), while those with LRTI were less strong but still clinically relevant, with fully adjusted IRRs up to 1.47 (95% CI, 1.34-1.62; Table 3). Fully adjusted rates and rate ratios using patients with eGFRs of 75 to 89 mL/min/1.73 m^2 as a reference group suggested a J-shaped relationship between eGFR and LRTI and pneumonia incidence (Fig 2; Table S1 [available as online supplementary material]).

Proteinuria was an independent risk marker for infection incidence after adjustment for eGFR (Table 4). In minimally adjusted analyses, patients with a history of proteinuria had a higher incidence of LRTI, pneumonia, and sepsis (IRRs of 1.13 [95% CI, 1.10-1.15], 1.37 [95% CI, 1.30-1.45], and 1.44 [95% CI, 1.30-1.59], respectively) compared with patients

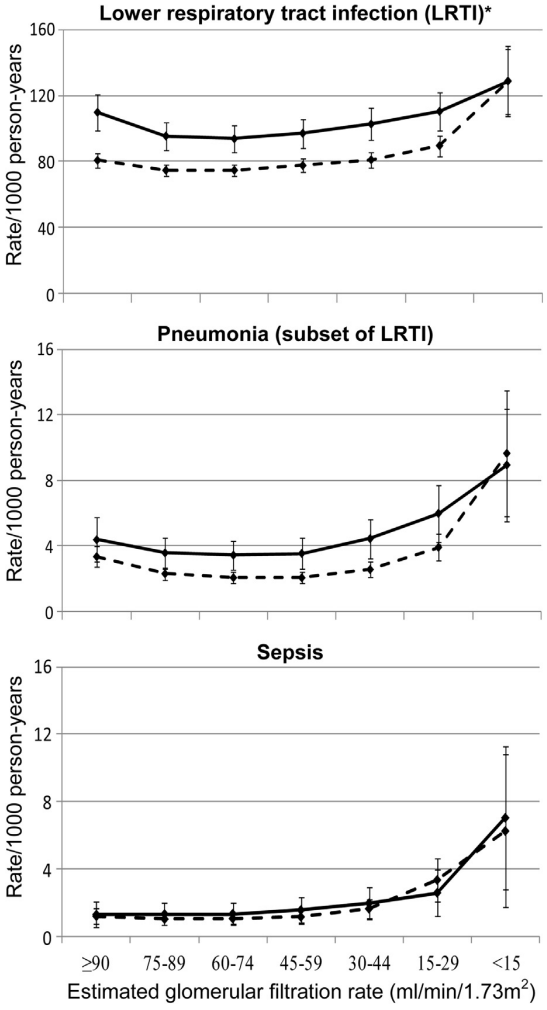


Figure 2. Fully adjusted infection rates/1,000 person-years (with 95% confidence intervals) against category of estimated glomerular filtration rate, by proteinuria status. Solid line, patients with a history of proteinuria; dashed line, patients with no history of proteinuria. Rates adjusted for age (updated), sex, socioeconomic status by practice, financial year prior to or post 2004, ischemic heart disease (updated), congestive cardiac failure (updated), hypertension (updated), cerebrovascular disease (updated), other dementia (updated), chronic lung disease (updated), smoking (baseline), hemoglobin A_{1c} level (baseline), and antidiabetes medication history (baseline). *LRTI y-axis scale is 10-fold greater than the pneumonia and sepsis scales.

without a history of proteinuria. These associations were diminished but persisted after adjustment for time-updated comorbid conditions and characteristics of diabetes for LRTI, pneumonia, and sepsis (IRRs of 1.07 [95% CI, 1.05-1.09], 1.26 [95% CI, 1.19-1.33], and 1.33 [95% CI, 1.20-1.47], respectively). The effect of proteinuria did not vary by eGFR category (Fig 2).

No clinically important interaction between age and eGFR was observed for LRTI or pneumonia. Sensitivity analyses limiting follow-up to post April 2004, restricting the data set to patients with HES linkage available, or using only the first infection as an outcome found similar results to the main analysis.

DISCUSSION

In our study population of older patients with diabetes, there was a high burden of community-acquired LRTI, pneumonia, and sepsis among those with CKD (manifested as reduced eGFR and/or history of proteinuria). Reduced eGFR and history of proteinuria represented independent risk markers for incidence of LRTI, pneumonia, and sepsis. Associations between eGFR and infection incidence were graded, with increased infection incidence at more severe stages of CKD. These associations persisted after adjustment for comorbid conditions, smoking status, and characteristics of diabetes mellitus. The association between eGFR and infection was not modified by age. Effect sizes were larger for sepsis than for pneumonia, and for pneumonia than for LRTI.

The strengths of the study include the following: first, consideration of both eGFR and proteinuria in mutually adjusted analyses; second, frequent monitoring of serum creatinine and proteinuria, allowing good ascertainment of CKD status; third, the detailed methods used to define infections, including exclusion of time in hospital from time at risk and treating recurrent consultations for infection within 28 days as a single episode; and fourth, inclusion of infections managed in primary care, not just those resulting in hospitalization or death.

The study is limited by the nature of routinely collected data; we may have underascertained proteinuria or comorbid conditions. The high prevalence of proteinuria and comorbid conditions observed in the study population suggests that neither is an extensive problem. Proteinuria monitoring has been financially incentivized in primary care for this population since 2004.[32] A small percentage of the potential study population had no available creatinine results and therefore were not included. These people did not appear to seek regular care, which may have led to us underestimating the association between CKD and infection. We do not have formal validation study data for our outcomes, but the advantage of linked data is capturing a more complete ascertainment of infections than in the stand-alone data sets used previously.

The complex relationships between CKD, infection, and cardiovascular disease limit interpretation of the direction of any causal association between CKD and infection. The same comorbid conditions, such as cardiovascular events, may both cause and be caused by CKD. Adjustment for baseline comorbid conditions risks underestimation of the association

Table 4. Infection Rates With Corresponding Rate Ratios by Proteinuria Status

Infection	No. of Events	Time at Risk (person-y)	Crude Rate/1,000 Person-y (95% CI)	Minimally Adjusted[a,b]	Adjusted for Comorbid Conditions at Baseline[b,c]	Adjusted for Time-Updated Comorbid Conditions[b,d]	Adjusted for Characteristics of Diabetes[b,e]
LRTI	115,080	808,194	155.8 (154.3-157.4)				
Proteinuria	31,823	202,658	178.7 (175.7-181.8)	1.13 (1.10-1.15)	1.15 (1.13-1.17)	1.09 (1.07-1.11)	1.07 (1.05-1.09)
No proteinuria	83,257	605,536	148.9 (147.3-150.5)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Pneumonia[f]	7,870	816,517	10.3 (10.1-10.6)				
Proteinuria	2,646	204,908	14.4 (13.8-15.1)	1.37 (1.30-1.45)	1.40 (1.33-1.48)	1.31 (1.24-1.39)	1.26 (1.19-1.33)
No proteinuria	5,224	611,609	9.0 (8.8-9.3)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)
Sepsis	1,980	816,826	2.5 (2.4-2.6)				
Proteinuria	712	205,002	3.6 (3.3-3.9)	1.44 (1.30-1.59)	1.46 (1.32-1.61)	1.41 (1.27-1.56)	1.33 (1.20-1.47)
No proteinuria	1,268	611,824	2.1 (2.0-2.2)	1.00 (reference)	1.00 (reference)	1.00 (reference)	1.00 (reference)

Note: Unless otherwise indicated, values are given as rate ratio (95% CI).

Abbreviations: CI, confidence interval; eGFR, estimated glomerular filtration rate; LRTI, lower respiratory tract infection.

[a]Adjusted for eGFR (updated), age (updated), sex, socioeconomic status by practice, and financial year prior to or post 2004.

[b]$P < 0.001$ for all rate ratios. Likelihood ratio test for inclusion of eGFR as a binary variable in the model.

[c]Adjusted for eGFR (updated), age (updated), sex, socioeconomic status by practice, financial year prior to or post 2004, ischemic heart disease, congestive cardiac failure, hypertension, cerebrovascular disease, other dementia, chronic lung disease, and smoking at baseline.

[d]Adjusted for eGFR (updated), age (updated), sex, socioeconomic status by practice, financial year prior to or post 2004, ischemic heart disease (updated), congestive cardiac failure (updated), hypertension (updated), cerebrovascular disease (updated), other dementia (updated), chronic lung disease (updated), and smoking (baseline).

[e]Adjusted for eGFR (updated), age (updated), sex, socioeconomic status by practice, financial year prior to or post 2004, ischemic heart disease (updated), congestive cardiac failure (updated), hypertension (updated), cerebrovascular disease (updated), other dementia (updated), chronic lung disease (updated), smoking (baseline), hemoglobin A_{1c} level (baseline), and antidiabetes medication history (baseline).

[f]Subset of LRTI.

between CKD and infection by residual confounding from new-onset comorbid conditions that may reflect baseline risk factors for these comorbid conditions, such as poorly treated hypertension prior to CKD. Adjustment for time-updated comorbid conditions is vulnerable to overadjusting for events that mediate any effect of CKD on infection. We present both models. The true association between CKD and infection is likely to lie between the 2 results. We avoided time-updating hemoglobin A_{1c} levels and smoking status because these are vulnerable to reverse causality from infection (eg, pneumonia may motivate smoking cessation). We also addressed the risk of reverse causality by conducting a sensitivity analysis in which we limited follow-up to the first infection, which found results similar to the main analysis. We did not adjust for vaccination status because interrelationships between CKD status, receipt of vaccination, and infection are likely to be complex and looking formally for vaccine effectiveness was beyond the scope of this report.

Our results may not be generalizable to the general population without diabetes because there may be a particular relationship between CKD and infection among older people with diabetes. However, older people with diabetes are a large and growing population with a high burden of CKD and infection, and so our findings in this population are important.

The associations we observed of preexisting eGFR with sepsis and pneumonia were similar to those observed between eGFR with bloodstream infections and pneumonia diagnoses in hospital records among the general population 65 years and older (although different outcome definitions mean the absolute rates are not comparable).[3,5] A large case-control study in the United Kingdom identified CKD as a risk factor for primary care diagnosis of pneumonia (adjusted odds ratio, 1.72; 95% CI, 1.3-2.07), but the unclear definition of CKD may have included patients with an increased risk of infection from renal replacement therapy and did not permit stage-specific rate ratios.[9] James et al[3] and Dalrymple et al[33] observed a J-shaped association between eGFR and infection risk, which was not present using cystatin C–based eGFR. Our results suggested that patients with eGFRs $\geq$ 90 mL/min/1.73 m^2 had a slightly increased risk of LRTI or pneumonia compared with those with eGFRs of 75 to 89 mL/min/1.73 m^2, but we did not observe a J shape for the association of eGFR with sepsis. James et al found that the association of eGFR with hospitalization for pneumonia was weaker among older age groups.[3] We did not observe an interaction between eGFR and age in the present

study, which may be due partly to our study population being limited to older people. In a previous study identifying albuminuria as a risk factor for infection-related hospitalization among patients with diabetes, eGFR was not an independent risk marker for infection, and infections managed in primary care were not included.[6] To our knowledge, our finding that proteinuria is a risk marker for incidence of LRTI, pneumonia, and sepsis, independently of eGFR, and including infections managed in primary care, is novel.

This study found that the associations of reduced eGFR and history of proteinuria with infection appeared to be due in part to underlying accrued cardiovascular and cerebrovascular comorbid conditions, but a strong graded association remained after adjustment for these. It has been suggested that CKD as a risk marker for infection may be a surrogate for chronicity of diabetes or poor glycemic control.[6] Adjustments for cardiovascular and cerebrovascular comorbid conditions will reflect chronicity and severity of diabetes to a certain extent, but further adjustment for severity and control of diabetes had little effect on the associations between CKD and infection, suggesting that the associations between CKD and infections are not explained fully by these factors. Stronger independent associations with CKD were observed for pneumonia than LRTI, and stronger still for sepsis, consistent with a view of CKD as a risk factor for poorer prognosis and greater incidence of infection. For clinicians managing older people with diabetes, our findings may help identify patients at increased risk of infection and inform discussions about infection risk and vaccination.[30] For policy makers, the association of CKD with a high burden of morbidity from infection is important for health service planning because the populations with diabetes and CKD are predicted to grow.[34]

Several unanswered questions remain. More research is needed to identify the biological mechanisms underlying the associations of proteinuria and eGFR with infection and to improve infection prevention strategies. For example, better understanding of vaccine effectiveness among people with CKD could inform whether pneumococcal and influenza vaccination recommendations should include people with proteinuria.[35]

ACKNOWLEDGEMENTS

Support: Dr Thomas reports a Career Development Fellowship grant (CDF 2010-03-32) from the National Institute for Health Research during the conduct of the study. Dr McDonald reports a PhD studentship grant from Kidney Research UK (grant reference ST2/2011) during the conduct of the study, for which Drs Thomas and Nitsch wrote the studentship application. The study sponsors had no role in the design or conduct of the study; the collection, management, analysis, or interpretation of data; the preparation, review, or approval of the manuscript; or the decision to submit it for publication. The views expressed in this publication are those of the authors and not necessarily those of the UK National Health Service, the National Institute for Health Research, the Department of Health, or Kidney Research UK.

Financial Disclosure: Dr Nitsch reports subcontract work for grants received by BMJ Informatica for the Healthcare Quality Improvement Partnership national CKD audit. The remaining authors declare that they have no other relevant financial interests.

Contributions: Research idea and study design: HIM, SLT, ERCM, DN; data acquisition: SLT; data analysis/interpretation: HIM, SLT, ERCM, DN; statistical analysis: HIM, ERCM; supervision or mentorship: SLT, DN. Each author contributed important intellectual content during manuscript drafting or revision and accepts accountability for the overall work by ensuring that questions pertaining to the accuracy or integrity of any portion of the work are appropriately investigated and resolved. HIM takes responsibility that this study has been reported honestly, accurately, and transparently; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

SUPPLEMENTARY MATERIAL

Table S1: Sensitivity analysis of association between eGFR and infection incidence.

Note: The supplementary material accompanying this article (http://dx.doi.org/10.1053/j.ajkd.2014.11.027) is available at www.ajkd.org

REFERENCES

1. 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):e002007.

2. Trotter CL, Stuart JM, George R, Miller E. Increasing hospital admissions for pneumonia, England. *Emerg Infect Dis*. 2008;14(5):727-733.

3. James MT, Quan H, Tonelli M, et al. CKD and risk of hospitalization and death with pneumonia. *Am J Kidney Dis*. 2009;54(1):24-32.

4. Collins AJ, Foley RN, Herzog C, et al. US Renal Data System 2010 Annual Data Report. *Am J Kidney Dis*. 2011;57(1)(suppl 1):A8, e1-e526.

5. James MT, Laupland KB, Tonelli M, Manns BJ, Culleton BF, Hemmelgarn BR. Risk of bloodstream infection in patients with chronic kidney disease not treated with dialysis. *Arch Intern Med*. 2008;168(21):2333-2339.

6. Hamilton EJ, Martin N, Makepeace A, Sillars BA, Davis WA, Davis TM. Incidence and predictors of hospitalization for bacterial infection in community-based patients with type 2 diabetes: the Fremantle Diabetes Study. *PLoS One*. 2013;8(3):e60502.

7. Karunajeewa H, McGechie D, Stuccio G, Stingemore N, Davis WA, Davis TM. Asymptomatic bacteriuria as a predictor of subsequent hospitalisation with urinary tract infection in diabetic adults: the Fremantle Diabetes Study. *Diabetologia*. 2005;48(7):1288-1291.

8. Viasus D, Garcia-Vidal C, Cruzado JM, et al. Epidemiology, clinical features and outcomes of pneumonia in patients with chronic kidney disease. *Nephrol Dial Transplant*. 2011;26(9):2899-2906.

9. 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-e338.

10. New JP, Middleton RJ, Klebe B, et al. Assessing the prevalence, monitoring and management of chronic kidney disease

in patients with diabetes compared with those without diabetes in general practice. *Diabet Med*. 2007;24(4):364-369.

11. De Lusignan S. Identification and management of chronic kidney disease. *Prescriber*. 2008;19(10):10-18.

12. Collins AJ, Foley RN, Chavers B. United States Renal Data System 2011 annual data report: atlas of chronic kidney disease and end-stage renal disease in the United States. *Am J Kidney Dis*. 2012;59(1)(suppl 1):e1-e420.

13. The NHS Information Centre Prescribing and Primary Care Services. *Quality and Outcomes Framework Achievement Data 2010/11*. Leeds, UK: The NHS Information Centre; 2011.

14. Campbell SM, Reeves D, Kontopantelis E, Sibbald B, Roland M. Effects of pay for performance on the quality of primary care in England. *N Engl J Med*. 2009;361(4):368-378.

15. McDonald HI, Nitsch D, Millett ER, 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*. 2013;31(5):606-614.

16. Holman N, Forouhi NG, Goyder E, Wild SH. The Association of Public Health Observatories (APHO) diabetes prevalence model: estimates of total diabetes prevalence for England, 2010-2030. *Diabet Med*. 2011;28(5):575-582.

17. Finlay S, Bray B, Lewington AJ, et al. Identification of risk factors associated with acute kidney injury in patients admitted to acute medical units. *Clin Med*. 2013;13(3):233-238.

18. CPRD. The Clinical Practice Research Datalink-CPRD. 2012. http://www.cprd.com/intro.asp. Accessed August 30, 2012.

19. 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.

20. Walley T, Mantgani A. The UK General Practice Research Database. *Lancet*. 1997;350(9084):1097-1099.

21. The Health and Social Care Information Centre. HESonline. 2012. http://www.hscic.gov.uk/hes. Accessed September 14, 2012.

22. Office for National Statistics. 2012. www.ons.gov.uk. Accessed September 14, 2012.

23. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*. 2009;150(9):604-612.

24. Matsushita K, Mahmoodi BK, Woodward M, et al. Comparison of risk prediction using the CKD-EPI equation and the MDRD Study equation for estimated glomerular filtration rate. *JAMA*. 2012;307(18):1941-1951.

25. National Kidney Foundation: K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Am J Kidney Dis*. 2002;39(2)(suppl 1): S1-S266.

26. Mathur R, Bhaskaran K, Chaturvedi N, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *J Public Health (Oxf)*. 2014;36(4):684-692.

27. de Lusignan S, Tomson C, Harris K, van Vlymen J, Gallagher H. Creatinine fluctuation has a greater effect than the formula to estimate glomerular filtration rate on the prevalence of chronic kidney disease. *Nephron Clin Pract*. 2011;117(3): c213-c224.

28. Millett ER, 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.

29. Ferguson J, Bauld L, Chesterman J, Judge K. The English smoking treatment services: one-year outcomes. *Addiction*. 2005;100(suppl 2):59-69.

30. Anandarajah S, Tai T, de Lusignan S, et al. The validity of searching routinely collected general practice computer data to identify patients with chronic kidney disease (CKD): a manual review of 500 medical records. *Nephrol Dial Transplant*. 2005;20(10):2089-2096.

31. Hobbs H, Stevens P, Klebe B, et al. Referral patterns to renal services: what has changed in the past 4 years? *Nephrol Dial Transplant*. 2009;24(11):3411-3419.

32. NHS Employers. Quality and outcomes framework. 2012. http://www.nhsemployers.org/payandcontracts/generalmedicalservicescontract/qof/pages/qualityoutcomesframework.aspx. Accessed September 12, 2012.

33. Dalrymple LS, Katz R, Kestenbaum B, et al. The risk of infection-related hospitalization with decreased kidney function. *Am J Kidney Dis*. 2012;59(3):356-363.

34. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu CY. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. *N Engl J Med*. 2004;351(13):1296-1305.

35. Kausz AT, Gilbertson DT. Overview of vaccination in chronic kidney disease. *Adv Chronic Kidney Dis*. 2006;13(3): 209-214.

Table S1: Sensitivity analysis of the association between estimated glomerular filtration rate (eGFR) and infection incidence, using patients with eGFR 75-90 ml/min/1.73m^2 as the reference group

	eGFR ml/min/1.73m^2	Number of infections	Rate ratio (95% confidence interval) [a]	P [b]
Lower respiratory tract infection (LRTI)				
	<15	607	1.48 (1.34–1.64)	
	15-29	5,153	1.18 (1.13–1.23)	
	30-44	16,557	1.08 (1.05–1.11)	
	45-59	29,783	1.03 (1.01–1.06)	<0.001
	60-74	31,538	1.00 (0.98–1.02)	
	75-90	24,110	1 (reference)	
	≥90	7,063	1.09 (1.06–1.13)	
	Total	114,811		
Pneumonia (subset of LRTI)				
	<15	99	2.97 (2.35–3.76)	
	15-29	650	1.69 (1.51–1.88)	
	30-44	1,523	1.16 (1.07–1.27)	
	45-59	1,980	0.93 (0.86–1.00)	<0.001
	60-74	1,815	0.92 (0.85–0.99)	
	75-90	1,397	1 (reference)	
	≥90	387	1.40 (1.24–1.58)	
	Total	7,851		
Sepsis				
	<15	41	5.72 (3.96–8.26)	
	15-29	186	2.57 (2.09–3.15)	
	30-44	387	1.55 (1.32–1.83)	
	45-59	499	1.15 (0.99–1.33)	<0.001
	60-74	449	1.03 (0.88–1.18)	
	75-90	324	1 (reference)	
	≥90	87	1.11 (0.87–1.42)	
	Total	1,973		

eGFR, estimated glomerular filtration rate

a. Adjusted for proteinuria (updated), age (updated), sex, socio-economic status by practice, financial year prior to or post 2004, ischaemic heart disease (updated), congestive cardiac failure (updated), hypertension (updated), cerebrovascular disease (updated), other dementia (updated), chronic lung disease (updated) and smoking (baseline).

b. Likelihood ratio test for inclusion of eGFR as a categorical variable in the model

8.3 Further discussion of adjusted estimates of CKD and infection incidence

The study presented several models for the association between markers of CKD and infection incidence, with increasingly full adjustment for patient characteristics. The "minimally adjusted model" was adjusted for time-updated age (within 5-year age bands up to age ≥85 years), sex, practice-level socio-economic status, financial year prior to or post 2004, and eGFR or proteinuria. The association between CKD and infection incidence was also modelled with additional adjustment for patient co-morbidities and smoking status at baseline (the "baseline co-morbidities model"); with adjustment for time-updated co-morbidities and smoking status at baseline (the "time-updated model"); and with further adjustment for characteristics of diabetes at baseline (the "characteristics of diabetes model"). The results of each analysis are presented in Table 3 of **Paper 3** (**8.2**).

This section offers an expanded discussion of the advantages and disadvantages of each model in describing the association between CKD and infection incidence, taking LRTI as an illustrative example, in the context of the conceptual model used to design the study.

8.3.1 How were effect estimates altered by adjustment for co-morbidities?

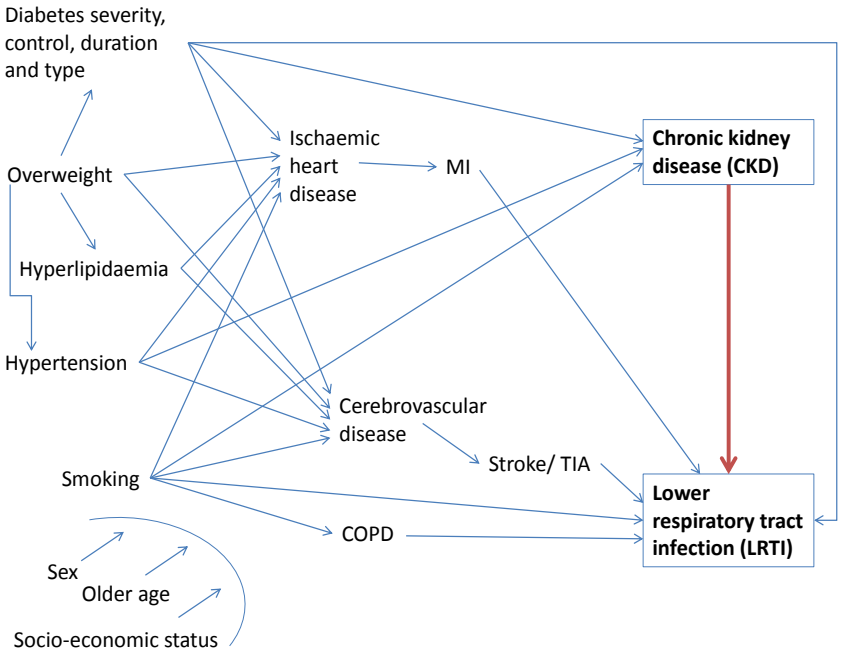
The minimally adjusted model provided effect estimates for the association of CKD with infection incidence which was not explained by age, sex, socio-economic status, or changes in recording practice pre and post-QOF. Estimates of the association between CKD and infection from the minimally adjusted model may be explained by other confounding patient characteristics.

Figure 8.1 shows a simplified conceptual diagram illustrating key relationships between selected patient characteristics at baseline, new-onset CKD, and a first subsequent LRTI. Age, sex and socio-economic status are associated with each component of the diagram, and these relationships are not individually shown. Baseline patient characteristics were considered *a priori* confounders if they were risk factors for community-acquired LRTI, and also associated with CKD, but not on any causal pathway between CKD and LRTI. Baseline patient characteristics identified as *a priori* confounders of the association between markers of chronic kidney disease and first incidence of community-acquired LRTI among older people with diabetes included: age, sex, socio-economic status, smoking status, ischaemic heart disease, hypertension, cerebrovascular disease, and chronic lung disease. The omission of other patient characteristics from this model is discussed in **8.3.6**. Several of these characteristics were expected *a priori* to have a strong confounding effect: for

example, smoking is an important risk factor CKD and for LRTI (directly and via co-morbidities including COPD and cardio- and cerebrovascular disease).

Overweight is represented on the diagram as an upstream cause of CKD and infection; however, the relationship between overweight and CKD varies over the lifecourse and is straightforward. Historical weight records were not uniformly available, and so mediators of the causal relationship of overweight with CKD and infection were adjusted for instead of overweight itself.

Figure 8.1 Key relationships between selected baseline patient characteristics, new-onset chronic kidney disease (CKD) and lower respiratory tract infection (LRTI)



MI, myocardial infarction; TIA, transient ischaemic attack; COPD, chronic obstructive pulmonary disease

Adjustment for these *a priori* confounders appeared to have a surprisingly small impact. Estimates for the association between eGFR and infection incidence were slightly lower in the baseline comorbidities model than the minimally adjustment model, but the confidence intervals for most estimates overlapped between the two models, and the size of each difference was too small to be of particular clinical relevance. Estimates of the association between CKD and infection incidence in the time-updated model were slightly lower again, but still similar in magnitude (**8.2: Table 1, Paper 3**).

8.3.2 Why were effect estimates so little altered by adjustment for co-morbidities?

There are three main explanations for the similarity between the results of different models. First, the association of CKD with infection incidence may in fact be relatively independent of confounding by these patient characteristics. This appears unlikely, given that a number of co-morbidities were identified which were expected to have a strong confounding effect on the association between CKD and infection incidence.

Second, misclassification of patient characteristics could have resulted in under-adjustment for confounding in the adjusted models. Classification of co-morbidities was vulnerable to under-ascertainment, as a negative could not be definitely established. Routinely-collected health records will positively identify the presence of a co-morbidity, but not its absence. Thus the absence of a co-morbidity was assumed for all patients without a relevant diagnostic code. In general, diagnostic codes recorded in CPRD have been found to have a high positive predictive value, in that a high proportion of patients with a diagnostic code have been confirmed as having the relevant co-morbidity. However, the proportion of patients with a co-morbidity who have a relevant diagnostic code (sensitivity) is less well established.[72] As most of the co-morbidities were *a priori* positive confounders of the association, under-ascertainment of co-morbidities could have resulted in over-estimation of the association between CKD and infection incidence in all adjusted models.

This risk was mitigated by studying a highly monitored population, by using detailed codelists, modelling each *a priori* confounder individually to allow adjustment for confounding by individual characteristics to be undiluted by any patient characteristics incorrectly identified as potential confounders, and taking a tailored approach to categorising co-morbidities where there were specific concerns. For example, congestive cardiac failure has been found to have a particularly strong association with LRTI and so was included as an independent variable separately from ischaemic heart disease.[56] Dementia was also separated into two independent variables: cerebrovascular dementia was included with stroke and TIA as evidence of cerebrovascular disease, which shares risk factors with CKD including a history of smoking, hypertension and age. Although only dementia of cerebrovascular aetiology is likely to confound the association between CKD and LRTI, the aetiology of dementia will not necessarily be identified, and may be less likely to be investigated or coded among patients with other severe co-morbidities or older age. To capture cerebrovascular dementia recorded using more general codes I also adjusted for 'other dementia', a category which included all dementia which did not have a

cerebrovascular aetiology specified (for example Eu02z00 *'[X] Unspecified dementia'*). This category will inevitably include patients with dementia of other aetiologies, and so to minimise misclassification of cerebrovascular dementia, 'other dementia' was retained as an independent variable.

A reassuringly high prevalence of co-morbidities was ascertained (**8.2: Table 1, Paper 3**). The effect estimates for the association of CKD and infection were large, and only slightly reduced by adjustment for the high prevalence of co-morbidities which was identified, suggesting that the confounding effect of these co-morbidities was not large, and a low level of under-ascertainment could perhaps be tolerated without invalidating the adjusted effect estimates. The possibility of residual confounding from under-ascertained co-morbidities resulting in over-estimation of the association between CKD and infection incidence cannot be excluded, but is unlikely to be a large enough issue to fully explain the similarity of observed effect estimates across adjusted models.

Third, the similarity may result from partial control for confounding already achieved in the minimally adjusted model. Partial control of two important *a priori* confounders, age and diabetes mellitus, had been achieved by restricting the study population to older people with diabetes. In addition, adjustment for age and socio-economic status may have partially adjusted for confounding by co-morbidities and smoking status. This may explain the small size of adjustment to effect estimates which was observed with additional adjustment for baseline or time-updated co-morbidities.

8.3.3 What are the advantages and disadvantages of adjustment for time-updated comorbidities?

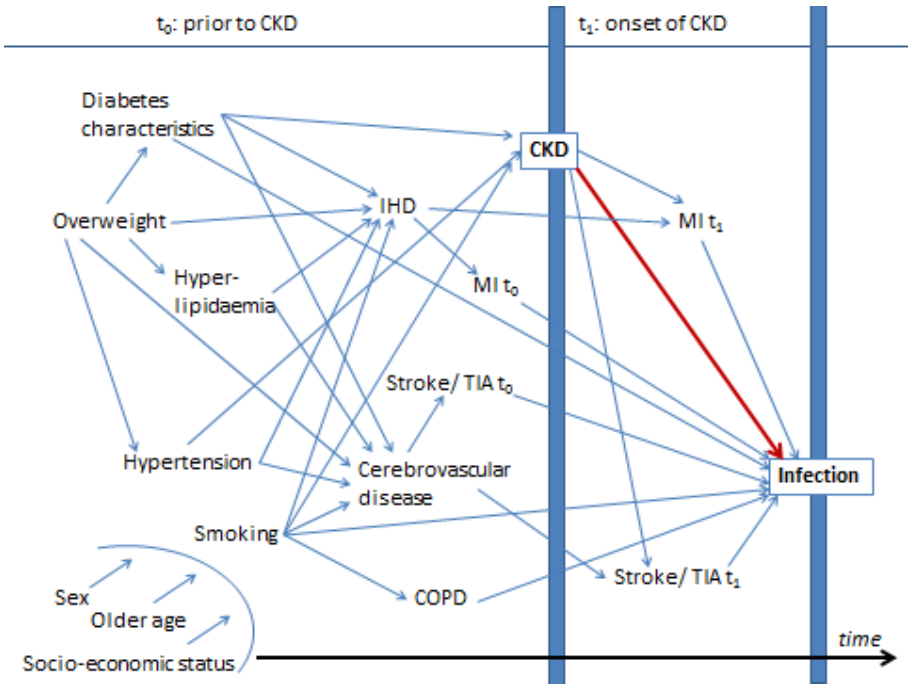
The baseline co-morbidities model estimates the association between CKD and infection incidence which is not explained by differences in patient characteristics at baseline. However, the baseline co-morbidities model is vulnerable to residual confounding from co-morbidities diagnosed during follow-up for two reasons. First, CKD status was updated during follow-up, and co-morbidities which developed or were diagnosed during follow-up may have influenced subsequent development or progression of CKD, and confounded the association between subsequent CKD status and later infection incidence. Second, several of the co-morbidities identified as confounders (including MI, stroke and COPD) may occur after a long time lag from the patient's initial exposure to their underlying risk factors (such as smoking or hypertension). Cardio- and cerebrovascular events, and diagnoses of COPD which occur after the onset or progression of CKD may therefore still indicate confounding

by risk factors (such as smoking or hypertension) which existed prior to CKD onset or progression.

This is illustrated in **Figure 8.2,** a simplified conceptual diagram illustrating key relationships of new-onset time-updated CKD and a first subsequent LRTI, with cardio- and cerebrovascular events occurring throughout follow-up. MI t_0 represents a myocardial infarction occurring prior to the onset of CKD (time t_0). This confounds the association between CKD and the first infection via several pathways of common risk factors, for example (CKD t_1←smoking→IHD→MI t_0→Infection).

MI t_1 represents a myocardial infarction which occurs after CKD onset but prior to a first infection. Despite occurring after CKD onset, MI t_1 also confounds the association between new-onset CKD and infection due to the lag time between patient exposure to risk factors for IHD (such as smoking and hypertension) and occurrence of MI, via several pathways (e.g. CKD t_1←smoking→IHD→MI t_1→infection). MI t_1 would additionally confound the association between any later progression of CKD status and subsequent infection via the same pathways as MI t_0. Adjustment for time-updated co-morbidities is designed to remove confounding by co-morbidities which occur during follow-up.

Figure 8.2: Key relationships between selected baseline patient characteristics, new-onset chronic kidney disease (CKD), cardio- and cerebrovascular events during follow-up, and first subsequent lower respiratory tract infection



MI, myocardial infarction; TIA, transient ischaemic attack; COPD, chronic obstructive pulmonary disease

However, adjusting for time-updated co-morbidities may represent over-adjustment of the association between CKD and infection. For example, both reduced eGFR and proteinuria are associated with MI and stroke, which can cause secondary LRTI: thus cerebro- and cardiovascular events may lie on the causal pathway between CKD status and LRTI.[137] This is represented in **Figure 8.2**: MI t_1 may mediate the association between CKD t_1 and infection (CKD t_1→ MI t_1→ infection).

Adjusting the association between CKD and infection for co-morbidities at baseline risks residual confounding of the association between CKD and infection from co-morbidities which occur or are diagnosed during follow-up. The time-updated model addresses the vulnerability of the baseline co-morbidities model to this residual confounding but removes the component of the association between CKD and infection which is mediated via these co-morbidities, which may result in over-adjustment. As the specified *a priori* confounders are predominantly risk factors for both CKD and infection, residual confounding would be expected to result in over-estimation of the association between CKD and infection, while over-adjustment would be expected to result in under-estimation of the association. Consistent with this expectation, the results of the time-updated model were slightly lower than those of the baseline co-morbidities model. While the best estimate is likely to lie between the two approaches, the two models produced very similar estimates for each infection and so interpretation of the association between CKD and infection appears robust to this difference in approach to adjustment for confounding.

8.3.4 What are the advantages and disadvantages of including repeat infections?

Some patients experienced multiple infections during follow-up, and repeat infections were included in the main analysis. The study included 115,080 LRTIs among 56,076 patients, 7,870 episodes of pneumonia among 7,095 patients, and 1,980 episodes of sepsis among 1,902 patients: thus 51.3% of LRTIs, 9.8% of pneumonia episodes, and 3.9% of sepsis episodes occurred as repeat episodes of the relevant infection for the patient.

As repeat infections represent a cluster of multiple outcomes occurring to an individual patient, including repeat infections requires appropriate analysis to avoid over-precise estimation of the association between CKD and infection. We addressed this by using a Poisson regression with a random effects model, as discussed in **4.6**. This results in less weight being given to infections which are part of a cluster of repeated infections occurring to one patient, which avoids over-precision of estimates.

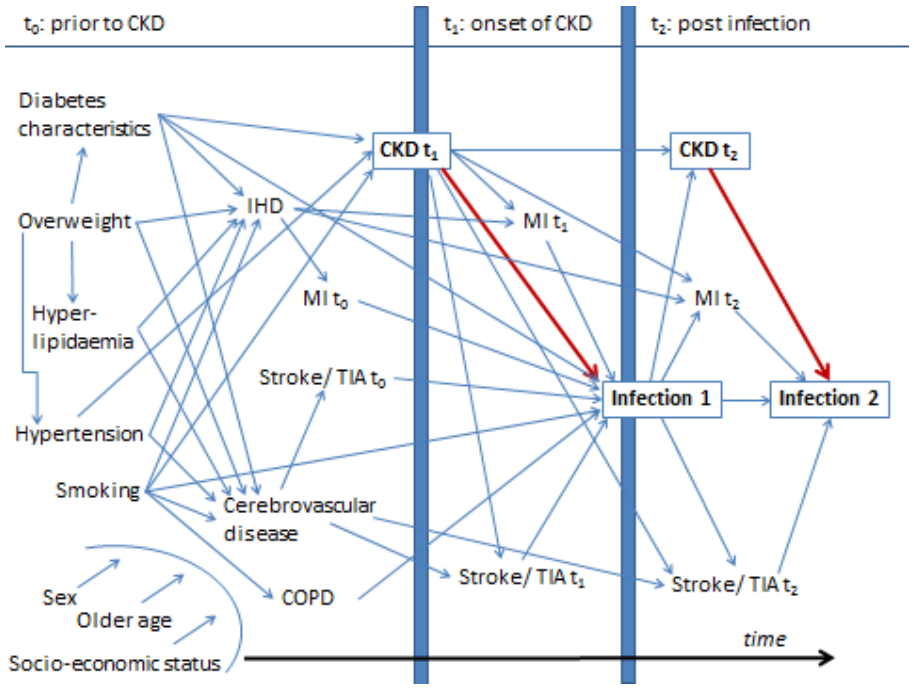
Including repeat outcomes in a model with time-updated variables risks introducing reverse causation to the estimate of the association of CKD and infection incidence. If a first infection is a risk marker for repeat infection and itself hastens the onset or progression of CKD (for example by causing acute kidney injury, AKI, which increases the subsequent risk of CKD progression),[139] then an association will be observed between CKD incidence and infections which is driven by a causal relationship from infection to CKD.

There is also a risk of introducing reverse causation by adjusting for other time-updated patient characteristics when including multiple infections. Smoking status is particularly vulnerable to reverse causation. For example, an episode of pneumonia may motivate a patient towards successful smoking cessation, which may reduce the risk of subsequent pneumonia and also reduce CKD progression. The confounding effects of a history of smoking are multi-dimensional, including the cumulative volume of smoking (measured in cigarette pack-years, which combine frequency and duration of smoking), and recentness of cessation, as well as immediate status as a current or former smoker. As only a crude measure of smoking status was consistently and routinely available within the data, classifying recent smokers who had been prompted to stop smoking by infection as ex-smokers (who generally have a lower risk of infection than current smokers) risked the category of ex-smokers among patients who had experienced a previous infection comprising a population at higher risk from recent smoking than other ex-smokers in the study. Adjustment for ex-smoking status applied equally to all patients would represent a greater under-adjustment among patients with a history of previous LRTI. The direction of effect from this would be unpredictable. The risks of this reverse causation were reduced by the decision not to time-update smoking status.

Inclusion of multiple infections with adjustment for time-updated co-morbidities which may be caused by either or both CKD and infection risks introducing collider bias. Collider bias is bias introduced by adjusting for a variable which is a common outcome of both the exposure of interest and the outcome of interest. An example is illustrated in **Figure 8.3**, a simplified conceptual diagram illustrating key relationships of time-updated CKD and subsequent LRTIs, with MI and stroke/TIA occurring during follow-up. MI t_2 represents a myocardial infarction which occurs after both the onset of CKD t_1 and a first subsequent infection. MI t_2 confounds the associations between CKD t_1 and the first infection and between CKD t_2 and the second infection, and may partially mediate the association between CKD t_1 and the second infection. In addition, since MI t_2 may be a result of CKD and infection, the effects of CKD and infection 'collide' along the way to producing MI t_2

(CKD t_2← CKD t_1→ MI t_2←Infection 1→Infection 2). Adjusting the association between CKD t_2 and Infection 2 for MI t_2 could introduce a collider bias which would bias the estimate of the association between CKD t_2 and Infection 2 downwards.

Figure 8.3: Key relationships between selected baseline patient characteristics, time-updated chronic kidney disease (CKD), cardio- and cerebrovascular events during follow-up, and lower respiratory tract infections over time



MI, myocardial infarction; TIA, transient ischaemic attack; COPD, chronic obstructive pulmonary disease

For simplicity, the ongoing effects of most baseline covariates on CKD t_2 and Infection 2 are shown via an association between CKD t_1 and CKD t_2, and an association between Infection 1 and Infection 2.

There are compelling advantages to including repeat infections within the main analysis. As LRTI is common, patients who entered the older age groups without a history of LRTI during follow-up would be a selected population, consisting either of patients at low risk of LRTI (for example patients with good overall health status), or patients who had entered the study at an older age (for example patients who had moved areas and newly registered with a GP at an older age, which may reflect better or worse health status than patients who are less geographically mobile). If patients had been followed up only to first occurrence of each infection, there would be a risk of introducing survivor bias among older age groups, particularly for LRTI, the direction of effect of which would be unpredictable. Including repeat infections also increased the power of the study to estimate the association between CKD and infection incidence precisely.

For these reasons, the main analysis included repeat infections, but a sensitivity analysis was conducted in which patients were followed-up only to the first infection (**Appendix E**). The effect estimates for the association between CKD and infection were unchanged, suggesting that the effects of reverse causation and collider bias on the main analysis were minimal.

8.3.5 Which model provides the best overall estimate of the association between CKD and infection?

Each model has a different profile of advantages and disadvantages for understanding the association between CKD and infection incidence. The minimally adjusted model provides an estimate of the size of the association (independent of demographics and changes in recording practice pre and post-QOF) which may be useful for health planners and economic analyses of the burden of excess morbidity among patients with CKD.

From a clinical perspective, it is perhaps more important to understand whether CKD is a useful risk marker of infection risk independently of other co-morbidities. The baseline co-morbidities and time-updated models estimate this association. While the baseline co-morbidities model may somewhat overestimate the association between CKD and infection due to residual confounding, and the time-updated model is vulnerable to over-adjustment for factors which mediate the relationship between CKD and infection incidence, both models produced similar results, and so either is a reasonable indicator of the use of eGFR and proteinuria as risk markers of infection incidence independently of other co-morbidities.

This is an observational study and thus had limited ability to explore causation of any observed association. However, one aspect of causation was explored: the role of severity of diabetes mellitus. It has been suggested that any use of CKD as a risk marker for infection among patients with diabetes mellitus is due to its status as a marker of diabetes severity. The characteristics of diabetes model adjusted for available characteristics of diabetes (HbA1C and history of antidiabetes medications), in addition to adjustment for cardiovascular disease which may also be a marker of macro-vascular complications of diabetes. Peripheral vascular disease and diabetic retinopathy were not adjusted for in this model. As these represent microvascular complications of diabetes, and this reflects the same aspect of diabetes severity as diabetic nephropathy, these were expected to have considerable collinearity with the association of CKD with infection among people with diabetes, and were not expected to have any separate causative association with the

infections studied. The association between CKD and infection incidence was slightly diminished after adjustment for characteristics of diabetes, but remained similar in magnitude, suggesting that the association between CKD and infection incidence is not merely explained by difference in severity of diabetes at baseline.

8.3.6 To what extent were all models vulnerable to residual confounding from omission of patient characteristics?

All models are vulnerable to confounding from patient characteristics not measured or adjusted for, if these confound the association between CKD and infection incidence.

Some baseline patient characteristics relevant to infection risk were omitted in the interests of parsimony as they were not believed to be relevant confounders. For example, ethnicity was considered a distal determinant of infection incidence. Ethnicity may be associated with infection incidence via prevalence of smoking, diabetes mellitus and cardiovascular disease. The study population was restricted to patients with diabetes, and the models adjusted for factors downstream of ethnicity (including smoking, cardiovascular disease and a range of co-morbidities). Ethnicity was not expected to be independently associated with infection incidence other than via the co-morbidities which had been adjusted for, and so was not included separately in the model. Black ethnicity may also affect estimation of eGFR from serum creatinine, but eGFR estimates were adjusted for Black ethnicity to reduce misclassification of eGFR (**5.4.3**). Body mass index and hyperlipidaemia were also considered distal determinants of infection, as the main confounding effects of these would be mediated via downstream characteristics (such as hypertension and cardiovascular disease) which were included in the model (**Figure 8.3.1**).

Risk factors for infection were also not adjusted for if their primary relationship between CKD and infection was on any causal pathway between CKD and infection incidence. Adjusting for variables on any causal pathway between CKD and infection would have resulted in over-adjustment and under-estimation of the association between CKD and infection incidence. For example, if anaemia and Vitamin D status are associated with both CKD and infection incidence this is likely to be as mediators of any causal relationship between CKD and infection incidence, and these patient characteristics were therefore not considered *a priori* confounders. The omission of functional status (which in practice is difficult to define using data available in CPRD) was also considered acceptable, as functional status is not a direct cause of CKD (except as a distal determinant) and is likely to have an important mediating role in any causal association between CKD and infection.

The approach throughout the thesis of adjusting for confounding by identifying diagnoses of co-morbidities rather than prescriptions of medication is discussed in **6.8.2,** and the rationale for identifying co-morbidities from primary care records but not secondary care records is discussed in **6.8.1**.

8.4 Additional sensitivity analysis of the association between CKD and infection incidence

An exclusion criterion for patients in study population B, who were the study population for Paper 3, was a history of renal replacement therapy in CPRD. This was revised for the subsequent studies comprising this thesis by additionally excluding patients with a history of renal replacement therapy identified in HES.

An additional sensitivity analysis was conducted post-publication exploring the impact of this change in Paper 3. The final model for the association of CKD with infection incidence adjusted for time-updated co-morbidities and the characteristics of diabetes was repeated for each infection, additionally excluding patients with a history of RRT in HES. This excluded an additional 62 patients, and 440 patients exited the study at an earlier date. The effect estimates of the association between CKD and infection incidence were unchanged (**Appendix E**).

8.5 How plausible is the observed prevalence of CKD?

The implications of any misclassification of CKD for estimating the association between CKD and infection incidence were discussed in **Paper 3**. This additional discussion considers the extent to which CKD may have been misclassified. First, the observed baseline prevalence of CKD is presented. The vulnerabilities of the study definition of CKD status to misclassification are briefly recapped, and the prevalence estimate is compared to the existing literature.

8.5.1 Observed prevalence of CKD

Table 8.1 presents the baseline prevalence of CKD for study population B, which comprised 218,688 patients with diabetes ≥65 years with no history of renal replacement therapy (including patients with missing data for smoking status, who were excluded from analysis in **Paper 3**). The identification of study population B was described in **3.9.2**.

Overall, 39.4% of the population had CKD of any stage as identified by either eGFR <60 ml/min/1.73m^2 or a history of proteinuria (86,192/218,688). The baseline prevalence of

CKD stages 3–5 (eGFR<60 ml/min/1.73m^2) was 31.6% (69,050/218,688). The prevalence of a history of proteinuria was 12.7%, and showed a graded increase with reduced eGFR: 15.5% of patients with CKD stages 3–5 had a history of proteinuria (10,714/69,050).

Table 8.1 Prevalence of markers of chronic kidney disease (eGFR and a history of proteinuria) at study entry

eGFR ml/min/1.73m^2	No history of proteinuria n	History of proteinuria n (row %)	Total n (column %)
<15	360	244 (40.4)	604 (0.3)
15-29	3,568	1,219 (25.5)	4,787 (2.2)
30-44	15,248	3,433 (18.4)	18,681 (8.5)
45-59	39,160	5,818 (12.9)	44,978 (20.6)
60-74	52,559	6,735 (11.4)	59,294 (27.1)
75-89	41,710	5,396 (11.5)	47,106 (21.5)
≥90	15,511	2,685 (14.8)	18,196 (8.3)
None recorded	22,716	2,326 (9.3)	25,042 (11.5)
Total	190,832	27,856 (12.7)	218,688

eGFR, estimated glomerular filtration rate based on the single serum creatinine result at study entry

8.5.2 Comparison of the observed CKD prevalence to the literature

The overall 39.4% prevalence of any CKD (eGFR <60 ml/min/m^2 or proteinuria) observed in the present study among people aged ≥65 years with diabetes appears to be broadly consistent with large US studies which have pro-actively tested for CKD. The National Health and Nutrition Examination Survey (NHANES) and the Kidney Early Evaluation Program (KEEP) both found approximately 44% prevalence of any CKD (identified as eGFR <60 ml/min/m^2 or albuminuria) among people aged ≥65 years in the US.[140] The NHANES study identified a 39% prevalence of any CKD (using the same definition) among people with diagnosed diabetes: although this estimate is not limited to older people, 44% of this population were aged ≥60 years.[22]

The 31.6% prevalence of stage 3-5 CKD observed in this study is at the upper end of international prevalence estimates of eGFR <60 ml/min/m^2 among people ≥65 years, which range from 4.7% to 35.8%,.[141, 142] In England, the 2009/2010 Health Survey for England pro-actively identified stage 3–5 CKD among 29% of men and 35% of women in the general population aged ≥75 years.[16] The present estimate is higher, as would be expected, as diabetes is as additional risk factor for CKD.

The 31.6% stage 3-5 CKD prevalence observed is also higher than estimates of CKD prevalence among people with the general population with diabetes in the UK. The National Diabetes Audit 2012/3 reported a prevalence of 20.1% (203,983/1,014,698) of

CKD stages 3–5 among patients with type 2 diabetes in England and Wales.[143] Studies of adults with diabetes in the UK using routinely-collected electronic health records have reported CKD stage 3-5 prevalence ranging from 18-31%.[18, 46, 144, 145] Although the population with diabetes is older than the general adult population, age is a strong risk factor for CKD and so the present study would still expect to observe a higher CKD prevalence than these. A study in Salford reported that the prevalence of CKD stages 3–5 rose from 27.5% among all patients with diabetes to 49% among those aged ≥70 years old.[144]

The 12.7% baseline proteinuria prevalence may be lower than expected for this study population. The Health Survey for England 2009 tested a randomly selected sample of the adult general population for albuminuria, and found a prevalence of 10% among men and 8% among women, rising to 26% of men and 18% of women aged ≥75 years.[16] In several studies using electronic health records, low recording of proteinuria in routine records has hindered assessment of proteinuria prevalence.[46, 144] However, in a study of patients with diabetes in East London a proteinuria test result was recorded for 75% of all patients. The prevalence of proteinuria was 8.6% among all patients tested, and 18.6% among all patients with eGFR <60 ml/min/1.73m^2.[145] Our study found comparable results but might be expected to be higher as the East London study was not limited to older people. In a 2014 study of the general population in CPRD, the prevalence of proteinuria was 7.7%, 12.2%, 20.1% and 38.0% among patients with CKD stages 3a, 3b, 4 and 5, respectively, and did not change over the study period, which spanned the introduction of QOF.[18] Our study found higher prevalences (as might be expected given that diabetic nephropathy will be more common in the study population of older people with diabetes), but both studies are equally vulnerable to under-ascertainment.

8.5.3 How likely is misclassification of CKD status according to reduced eGFR?

GFR was estimated from serum creatinine test results using the CKD-EPI equation, adjusted for Black ethnicity as described in **5.4.3**.[102] This offers a precise and granular estimate of eGFR status at the time of testing. There is still a risk of misclassification, however, as serum creatinine testing was driven by the clinical status of the patient rather than the needs of the study. The implications of this are discussed in depth in **5.4.4**. Briefly, the timing and frequency of creatinine testing may reflect a patient's overall health status and could even be directly related to incidence of infections. Serum creatinine levels fluctuate, and may in particular be raised during a period of acute illness (for example due to AKI), without existence of CKD. Basing classification on a single creatinine result tends to

overestimate CKD prevalence.[118] Attempts to reduce this risk by requiring multiple results over time to confirm chronicity of reduced eGFR risk under-estimation of CKD among patients lacking repeat tests, which could result in ascertainment bias affecting estimates of the association between CKD and infection incidence.

The main defence against this risk is complete and frequent routine testing of eGFR. Older people with diabetes were expected to be a highly-monitored population, and the study found that serum creatinine tests offered a relatively complete and frequently updated source of GFR estimates for this population: 88.5% of study population B had a creatinine test during follow-up, and 89% of creatinine tests were followed by another within the next year. This summary figure masks a marked change in serum creatinine test completeness before and after the introduction of QOF (**5.4.4**).

In **Paper 3**, GFR status at any given time was estimated from the single most recent serum creatinine test (the last-carried-forward method). This is vulnerable to over-estimation of CKD due to creatinine fluctuation and misclassification of AKI. Mimicking this approach, in **Table 8.1** baseline GFR was estimated from the first creatinine result after meeting eligibility criteria for entry to the study population to allow comparison of this estimate to existing literature.

8.5.4 How likely is misclassification of CKD status according to identification of persistent proteinuria?

Proteinuria status was defined from a proteinuria test results and diagnostic Read codes for proteinuria or proteinuric disease, as described in **5.5**. Methodological issues in identifying proteinuria status from routinely-collected primary care records are discussed in detail in **5.2** and **5.5**. To summarise, there were three main issues. First, test results are incompletely recorded in the coded dataset available to researchers, and negative test results are known to be particularly incomplete. Second, a patient with persistent proteinuria may have subsequent negative proteinuria tests (for example due to treatment with ACE inhibitors), but the history of persistent proteinuria may still be a relevant marker of increased clinical risk. Third, many records of proteinuria (such as test results) could represent either transient or persistent proteinuria.

To mitigate the risk of ascertainment bias from uncoded negative test results, it was assumed that absence of a positive test result implied a negative proteinuria status. This may have resulted in under-estimation of proteinuria prevalence due to misclassification of patients with proteinuria but no recorded test result.

To address the second issue, a single positive result defined a history of proteinuria for the rest of the patient's time in the study. Any misclassification of transient proteinuria as persistent proteinuria therefore had the potential to cause appreciable over-estimation of proteinuria status over the study period. To address this, proteinuria test results (and any Read codes which did not explicitly define persistent proteinuria) were excluded if recorded on the same day as a urinary tract infection, as this is a major cause of transient proteinuria. In addition, a cautious approach was taken to identifying positive results which included: not classifying trace results as positive, extensive data cleaning for internal consistency of records, and not including diagnostic codes for diabetic nephropathy in the codelist for proteinuria identification. If this approach was overly cautious, this may have resulted in under-ascertainment of proteinuria due to exclusion of positive records.

Proteinuria prevalence was therefore vulnerable to both under- and over-estimation. Many of the patients with an identified history of proteinuria had this status defined by a single record, many of which were test results, which potentially identify only transient proteinuria (**5.5.4**). It is hard to interpret whether this is due to the cautious approach to including positive proteinuria records (which would suggest under-estimation of proteinuria prevalence) or misclassification of some patients with transient proteinuria as having persistent proteinuria (which would cause over-estimation of proteinuria prevalence) but both may be partly responsible.

Chapter 9. The effectiveness of pneumococcal and influenza vaccinations in preventing community-acquired LRTI and pneumonia

This chapter presents a study of the extent to which the burden of community-acquired LRTI and pneumonia among older people with diabetes may be preventable with pneumococcal and influenza vaccination, and whether this varies according to CKD status (objective 4). The study-specific methods and results are presented in a draft article.

An ethics application has been made to the Royal College of General Practitioners Research and Surveillance centre requesting the dates of the first week in which 10% of nose and throat specimens sent as part of the RCGP influenza surveillance scheme tested positive for influenza virus, for each year of the study period.[146] This data would allow the definition of 'winter' in the analysis of influenza vaccine effectiveness to be tailored to the influenza season for each year. If the application is accepted, these data will be used to refine the draft article before submission. If not, the article will be submitted for publication as it stands. These data will not be obtained within the timeframe of the thesis, and so the draft paper is included in this thesis.

9.1 Introduction to Paper 4

This draft paper presents a retrospective cohort study to identify the vaccine effectiveness of pneumococcal vaccine against all community-acquired pneumonia, and of influenza vaccine against community-acquired LRTI, and whether these vary according to markers of CKD. The study population comprised 191,672 patients aged ≥65 years with diabetes mellitus and a valid serum creatinine result during follow-up, with no history of renal replacement therapy (study population B), identified as described in **3.9.2**.

The outcomes were acute, community-acquired lower respiratory tract infection (LRTI), and pneumonia (as a subset of LRTI). Methods used to identify community-acquired infection incidence were detailed in **Chapter 4**. Pneumococcal and influenza vaccinations were identified as described in the draft paper. CKD was identified by a history of proteinuria or reduced estimated glomerular filtration rate (eGFR) as described in **Chapter 5**. Incidence rate ratios were adjusted for *a priori* potential confounders of the association between markers of vaccination and infection. Detailed definitions of these confounders are described in **Chapter 6**.

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	Helen McDonald
Principal Supervisor	Dr Dorothea Nitsch
Thesis Title	The epidemiology of infections among older people with diabetes mellitus and chronic kidney disease

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?			
When was the work published?			
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion			
Have you retained the copyright for the work?*	Choose an item.	Was the work subject to academic peer review?	Choose an item.

**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?	Will be submitted first to Diabetes Care
Please list the paper's authors in the intended authorship order:	McDonald HI, Thomas SL, Millett ERC, Quint JK, Nitsch D.
Stage of publication	**Not yet submitted**

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)	The original idea for this study was conceived by Dorothea Nitsch and Sara Thomas, who wrote the study proposal and obtained ethics approval and funding for the study. Sara Thomas obtained the study data from CPRD. I developed a detailed study design from the original study proposal, with advice and supervision from Dorothea Nitsch

	and Sara Thomas. This study used the dataset of community-acquired infections among study population B which I had constructed for objective 2 using methods developed by Elizabeth Millett, as presented in Paper 2 (Chapter 7). Jennifer Quint is an advisory committee member to Elizabeth Millett, and advised on the development of these methods. I developed an approach to identifying markers of chronic kidney disease in CPRD (described in Chapter 5), with advice and supervision from Dorothea Nitsch and Sara Thomas. I extracted, cleaned and classified the data to define other variables as described in Chapter 6, and was fortunate to be able to use codelists developed by colleagues at LSHTM (each attributed individually in Chapter 6). I conducted all data analysis, and drafted the manuscript, which was then commented on by all co-authors.

Student Signature: K McDonald **Date:** 10 April 2015

Supervisor Signature: Dorothea Nitsch **Date:** 14/4/15

Title: How helpful are influenza and pneumococcal vaccines in the bid to prevent community-acquired LRTI among older people with diabetes and does this vary with chronic kidney disease? A cohort study using electronic health records

Authors: Helen I McDonald,[1] Sara L Thomas,[2] Elizabeth RC Millett,[2] Jennifer K Quint,[1] Dorothea Nitsch.[1]

1. Department of Non-Communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK.

2. Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK.

9.2.1 Abstract

Objective: We aimed to estimate the effectiveness of influenza and pneumococcal vaccination to reduce the burden of community-acquired lower respiratory tract infection (LRTI) among older people with diabetes, and whether this varied with chronic kidney disease.

Research design and methods: We used linked UK electronic health records for a retrospective cohort study of 190,492 patients ≥65 years with diabetes mellitus and no history of renal replacement therapy, 1997–2011. We included community-acquired LRTIs managed in primary or secondary care. Infection incidence rate ratios were estimated using Poisson regression. Pneumococcal vaccine effectiveness (VE) was calculated as (1 - effect measure). To estimate influenza VE a ratio-of-ratios analysis (winter effectiveness/summer effectiveness) was used to address confounding by indication. Final estimates of VE were stratified according to estimated glomerular filtration rate and proteinuria status.

Results*:* Neither influenza not pneumococcal vaccine uptake varied according to CKD status among older people with diabetes. Pneumococcal VE was 22% (95%CI 11–31) against community-acquired pneumonia for the first year after vaccination, but was negligible after five years. Using a ratio-of-ratios analysis, current influenza vaccination had 7% effectiveness for preventing community-acquired LRTI (95%CI 3–12). Pneumococcal vaccine effectiveness was lower among patients with a history of proteinuria than patients without proteinuria (p=0.04), but otherwise this study did not identify variation in pneumococcal or influenza VE according to markers of CKD.

Conclusions: The public health benefits of influenza vaccine may be modest among this population. Pneumococcal vaccination protection against community-acquired pneumonia declines swiftly: annual vaccination schedules should be investigated.

9.2.2 Introduction

Hospital admissions for pneumonia are rising rapidly in the UK, most steeply among older people.[9] Older people with diabetes have a particularly high burden of lower respiratory tract infection (LRTI) and pneumonia.[147]

Directly or indirectly, *Streptococcus pneumoniae* ('pneumococcus') and seasonal influenza viruses are responsible for a large burden of community-acquired pneumonia. Pneumococcus is the commonest cause of community-acquired pneumonia among older people.[91] Up to a third of community-acquired pneumonia may be influenza-related, due to bacterial co-infection or secondary bacterial pneumonia.[148] Vaccination is available against both these pathogens, and recommended in the UK for everyone aged ≥65 years.[149] However, the extent to which these vaccines protect against pneumonia among older people remains unclear for both vaccines.

The effectiveness of pneumococcal vaccination against all-cause pneumonia among older people is unclear, and meta-analyses have been hampered by unexplained heterogeneity.[150-152] Waning immunity among vaccinated participants has been suggested as a possible cause, but few estimates are available of pneumococcal vaccine effectiveness (VE) according to time since vaccination.[153]

Traditional observational studies of influenza VE among older people may have over-estimated influenza VE due to uncontrolled confounding by indication, in which the patient's functional status affects vaccine uptake.[154-156] Observational studies which have used strategies to control confounding by indication (such as a "ratio-of-ratios" analysis in which the excess influenza VE during winter compared to summer is calculated) have suggested a null or modest influenza VE against community-acquired pneumonia among older people.[154, 157-159]

Older people with diabetes have a high prevalence of chronic kidney disease (CKD).[160] Even at early stages, patients with CKD have increased incidence of LRTI and pneumonia.[125, 160, 161] Patients with CKD have a generally reduced response to vaccines, and a faster decline in antibody levels following vaccination.[30] A ratio-of-ratios analysis of influenza VE found no evidence of any protection against influenza-like-illness, influenza/pneumonia hospitalisation, or mortality among patients receiving haemodialysis.[162] Influenza VE at earlier stages of CKD is unclear, and still less is known about pneumococcal VE among patients with CKD.[27, 30]

We aimed to describe the extent to which the burden of community-acquired LRTI and pneumonia among older people with diabetes may be preventable with pneumococcal and influenza vaccination, and whether this varied according to CKD status. We conducted a retrospective cohort study using linked primary and secondary care electronic health record to calculate the vaccine effectiveness of pneumococcal vaccine against all community-acquired pneumonia. Since influenza vaccine may potentially reduce incidence of both influenza infection and secondary pneumonia, we calculated the influenza VE to prevent all community-acquired LRTI (considered as a broad category of all 'chest infections', including influenza infections, and possible secondary infections such as bronchitis and pneumonia), using a ratio-of-ratios analysis to address confounding by indication.

9.2.3 Research design and methods

Data sources

We analysed data from the Clinical Practice Research Datalink (CPRD), a database of anonymised primary care medical records. Data were extracted in May 2011, and contained records for 12.8 million patients at 627 practices across the UK.[68] Records include patient demographics, health behaviours, test results, diagnoses, and prescriptions. Diagnoses are recorded using Read codes, and have generally been found to have good positive predictive value in validations.[72] The CPRD population is similar to the general UK population in terms of age and sex.[66, 67]

Linked data are available for patients in England, subject to practice-level consent. This study used linked data on all hospital inpatient admissions to NHS hospitals in England from Hospital Episodes Statistics (HES), and socio-economic status from the Office for National Statistics (ONS).[75, 163]

Study population

The study population comprised all patients in CPRD with diabetes mellitus, aged ≥65 years, with no history of renal replacement therapy, who had at least one valid serum creatinine result recorded in primary care. Diabetes was identified by diagnostic Read codes. For less definitive Read codes, we required confirmation with an antidiabetes medication prescription, as described in detail previously.[147]

Patients met eligibility criteria at the latest time-point of: diabetes diagnosis, 65th birthday, one year after practice registration, practice fulfilling CPRD quality control standards, or 1 April 1997. Their study entry date was their first valid serum creatinine result after the

eligibility criteria were met. Patients left the study at the first time-point of: death, leaving the practice, last data collection from the practice, renal replacement therapy (dialysis or renal transplant), or 31 March 2011. Patients with a diagnosis of HIV or hyposplenia (including coeliac disease or sickle cell disease) at any point in their medical record were excluded from the study.

Definition of infections

Lower respiratory tract infection was defined as a broad category of all infections of the lower respiratory tract, including influenza infections, bronchitis and pneumonia.

A clinical diagnosis of infection was identified by a diagnostic Read code in primary care records, or a diagnostic International Classification of Disease 10 (ICD-10) code as the primary cause of hospital admission in secondary care records. To avoid overestimation from repeat attendances for the same infection, diagnostic codes recorded within 28 days of one another were attributed to a single episode of infection. The first consultation for infection was treated as the date of infection onset, and the infection had duration until 28 days after the latest of the last diagnostic code or hospital discharge. All infections with onset date during a HES hospitalisation spell, or within 14 days following hospital discharge, or which included a code for postoperative infection, were designated hospital-acquired, and excluded. These methods have also been described in detail previously.[5]

Time at risk

Patients were not at risk of incident community-acquired infection during ongoing infection (community- or hospital-acquired), during any hospitalisation, or within 14 days following hospital discharge. These time periods were removed from time at risk. As pneumonia was a subset of LRTI, a patient could be at risk of pneumonia during an ongoing LRTI.

Assignment of vaccination status

Vaccination status was identified from primary care records using Read codes, prescription data, and immunisation record forms.

For pneumococcal vaccination, any of these records could define a first vaccination, and any subsequent prescription could identify a booster vaccination. Time-updated pneumococcal vaccination status was classified according to time since the latest pneumococcal vaccination (<1 year, 1 – 5 years, ≥5 years, never vaccinated).

Time-updated influenza vaccination status was assigned within vaccination years (1 September to 31 August). Within each vaccination year, influenza vaccination status was

current from the first vaccination record to the subsequent 31 August. Patients without a current vaccination who had received an influenza vaccination within any of the previous five vaccination years were classified as having 'residual' influenza vaccination status, and other patients were categorised as unvaccinated.

Definition of CKD

We studied two markers of CKD: estimated glomerular filtration rate (eGFR) and proteinuria. Estimated GFR was calculated from serum creatinine test results in primary care, using the CKD-EPI equation, including adjustment for Black ethnicity.[102] Estimated GFR status was time-updated using a last-carried-forward method, with eGFR status assigned according to the most recent creatinine result.[125]

A history of proteinuria was established from a Read code for persistent proteinuria or proteinuric disease, or a positive test result which did not coincide with a urinary tract infection diagnosis.

Definition of covariates

Age was categorised in five-year bands up to a final category of ≥85 years. Socio-economic status was assigned at a practice level, using 2007 ONS estimates of the Index of Multiple Deprivation, a composite area-level marker of deprivation.[163] Smoking status was identified as current, ex-smoker or non-smoker from HES or CPRD records. Co-morbidities were identified from diagnostic Read codes in CPRD and were modelled as separate variables which were: ischaemic heart disease, congestive cardiac failure, hypertension, cerebrovascular disease, other dementia, chronic lung disease (which included chronic obstructive pulmonary disease and other non-reversible lung disease, but not asthma), chronic liver disease. Baseline HbA1C was defined by the most recent HbA1C test result in CPRD prior to (or on) the study entry date. Baseline medication history was identified from CPRD prescription records.

Data analysis

Analysis was conducted separately for pneumococcal vaccine effectiveness against pneumonia and for influenza vaccine effectiveness against LRTI.

We excluded patients with missing smoking status. For co-morbidities and proteinuria status, absence of a positive record was treated as absence of disease. Absence of a recorded HbA1C test result was included as indicating a relevant category of control.

Incidence rates and rate ratios were calculated for each infection using Poisson regression with lexis expansions for age, and a random effects model to adjust for multiple infection episodes. We adjusted models for pre-specified *a priori* confounders of the association between vaccination status and respiratory infection, and/or the relationship between chronic kidney disease and respiratory infection. These were: age, sex, socio-economic status at practice level, residential or nursing home care, baseline smoking status, time-updated co-morbidities, steroid use in the 3 months prior to study entry, HbA1C and diabetic medication history at baseline, and date prior to or post 1 April 2004 (when Quality Outcomes Framework guidelines introduced financial incentives for recording CKD status among people with diabetes in primary care which may have improved ascertainment of CKD in primary care).[20]

For pneumococcal vaccine, vaccine effectiveness was calculated as (1 – effect estimate). To explore waning of immunity we described pneumococcal VE according to time since vaccination.

To control for confounding by indication in influenza vaccination we estimated the ratio of influenza VE in summer to influenza VE in winter in a "ratio-of-ratios" analysis by including an interaction term between influenza vaccination status and season, and reporting the antilog of the beta coefficient for the interaction term.[162] Winter was defined as 1 September to 31 March, to capture excess winter influenza-like-illness.[164]

Final estimates of VE were stratified by time-updated eGFR and history of proteinuria, as markers of CKD.

Stata version 13.1 was used for data analyses. All code lists are available on request.

Sensitivity analyses

Pneumococcal vaccination has been recommended for patients with CKD in the UK since 1992, but in 2003 the recommendation was extended to everyone aged ≥65 years.[149] As a sensitivity analysis we estimated pneumococcal VE separately for the periods before and after 31 March 2003 (to avoid separating the 2002-3 winter season) to check for bias from secular changes in vaccine uptake.

The match of influenza vaccine strain to circulating influenza varies each year, which affects vaccine effectiveness.[162] As a sensitivity analysis, we estimated influenza VE separately for each winter.

We conducted a sensitivity analysis of influenza VE excluding patients with chronic lung disease or congestive heart failure, as the relationship of influenza to LRTI aetiology for these patients may differ from that among the general population.

Ethics

The study was approved by the Independent Scientific Advisory Group of the CPRD (ISAC reference 11_033A) and the London School of Hygiene & Tropical Medicine Ethics Committee (LSHTM reference 6116).

9.2.4 Results

Of 193,470 eligible patients, 1,049 patients with a diagnosis of HIV or hyposplenia, 1,764 patients with no smoking status available, and 165 patients who had a record of pneumococcal vaccine administration with a missing date, were excluded from both analyses **(Figure 1)**. For both pneumococcal and influenza vaccinations, unvaccinated patients had a lower recorded prevalence of ischaemic heart disease and chronic lung disease than vaccinated patients. Unvaccinated patients may have had poorer diabetic control than vaccinated patients: a higher proportion had poor or unrecorded HbA1C status, and a lower proportion had a history of both oral antidiabetes medication and insulin prescription than vaccinated patients. Prevalence of chronic kidney disease was similar for vaccinated and unvaccinated patients, although unvaccinated patients had a slightly lower prevalence of a recorded history of proteinuria **(Table 1)**.

Pneumococcal vaccine

190,492 patients contributed 811,498 person-years to the pneumococcal vaccine analysis, during which there were 7,805 pneumonia episodes among 7,036 people. At study entry, 58.3% of patients (111,016/190,492) were vaccinated against pneumococcal disease **(Table 1)**. Baseline pneumococcal vaccination increased among patients who entered the study after 2003–4, and did not differ according to eGFR at baseline **(Figure S1)**.

Crude rates of pneumonia were lowest among patients within a year of pneumococcal vaccine. The adjusted effectiveness of pneumococcal vaccine for preventing pneumonia was 22% (95% CI 11–31) within the first year after vaccination, and fell with increasing time since vaccination. Pneumonia incidence among patients vaccinated more than 5 years previously was similar to that among patients with no record of vaccination (incidence rate ratio, IRR 1.03: 95% CI 0.95–1.11). There was the suggestion of a trend of decreased pneumococcal vaccine effectiveness among patients with reduced eGFR, but this was not statistically significant. There was evidence for a greater protective effect of pneumococcal

vaccine among patients without a history of proteinuria than with a history of proteinuria **(Table 2)**.

A sensitivity analysis of pneumococcal VE stratified by date before or after 1 April 2003 suggested that the estimate was not affected by the change in vaccine recommendation in 2003 **(Table S1)**.

Influenza vaccine

For the influenza vaccine effectiveness analysis, 190,459 patients contributed 803,230 person-years to time at risk, during which there were 114,313 cases of LRTI among 55,685 patients. At study entry, 65.2% of patients (124,130/190,459) had received a current vaccination against influenza **(Table 1)**. Baseline influenza vaccination status increased slightly over time, and did not differ by eGFR status **(Figure S2)**.

Vaccinated patients had a higher crude incidence of LRTI than unvaccinated patients, in winter and summer. After adjustment for age, sex, co-morbidities, pneumococcal vaccination, and characteristics of diabetes, the winter incidence rate of LRTI was higher among patients with a current influenza vaccine than unvaccinated patients (IRR 1.19: 95%CI 1.15–1.23) and among patients with residual influenza vaccination than unvaccinated patients (IRR 1.23: 95%CI 1.18–1.28). Similar, or higher, adjusted incidence rate ratios were observed in summer. Using the ratio-of-ratios analysis, a 7% effectiveness of current influenza vaccine (95% CI 3–12) and a 12% effectiveness of residual influenza vaccination (95% CI 7–17) to prevent community-acquired LRTI were observed. There was no evidence to suggest a relationship between vaccine effectiveness and eGFR nor proteinuria **(Table 3)**.

Similar results were obtained in sensitivity analyses of influenza VE stratified by year **(Table S2)**, and excluding patients with chronic lung disease and congestive heart failure **(Table S3)**.

9.2.5 Conclusions

Pneumococcal vaccine had 22% (95%CI 11–31) effectiveness against community-acquired pneumonia within the first year after vaccination. Pneumonia incidence among patients vaccinated more than 5 years previously was similar to that among patients with no record of vaccination (incidence rate ratio, IRR 1.03: 95%CI 0.95–1.11). Community-acquired LRTI rates were higher among patients who received an influenza vaccination than among patients who did not, and this relationship remained after adjustment for age, sex, co-morbidities and characteristics of diabetes, and was observed in both summer and winter.

Using a traditional analysis, a negative vaccine effectiveness of influenza vaccine to prevent community-acquired LRTI would have been observed. However, using a ratio-of-ratios analysis, a 7% effectiveness (95% CI 3–12) of current influenza vaccine against community-acquired LRTI was observed. There was no evidence of a trend in influenza vaccine effectiveness according to CKD status, however there was evidence that the protective effect of pneumococcal vaccine was greater among patients without a history of proteinuria than patients with a history of proteinuria.

Previous meta-analyses have found insufficient evidence for a protective effect of pneumococcal vaccine against all-cause pneumonia among the adult population due to heterogeneity.[150-152] A subgroup analysis of a large Spanish cohort study found that only recent pneumococcal vaccination (<5 years) protected against hospitalisation for all-cause community-acquired pneumonia (hazard ratio 0.75; 95% CI 0.58–0.98) among the general population aged ≥60 years.[153] The authors suggested that the heterogeneity observed in meta-analyses might be explained by waning immunity among the vaccinated population. Our results support this view and suggest that pneumococcal vaccination appears to be effective against all-cause community-acquired pneumonia for a year following vaccination among people aged ≥65 years with diabetes, after which time we observed a decrease in pneumococcal vaccine effectiveness to a null effect after five years.

Previous cohort studies among older people have provided evidence of a “healthy vaccinee effect”, in which higher vaccine uptake among healthier patients resulted in likely over-estimation of influenza VE.[154-156, 165] Evidence suggesting a healthy vaccinee effect has also previously been found among older people with diabetes.[166-170] In contrast, we observed higher rates of LRTI among patients who had received an influenza vaccination than among unvaccinated patients: our vaccinated patients appear, on this outcome measure, to be *less* healthy than unvaccinated patients. This finding is intriguing. The major difference between our study and most previous studies of this question is that we have included community-acquired LRTIs diagnosed and managed in both primary and secondary care. One possible explanation of the difference is that vaccination may reflect health-seeking behaviour in primary care. When patients develop symptoms of LRTI, patients who attend primary care for diagnosis and treatment may also be patients who were more likely to take up the influenza vaccine. This ascertainment bias may be less relevant to studies with hospitalisation as an outcome - or could even be reversed, as vaccinated patients who attended primary care promptly with LRTI may be less likely to require hospital admission. An alternative explanation is that the healthy vaccinee effect

observed in studies of hospitalisation for LRTI/pneumonia may reflect residual confounding by 'frailty' in which frailer patients are less likely to take up vaccination and more likely to be admitted to hospital when they develop infection. This would be less relevant to diagnosis of LRTI in primary care, and so our outcome may be less vulnerable to residual confounding by indication.

Our "ratio-of-ratios" estimate suggested 7% VE of current influenza vaccination against LRTI among older people with diabetes (95% CI 3–12). Previous studies using similar strategies among the general population of older people, have found no evidence of influenza VE against community-acquired pneumonia (VE 8%: 95% CI -10–23%), and evidence of a modest protection against influenza-related excess hospitalisation with pneumonia/influenza (VE 19%: 95% CI 4–31%).[158, 159] Our estimate is consistent with both these estimates, and the difference may be due to the higher precision available for the present study due to the large cohort size.

Our results suggested that pneumococcal VE may be reduced among patients with a history of proteinuria. We did not find any evidence of altered influenza VE among patients with CKD, but this may be due to limited power for the stratified ratio-of-ratios analysis. To the best of our knowledge, neither pneumococcal VE against pneumonia nor influenza VE against LRTI using methods to control for confounding by indication have previously been studied among patients with CKD who are not receiving dialysis. Studies of patients receiving dialysis may give some indication as to whether alteration of VE with CKD status is likely. A large observational study of pneumococcal vaccine found no evidence of effectiveness against hospitalisation for pneumonia or respiratory infections among patients receiving dialysis.[171] A study of influenza vaccine which calculated a ratio-of-ratios VE comparing influenza effectiveness in years with good match between the vaccine and circulating strain to effectiveness in a poorly-matched 'placebo year' found no evidence of protection against influenza/pneumonia hospitalisation among patients receiving haemodialysis (VE 2%: 95% CI −2– 5).[162] These studies suggest that the suggestion of reduced pneumococcal VE associated with CKD is plausible.

This study has several strengths. We used large, linked datasets with a careful definition of infection episodes to identify community-acquired infections managed in primary or secondary care, and excluded hospital-acquired infections and hospitalisation from time at risk. This avoids differential hospital attendance patterns biasing estimates of VE according to markers of CKD.[107] We adjusted for a wide range of co-morbidities, and conducted a ratio-of-ratios analysis for influenza VE to address confounding by indication. We described

the effect of pneumococcal vaccine according to time since vaccination, including booster doses, to identify waning immunity following vaccination. Our study population of older people with diabetes is well-monitored for CKD,[46] and this permitted us to explore the relationship of influenza and pneumococcal VE with CKD among patients not receiving dialysis, which we believe is novel.

As an observational study of vaccine effectiveness using routinely-collected health record data, the study has limitations. We may have under-ascertained vaccinations, proteinuria and co-morbidities: however, the selection of a highly monitored study population should minimise this risk, and the high prevalence of each we observed suggests that this was not a major source of misclassification. Despite adjustment for multiple co-morbidities, residual confounding by indication may remain in the pneumococcal VE analysis. Despite our use of large, linked datasets, we had limited power to estimate the relationship of VE according to CKD status, especially in a ratio-of-ratios influenza VE analysis.

Our findings have implications for clinical practice, public health and future research. As LRTI and pneumonia are typically diagnosed clinically in general practice, without microbiological testing for the causative pathogen, we studied broad LRTI/pneumonia outcomes, in common with previous observational studies of influenza VEff. Our results should not be interpreted as demonstrating that influenza and pneumococcal vaccines are ineffective against their specific pathogens among this population. As such, the results should not discourage patients nor health professionals from influenza and pneumococcal vaccination.

Our study question was the extent to which the burden of community-acquired LRTI may be preventable with vaccination and our results suggest that the growing burden of community-acquired LRTI and pneumonia among this population cannot be easily tackled by increasing uptake of existing routine vaccination programmes. This is relevant for public health – both in planning health service provision and designing effective strategies to prevent illness. It should also prompt a call for research into more effective immunisation strategies and vaccination schedules. Our findings suggest that protection against pneumonia may be improved by a more frequent, perhaps even annual, pneumococcal vaccination schedule among older people with diabetes. The low influenza VE we observed against community-acquired LRTI, when contrasted with the large burden of infection directly and indirectly attributed to influenza, suggests scope for improved influenza immunisation among this population, for example with adjuvants. The finding of reduced

pneumococcal VE among patients with proteinuria is novel, and further research is needed to confirm the relationship between CKD and influenza and pneumococcal VE.

9.2.6 Acknowledgements

This work was supported by National Institute for Health Research [CDF 2010-03-32 to S.L.T.] and Kidney Research UK [ST2/2011 to H.I.M.]. The study funders had no role in the design nor conduct of the study, nor the collection, management, analysis nor interpretation of data, nor the preparation, review nor approval of the manuscript, nor the decision to submit it for publication. The views expressed in this publication are those of the authors and not necessarily those of the UK National Health Service, the National Institute for Health Research, the Department of Health nor Kidney Research UK. The results presented in this paper have not been published previously in whole or part, except in abstract format.

Table 9.1 Baseline description of study population

	Pneumococcal vaccine status at baseline n=190,492		Influenza vaccine status at baseline n=190,459		
	Never vaccinated n=79,476	**Vaccinated** n=111,016	**Unvaccinated** [1] n=32,552	**Currently vaccinated** n=124,130	**Residual 1–5 years** n=33,777
Age in years, median (IQR)	71 (66–77)	72 (66–78)	71 (66–77)	72 (66–78)	71 (66–78)
	n (%)	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Female gender	40,308 (50.7)	53,146 (47.9)	16,603 (51.0)	60,019 (48.4)	16,813 (49.8)
Socio-economic status[2]					
1 (least deprived)	13,701 (17.2)	19,912 (17.9)	5,618 (17.3)	22,181 (17.9)	5,809 (17.2)
2	14,666 (18.5)	19,591 (17.7)	5,799 (17.8)	22,394 (18.0)	6,058 (17.9)
3	16,156 (20.3)	23,329 (21.0)	6,567 (20.2)	25,957 (20.9)	6,956 (20.6)
4	17,758 (22.3)	25,481 (23.0)	7,312 (22.5)	28,128 (22.7)	7,789 (23.1)
5 (most deprived)	17,195 (21.6)	22,703 (20.5)	7,256 (22.3)	25,470 (20.5)	7,165 (21.2)
Residential care	1,697 (2.1)	3,274 (3.0)	436 (1.3)	3,516 (2.8)	1,016 (3.0)
Smoking status					
Non-smoker	38,078 (47.9)	44,713 (40.3)	15,449 (47.5)	52,782 (42.5)	14,543 (43.1)
Current smoker	13,901 (17.5)	16,439 (14.8)	6,252 (19.2)	18,327 (14.8)	5,756 (17.0)
Ex-smoker	27,497 (34.6)	49,864 (44.9)	10,851 (33.3)	53,021 (42.7)	13,478 (39.9)
Co-morbidities					
Ischaemic heart disease	18,886 (23.8)	34,415 (31.0)	6,825 (21.0)	36,761 (29.6)	9,713 (28.8)
Congestive cardiac failure	5,935 (7.5)	10,018 (9.0)	2,175 (6.7)	10,721 (8.6)	3,065 (9.1)
Hypertension	46,626 (58.7)	71,311 (64.2)	18,644 (57.3)	78,252 (63.0)	21,024 (62.2)
Cerebrovascular disease	9,714 (12.2)	14,469 (13.0)	3,612 (11.1)	16,021 (12.9)	4,540 (13.4)
Other dementia	1,437 (1.8)	1,956 (1.8)	322 (1.0)	2,326 (1.9)	729 (2.2)
Chronic lung disease	4,016 (5.1)	10,881 (9.8)	1,515 (4.7)	10,530 (8.5)	2,851 (8.4)
Chronic liver disease	402 (0.5)	734 (0.7)	168 (0.5)	729 (0.6)	240 (0.7)
Steroid use in previous 3 months	2,870 (3.6)	5,560 (5.0)	1,039 (3.2)	5,841 (4.7)	1,544 (4.6)
Latest HbA1C status					
None recorded	11,202 (14.1)	10,620 (9.6)	4,872 (15.0)	13,317 (10.7)	3,627 (10.7)
Good <7%	34,669 (43.6)	53,305 (48.0)	13,621 (41.8)	58,741 (47.3)	15,596 (46.2)
Intermediate 7-10%	27,935 (35.2)	41,354 (37.3)	11,389 (35.0)	45,383 (36.6)	12,509 (37.0)
Poor >10%	5,670 (7.1)	5,737 (5.2)	2,670 (8.2)	6,689 (5.4)	2,045 (6.1)
Antidiabetes medication history					
None	38,755 (48.8)	50,517 (45.5)	16,463 (50.6)	58,195 (46.9)	14,598 (43.2)
Oral	33,623 (42.3)	46,949 (42.3)	13,613 (41.8)	51,914 (41.8)	15,031 (44.5)
Insulin	2,889 (3.6)	4,136 (3.7)	1,024 (3.2)	4,687 (3.8)	1,314 (3.9)
Oral and insulin	4,209 (5.3)	9,414 (8.5)	1,452 (4.5)	9,334 (7.5)	2,834 (8.4)
Latest eGFR					
<30	2,098 (2.6)	2,986 (2.7)	767 (2.4)	3,337 (2.7)	977 (2.9)
30-44	7,558 (9.5)	10,607 (9.6)	2,932 (9.0)	11,964 (9.6)	3,254 (9.6)
45-59	18,678 (23.5)	25,508 (23.0)	7,337 (22.5)	29,039 (23.4)	7,806 (23.1)
≥60	51,142 (64.4)	71,915 (64.8)	21,516 (66.1)	79,790 (64.3)	21,740 (64.4)
History of proteinuria					
No	71,095 (89.5)	94,128 (84.8)	29,231 (89.8)	107,212 (86.4)	28,735 (85.1)
Yes	8,381 (10.6)	16,888 (15.2)	3,321 (10.2)	16,918 (13.6)	5,042 (14.9)

HbA1C, glycated haemoglobin; eGFR, estimated glomerular filtration rate ml/min/1.73m^2

1. Not vaccinated within the 5 previous years.

2. Index of multiple deprivation for primary care practice location

Table 9.2 Pneumococcal vaccine effectiveness against pneumonia (n=190,492)

		Pneumococcal vaccination status			
		Never	**< 1 year**	**1–4 years**	**≥ 5 years**
Person-time (years)		189,776	51,397	275,841	294,484
Infections (n)		1,661	326	2,255	3,563
Crude pneumonia rate /1,000 person-years (95% CI)		9.0 (8.6–9.5)	6.6 (5.9–7.3)	8.7 (8.3–9.1)	13.6 (13.1–14.1)
Adjusted[1] pneumonia rate ratio (95% CI)		1 (reference)	0.78 (0.69–0.89)	0.92 (0.85–0.99)	1.03 (0.95–1.11)
Vaccine effectiveness [1] % (95% CI)		0 (reference)	22 (11–31)	8 (1–15)	-3 (-11–5)
Vaccine effectiveness[1] % (95% CI) stratified by eGFR (ml/min/1.73m^2)	eGFR <30	0 (reference)	6 (-40–37)	4 (-22–25)	6 (-19–26)
	eGFR 30–44	0 (reference)	16 (-12–37)	1 (-18–17)	-7 (-27–11)
	eGFR 45–59	0 (reference)	21 (-1–38)	9 (-6–21)	-1 (-17–14)
	eGFR ≥60	0 (ref)	26 (11–39)	12 (2–22)	-3 (-15–8)
	P (test for trend) [2]	—	0.25	0.49	0.07
Vaccine effectiveness[1] % (95% CI) stratified by proteinuria status	No proteinuria	0 (ref)	28 (16–38)	13 (5–20)	1 (-8–10)
	Proteinuria	0 (ref)	2 (-25–23)	-6 (-23–9)	-19 (-38– -3)
	P (interaction) [3]	—	0.04	0.03	0.04

1. Adjusted for: age, sex, socio-economic status at practice level, residential care, date post 1 April 2004, smoking status, time-updated co-morbidities (ischaemic heart disease, congestive cardiac failure, hypertension, cerebrovascular disease, other dementia, chronic lung disease, chronic liver disease), time-updated CKD status (eGFR, proteinuria), steroid use in the 3 months prior to study entry, influenza vaccination status, and HbA1C and diabetic medication history at baseline.
2. Wald test for interaction term of pneumococcal vaccine with eGFR.

Table 9.3 LRTI rates and influenza vaccine effectiveness to prevent LRTI by season (n=190,459)

		Summer			Winter		
		Influenza vaccination status			Influenza vaccination status		
		>5 years/ never	Current	Residual 1-5 years	>5 years/ never	Current	Residual 1-5 years
Person-time (years)		35,233	219,456	74,704	47,352	355,766	70,718
Infections (n)		2,363	22,726	8,496	5,751	62,077	12,900
Crude LRTI rate /1000 py (95% CI)		73.0 (69.5–76.5)	111.0 (109.2–112.9)	121.3 (118.4–124.1)	134.8 (130.3–139.2)	187.2 (185.4–189.4)	195.5 (191.6–199.4)
Crude LRTI rate ratio (95% CI)		1 (reference)	1.52 (1.45–1.60)	1.66 (1.58–1.75)	1 (reference)	1.38 (1.34–1.44)	1.45 (1.40–1.51)
Adjusted[1] LRTI rate ratio (95% CI)							
Overall		1 (reference)	1.28 (1.21–1.35)	1.39 (1.32–1.47)	1 (reference)	1.19 (1.15–1.23)	1.23 (1.18–1.28)
Stratified by	eGFR <30	1 (reference)	1.20 (0.94–1.52)	1.29 (1.01–1.65)	1 (reference)	1.17 (0.99–1.38)	1.20 (1.01–1.43)
eGFR	eGFR 30–44	1 (reference)	1.27 (1.10–1.45)	1.31 (1.13–1.51)	1 (reference)	1.20 (1.10–1.32)	1.20 (1.08–1.33)
ml/min/1.73m^2	eGFR 45–59	1 (reference)	1.32 (1.19–1.46)	1.40 (1.26–1.56)	1 (reference)	1.16 (1.09–1.25)	1.20 (1.12–1.30)
	eGFR ≥ 60	1 (reference)	1.27 (1.19–1.37)	1.42 (1.32–1.53)	1 (reference)	1.21 (1.15–1.27)	1.26 (1.20–1.33)
Stratified by	No proteinuria	1 (reference)	1.27 (1.20–1.35)	1.36 (1.30–1.45)	1 (reference)	1.18 (1.13–1.23)	1.22 (1.16–1.27)
proteinuria	Proteinuria	1 (reference)	1.33 (1.19–1.49)	1.50 (1.33–1.68)	1 (reference)	1.23 (1.14–1.33)	1.27 (1.17–1.38)
Ratio of incidence rate ratios[1] winter/summer (95% CI)							
Overall					1 (reference)	0.93 (0.88–0.97)	0.88 (0.83–0.93)
Stratified by	eGFR <30				1 (reference)	0.93 (0.73–1.17)	0.90 (0.70–1.16)
eGFR	eGFR 30–44				1 (reference)	0.88 (0.77–1.01)	0.86 (0.75–1.00)
ml/min/1.73m^2	eGFR 45–59				1 (reference)	0.90 (0.81–0.99)	0.87 (0.78–0.98)
	eGFR ≥ 60				1 (reference)	0.95 (0.89–1.02)	0.89 (0.83–0.96)
Stratified by	No proteinuria				1 (reference)	0.93 (0.88–0.99)	0.90 (0.84–0.95)
proteinuria	Proteinuria				1 (reference)	0.90 (0.81–1.00)	0.83 (0.74–0.93)
VE[1] based on ratio of incidence rate ratios % (95% CI)							
Overall					0 (reference)	7 (3–12)	12 (7–17)
	eGFR <30				0 (reference)	7 (-17–27)	10 (-16–30)
	eGFR 30–44				0 (reference)	12 (-1–23)	14 (0–25)

Stratified by eGFR ml/min/1.73m^2	eGFR 45–59	0 (reference)	10 (1–19)	13 (2–22)
	eGFR ≥ 60	0 (reference)	5 (-2–11)	11 (4–17)
	P (test for trend) [2]	—	0.31	0.79
Stratified by proteinuria	No proteinuria	0 (reference)	7 (1–12)	10 (5–16)
	Proteinuria	0 (reference)	10 (0–19)	17 (7–26)
	P (interaction) [3]	—	0.56	0.26

LRTI, lower respiratory tract infection; py, person-years; VE, vaccine effectiveness

1. Adjusted for: age, sex, socio-economic status at practice level, residential care, date post 1 April 2004, smoking status, time-updated co-morbidities (ischaemic heart disease, congestive cardiac failure, hypertension, cerebrovascular disease, other dementia, chronic lung disease, chronic liver disease), time-updated CKD status (eGFR, proteinuria), steroid use in the 3 months prior to study entry, pneumococcal vaccination, and HbA1C and diabetic medication history at baseline.

2. Wald test for interaction of eGFR with both influenza vaccination status and season, with eGFR modelled as a linear variable.

3. Wald test for interaction of proteinuria with both influenza vaccination status and season

S 1 Association of pneumococcal vaccine status with community-acquired pneumonia incidence before and after 1 April 2003 (n=190,492)

Adjusted [1] pneumonia rate ratio (95% CI)	Pneumococcal vaccination status			
	Never	< 1 year	1–4 years	≥ 5 years
1 April 1997 – 31 March 2003	1 (ref)	0.76 (0.59–0.99)	0.88 (0.75–1.02)	0.97 (0.79–1.18)
1 April 2003 – 31 March 2011	1 (ref)	0.78 (0.68–0.90)	0.92 (0.84–1.01)	1.03 (0.94–1.12)

1. Adjusted for: age, sex, socio-economic status at practice level, residential care, smoking status, time-updated co-morbidities (ischaemic heart disease, congestive cardiac failure, hypertension, cerebrovascular disease, other dementia, chronic lung disease, chronic liver disease), time-updated CKD status (eGFR, proteinuria), steroid use in the 3 months prior to study entry, influenza vaccination status, and HbA1C and diabetic medication history at baseline.

S 2 Association of influenza vaccine status with community-acquired LRTI incidence for each winter (n=190,459)

Year	Number of LRTIs in winter	Adjusted [1] LRTI incidence rate ratio (95% CI) according to influenza vaccination status in winter [2]		
		>5 years/ never	Current	Residual 1-5 years
1997-8	620	1 (ref)	1.33 (1.03–1.72)	1.45 (1.08–1.94)
1998-9	1,137	1 (ref)	1.45 (1.22–1.77)	1.37 (1.11–1.70)
1999-2000	1,821	1 (ref)	1.23 (1.06–1.42)	1.24 (1.05–1.47)
2000-1	2,583	1 (ref)	1.21 (1.05–1.39)	1.40 (1.17–1.66)
2001-2	3,449	1 (ref)	1.33 (1.16–1.53)	1.44 (1.22–1.69)
2002-3	4,641	1 (ref)	1.18 (1.04–1.34)	1.30 (1.13–1.50)
2003-4	6,266	1 (ref)	1.35 (1.20–1.51)	1.52 (1.33–1.73)
2004-5	7,656	1 (ref)	1.23 (1.11–1.37)	1.19 (1.05–1.34)
2005-6	8,084	1 (ref)	1.14 (1.01–1.28)	1.14 (1.00–1.28)
2006-7	8,666	1 (ref)	1.17 (1.05–1.32)	1.09 (0.97–1.23)
2007-8	9,134	1 (ref)	1.12 (0.99–1.25)	1.20 (1.07–1.35)
2008-9	10,061	1 (ref)	1.24 (1.11–1.38)	1.25 (1.11–1.40)
2009-10	8,945	1 (ref)	1.33 (1.18–1.49)	1.42 (1.25–1.61)
2010-11	7,661	1 (ref)	1.12 (0.99–1.26)	1.16 (1.02–1.31)

1. Adjusted for: age, sex, socio-economic status at practice level, residential care, smoking status, time-updated co-morbidities (ischaemic heart disease, congestive cardiac failure, hypertension, cerebrovascular disease, other dementia, chronic lung disease, chronic liver disease), time-updated CKD status (eGFR, proteinuria), steroid use in the 3 months prior to study entry, pneumococcal vaccination, and HbA1C and diabetic medication history at baseline.
2. Winter was defined as 1 September to 31 March.

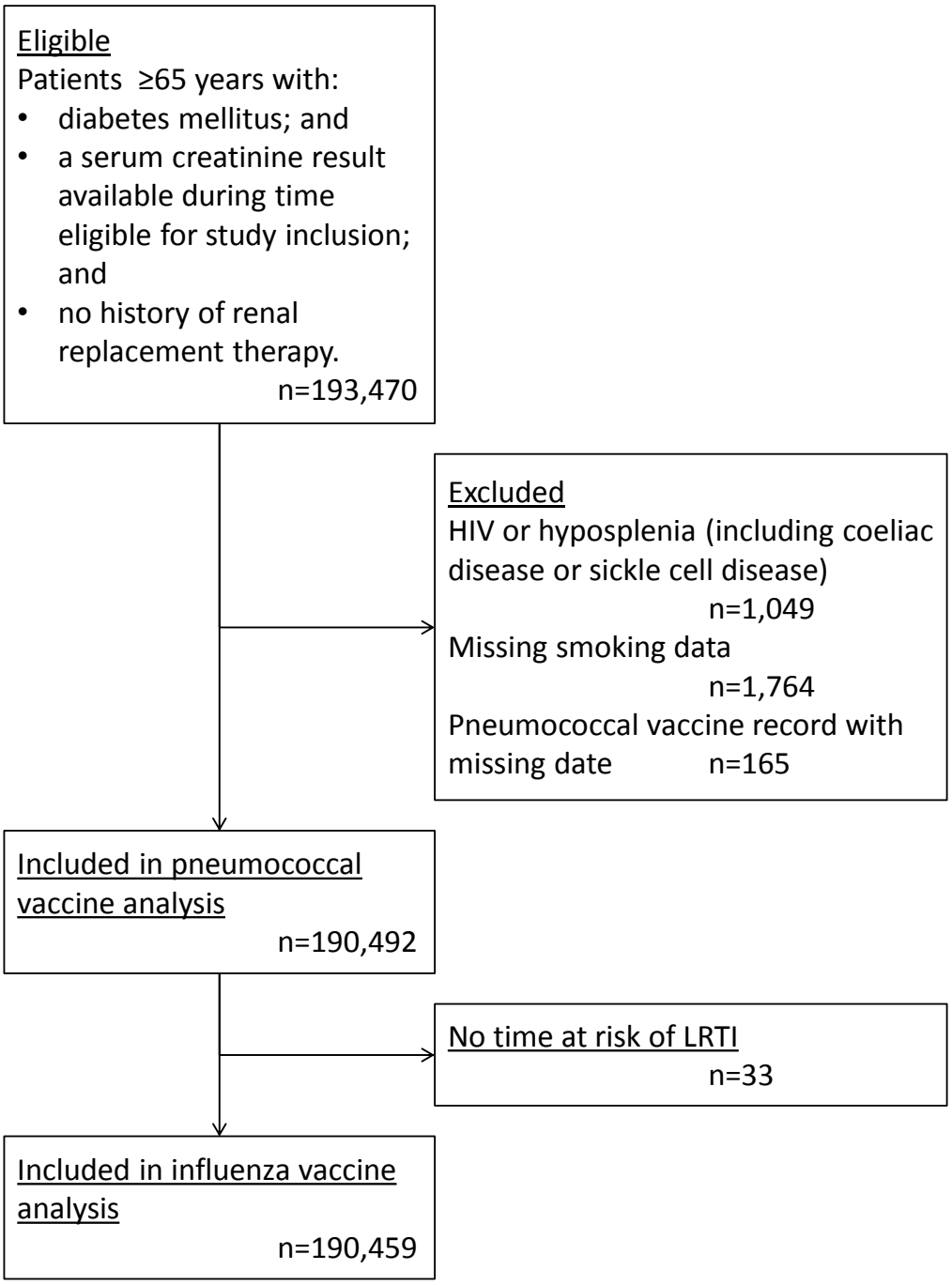
S 3 Adjusted LRTI rate ratio according to influenza vaccination status excluding patients with congestive heart failure or chronic lung disease (n=162,412)

	Summer			Winter		
	Influenza vaccination status			Influenza vaccination status		
	>5 years/never	Current	Residual 1-5 years	>5 years/never	Current	Residual 1-5 years
Person-time (years)	30,904	182,214	61,928	41,582	295,921	58,684
Infections	1,578	13,587	5,020	3,976	39,304	8,054
Adjusted[1] LRTI rate ratio (95% CI)	1 (ref)	1.32 (1.24–1.41)	1.42 (1.32–1.51)	1 (ref)	1.23 (1.17–1.28)	1.25 (1.19–1.31)
Ratio of incidence rate ratios[1] winter/summer (95% CI)				1 (ref)	0.95 (0.89–1.01)	0.90 (0.84–0.96)
VE[1] based on ratio of incidence rate ratios % (95% CI)				0 (ref)	5 (-1–11)	10 (4–16)

LRTI, lower respiratory tract infection; VE, vaccine effectiveness

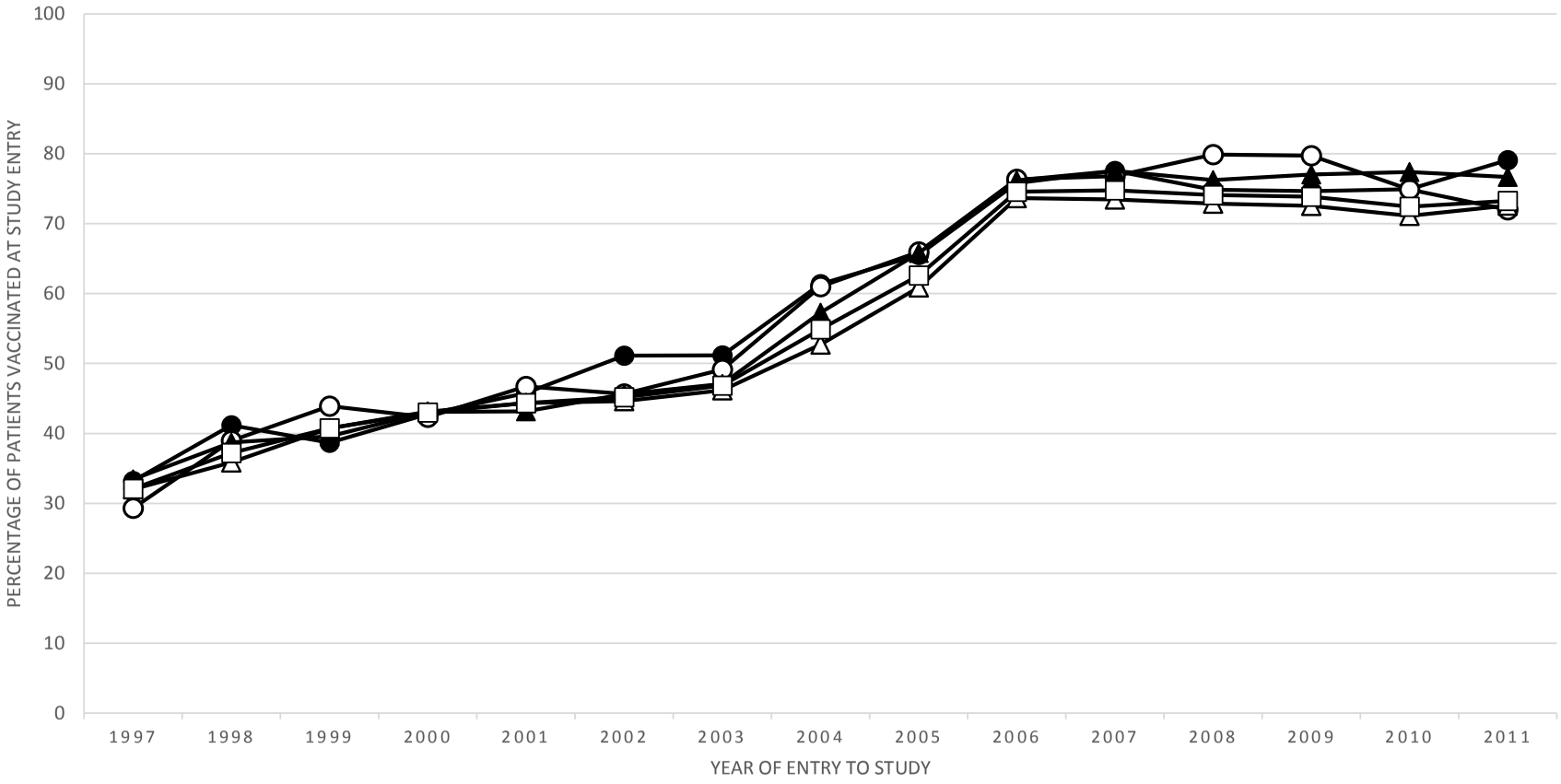
1. Adjusted for: age, sex, socio-economic status at practice level, residential care, date post 1 April 2004, smoking status, time-updated co-morbidities (ischaemic heart disease, hypertension, cerebrovascular disease, other dementia, chronic liver disease), time-updated CKD status (eGFR, proteinuria), steroid use in the 3 months prior to study entry, pneumococcal vaccination, and HbA1C and diabetic medication history at baseline.

Figure 9.1 Flowchart of study inclusion



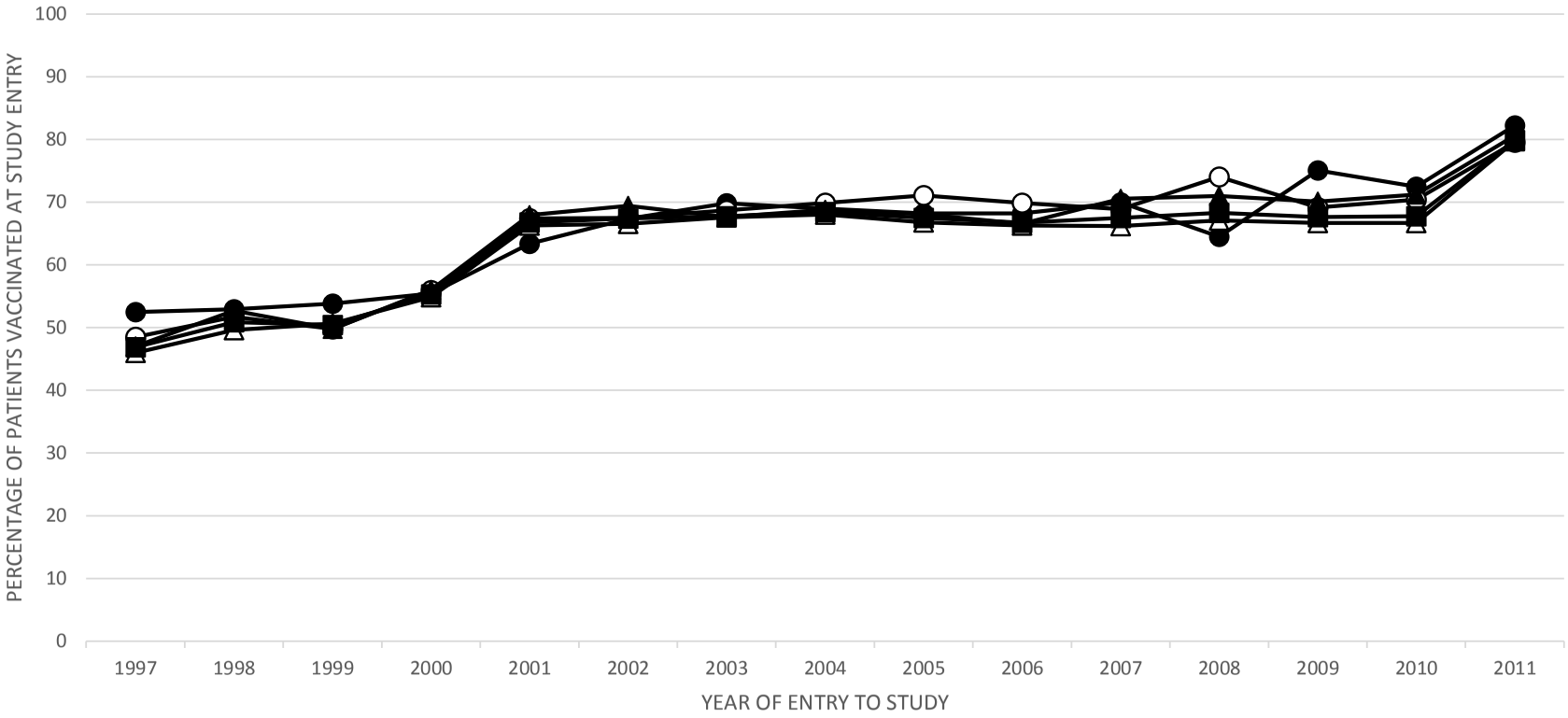
HIV, human immunodeficiency virus infection; LRTI, lower respiratory tract infection

Figure S 1Pneumococcal vaccination status at baseline according to year of study entry and estimated glomerular filtration rate status (eGFR)



Black circles, eGFR<30; white circles, eGFR 30-44; black triangles, eGFR 45-59; white triangles, eGFR ≥60, black squares, overall.

Figure S 2 Influenza vaccination status at baseline according to year of study entry and estimated glomerular filtration rate status (eGFR)



Black circles, eGFR<30; white circles, eGFR 30-44; black triangles, eGFR 45-59; white triangles, eGFR ≥60, black squares, overall.

Chapter 10. The association of markers of chronic kidney disease with mortality following community-acquired pneumonia and sepsis

This chapter presents a study of the association of chronic kidney disease (CKD) with short-term mortality following community-acquired sepsis and pneumonia (objective 5). The results of two sensitivity analyses discussed in the article are presented in **10.3**, followed by a brief additional discussion of the identification of CKD status in this study.

10.1 Introduction to Paper 5

This paper presents a retrospective cohort study to identify the risk of mortality following a first diagnosis of community-acquired pneumonia or sepsis according to markers of CKD among patients aged ≥65 years with diabetes mellitus, a valid serum creatinine result, and no history of renal replacement therapy. To ensure good ascertainment of mortality the study population was restricted to patients in CPRD who had data linkage available to an updated linked ONS mortality dataset. Identification of the study population was described in **3.9.3.**

The primary outcome was 28-day all-cause mortality following diagnosis of pneumonia or sepsis. The methods used to identify episodes of community-acquired pneumonia and sepsis were detailed in **4.1** and **4.2**. The exposure of interest was CKD, identified by a history of proteinuria or reduced estimated glomerular filtration rate (eGFR) as described in **Chapter 5**. As CKD status was not time-updated over the 28 day follow-up period, eGFR status was identified using the best-of-two method as described in **5.4.5**.

Mortality risk ratios were calculated to estimate the risk of mortality according to CKD status. The calculation of risk ratios rather than rate ratios reflects the selected study question: whether CKD was associated with a difference in the risk of death at any point within 28 days after infection, as opposed to whether CKD was associated with differences in timing of mortality within the 28 days following infection.

Mortality risk ratios were adjusted for *a priori* potential confounders of the association between markers of CKD and post-infection mortality. Detailed definitions of these confounders are described in **Chapter 6**.

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	Helen McDonald
Principal Supervisor	Dr Dorothea Nitsch
Thesis Title	The epidemiology of infections among older people with diabetes mellitus and chronic kidney disease

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?	Nephrology Dialysis Transplantation		
When was the work published?	2015		
If the work was published prior to registration for your research degree, give a brief rationale for its inclusion	N/A. Re copyright: this is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial Licence (CC-BY-NC).		
Have you retained the copyright for the work?*	Choose an item.	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)	The original idea for this study was conceived by Dorothea Nitsch and Sara Thomas, who wrote the study proposal and obtained ethics approval and funding for the study. Sara Thomas obtained the study data from CPRD. I developed a detailed study design and planned the analysis with advice and supervision from Sara Thomas and

	Dorothea Nitsch. Sara Thomas provided the main supervision for this study. The study used the dataset of community-acquired infections among study population B which I had constructed for objective 2 using methods developed by Elizabeth Millett, as presented in Chapter 4. I developed an approach to identifying markers of chronic kidney disease in CPRD (described in Chapter 5), with advice and supervision from Dorothea Nitsch and Sara Thomas. I extracted, cleaned and classified the data to define other variables as described in Chapter 6, and was fortunate to be able to use codelists developed by colleagues at LSHTM (each attributed individually in Chapter 6). Body mass index data was extracted, cleaned and classified using a do-file written by Krishnan Bhaskaran. I conducted all data analysis, and drafted the manuscript, which was then commented on by all co-authors. The manuscript was peer-reviewed, and I incorporated suggestions from reviewers into the final manuscript.

Student Signature: KMcDonald **Date:** 10 April 2015

Supervisor Signature: Dorothea Nitsch **Date:** 14/4/15

NDT Advance Access published January 20, 2015

Nephrol Dial Transplant (2015) 0: 1–8
doi: 10.1093/ndt/gfu401

Original Article

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

Helen I. McDonald[1], Dorothea Nitsch[1], Elizabeth R. C. Millett[2], Alan Sinclair[3] and Sara L. Thomas[2]

[1]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK, [2]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK and [3]Institute of Diabetes for Older People (IDOP), University of Bedfordshire, Luton, Bedfordshire LU1 3JU, UK

Correspondence and offprint requests to: Helen I. McDonald; E-mail helen.mcdonald@lshtm.ac.uk

ABSTRACT

Background. We aimed to examine whether pre-existing impaired estimated glomerular filtration rate (eGFR) and proteinuria were associated with mortality following community-acquired pneumonia or sepsis among people aged ≥65 years with diabetes mellitus, without end-stage renal disease.

Methods. Patients were followed up from onset of first community-acquired pneumonia or sepsis episode in a cohort study using large, linked electronic health databases. Follow-up was for up to 90 days, unlimited by hospital discharge. We used generalized linear models with log link, normal distribution and robust standard errors to calculate risk ratios (RRs) for all-cause 28- and 90-day mortality according to two markers of chronic kidney disease: eGFR and proteinuria.

Results. All-cause mortality among the 4743 patients with pneumonia was 29.6% after 28 days and 37.4% after 90 days. Among the 1058 patients with sepsis, all-cause 28- and 90-day mortality were 35.6 and 44.2%, respectively. eGFR <30 mL/min/1.73 m^2 was a risk marker of higher 28-day mortality for pneumonia (RR 1.27: 95% CI 1.12–1.43) and sepsis (RR 1.32: 95% CI 1.07–1.64), adjusted for age, sex, socio-economic status, smoking status and co-morbidities. Neither moderately impaired eGFR nor proteinuria were associated with short-term mortality following either infection.

Conclusions. People with pre-existing low eGFR but not on dialysis are at higher risk of death following pneumonia and sepsis. This association was not explained by existing co-morbidities. These patients need to be carefully monitored to prevent modifiable causes of death.

Keywords: chronic kidney disease, community-acquired infections, electronic health records, infection/mortality, proteinuria

INTRODUCTION

Chronic kidney disease (CKD) affects an estimated 1.8 million people in England, 98% of whom do not require renal replacement therapy [1]. CKD is defined by reduced estimated glomerular filtration rate (eGFR) or evidence of kidney damage such as proteinuria and is commonest among older people [2, 3]. Most patients with CKD are managed in primary care [4].

Infection is an important cause of mortality among older people [5, 6]. Both reduced eGFR and proteinuria are associated with an increased rate of infection-related mortality, which could be partly explained by increased incidence of infection [7–9]. It is less clear whether CKD is also associated with poorer prognosis following infection. When clinicians assess patients with community-acquired infection, developing complications such as acute kidney injury (AKI) may not yet be apparent, but they will know which patients have pre-existing CKD. If pre-existing CKD is a risk marker for short-term mortality, this

would be useful for risk stratification and clinical management of patients with infections, especially in primary care where clinicians may not have access to immediate laboratory tests. While the implications of acute changes in eGFR during infection are a focus of current research, few studies have investigated the role of pre-existing CKD [10]. Low baseline eGFR has been found to be associated with mortality following sepsis and community-acquired pneumonia, but rarely examined according to clinically meaningful categories of eGFR [11–15]. To the best of our knowledge, proteinuria has not been examined as a potential risk marker for mortality following infection [12, 15, 16].

Among older people, CKD frequently co-exists with other co-morbidities [3]. An association of CKD with poor prognosis of infections could thus be due to confounding from these co-morbidities. For example, CKD is strongly associated with cardiovascular disease, which may be complicated by infection, resulting in post-infection mortality driven by the underlying cardiovascular disease [17]. Such deaths would largely follow hospitalization for cardiovascular events. CKD is associated with healthcare-associated pneumonia, which carries a worse prognosis than community-acquired pneumonia [18, 19]. Focusing on community-acquired infections should exclude infections arising as short-term sequelae of cardiovascular events and improve understanding of the relationship between pre-existing CKD markers and infection prognosis.

Older people with diabetes mellitus form a large and growing population in primary care who suffer a high incidence of community-acquired pneumonia and sepsis [20]. Forty per cent of adults with diabetes have CKD, of whom three-quarters have proteinuria, and CKD among these patients is associated with a greater all-cause excess mortality than among patients without diabetes [21]. If proteinuria is a risk marker for mortality among older patients with diabetes who develop community-acquired pneumonia or sepsis, this could inform clinical management of a large primary care patient population with appreciable mortality following infection [20].

This study aimed to examine whether baseline eGFR and proteinuria were independent risk markers for short-term mortality following community-acquired pneumonia or sepsis among older people with diabetes mellitus, using large, linked electronic health record databases.

MATERIALS AND METHODS

Data sources

The Clinical Practice Research Datalink (CPRD) is an anonymized UK dataset, comprising primary care records (including diagnoses, prescriptions and test results) for 12.8 million patients in May 2011 when data were extracted. The CPRD population is representative of the general UK population and validity of recorded diagnoses is generally high [22, 23]. Monitoring of eGFR and proteinuria in primary care is standard practice for people with diabetes and has been financially incentivized by the Quality Outcomes Framework since April 2004 [24].

Data linkage is available within England subject to practice-level consent. Records of all patients in CPRD with available linkage to Office for National Statistics (ONS) mortality data formed the study dataset [25]. We additionally used linked Hospital Episodes Statistics admissions data, which were available for all patients [26].

Study population

The study population was a subset of a population described in more detail previously [20]. It comprised people aged ≥65 years with diabetes mellitus who experienced a first community-acquired pneumonia or sepsis, with a valid serum creatinine result and no history of renal replacement therapy.

A valid serum creatinine result was one recorded in primary care after the latest time-point of diabetes diagnosis, 65th birthday, 1 year after patients' practice registration, date the practice reached CPRD quality control standards or 1 January 1998. Study exit occurred at the first time-point of death, patient leaving the practice, last data collection from the practice, last ONS data linkage date, renal replacement therapy (kidney transplant or dialysis) or 31 March 2011. Patients with a history of renal replacement therapy were excluded.

Definition of infections

Infection was identified by a diagnostic Read code in primary care records, or a diagnostic International Classification of Disease 10 (ICD-10) code as the primary cause of hospital admission in secondary care records. The first consultation for infection was treated as the date of infection onset. Any community-acquired infection with onset at least 28 days after the first valid serum creatinine result, and before study exit, was included in the study.

Hospital-acquired infections were identified and excluded as described previously [20]. Briefly, infections were designated as hospital-acquired if onset was during or within 14 days of discharge from a hospitalization. Hospital-acquired infections continued until 28 days had passed without a diagnostic code for the infection or 28 days after hospital discharge, whichever was the later. After this, patients re-entered follow-up for community-acquired infection.

Study outcomes

The outcomes were death from any cause recorded in ONS mortality data within 28 days (primary outcome) or within 90 days (secondary outcome) of infection onset.

Definition of CKD

CKD was described in terms of eGFR and proteinuria, using primary care records. We estimated eGFR from serum creatinine test results using the CKD Epidemiology Collaboration (CKD-EPI) equation including adjustment for ethnicity [27]. We excluded serum creatinine results <28 days prior to infection onset to avoid misclassification of CKD status, as a developing infection could disrupt serum creatinine levels.

Clinically, CKD diagnosis is based on two GFR estimates at least 3 months apart [2]. Using a single GFR estimate can result in over-ascertainment of CKD due to creatinine fluctuation [28]. If more than one serum creatinine result was recorded between the start of patient follow-up and 28 days prior to infection onset, we used the higher eGFR from the latest two

results that were at least 3 months apart, to obtain conservative estimates of eGFR [28].

We aimed to categorize eGFR according to thresholds corresponding to those used in diagnosing CKD stage. Due to the small number of outcomes in the category eGFR < 15 mL/min/1.73 m^2, we collapsed Stages 4 and 5 to categorize eGFR as < 30, 30–44, 45–59 and ≥60 mL/min/1.73 m^2 [2].

A history of proteinuria was defined by either a positive urine protein test result (excluding results on the same day as a urinary tract infection diagnosis) or a diagnosis of proteinuric renal disease. We did not count trace results as positive.

Other variables

Age was defined in 5-year age-bands up to a final category of ≥85 years. Socio-economic status was assigned by quintile at an individual level, using 2007 ONS estimates of the Index of Multiple Deprivation, a composite area-level marker of deprivation [25]. If this was not available, it was supplemented by the socio-economic status for the patient's primary care practice. Smoking status was defined by the most recent record before infection onset when available, otherwise by the first subsequent record. Non-cardiovascular co-morbidities (chronic lung disease, dementia, cancer, connective tissue disorders, hypertension and cerebrovascular disease) and cardiovascular co-morbidities (congestive heart failure and ischaemic heart disease) were defined by diagnostic CPRD Read codes, and diabetic medication history by CPRD prescription records prior to infection onset. HbA1C test results within 28 days before infection were excluded as these could reflect disturbed glycaemic control during the early stages of infection.

Data analysis

Pneumonia and sepsis analyses were conducted separately. A patient could be included in both the pneumonia and sepsis analyses if they experienced both infections, but not for multiple episodes of either pneumonia or sepsis. We excluded patients with no smoking status or HbA1C result available.

We described mortality using Kaplan–Meier survival curves stratified by eGFR status. We calculated risk ratios (RRs) using a generalized linear model with log link, normal distribution and robust standard errors, according to a pre-specified analysis plan [29]. Our first model adjusted for age, sex, socio-economic status and infection onset prior to 1 April 2004 (when Quality Outcomes Framework guidelines financially incentivizing recording of CKD status among people with diabetes in primary care were introduced) [24]. Our second model adjusted for confounding by smoking status, characteristics of diabetes (HbA1C and diabetic medication history) and non-cardiovascular co-morbidities. Our final model additionally adjusted for congestive heart failure and ischaemic heart disease, which could confound or mediate an association between CKD and post-infection mortality. We repeated the final model with additional adjustment for peripheral vascular disease as a sensitivity analysis. We looked for effect modification between eGFR and proteinuria in the final model.

We focused on whether pre-existing CKD was a risk marker for short-term mortality following infection. Data on acute electrolyte changes during infection were not routinely available, and so potential causal mechanisms such as AKI could not be explored [30].

Causes of death in ONS mortality data are recorded using ICD-10 codes from 1 January 2001, and ICD-9 codes prior to this, which are not easily comparable [31]. We therefore described cause of death among patients who died after 1 January 2001.

Stata version 13.1 was used for data analysis. All code lists are available on request.

Ethics

The study was approved by the Independent Scientific Advisory Group of the CPRD (ISAC reference 11_033A) and the London School of Hygiene & Tropical Medicine Ethics Committee (LSHTM reference 6116).

RESULTS

We identified 4957 individuals with community-acquired pneumonia and 1114 individuals with community-acquired sepsis. Data were missing (for smoking status and/or HbA1C results) for 212/4957 individuals with pneumonia (4.3%) and 56/1114 individuals with sepsis (5.0%). These patients were excluded. Among patients with pneumonia, patients with missing data were older (median age 83 years, IQR: 78–88) compared with those included (median age 80 years, IQR: 74–85), with a higher 28-day mortality (excluded 88/214, 41.1%: included 1406/4743, 29.6%) but a similar distribution of baseline eGFR. A similar pattern was seen for sepsis.

Estimated GFR was based on the higher of two results for 4029 patients with pneumonia (84.9%) and 919 patients with sepsis (86.9%); for the remaining patients, only a single valid serum creatinine result was available. CKD prevalence was high: almost half of the patients had eGFR <60 mL/min/1.73 m^2 and a third had proteinuria. Patients with eGFR <60 mL/min/1.73 m^2 were older, with a higher prevalence of ischaemic heart disease and congestive heart failure than patients with eGFR ≥60 mL/min/1.73 m^2 (Table 1).

Patients with pneumonia experienced 29.6% 28-day all-cause mortality (1406 deaths). Patients with sepsis experienced 35.6% 28-day all-cause mortality (377 deaths) (Table 2). Survival curves showed high mortality at infection onset, declining over ~30 days to a more stable rate for the next 60 days following both pneumonia and sepsis (Figure 1). RRs for 28-day mortality were higher among people with eGFR <30 mL/min/1.73 m^2 compared with people with eGFR ≥60 mL/min/1.73 m^2 for pneumonia (RR = 1.27; 95% CI 1.10–1.47) and sepsis (RR = 1.42; 1.10–1.84), adjusted for age, sex, socio-economic status and onset prior to April 2004. Adjustment for smoking status, co-morbidities and characteristics of diabetes had minimal effect on these RRs for pneumonia (fully adjusted RR = 1.27; 1.12–1.43) or sepsis (fully adjusted RR = 1.32; 1.07–1.64). There was no evidence of associations between intermediate levels of eGFR and 28-day mortality for either infection, nor for an association between proteinuria and 28-day mortality. The pattern of associations of eGFR and proteinuria with 90-day mortality was similar to those for 28-day mortality (Table 2). Results were unchanged by additional adjustment for

Table 1. Baseline characteristics of the study population

	Pneumonia		Sepsis	
	eGFR < 60 mL/min/1.73 m^2	eGFR ≥ 60 mL/min/1.73 m^2	eGFR < 60 mL/min/1.73 m^2	eGFR ≥ 60 mL/min/1.73 m^2
Age (years)				
Median (IQR)	82 (77–87)	78 (72–83)	81 (75–86)	76 (71–82)
	n (col %)	*n* (col %)	*n* (col %)	*n* (col %)
Gender				
Female	1231 (53.0)	1106 (42.0)	294 (55.7)	268 (45.7)
Onset prior to 1 April 2004	584 (25.1)	527 (20.0)	124 (23.5)	120 (20.5)
Socio-economic status (IMD quintile)[a]				
1 (least deprived)	383 (16.5)	438 (16.6)	109 (20.6)	106 (18.1)
2	554 (23.8)	595 (22.6)	116 (22.0)	127 (21.7)
3	494 (21.3)	582 (22.1)	120 (22.7)	137 (23.4)
4	502 (21.6)	574 (21.8)	110 (20.8)	108 (18.4)
5 (most deprived)	391 (16.8)	444 (16.9)	73 (13.8)	108 (18.4)
Smoking status				
Current	321 (13.8)	549 (20.9)	70 (13.3)	103 (17.6)
Ex-smoker	1185 (51.0)	1351 (51.3)	235 (44.5)	284 (48.5)
Non-smoker	776 (33.4)	698 (26.5)	207 (39.2)	193 (32.9)
Missing	42 (1.8)	35 (1.3)	16 (3.0)	6 (1.0)
Comorbidities				
Chronic lung disease	503 (21.6)	686 (26.1)	59 (11.2)	95 (16.2)
Hypertension	1623 (69.8)	1662 (63.1)	372 (70.5)	381 (65.0)
Congestive heart failure	734 (31.6)	430 (16.3)	147 (27.8)	95 (16.2)
Ischaemic heart disease	966 (41.6)	887 (33.7)	223 (42.2)	207 (35.3)
Cerebrovascular disease	692 (29.8)	638 (24.2)	152 (28.8)	145 (24.7)
Other dementia	174 (7.5)	190 (7.2)	42 (8.0)	23 (3.9)
Cancer	382 (16.4)	474 (18.0)	99 (18.8)	120 (20.5)
Connective tissue disorders	228 (9.8)	215 (8.2)	36 (6.8)	52 (8.9)
HbA1C				
Good <7%	1176 (50.6)	1401 (53.2)	251 (47.5)	308 (52.6)
Borderline 7–10%	966 (41.6)	1033 (39.2)	213 (40.3)	230 (39.3)
Poor >10%	108 (4.7)	131 (5.0)	41 (7.8)	34 (5.8)
None recorded	74 (3.2)	68 (2.6)	23 (4.4)	14 (2.4)
Prior antidiabetes medication				
Insulin	122 (5.3)	126 (4.8)	34 (6.4)	23 (3.9)
Oral medications	1139 (49.0)	1426 (54.2)	247 (46.8)	312 (53.2)
Both	462 (19.9)	384 (14.6)	130 (24.6)	114 (19.5)
None	601 (25.9)	697 (26.5)	117 (22.2)	137 (23.4)
Total	2324	2633	528	586

eGFR, estimated glomerular filtration rate.
[a]Index of multiple deprivation (IMD) score for patient's postcode where available, otherwise, practice-level IMD score.

peripheral vascular disease. There was no good evidence of effect modification between eGFR and proteinuria, and in particular no evidence of any association of proteinuria with 28-day mortality within any category of eGFR status (data not shown).

The underlying causes of death following pneumonia and sepsis were similar for patients with eGFR above and below 60 mL/min/1.73 m^2. Causes of 28-day mortality following sepsis were predominantly sources of infection (Table 3). Following pneumonia onset, pneumonia was recorded as an underlying or contributory cause of death for 83.9% (1191/1419) of deaths within 28 days and 76.3% (1366/1790) of deaths within 90 days. Among patients with eGFR <60 mL/min/1.73 m^2, renal disease was recorded as an underlying or contributory cause for 10.6% (77/724) of those who died within 28 days of pneumonia onset and 16.6% (152/913) of those who died within 90 days. Recording of CKD as a cause of death increased with lower eGFR (Table 4).

DISCUSSION

Among this population of older people with diabetes mellitus, eGFR <30 mL/min/1.73 m^2 was a risk marker of higher 28- and 90-day mortality following community-acquired pneumonia and sepsis, compared with patients with eGFR ≥60 mL/min/1.73 m^2. The relationship between eGFR and mortality did not change with adjustment for co-morbidities. Neither moderately impaired eGFR nor proteinuria was associated with higher short-term mortality following either infection.

The strengths of this study follow from the analysis of a focused question using large, linked datasets for a highly monitored primary care population with a cohort study design. Our study identifies that the association between eGFR and post-infection mortality persists when patients with end-stage renal disease (ESRD) are excluded (and is not explained by renal replacement therapy), when considering fixed-term

Table 2. Short-term mortality by CKD status (n = 4743 for pneumonia, n = 1058 for sepsis)[a]

		Number (column %)	28-Day mortality (row %)	90-Day mortality (row %)	RRs for 28-day mortality (95% CI)			RRs for 90-day mortality (95% CI)		
					Adjusted for demographics[b]	adjusted for non-CVD co-morbidities[c]	Fully adjusted[d]	Adjusted for demographics[b]	Adjusted for non-CVD co-morbidities[c]	Fully adjusted[d]
Pneumonia										
Proteinuria (adjusted for eGFR)										
Yes		1611 (34.0)	499 (31.0)	625 (38.8)	1.07 (0.97–1.17)	1.07 (0.98–1.18)	1.07 (0.98–1.18)	1.05 (0.97–1.13)	1.06 (0.98–1.14)	1.05 (0.98–1.14)
No		3132 (66.0)	907 (29.0)	1150 (36.7)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
eGFR (mL/min/1.73 m^2) (adjusted for proteinuria)										
<30	<15	23 (0.5)	12 (52.2)	12 (52.2)	1.27 (1.10–1.47)	1.31 (1.13–1.52)	1.30 (1.12–1.52)	1.26 (1.12–1.42)	1.29 (1.14–1.45)	1.27 (1.12–1.43)
	15–29	263 (5.6)	110 (41.8)	139 (52.9)						
30–44		764 (16.1)	265 (34.7)	336 (44.0)	0.98 (0.86–1.10)	0.99 (0.88–1.13)	0.99 (0.88–1.12)	1.02 (0.93–1.13)	1.04 (0.94–1.15)	1.03 (0.94–1.14)
45–60		1162 (24.5)	332 (28.6)	418 (36.0)	0.91 (0.82–1.02)	0.94 (0.84–1.04)	0.93 (0.84–1.04)	0.91 (0.83–1.00)	0.93 (0.85–1.02)	0.92 (0.84–1.01)
≥60		2531 (53.4)	687 (27.1)	870 (34.4)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Total		4743	1406 (29.6)	1775 (37.4)						
Sepsis										
Proteinuria (adjusted for eGFR)										
Yes		358 (33.8)	128 (35.8)	159 (44.4)	0.98 (0.82–1.17)	1.01 (0.85–1.21)	1.01 (0.85–1.21)	0.98 (0.86–1.13)	0.96 (0.84–1.10)	0.96 (0.84–1.10)
No		700 (66.2)	249 (35.6)	309 (44.1)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
eGFR (mL/min/1.73 m^2) (adjusted for proteinuria)										
<30	<15	8 (0.8)	2 (25.0)	5 (62.5)	1.42 (1.10–1.84)	1.41 (1.08–1.84)	1.37 (1.05–1.79)	1.39 (1.14–1.70)	1.32 (1.07–1.63)	1.32 (1.07–1.64)
	15–29	62 (5.9)	32 (51.6)	39 (62.9)						
30–44		190 (18.0)	88 (46.3)	103 (54.2)	1.25 (1.01–1.55)	1.24 (0.99–1.55)	1.24 (0.99–1.54)	1.14 (0.96–1.36)	1.12 (0.94–1.33)	1.11 (0.94–1.32)
45–60		232 (21.9)	76 (32.8)	95 (41.0)	0.95 (0.75–1.19)	0.91 (0.72–1.15)	0.91 (0.72–1.14)	0.92 (0.77–1.11)	0.89 (0.74–1.07)	0.89 (0.75–1.07)
≥60		566 (53.5)	179 (31.6)	226 (39.9)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)	1 (reference)
Total		1058	377 (35.6)	468 (44.2)						

[a]Excluding patients with missing smoking or HbA1C data.
[b]Age, gender, socio-economic status, onset prior to 1 April 2004.
[c]Age, gender, socio-economic status, onset prior to 1 April 2004, smoking status, chronic lung disease, dementia, cancer, connective tissue disorders, hypertension, cerebrovascular disease, diabetes medications, latest HbA1C.
[d]Age, gender, socio-economic status, onset prior to 1 April 2004, smoking status, chronic lung disease, dementia, cancer, connective tissue disorders, hypertension, cerebrovascular disease, diabetes medications, latest HbA1C, congestive heart failure, ischaemic heart disease.

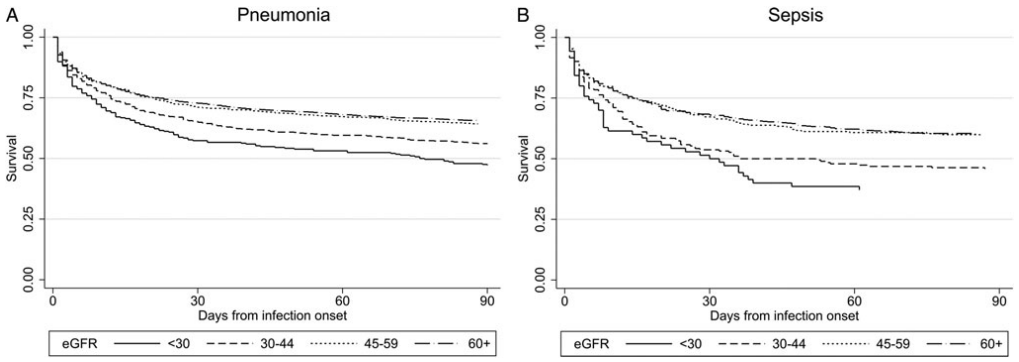


FIGURE 1: Survival curves of short-term mortality following infection onset by eGFR status for (**A**) pneumonia and (**B**) sepsis.

Table 3. Top five underlying causes of death by ICD-10 code for short-term mortality following pneumonia and sepsis (deaths after 2001)[a]

	eGFR	
	eGFR <60 mL/min/1.73 m^2	eGFR ≥60 mL/min/1.73 m^2
28-Day mortality following pneumonia, *n* = 1419	J18 Pneumonia, organism unspecified, *n* = 256, 35.4% J44 Other chronic obstructive pulmonary disease, *n* = 63, 8.7% I25 Chronic ischaemic heart disease, *n* = 47, 6.5% I50 Heart failure, *n* = 43, 5.9% I64 Stroke, not specified as haemorrhage or infarction, *n* = 37, 5.1% Total = 724	J18 Pneumonia, organism unspecified, *n* = 216, 31.1% J44 Other chronic obstructive pulmonary disease, *n* = 63, 9.1% I64 Stroke, not specified as haemorrhage or infarction, *n* = 35, 5.0% C34 Malignant neoplasm of bronchus and lung, *n* = 34, 4.9% I25 Chronic ischaemic heart disease, *n* = 29, 4.2% Total = 695
29–90 Day mortality following pneumonia, *n* = 371	J18 Pneumonia, organism unspecified, *n* = 38, 20.1% I25 Chronic ischaemic heart disease, *n* = 24, 12.7% J44 Other chronic obstructive pulmonary disease, *n* = 13, 6.9% C34 Malignant neoplasm of bronchus and lung, *n* = 12, 6.4% I21 Acute myocardial infarction, *n* = 12, 6.4% Total = 189	J18 Pneumonia, organism unspecified, *n* = 20, 11.0% C34 Malignant neoplasm of bronchus and lung, *n* = 20, 11.0% J44 Other chronic obstructive pulmonary disease, *n* = 19, 10.4% I25 Chronic ischaemic heart disease, *n* = 9, 5.0% I64 Stroke, not specified as haemorrhage or infarction, *n* = 9, 5.0% Total = 182
28-Day mortality following sepsis, *n* = 387	N39 Other disorders of urinary system[b], *n* = 33, 15.9% J18 Pneumonia, organism unspecified, *n* = 30, 14.4% E14 Unspecified diabetes mellitus, *n* = 12, 5.8% A41 Other sepsis, *n* = 11, 5.3% L03 Cellulitis, *n* = 11, 5.3% Total = 208	N39 Other disorders of urinary system[b], *n* = 20, 11.2% A41 Other sepsis, *n* = 18, 10.1% J18 Pneumonia, organism unspecified, *n* = 18, 10.1% E14 Unspecified diabetes mellitus, *n* = 10, 5.6% J44 Other chronic obstructive pulmonary disease, *n* = 6, 3.4% =K55 Vascular disorders of intestine, *n* = 6, 3.4% =L03 Cellulitis, *n* = 6, 3.4% Total = 179

[a]Deaths prior to 2001 were recorded using ICD-9 codes and have not been included.
[b]All incidences of code N39 were N39.0 *Urinary tract infection, site not specified.*

Table 4. Recording of renal disease as a cause of death[a] following pneumonia among patients with reduced eGFR

eGFR (mL/min/1.73 m^2)	Deaths within 28 days of pneumonia onset			Deaths within 90 days of pneumonia onset		
	Total	N18 CKD, *n* (%)	Renal disease,[b] *n* (%)	Total	N18 CKD, *n* (%)	Renal disease,[b] *n* (%)
<15	8	6 (75.0)	7 (87.5)	8	6 (75.0)	7 (87.5)
15–29	109	19 (17.4)	41 (37.6)	139	26 (18.7)	53 (38.1)
30–44	275	14 (5.1)	2 (15.3)	347	20 (5.8)	56 (16.1)
35–59	332	8 (2.4)	27 (8.1)	419	10 (2.4)	36 (8.6)
Total	724	47 (6.5)	77 (10.6)	913	62 (6.8)	152 (16.6)

[a]Deaths prior to 2001 were recorded using ICD-9 codes and have not been included.
[b]Any ICD-10 code from Chapter XIV 'Diseases of the genitourinary system' except N10 'Acute tubule-interstitial nephritis', which is used for pyelonephritis, N30 'Cystitis', N34 'Urethritis' or N39.0 'Urinary tract infection, site not specified'.

rather than in-hospital mortality (thus is not due to differences in hospital stay) and when exclusively community-acquired infections are considered (so does not result from increased risk of healthcare-associated infections). The linked datasets allowed us to identify infections both among patients presenting directly to hospital and those managed in the community, maximized ascertainment of mortality and enabled description of the causes of death. The highly monitored population allowed good ascertainment of CKD status. The cohort study design has less potential for selection bias than an equivalent case–control study.

A limitation is our assumption that the absence of a record implies a negative status for proteinuria and co-morbidities. Under-ascertainment of co-morbidities could result in residual confounding, with unpredictable effects, but the high prevalence of co-morbidities observed suggests that ascertainment was not markedly incomplete. We observed a high prevalence of proteinuria, and this is a highly monitored population (with financial incentives for standardized recording of proteinuria since 2004), but under-ascertainment of proteinuria could result in underestimation of any association between proteinuria and mortality [24]. Residual confounding from undiagnosed cardiovascular disease should have been minimized by adjustment for cardiovascular disease risk factors including smoking, hypertension and characteristics of diabetes.

Our findings for eGFR provide further detail to build on previous findings that baseline eGFR <60 mL/min/1.73 m^2 or renal disease are risk factors for short-term mortality following (hospital- or community-acquired) sepsis and for in-hospital mortality following community-acquired pneumonia (including patients receiving dialysis) [11–15]. A more comparable Canadian study examined the associations between eGFR and 30-day mortality following community-acquired pneumonia among the general population aged ≥65 years, excluding patients with ESRD [16]. Fully adjusted hazard ratios for 30-day mortality were 1.22 (95% CI 1.01–1.49) for eGFR 45–59, 2.03 (1.64–2.50) for eGFR 30–44 and 4.94 (3.94–6.19) for eGFR < 30, compared with eGFR 60–104 mL/min/1.73 m^2. These are somewhat greater than the associations we observed. The difference may be explained by the different study populations. Both studies required a baseline serum creatinine result for inclusion. Our study population of older people with diabetes were routinely monitored for CKD (with financial incentivization in primary care) [32]. Creatinine testing of the Canadian study population may have been encouraged by co-morbidities or health-behaviours associated with CKD (such as smoking), which increase post-infection mortality, resulting in overestimation of the association of eGFR and post-infection mortality. Our study population is less vulnerable to differential ascertainment of CKD. Alternatively, the association between eGFR and post-infection mortality may be smaller among patients with diabetes.

To the best of our knowledge, our examination of any association between proteinuria and mortality following community-acquired pneumonia and sepsis is novel. A history of proteinuria, although a marker for mortality in general, does not appear to be a risk marker for short-term mortality following community-acquired infection. This is unlikely to be due to chance, as the study was large, with findings consistent across both infections. We designed our study to produce conservative estimates and may have under-estimated the association between proteinuria and post-infection mortality due to under-ascertainment of proteinuria. Alternatively, any potential relationship between proteinuria and mortality may have been mitigated by clinical care of patients with infection who had pre-existing proteinuria, for example, through swift recognition of AKI.

The survival curves demonstrate a steep initial mortality following infection onset, and a high proportion of deaths had the underlying cause assigned to infection. Since 2001 in England, co-morbidities are assigned as the underlying cause of death when pneumonia has occurred in the context of, for example, malignancy or respiratory disease [25]. This suggests that the associations we observed are driven by an association between eGFR and infection prognosis, not merely high underlying baseline mortality among patients with impaired eGFR. This is supported by previous research which found 7.7-fold elevated mortality in the 30 days following community-acquired pneumonia [19]. Estimates for associations between eGFR and mortality were not substantially altered by adjustment for co-morbidities, suggesting that any causal relationship between eGFR and mortality is not mediated through co-morbidities.

Our findings apply to the large population of older patients with community-acquired pneumonia or sepsis with diabetes mellitus who do not have ESRD. Inclusion criteria are unlikely to have limited generalizability appreciably. Practices which consent to data linkage could be more research oriented, providing good primary care management of risk factors for infection-related mortality (such as smoking cessation), but this is unlikely to affect the relationship between CKD and short-term mortality post-infection. Lack of pre-existing creatinine test results is likely to reflect limited time potentially eligible for the study rather than CKD status among this highly monitored population. Missing data on smoking status and HbA1C may be a marker of low patient engagement: caution should be used in generalizing our results to patients who are not actively managed in primary care.

We found that CKD is a useful clinical risk marker for post-infectious mortality. Whether this relationship is causal is less clear; but the association does not appear to be explained by age, co-morbidities or hospital attendance. Potential mechanisms include immune system dysfunction, but also more preventable complications such as AKI. Combinations of risk factors may be important: for example, patients with post-operative AKI have higher mortality if they also have pre-existing CKD [30].

Our results have implications for patient management and future research. Patients with baseline eGFR <30 mL/min/1.73 m^2 and community-acquired infection need careful monitoring, particularly in the 28 days following infection. Future research should investigate preventable mechanisms by which low baseline eGFR could be related to post-infection mortality, for example, fluid management, AKI and drug dosage in patients with low renal clearance.

ACKNOWLEDGEMENTS

This work was supported by National Institute for Health Research (CDF 2010-03-32 to S.L.T.) and Kidney Research UK (ST2/2011 to H.I.M.). The study funders had no role in the design nor conduct of the study, nor the collection, management, analysis, nor interpretation of data, nor the preparation, review, nor approval of the manuscript, nor the decision to submit it for publication. The views expressed in this publication are those of the authors and not necessarily those of the UK National Health Service, the National Institute for Health Research, the Department of Health or Kidney Research UK. The results presented in this paper have not been published previously in whole or part, except in abstract format.

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Kerr M, Bray B, Medcalf J *et al.* Estimating the financial cost of chronic kidney disease to the NHS in England. Nephrol Dial Transplant 2012; 27 (Suppl. 3): iii73–iii80.
2. National Kidney Foundation. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Am J Kidney Dis 2002; 39: S1–S266
3. Roderick PJ, Atkins RJ, Smeeth L *et al.* Detecting chronic kidney disease in older people; what are the implications? Age Ageing 2008; 37: 179–186
4. Burden R, Tomson C. Identification, management and referral of adults with chronic kidney disease: concise guidelines. Clin Med 2005; 5: 635–642
5. Davies S. Annual Report of the Chief Medical Officer, Volume 1, 2011, On the State of the Public's Health. Department of Health, London, 2012
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. Wang HE, Gamboa C, Warnock DG *et al.* Chronic kidney disease and risk of death from infection. Am J Nephrol 2011; 34: 330–336
8. Fried LF, Katz R, Sarnak MJ *et al.* Kidney function as a predictor of non-cardiovascular mortality. J Am Soc Nephrol 2005; 16: 3728–3735
9. McDonald HI, Thomas SL, Nitsch D. Chronic kidney disease as a risk factor for acute community-acquired infections in high-income countries: a systematic review. BMJ Open 2014; 4: e004100
10. Rewa O, Bagshaw SM. Acute kidney injury-epidemiology, outcomes and economics. Nat Rev Nephrol 2014; 10: 193–207
11. Maizel J, Deransy R, Dehedin B *et al.* Impact of non-dialysis chronic kidney disease on survival in patients with septic shock. BMC Nephrol 2013; 14: 77
12. James MT, Laupland KB, Tonelli M *et al.* Risk of bloodstream infection in patients with chronic kidney disease not treated with dialysis. Arch Intern Med 2008; 168: 2333–2339
13. Kaplan V, Angus DC, Griffin MF *et al.* Hospitalized community-acquired pneumonia in the elderly: age- and sex-related patterns of care and outcome in the United States. Am J Respir Crit Care Med 2002; 165: 766–772
14. Marrie TJ, Carriere KC, Jin Y *et al.* Factors associated with death among adults <55 years of age hospitalized for community-acquired pneumonia. Clin Infect Dis 2003; 36: 413–421
15. Viasus D, Garcia-Vidal C, Cruzado JM *et al.* Epidemiology, clinical features and outcomes of pneumonia in patients with chronic kidney disease. Nephrol Dial Transplant 2011; 26: 2899–2906
16. James MT, Quan H, Tonelli M *et al.* CKD and risk of hospitalization and death with pneumonia. Am J Kidney Dis 2009; 54: 24–32
17. Go AS, Chertow GM, Fan D *et al.* Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. N Engl J Med 2004; 351: 1296–1305
18. Chalmers JD, Taylor JK, Singanayagam A *et al.* Epidemiology, antibiotic therapy, and clinical outcomes in health care-associated pneumonia: a UK cohort study. Clin Infect Dis 2011; 53: 107–113
19. Mortensen EM, Coley CM, Singer DE *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: 1059–1064
20. McDonald HI, Nitsch D, Millett ER *et al.* 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 2013; 31: 606–614
21. Afkarian M, Sachs MC, Kestenbaum B *et al.* Kidney disease and increased mortality risk in type 2 diabetes. J Am Soc Nephrol 2013; 24: 302–308
22. 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
23. Walley T, Mantgani A. The UK General Practice Research Database. Lancet 1997; 350: 1097–1099
24. Hobbs H, Stevens P, Klebe B *et al.* Referral patterns to renal services: what has changed in the past 4 years? Nephrol Dial Transplant 2009; 24: 3411–3419
25. Office for National Statistics. *Office for National Statistics, 2012.* www.ons.gov.uk. (14 September 2012, date last accessed)
26. The Health and Social Care Information Centre. *Hospital Episode Statistics, 2014*; http://www.hscic.gov.uk/hes. (11 October 2014, date last accessed)
27. Levey AS, Stevens LA, Schmid CH *et al.* A new equation to estimate glomerular filtration rate. Ann Intern Med 2009; 150: 604–612
28. de Lusignan S, Tomson C, Harris K *et al.* Creatinine fluctuation has a greater effect than the formula to estimate glomerular filtration rate on the prevalence of chronic kidney disease. Nephron Clin Pract 2011; 117: c213–c224
29. Cummings P. Methods for estimating adjusted risk ratios. Stata J 2009; 9: 175–196
30. Wu VC, Huang TM, Lai CF *et al.* Acute-on-chronic kidney injury at hospital discharge is associated with long-term dialysis and mortality. Kidney Int 2011; 80: 1222–1230
31. Brock A, Griffiths C, Rooney C. The impact of introducing ICD-10 on analysis of respiratory mortality trends in England and Wales. Health Stat Q. 2006; 29: 9–17
32. NHS Employers. *Quality and outcomes framework, 2012.* http://www.nhsemployers.org/payandcontracts/generalmedicalservicescontract/qof/pages/qualityoutcomesframework.aspx. (12 September 2012, date last accessed)

Received for publication: 13.8.2014; Accepted in revised form: 2.12.2014

10.3 Sensitivity analysis adjusting for peripheral vascular disease

The relationship between peripheral vascular disease (PVD) and the association between CKD and post-infectious mortality is an interesting one. PVD is associated with excess mortality from cancer and cardiovascular disease.[172] PVD and diabetic nephropathy are both microvascular complications of diabetes, and PVD is strongly correlated with CKD among people with diabetes.[173] PVD could therefore be considered a confounder of the association of CKD with mortality. However, if PVD is a marker of underlying microvascular disease rather than a direct cause of mortality, the statistical association between PVD and mortality could in part be driven by a causal association between diabetic nephropathy and mortality. In this case, adjusting for peripheral vascular disease would risk obscuring the component of a causal association between CKD and mortality. In the main analysis I was therefore initially wary of adjusting for peripheral vascular disease, as this could result in over-adjustment.

In response to a reviewer's suggestion, I conducted a sensitivity analysis in which adjustment for PVD was included in the final model. Surprisingly, this did not change the estimates of the association between CKD and post-infectious mortality. The results are not presented, as every point estimate and 95% confidence limit was unchanged. This model enhances the robustness of the results, as it does not make assumptions about the causality of the association between PVD and mortality.

10.4 No effect modification was observed between eGFR and proteinuria

Little is understood about the potential causal mechanisms between CKD and post-infectious mortality, nor how eGFR and proteinuria may relate to one another as risk markers. In addition to investigating whether eGFR and a history of proteinuria were independent risk factors for post-infection mortality, I therefore examined whether they combined multiplicatively or extra-multiplicatively.

To look for effect modification I constructed the final models with an interaction term between eGFR and proteinuria. **Table 10.1** presents the resulting estimates of the association of eGFR with mortality stratified by proteinuria status, and estimates of the association of proteinuria with mortality stratified by eGFR status.

Table 10.1 Risk ratios for mortality in a model with interaction between eGFR and proteinuria

		Association of eGFR with mortality within each category of proteinuria status		Association of proteinuria with mortality within each category of eGFR status	
		No proteinuria	**History of proteinuria**	**No proteinuria**	**History of proteinuria**
		Risk ratio (95% CI)	Risk ratio (95% CI)	Risk ratio (95% CI)	Risk ratio (95% CI)
Pneumonia					
eGFR	<30	1.47 (1.20–1.73)	1.15 (0.89–1.42)	1 (reference)	0.86 (0.64–1.08)
	30-44	0.99 (0.84–1.15)	0.99 (0.79–1.18)	1 (reference)	1.09 (0.87–1.32)
	45-59	0.92 (0.79–1.04)	0.96 (0.79–1.14)	1 (reference)	1.15 (0.94–1.36)
	≥60	1 (reference)	1 (reference)	1 (reference)	1.10 (0.95–1.24)
Sepsis					
eGFR	<30	1.58 (1.07–2.08)	1.05 (0.60–1.51)	1 (reference)	0.81 (0.43–1.19)
	30-44	1.34 (0.98–1.70)	1.03 (0.67–1.39)	1 (reference)	0.93 (0.61–1.26)
	45-59	1.00 (0.73–1.28)	0.72 (0.42–1.03)	1 (reference)	0.88 (0.49–1.26)
	≥60	1 (reference)	1 (reference)	1 (reference)	1.21 (0.91–1.52)

eGFR, estimated glomerular filtration rate ml/min/1.73m^2; 95% CI, 95% confidence interval.

There was no evidence that the effect of proteinuria was modified by eGFR status, and in particular no evidence of any association of proteinuria with 28-day mortality within any category of eGFR status. For both pneumonia and sepsis, the association of eGFR <30 ml/min/1.73m^2 with mortality may be slightly larger among patients without proteinuria than among patients with a history of proteinuria: however the confidence intervals of these two estimates were wide and overlapping. The effect was not seen in any other category of eGFR, suggesting that it may reflect the fact that the subgroups of patients with eGFR <30 ml/min/1.73m^2 once stratified by proteinuria status were small. Wald tests for each interaction term did not provide any evidence against the null hypothesis that there was no interaction between eGFR and proteinuria. Overall, there was no good evidence of a clinically relevant effect modification.

10.5 Why did this study focus on the first episode of community-acquired pneumonia and sepsis?

Pneumonia and sepsis are severe infections, associated with a high short-term mortality among this study population, as described in **Paper 2**. They are therefore infections where any association of CKD with mortality would have particular relevance.

There are epidemiological advantages to studying the first episode of infection during follow-up when mortality is the outcome of interest. If repeat infections were included in this study without an adjusted analysis, an assumption of the model would be that having survived a previous episode of infection does not affect the probability of surviving a further episode of infection. The frailty models used to adjust for clustering in the analysis of CKD with infection incidence (**4.6**) are not available to this study, as mortality is a unique event for each patient. One option would be to include previous infections as a variable in the model, with robust standard errors to adjust confidence intervals. This model would assume that the relationship between prior infection (and survival to encounter a second infection) with short-term mortality at a repeat infection was the same for all patients, which also seems unlikely. A simpler approach, without such assumptions, was to include only the first infection during follow-up. For a common infection, this might have the disadvantage of risking survivor bias among older age groups and in age-adjusted estimates, as patients who encountered a first infection at an older age might be unusually robust (as discussed in **8.3.4**). However, as pneumonia and sepsis are uncommon infections

and repeat infection is rare, the risk of survivor bias when considering the first episode of these infections should be minimal.

10.6 Identifying eGFR status using the 'best-of-two' approach

The exposure of interest in this study was baseline CKD status at onset of infection, which is not a time-updated variable. It was therefore suitable for the 'best-of-two' method for identifying eGFR status. This method was discussed in **5.4.5**, but briefly the approach is, from the diagnosis date of infection, to identify the latest two serum creatinine tests with at least a three month interval between them. The serum creatinine test which results in the highest GFR estimate of the pair is the 'defining' creatinine test, and is used to identify a 'best-of-two' estimate of eGFR status.

The advantages of this method are that is approximates the clinical definition of CKD, prevents overestimation of CKD prevalence from fluctuation in creatinine levels, and minimises misclassification of acute kidney injury as CKD. The method risks under-estimating CKD status, as if the patient's CKD status has progressed between the two serum creatinine results this will not be reflected in the CKD status assigned.[118] If only patients with two results available were included, this would risk ascertainment bias, but this can be addressed by categorising eGFR based on a single result where only one serum creatinine test result is available.

This 'best-of-two' method is selected from a range of methods for identifying CKD explored by de Lusignan *et al.* among the general population, who found that when using primary care records for the general population the best-of-two method assigned fewer patients to stage 3 CKD than the single latest GFR estimate, and adjusting for discrepant interim readings in the 3 month interval made little difference to CKD status estimates.[118]

Among the general population, de Lusignan *et al.* found that GFR estimates were as likely to rise as fall over a two year period, but that eGFR tended to decline over five years.[118] This suggested that fluctuation rather than progression was the dominant phenomenon in serum creatinine changes over a two year timescale. If this was also the case among the present study population of older people with diabetes mellitus, it would suggest that the best-of-two method was more suitable than a single measurement to estimate baseline CKD status.

Among the 4,969 patients with pneumonia, eGFR was based on a single creatinine result for 813 (16.3%) and on the best-of-two for 4,173 patients (83.7%). The eGFR distribution obtained was similar whether defined using a single result or the best-of-two (**Table 10.2**). The serum creatinine test used to define eGFR status (whether a single result or the best of a pair) was within two years of infection diagnosis for 94.5% of patients with pneumonia (4714/4,969). Among the 4,173 patients with a pair of results, 4076 (97.7%) of the paired tests were within 2 years of one another. The first of the pair was the highest GFR estimate (used to define eGFR status) for 2,095 patients, and the second of the pair was the defining creatinine result for 1,947 patients (the values were the same for 131 patients). A histogram of the difference between pairs had a good approximation to a normal distribution (not shown).

Table 10.2 Characteristics of eGFR classification using the best-of-two method

	Pneumonia n=4,986		**Sepsis** n=1,119	
	Median (IQR)	**Mean [SD]**	**Median (IQR)**	**Mean [SD]**
eGFR (all patients)	61.6 (47.3–76.9)	61.4 [19.5]	61.5 (44.8–75.7)	60.4 [19.8]
eGFR (patients with a single serum creatinine test)	61.0 (44.8–75.8)	60.2 [20.8]	56.6 (41.6–73.4)	57.2 [20.9]
Best-of-two eGFR [1]	61.8 (47.8–77.2)	61.6 [19.2]	62.1 (46.3–76.3)	61.0 [19.6]
Time between the pairs (days) [1]	211 (143–343)	262 [178]	215 (143–351)	264 [182]
Absolute difference in serum creatinine from the first to the second of each pair [1]	+1 (-8 to +9)	1.9 (22.6)	+1 (-8 to +11)	+3.13 [35.4]

IQR, interquartile range boundaries; SD, standard deviation; eGFR, estimated glomerular filtration rate ml/min/1.73m^2.
1. Among patients who had two eligible serum creatinine tests, n=4,173 for pneumonia; n=948 for sepsis.

Among the 1,119 patients with sepsis, eGFR was based on a single creatinine result for 171 patients (15.3%), and on the best-of-two for 948 patients (84.7%). As with pneumonia, the eGFR distribution obtained was similar whether defined using a single result or the best-of-two (**Table 10.2**). The serum creatinine test used to define eGFR status (whether a single result or the best of a pair) was within two years of infection diagnosis for 94.3% of patients with sepsis (1,055/1,119). Among the 948 patients who had a pair of results, 928 (97.9 %) of the paired tests were within 2 years of each other. The first of the pair was the highest GFR estimate (used to define eGFR status) for 487 patients, and the second of the pair was the defining creatinine result for 419 patients (the values were the same for 42 patients). A histogram of the difference between pairs had a good approximation to a normal distribution (not shown).

Thus, for both infections studied, most patients had eGFR based on the best-of-two serum creatinine tests, but patients with a single eGFR test result did not appear to differ in terms of assigned CKD status. Almost all tests used to define eGFR status were within 2 years of infection diagnosis, and this appeared to be a sufficiently short timescale that serum creatinine fluctuation dominated over progression, suggesting that the best-of-two was an appropriate method to this study.

Chapter 11. Overall discussion

This thesis used routinely-collected electronic health records to investigate the epidemiology of acute, community-acquired infections according to markers of chronic kidney disease (CKD) prior to end-stage renal disease. Among patients aged ≥65 years with diabetes mellitus, the objectives (**1.4.2**) were to describe: the burden of acute community-associated infections (objective 2); the associations between markers of CKD with incidence of, and short-term mortality following, selected community-acquired infections (objectives 3 and 5, respectively); and the extent to which routine influenza and pneumococcal vaccination could prevent community-acquired LRTI and pneumonia according to CKD status (objective 4).

This chapter reviews the main findings, briefly surveys the overall strengths and weaknesses of the thesis, and considers the implications of the results for clinical practice and research.

11.1 Summary of main results

11.1.1 Incidence of community-acquired infections among older people with diabetes

What was known

Hospital admission rates for community-acquired infections such as pneumonia are rising rapidly in the UK, particularly among older age groups. One suggested explanation is the increasing prevalence of co-morbidities such as diabetes. The estimated number of adults in England with diabetes mellitus was 3.1 million in 2010 and is predicted to have risen to 4.6 million by 2030. Although diabetes mellitus is a risk factor for infection-related hospitalisation and mortality, the full burden of community-acquired infections among older people with diabetes mellitus had not been described.

What this study adds

There was a high burden of community-acquired infection among older people with diabetes, LRTI having the highest incidence (crude rate 152.7/1,000 years) followed by UTI (crude rates 51.4 and 147.9/1,000 years for males and females respectively). The incidence of all infections increased over the study period (1997–2011) which appeared to be driven by the changing age structure of the population. Although the proportion of patients hospitalised within four weeks of diagnosis was low for UTI compared to pneumonia or

sepsis, the absolute number of patients hospitalised within four weeks of diagnosis was higher for UTI than for pneumonia and sepsis combined. All-cause 28-day mortality was 32.1% for pneumonia, 31.7% for sepsis, 4.1% for LRTI and 1.6% for UTI.

11.1.2 Association of CKD with infection incidence

What was known

Chapter 2 presented a systematic review of the association between CKD prior to end-stage renal disease with the incidence of acute, community-acquired UTI, LRTI, central nervous system infection or sepsis. The review identified 14 eligible studies, of which only three included infections managed in the community. Probable misclassification of kidney disease status and poor adjustment for confounding were common. There was evidence from a few large, high-quality studies of a graded association between reduced eGFR and increased hospitalisation for infection. There were little data available on the association of CKD with infection incidence using less severe outcome measures than hospitalisation, and it was not possible to distinguish an effect on susceptibility to infection from an effect on the severity of infection. There was no evidence on the relationship between proteinuria and infection incidence independent of eGFR.

What this study adds

This was the first large study to explore the association of both estimated glomerular filtration rate and a history of proteinuria with incidence of these community-acquired infections. The inclusion of infections managed in primary care, as well as those resulting in hospitalization, allowed estimation of the association of CKD with infection incidence that was not driven by hospital admission thresholds or infection severity. Among older people with diabetes mellitus and no history of renal replacement, reduced eGFR was associated with a strong and graded increased incidence of community-acquired LRTI, pneumonia (as a subset of LRTI) and sepsis. The association was steeper for more severe infections: the association between reduced eGFR and infection incidence was greater for sepsis than pneumonia, and for pneumonia than for LRTI more widely. Proteinuria was a risk marker of increased infection incidence independently of eGFR, for LRTI (rate ratio 1.07: 95% CI 1.05–1.09), pneumonia (1.26: 1.19–1.33), and sepsis (1.33: 1.20–1.47), after adjustment for age, sex, co-morbidities, smoking status and characteristics of diabetes.

11.1.3 The use of pneumococcal and influenza vaccination against community-acquired LRTI

What was known

Older people with diabetes have a high burden of morbidity and mortality from community-acquired LRTI and pneumonia. *Streptococcus pneumoniae* ('pneumococcus') and bacterial co-infections or secondary infections following seasonal influenza infection together account for a high proportion of cases of community-acquired pneumonia, especially among older people. Vaccines are available against both pneumococcus and influenza, but their effectiveness among this population is unclear. Patients with CKD have a generally reduced response to vaccines; it was unknown whether influenza and pneumococcal vaccine effectiveness (VE) was reduced among patients with CKD, especially prior to end-stage renal disease.

What this study adds

The adjusted effectiveness of pneumococcal vaccine for preventing pneumonia was 22% (95% CI 11–31) within the first year after vaccination, but protection appeared to wane swiftly: pneumonia incidence among patients vaccinated more than 5 years previously was similar to that among patients with no record of vaccination (incidence rate ratio 1.03: 95% CI 0.95–1.11). There was some evidence for reduced pneumococcal vaccine effectiveness among patients with proteinuria, and possibly among patients with impaired eGFR.

The incidence of LRTI was higher among patients with a current influenza vaccine than unvaccinated patients before and after adjustment for co-morbidities, in winter and summer. Using a traditional analysis, a negative influenza vaccine effectiveness against community-acquired LRTI would have been observed. Using the ratio-of-ratios analysis to address confounding by indication, a current influenza vaccination had 7% effectiveness (95% CI 3–12) and residual influenza vaccination 12% effectiveness (95% CI 7–17) to prevent community-acquired LRTI. There was no evidence to suggest that influenza vaccine effectiveness varied according to eGFR or proteinuria status.

11.1.4 Association of CKD with short-term mortality following infection

What was known

Reduced eGFR and proteinuria are associated with an increased rate of infection-related mortality. This could be partly explained by increased incidence of community-acquired infection, but CKD may also be associated with poorer prognosis following infection. Low

baseline eGFR had been found to be associated with mortality following sepsis and community-acquired pneumonia, but rarely examined according to clinically meaningful categories of eGFR. To the best of our knowledge, proteinuria had not been examined as a potential risk marker for mortality following infection.

What this study adds

People with eGFR <30 ml/min/1.73m^2 were at higher risk of death compared to people with eGFR≥60 ml/min/1.73m^2 in the 28 days following diagnosis of pneumonia (adjusted risk ratio, RR 1.27: 95% CI 1.12–1.43) and sepsis (adjusted RR 1.32: 1.07–1.64), which was not explained by existing co-morbidities (adjusted for age, sex, socio-economic status, smoking and co-morbidities). Neither moderately impaired eGFR nor proteinuria were associated with short-term mortality following either infection.

11.2 Strengths

The strengths of the individual studies were discussed in **Chapters 7 to 10**. The overall strengths of this thesis follow from using large, linked datasets among a highly monitored population with study definitions, design and analysis tailored to a series of focused study questions designed to address identified gaps in the literature.

11.2.1 The use of large, linked datasets

The thesis used large, linked datasets of primary care and hospital admission records, with additional linkages to Office for National Statistics mortality data. Large cohorts permitted precise estimates of the main associations of interest with eGFR categorised into narrow categories to improve the clinical relevance of estimates and allow detection of graded associations.

Data linkage allowed exploitation of the different strengths of a variety of datasets. Primary care datasets offered a rich source of patient characteristics and co-morbidities, including results of laboratory tests sent from primary care. Data linkage to secondary care allowed identification of infections both among patients presenting directly to hospital and those managed in the community, and the distinction of community-acquired from hospital-acquired infections. Linkage to ONS mortality data maximized ascertainment of mortality and enabled description of the causes of death.

11.2.2 Selection of a highly monitored and clinically relevant study population

All study populations were restricted to older people with diabetes. Among this population, a high proportion had regular serum creatinine tests recorded, and this reduces the risk of selection bias in studies in which serum creatinine tests were an inclusion criterion, of misclassification of eGFR status from infrequent monitoring, and of ascertainment bias in estimated associations of eGFR with outcomes of interest. Age and diabetes are important *a priori* confounders of any associations between CKD and infection incidence or mortality, and restriction of the study population helps control confounding by these characteristics (**1.4.1**).

Older people with diabetes are a large population, who bear a high burden of morbidity and mortality from community-acquired infections, and have a high prevalence of CKD: the association of CKD with infection-related morbidity and mortality is thus particularly salient among this population, from clinical and public health perspectives. The population is ageing and the prevalence of diabetes is increasing, and the health service needs of this population are of growing relevance to health service planning and health promotion strategies.

11.2.3 Tailored study definitions, design and analysis for a series of focused questions

This thesis presents the first studies to describe mutually-adjusted associations of eGFR and proteinuria with incidence of these community-acquired infections, and with post-infectious mortality. This allows interpretation of eGFR and proteinuria as independent markers of risk among older people with diabetes.

Episodes of community-acquired infections were defined according to detailed methods developed by Elizabeth Millett for LRTI and pneumonia, and adapted in this thesis to sepsis and UTI, as described in **Chapter 4**. The methods were designed to minimise misclassification of infections, and the identification of community-acquired infections specifically allows the study of the association of CKD with infection incidence to be independent of more frequent hospital attendance among patients with CKD, and the association of CKD with post-infectious mortality to be independent of poorer prognosis of hospital-acquired infections.

CKD was identified sensitively to the historical context of the data, among a highly monitored study population and data handling decisions were taken in the context of the thesis objectives, to minimise misclassification and ascertainment bias in the estimation of

the association of CKD with infection-related morbidity and mortality, as discussed in **Chapter 5**.

Cohort study designs were used for all study objectives, and these have less potential for selection bias than an equivalent case-control study. Within this design, each study was analysed with respect to the main biases of concern to each specific study question, for example in the ratio-of-ratios analysis to address confounding by indication in influenza vaccine effectiveness estimates. The data permitted adjustment for a wide range of co-morbidities. Confounders were identified using a conceptual framework and each confounder adjusted for individually, to adjust for confounding as accurately as possible while maintaining a parsimonious model.

The systematic review presented in **Chapter 2** identified that the roles of increased susceptibility to infection and poorer prognosis from infection in driving excess infection-related mortality and hospitalization among patients with CKD could not be separately identified from existing studies of the association between eGFR and infection. This thesis addressed that gap, by conducting separate studies of the association of CKD with infection incidence (including infections managed in the community) and with short-term mortality following infection diagnosis. The thesis also presented novel analyses of the associations of proteinuria with infection-related morbidity and mortality independently of eGFR, which may advance the use of proteinuria as a marker of clinical risk.

11.3 Limitations

The potential for misclassification, information bias, selection bias and reverse causation specific to each study was discussed in **Chapters 7 to 10**. The overarching limitations of the thesis mostly arise from its reliance on historical data collected for clinical purposes. Data validity depended on full and accurate clinical investigation, diagnosis and coding.

11.3.1 Misclassification and information bias

All variables were subject to misclassification from clinical error or inaccurate coding. In general, the positive predictive value of diagnostic records in CPRD has been found to be high, but the negative predictive value is less clear.[72] Patient investigation, diagnosis and monitoring are driven by clinical criteria, which is inherently differential according to overall health status. As a wide range of factors inform CKD status and infection risk which are likely to be correlated with clinical management and coding (such as smoking status, age and co-morbidities), all analyses of the associations between CKD and infection-related

outcomes are vulnerable to information bias from misclassification, and poorly controlled confounding from misclassified confounders. Data recording also reflects clinical imperatives. In particular, negative results are often less clinically relevant to patient management than positive results, which distorts recording. The resulting missing data may cause misclassification, information bias, selection bias and uncontrolled confounding, depending on the context and how the data are handled.

Misclassification resulting from missing data is particularly relevant to patient co-morbidities and proteinuria status, as a negative status was inferred from absence of a positive record (**5.2.4** and **6.5**). Any misclassification of co-morbidities would impair control for confounding, biasing estimates of CKD with infection-related outcomes in either direction. Misclassification of proteinuria status that was non-differential by CKD status and infection-related outcomes is likely to have resulted in under-estimation of the association of proteinuria with infection-related outcomes. Information bias from differential misclassification of proteinuria status could bias estimates of the association of proteinuria with infection-related outcomes in either direction.

It was not always possible to categorise variables finely. For example, smoking status was classified into crude categories based on current status, when ideally smoking pack year history would have allowed adjustment for cumulative exposure in addition to current status (**6.3**). This may have resulted in residual confounding by smoking status. The data did not permit confident identification of albuminuria, or quantification of proteinuria, both of which would have improved the detail in which the associations between proteinuria and infection-related outcomes could be described (**5.5.5**).

Estimated GFR status was updated over time for two analyses (objectives 3 and 4). As the frequency and timing of serum creatinine testing reflects clinical imperatives, this risked misclassification over time which was likely to be differential with respect to infection incidence, potentially introducing ascertainment bias. This risk was mitigated by selecting a highly-monitored population. The high frequency of testing (**5.4.4**), and results of sensitivity analysis restricted to the period after the introduction of Quality Outcomes Framework incentivised annual testing for all (**Paper 3**), suggested that ascertainment bias was not a major problem among this population.

11.3.2 External validity

All studies included only a subset of older people with diabetes mellitus. The inclusion criteria for all studies required a diagnosis of diabetes mellitus; studies of the associations

of CKD with infection-related outcomes also required a valid serum creatinine test result and non-missing smoking status (objectives 3-5); and the study of post-infectious mortality additionally required an available HbA1C result (objective 5).

Patients with undiagnosed diabetes mellitus do not receive care to manage their diabetes, thus patients excluded from the study by lack of diabetes diagnosis may be at greater risk of CKD and infection.[174] Serum creatinine testing, smoking cessation advice, and HbA1C monitoring are also key components of care for patients with diabetes, and while these data may be missing for a variety of reasons, excluded patients are likely to be a population with poorly managed diabetes compared to included patients. These results may not be valid to be generalised to all older people with diabetes mellitus. However, the results may be generalised to the population of older people with a diagnosis of diabetes mellitus who are actively managed in primary care, which is a large and reasonably identifiable population.

11.3.3 Confounding

Adjustment for confounding is a strength of this thesis. Confounders were identified using a conceptual framework and a wide range of confounders were adjusted for individually, to adjust for confounding as accurately as possible while maintaining a parsimonious model.

All adjusted estimates are of course still vulnerable to residual confounding from crude categorisation of confounders (**11.3.1**), or from omission of relevant confounders (**8.3.6**).

11.3.4 Reverse causation

Acute community-acquired infections may alter kidney function directly (for example by triggering acute kidney injury) or indirectly (for example by disrupting glycaemic control, or prompting smoking cessation). This results in a risk of reverse causation in estimates of the association of CKD and acute community-acquired infection incidence (objective 2), but the study was designed to mitigate this, and sensitivity analysis suggested this was not a major issue (**8.3.4**).

11.3.5 Restricted lines of investigation

The content and coding of the data reflected changing clinical practice, diagnostic categories and coding regulations during the study period, and this required consideration in the study designs. For example, the plan for identification and classification of CKD status was developed with careful reference to the clinical, historical and coding context of the study period (**Chapter 5**). In studying the association of CKD with post-infection mortality

(objective 5), all-cause 28-day mortality was selected as an outcome over deaths coded as caused by infection as this outcome was better supported by the data and coding rules, rather than due to its greater intrinsic interest (**3.4.3**).

Some information was not consistently available to the study and this limited the feasible study questions. Infectious aetiology would have allowed the studies to identify the main causative pathogens driving the burden of infection-related mortality among older people with diabetes, to examine whether CKD status is associated with different infectious aetiologies than those among the general population, and to allow estimation of influenza and pneumococcal vaccine effectiveness against the specific outcomes of influenza and pneumococcal infections. However, the aetiology of community-acquired infections could not be uniformly and routinely identified from the available datasets (**4.4**).

Data available in Hospital Episode Statistics were less rich than those in CPRD. For example, treatment and laboratory test results in secondary care were not available. This prevented an exploration of the role of events in secondary care (such as antibiotic prescription or acute changes in serum creatinine or electrolytes) as potential mechanisms for the association of CKD with prognosis among patients with acute community-acquired infections.

11.4 Interpretation

Among older people with diagnosed diabetes mellitus who are actively managed in primary care, both proteinuria and even mildly reduced eGFR are associated with increased incidence of acute community-acquired infections. The differences between this thesis and previous studies help clarify the possibly explanations for this. This study excluded patients with end-stage renal disease (identifying that the association is not due to renal replacement therapy), considered exclusively community-acquired infections (establishing that the association is not explained by increased risk of healthcare-associated infections), included infections managed in the community (identifying that CKD is truly associated with increased infection incidence, and the association is not solely a function of hospital admission thresholds or infection severity) and adjusted for a range of co-morbidities and characteristics of diabetes to limit the extent to which the association may be explained by confounding.

An association between CKD and post-infection mortality was observed only among patients with severely reduced eGFR. This is not evidence that early stages of CKD are not

associated with poorer prognosis of infection. As the association between CKD and infection incidence was stronger for more severe infections (sepsis and pneumonia, compared to LRTI), it seems unlikely that there is no underlying relationship between CKD and infection severity. It is possible, for example, that an inherently poorer prognosis for patients with CKD may be unobserved if mitigated by good clinical care following diagnosis.

A high burden of LRTI and pneumonia was observed among the population of older people with diabetes, and particularly patients with CKD. Among the general population, a large burden of LRTI and pneumonia is related to two pathogens against which immunisation is available. The results suggested that increased uptake of routine influenza and pneumococcal vaccine would have limited impact on the burden of community-acquired LRTI and pneumonia among older people with diabetes. This may be due to a high burden of non-vaccine preventable LRTI/pneumonia, but it may also be that vaccine effectiveness could be improved among older people with diabetes using current vaccine schedules.

Are the associations of CKD with infection incidence and post-infectious mortality causal? The graded association of eGFR with infection incidence would be consistent with a causal explanation, although it could also be explained alternatively. There are plausible biological mechanisms for a causal relationship between CKD and infection incidence and CKD with post-infection mortality, but this thesis was not designed to investigate these (**1.2.3**). Even if CKD is not the cause of increased incidence of, and mortality from, acute community-acquired infections, both eGFR and proteinuria are observable and quantifiable, and may serve as useful markers for identifying patients at risk of infection-related ill health.

11.5 Implications for clinical practice

This thesis may have some direct applications to clinical practice. For example, the quantification of incidence of selected acute community-acquired infections among older people with diabetes may allow primary care clinicians to offer more detailed information to patients than previously when discussing infection risk for patients with diabetes, and how CKD relates to this. This may, for example, inform and support vaccination uptake. The description of the association between baseline CKD status and short-term mortality may also inform risk-stratification of older patients with diabetes at diagnosis of acute, community-acquired infection.

The UK population is ageing, and the prevalence of diabetes is increasing: understanding the health needs of older people with diabetes is of growing importance to health service

planning and public health. Quantifying the burden of community-acquired infection among older people with diabetes, and describing the role of CKD in this, should inform health service planning and help refine health economics analyses of the health and financial costs of CKD.

11.6 Implications for research

11.6.1 Investigating CKD using electronic health records

This thesis worked within the limitations of historical routinely-collected data, by selecting a highly-monitored study population and by developing the study classification of CKD within the context of the available data and its clinical and historical context. It might be possible to improve usefulness of historical records for identifying and categorising proteinuria, for example, by developing tools to scan and automatically encode proteinuria results recorded in free text. There is potential to increase the scope of renal research using electronic health records by altering the capture and curation of CKD data in prospectively collected electronic health records. For example, data structure and curation which standardised the recording of proteinuria (using templates for data entry, and retaining detail in dataset curation) might allow quantification of proteinuria and albuminuria, improving the granularity of research into the relevance of proteinuria.

11.6.2 Unanswered questions

The ability to study potential mechanisms for any causal association of CKD with infection incidence and mortality in this thesis was limited, but future research investigating whether potential causal mechanisms are in fact observed would help to further answer the question of causality and identify the relevance of any modifiable mechanisms, such as acute kidney injury.

Single pneumococcal and influenza vaccination appears to have limited scope to reduce the overall burden of community-acquired pneumonia and LRTI. This may partly be explained by the non-specificity of pneumonia and LRTI aetiology to these vaccines among the study population. However, enhanced vaccines (for example, using adjuvants) and schedules (for example, annual pneumococcal vaccination) may improve vaccine effectiveness, and should be explored for this high-risk population.

11.7 Personal learning

Preparing this thesis has been an opportunity for me to reflect on my personal learning. The course of my studies has not always been efficient. For example, the identification and classification of Read codes for chronic renal diseases was time-consuming. This code list is now a resource available to other researchers, but an earlier assessment of their strengths and weaknesses relative to the completeness of serum creatinine testing might have allowed me to describe CKD status more quickly.

I have gained new knowledge and skills during my studies. Practically, I have developed improved data management skills, learned to conduct data analyses which were new to me and responded to some analytical challenges, such as developing an approach to non-converging models when calculate mortality risk ratios. I have also gained a better theoretical understanding of the challenges of handling, analysing and interpreting routinely-collected electronic health records. In particular, developing detailed study design and data analysis plans for each study objective required me to move from identifying bias in studies to managing it as far as possible.

I have been fortunate to have my studies funded by Kidney Research UK. Attending the annual KRUK Fellows' Days, I have been struck by how widely the scope of kidney research ranges: in discipline from epigenetics and immunology to epidemiology; in aim from behaviour change for CKD prevention to refining dialysis delivery arrangements; in scale from studies of muscle-cell models to interactions of CKD with the cardiovascular system. Patients have generously shared their stories of living with the psychological, as well as physical, demands of kidney disease. Homer W. Smith argued that it was the kidney's salt and water regulation that had won humankind the freedom from the sea to become philosophers rather than fish.[1] From my contact with the Kidney Research UK community, I have gained some appreciation of the nuanced and pervasive roles of the kidney in supporting wellbeing, and learned to see the kidney as more than a physiologically impressive organ working in isolation from the rest of the person.

[1] A quote from Homer W. Smith's book, 'From Fish to Philosopher' (1953) appears on page 3 of this thesis.

11.8 Conclusions

Electronic health records facilitate fast, large, low-cost observational studies. They also raise challenging methodological issues, such as how to identify health status from a wealth of data collected for the purposes of routine care with varying validity. This is a particular challenge for CKD, as a silent disease.

The population of older people with diabetes has a high burden of morbidity and mortality from community-acquired infections, and among this population, CKD is associated at early stages with an increased risk of infection incidence and, at later stages, with short-term mortality post-infection.

There has traditionally been some hesitancy in diagnosing early stages of CKD among older people. The relevance of CKD is increasingly accepted as extending beyond the risk of progression to end-stage renal failure. Early stages of CKD are now recognised as important risk markers for cardiovascular and cerebrovascular events, and are used to prompt aggressive management of cardio- and cerebrovascular risk factors. A parallel relationship appears to exist between CKD and community-acquired infection incidence, and I hope that this thesis adds detail to that picture.

Several major causes of CKD (diabetes, hypertension and smoking) are modifiable: CKD is predominantly a preventable disease. It is to be hoped that this further evidence of the systemic importance of early CKD will contribute another element to recognition of the burden of morbidity and mortality associated with even early stages of CKD, and the potential benefits of CKD prevention.

References

1. Ruben, F.L., et al., *Clinical infections in the noninstitutionalized geriatric age group: methods utilized and incidence of infections. The Pittsburgh Good Health Study.* Am J Epidemiol, 1995. **141**(2): p. 145-57.
2. Laupland, K.B., et al., *Burden of community-onset bloodstream infection: a population-based assessment.* Epidemiol Infect, 2007. **135**(6): p. 1037-42.
3. Laupland, K.B., et al., *Community-onset urinary tract infections: a population-based assessment.* Infection, 2007. **35**(3): p. 150-3.
4. Kaplan, V., et al., *Hospitalized community-acquired pneumonia in the elderly: age- and sex-related patterns of care and outcome in the United States.* Am J Respir Crit Care Med, 2002. **165**(6): p. 766-72.
5. Millett, E.R., 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**(9): p. e75131.
6. Ruth, K. and J. Verne, *Deaths in older adults in England*. 2010, National End of Life Care Intelligence Network.
7. Office for National Statistics. *Population Ageing in the United Kingdom, its Constituent Countries and the European Union*. 2012 13 January 2015]; Available from: http://www.ons.gov.uk/ons/dcp171776_258607.pdf.
8. Bardsley, M., 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**(1): p. e002007.
9. Trotter, C.L., et al., *Increasing hospital admissions for pneumonia, England.* Emerg Infect Dis, 2008. **14**(5): p. 727-33.
10. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group, *KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease.* Kidney inter., 2013. **Suppl.** (3): p. 1-150.
11. NICE, *Chronic kidney disease: early identification and management of chronic kidney disease in adults in primary and secondary care.* 2008, NICE clinical guideline 73.
12. De Lusignan, S., *Identification and management of chronic kidney disease.* Prescriber, 2008: p. 10-18.
13. *K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification.* Am J Kidney Dis, 2002. **39**(2 Suppl 1): p. S1-266.
14. Kerr, M., et al., *Estimating the financial cost of chronic kidney disease to the NHS in England.* Nephrol Dial Transplant, 2012. **27**(Suppl 3): p. iii73-80.
15. The NHS Information Centre Prescribing and Primary Care Services, *Quality and Outcomes Framework Achievement Data 2010/11*. 2011, The NHS Information Centre: Leeds, UK.
16. Roth, M., P. Roderick, and J. Mindell, *Kidney disease and renal function*, in *Health Survey for England 2010*, R. Craig and J. Mindell, Editors. 2011, The Health and Social Care Information Centre: England, UK.
17. Stevens, P.E., et al., *Chronic kidney disease management in the United Kingdom: NEOERICA project results.* Kidney Int, 2007. **72**(1): p. 92-9.
18. Jameson, K., et al., *Prevalence and management of chronic kidney disease in primary care patients in the UK.* Int J Clin Pract, 2014. **68**(9): p. 1110-21.
19. Silverwood, R.J., et al., *Association between younger age when first overweight and increased risk for CKD.* J Am Soc Nephrol, 2013. **24**(5): p. 813-21.
20. Hobbs, H., et al., *Referral patterns to renal services: what has changed in the past 4 years?* Nephrol Dial Transplant, 2009. **24**(11): p. 3411-9.

21. Kerr, M., *Chronic kidney disease in England: the human and financial cost*. 2012, Insight Health Economics Ltd: Haverhill, Suffolk.
22. Stevens, L.A., G. Viswanathan, and D.E. Weiner, *Chronic kidney disease and end-stage renal disease in the elderly population: current prevalence, future projections, and clinical significance.* Adv Chronic Kidney Dis, 2010. **17**(4): p. 293-301.
23. Go, A.S., et al., *Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization.* N Engl J Med, 2004. **351**(13): p. 1296-305.
24. Tonelli, M., et al., *Chronic kidney disease and mortality risk: a systematic review.* J Am Soc Nephrol, 2006. **17**(7): p. 2034-47.
25. Sarnak, M.J. and B.L. Jaber, *Pulmonary infectious mortality among patients with end-stage renal disease.* Chest, 2001. **120**(6): p. 1883-7.
26. Sarnak, M.J. and B.L. Jaber, *Mortality caused by sepsis in patients with end-stage renal disease compared with the general population.* Kidney Int, 2000. **58**(4): p. 1758-64.
27. Dalrymple, L.S. and A.S. Go, *Epidemiology of acute infections among patients with chronic kidney disease.* Clin J Am Soc Nephrol, 2008. **3**(5): p. 1487-93.
28. Collins, A.J., et al., *United States Renal Data System 2007 Annual Data Report Abstract.* Am J Kidney Dis, 2008. **51**(1): p. A6-A7.
29. Allon, M., et al., *Impact of dialysis dose and membrane on infection-related hospitalization and death: results of the HEMO Study.* J Am Soc Nephrol, 2003. **14**(7): p. 1863-70.
30. Kausz, A.T. and D.T. Gilbertson, *Overview of vaccination in chronic kidney disease.* Adv Chronic Kidney Dis, 2006. **13**(3): p. 209-14.
31. Janus, N., et al., *Vaccination and chronic kidney disease.* Nephrol Dial Transplant, 2008. **23**(3): p. 800-7.
32. Yin, K. and D.K. Agrawal, *Vitamin D and inflammatory diseases.* J Inflamm Res, 2014. **7**: p. 69-87.
33. Wang, H.E., et al., *Chronic kidney disease and risk of death from infection.* Am J Nephrol, 2011. **34**(4): p. 330-6.
34. Fried, L.F., et al., *Kidney function as a predictor of noncardiovascular mortality.* J Am Soc Nephrol, 2005. **16**(12): p. 3728-35.
35. Collins, A.J., et al., *CKD surveillance using administrative data: impact on the health care system.* Am J Kidney Dis, 2009. **53**(3 Suppl 3): p. S27-36.
36. Naqvi, S.B. and A.J. Collins, *Infectious complications in chronic kidney disease.* Adv Chronic Kidney Dis, 2006. **13**(3): p. 199-204.
37. Foley, R.N., *Infections in patients with chronic kidney disease.* Infect Dis Clin North Am, 2007. **21**(3): p. 659-72, viii.
38. Foley, R.N., *Infections and cardiovascular disease in patients with chronic kidney disease.* Adv Chronic Kidney Dis, 2006. **13**(3): p. 205-8.
39. Coresh, J., *CKD prognosis: beyond the traditional outcomes.* Am J Kidney Dis, 2009. **54**(1): p. 1-3.
40. Holman, N., et al., *The Association of Public Health Observatories (APHO) Diabetes Prevalence Model: estimates of total diabetes prevalence for England, 2010-2030.* Diabet Med, 2011. **28**(5): p. 575-82.
41. Seshasai, S.R., et al., *Diabetes mellitus, fasting glucose, and risk of cause-specific death.* N Engl J Med, 2011. **364**(9): p. 829-41.
42. Hex, N., et al., *Estimating the current and future costs of Type 1 and Type 2 diabetes in the UK, including direct health costs and indirect societal and productivity costs.* Diabet Med, 2012. **29**(7): p. 855-62.
43. The Renal Association, *UK Renal Registry: the sixteenth annual report*. 2013.

44. Middleton, R.J., et al., *The unrecognized prevalence of chronic kidney disease in diabetes.* Nephrol Dial Transplant, 2006. **21**(1): p. 88-92.
45. Jick, H. and L.E. Derby, *A large population-based follow-up study of trimethoprim-sulfamethoxazole, trimethoprim, and cephalexin for uncommon serious drug toxicity.* Pharmacotherapy, 1995. **15**(4): p. 428-32.
46. New, J.P., et al., *Assessing the prevalence, monitoring and management of chronic kidney disease in patients with diabetes compared with those without diabetes in general practice.* Diabet Med, 2007. **24**(4): p. 364-9.
47. Joshi, N., et al., *Infections in patients with diabetes mellitus.* N Engl J Med, 1999. **341**(25): p. 1906-12.
48. Knapp, S., *Diabetes and Infection: Is There a Link? - A Mini-Review.* Gerontology, 2013. **59**(2): p. 99-104.
49. Kornum, J.B., et al., *Diabetes, glycemic control, and risk of hospitalization with pneumonia: a population-based case-control study.* Diabetes Care, 2008. **31**(8): p. 1541-5.
50. Shah, B.R. and J.E. Hux, *Quantifying the risk of infectious diseases for people with diabetes.* Diabetes Care, 2003. **26**(2): p. 510-3.
51. Vinogradova, Y., J. Hippisley-Cox, and C. Coupland, *Identification of new risk factors for pneumonia: population-based case-control study.* Br J Gen Pract, 2009. **59**(567): p. e329-38.
52. Muller, L.M., et al., *Increased risk of common infections in patients with type 1 and type 2 diabetes mellitus.* Clin Infect Dis, 2005. **41**(3): p. 281-8.
53. Caljouw, M.A., et al., *Predictive factors of urinary tract infections among the oldest old in the general population. A population-based prospective follow-up study.* BMC Med, 2011. **9**: p. 57.
54. Sliedrecht, A., et al., *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): p. 817-22.
55. Loeb, M., et al., *Environmental risk factors for community-acquired pneumonia hospitalization in older adults.* J Am Geriatr Soc, 2009. **57**(6): p. 1036-40.
56. Farr, B.M., et al., *Risk factors for community-acquired pneumonia diagnosed by general practitioners in the community.* Respiratory Medicine, 2000. **94**(5): p. 422-427.
57. Almirall, J., et al., *Risk factors for community-acquired pneumonia in adults: a population-based case-control study.* Eur Respir J, 1999. **13**(2): p. 349-55.
58. de Lusignan, S. and T. Chan, *The development of primary care information technology in the United kingdom.* J Ambul Care Manage, 2008. **31**(3): p. 201-10.
59. Simon, C., *Overview of the GP contract.* InnovAiT, 2008. **1**: p. 134-139.
60. Audit Commission, *A focus on General Practice in England*. 2002, Audit Commission Publications: Wetherby, UK.
61. Gnani, S. and A. Majeed, *A user's guide to data collected in primary care in England*. 2006: Cambridge, UK.
62. Rothman, K.J., S. Greenland, and T.L. Lash, *Modern epidemiology*. 3rd ed. ed. 2008, Philadelphia, Pa. ; London: Lippincott Williams & Wilkins.
63. Hall, G.C., et al., *Guidelines for good database selection and use in pharmacoepidemiology research.* Pharmacoepidemiol Drug Saf, 2012. **21**(1): p. 1-10.
64. de Lusignan, S. and C. van Weel, *The use of routinely collected computer data for research in primary care: opportunities and challenges.* Fam Pract, 2006. **23**(2): p. 253-63.

65. NHS Employers. *Quality and outcomes framework*. 2012 12 September 2012]; Available from: http://www.nhsemployers.org/payandcontracts/generalmedicalservicescontract/qof/pages/qualityoutcomesframework.aspx.
66. Lawson, D.H., V. Sherman, and J. Hollowell, *The General Practice Research Database. Scientific and Ethical Advisory Group.* QJM, 1998. **91**(6): p. 445-52.
67. Walley, T. and A. Mantgani, *The UK General Practice Research Database.* Lancet, 1997. **350**(9084): p. 1097-9.
68. CPRD. *The Clinical Practice Research Datalink - CPRD*. 2012 30 August 2012]; Available from: http://www.cprd.com/intro.asp.
69. Williams, T., et al., *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): p. 89-99.
70. de Lusignan, S., *Codes, classifications, terminologies and nomenclatures: definition, development and application in practice.* Inform Prim Care, 2005. **13**(1): p. 65-70.
71. Lewis, J.D., et al., *The relationship between time since registration and measured incidence rates in the General Practice Research Database.* Pharmacoepidemiol Drug Saf, 2005. **14**(7): p. 443-51.
72. Herrett, E., et al., *Validation and validity of diagnoses in the General Practice Research Database: a systematic review.* Br J Clin Pharmacol, 2010. **69**(1): p. 4-14.
73. Joint Formulary Committee. *British National Formulary* 31 January 2012]; Available from: http://www.bnf.org.
74. The Health and Social Care Information Centre. *Data Access Request Service*. 2014 [cited 2014 30 October 2014]; Available from: http://www.hscic.gov.uk/dars.
75. The Health and Social Care Information Centre. *Hospital Episode Statistics*. 2014 11 October 2014]; Available from: http://www.hscic.gov.uk/hes.
76. Hospital Episode Statistics Analysis, H.a.S.C.I.C., *Hospital Episode Statistics: Admitted Patient Care – 2012-13*. 2013.
77. The Health and Social Care Information Centre *The Quality of Nationally Submitted Health and Social Care Data in England - 2012, First annual report, Experimental statistics*. 2012.
78. Audit Commission, *Improving coding, costing and commissioning: annual report on the Payment by Results data assurance programme 2010/11*. 2011, Audit Commission: London, UK.
79. World Health Organization, *International Statistical Classification of Diseases and Related Health Problems 10th Revision*. 2010.
80. Office for National Statistics. *Mortality statistics: metadata*. 2014 12 October 2014]; Available from: http://www.ons.gov.uk/ons/guide-method/user-guidance/health-and-life-events/index.html.
81. Mathers, C.D., et al., *Counting the dead and what they died from: an assessment of the global status of cause of death data.* Bulletin of the World Health Organization, 2005. **83**(3): p. 171-177.
82. Brock, A., C. Griffiths, and C. Rooney, *The impact of introducing ICD-10 on analysis of respiratory mortality trends in England and Wales.* Health Stat Q, 2006. **29**: p. 9-17.
83. Noble, M., et al., *The English Indices of Deprivation*. 2008, Department for Communities and Local Government: London, UK.
84. Torres, A., et al., *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): p. 1065–1079.

85. Bone, R.C., et al., *Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine.* Chest, 1992. **101**(6): p. 1644-55.
86. Johansen, T.E., et al., *Critical review of current definitions of urinary tract infections and proposal of an EAU/ESIU classification system.* Int J Antimicrob Agents, 2011. **38 Suppl**: p. 64-70.
87. Schmiemann, G., et al., *The diagnosis of urinary tract infection: a systematic review.* Dtsch Arztebl Int, 2010. **107**(21): p. 361-7.
88. 59, S., *Community management of lower respiratory tract infection in adults*. 2002, Scottish Intercollegiate Guidelines Network: Edinburgh.
89. Bartlett, J.G., *Diagnostic tests for agents of community-acquired pneumonia.* Clin Infect Dis, 2011. **52 Suppl 4**: p. S296-304.
90. 88, S., *Management of suspected bacterial urinary tract infection in adults*. 2006 (updated 2012), Scottish Intercollegiate Guidelines Network: Edinburgh.
91. Lim, W.S., et al., *BTS guidelines for the management of community acquired pneumonia in adults: update 2009.* Thorax, 2009. **64**: p. iii1-iii55
92. Matsushita, K., et al., *Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis.* Lancet, 2010. **375**(9731): p. 2073-81.
93. Dinneen, S.F. and H.C. Gerstein, *The association of microalbuminuria and mortality in non-insulin-dependent diabetes mellitus: a systematic overview of the literature.* Arch Intern Med., 1997. **157**(13): p. 1413-8.
94. van der Velde, M., et al., *Lower estimated glomerular filtration rate and higher albuminuria are associated with all-cause and cardiovascular mortality. A collaborative meta-analysis of high-risk population cohorts.* Kidney Int, 2011. **79**(12): p. 1341-52.
95. DH Renal NSF Team, *The National Service Framework for Renal Services – Part Two: Chronic Kidney Disease, Acute Renal Failure and End of Life Care*, D.o. Health, Editor. 2005, DH Publications: London.
96. NICE, *Chronic kidney disease: early identification and management of chronic kidney disease in adults in primary and secondary care*. 2014.
97. Levey, A.S., L.A. Inker, and J. Coresh, *GFR estimation: from physiology to public health.* Am J Kidney Dis, 2014. **63**(5): p. 820-834.
98. Duncan, L., et al., *Screening for renal disease using serum creatinine: who are we missing?* Nephrol Dial Transplant, 2001. **16**(5): p. 1042-1046.
99. Levey, A.S., et al., *A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group.* Ann Intern Med, 1999. **130**(6): p. 461-70.
100. Burden, R. and C. Tomson, *Identification, management and referral of adults with chronic kidney disease: concise guidelines.* Clin Med, 2005. **5**(6): p. 635-42.
101. Klebe, B., et al., *Kidney disease management in UK primary care: guidelines, incentives and information technology.* Fam Pract, 2007. **24**(4): p. 330-5.
102. Levey, A.S., et al., *A new equation to estimate glomerular filtration rate.* Ann Intern Med, 2009. **150**(9): p. 604-12.
103. Department of Health, *Estimating glomerular filtration rate (GFR): information for laboratories*, Department of Health, Editor. 2007, Central Office of Information: London, UK.
104. Levey, A.S., et al., *Expressing the Modification of Diet in Renal Disease Study equation for estimating glomerular filtration rate with standardized serum creatinine values.* Clin Chem, 2007. **53**(4): p. 766-72.

105. Roderick, P.J., et al., *Detecting chronic kidney disease in older people; what are the implications?* Age Ageing, 2008. **37**(2): p. 179-86.
106. Nitsch, D., et al., *The association of renal impairment with all-cause and cardiovascular disease mortality.* Nephrol Dial Transplant, 2010. **25**(4): p. 1191-9.
107. Nitsch, D., et al., *CKD and hospitalization in the elderly: a community-based cohort study in the United Kingdom.* Am J Kidney Dis, 2011. **57**(5): p. 664-72.
108. Roderick, P.J., et al., *CKD and mortality risk in older people: a community-based population study in the United Kingdom.* Am J Kidney Dis, 2009. **53**(6): p. 950-60.
109. Jafar, T.H., et al., *Proteinuria as a modifiable risk factor for the progression of non-diabetic renal disease.* Kidney Int, 2001. **60**(3): p. 1131-40.
110. Klebe, B., et al., *The cost of implementing UK guidelines for the management of chronic kidney disease.* Nephrol Dial Transplant, 2007. **22**(9): p. 2504-2512.
111. Swedko, P.J., et al., *Serum creatinine is an inadequate screening test for renal failure in elderly patients.* Arch Intern Med, 2003. **163**(3): p. 356-60.
112. Foote, C., et al., *Impact of Estimated GFR Reporting on Late Referral Rates and Practice Patterns for End-Stage Kidney Disease Patients: A Multilevel Logistic Regression Analysis Using the Australia and New Zealand Dialysis and Transplant Registry (ANZDATA).* Am J Kidney Dis, 2014. **64**(3): p. 359-66.
113. Anandarajah, S., et al., *The validity of searching routinely collected general practice computer data to identify patients with chronic kidney disease (CKD): a manual review of 500 medical records.* Nephrol Dial Transplant, 2005. **20**(10): p. 2089-96.
114. Grams, M.E., et al., *Validation of CKD and related conditions in existing data sets: A systematic review.* Am J Kidney Dis, 2011. **57**(1): p. 44-54.
115. Vlasschaert, M.E., et al., *Validity of administrative database coding for kidney disease: a systematic review.* Am J Kidney Dis, 2011. **57**(1): p. 29-43.
116. van Staa, T.P. and L. Anbenhaim, *The quality of information recorded on a UK database of primary care records: a study of hospitalizations due to hypoglycaemia and other conditions.* Pharmacoepidemiol Drug Saf, 1994. **3**: p. 15-21.
117. Denburg, M.R., et al., *Validation of The Health Improvement Network (THIN) database for epidemiologic studies of chronic kidney disease.* Pharmacoepidemiol Drug Saf, 2011. **20**(11): p. 1138-49.
118. de Lusignan, S., et al., *Creatinine fluctuation has a greater effect than the formula to estimate glomerular filtration rate on the prevalence of chronic kidney disease.* Nephron Clin Pract, 2011. **117**(3): p. c213-24.
119. Health & Social Care Information Centre. *National Diabetes Audit*. 2014 [cited 2014 12 September 2014]; Available from: http://www.hscic.gov.uk/nda.
120. Bellomo, R., J.A. Kellum, and C. Ronco, *Acute kidney injury.* Lancet, 2012. **380**(9843): p. 756-66.
121. Grams, M.E., et al., *Albuminuria and estimated glomerular filtration rate independently associate with acute kidney injury.* J Am Soc Nephrol, 2010. **21**(10): p. 1757-64.
122. Whaley-Connell, A.T., et al., *Advances in CKD detection and determination of prognosis: executive summary of the National Kidney Foundation-Kidney Early Evaluation Program (KEEP) 2012 annual data report.* Am J Kidney Dis, 2013. **61**(4 Suppl 2): p. S1-3.
123. Suissa, S., *Immortal time bias in pharmaco-epidemiology.* Am J Epidemiol, 2008. **167**(4): p. 492-9.
124. James, M.T., et al., *Risk of bloodstream infection in patients with chronic kidney disease not treated with dialysis.* Arch Intern Med, 2008. **168**(21): p. 2333-9.
125. James, M.T., et al., *CKD and risk of hospitalization and death with pneumonia.* Am J Kidney Dis, 2009. **54**(1): p. 24-32.

126. Huerta, C., et al., *Nonsteroidal anti-inflammatory drugs and risk of ARF in the general population.* Am J Kidney Dis, 2005. **45**(3): p. 531-9.
127. Haynes, J. and R. Haynes, *Proteinuria.* BMJ, 2006. **332**(7536): p. 284.
128. NICE, *Acute kidney injury: prevention, detection and management of acute kidney injury up to the point of renal replacement therapy*. 2013, National Clinical Guideline Centre.
129. van Walraven, C., et al., *The usefulness of administrative databases for identifying disease cohorts is increased with a multivariate model.* J Clin Epidemiol, 2010. **63**(12): p. 1332-41.
130. Mathur, R., et al., *Completeness and usability of ethnicity data in UK-based primary care and hospital databases.* J Public Health (Oxf), 2013. **doi:10.1093/pubmed/fdt116**.
131. Ferguson, J., et al., *The English smoking treatment services: one-year outcomes.* Addiction, 2005. **100 Suppl 2**: p. 59-69.
132. NHS Diabetes, Diabetes UK, and Association for Clinical Biochemistry. *HbA1c Standardisation For Clinical Health Care Professionals*. 2009 2 October 2014]; Available from: http://www.acb.org.uk/docs/default-source/guidelines/hba1chealthcareprofessionaloct2011.pdf?sfvrsn=2.
133. NICE, *Type 2 diabetes: the management of type 2 diabetes*. 2009, Clinical Guideline 87. .
134. NHS Diabetes and Royal College of General Practitioners, *Coding, classification and diagnosis of diabetes: a review of the coding, classification and diagnosis of diabetes in primary care in England with recommendations for improvement*. 2011.
135. de Lusignan, S., et al., *A method of identifying and correcting miscoding, misclassification and misdiagnosis in diabetes: a pilot and validation study of routinely collected data.* Diabet Med, 2010. **27**(2): p. 203-9.
136. Matthews, S.J. and J.W. Lancaster, *Urinary tract infections in the elderly population.* Am J Geriatr Pharmacother, 2011. **9**(5): p. 286-309.
137. Smeeth, L., et al., *Risk of myocardial infarction and stroke after acute infection or vaccination.* N Engl J Med, 2004. **351**(25): p. 2611-8.
138. Davies, S., *Annual Report of the Chief Medical Officer, Volume One, 2011, On the State of the Public's Health*. 2012, Department of Health: London.
139. Ishani, A., et al., *Acute kidney injury increases risk of ESRD among elderly.* J Am Soc Nephrol, 2009. **20**(1): p. 223-8.
140. Stevens, L.A., et al., *Prevalence of CKD and comorbid illness in elderly patients in the United States: results from the Kidney Early Evaluation Program (KEEP).* Am J Kidney Dis, 2010. **55**(3 Suppl 2): p. S23-33.
141. Zhang, Q.L. and D. Rothenbacher, *Prevalence of chronic kidney disease in population-based studies: systematic review.* BMC Public Health, 2008. **8**: p. 117.
142. McCullough, K., et al., *Measuring the population burden of chronic kidney disease: a systematic literature review of the estimated prevalence of impaired kidney function.* Nephrol Dial Transplant, 2012. **27**(5): p. 1812-21.
143. Health and Social Care Information Centre, *The National Diabetes Audit - 2012-2013: Report 2, Complications and Mortality*. 2015: Leeds.
144. Middleton, R.J., et al., *The unrecognized prevalence of chronic kidney disease in diabetes.* Nephrol Dial Transplant, 2006. **21**(1): p. 88-92.
145. Dreyer, G., et al., *The effect of ethnicity on the prevalence of diabetes and associated chronic kidney disease.* QJM, 2009. **102**(4): p. 261-9.
146. Royal College of General Practitioners. *Research & Surveillance Centre*. [cited 2015 7 April 2015]; Available from: http://www.rcgp.org.uk/clinical-and-research/research-and-surveillance-centre.aspx.

147. McDonald, H.I., et al., *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, 2013. **31**(5): p. 606-614.
148. Joseph, C., Y. Togawa, and N. Shindo, *Bacterial and viral infections associated with influenza.* Influenza Other Respir Viruses, 2013. **7 Suppl 2**: p. 105-13.
149. Public Health England. *Immunisation against infectious disease: the Green Book*. July 2013 2 September 2014 [cited 2014 18 November 2014]; Available from: https://www.gov.uk/government/collections/immunisation-against-infectious-disease-the-green-book.
150. Huss, A., et al., *Efficacy of pneumococcal vaccination in adults: a meta-analysis.* CMAJ, 2009. **180**(1): p. 48-58.
151. Moberley, S., et al., *Vaccines for preventing pneumococcal infection in adults.* Cochrane Database Syst Rev, 2013. **1**: p. CD000422.
152. Mangtani, P., F. Cutts, and A.J. Hall, *Efficacy of polysaccharide pneumococcal vaccine in adults in more developed countries: the state of the evidence.* Lancet Infect Dis, 2003. **3**(2): p. 71-8.
153. Ochoa-Gondar, O., et al., *Effectiveness of the 23-valent pneumococcal polysaccharide vaccine against community-acquired pneumonia in the general population aged >/= 60 years: 3 years of follow-up in the CAPAMIS study.* Clin Infect Dis, 2014. **58**(7): p. 909-17.
154. Nelson, J.C., et al., *New strategies are needed to improve the accuracy of influenza vaccine effectiveness estimates among seniors.* J Clin Epidemiol., 2009. **62**(7): p. 687-94.
155. Jackson, L.A., et al., *Evidence of bias in estimates of influenza vaccine effectiveness in seniors.* Int J Epidemiol, 2006. **35**(2): p. 337-44.
156. Jackson, L.A., et al., *Functional status is a confounder of the association of influenza vaccine and risk of all cause mortality in seniors.* Int J Epidemiol, 2006. **35**(2): p. 345-52.
157. Hak, E., et al., *Confounding by indication in non-experimental evaluation of vaccine effectiveness: the example of prevention of influenza complications.* J Epidemiol Community Health 2002. **56**(12): p. 951-955.
158. Jackson, M.L., et al., *Influenza vaccination and risk of community-acquired pneumonia in immunocompetent elderly people: a population-based, nested case-control study.* Lancet, 2008. **372**(9636): p. 398-405.
159. Ridenhour, B.J., et al., *Effectiveness of inactivated influenza vaccines in preventing influenza-associated deaths and hospitalizations among Ontario residents aged >/= 65 years: estimates with generalized linear models accounting for healthy vaccinee effects.* PLoS One, 2013. **8**(10): p. e76318.
160. McDonald, H.I., et al., *Chronic kidney disease and the risk of acute, community-acquired infections among older people with diabetes mellitus: a retrospective cohort study using electronic health records.* American Journal of Kidney Diseases, 2015. **In press.**(Available online 30 January 2015.): p. doi:10.1053/j.ajkd.2014.11.027.
161. Dalrymple, L.S., et al., *The risk of infection-related hospitalization with decreased kidney function.* Am J Kidney Dis, 2012. **59**(3): p. 356-63.
162. McGrath, L., et al., *Influenza vaccine effectiveness in patients on hemodialysis.* Arch Intern Med., 2012. **172**(7): p. 548-554.
163. Office for National Statistics. *Office for National Statistics*. 2012 14 September 2012]; Available from: www.ons.gov.uk.

164. Elliot, A.J., K.W. Cross, and D.M. Fleming, *Acute respiratory infections and winter pressures on hospital admissions in England and Wales 1990–2005.* J Public Health (Oxf), 2008. **30**(1): p. 91-98.
165. Michiels, B., et al., *A systematic review of the evidence on the effectiveness and risks of inactivated influenza vaccines in different target groups.* Vaccine, 2011. **29**(49): p. 9159-70.
166. Lau, D., et al., *Effectiveness of influenza vaccination in working-age adults with diabetes: a population-based cohort study.* Thorax, 2013. **68**(7): p. 658-63.
167. Colquhoun, A.J., et al., *Effectiveness of influenza vaccine in reducing hospital admissions in people with diabetes.* Epidemiol Infect, 1997. **119**(3): p. 335-41.
168. Looijmans-Van den Akker, I., et al., *Clinical effectiveness of first and repeat influenza vaccination in adult and elderly diabetic patients.* Diabetes Care, 2006. **29**(8): p. 1771-6.
169. Wang, I.K., et al., *Effectiveness of influenza vaccination in elderly diabetic patients: a retrospective cohort study.* Vaccine, 2013. **31**(4): p. 718-24.
170. Hak, E., et al., *Influence of high-risk medical conditions on the effectiveness of influenza vaccination among elderly members of 3 large managed-care organizations.* Clin Infect Dis, 2002. **35**(4): p. 370-7.
171. Gilbertson, D.T., et al., *The association of pneumococcal vaccination with hospitalization and mortality in hemodialysis patients.* Nephrol Dial Transplant, 2011. **26**(9): p. 2934-9.
172. van Kruijsdijk, R.C., et al., *Cause-specific mortality and years of life lost in patients with different manifestations of vascular disease.* Eur J Prev Cardiol, 2015.
173. Tsai, C.W., et al., *Cystatin C- and creatinine-based estimated glomerular filtration rate, vascular disease, and mortality in persons with diabetes in the U.S.* Diabetes Care, 2014. **37**(4): p. 1002-8.
174. Hassan Sadek, N., et al., *Evaluating tools to support a new practical classification of diabetes: excellent control may represent misdiagnosis and omission from disease registers is associated with worse control.* Int J Clin Pract, 2012. **66**(9): p. 874-82.
175. Stevens, L.A., et al., *Comparison of the CKD Epidemiology Collaboration (CKD-EPI) and Modification of Diet in Renal Disease (MDRD) study equations: risk factors for and complications of CKD and mortality in the Kidney Early Evaluation Program (KEEP).* Am J Kidney Dis, 2011. **57**(3 Suppl 2): p. S9-16.
176. Campbell, C.N.J., et al., *Hospitalization in two waves of pandemic influenza A(H1N1) in England.* Epidemiology & Infection, 2011. **139**(10): p. 1560-9.

Appendix A: Supplementary material for Paper 1 - Chronic kidney disease as a risk factor for acute community-acquired infections in high-income countries: a systematic review

This appendix contains the supplementary material referred to in the systematic review presented in **Chapter 2**. It comprises: full Medline, Embase and Cochrane search strategies (**Tables 1–3)**; inclusion and exclusion criteria used to determine study eligibility (**Table 4**); details of the quality assessment of each study (**Table 5**); and a funnel plot showing the relationship between relative risk and standard error for the 17 estimates from all 12 studies considered for meta-analysis (**Figure 1**).

Table A 1 Medline search strategy

	Search	Results
1	(sepsis* or septic?emia or bacter?emia or fung?emia or pneumonia* or bronchopneumonia* or pleuropneumonia* or LRTI or empy?ema or influenza* or legionell* or bacteriuri* or pyelonephriti* or cystitis* or pyelocystitis* or pyelitis* or urethriti* or UTI or meningiti* or meningococc* or encephaliti* or poliomyeliti* or septic shock).tw.	343181
2	(CNS adj4 infection*).tw.	2545
3	(central nervous adj4 infection*).tw.	3805
4	exp cerebral phaeohyphomycosis/ or central nervous system infections/ or exp brain abscess/ or exp toxoplasmosis, cerebral/ or central nervous system bacterial infections/ or exp empyema, subdural/ or exp epidural abscess/ or exp lyme neuroborreliosis/ or exp meningitis, bacterial/ or exp meningitis, escherichia coli/ or exp meningitis, haemophilus/ or exp meningitis, listeria/ or exp meningitis, meningococcal/ or exp meningitis, pneumococcal/ or exp central nervous system fungal infections/ or exp meningitis, fungal/ or exp meningitis, cryptococcal/ or exp neuroaspergillosis/ or central nervous system viral diseases/ or exp encephalitis/ or exp encephalitis, viral/ or exp encephalitis, arbovirus/ or exp encephalitis, california/ or exp encephalitis, japanese/ or exp "encephalitis, st. louis"/ or exp encephalitis, tick-borne/ or exp west nile fever/ or exp encephalitis, herpes simplex/ or exp encephalitis, varicella zoster/ or exp encephalomyelitis, equine/ or exp meningitis, viral/ or exp lymphocytic choriomeningitis/ or exp meningitis, aseptic/ or exp paraparesis, tropical spastic/ or poliomyelitis/ or exp poliomyelitis, bulbar/ or exp encephalomyelitis/ or exp meningitis/	102876

5	exp endocarditis, bacterial/ or exp endocarditis, subacute bacterial/ or exp pneumococcal infections/ or catheter-related infections/ or exp coinfection/ or communicable diseases/ or exp community-acquired infections/ or exp sepsis/ or exp bacteremia/ or exp hemorrhagic septicemia/ or exp fungemia/ or exp shock, septic/ or exp empyema/ or exp viremia/ or exp parasitemia/	139752
6	((renal or kidney) adj4 chronic adj4 failure*).tw.	21053
7	((renal or kidney) adj4 chronic adj4 disease*).tw.	15978
8	((renal or kidney) adj4 chronic adj4 insufficienc*).tw.	4448
9	((renal or kidney) adj4 chronic adj4 injury).tw.	454
10	((renal or kidney) adj4 chronic adj4 impairment*).tw.	336
11	(creatinine* or GFR or eGFR or albuminuri* or proteinuri* or microalbuminuri* or nephropath* or glomerulo* or nephr#ti* or nephrosi* or ur?emia or ESRD or CKD or cardio-renal or Kimmelstiel-Wilson).tw.	194742
12	Creatinine/bl [Blood]	25724
13	Kidney Diseases/co, ep [Complications, Epidemiology]	11809
14	exp diabetic nephropathies/ or exp hypertension, renal/ or exp nephritis/ or exp anti-glomerular basement membrane disease/ or exp glomerulonephritis, iga/ or exp glomerulonephritis, membranoproliferative/ or exp glomerulonephritis, membranous/ or exp lupus nephritis/ or exp nephrosclerosis/ or exp nephrosis/ or exp renal insufficiency/ or exp cardio-renal syndrome/ or exp uremia/ or exp azotemia/ or exp proteinuria/	234481
15	kidney function tests/ or exp glomerular filtration rate/	44837
16	Animals/	4889105
17	Humans/	12139628
18	16 not (16 and 17)	3594930
19	Adult/	3567838
20	exp child/ or exp child, preschool/ or exp infant/	1849722
21	20 not (19 and 20)	1265383
22	Case reports/	1557478
23	developing countries/ or exp africa/ or cuba/ or dominica/ or dominican republic/ or grenada/ or guadeloupe/ or haiti/ or jamaica/ or exp central america/ or "gulf of mexico"/ or latin america/ or exp south america/ or exp asia, central/ or borneo/ or cambodia/ or east timor/ or indonesia/ or laos/ or malaysia/ or mekong valley/ or myanmar/ or philippines/ or thailand/ or vietnam/ or bangladesh/ or india/ or afghanistan/ or iran/ or iraq/ or jordan/ or lebanon/ or syria/ or turkey/ or yemen/ or nepal/ or	620630

	pakistan/ or sri lanka/ or exp china/ or "democratic people's republic of korea"/ or mongolia/ or albania/ or latvia/ or lithuania/ or bosnia-herzegovina/ or bulgaria/ or "republic of belarus"/ or romania/ or exp russia/ or serbia/ or ukraine/ or yugoslavia/ or exp transcaucasia/ or exp indian ocean islands/ or fiji/ or papua new guinea/ or vanuatu/ or palau/ or hawaii/	
24	developed countries/ or bahamas/ or barbados/ or netherlands antilles/ or puerto rico/ or "trinidad and tobago"/ or "virgin islands of the united states"/ or canada/ or greenland/ or united states/ or brunei/ or singapore/ or bahrain/ or israel/ or kuwait/ or oman/ or qatar/ or saudi arabia/ or united arab emirates/ or hong kong/ or macau/ or exp japan/ or "republic of korea"/ or bermuda/ or exp australia/ or andorra/ or austria/ or belgium/ or estonia/ or croatia/ or czech republic/ or hungary/ or poland/ or slovakia/ or slovenia/ or finland/ or exp france/ or exp germany/ or gibraltar/ or exp great britain/ or greece/ or iceland/ or ireland/ or exp italy/ or liechtenstein/ or luxembourg/ or cyprus/ or malta/ or monaco/ or netherlands/ or portugal/ or san marino/ or exp scandinavia/ or spain/ or switzerland/ or new zealand/ or new caledonia/ or guam/	1800832
25	23 not (23 and 24)	556094
26	Postoperative complications.sh.	263650
27	(incidence* or odds ratio or risk ratio or risk factor or relative risk).tw.	608698
28	(respiratory adj3 infection*).tw.	28563
29	(lower respiratory adj3 infection*).tw.	4633
30	(urinary adj3 infection*).tw.	28333
31	(upper urinary adj3 infection*).tw.	312
32	6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15	366856
33	(exp incidence/ or exp multivariate analysis/ or exp odds ratio/ or exp logistic models/ or exp risk factors/ or exp epidemiologic studies/).sh.	1799348
34	(exp pyelitis/ or exp pyelocystitis/ or exp pyelonephritis/ or exp urethritis/ or cystitis or urinary tract infections or exp pyuria/).sh.	50526
35	(respiratory tract infections or exp bronchiolitis/ or exp bronchiolitis, viral/ or empyema, pleural or exp influenza, human/ or exp legionellosis/ or exp legionnaires' disease/ or exp lung abscess/ or exp lung diseases, fungal/ or exp lung diseases, parasitic/ or exp pneumonia/ or exp bronchopneumonia/ or exp pleuropneumonia/ or exp pneumonia, bacterial/ or exp chlamydial pneumonia/ or exp pneumonia, mycoplasma/ or exp pneumonia, pneumococcal/ or exp pneumonia, rickettsial/ or exp pneumonia, staphylococcal/ or exp pneumonia, pneumocystis/ or exp	155035

	pneumonia, viral/ or exp severe acute respiratory syndrome/ or exp tracheitis/ or exp whooping cough/).sh.	
36	1 or 2 or 3 or 4 or 5 or 28 or 29 or 30 or 31 or 34 or 35	585963
37	27 or 33	2098986
38	32 and 36 and 37	5940
39	38 not 18 not 21 not 22 not 25 not 26	3514
40	limit 39 to (english or french or german)	3163

Table A 2 Embase search strategy

	Search	Results
1	kidney failure/co, ep [Complication, Epidemiology]	13218
2	chronic kidney failure/co, ep [Complication, Epidemiology]	10827
3	exp proteinuria/co, ep [Complication, Epidemiology]	5456
4	uremia/co, ep [Complication, Epidemiology]	3030
5	glomerulus filtration rate/	43185
6	creatinine clearance/	17973
7	glomerulosclerosis/co, ep [Complication, Epidemiology]	450
8	kidney disease/co, ep [Complication, Epidemiology]	10406
9	analgesic nephropathy/co, ep [Complication, Epidemiology]	39
10	chronic kidney disease/co, ep [Complication, Epidemiology]	1733
11	diabetic nephropathy/co, ep [Complication, Epidemiology]	9683
12	allergic glomerulonephritis/co, ep [Complication, Epidemiology]	109
13	immune complex nephritis/co, ep [Complication, Epidemiology]	77
14	immunoglobulin A nephropathy/co, ep [Complication, Epidemiology]	678
15	kidney amyloidosis/co, ep [Complication, Epidemiology]	228
16	nephritis/co, ep [Complication, Epidemiology]	1053
17	glomerulitis/	456
18	Goodpasture syndrome/co, ep [Complication, Epidemiology]	31
19	immune complex nephritis/co, ep [Complication, Epidemiology]	77
20	interstitial nephritis/co, ep [Complication, Epidemiology]	901
21	lupus erythematosus nephritis/co, ep [Complication, Epidemiology]	850
22	nephrosis/co, ep [Complication, Epidemiology]	189
23	nephrotic syndrome/co, ep [Complication, Epidemiology]	2133
24	exp glomerulopathy/co, ep [Complication, Epidemiology]	5475
25	exp glomerulonephritis/co, ep [Complication, Epidemiology]	4296
26	exp kidney dysfunction/co, ep [Complication, Epidemiology]	1322
27	(creatinine* or GFR or eGFR or albuminuri* or proteinuri* or microalbuminuri* or nephropath* or glomerulo* or nephr#ti* or nephrosi* or ur?emia or ESRD or CKD or cardio-renal or Kimmelstiel-Wilson).tw.	282722
28	((renal or kidney) adj4 chronic adj4 failure*).tw.	28639
29	((renal or kidney) adj4 chronic adj4 disease*).tw.	23893
30	((renal or kidney) adj4 chronic adj4 insufficienc*).tw.	6425
31	((renal or kidney) adj4 chronic adj4 injury).tw.	631
32	((renal or kidney) adj4 chronic adj4 impairment*).tw.	501

33	exp infectious pneumonia/ or bacterial pneumonia/ or chlamydial pneumonia/ or group b streptococcal pneumonia/ or legionnaire disease/ or mycoplasma pneumonia/ or pneumocystis pneumonia/ or pulmonary candidiasis/ or severe acute respiratory syndrome/ or staphylococcal pneumonia/ or virus pneumonia/	50671
34	respiratory tract infection/ or exp influenza/ or laryngotracheobronchitis/ or lower respiratory tract infection/ or parainfluenza virus infection/ or respiratory syncytial virus infection/ or viral respiratory tract infection/	106624
35	avian influenza/	5081
36	chest infection/ or pertussis/	13997
37	bronchiolitis/ or laryngotracheobronchitis/ or tracheobronchitis/	10003
38	pleura empyema/	3703
39	pyuria/ or urinary tract infection/	66023
40	candiduria/ or kidney infection/	1502
41	kidney abscess/ or pyonephrosis/	1666
42	cystitis/	11865
43	pyelonephritis/ or acute pyelonephritis/	22138
44	brain infection/ or brain abscess/ or herpes simplex encephalitis/ or herpes zoster encephalitis/ or subdural empyema/ or tick borne encephalitis/ or virus encephalitis/	24862
45	central nervous system infection/ or epidural abscess/ or poliomyelitis/	38386
46	meningitis/ or bacterial meningitis/ or exp fungal meningitis/ or haemophilus meningitis/ or lymphocytic choriomeningitis/ or subdural empyema/ or virus meningitis/	57864
47	encephalitis/ or brain ventriculitis/ or eastern equine encephalitis/ or encephalomyelitis/ or epidemic encephalitis/ or meningoencephalitis/ or panencephalitis/ or primary amebic meningoencephalitis/	47288
48	exp meningococcosis/	11231
49	exp pneumococcal infection/	5729
50	exp group b streptococcal infection/ or group b streptococcal pneumonia/	405
51	exp bacteremia/ or staphylococcal bacteremia/	29638
52	bloodstream infection/	2518
53	candidemia/	1358
54	systemic mycosis/ or fungemia/ or invasive aspergillosis/ or invasive candidiasis/	5182
55	sepsis/ or bacteremia/ or septic shock/ or septicemia/ or urosepsis/	140091
56	viremia/	12287
57	parasitemia/	6918

58	(sepsis* or septic?emia or bacter?emia or fung?emia or pneumonia* or bronchopneumonia* or pleuropneumonia* or LRTI or empy?ema or influenza* or legionell* or bacteriuri* or pyelonephritis or cystitis or pyelocystitis or pyelitis or urethriti* or meningiti* or meningococc* or encephaliti* or poliomyeliti* or septic shock).tw.	497436
59	(CNS adj4 infection*).tw.	3591
60	(central nervous adj4 infection*).tw.	4861
61	UTI.tw.	6684
62	bronchopneumonia/	8394
63	arachnoiditis/ or aseptic meningitis/ or epidemic meningitis/ or group b streptococcal meningitis/ or meningoencephalitis/ or pneumococcal meningitis/	21305
64	exp epidemiology/ or exp incidence/	1705072
65	exp risk factor/	513022
66	exp attributable risk/	1487
67	exp hazard ratio/	11362
68	statistical model/	87903
69	(odds adj1 ratio).tw.	101865
70	(relative adj2 ratio).tw.	2736
71	case report/	1892302
72	developing country/	71459
73	developed country/	25618
74	postoperative complication/ or postoperative infection/ or surgical infection/	272218
75	exp Africa/	196804
76	argentina/ or bolivia/ or brazil/ or chile/ or colombia/ or ecuador/ or french guiana/ or guyana/ or paraguay/ or peru/ or suriname/ or uruguay/ or venezuela/	98392
77	exp Central America/	15618
78	china/ or mongolia/ or philippines/	82530
79	borneo/ or cambodia/ or indonesia/ or laos/ or malaysia/ or myanmar/ or papua new guinea/ or thailand/ or timor-leste/ or viet nam/	53670
80	North Korea/	237
81	latvia/ or lithuania/	3316
82	albania/ or armenia/ or azerbaijan/ or belarus/ or "bosnia and herzegovina"/ or bulgaria/ or "georgia (republic)"/ or "macedonia (republic)"/ or romania/ or russian federation/ or serbia/ or ukraine/	83374
83	USSR/	50149

84	iran/ or iraq/ or jordan/ or lebanon/ or "turkey (republic)"/ or yemen/	49920
85	kazakhstan/ or kyrgyzstan/ or tajikistan/ or turkmenistan/ or uzbekistan/	5682
86	afghanistan/ or bangladesh/ or india/ or nepal/ or pakistan/ or sikkim/ or sri lanka/	105351
87	cuba/ or dominica/ or dominican republic/ or grenada/ or guadeloupe/ or haiti/ or jamaica/	11346
88	fiji/ or philippines/ or polynesia/	8607
89	exp Indian Ocean/	2505
90	Mexico/	28748
91	72 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	789122
92	exp Western Europe/	911511
93	croatia/ or czech republic/ or hungary/ or poland/ or slovakia/ or slovenia/	73494
94	Estonia/	2056
95	canada/ or united states/	1031054
96	japan/ or macao/	115065
97	South Korea/	4982
98	bahrain/ or cyprus/ or israel/ or kuwait/ or oman/ or qatar/ or saudi arabia/ or united arab emirates/	37707
99	exp "Australia and New Zealand"/	129186
100	brunei darussalam/ or hong kong/ or singapore/	21427
101	73 or 92 or 93 or 94 or 95 or 96 or 97 or 98 or 99 or 100	2259038
102	91 not (91 and 101)	710496
103	treatment outcome/	579285
104	editorial/	438527
105	embryo/	177038
106	infant/	533322
107	child/	1295310
108	preschool child/	469034
109	school child/	217344
110	adolescent/	1180705
111	adult/	4186945
112	105 or 106 or 107 or 108 or 109 or 110	2546570
113	112 not (112 and 111)	1658687
114	animal model/	630310
115	animal experiment/	1606715
116	nonhuman/	3807183
117	animal/	1773703

118	human/	13422168
119	114 or 115 or 116 or 117	5921124
120	119 not (119 and 118)	4747089
121	pneumonia/	97950
122	lung infection/ or hantavirus pulmonary syndrome/ or lung abscess/ or viral bronchiolitis/	21795
123	(respiratory adj3 infection*).tw.	43371
124	(lower respiratory adj3 infection*).tw.	6553
125	(urinary adj3 infection*).tw.	44177
126	(upper urinary adj3 infection*).tw.	444
127	(epidemiolog$ or incidence).tw.	878025
128	(relative adj risk*).tw.	55195
129	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32	364340
130	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 121 or 122 or 123 or 124 or 125 or 126	851259
131	64 or 65 or 66 or 67 or 68 or 69 or 70 or 127 or 128	2659100
132	129 and 130 and 131	7357
133	132 not 120 not 113 not 104 not 71 not 74 not 102	4970
134	limit 133 to (english or french or german)	4602
135	limit 134 to embase	4247

Table A 3 Cochrane library search strategy

	Search	Results
1	sepsis* or septic*mia or bacter*mia or fung*mia or pneumonia* or bronchopneumonia* or pleuropneumonia* or LRTI or empy*ma or influenza* or legionell* or bacteriuri* or pyelonephriti* or cystitis* or pyelocystitis* or pyelitis* or urethriti* or UTI or meningiti* or meningococc* or encephaliti* or poliomyeliti* or "septic shock"	19098
2	CNS near/4 infection*	47
3	"central nervous" near/4 infection*	92
4	[mh "cerebral phaeohyphomycosis"] or [mh ^"central nervous system infections"] or [mh "brain abscess"] or [mh "toxoplasmosis, cerebral"] or [mh ^"central nervous system bacterial infections"] or [mh "empyema, subdural"] or [mh "epidural abscess"] or [mh "lyme neuroborreliosis"] or [mh "meningitis, bacterial"] or [mh "meningitis, escherichia coli"] or [mh "meningitis, haemophilus"] or [mh "meningitis, listeria"] or [mh "meningitis, meningococcal"] or [mh "meningitis, pneumococcal"] or [mh "central nervous system fungal infections"] or [mh "meningitis, fungal"] or [mh "meningitis, cryptococcal"] or [mh neuroaspergillosis] or [mh ^"central nervous system viral diseases"] or [mh encephalitis] or [mh "encephalitis, viral"] or [mh "encephalitis, arbovirus"] or [mh "encephalitis, california"] or [mh "encephalitis, japanese"] or [mh "encephalitis, st. louis"] or [mh "encephalitis, tick-borne"] or [mh "west nile fever"] or [mh "encephalitis, herpes simplex"] or [mh "encephalitis, varicella zoster"] or [mh "encephalomyelitis, equine"] or [mh "meningitis, viral"] or [mh "lymphocytic choriomeningitis"] or [mh "meningitis, aseptic"] or [mh "paraparesis, tropical spastic"] or [mh ^poliomyelitis] or [mh "poliomyelitis, bulbar"] or [mh encephalomyelitis] or [mh meningitis]	1015
5	[mh "endocarditis, bacterial"] or [mh "endocarditis, subacute bacterial"] or [mh "pneumococcal infections"] or [mh ^"catheter-related infections"] or [mh coinfection] or [mh ^"communicable diseases"] or [mh "community-acquired infections"] or [mh sepsis] or [mh bacteremia] or [mh "hemorrhagic septicemia"] or [mh fungemia] or [mh "shock, septic"] or [mh empyema] or [mh viremia] or [mh parasitemia]	4033
6	respiratory near/3 infection*	4398
7	urinary near/3 infection*	3732
8	[mh pyelitis] or [mh pyelocystitis] or [mh pyelonephritis] or [mh urethritis] or [mh ^cystitis] or [mh ^"urinary tract infections"] or [mh pyuria]	2143
9	[mh ^"respiratory tract infections"] or [mh bronchiolitis] or [mh "bronchiolitis, viral"] or [mh ^"empyema, pleural"] or [mh "influenza, human"] or [mh legionellosis] or [mh "legionnaires' disease"] or [mh "lung abscess"] or [mh "lung diseases, fungal"] or exp [mh "lung diseases, parasitic"] or [mh pneumonia] or [mh bronchopneumonia] or [mh pleuropneumonia] or [mh "pneumonia, bacterial"] or [mh "chlamydial pneumonia"] or [mh "pneumonia, mycoplasma"] or [mh "pneumonia, pneumococcal"] or [mh "pneumonia, rickettsial"] or [mh "pneumonia, staphylococcal"] or [mh "pneumonia, pneumocystis"] or [mh "pneumonia, viral"] or [mh "severe acute respiratory syndrome"] or [mh tracheitis] or [mh "whooping cough"]	5402
10	(renal or kidney) near/4 chronic near/4 failure*	4476
11	(renal or kidney) near/4 chronic near/4 disease*	1647
12	(renal or kidney) near/4 chronic near/4 insufficienc*	510
13	(renal or kidney) near/4 chronic near/4 injury	29
14	(renal or kidney) near/4 chronic near/4 impairment*	34

15	creatinine* or GFR or eGFR or albuminuri* or proteinuri* or microalbuminuri* or nephropath* or glomerulo* or nephr?ti* or nephrosi* or ur*mia or ESRD or CKD or cardio-renal or Kimmelstiel-Wilson	16810
16	[mh ^creatinine/BL]	2042
17	[mh ^"kidney diseases"/CO,EP]	341
18	[mh "diabetic nephropathies"] or [mh "hypertension, renal"] or [mh nephritis] or [mh "anti-glomerular basement membrane disease"] or [mh "glomerulonephritis, iga"] or [mh "glomerulonephritis, membranoproliferative"] or [mh "glomerulonephritis, membranous"] or [mh "lupus nephritis"] or [mh nephrosclerosis] or [mh nephrosis] or [mh "renal insufficiency"] or [mh "cardio-renal syndrome"] or [mh uremia] or [mh azotemia] or [mh proteinuria]	7117
19	[mh ^"kidney function tests"] or [mh "glomerular filtration rate"]	2417
20	{or #1-#9}	25511
21	[175-#19]	21120
22	{and #20-#21}	1422
23	incidence* or "odds ratio" or "risk ratio" or "risk factor" or "relative risk"	69239
24	[mh incidence] or [mh "multivariate analysis"] or [mh "odds ratio"] or [mh "logistic models"] or [mh "risk factors"] or [mh "epidemiologic studies"]	122866
25	{or #23-#24}	165844
26	{and #22, #25}	953

Table A 4 Inclusion and exclusion criteria for determining study eligibility

	Included	Excluded
Participants	Adult human participants.	Populations exclusively of: - pregnant women; - kidney transplant recipients or patients receiving renal replacement therapy; - patient groups usually managed in secondary care unless this was for chronic kidney disease, or routinely treated with immunosuppressive medication.
Study settings	High income countries (World Bank classification). Community settings, including adults living in institutional care.	
Exposure of interest	Chronic acquired kidney disease, indicated by any of the following: - medical diagnosis; - reduced estimated glomerular filtration rate; - reduced creatinine clearance; - elevated creatinine; - proteinuria, micro- or macro-albuminuria; - renal structural abnormalities. Where there was no 'unexposed' group without kidney disease, comparison between stages 1-2 and stages 3-5 CKD was accepted.	
Outcomes of interest	Incidence rate ratio, risk ratio or odds ratio estimates of the effect of kidney disease on any of the following community-acquired acute infections: - lower respiratory tract infections; - urinary tract infections (UTIs); - central nervous system infections; - sepsis. Urinary catheter-associated UTIs from community settings, and incidence of severe disease (such as hospitalisation for infection) were accepted.	Outcomes not accepted: - infection prevalence; - hospital-associated infection rates; - post-operative follow up outcomes; - incidence of infection-related mortality; - prognosis among infected patients.
Study methodology	Trials, case-control studies, cohort studies or other observational study designs containing original data. Relevant review articles without original data were identified for reference list screening.	Case reports. Descriptive studies without a comparison group. Studies with fewer than 30 participants in either the exposed or unexposed categories.
Publication details	Any publication date. Languages: English, German, French.	

Table A 5 Quality assessment of studies including rationale (n=14)

Case-control studies	Vinogradova 2009 (16)	Watt 2007 (17)	Loeb 2009 (18)	Schnoor 2007 (19)
Selection bias				
Selection of controls [1]	Low: matched selection of primary care registered patients	Low: neighbourhood controls selected systematically by proximity	Low: random digit dialling of hospital catchment area residents	Low: random selection from population register
Participation bias [2]	Low: automatic participation	Low: participation 83%	Uncertain: participation not reported	High: Participation <60% overall
Loss to follow up [3]	N/A: case-control study	N/A: case-control study	N/A: case-control study	N/A: case-control study
Non-differential misclassification of exposure [4]	High: relies on medical diagnosis of chronic renal disease in medical records.	High: relied on medical diagnosis of chronic renal disease in medical records.	High: ascertained medical diagnosis of chronic renal disease in participant interview.	High: ascertained medical diagnosis of chronic renal disease in questionnaire for controls
Information bias: exposure				
Recall bias [5]	Low: kidney disease diagnosis ascertained from pre-existing medical records	Low: kidney disease diagnosis ascertained from pre-existing medical records	High: ascertained medical diagnosis of kidney disease in hospital interview for cases and at home for controls	High: ascertained medical diagnosis of kidney disease at home for controls
Observer bias [6]	Low: used pre-specified codes to define kidney disease status	Uncertain: Medical record abstractors not blinded to case-control status and criteria for assigning kidney disease status not reported	High: interviewers aware of case status (interviewed in hospital) or control status (telephone interview at home)	High: decision to list diagnosis of kidney disease in case report made in context of illness for cases
Ascertainment [7]	Low: chronic kidney disease diagnosis would have to predate current acute infection	Low: chronic kidney disease diagnosis would have to predate current acute infection	Low: chronic kidney disease diagnosis would have to predate current acute infection	High: ascertainment entirely different for cases than controls
Non-differential misclassification of outcome [8]	Low: medical diagnosis of severe outcome	Low: active surveillance with clear criteria	Low: severe outcome with clear criteria	Low: severe outcome with clear criteria
Information bias: outcome				
Recall bias [9]	Low: cases identified from medical records based on GP diagnosis	Low: cases identified by laboratory surveillance	Low: cases determined by medical diagnosis in hospital	Low: Low: realtime reporting system through established surveillance network
Observer bias [10]	Low: clinical diagnosis of severe outcome unlikely to be severely affected by kidney disease comorbidity	Low: Laboratory based surveillance system with clear criteria for cases	Low: CKD status unlikely to severely affect physician application of clear criteria	Low: surveillance system with clear criteria for cases
Ascertainment[11]	Low: kidney disease status unlikely to affect primary care attendance with severe outcome	Low: active surveillance with clear criteria, testing for IPD unlikely to be markedly influenced by CKD status in context of known high incidence among the Navajo Nation	Low: kidney disease status unlikely to affect hospital attendance with severe outcome	Low: kidney disease status unlikely to affect primary care or hospital attendance with severe outcome
Confounding [12]	Low: controls matched to cases on age and sex, estimate adjusted for wide range of confounders including diabetes[13]	Low: controls matched for age and sex. Diabetes eligible for inclusion in final model [14]	Low. Age, sex and diabetes eligible for inclusion in final model [15]	High: unadjusted
Reverse causation [18]	Low: pre-existing kidney disease reported at time of infection	Low: pre-existing kidney disease reported at time of infection	Low: pre-existing kidney disease reported at time of infection	Low: pre-existing kidney disease reported at time of infection

Cohort studies	Higgins 1989 (22)	Hackam 2006 (24)	Dalrymple 2012 (23)	Karunajeewa 2005 (25)	James 2008 (26)	James 2009 (27)	Wang 2012 (28)	Caljouw 2011 (29)	Campbell 2011 * (21)	USRDS 2010(20)
Selection bias										
Selection of controls [1]	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study
Participation bias [2]	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study	N/A: cohort study
Loss to follow up [3]	Uncertain: not reported and clinically determined. May be differential, but study period only one year.	Low: automated follow up	Low: >80% follow up	Low: automated follow up	Low: automated follow up	Low: automated follow up	Low: >80% follow-up	Low: followed up 479 of 551 participants alive aged 86 years: 86% follow up	Low: active case-finding applied to national census figure (no follow up required)	Low: automated follow up
Non-differential misclassification of exposure [4]	Uncertain: not reported when creatinine measured.	High: relies on medical diagnosis of chronic renal disease in medical records.	Low: determined prospectively from blood results	Low: determined prospectively from test results	Low: determined prospectively from blood results	Low: determined prospectively from blood results	Low: determined prospectively from blood results	Low: determined prospectively from blood results	High: relies on medical diagnosis of chronic renal disease in medical records.	High: relies on medical diagnosis of chronic renal disease in insurance claims
Information bias: exposure										
Recall bias [5]	Low: determined from serum creatinine with clear cut-off (objective measure)	Low: kidney disease diagnosis ascertained from pre-existing medical records	Low: determined prospectively from blood results.	Low: determined prospectively from test results.	Low: determined prospectively from blood results.	Low: determined prospectively from blood results.	Low: determined prospectively from blood results.	Low: determined prospectively from blood results.	Low: kidney disease diagnosis ascertained from pre-existing medical records	Low: kidney disease diagnosis ascertained from pre-existing insurance records
Observer bias [6]	Low: determined from serum creatinine with clear cut-off (objective measure)	Uncertain: source of kidney disease status data not reported. If hospital records are used, decision to list diagnosis in discharge record made in context	Low: determined from serum cystatin C (objective measure)	Low: determined from blood and urine test results (objective measure)	Low: determined from serum creatinine (objective measure)	Low: determined from serum creatinine (objective measure)	Low: determined from serum creatinine (objective measure)	Low: determined from serum creatinine (objective measure)	High: decision to list diagnosis of kidney disease in case report made in context of illness for cases	Low: used pre-specified codes to define kidney disease status

		of illness for cases.								
Ascertainment [7]	Uncertain: not reported when creatinine measured, or whether this is recurrent/ prompted by illness	Uncertain: source of kidney disease status data not reported. If hospital records used, patients with infection-related hospitalisations more likely to have CKD status recorded.	Low: all participants tested at baseline.	Low: participants monitored annually.	Low: baseline measure used (that only patients with a result were eligible was considered a limitation to generalisability)	Low: sensitivity analysis using only the baseline creatinine test found similar results to the last-carried forward method	Low: all participants tested at baseline.	Low: all participants tested at baseline.	High: ascertainment entirely different for cases than non-cases	Low: kidney disease status ascertained in year prior to study
Non-differential misclassification of outcome [8]	Uncertain: methods for ascertaining infection not reported	Low: severe outcome with widely accepted clinical criteria	Low: severe outcome with clear criteria	Low: severe outcome with widely accepted clinical criteria	Low: severe outcome with clear criteria	Low: severe outcome with widely accepted clinical criteria	Low: severe outcome with clear criteria	Uncertain: kidney disease status may affect healthcare attendance for minor illness such as UTI	Uncertain: sending of PCR test during influenza pandemic vulnerable to be influenced by kidney disease status	Low: severe outcome unlikely to be missed
Information bias: outcome										
Recall bias [9]	Uncertain: methods for ascertaining infection not reported	Low: monitoring of all hospital discharge reports	Low: semi-annual cohort monitoring	Low: monitoring of all hospital discharge reports	Low: monitoring of all biochemistry results	Low: monitoring of all hospital discharge reports	Low: semi-annual cohort monitoring	Low: annual clinician interviews supplemented with medical record review	Low: realtime case finding system through laboratory results	Low: monitoring of all hospital insurance claims
Observer bias [10]	Uncertain: standard definition of APN is vague and not reported whether any observer blinded to renal status	Low: kidney disease status unlikely to affect choice of hospital diagnosis code for severe outcome	Low: kidney disease status unlikely to affect choice of hospital diagnosis code for severe outcome	Low: kidney disease status unlikely to affect choice of hospital diagnosis code for severe outcome	Low: objective definition of outcome independent of exposure status	Low: kidney disease status unlikely to affect choice of hospital diagnosis code for severe outcome	Low: CKD status unlikely to severely affect application of clear criteria	Low: kidney disease status unlikely to strongly influence diagnosis of UTI at age 86-89 years, given case criteria include symptoms and urinary analysis	Low: objective criteria for cases once tested	Low: kidney disease status unlikely to affect choice of hospital diagnosis code for severe outcome

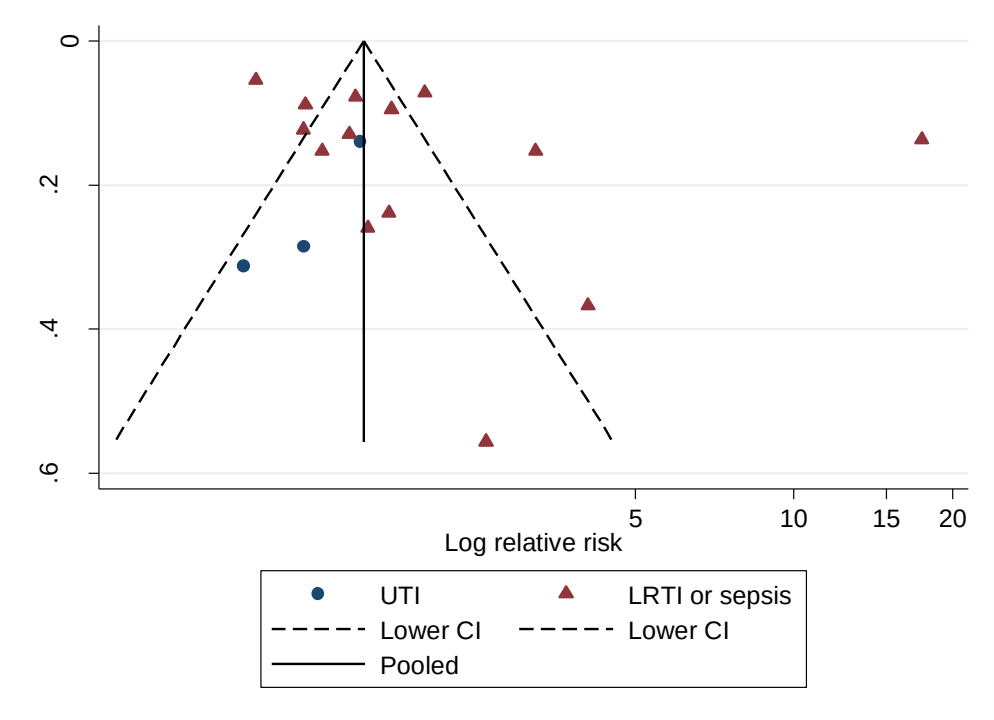
Ascertainment[11]	Uncertain: methods for ascertaining infection not reported	Low: kidney disease status unlikely to affect diagnosis of severe outcome with widely accepted clinical criteria	Low: kidney disease status unlikely to affect hospital attendance with severe outcome	Low: kidney disease status unlikely to affect hospital attendance with severe outcome	Low: sending of blood culture unlikely to be influenced by kidney disease in context of severe illness	Low: kidney disease status unlikely to affect diagnosis of severe outcome with widely accepted clinical	Low: kidney disease status unlikely to affect hospital attendance with severe outcome	Uncertain: kidney disease status may affect healthcare attendance for minor illness such as UTI	Uncertain: sending of PCR test during influenza pandemic vulnerable to be influenced by comorbidities	Low: kidney disease status unlikely to affect hospital attendance with severe outcome
Confounding [12]	High: unadjusted estimate. In particular, high immunosuppressant use among the study population	Low: adjusted for age, sex, nature of index event, charlson index, healthcare use, and other comorbidities	Low: adjusted for age, sex, race, smoking, BMI, diabetes mellitus, and multiple co-morbidities.	High: no adjustment for sex [16]	Low: adjusted for age, sex, diabetes, comorbidity score, care in a dedicated renal clinic	Low: adjusted for age, sex, socio-economic status, ethnicity, diabetes mellitus, Charlson comorbidity score	High: adjusted for age, sex, alcohol, smoking and demographic factors but no comorbitidies.	High: no adjustment for sex or diabetes [17]	High: adjusted for age only	High: unadjusted [19]
Reverse causation [18]	Uncertain: Timing of creatinine measurement relative to infections not specified	Low: chronic renal failure should not be diagnosed within one hospital episode for infection	Low: baseline serum cystatin C used	Low: serum biochemistry tested at screening	Low: baseline creatinine used	Low: only followed to first outcome event, creatinine result predates qualifying infection	Low: baseline creatinine used	Low: baseline creatinine used	Low: pre-existing kidney disease reported at time of infection	Low: kidney disease status established in year prior to study

*The unusual design of the artificial cohort study by Campbell *et al*. is worth clarification. During the 2009–2010 influenza pandemic, hospitalised cases of laboratory confirmed pandemic influenza A (H1N1) were reported by a laboratory-based national surveillance system. The surveillance report, completed by the microbiologist, asked whether the case had a diagnosis of CKD. Denominators for infection rates were obtained from primary care registers of patients eligible for pandemic influenza vaccination by virtue of a CKD diagnosis (for CKD): and from the national census (for non-CKD).[176] The effect of CKD on influenza may be overestimated in this study, because CKD was advertised as a possible risk factor for pandemic influenza to encourage vaccine uptake among this group, and patients with flu-like symptoms could have been more likely to attend hospital or be tested for pandemic influenza A (H1N1) if they had a diagnosis of CKD.

1. High risk: probability of selection as a control likely to be affected by kidney disease status (known or unknown).
Low risk: controls selected using random sampling (or other system unlikely to be biased by kidney disease status) from the population from which the cases arose.
2. Low risk: (1) automated participation (e.g. medical record review), or (2) ≥80% participation, or (3) 70-80% participation with a comparison (min age, sex, death/morbidity) showing similar characteristics between those included and those not included in the study.
3. Low risk: (1) automated follow up (e.g. through record linkage), or (2) ≥80% follow up, or (3) 70-80% follow up with a comparison (min age, sex, death/morbidity) showing similar characteristics between those included and those not included in the study.
4. High risk: allocation of kidney disease status relies on existing kidney disease having been diagnosed as part of routine medical care.
Low risk: All members of study assessed for kidney disease at baseline.
5. High risk: kidney disease status defined by patient recall of CKD diagnosis in context of recent infection.
6. High risk: kidney disease status defined by observer unblinded to case status, without clear objective criteria to apply.
7. High risk: participants with infections are more or less likely to be tested for kidney disease.
8. Low risk: Active screening for infection, or severe outcome unlikely to be missed or presents validation results of >70% sensitivity and specificity

9. High risk: infection status defined by patient recall of infection in context of kidney disease e.g. participants with kidney disease asked to recall infections while at renal clinic.
10. High risk: infection status defined by observer in context likely to be influenced by kidney disease status (assumed that clinical diagnosis of severe infections was unlikely to be strongly influenced by awareness of CKD as a comorbidity but that less severe infections may be influenced by this in the absence of clear diagnostic criteria).
11. High risk: ascertainment of infections likely to be influenced by kidney disease status (assumed that attendance to healthcare facility for severe infections was unlikely to be strongly influenced by awareness of CKD as a comorbidity but that attendance for less severe infections may be influenced by this in the absence of active surveillance).
12. Low risk: At least age, sex and diabetes must have been eligible and considered for the final model.
13. Controls matched to cases on age at index data (within 1 year), sex, general practice, and calendar time. Estimate adjusted for smoking status, Townsend deprivation score, use of influenza vaccine in previous 12 months, use of pneumococcal vaccine in previous 5 years, number of years of medical records data available in database, and comorbidities including: diabetes, chronic heart disease, chronic respiratory disease, asplenia, cerebrospinal shunt, chronic liver disease, sickle cell disease or coeliac disease, cochlear implant, HIV/AIDS, immunosuppression, stroke or transient ischaemic attack, rheumatoid arthritis, Parkinson's disease, multiple sclerosis, dementia, osteoporosis, and any cancer.
14. Controls matched for age and sex. Diabetes eligible for inclusion in final model, which was adjusted for age, pneumococcal vaccine, congestive heart failure, alcohol use, BMI and unemployment.
15. Age, sex and diabetes eligible for inclusion in final model. Final model adjusted for age, non-English language at home, living in a detached house, living alone, congestive heart failure, chronic obstructive pulmonary disease, dysphagia, Barthel index of functional status, immunosuppressive medications, nutritional score, tobacco use, alcohol use, and exposure to fumes.
16. The best adjusted estimate for urinary sepsis is adjusted for asymptomatic bacteriuria and age, and restricted to diabetics, but not adjusted for sex.
17. High risk: age-restricted and stratified by long term care facility (LCTF) residency, but no management of confounding by sex or co-morbidities.
18. High risk: exposure defined after the infection defined as the study outcome.
19. Rate ratios for hospitalisation with UTI, pneumonia and bacteraemia/ sepsis unadjusted. Rate ratios for hospitalisation with any infection calculated as the ratio of the rate among participants with CKD, compared with no CKD, each rate having been adjusted for gender, prior hospitalisation, ASHD, CHF, CVA, PVD, dysrhythmia, other cardiac disease, diabetes, COPD, hypertension, liver disease, gastrointestinal disease, cancer, and anaemia.

Figure A 1 Funnel plot showing the relationship between relative risk and standard error for the 17 estimates from all 12 studies considered for meta-analysis (all infections combined)



UTI = urinary tract infection LRTI= lower respiratory tract infections

Appendix B: Codelists for identifying diabetes mellitus

This appendix contains the Read codes used to identify a definite or possible diagnosis of diabetes mellitus, and the medication codes used to identify insulin and oral antidiabetes medications.

Table B 1 Read codes used to identify a 'definite' diagnosis of diabetes mellitus

Compiled by Sara Thomas.

Read code	Read term
R054200	[D]Gangrene of toe in diabetic
R054300	[D]Widespread diabetic foot gangrene
ZV6DA00	[V]Admitted for commencement of insulin
ZV6DB00	[V]Admitted for conversion to insulin
ZV65312	[V]Dietary counselling in diabetes mellitus
U602312	[X] Adverse reaction to insulins
U602311	[X] Adverse reaction to insulins and antidiabetic agents
U60231E	[X] Adverse reaction to insulins and antidiabetic agents NOS
Cyu2.00	[X]Diabetes mellitus
Kyu0300	[X]Glomerular disorders in diabetes mellitus
Cyu2000	[X]Other specified diabetes mellitus
Cyu2300	[X]Unspecified diabetes mellitus with renal complications
F372000	Acute painful diabetic neuropathy
8H2J.00	Admit diabetic emergency
F420300	Advanced diabetic maculopathy
F420500	Advanced diabetic retinal disease
TJ23000	Adverse reaction to insulins
TJ23.00	Adverse reaction to insulins and antidiabetic agents
TJ23z00	Adverse reaction to insulins and antidiabetic agents NOS
66AT.00	Annual diabetic blood test
F372200	Asymptomatic diabetic neuropathy
9OLH.00	Attended DAFNE diabetes structured education programme
9OLG.00	Attended XPERT diabetes structured education programme
9NM0.00	Attending diabetes clinic
F171100	Autonomic neuropathy due to diabetes
F420000	Background diabetic retinopathy
66AJ100	Brittle diabetes
C350011	Bronzed diabetes
M037200	Cellulitis in diabetic foot
F372100	Chronic painful diabetic neuropathy
7L10000	Continuous subcutaneous infusion of insulin
66AH000	Conversion to insulin
ZRB4.11	CSQ - Diabetes clinic satisfaction questionnaire
C10N100	Cystic fibrosis related diabetes mellitus
9OLJ.00	DAFNE diabetes structured education programme completed
66AN.00	Date diabetic treatment start
889A.00	Diab mellit insulin-glucose infus acute myocardial infarct
66AU.00	Diabetes care by hospital only

8CS0.00	Diabetes care plan agreed
ZRB4.00	Diabetes clinic satisfaction questionnaire
8CR2.00	Diabetes clinical management plan
66AR.00	Diabetes management plan given
8B3l.00	Diabetes medication review
C104z00	Diabetes mellitis with nephropathy NOS
C10..00	Diabetes mellitus
C10C.00	Diabetes mellitus autosomal dominant
C10D.00	Diabetes mellitus autosomal dominant type 2
C102z00	Diabetes mellitus NOS with hyperosmolar coma
C101z00	Diabetes mellitus NOS with ketoacidosis
C103z00	Diabetes mellitus NOS with ketoacidotic coma
C106z00	Diabetes mellitus NOS with neurological manifestation
C100z00	Diabetes mellitus NOS with no mention of complication
C105z00	Diabetes mellitus NOS with ophthalmic manifestation
C10yz00	Diabetes mellitus NOS with other specified manifestation
C107z00	Diabetes mellitus NOS with peripheral circulatory disorder
C10zz00	Diabetes mellitus NOS with unspecified complication
C107.11	Diabetes mellitus with gangrene
C102.00	Diabetes mellitus with hyperosmolar coma
C101.00	Diabetes mellitus with ketoacidosis
C103.00	Diabetes mellitus with ketoacidotic coma
C106.00	Diabetes mellitus with neurological manifestation
C106.12	Diabetes mellitus with neuropathy
C100.00	Diabetes mellitus with no mention of complication
C105.00	Diabetes mellitus with ophthalmic manifestation
C10y.00	Diabetes mellitus with other specified manifestation
C107.00	Diabetes mellitus with peripheral circulatory disorder
C106.13	Diabetes mellitus with polyneuropathy
C104.00	Diabetes mellitus with renal manifestation
C10z.00	Diabetes mellitus with unspecified complication
C106100	Diabetes mellitus, adult onset, + neurological manifestation
C105100	Diabetes mellitus, adult onset, + ophthalmic manifestation
C10z100	Diabetes mellitus, adult onset, + unspecified complication
C100100	Diabetes mellitus, adult onset, no mention of complication
C102100	Diabetes mellitus, adult onset, with hyperosmolar coma
C101100	Diabetes mellitus, adult onset, with ketoacidosis
C103100	Diabetes mellitus, adult onset, with ketoacidotic coma
C104100	Diabetes mellitus, adult onset, with renal manifestation
C107200	Diabetes mellitus, adult with gangrene
C10y100	Diabetes mellitus, adult, + other specified manifestation
C107100	Diabetes mellitus, adult, + peripheral circulatory disorder
C107000	Diabetes mellitus, juvenile +peripheral circulatory disorder
C105000	Diabetes mellitus, juvenile type, + ophthalmic manifestation
C10z000	Diabetes mellitus, juvenile type, + unspecified complication
C100000	Diabetes mellitus, juvenile type, no mention of complication
C102000	Diabetes mellitus, juvenile type, with hyperosmolar coma
C101000	Diabetes mellitus, juvenile type, with ketoacidosis
C103000	Diabetes mellitus, juvenile type, with ketoacidotic coma
C104000	Diabetes mellitus, juvenile type, with renal manifestation
C106000	Diabetes mellitus, juvenile, + neurological manifestation

ZRB5.00	Diabetes treatment satisfaction questionnaire
3883	Diabetes treatment satisfaction questionnaire
66An.00	Diabetes type 1 review
66Ao.00	Diabetes type 2 review
3882	Diabetes well being questionnaire
ZRB6.00	Diabetes wellbeing questionnaire
C107.12	Diabetes with gangrene
66AP.00	Diabetes: practice programme
66AQ.00	Diabetes: shared care programme
66AK.00	Diabetic - cooperative patient
66AM.00	Diabetic - follow-up default
66AI.00	Diabetic - good control
66AJ.00	Diabetic - poor control
66AJz00	Diabetic - poor control NOS
66Ai.00	Diabetic 6 month review
F381311	Diabetic amyotrophy
C106.11	Diabetic amyotrophy
66AS.00	Diabetic annual review
F464000	Diabetic cataract
N030100	Diabetic Charcot arthropathy
N030000	Diabetic cheiroarthropathy
N030011	Diabetic cheiropathy
8A12.00	Diabetic crisis monitoring
68AB.00	Diabetic digital retinopathy screening offered
66AG.00	Diabetic drug side effects
66Ab.00	Diabetic foot examination
8I3W.00	Diabetic foot examination declined
8I6G.00	Diabetic foot examination not indicated
66AW.00	Diabetic foot risk assessment
66Aq.00	Diabetic foot screen
F440700	Diabetic iritis
F420400	Diabetic maculopathy
F345000	Diabetic mononeuritis multiplex
F35z000	Diabetic mononeuritis NOS
F3y0.00	Diabetic mononeuropathy
C104.11	Diabetic nephropathy
F372.12	Diabetic neuropathy
66A3.00	Diabetic on diet only
66A5.00	Diabetic on insulin
66AV.00	Diabetic on insulin and oral treatment
66A4.00	Diabetic on oral treatment
9OLD.00	Diabetic patient unsuitable for digital retinal photography
G73y000	Diabetic peripheral angiopathy
66Ac.00	Diabetic peripheral neuropathy screening
F372.11	Diabetic polyneuropathy
6761	Diabetic pre-pregnancy counselling
F420.00	Diabetic retinopathy
8HBG.00	Diabetic retinopathy 12 month review
8HBH.00	Diabetic retinopathy 6 month review
F420z00	Diabetic retinopathy NOS
68A7.00	Diabetic retinopathy screening

8I6F.00	Diabetic retinopathy screening not indicated
68A9.00	Diabetic retinopathy screening offered
8I3X.00	Diabetic retinopathy screening refused
8A13.00	Diabetic stabilisation
66AH.00	Diabetic treatment changed
13AC.00	Diabetic weight reducing diet
66AL.00	Diabetic-uncooperative patient
8HLE.00	Diabetology D.V. done
8HKE.00	Diabetology D.V. requested
9NiC.00	Did not attend DAFNE diabetes structured education programme
9N4p.00	Did not attend diabetic retinopathy clinic
9NiE.00	Did not attend XPERT diabetes structured education programme
8I82.00	Did not complete DAFNE diabetes structured education program
8I84.00	Did not complete XPERT diabetes structured education program
ZC2C800	Dietary advice for diabetes mellitus
ZC2C900	Dietary advice for type I diabetes
ZC2CA00	Dietary advice for type II diabetes
ZLD7500	Discharge by diabetic liaison nurse
8Hg4.00	Discharged from care of diabetes specialist nurse
ZRB5.11	DTSQ - Diabetes treatment satisfaction questionnaire
ZRB6.11	DWBQ - Diabetes wellbeing questionnaire
2G5C.00	Foot abnormality - diabetes related
2G51000	Foot abnormality - diabetes related
14F4.00	H/O: Admission in last year for diabetes foot problem
14P3.00	H/O: insulin therapy
42W..00	Hb. A1C - diabetic control
42WZ.00	Hb. A1C - diabetic control NOS
42c..00	HbA1 - diabetic control
F420800	High risk non proliferative diabetic retinopathy
F420700	High risk proliferative diabetic retinopathy
C10FK00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus
C109K00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus
C110000	Iatrogenic hyperinsulinism
C108.11	IDDM-Insulin dependent diabetes mellitus
9M00.00	Informed consent for diabetes national audit
9M10.00	Informed dissent for diabetes national audit
66AA.11	Injection sites - diabetic
C110.11	Insulin coma
C108900	Insulin dependant diabetes maturity onset
C108800	Insulin dependant diabetes mellitus - poor control
C108J00	Insulin dependent diab mell with neuropathic arthropathy
C108G00	Insulin dependent diab mell with peripheral angiopathy
C10E912	Insulin dependent diabetes maturity onset
C100011	Insulin dependent diabetes mellitus
C108.00	Insulin dependent diabetes mellitus
C10E.12	Insulin dependent diabetes mellitus
C10E812	Insulin dependent diabetes mellitus - poor control
C108H00	Insulin dependent diabetes mellitus with arthropathy
C108F00	Insulin dependent diabetes mellitus with diabetic cataract
C10EF12	Insulin dependent diabetes mellitus with diabetic cataract
C108600	Insulin dependent diabetes mellitus with gangrene

C108E00	Insulin dependent diabetes mellitus with hypoglycaemic coma
C10EE12	Insulin dependent diabetes mellitus with hypoglycaemic coma
C108B00	Insulin dependent diabetes mellitus with mononeuropathy
C10E312	Insulin dependent diabetes mellitus with multiple complicat
C108300	Insulin dependent diabetes mellitus with multiple complicatn
C108D00	Insulin dependent diabetes mellitus with nephropathy
C108C00	Insulin dependent diabetes mellitus with polyneuropathy
C10EC12	Insulin dependent diabetes mellitus with polyneuropathy
C108700	Insulin dependent diabetes mellitus with retinopathy
C10E712	Insulin dependent diabetes mellitus with retinopathy
C108500	Insulin dependent diabetes mellitus with ulcer
C10E512	Insulin dependent diabetes mellitus with ulcer
66Am.00	Insulin dose changed
9kL..00	Insulin initiation - enhanced services administration
M21yC00	Insulin lipohypertrophy
66Ag.00	Insulin needles changed daily
66Ah.00	Insulin needles changed for each injection
66Aj.00	Insulin needles changed less than once a day
M21yC11	Insulin site lipohypertrophy
8I3k.00	Insulin therapy declined
C109J11	Insulin treated non-insulin dependent diabetes mellitus
C10FJ00	Insulin treated Type 2 diabetes mellitus
C109J00	Insulin treated Type 2 diabetes mellitus
C109J12	Insulin treated Type II diabetes mellitus
C10FJ11	Insulin treated Type II diabetes mellitus
66Ap.00	Insulin treatment initiated
C108200	Insulin-dependent diabetes mellitus with neurological comps
C108100	Insulin-dependent diabetes mellitus with ophthalmic comps
C10E112	Insulin-dependent diabetes mellitus with ophthalmic comps
C108000	Insulin-dependent diabetes mellitus with renal complications
C108A00	Insulin-dependent diabetes without complication
C10EA12	Insulin-dependent diabetes without complication
M271000	Ischaemic ulcer diabetic foot
C10ER00	Latent autoimmune diabetes mellitus in adult
C10M.00	Lipoatrophic diabetes mellitus
8HME.00	Listed for Diabetology admissn
C100111	Maturity onset diabetes
C10C.11	Maturity onset diabetes in youth
C10C.12	Maturity onset diabetes in youth type 1
C10D.11	Maturity onset diabetes in youth type 2
M271200	Mixed diabetic ulcer - foot
F381300	Myasthenic syndrome due to diabetic amyotrophy
K01x100	Nephrotic syndrome in diabetes mellitus
M271100	Neuropathic diabetic ulcer - foot
C109.11	NIDDM - Non-insulin dependent diabetes mellitus
F420600	Non proliferative diabetic retinopathy
C109E00	Non-insulin depend diabetes mellitus with diabetic cataract
C109700	Non-insulin dependent diabetes mellitus - poor control
C109H00	Non-insulin dependent d m with neuropathic arthropathy
C100112	Non-insulin dependent diabetes mellitus
C109.00	Non-insulin dependent diabetes mellitus

C109G00	Non-insulin dependent diabetes mellitus with arthropathy
C109500	Non-insulin dependent diabetes mellitus with gangrene
C109D00	Non-insulin dependent diabetes mellitus with hypoglyca coma
C109A00	Non-insulin dependent diabetes mellitus with mononeuropathy
C109C00	Non-insulin dependent diabetes mellitus with nephropathy
C109B00	Non-insulin dependent diabetes mellitus with polyneuropathy
C109400	Non-insulin dependent diabetes mellitus with ulcer
C109F00	Non-insulin-dependent d m with peripheral angiopath
C109300	Non-insulin-dependent diabetes mellitus with multiple comps
C109200	Non-insulin-dependent diabetes mellitus with neuro comps
C109100	Non-insulin-dependent diabetes mellitus with ophthalm comps
C109000	Non-insulin-dependent diabetes mellitus with renal comps
C109600	Non-insulin-dependent diabetes mellitus with retinopathy
C109900	Non-insulin-dependent diabetes mellitus without complication
8H3O.00	Non-urgent diabetic admission
2BBL.00	O/E - diabetic maculopathy present both eyes
2G5W.00	O/E - left chronic diabetic foot ulcer
2G5L.00	O/E - Left diabetic foot - ulcerated
2G5K.00	O/E - Left diabetic foot at high risk
2G5I.00	O/E - Left diabetic foot at low risk
2G5J.00	O/E - Left diabetic foot at moderate risk
2G5B.00	O/E - Left diabetic foot at risk
2BBQ.00	O/E - left eye background diabetic retinopathy
2BBX.00	O/E - left eye diabetic maculopathy
2BBS.00	O/E - left eye preproliferative diabetic retinopathy
2BBV.00	O/E - left eye proliferative diabetic retinopathy
2BBl.00	O/E - left eye stable treated prolif diabetic retinopathy
2G5V.00	O/E - right chronic diabetic foot ulcer
2G5H.00	O/E - Right diabetic foot - ulcerated
2G5G.00	O/E - Right diabetic foot at high risk
2G5E.00	O/E - Right diabetic foot at low risk
2G5F.00	O/E - Right diabetic foot at moderate risk
2G5A.00	O/E - Right diabetic foot at risk
2BBP.00	O/E - right eye background diabetic retinopathy
2BBW.00	O/E - right eye diabetic maculopathy
2BBR.00	O/E - right eye preproliferative diabetic retinopathy
2BBT.00	O/E - right eye proliferative diabetic retinopathy
2BBk.00	O/E - right eye stable treated prolif diabetic retinopathy
2BBo.00	O/E - sight threatening diabetic retinopathy
7276	Pan retinal photocoagulation for diabetes
93C4.00	Patient consent given for addition to diabetic register
9360	Patient held diabetic record issued
8BL2.00	Patient on maximal tolerated therapy for diabetes
ZRbH.00	Perceived control of insulin-dependent diabetes
F372.00	Polyneuropathy in diabetes
L180500	Pre-existing diabetes mellitus, insulin-dependent
L180600	Pre-existing diabetes mellitus, non-insulin-dependent
L180X00	Pre-existing diabetes mellitus, unspecified
F420200	Preproliferative diabetic retinopathy
F420100	Proliferative diabetic retinopathy
8H7r.00	Refer to diabetic foot screener

8Hl1.00	Referral for diabetic retinopathy screening
8Hj3.00	Referral to DAFNE diabetes structured education programme
8HTk.00	Referral to diabetic eye clinic
8HHy.00	Referral to diabetic register
8Hj5.00	Referral to XPERT diabetes structured education programme
2BBF.00	Retinal abnormality - diabetes related
C10G.00	Secondary pancreatic diabetes mellitus
C10G000	Secondary pancreatic diabetes mellitus without complication
9N1v.00	Seen in diabetic eye clinic
9N1i.00	Seen in diabetic foot clinic
7L19800	Subcutaneous injection of insulin
8CP2.00	Transition of diabetes care options discussed
C10E.00	Type 1 diabetes mellitus
C108.12	Type 1 diabetes mellitus
C10E800	Type 1 diabetes mellitus - poor control
C108812	Type 1 diabetes mellitus - poor control
C10E900	Type 1 diabetes mellitus maturity onset
C108912	Type 1 diabetes mellitus maturity onset
C10EH00	Type 1 diabetes mellitus with arthropathy
C10EF00	Type 1 diabetes mellitus with diabetic cataract
C10EP00	Type 1 diabetes mellitus with exudative maculopathy
C10E600	Type 1 diabetes mellitus with gangrene
C10EQ00	Type 1 diabetes mellitus with gastroparesis
C10EE00	Type 1 diabetes mellitus with hypoglycaemic coma
C108E12	Type 1 diabetes mellitus with hypoglycaemic coma
C10EM00	Type 1 diabetes mellitus with ketoacidosis
C10EN00	Type 1 diabetes mellitus with ketoacidotic coma
C10EB00	Type 1 diabetes mellitus with mononeuropathy
C10E300	Type 1 diabetes mellitus with multiple complications
C10ED00	Type 1 diabetes mellitus with nephropathy
C10E200	Type 1 diabetes mellitus with neurological complications
C108212	Type 1 diabetes mellitus with neurological complications
C108J12	Type 1 diabetes mellitus with neuropathic arthropathy
C10EJ00	Type 1 diabetes mellitus with neuropathic arthropathy
C10E100	Type 1 diabetes mellitus with ophthalmic complications
C10EG00	Type 1 diabetes mellitus with peripheral angiopathy
C10EL00	Type 1 diabetes mellitus with persistent microalbuminuria
C10EK00	Type 1 diabetes mellitus with persistent proteinuria
C10EC00	Type 1 diabetes mellitus with polyneuropathy
C108012	Type 1 diabetes mellitus with renal complications
C10E000	Type 1 diabetes mellitus with renal complications
C10E700	Type 1 diabetes mellitus with retinopathy
C108712	Type 1 diabetes mellitus with retinopathy
C10E500	Type 1 diabetes mellitus with ulcer
C108512	Type 1 diabetes mellitus with ulcer
C10EA00	Type 1 diabetes mellitus without complication
C10F.00	Type 2 diabetes mellitus
C109.12	Type 2 diabetes mellitus
C10F700	Type 2 diabetes mellitus - poor control
C109712	Type 2 diabetes mellitus - poor control
C109G12	Type 2 diabetes mellitus with arthropathy

C10FG00	Type 2 diabetes mellitus with arthropathy
C109E12	Type 2 diabetes mellitus with diabetic cataract
C10FE00	Type 2 diabetes mellitus with diabetic cataract
C10FQ00	Type 2 diabetes mellitus with exudative maculopathy
C10F500	Type 2 diabetes mellitus with gangrene
C109512	Type 2 diabetes mellitus with gangrene
C10FR00	Type 2 diabetes mellitus with gastroparesis
C10FD00	Type 2 diabetes mellitus with hypoglycaemic coma
C109D12	Type 2 diabetes mellitus with hypoglycaemic coma
C10FN00	Type 2 diabetes mellitus with ketoacidosis
C10FP00	Type 2 diabetes mellitus with ketoacidotic coma
C10FA00	Type 2 diabetes mellitus with mononeuropathy
C10F300	Type 2 diabetes mellitus with multiple complications
C10FC00	Type 2 diabetes mellitus with nephropathy
C109C12	Type 2 diabetes mellitus with nephropathy
C10F200	Type 2 diabetes mellitus with neurological complications
C109212	Type 2 diabetes mellitus with neurological complications
C10FH00	Type 2 diabetes mellitus with neuropathic arthropathy
C109H12	Type 2 diabetes mellitus with neuropathic arthropathy
C10F100	Type 2 diabetes mellitus with ophthalmic complications
C109112	Type 2 diabetes mellitus with ophthalmic complications
C10FF00	Type 2 diabetes mellitus with peripheral angiopathy
C109F12	Type 2 diabetes mellitus with peripheral angiopathy
C10FM00	Type 2 diabetes mellitus with persistent microalbuminuria
C10FL00	Type 2 diabetes mellitus with persistent proteinuria
C10FB00	Type 2 diabetes mellitus with polyneuropathy
C109012	Type 2 diabetes mellitus with renal complications
C10F000	Type 2 diabetes mellitus with renal complications
C10F600	Type 2 diabetes mellitus with retinopathy
C109612	Type 2 diabetes mellitus with retinopathy
C10F400	Type 2 diabetes mellitus with ulcer
C109412	Type 2 diabetes mellitus with ulcer
C10F900	Type 2 diabetes mellitus without complication
C10E.11	Type I diabetes mellitus
C108.13	Type I diabetes mellitus
C108811	Type I diabetes mellitus - poor control
C108911	Type I diabetes mellitus maturity onset
C10E911	Type I diabetes mellitus maturity onset
C108H11	Type I diabetes mellitus with arthropathy
C108F11	Type I diabetes mellitus with diabetic cataract
C10EP11	Type I diabetes mellitus with exudative maculopathy
C108E11	Type I diabetes mellitus with hypoglycaemic coma
C10EM11	Type I diabetes mellitus with ketoacidosis
C10EN11	Type I diabetes mellitus with ketoacidotic coma
C108B11	Type I diabetes mellitus with mononeuropathy
C10E311	Type I diabetes mellitus with multiple complications
C108D11	Type I diabetes mellitus with nephropathy
C108211	Type I diabetes mellitus with neurological complications
C108J11	Type I diabetes mellitus with neuropathic arthropathy
C10E111	Type I diabetes mellitus with ophthalmic complications
C10EC11	Type I diabetes mellitus with polyneuropathy

C108011	Type I diabetes mellitus with renal complications
C108711	Type I diabetes mellitus with retinopathy
C10E711	Type I diabetes mellitus with retinopathy
C108511	Type I diabetes mellitus with ulcer
C10E511	Type I diabetes mellitus with ulcer
C10EA11	Type I diabetes mellitus without complication
C108A11	Type I diabetes mellitus without complication
C109.13	Type II diabetes mellitus
C10F.11	Type II diabetes mellitus
C109711	Type II diabetes mellitus - poor control
C10F711	Type II diabetes mellitus - poor control
C109G11	Type II diabetes mellitus with arthropathy
C109E11	Type II diabetes mellitus with diabetic cataract
C10FE11	Type II diabetes mellitus with diabetic cataract
C109511	Type II diabetes mellitus with gangrene
C109D11	Type II diabetes mellitus with hypoglycaemic coma
C10FD11	Type II diabetes mellitus with hypoglycaemic coma
C109A11	Type II diabetes mellitus with mononeuropathy
C10FA11	Type II diabetes mellitus with mononeuropathy
C10F311	Type II diabetes mellitus with multiple complications
C109C11	Type II diabetes mellitus with nephropathy
C109211	Type II diabetes mellitus with neurological complications
C10F211	Type II diabetes mellitus with neurological complications
C109H11	Type II diabetes mellitus with neuropathic arthropathy
C109111	Type II diabetes mellitus with ophthalmic complications
C10F111	Type II diabetes mellitus with ophthalmic complications
C109F11	Type II diabetes mellitus with peripheral angiopathy
C10FM11	Type II diabetes mellitus with persistent microalbuminuria
C10FL11	Type II diabetes mellitus with persistent proteinuria
C109B11	Type II diabetes mellitus with polyneuropathy
C10FB11	Type II diabetes mellitus with polyneuropathy
C109011	Type II diabetes mellitus with renal complications
C10F011	Type II diabetes mellitus with renal complications
C10F611	Type II diabetes mellitus with retinopathy
C109611	Type II diabetes mellitus with retinopathy
C109411	Type II diabetes mellitus with ulcer
C10F411	Type II diabetes mellitus with ulcer
C10F911	Type II diabetes mellitus without complication
9NND.00	Under care of diabetic foot screener
9NN8.00	Under care of diabetologist
C108z00	Unspecified diabetes mellitus with multiple complications
66AJ.11	Unstable diabetes
C108400	Unstable insulin dependant diabetes mellitus
C10E412	Unstable insulin dependent diabetes mellitus
C10E400	Unstable type 1 diabetes mellitus
C108412	Unstable type 1 diabetes mellitus
C10E411	Unstable type I diabetes mellitus
C108411	Unstable type I diabetes mellitus
9OLL.00	XPERT diabetes structured education programme completed
2BBr.00	Impaired vision due to diabetic retinopathy
C10E611	Type I diabetes mellitus with gangrene

C10E212	Insulin-dependent diabetes mellitus with neurological comps
C10ED12	Insulin dependent diabetes mellitus with nephropathy
C10FC11	Type II diabetes mellitus with nephropathy

Table B 2 Read codes used to identify a 'possible' diagnosis of diabetes mellitus

Compiled by Sara Thomas.

Read code	Read term
9OLB.00	Attended diabetes structured education programme
9OL1.00	Attends diabetes monitoring
9OLK.00	DESMOND diabetes structured education programme completed
9OL..11	Diabetes clinic administration
9OLA.00	Diabetes monitor. check done
9OL8.00	Diabetes monitor.phone invite
9OL7.00	Diabetes monitor.verbal invite
9OLA.11	Diabetes monitored
9OL4.00	Diabetes monitoring 1st letter
9OL5.00	Diabetes monitoring 2nd letter
9OL6.00	Diabetes monitoring 3rd letter
9OL..00	Diabetes monitoring admin.
9OLZ.00	Diabetes monitoring admin.NOS
9OL3.00	Diabetes monitoring default
9OL9.00	Diabetes monitoring deleted
9OLF.00	Diabetes structured education programme completed
9OLM.00	Diabetes structured education programme declined
13B1.00	Diabetic diet
66AY.00	Diabetic diet - good compliance
66Aa.00	Diabetic diet - poor compliance
13AB.00	Diabetic lipid lowering diet
66A..00	Diabetic monitoring
66Al.00	Diabetic monitoring - higher risk albumin excretion
66Ak.00	Diabetic monitoring - lower risk albumin excretion
66AZ.00	Diabetic monitoring NOS
66b1.00	Diabetic monitoring not required
9NiD.00	Did not attend DESMOND diabetes structured education program
9NiA.00	Did not attend diabetes structured education programme
8I83.00	Did not complete DESMOND diabetes structured educat program
8I81.00	Did not complete diabetes structured education programme
9N4I.00	DNA - Did not attend diabetic clinic
3881	Education score - diabetes
ZRBa.00	Education score - diabetes
9h41.00	Excepted from diabetes qual indicators: Patient unsuitable
9h42.00	Excepted from diabetes quality indicators: Informed dissent
9h4..00	Exception reporting: diabetes quality indicators
66A2.00	Follow-up diabetic assessment
66AD.00	Fundoscopy - diabetic check
44V3.00	Glucose tol. test diabetic
66A8.00	Has seen dietician - diabetes

66A1.00	Initial diabetic assessment
2BBM.00	O/E - diabetic maculopathy absent both eyes
2BBK.00	O/E - no left diabetic retinopathy
2BBJ.00	O/E - no right diabetic retinopathy
C103y00	Other specified diabetes mellitus with coma
C101y00	Other specified diabetes mellitus with ketoacidosis
C108y00	Other specified diabetes mellitus with multiple comps
C106y00	Other specified diabetes mellitus with neurological comps
C105y00	Other specified diabetes mellitus with ophthalmic complicatn
C10yy00	Other specified diabetes mellitus with other spec comps
C104y00	Other specified diabetes mellitus with renal complications
C10zy00	Other specified diabetes mellitus with unspecified comps
66Af.00	Patient diabetes education review
8I57.00	Patient held diabetic record declined
679R.00	Patient offered diabetes structured education programme
8HVU.00	Private referral to diabetologist
8CA4100	Pt advised re diabetic diet
8H7C.00	Refer, diabetic liaison nurse
8Hl4.00	Referral to community diabetes specialist nurse
8Hj4.00	Referral to DESMOND diabetes structured education programme
8H7f.00	Referral to diabetes nurse
ZL62500	Referral to diabetes nurse
8HTe.00	Referral to diabetes preconception counselling clinic
8H4e.00	Referral to diabetes special interest general practitioner
8Hj0.00	Referral to diabetes structured education programme
ZL62600	Referral to diabetic liaison nurse
8H4F.00	Referral to diabetologist
8HTi.00	Referral to multidisciplinary diabetic clinic
9OL2.00	Refuses diabetes monitoring
ZLA2500	Seen by diabetic liaison nurse
9N2i.00	Seen by diabetic liaison nurse
9N2d.00	Seen by diabetologist
9Nl4.00	Seen by general practitioner special interest in diabetes
9N0n.00	Seen in community diabetes specialist clinic
9N0o.00	Seen in community diabetic specialist nurse clinic
9N1Q.00	Seen in diabetic clinic
9N0m.00	Seen in diabetic nurse consultant clinic
9N1o.00	Seen in multidisciplinary diabetic clinic
9NN9.00	Under care of diabetes specialist nurse
ZL22500	Under care of diabetic liaison nurse
66A9.00	Understands diet - diabetes

Table B 3 Medication product codes identifying insulin prescriptions

Product code	Product name
22823	INSULIN ISOPHANE (PURIFIED) 100 I/U INJ
24722	INSULIN ISOPHANE 50%/NEUTRAL 50% 100 I/U INJ
9079	INSULIN SOLUBLE 100 I/U INJ
20196	INSULIN SOLUBLE 40 I/U INJ
26784	INSULIN ZINC SEMILENTE SUSP BP 100 I/U INJ
20195	INSULIN BOVINE PROTAMINE ZINC 40 I/U INJ
24866	INSULIN INSULATARD (LEO RETARD) 40 I/U INJ
23003	INSULIN ISOPHANE (NPH) 40 I/U
24845	INSULIN PUR-IN ISOPHANE 100 I/U INJ
28978	INSULIN PUR-IN MIX 15/85 100 I/U INJ
31267	INSULIN PUR-IN MIX 50/50 100 I/U INJ
7861	INSULIN HUMULIN S (NEUTRAL) CARTRIDGE 100 I/U
8646	INSULIN ZINC CRYSTALLINE susp 100 I/U INJ
7765	INSULIN NEUTRAL (HUMAN) 100 I/U INJ
10566	INSULIN HUMULIN M CARTRIDGE 100 I/U
12060	INSULIN QUICKSOL (SOLUBLE NEUTRAL) 100 I/U INJ
12244	INSULIN ZINC BOVINE susp 100 I/U INJ
15624	INSULIN ISOPHANE (HIGHLY PURIFIED) 100 I/U INJ
15040	INSULIN MONOPHANE (ISOPHANE) 100 I/U INJ
18645	INSULIN NEUTRAL (PURIFIED) 100 I/U INJ
4248	INSULIN NOVO ULTRATARD MC 100 I/U INJ
1839	INSULIN HUMULIN I (ISOPHANE) 100 I/U INJ
2373	INSULIN HUMAN VELOSULIN 100 I/U INJ
14504	INSULIN HYPURIN PROTAMINE ZINC 100 I/U INJ
10546	INSULIN HUMULIN M4 100 I/U INJ
10691	INSULIN ISOPHANE (NPH) 100 I/U INJ
8354	INSULIN ISOPHANE 70%/NEUTRAL 30% 100 I/U INJ
14506	INSULIN BOVINE PROTAMINE ZINC 100 I/U INJ
8839	INSULIN SEMITARD 100 I/U INJ
10545	INSULIN HUMULIN M4 CARTRIDGE 100 I/U
13550	INSULIN BP 100 I/U
2808	INSULIN LENTARD INJ
8838	INSULIN SEMITARD 40 I/U INJ
7959	INSULIN MIXTARD 30/70 40 I/U INJ
7757	INSULIN NEULENTE (ZINC SUSP)(PURIFIED) 100 I/U INJ
7763	INSULIN NEUPHANE (ISOPHANE)(PURIFIED) 100 I/U INJ
1645	INSULIN NOVO ACTRAPID MC 100 I/U INJ
321	INSULIN HUMAN ACTRAPID (NEUTRAL) 40 I/U INJ
16209	INSULIN HYPURIN SOLUBLE 100 I/U INJ
7764	INSULIN NEUSULIN (NEUTRAL)(PURIFIED) 100 I/U INJ
1643	INSULIN NOVO MONOTARD MC 100 I/U INJ
18301	INSULIN SOLUBLE INJ I/U^2
8376	INSULIN ISOPHANE 100 I/U

7783 INSULIN ISOPHANE (HUMAN) 100 I/U INJ
20672 INSULIN HUM/ACTRAPID
20671 INSULIN HUM/ACTRAPHANE
27911 INSULIN HUMAN ACTRAPID PENFILL
25006 INSULIN HUMAN ACTRAPID (NEUTRAL)
19707 INSULIN HUMULIN S (NEUTRAL SOLUBLE)
22161 INSULIN HUMULIN M1 VIAL
22094 INSULIN HUMULIN M2 VIAL
21945 INSULIN PORK INSULATARD
19829 INSULIN NOVO MONOTARD MC
28723 INSULIN ZINC BOVINE SUSPENSION
24485 INSULIN ZINC ANIMAL SUSPENSION
30861 INSULIN ZINC HUMAN SUSPENSION
22496 INSULIN ZINC LENTE PURIFIED SUSPENSION
22806 INSULIN PORK ACTRAPID
32053 INSULIN HUMALOG MIX 25
6209 NOVORAPID VIAL injection solution 100 units/ml [NOVO]
5021 NOVORAPID PENFILL injection solution 100 units/ml [NOVO]
11337 NOVORAPID NOVOLET injection 100 iu/ml [NOVO]
5892 NOVORAPID FLEXPEN injection solution 100 units/ml [NOVO]
6447 insulin aspart human pyr injection 100 iu/ml
322 HUMALOG injection 100 iu/ml [LILLY]
5214 insulin lispro human prb injection 100 iu/ml
28442 insulin glulisine injection solution 100 units/ml
19491 APIDRA VIAL injection solution 100 units/ml [SANOFI/AVE]
14345 APIDRA CARTRIDGE injection solution 100 units/ml [SANOFI/AVE]
29567 insulin aspart vial injection solution 100 units/ml
16142 insulin aspart cartridge injection solution 100 units/ml
19877 insulin aspart disposable pen injection solution 100 units/ml
26060 insulin lispro vial injection solution 100 units/ml
14313 insulin lispro cartridge injection solution 100 units/ml
14362 insulin lispro disposable pen injection solution 100 units/ml
18224 HUMALOG VIAL injection solution 100 units/ml [LILLY]
7318 HUMALOG CARTRIDGE injection solution 100 units/ml [LILLY]
10264 HUMALOG DISPOSABLE PEN injection solution 100 units/ml [LILLY]
28101 insulin glulisine vial injection solution 100 units/ml
14299 insulin glulisine cartridge injection solution 100 units/ml
21583 APIDRA OPTISET injection solution 100 units/ml [SANOFI/AVE]
21590 insulin glulisine disposable pen injection solution 100 units/ml
36355 insulin human powder for inhalation 1mg
36356 insulin human powder for inhalation 3mg
31465 EXUBERA powder for inhalation 1mg [PFIZER]
31467 EXUBERA powder for inhalation 3mg [PFIZER]
29953 APIDRA OPTICLIK injection solution 100 units/ml [SANOFI/AVE]
36920 APIDRA SOLOSTAR injection solution 100 units/ml [SANOFI/AVE]
38986 HUMALOG KWIKPEN injection solution 100 units/ml [LILLY]

1587 MONOTARD injection 100 units/ml [NOVO]
4760 HUMULIN I injection 100 units/ml [LILLY]
8895 INITARD 50/50 injection 100 units/ml [NOVO]
1843 PORK INSULATARD VIAL injection suspension 100 units/ml [NOVO]
4784 LENTARD MC injection 100 units/ml [NOVO]
2459 PORK MIXTARD 30 VIAL injection suspension 100 units/ml [NOVO]
34031 MONOTARD MC injection 100 units/ml [NOVO]
4163 RAPITARD MC injection 100 units/ml [NOVO]
12299 SEMITARD MC injection 100 units/ml [NOVO]
4715 HUMALOG MIX 25 injection 25:75; 100 units/ml [LILLY]
7537 HUMULIN ZN injection 100 units/ml [LILLY]
41834 INSULIN ZINC SUSPENSION LENTE injection 100 iu/ml [CELLTECH]
16682 TEMPULIN injection 100 units/ml [KNOLL]
7772 HUMAN PROTAPHANE injection 100 units/ml [NOVO]
7771 HUMAN PROTAPHANE PENFILL 100 units/ml [NOVO]
1844 ULTRATARD injection 100 units/ml [NOVO]
33966 INSULATARD injection 100 units/ml [NOVO]
1886 INSULATARD ge injection 100 iu/ml [NOVO]
1593 INSULATARD PENFILL 100 iu/ml [NOVO]
1805 MIXTARD 30/70 injection 100 units/ml [NOVO]
34097 HUMAN INITARD 50/50 injection 100 units/ml [NOVO]
38422 ISOPHANE injection 100 iu/ml [CELLTECH]
5501 INSUMAN BASAL injection 100 iu/ml [AVENTIS]
9503 HYPURIN BOVINE PROTAMINE ZINC VIAL injection suspension 100 units/ml [CP PHARM]
17712 HYPURIN BOVINE LENTE VIAL injection suspension 100 units/ml [CP PHARM]
13516 HYPURIN BOVINE ISOPHANE injection 100 units/ml [CP PHARM]
1649 HUMAN ACTRAPHANE injection 100 iu/ml [NOVO]
10547 HUMULIN LENTE injection 100 units/ml [LILLY]
4199 HUMULIN M1 injection 100 units/ml [LILLY]
4093 HUMULIN M2 injection 100 units/ml [LILLY]
4198 HUMULIN M3 injection 100 units/ml [LILLY]
11107 HUMULIN M4 injection 100 units/ml [LILLY]
8841 HUMULIN M5 injection 100 units/ml [LILLY]
13416 insulin biphasic injection 100 units/ml
16700 insulin zinc mixed bovine vial injection suspension 100 units/ml
8322 insulin zinc suspension mixed human pyr injection 100 units/ml
9376 insulin zinc suspension crystalline human pyr - long acting injection 100 units/ml
18931 insulin zinc suspension crystalline human prb - intermediate acting injection 100 units/ml
15484 insulin isophane bovine injection 100 units/ml
4247 insulin isophane porcine injection 100 units/ml
14619 insulin biphasic isophane porcine injection 30:70; 100 units/ml
14505 insulin protamine zinc bovine vial injection suspension 100 units/ml
12035 insulin zinc lente bovine vial injection suspension 100 units/ml
30236 isophane insulin injection 100 iu/ml

8203 PENMIX 50/50 PENFILL injection 100 iu/ml [NOVO]
10887 PENMIX 40/60 PENFILL injection 100 iu/ml [NOVO]
27614 PENMIX 30/70 injection 100 iu/ml [NOVO]
10484 PENMIX 20/80 PENFILL PENFILL [NOVO]
15961 insulin isophane human crb injection 100 iu/ml
3396 PENMIX 10/90 PENFILL PENFILL [NOVO]
26403 PUR-IN MIX 25/75 injection [CP PHARM]
22058 PUR-IN MIX 15/85 injection [CP PHARM]
1806 PENMIX 30/70 PENFILL injection 100 iu/ml [NOVO]
3439 PENMIX 10/90 pen [NOVO]
2220 PENMIX 20/80 pen [NOVO]
21347 PENMIX 40/60 injection 100 iu/ml [NOVO]
17731 PENMIX 50/50 injection 100 iu/ml [NOVO]
6057 LANTUS injection 100 iu/ml [AVENTIS]
33167 insulin biphasic isophane human crb injection 25:75; 100 units/ml
33232 insulin biphasic isophane human crb injection 50:50; 100 units/ml
10001 HUMALOG MIX 50 DISPOSABLE PEN injection suspension 100 units/ml [LILLY]
5953 insulin glargine injection 100 iu/ml
15199 INSUMAN COMB 25 injection 100 iu/ml [AVENTIS]
21554 INSUMAN COMB 50 injection 100 iu/ml [AVENTIS]
5891 INSULATARD FLEXPEN injection 100 iu/ml [NOVO]
1595 INSULATARD NOVOLET 100 iu/ml [NOVO]
5255 MIXTARD 10 PENFILL 100 iu/ml [NOVO]
2456 MIXTARD 10 NOVOLET 100 iu/ml [NOVO]
3551 MIXTARD 20 PENFILL 100 iu/ml [NOVO]
2455 MIXTARD 20 NOVOLET 100 iu/ml [NOVO]
2454 MIXTARD 30 PENFILL 100 iu/ml [NOVO]
2221 MIXTARD 30 NOVOLET 100 iu/ml [NOVO]
2929 MIXTARD 30 ge injection 100 iu/ml [NOVO]
3550 MIXTARD 40 PENFILL 100 iu/ml [NOVO]
2812 MIXTARD 40 NOVOLET 100 iu/ml [NOVO]
4790 MIXTARD 50 PENFILL 100 iu/ml [NOVO]
5933 MIXTARD 50 NOVOLET 100 iu/ml [NOVO]
12818 MIXTARD 50 injection 50:50; 100 units/ml [NOVO]
20422 INSUMAN COMB 15 injection 100 iu/ml [AVENTIS]
8118 HUMAJECT I pen 100 iu/ml [LILLY]
10915 HUMAJECT M1 pen 100 iu/ml [LILLY]
10910 HUMAJECT M2 pen 100 iu/ml [LILLY]
7793 HUMAJECT M3 pen 100 iu/ml [LILLY]
17809 HUMAJECT M4 pen 100 iu/ml [LILLY]
22155 HUMAJECT M5 pen 100 iu/ml [LILLY]
44251 insulin zinc suspension mixed porcine injection 100 units/ml
26498 insulin zinc suspension mixed bovine and porcine injection 100 units/ml
13837 insulin biphasic isophane human prb injection 10:90; 100 units/ml
14644 insulin biphasic isophane human prb injection 20:80; 100 units/ml
9341 insulin biphasic isophane human prb injection 30:70; 100 units/ml

10175 insulin isophane human pyr injection 100 iu/ml
11080 insulin isophane human prb injection 100 iu/ml
13729 insulin isophane human emp injection 100 units/ml
18461 insulin zinc suspension mixed human prb injection 100 units/ml
14649 insulin biphasic isophane human pyr injection 10:90; 100 units/ml
11055 insulin biphasic isophane human pyr injection 20:80; 100 units/ml
11056 insulin biphasic isophane human pyr injection 30:70; 100 units/ml
21395 insulin biphasic isophane human pyr injection 40:60; 100 units/ml
22697 insulin biphasic isophane human pyr injection 50:50; 100 units/ml
21374 insulin biphasic isophane human prb injection 40:60; 100 units/ml
21110 insulin biphasic isophane human prb injection 50:50; 100 units/ml
5845 MIXTARD 30 INNOLET injection suspension 30:70; 100 units/ml [NOVO]
9737 INSULATARD INNOLET injection 100 iu/ml [NOVO]
5250 insulin biphasic lispro human prb injection 25:75; 100 units/ml
27177 insulin biphasic lispro human prb injection 50:50; 100 units/ml
10067 insulin biphasic aspart human pyr injection 30:70; 100 units/ml
6061 NOVOMIX 30 injection 30:70; 100 units/ml [NOVO]
13819 HYPURIN PORCINE ISOPHANE injection 100 units/ml [CP PHARM]
9618 HYPURIN PORCINE 30/70 MIX injection 100 iu/ml [CP PHARM]
29837 insulin biphasic isophane human prb injection 25:75; 100 units/ml
10184 insulin detemir injection solution 100 iu/ml
6965 LEVEMIR PENFILL injection solution 100 units/ml [NOVO]
6958 LEVEMIR FLEXPEN injection solution 100 units/ml [NOVO]
25736 insulin biphasic isophane human cartridge injection suspension 10:90; 100 units/ml
10245 MIXTARD 10 PENFILL injection suspension 100 units/ml [NOVO]
25735 insulin biphasic isophane human cartridge injection suspension 20:80; 100 units/ml
7319 MIXTARD 20 PENFILL injection suspension 100 units/ml [NOVO]
16152 insulin biphasic isophane human cartridge injection suspension 30:70; 100 units/ml
21232 insulin biphasic isophane human vial injection suspension 30:70; 100 units/ml
19878 insulin biphasic isophane human disposable pen injection suspension 30:70; 100 units/ml
7231 MIXTARD 30 PENFILL injection suspension 100 units/ml [NOVO]
10277 HUMULIN M3 CARTRIDGE injection suspension 100 units/ml [LILLY]
19513 HUMULIN M3 VIAL injection suspension 100 units/ml [LILLY]
16160 HUMULIN M3 DISPOSABLE PEN injection suspension 100 units/ml [LILLY]
21422 insulin biphasic isophane human cartridge injection suspension 40:60; 100 units/ml
10244 MIXTARD 40 PENFILL injection suspension 100 units/ml [NOVO]
28096 insulin biphasic isophane human cartridge injection suspension 50:50; 100 units/ml
41120 insulin biphasic isophane human disposable pen injection suspension 50:50; 100 units/ml
35253 INSUMAN COMB 50 CARTRIDGE injection suspension 100 units/ml [AVENTIS]
31205 INSUMAN COMB 50 OPTISET injection suspension 100 units/ml [AVENTIS]
30819 INSUMAN COMB 15 OPTISET injection suspension 100 units/ml [AVENTIS]
42954 insulin biphasic isophane human vial injection suspension 25:75; 100 units/ml
36194 insulin biphasic isophane human cartridge injection suspension 25:75; 100 units/ml
44378 insulin biphasic isophane human disposable pen injection suspension 25:75; 100 units/ml

24002 INSUMAN COMB 25 VIAL injection suspension 100 units/ml [AVENTIS]
24993 INSUMAN COMB 25 CARTRIDGE injection suspension 100 units/ml [AVENTIS]
25133 INSUMAN COMB 25 OPTISET injection suspension 100 units/ml [AVENTIS]
14925 insulin isophane human vial injection suspension 100 units/ml
10207 insulin isophane human cartridge injection suspension 100 units/ml
25812 insulin isophane human disposable pen injection suspension 100 units/ml
10225 LANTUS OPTICLIK injection solution 100 units/ml [AVENTIS]
14918 HUMULIN I VIAL injection suspension 100 units/ml [LILLY]
14357 HUMULIN I CARTRIDGE injection suspension 100 units/ml [LILLY]
10229 HUMULIN I DISPOSABLE PEN injection suspension 100 units/ml [LILLY]
27461 INSUMAN BASAL CARTRIDGE injection suspension 100 units/ml [AVENTIS]
35468 INSUMAN BASAL VIAL injection suspension 100 units/ml [AVENTIS]
23992 INSUMAN BASAL OPTISET injection suspension 100 units/ml [AVENTIS]
24795 insulin biphasic aspart cartridge injection suspension 30:70; 100 units/ml
23099 insulin biphasic aspart disposable pen injection suspension 30:70; 100 units/ml
7267 NOVOMIX 30 PENFILL injection suspension 100 units/ml [NOVO]
7228 NOVOMIX 30 FLEXPEN injection suspension 100 units/ml [NOVO]
28185 insulin biphasic lispro cartridge injection suspension 25:75; 100 units/ml
31258 insulin biphasic lispro disposable pen injection suspension 25:75; 100 units/ml
10243 HUMALOG MIX 25 CARTRIDGE injection suspension 100 units/ml [LILLY]
14270 HUMALOG MIX 25 DISPOSABLE PEN injection suspension 100 units/ml [LILLY]
14301 insulin detemir cartridge injection solution 100 units/ml
14330 insulin detemir disposable pen injection solution 100 units/ml
10259 insulin glargine vial injection solution 100 units/ml
7400 insulin glargine disposable pen injection solution 100 units/ml
7393 insulin glargine cartridge injection solution 100 units/ml
7402 LANTUS VIAL injection solution 100 units/ml [AVENTIS]
7237 LANTUS OPTISET injection solution 100 units/ml [AVENTIS]
7266 LANTUS CARTRIDGE injection solution 100 units/ml [AVENTIS]
7350 insulin isophane porcine vial injection suspension 100 units/ml
30686 insulin isophane porcine cartridge injection suspension 100 units/ml
28183 HYPURIN PORCINE ISOPHANE VIAL injection suspension 100 units/ml [CP PHARM]
14933 HYPURIN PORCINE ISOPHANE CARTRIDGE injection suspension 100 units/ml [CP PHARM]
27280 insulin biphasic isophane porcine vial injection suspension 30:70; 100 units/ml
36031 insulin biphasic isophane porcine cartridge injection suspension 30:70; 100 units/ml
24800 HYPURIN PORCINE 30/70 MIX VIAL injection suspension 100 units/ml [CP PHARM]
20995 HYPURIN PORCINE 30/70 MIX CARTRIDGE injection suspension 100 units/ml [CP PHARM]
18590 insulin isophane bovine vial injection suspension 100 units/ml
36066 insulin isophane bovine cartridge injection suspension 100 units/ml
14340 HYPURIN BOVINE ISOPHANE VIAL injection suspension 100 units/ml [CP PHARM]
28588 HYPURIN BOVINE ISOPHANE CARTRIDGE injection suspension 100 units/ml [CP PHARM]
13277 MIXTARD 50 PENFILL injection suspension 100 units/ml [NOVO]
14290 INSULATARD PENFILL injection suspension 100 units/ml [NOVO]

14928 INSULATARD VIAL injection suspension 100 units/ml [NOVO]
10208 INSULATARD INNOLET injection suspension 100 units/ml [NOVO]
7300 MIXTARD 30 VIAL injection suspension 100 units/ml [NOVO]
35701 insulin biphasic lispro disposable pen injection suspension 50:50; 100 units/ml
36146 insulin biphasic lispro cartridge injection suspension 50:50; 100 units/ml
18593 HUMALOG MIX 50 CARTRIDGE injection suspension 100 units/ml [LILLY]
35260 LEVEMIR INNOLET injection solution 100 units/ml [NOVO]
36853 LANTUS SOLOSTAR injection solution 100 units/ml [SANOFI/AVE]
39086 HUMALOG MIX 50 KWIKPEN injection suspension 100 units/ml [LILLY]
39006 HUMALOG MIX 25 KWIKPEN injection suspension 100 units/ml [LILLY]
43953 insulin biphasic lispro vial injection suspension 25:75; 100 units/ml
42395 HUMALOG MIX 25 VIAL injection suspension 100 units/ml [LILLY]
43991 HUMULIN M3 KWIKPEN injection suspension 100 units/ml [LILLY]
43950 HUMULIN I KWIKPEN injection suspension 100 units/ml [LILLY]
44480 INSUMAN COMB 25 SOLOSTAR injection suspension 100 units/ml [AVENTIS]
45158 INSUMAN COMB 15 CARTRIDGE injection suspension 100 units/ml [AVENTIS]
1588 ACTRAPID injection 100 iu/ml [NOVO]
1840 HUMULIN S injection 100 units/ml [LILLY]
1842 PORK VELOSULIN injection 100 units/ml [NOVO]
30209 ACTRAPID MC injection 100 units/ml [ARUN]
4706 VELOSULIN VIAL injection solution 100 units/ml [NOVO]
12297 HYPURIN BOVINE NEUTRAL injection 100 units/ml [CP PHARM]
1592 ACTRAPID PENFILL 100 iu/ml [NOVO]
1594 ACTRAPID NOVOLET 100 iu/ml [NOVO]
10572 insulin soluble bovine injection 100 units/ml
24593 neutral insulin bovine injection 100 iu/ml
36513 VELOSULIN CARTRIDGE injection 100 units/ml [NOVO]
17336 NOVOPEN injection device 100 units/ml [NOVO]
41959 PENJECT injection device 100 units/ml [HYPOGUARD]
4129 insulin soluble porcine injection 100 units/ml
24846 PUR-IN NEUTRAL injection 100 units/ml [CP PHARM]
9565 HUMAJECT S DISPOSABLE PEN injection solution 100 units/ml [LILLY]
12638 insulin soluble human pyr injection 100 units/ml
15710 insulin soluble human emp injection 100 units/ml
12654 insulin soluble human prb injection 100 units/ml
13622 HYPURIN PORCINE NEUTRAL injection 100 units/ml [CP PHARM]
22945 INSUMAN RAPID injection 100 iu/ml [AVENTIS]
9521 PORK ACTRAPID VIAL injection solution 100 units/ml [NOVO]
26621 insulin soluble human crb injection 100 iu/ml
27402 insulin soluble human vial injection solution 100 units/ml
36430 insulin soluble human disposable pen injection solution 100 units/ml
16129 insulin soluble human cartridge injection solution 100 units/ml
21235 HUMULIN S VIAL injection solution 100 units/ml [LILLY]
14944 HUMULIN S CARTRIDGE injection solution 100 units/ml [LILLY]
22983 INSUMAN RAPID CARTRIDGE injection solution 100 units/ml [AVENTIS]
23993 INSUMAN RAPID OPTISET injection solution 100 units/ml [AVENTIS]

7349	ACTRAPID VIAL injection solution 100 units/ml [NOVO]
27396	insulin soluble porcine vial injection solution 100 units/ml
25479	insulin soluble porcine cartridge injection solution 100 units/ml
26098	HYPURIN PORCINE NEUTRAL VIAL injection solution 100 units/ml [CP PHARM]
14930	HYPURIN PORCINE NEUTRAL CARTRIDGE injection solution 100 units/ml [CP PHARM]
18592	insulin soluble bovine vial injection solution 100 units/ml
14938	insulin soluble bovine cartridge injection solution 100 units/ml
23231	HYPURIN BOVINE NEUTRAL CARTRIDGE injection solution 100 units/ml [CP PHARM]
14339	HYPURIN BOVINE NEUTRAL VIAL injection solution 100 units/ml [CP PHARM]

Table B 4 Product codes identifying oral antidiabetes medication prescriptions

Product code	Product name	Drug substance
9108	TOLBUTAMIDE 250 MG TAB	
16211	TOLBUTAMIDE 100 MG TAB	
22636	TOLBUTAMIDE 1 GM TAB	
41898	GLIBENCLAMIDE	
1254	glibenclamide tablets 5mg	glibenclamide
2219	glibenclamide tablets 2.5mg	glibenclamide
4862	DIABETAMIDE tablets 2.5mg [ASHBOURNE]	glibenclamide
7744	DAONIL tablets 5mg [AVENTIS]	glibenclamide
7912	SEMI-DAONIL tablets 2.5mg [AVENTIS]	glibenclamide
8976	EUGLUCON tablets 2.5mg [AVENTIS]	glibenclamide
13331	EUGLUCON tablets 5mg [AVENTIS]	glibenclamide
16602	CALABREN tablets 2.5mg [BERK]	glibenclamide
21424	glibenclamide oral suspension 5mg/5ml	glibenclamide
21832	DIABETAMIDE tablets 5mg [ASHBOURNE]	glibenclamide
25636	LIBANIL tablets 2.5mg [APS]	glibenclamide
26218	CALABREN tablets 5mg [BERK]	glibenclamide
28708	MALIX tablets 2.5mg [LAGAP]	glibenclamide
30460	MALIX tablets 5mg [LAGAP]	glibenclamide
31474	LIBANIL tablets 5mg [APS]	glibenclamide
34507	GLIBENCLAMIDE tablets 2.5mg [CP PHARM]	glibenclamide
34563	GLIBENCLAMIDE tablets 5mg [CP PHARM]	glibenclamide
34676	GLIBENCLAMIDE tablets 2.5mg [HILLCROSS]	glibenclamide
34706	GLIBENCLAMIDE tablets 2.5mg [IVAX]	glibenclamide
41558	GLIBENCLAMIDE tablets 5mg [TEVA]	glibenclamide
41559	GLIBENCLAMIDE tablets 5mg [HILLCROSS]	glibenclamide
41593	GLIBENCLAMIDE tablets 2.5mg [TEVA]	glibenclamide
32	gliclazide tablets 80mg	gliclazide
1964	DIAMICRON tablets 80mg [SERVIER]	gliclazide
5627	gliclazide modified release tablet 30mg	gliclazide
11695	DIAMICRON MR tablets 30mg [SERVIER]	gliclazide
15374	gliclazide oral suspension 40mg/5ml	gliclazide
17343	GLICLAZIDE tablets 80mg [HILLCROSS]	gliclazide
21564	GLICLAZIDE tablets 80mg [WOCKHARDT]	gliclazide
21892	DIAGLYK tablets 80mg [ASHBOURNE]	gliclazide
29939	GLICLAZIDE tablets 80mg [GEN (UK)]	gliclazide
31212	GLICLAZIDE tablets 80mg [ACTAVIS]	gliclazide
33562	DUCLAZIDE tablets 80mg [DUMEX]	gliclazide
34399	GLICLAZIDE tablets 80mg [IVAX]	gliclazide
34932	GLICLAZIDE tablets 80mg [GENUS]	gliclazide
36856	GLICLAZIDE tablets 80mg [SANDOZ]	gliclazide
40425	NAZDOL MR tablets 30mg [TEVA]	gliclazide
42790	GLICLAZIDE tablets 80mg [MERCK-GEN]	gliclazide

43065	gliclazide tablets 40mg	gliclazide
43465	ZICRON tablets 40mg [BRISTOL LB]	gliclazide
44473	EDICIL MR tablets 30mg [RATIOPHARM]	gliclazide
45215	GLICLAZIDE tablets 80mg [NEOLAB]	gliclazide
5276	glimepiride tablets 1mg	glimepiride
5316	glimepiride tablets 4mg	glimepiride
5353	glimepiride tablets 2mg	glimepiride
6337	glimepiride tablets 3mg	glimepiride
7284	AMARYL tablets 2mg [AVENTIS]	glimepiride
7332	AMARYL tablets 1mg [AVENTIS]	glimepiride
7409	AMARYL tablets 3mg [AVENTIS]	glimepiride
11284	AMARYL tablets 4mg [AVENTIS]	glimepiride
40365	GLIMEPIRIDE tablets 1mg [ACTAVIS]	glimepiride
44738	NIDDARYL tablets 1mg [DEE]	glimepiride
547	glipizide tablets 2.5mg	glipizide
5636	glipizide tablets 5mg	glipizide
12513	GLIBENESE tablets 5mg [PFIZER]	glipizide
17698	MINODIAB tablets 5mg [PHARMACIA]	glipizide
17706	MINODIAB tablets 2.5mg [PHARMACIA]	glipizide
29326	GLIPIZIDE tablets 5mg [GEN (UK)]	glipizide
34802	GLIPIZIDE tablets 5mg [IVAX]	glipizide
1965	tolbutamide tablets 500mg	tolbutamide
11946	tolbutamide injection 50mg/ml	tolbutamide
12455	RASTINON tablets 500mg [HOECHSTMAR]	tolbutamide
33673	TOLBUTAMIDE tablets 500mg [ACTAVIS]	tolbutamide
34957	TOLBUTAMIDE tablets 500mg [HILLCROSS]	tolbutamide
44304	GLYCONON tablets 500mg [DDSA]	tolbutamide
35144	BYETTA injection 5 micrograms [LILLY]	exenatide
35149	exenatide injection 10micrograms	exenatide
35150	BYETTA injection 10micrograms [LILLY]	exenatide
35251	exenatide injection 5 micrograms	exenatide
40642	VICTOZA injection 18mg/3ml [NOVO]	liraglutide
40693	liraglutide injection 18mg/3ml	liraglutide
5678	nateglinide tablets 120mg	nateglinide
5989	nateglinide tablets 180mg	nateglinide
11483	nateglinide tablets 60mg	nateglinide
15955	STARLIX tablets 120mg [NOVARTIS]	nateglinide
23945	STARLIX tablets 60mg [NOVARTIS]	nateglinide
27125	STARLIX tablets 180mg [NOVARTIS]	nateglinide
548	pioglitazone tablets 15mg	pioglitazone hydrochloride
9699	pioglitazone tablets 30mg	pioglitazone hydrochloride
10051	pioglitazone tablets 45mg	pioglitazone hydrochloride
19472	ACTOS tablets 45mg [TAKEDA]	pioglitazone hydrochloride

20287	ACTOS tablets 15mg [TAKEDA]	pioglitazone hydrochloride
20889	ACTOS tablets 30mg [TAKEDA]	pioglitazone hydrochloride
9707	repaglinide tablets 1mg	repaglinide
9748	repaglinide tablets 2mg	repaglinide
9865	repaglinide tablets 500 micrograms	repaglinide
11316	NOVONORM tablets 500 micrograms [NOVO]	repaglinide
11321	NOVONORM tablets 1mg [NOVO]	repaglinide
11366	NOVONORM tablets 2mg [NOVO]	repaglinide
35561	PRANDIN tablets 2mg [NOVO]	repaglinide
36774	PRANDIN tablets 1mg [NOVO]	repaglinide
36948	PRANDIN tablets 500 micrograms [NOVO]	repaglinide
41204	saxagliptin tablets 5mg	saxagliptin hydrochloride
41431	ONGLYZA tablets 5mg [BMS]	saxagliptin hydrochloride
35022	sitagliptin tablets 100mg	sitagliptin phosphate monohydrate
35462	JANUVIA tablets 100mg [M S D]	sitagliptin phosphate monohydrate
37875	vildagliptin tablets 50mg	vildagliptin
39149	GALVUS tablets 50mg [NOVARTIS]	vildagliptin
43619	sitagliptin with metformin tablets 50mg + 1000mg	metformin hydrochloride/sitaglipti n phosphate monohydrate
43684	JANUMET tablets 50mg + 1000mg [M S D]	metformin hydrochloride/sitaglipti n phosphate monohydrate
37874	vildagliptin with metformin tablets 50mg + 850mg	metformin hydrochloride/vildaglipt in
37902	vildagliptin with metformin tablets 50mg + 1000mg	metformin hydrochloride/vildaglipt in
38551	EUCREAS tablets 50mg + 1000mg [NOVARTIS]	metformin hydrochloride/vildaglipt in
39203	EUCREAS tablets 50mg + 850mg [NOVARTIS]	metformin hydrochloride/vildaglipt in
18220	pioglitazone with metformin tablets 15mg + 850mg	metformin/pioglitazone
30316	metformin with pioglitazone tablets 850mg + 15mg	metformin/pioglitazone
31077	COMPETACT film coated tablets [TAKEDA]	metformin/pioglitazone
2928	METFORMIN HCl 850 MG TAB	
3252	METFORMIN HCl 500 MG TAB	
7815	METFORMIN 800 MG TAB	

16213	METFORMIN 250 MG TAB	
20810	METFORMIN	
23	metformin tablets 500mg	metformin hydrochloride
93	metformin tablets 850mg	metformin hydrochloride
735	metformin oral suspension 100mg/ml	metformin hydrochloride
7048	metformin modified release tablet 500mg	metformin hydrochloride
7166	GLUCOPHAGE tablets 500mg [MERCK SER]	metformin hydrochloride
7610	GLUCOPHAGE tablets 850mg [MERCK SER]	metformin hydrochloride
11990	metformin oral solution 500mg/5ml	metformin hydrochloride
16044	GLUCOPHAGE SR tablets 500mg [MERCK SER]	metformin hydrochloride
25678	GLUCAMET tablets 500mg [OPUS]	metformin hydrochloride
26258	GLUCAMET tablets 850mg [OPUS]	metformin hydrochloride
27501	ORABET tablets 500mg [LAGAP]	metformin hydrochloride
31146	METSOL oral solution 500mg/5ml [ORBIS]	metformin hydrochloride
33087	METFORMIN tablets 500mg [ACTAVIS]	metformin hydrochloride
33674	METFORMIN tablets 850mg [HILLCROSS]	metformin hydrochloride
34004	METFORMIN tablets 500mg [IVAX]	metformin hydrochloride
34020	METFORMIN tablets 850mg [IVAX]	metformin hydrochloride
34135	METFORMIN tablets 500mg [M&A PHARM]	metformin hydrochloride
34323	METFORMIN tablets 500mg [HILLCROSS]	metformin hydrochloride
34504	METFORMIN tablets 500mg [WOCKHARDT]	metformin hydrochloride
34598	METFORMIN tablets 500mg [GEN (UK)]	metformin hydrochloride
34697	METFORMIN tablets 850mg [WOCKHARDT]	metformin hydrochloride
34742	METFORMIN tablets 850mg [TEVA]	metformin hydrochloride
34836	METFORMIN tablets 850mg [ACTAVIS]	metformin hydrochloride
34917	METFORMIN tablets 500mg [TEVA]	metformin hydrochloride

38355	metformin modified release tablet 750mg	metformin hydrochloride
38400	GLUCOPHAGE SR tablets 750mg [MERCK SER]	metformin hydrochloride
39560	BOLAMYN SR tablets 500mg [TEVA]	metformin hydrochloride
39598	metformin modified release tablet 1000mg	metformin hydrochloride
39729	GLUCOPHAGE SR tablets 1000mg [MERCK SER]	metformin hydrochloride
39988	metformin oral powder 500mg	metformin hydrochloride
40007	GLUCOPHAGE sachets 1000mg [MERCK SER]	metformin hydrochloride
40110	GLUCOPHAGE sachets 500mg [MERCK SER]	metformin hydrochloride
40233	metformin oral powder 1000mg	metformin hydrochloride
42161	ORABET tablets 500mg [SANDOZ]	metformin hydrochloride
43270	METFORMIN sugar free oral solution 500mg/5ml [ROSEMONT]	metformin hydrochloride
44250	METFORMIN oral solution 500mg/5ml [HILLCROSS]	metformin hydrochloride
45581	METABET SR tablets 500mg [MORNINGSID]	metformin hydrochloride
22858	acetohexamide tablets 500mg	acetohexamide
26118	DIMELOR tablets 500mg [LILLY]	acetohexamide
1253	chlorpropamide tablets 100mg	chlorpropamide
1847	chlorpropamide tablets 250mg	chlorpropamide
8034	DIABINESE tablets 100mg [PFIZER]	chlorpropamide
8168	DIABINESE tablets 250mg [PFIZER]	chlorpropamide
27969	GLYMESE tablets 250mg [DDSA]	chlorpropamide
12245	GLUTRIL tablets 25mg [ROCHE]	glibornuride
12259	glibornuride tablets 25mg	glibornuride
8390	gliquidone tablets 30mg	gliquidone
19658	GLURENORM tablets 30mg [SANOFI S]	gliquidone
10427	tolazamide tablets 250mg	tolazamide
19336	tolazamide tablets 100mg	tolazamide
21489	TOLANASE tablets 250mg [PHARMACIA]	tolazamide
22145	TOLANASE tablets 100mg [PHARMACIA]	tolazamide
13628	ROMOZIN tablets 400mg [GLAXO]	troglitazone
24848	glymidine sodium tablets 500mg	glymidine sodium
469	rosiglitazone tablets 4mg	rosiglitazone maleate
5227	rosiglitazone tablets 8mg	rosiglitazone maleate
9662	AVANDIA tablets 4mg [GLAXSK PHA]	rosiglitazone maleate
15232	AVANDIA tablets 8mg [GLAXSK PHA]	rosiglitazone maleate
37617	rosiglitazone tablets 2mg	rosiglitazone maleate

6855	AVANDAMET tablets 2mg + 500mg [GLAXSK PHA]	metformin hydrochloride/rosiglitazone maleate
7325	AVANDAMET tablets 4mg + 1000mg [GLAXSK PHA]	metformin hydrochloride/rosiglitazone maleate
7375	rosiglitazone with metformin tablets 4mg + 1000mg	metformin hydrochloride/rosiglitazone maleate
11601	rosiglitazone with metformin tablets 2mg + 500mg	metformin hydrochloride/rosiglitazone maleate
11604	rosiglitazone with metformin tablets 1mg + 500mg	metformin hydrochloride/rosiglitazone maleate
11609	metformin with rosiglitazone tablets 500mg + 1mg	metformin hydrochloride/rosiglitazone maleate
11610	metformin with rosiglitazone tablets 500mg + 2mg	metformin hydrochloride/rosiglitazone maleate
11717	rosiglitazone with metformin tablets 2mg + 1000mg	metformin hydrochloride/rosiglitazone maleate
11737	metformin with rosiglitazone tablets 1000mg + 4mg	metformin hydrochloride/rosiglitazone maleate
11760	metformin with rosiglitazone tablets 1000mg + 2mg	metformin hydrochloride/rosiglitazone maleate
14164	AVANDAMET tablets 2mg + 1000mg [GLAXSK PHA]	metformin hydrochloride/rosiglitazone maleate
17580	AVANDAMET tablets 1mg + 500mg [GLAXSK PHA]	metformin hydrochloride/rosiglitazone maleate

Appendix C: Codelists for identifying diagnoses of infection

This Appendix contains the Read codes and ICD-10 codes used to identify diagnoses of acute, community-acquired lower respiratory tract infections (and the subset of these which identified a pneumonia diagnosis), sepsis, and urinary tract infections.

Table C 1 Read codes identifying a diagnosis of lower respiratory tract infection but not pneumonia

Read code	Read term
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
65VA.00	Notification of whooping cough
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
A39y000	Pulmonary nocardiosis
A3BXA00	Mycoplasma pneumoniae [PPLO] cause/dis classifd/oth chaptr
A3By100	Eaton's agent infection
A3By400	Pleuropneumonia-like organism (PPLO) infection
AB40600	Acute pulmonary histoplasmosis capsulati
AB42.00	Pulmonary histoplasmosis
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
H042.11	Laryngotracheitis
H042000	Acute laryngotracheitis without obstruction
H042100	Acute laryngotracheitis with obstruction
H042z00	Acute laryngotracheitis NOS

H04z.00	Acute laryngitis and tracheitis NOS
H052.00	Pharyngotracheitis
H053.00	Tracheopharyngitis
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
H060v00	Subacute bronchitis unspecified
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
H24..11	Chest infection with infectious disease EC
H27..00	Influenza
H271.00	Influenza with other respiratory manifestation
H271000	Influenza with laryngitis
H271100	Influenza with pharyngitis
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
H27z.11	Flu like illness
H27z.12	Influenza like illness

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
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
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
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
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

Table C 2 Read codes identifying a diagnosis of lower respiratory tract infection and a diagnosis of pneumonia

Read code H262.00 *'Postoperative pneumonia'* identified a post-operative pneumonia. Episodes of pneumonia which included this code were classified as hospital-acquired infections.

Read code	Read term
A022200	Salmonella pneumonia
A203.00	Primary pneumonic plague
A205.00	Pneumonic plague, unspecified
A521.00	Varicella pneumonitis
A54x400	Herpes simplex pneumonia
A551.00	Postmeasles pneumonia
A730.00	Ornithosis with pneumonia
A785000	Cytomegaloviral pneumonitis
A789300	HIV disease resulting in Pneumocystis carinii pneumonia
AB24.11	Pneumonia - candidal
AB40500	Histoplasma capsulatum with pneumonia
AB41500	Histoplasma duboisii with pneumonia
AD04.00	Toxoplasma pneumonitis
AD63.00	Pneumocystosis
H2...00	Pneumonia and influenza
H20..00	Viral pneumonia
H20..11	Chest infection - viral pneumonia
H200.00	Pneumonia due to adenovirus
H201.00	Pneumonia due to respiratory syncytial virus
H202.00	Pneumonia due to parainfluenza virus
H20y.00	Viral pneumonia NEC
H20y000	Severe acute respiratory syndrome
H20z.00	Viral pneumonia NOS
H21..00	Lobar (pneumococcal) pneumonia
H21..11	Chest infection - pneumococcal pneumonia
H22..00	Other bacterial pneumonia
H22..11	Chest infection - other bacterial pneumonia
H220.00	Pneumonia due to klebsiella pneumoniae
H221.00	Pneumonia due to pseudomonas
H222.00	Pneumonia due to haemophilus influenzae
H222.11	Pneumonia due to haemophilus influenzae
H223.00	Pneumonia due to streptococcus
H223000	Pneumonia due to streptococcus, group B
H224.00	Pneumonia due to staphylococcus
H22y.00	Pneumonia due to other specified bacteria
H22y000	Pneumonia due to escherichia coli
H22y011	E.coli pneumonia
H22y100	Pneumonia due to proteus
H22y200	Pneumonia - Legionella
H22yX00	Pneumonia due to other aerobic gram-negative bacteria
H22yz00	Pneumonia due to bacteria NOS
H22z.00	Bacterial pneumonia NOS
H23..00	Pneumonia due to other specified organisms

H23..11	Chest infection - pneumonia organism OS
H230.00	Pneumonia due to Eaton's agent
H231.00	Pneumonia due to mycoplasma pneumoniae
H232.00	Pneumonia due to pleuropneumonia like organisms
H233.00	Chlamydial pneumonia
H23z.00	Pneumonia due to specified organism NOS
H24..00	Pneumonia with infectious diseases EC
H240.00	Pneumonia with measles
H241.00	Pneumonia with cytomegalic inclusion disease
H242.00	Pneumonia with ornithosis
H243.00	Pneumonia with whooping cough
H243.11	Pneumonia with pertussis
H246.00	Pneumonia with aspergillosis
H247000	Pneumonia with candidiasis
H247z00	Pneumonia with systemic mycosis NOS
H24y.00	Pneumonia with other infectious diseases EC
H24y000	Pneumonia with actinomycosis
H24y100	Pneumonia with nocardiasis
H24y200	Pneumonia with pneumocystis carinii
H24y300	Pneumonia with Q-fever
H24y400	Pneumonia with salmonellosis
H24y500	Pneumonia with toxoplasmosis
H24y600	Pneumonia with typhoid fever
H24y700	Pneumonia with varicella
H24yz00	Pneumonia with other infectious diseases EC NOS
H24z.00	Pneumonia with infectious diseases EC NOS
H25..00	Bronchopneumonia due to unspecified organism
H25..11	Chest infection - unspecified bronchopneumonia
H26..00	Pneumonia due to unspecified organism
H26..11	Chest infection - pnemonia due to unspecified organism
H260.00	Lobar pneumonia due to unspecified organism
H260000	Lung consolidation
H261.00	Basal pneumonia due to unspecified organism
H270.00	Influenza with pneumonia
H270.11	Chest infection - influenza with pneumonia
H270000	Influenza with bronchopneumonia
H270100	Influenza with pneumonia, influenza virus identified
H270z00	Influenza with pneumonia NOS
H28..00	Atypical pneumonia
H530200	Gangrenous pneumonia
H530300	Abscess of lung with pneumonia
H540000	Hypostatic pneumonia
H540100	Hypostatic bronchopneumonia
H564.00	Bronchiolitis obliterans organising pneumonia
Hyu0800	[X]Other viral pneumonia
Hyu0A00	[X]Other bacterial pneumonia
Hyu0B00	[X]Pneumonia due to other specified infectious organisms
Hyu0D00	[X]Pneumonia in viral diseases classified elsewhere
Hyu0H00	[X]Other pneumonia, organism unspecified

Table C 3 ICD-10 codes identifying a diagnosis of lower respiratory tract infection but not pneumonia

ICD-10 code	Diagnostic name
A37	Whooping cough
A37.0	Whooping cough due to Bordetella pertussis
A37.1	Whooping cough due to Bordetella parapertussis
A37.8	Whooping cough due to other Bordetella species
A37.9	Whooping cough, unspecified
B96.0	Mycoplasma pneumoniae as cause dis class oth chaps
J04.1	Acute tracheitis
J04.2	Acute laryngotracheitis
J09	Influenza due to other identified influenza virus
J10	Influenza due to identified influenza virus
J10.1	Influenza with oth resp manifest influenza virus identified
J10.8	Influenza with other manifest influenza virus identified
J11	Influenza, virus not identified
J11.1	Influenza with oth resp manifestation virus not identified
J11.8	Influenza with other manifestations, virus not identified
J20	Acute bronchitis
J20.0	Acute bronchitis due to Mycoplasma pneumoniae
J20.1	Acute bronchitis due to Haemophilus influenzae
J20.2	Acute bronchitis due to streptococcus
J20.3	Acute bronchitis due to coxsackievirus
J20.4	Acute bronchitis due to parainfluenza virus
J20.5	Acute bronchitis due to respiratory syncytial virus
J20.6	Acute bronchitis due to rhinovirus
J20.7	Acute bronchitis due to echovirus
J20.8	Acute bronchitis due to other specified organisms
J20.9	Acute bronchitis, unspecified
J21	Acute bronchiolitis
J21.0	Acute bronchiolitis due to respiratory syncytial virus
J21.8	Acute bronchiolitis due to other specified organisms
J21.9	Acute bronchiolitis, unspecified
J22	Unspecified acute lower respiratory infection
J44.0	Chronic obstructive pulmonary disease with acute lower respiratory infection
J86	Pyothorax
J86.0	Pyothorax with fistula
J86.9	Pyothorax without fistula

Table C 4 ICD-10 codes identifying a diagnosis of lower respiratory tract infection and pneumonia

ICD-10 code	Diagnostic name
B01.2	Varicella pneumonia
B05.2	Measles complicated by pneumonia
B20.6	HIV disease resulting in Pneumocystis carinii pneumonia
B25.0	Cytomegaloviral pneumonitis
J10.0	Influenza with pneumonia, influenza virus identified
J11.0	Influenza with pneumonia, virus not identified
J12	Viral pneumonia, not elsewhere classified
J12.0	Adenoviral pneumonia
J12.1	Respiratory syncytial virus pneumonia
J12.2	Parainfluenza virus pneumonia
J12.8	Other viral pneumonia
J12.9	Viral pneumonia, unspecified
J13	Pneumonia due to Streptococcus pneumoniae
J14	Pneumonia due to Haemophilus influenzae
J15	Bacterial pneumonia, not elsewhere classified
J15.0	Pneumonia due to Klebsiella pneumoniae
J15.1	Pneumonia due to Pseudomonas
J15.2	Pneumonia due to staphylococcus
J15.3	Pneumonia due to streptococcus, group B
J15.4	Pneumonia due to other streptococci
J15.5	Pneumonia due to Escherichia coli
J15.6	Pneumonia due to other aerobic Gram-negative bacteria
J15.7	Pneumonia due to Mycoplasma pneumoniae
J15.8	Other bacterial pneumonia
J15.9	Bacterial pneumonia, unspecified
J16	Pneumonia due to other infectious organisms NEC
J16.0	Chlamydial pneumonia
J16.8	Pneumonia due to other specified infectious organisms
J17	Pneumonia in diseases classified elsewhere
J17.0	Pneumonia in bacterial diseases classified elsewhere
J17.1	Pneumonia in viral diseases classified elsewhere
J17.2	Pneumonia in mycoses
J17.3	Pneumonia in parasitic diseases
J17.8	Pneumonia in other diseases classified elsewhere
J18	Pneumonia, organism unspecified
J18.0	Bronchopneumonia, unspecified
J18.1	Lobar pneumonia, unspecified
J18.2	Hypostatic pneumonia, unspecified
J18.8	Other pneumonia, organism unspecified
J18.9	Pneumonia, unspecified
J85.1	Abscess of lung with pneumonia
U04	Severe acute respiratory syndrome [SARS]
U04.9	Severe acute respiratory syndrome, unspecified

Table C 5 Read codes identifying a diagnosis of sepsis

Read codes SP20000 '*Postoperative endotoxic shock'*, SP20100 *'Postoperative septic shock'*, SP25400 *'Postoperative septicaemia'* and SP38000 *'Septic shock due to transfusion'* identified sepsis with an iatrogenic source, and episodes of sepsis which contained any instances of these codes would have been classified as hospital-acquired infections.

Read code	Read term
A271100	Erysipelothrix septicaemia
A384100	Haemophilus influenzae septicaemia
A384400	Serratia septicaemia
A3Ay100	Toxic shock syndrome
A545.00	Herpes simplex septicaemia
AB2y300	Candidal septicaemia
R106.00	[D]Unspecified bacteraemia
R107.00	[D]Unspecified viraemia
A021.00	Salmonella septicaemia
A202.00	Septicaemic plague
A270100	Listeria septicaemia
A362.00	Meningococcal septicaemia
A365.00	Meningococcal meningitis with acute meningococcal septicaem
A366.00	Meningococcal meningitis with meningococcal septicaemia
A38..00	Septicaemia
A380.00	Streptococcal septicaemia
A380000	Septicaemia due to streptococcus group A
A380100	Septicaemia due to streptococcus group B
A380300	Septicaemia due to streptococcus pneumoniae
A380400	Septicaemia due to enterococcus
A380500	Vancomycin resistant enterococcal septicaemia
A381.00	Staphylococcal septicaemia
A381000	Septicaemia due to Staphylococcus aureus
A381100	Septicaemia due to coagulase-negative staphylococcus
A382.00	Pneumococcal septicaemia
A383.00	Septicaemia due to anaerobes
A384.00	Septicaemia due to other gram negative organisms
A384000	Gram negative septicaemia NOS
A384200	Escherichia coli septicaemia
A384211	E.coli septicaemia
A384300	Pseudomonas septicaemia
A384z00	Other gram negative septicaemia NOS
A38y.00	Other specified septicaemias
A38z.00	Septicaemia NOS
A38z.11	Sepsis
A98yz12	Gonococcal septicaemia
Ayu3E00	[X]Other streptococcal septicaemia
Ayu3F00	[X]Streptococcal septicaemia unspecified

Ayu3G00	[X]Septicaemia due to other gram-negative organisms
Ayu3H00	[X]Other specified septicaemia
Ayu3J00	[X]Septicaemia unspecified
H5y0100	Tracheostomy sepsis
J666.00	Biliary sepsis
R055200	[D]Endotoxic shock
R055300	[D]Gram-negative shock
R055500	[D]Septic shock
R055511	[D]Septicaemic shock
A362000	Acute meningococcaemia
A98yz11	Gonococcaemia NOS
AB25.00	Disseminated, systemic candida
Ayu3B00	[X]Meningococcaemia, unspecified

Table C 6 ICD-10 codes used to identify diagnoses of sepsis

In addition to these codes, any instances of A49.9 *'Bacterial infection, unspecified'* or B24.9 *'Viral infection, unspecified'* with hospital admission date on the same day as a Read code for a diagnosis of sepsis in CPRD would have been taken as confirming the primary care diagnosis of sepsis on that day.

ICD-10 code	ICD-10 diagnostic name
A02.1	Salmonella septicaemia
A20.7	Septicaemic plague
A22.7	Anthrax septicaemia
A24.1	Acute and fulminating melioidosis
A26.7	Erysipelothrix septicaemia
A32.7	Listerial septicaemia
A39.1	Waterhouse-Friderichsen syndrome
A39.2	Acute meningococcaemia
A39.4	Meningococcaemia, unspecified
A40.0	Septicaemia due to streptococcus, group A
A40.1	Septicaemia due to streptococcus, group B
A40.2	Septicaemia due to streptococcus, group D
A40.3	Septicaemia due to Streptococcus pneumoniae
A40.8	Other streptococcal septicaemia
A40.9	Streptococcal septicaemia, unspecified
A41.0	Septicaemia due to Staphylococcus aureus
A41.1	Septicaemia due to other specified staphylococcus
A41.2	Septicaemia due to unspecified staphylococcus
A41.3	Septicaemia due to Haemophilus influenzae
A41.4	Septicaemia due to anaerobes
A41.5	Septicaemia due to other Gram-negative organisms
A41.8	Other specified septicaemia
A41.9	Septicaemia, unspecified
A42.7	Actinomycotic septicaemia
A48.3	Toxic shock syndrome
B00.7	Disseminated herpesviral disease
B37.7	Candidal septicaemia
R57.2	Septic shock
R57.8	Other shock
R65.0	Systemic Inflammatory response syndrome of infectious origin without organ failure
R65.1	Systemic Inflammatory response syndrome of infectious origin with organ failure

Table C 7 Read codes used to identify a diagnosis of urinary tract infection

Read code K190200 '*Post operative urinary tract infection'* was used to identify episodes containing a postoperative urinary tract infection. Any episodes containing this code were classified as a hospital-acquired infection.

Read codes 1J4..00 *'Suspected UTI'* was used to link other UTI Read codes within a single episode, but did not on its own define a UTI.

Read code	Read term
1AG..00	Recurrent urinary tract infections
K190.00	Urinary tract infection, site not specified
K190.11	Recurrent urinary tract infection
K190300	Recurrent urinary tract infection
K190311	Recurrent UTI
K190500	Urinary tract infection
K190z00	Urinary tract infection, site not specified NOS
A32y300	Diphtheritic cystitis
A981100	Acute gonococcal cystitis
A981111	Bladder gonorrhoea - acute
K15..00	Cystitis
K150.00	Acute cystitis
K152000	Subacute cystitis
K154500	Cystitis in gonorrhoea
K154600	Cystitis in moniliasis
K154700	Cystitis in trichomoniasis
K155.00	Recurrent cystitis
K15y200	Abscess of bladder
K15z.00	Cystitis NOS
K213.00	Prostatocystitis
K101.00	Acute pyelonephritis
K101000	Acute pyelonephritis without medullary necrosis
K101200	Acute pyelitis
K101300	Acute pyonephrosis
K101z00	Acute pyelonephritis NOS
K102.00	Renal and perinephric abscess
K102000	Renal abscess
K102100	Perinephric abscess
K102200	Renal carbuncle
K102z00	Renal and perinephric abscess NOS
K104.00	Xanthogranulomatous pyelonephritis
K10y.00	Pyelonephritis and pyonephrosis unspecified
K10y000	Pyelonephritis unspecified
K10y100	Pyelitis unspecified
K10y200	Pyonephrosis unspecified
K10y300	Pyelonephritis in diseases EC
K10y400	Pyelitis in diseases EC
K10yz00	Unspecified pyelonephritis NOS

Table C 8 Read codes for urethritis used to exclude proteinuria records from defining CKD status

The codes in **Table C 7**, including Read code K190200 '*Post operative urinary tract infection'* and 1J4..00 *'Suspected UTI',* together with the following codes for urethritis, were all used to exclude proteinuria records from defining CKD status.

Read code	Read term
A980100	Acute gonococcal urethritis
A994.00	Nonspecific urethritis
AD10200	Trichomonal urethritis
K17..00	Urethritis due to non venereal causes
K17..11	Periurethritis
K170.00	Urethral and periurethral abscess
K170.11	Urethral abscess
K170000	Urethral abscess unspecified
K170111	Cowper's gland abscess
K170200	Urethral gland abscess
K170300	Periurethral cellulitis
K170311	Periurethritis
K170400	Periurethral abscess
K170z00	Urethral abscess NOS
K172.00	Candidal urethritis
K17y.00	Other urethritis
K17y000	Urethritis unspecified
K17y100	Urethral syndrome NOS
K17y200	Skene's glands adenitis
K17y300	Cowperitis
K17y600	Verumontanitis
K17z.00	Urethritis due to non venereal cause NOS

Table C 9 ICD-10 codes used to identify a diagnosis of urinary tract infection

ICD-10 Code	Name *WHO guidance for use (selected)*
N10	Acute tubulo-interstitial nephritis *Incl.: Acute:* • *infectious interstitial nephritis* • *pyelitis* • *pyelonephritis* *Use additional code (B95-B97), if desired, to identify infectious agent.*
N12	Tubulo-interstitial nephritis not spec as acute or chronic *Incl.:Interstitial nephritis NOS* *Pyelitis NOS* *Pyelonephritis NOS* *Excl.: calculous pyelonephritis (N20.9)*
N136	Pyonephrosis *Conditions in N13.0-N13.5 with infection* *Obstructive uropathy with infection*
N151	Renal and perinephric abscess
N159	Renal tubulo-interstitial disease, unspecified *Infection of kidney NOS* *Excl.: urinary tract infection NOS (N39.0)*
N300	Acute cystitis *Use additional code, if desired, to identify infectious agent (B95-B97) or responsible external agent (Chapter XX).* *Excl.: prostatocystitis (N41.3)*
N308	Other cystitis *Abscess of bladder*
N309	Cystitis, unspecified
N390	Urinary tract infection, site not specified

Appendix D: Codelists for identifying chronic kidney disease

This appendix includes Read and ICD-10 codes identifying renal replacement therapy (kidney transplant or dialysis), Read codes used to identify proteinuria status, and Read codes identifying staged CKD status.

Table D 1 Read codes identifying renal replacement therapy

Read code	Read term
14S2.00	H/O: kidney recipient
7B00.00	Transplantation of kidney
7B00100	Transplantation of kidney from live donor
7B00111	Allotransplantation of kidney from live donor
7B00200	Transplantation of kidney from cadaver
7B00211	Allotransplantation of kidney from cadaver
7B00300	Allotransplantation of kidney from cadaver heart-beating
7B00400	Allotransplantation kidney from cadaver heart non-beating
7B00y00	Other specified transplantation of kidney
7B00z00	Transplantation of kidney NOS
7B01511	Excision of rejected transplanted kidney
7B06300	Exploration of renal transplant
7B0F.00	Interventions associated with transplantation of kidney
7B0F100	Pre-transplantation of kidney work-up recipient
7B0F200	Pre-transplantation of kidney work-up live donor
7B0F400	Post-transplantation of kidney examination live donor
8L50.00	Renal transplant planned
K0B5.00	Renal tubulo-interstitial disordrs in transplant rejectn
Kyu1C00	[X]Renal tubulo-interstitial disorders/transplant rejection
SP08011	Det.ren.func.after ren.transpl
SP08300	Kidney transplant failure and rejection
TB00100	Kidney transplant with complication without blame
TB00111	Renal transplant with complication without blame
ZV42000	[V]Kidney transplanted
14V2.00	H/O: renal dialysis
14V2.11	H/O: kidney dialysis
7A60600	Creation of graft fistula for dialysis
7L1A.00	Compensation for renal failure
7L1A.11	Dialysis for renal failure
7L1A000	Renal dialysis
7L1A011	Thomas intravascular shunt for dialysis
7L1A100	Peritoneal dialysis
7L1A200	Haemodialysis NEC
7L1A300	Haemofiltration
7L1A400	Automated peritoneal dialysis
7L1A500	Continuous ambulatory peritoneal dialysis
7L1A600	Peritoneal dialysis NEC
7L1Ay00	Other specified compensation for renal failure
7L1Az00	Compensation for renal failure NOS
7L1B.00	Placement ambulatory apparatus compensation renal failure
7L1B.11	Placement ambulatory dialysis apparatus - compens renal fail

7L1B000	Insertion of ambulatory peritoneal dialysis catheter
7L1B100	Removal of ambulatory peritoneal dialysis catheter
7L1By00	Placement ambulatory apparatus- compensate renal failure OS
7L1C.00	Placement other apparatus for compensation for renal failure
7L1C000	Insertion of temporary peritoneal dialysis catheter
7L1Cz00	Placement other apparatus- compensate for renal failure NOS
TA02000	Accid cut puncture perf h'ge - kidney dialysis
TA02z00	Accid cut puncture perf h'ge - perfusion NOS
TA22.00	Failure of sterile precautions during perfusion
TA22000	Failure of sterile precautions during kidney dialysis
TB11.00	Kidney dialysis with complication without blame
TB11.11	Renal dialysis with complication without blame
U612200	[X]Failure sterile precautions dur kidney dialys/other perf
ZV45100	[V]Renal dialysis status
ZV56.00	[V]Aftercare involving intermittent dialysis
ZV56011	[V]Aftercare involving renal dialysis NOS
ZVu3G00	[X]Other dialysis

Table D 2 ICD-10 codes identifying renal replacement therapy

ICD-10 code	ICD-10 diagnostic name
T86.1	Kidney transplant failure and rejection
Z94.0	Kidney transplant status
Y84.1	Kidney dialysis
Z49.0	Preparatory care for dialysis
Z49.1	Extracorporeal dialysis
Z49.2	Other dialysis
Z99.2	Dependence on renal dialysis

Table D 3 Read codes identifying a positive test for proteinuria (defined proteinuria if no concurrent urinary tract infection)

Read code	Quality Outcomes Framework*	Read term
4674		Urine protein test = +
4675		Urine protein test = ++
4676		Urine protein test = +++
4677		Urine protein test = ++++
4678		Proteinuria
46W0.00		Urine microalbumin positive
R110.00	Yes	[D]Proteinuria
R110000	Yes	[D]Albuminuria
R110200		[D]Exercise proteinuria
R110300	Yes	[D]Microalbuminuria
R110z00	Yes	[D]Proteinuria NOS

* The Quality Outcomes Framework financially incentivises recording of these codes

Table D 4 Read codes identifying persistent proteinuria or proteinuric disease (defined proteinuria irrespective of urinary tract infections)

Read code	Quality Outcomes Framework*	Read term
1Z17.00		Chronic kidney disease stage 1 with proteinuria
1Z17.11		CKD stage 1 with proteinuria
1Z19.00		Chronic kidney disease stage 2 with proteinuria
1Z19.11		CKD stage 2 with proteinuria
1Z1B.00	Yes	Chronic kidney disease stage 3 with proteinuria
1Z1B.11	Yes	CKD stage 3 with proteinuria
1Z1D.00	Yes	Chronic kidney disease stage 3A with proteinuria
1Z1D.11	Yes	CKD stage 3A with proteinuria
1Z1F.00	Yes	Chronic kidney disease stage 3B with proteinuria
1Z1F.11	Yes	CKD stage 3B with proteinuria
1Z1H.00	Yes	Chronic kidney disease stage 4 with proteinuria
1Z1H.11	Yes	CKD stage 4 with proteinuria
1Z1K.00	Yes	Chronic kidney disease stage 5 with proteinuria
1Z1K.11	Yes	CKD stage 5 with proteinuria
A844100		Plasmodium malariae malaria with nephropathy
C10EK00		Type 1 diabetes mellitus with persistent proteinuria
C10EL00		Type 1 diabetes mellitus with persistent microalbuminuria
C10EL11		Type I diabetes mellitus with persistent microalbuminuria
C10FL00		Type 2 diabetes mellitus with persistent proteinuria
C10FL11		Type II diabetes mellitus with persistent proteinuria
C10FM00		Type 2 diabetes mellitus with persistent microalbuminuria
C10FM11		Type II diabetes mellitus with persistent microalbuminuria
C373600		Nephropathic amyloidosis
D011100		Vit B12 defic anaemia due to malabsorption with proteinuria
K00..00		Acute glomerulonephritis
K00..12		Bright's disease
K000.00		Acute proliferative glomerulonephritis
K001.00		Acute nephritis with lesions of necrotising glomerulitis
K00y.00		Other acute glomerulonephritis
K00y000		Acute glomerulonephritis in diseases EC
K00yz00		Other acute glomerulonephritis NOS
K00z.00		Acute glomerulonephritis NOS
K01..00		Nephrotic syndrome
K010.00		Nephrotic syndrome with proliferative glomerulonephritis
K011.00		Nephrotic syndrome with membranous glomerulonephritis
K012.00		Nephrotic syndrome+membranoproliferative glomerulonephritis
K013.00		Nephrotic syndrome with minimal change glomerulonephritis
K013.11		Lipoid nephrosis
K013.12		Steroid sensitive nephrotic syndrome
K014.00		Nephrotic syndrome, minor glomerular abnormality
K015.00		Nephrotic syndrome, focal and segmental glomerular lesions
K016.00		Nephrotic syndrome, diffuse membranous glomerulonephritis
K017.00		Nephrotic syn difus mesangial prolifertiv glomerulonephritis

K018.00	Nephrotic syn,difus endocapilary proliftv glomerulonephritis
K019.00	Nephrotic syn,diffuse mesangiocapillary glomerulonephritis
K01A.00	Nephrotic syndrome, dense deposit disease
K01B.00	Nephrotic syndrome, diffuse crescentic glomerulonephritis
K01x000	Nephrotic syndrome in amyloidosis
K01x100	Nephrotic syndrome in diabetes mellitus
K01x200	Nephrotic syndrome in malaria
K01x300	Nephrotic syndrome in polyarteritis nodosa
K01x400	Nephrotic syndrome in systemic lupus erythematosus
K01x411	Lupus nephritis
K01y.00	Nephrotic syndrome with other pathological kidney lesions
K01z.00	Nephrotic syndrome NOS
K020.00	Chronic proliferative glomerulonephritis
K021.00	Chronic membranous glomerulonephritis
K022.00	Chronic membranoproliferative glomerulonephritis
K023.00	Chronic rapidly progressive glomerulonephritis
K02y.00	Other chronic glomerulonephritis
K02y000	Chronic glomerulonephritis + diseases EC
K02y200	Chronic focal glomerulonephritis
K02y300	Chronic diffuse glomerulonephritis
K02yz00	Other chronic glomerulonephritis NOS
K02z.00	Chronic glomerulonephritis NOS
K030.00	Proliferative nephritis unspecified
K031.00	Membranous nephritis unspecified
K032.00	Membranoproliferative nephritis unspecified
K032000	Focal membranoproliferative glomerulonephritis
K032300	Anaphylactoid glomerulonephritis
K032400	Familial glomerulonephritis in Alport's syndrome
K032500	Other familial glomerulonephritis
K032600	Berger's IgA or IgG nephropathy
K032y00	Nephritis unsp+OS membranoprolif glomerulonephritis lesion
K032y11	Hypocomplementaemic persistent glomerulonephritis NEC
K032y13	Mesangioproliferative glomerulonephritis NEC
K032y14	Mesangiocapillary glomerulonephritis NEC
K032y15	Mixed membranous and proliferative glomerulonephritis NEC
K032z00	Nephritis unsp+membranoprolif glomerulonephritis lesion NOS
K033.00	Rapidly progressive nephritis unspecified
K03T.00	Tubulo-interstit nephritis, not specif as acute or chron
K03U.00	Unspecif nephr synd, diff concentric glomerulonephritis
K03V.00	Unspecified nephritic syndrome, dense deposit disease
K03W.00	Unsp nephrit synd, diff endocap prolif glomerulonephritis
K03X.00	Unsp nephrit synd, diff mesang prolif glomerulonephritis
K03y200	Other interstitial nephritis
K03z.00	Unspecified glomerulonephritis NOS
K072.00	Glomerulosclerosis
K08y500	Acute interstitial nephritis
K0A..00	Glomerular disease
K0A0.00	Acute nephritic syndrome
K0A0000	Acute nephritic syndrome minor glomerular abnormality

K0A0100	Acute nephritic syndrome, focal+segmental glomerular lesions
K0A0200	Acute nephritic syn, diffuse membranous glomerulonephritis
K0A0300	Acut neph syn, diffuse mesangial prolifrative glomnephritis
K0A0400	Ac neph syn difus endocaplry prolifrative glomerulonephritis
K0A0500	Acute neph syn, diffuse mesangiocapillary glomerulonephritis
K0A0600	Acute nephritic syndrome dense deposit disease
K0A0700	Acute nephrotic syndrm diffuse crescentic glomerulonephritis
K0A1.00	Rapidly progressive nephritic syndrome
K0A1100	Rapid progres nephritic syn focal+segmental glomerulr lesion
K0A1200	Rapid progres neph syn diffuse membranous glomerulonephritis
K0A1300	Rpd prog neph syn df mesangial prolifratv glomerulonephritis
K0A1600	Rapid progressive nephritic syndrome, dense deposit disease
K0A1700	Rapid progres nephritic syn df crescentic glomerulonephritis
K0A2100	Recur+persist haematuria, focal+segmental glomerular lesions
K0A2200	Recur+persist haematuria difus membranous glomerulonephritis
K0A2300	Recur+persist haemuria df mesangial prolif glomerulnephritis
K0A2500	Recur+persist hmuria df mesangiocapilary glomerulonephritis
K0A2700	Recur+persist haematuria difus crescentic glomerulonephritis
K0A3.00	Chronic nephritic syndrome
K0A3000	Chronic nephritic syndrome, minor glomerular abnormality
K0A3100	Chronic nephritic syndrm focal+segmental glomerular lesions
K0A3200	Chron nephritic syndrom difuse membranous glomerulonephritis
K0A3300	Chron neph syn difus mesangial prolifrtiv glomerulonephritis
K0A3500	Chronic neph syn difus mesangiocapillary glomerulonephritis
K0A3600	Chronic nephritic syndrome, dense deposit disease
K0A3700	Chronic nephritic syn diffuse crescentic glomerulonephritis
K0A4.00	Isolated proteinuria with specified morphological lesion
K0A4100	Isolatd proteinur/specifd morphlgcl les foc+seg glom lesn
K0A4200	Isolatd proteinur/specfd morphlgcl les df membrn glomneph
K0A4300	Isoltd prteinur/spcfd morph lesn df mesngl prolf glomneph
K0A4500	Isoltd prteinur+specfd morph les df mesangiocap glomnephr
K0A4W00	Isolated proteinuria with unspecified morpholog changes
K0A4X00	Isolated proteinuria with oth specif morpholog changes
K0A5000	Hereditary nephropathy NEC, minor glomerular abnormality
K0A5100	Hereditary nephropathy NEC,focal+segmnt glomerular lesion
K0A5200	Hereditry nephropathy NEC,difus membran glomerulnephritis
K0A5300	Heredtry nephrpthy NEC difus mesangial prolif glomnephrit
K0A5600	Hereditary nephropathy, NEC, dense deposit disease
K136.00	Benign postural proteinuria
K136.11	Orthostatic proteinuria
K190X00	Persistent proteinuria, unspecified
Kyu0900	[X]Unsp nephrit synd, diff mesang prolif glomerulonephritis
Kyu5G00	[X]Persistent proteinuria, unspecified

* The Quality Outcomes Framework financially incentivises recording of these codes

Table D 5 Read codes identifying chronic kidney disease by stage

Read code	Quality Outcomes Framework*	Read term
1Z10.00		Chronic kidney disease stage 1
1Z18.00		Chronic kidney disease stage 1 without proteinuria
1Z17.00		Chronic kidney disease stage 1 with proteinuria
1Z17.11		CKD stage 1 with proteinuria
1Z11.00		Chronic kidney disease stage 2
1Z1A.00		Chronic kidney disease stage 2 without proteinuria
1Z19.00		Chronic kidney disease stage 2 with proteinuria
1Z19.11		CKD stage 2 with proteinuria
1Z1A.11		CKD stage 2 without proteinuria
1Z1F.11	Yes	CKD stage 3B with proteinuria
1Z12.00	Yes	Chronic kidney disease stage 3
1Z1C.00	Yes	Chronic kidney disease stage 3 without proteinuria
1Z1E.00	Yes	Chronic kidney disease stage 3A without proteinuria
1Z15.00	Yes	Chronic kidney disease stage 3A
1Z16.00	Yes	Chronic kidney disease stage 3B
1Z1G.00	Yes	Chronic kidney disease stage 3B without proteinuria
1Z1B.00	Yes	Chronic kidney disease stage 3 with proteinuria
1Z1D.00	Yes	Chronic kidney disease stage 3A with proteinuria
1Z1F.00	Yes	Chronic kidney disease stage 3B with proteinuria
1Z1C.11	Yes	CKD stage 3 without proteinuria
1Z1B.11	Yes	CKD stage 3 with proteinuria
1Z1E.11	Yes	CKD stage 3A without proteinuria
1Z1D.11	Yes	CKD stage 3A with proteinuria
1Z1G.11	Yes	CKD stage 3B without proteinuria
1Z13.00	Yes	Chronic kidney disease stage 4
1Z1J.00	Yes	Chronic kidney disease stage 4 without proteinuria
1Z1H.00	Yes	Chronic kidney disease stage 4 with proteinuria
1Z1H.11	Yes	CKD stage 4 with proteinuria
1Z1J.11	Yes	CKD stage 4 without proteinuria
1Z14.00	Yes	Chronic kidney disease stage 5
1Z1K.00	Yes	Chronic kidney disease stage 5 with proteinuria
1Z1L.00	Yes	Chronic kidney disease stage 5 without proteinuria
1Z1K.11	Yes	CKD stage 5 with proteinuria
1Z1L.11	Yes	CKD stage 5 without proteinuria
K05..12		End stage renal failure
K050.00		End stage renal failure
K0D..00		End-stage renal disease

* The Quality Outcomes Framework financially incentivises recording of these codes

Appendix E. Supplementary analyses to Paper 3, examining the association of CKD with incidence of community-acquired infections

This appendix presents sensitivity analyses of the association of CKD with incidence of community-acquired infection as discussed in **Chapter 8**.

Tables E1 and E2 present an exploration of whether there was interaction between (1) and and eGFR or (2) proteinuria and eGFR in the models of LRTI and pneumonia incidence.

Table E 1 Association of estimated glomerular filtration rate (eGFR) with infection incidence among different categories of age and proteinuria status

eGFR	Age (years)		History of proteinuria	
	<75	≥75	No	Yes
	IRR (95% CI)	IRR (95% CI)	IRR (95% CI)	IRR (95% CI)
LRTI				
<15	1.54 (1.31–1.81)	1.51 (1.34–1.70)	1.81 (1.55–2.12)	1.37 (1.21–1.55)
15-29	1.08 (1.01–1.17)	1.27 (1.22–1.33)	1.27 (1.21–1.33)	1.17 (1.11–1.24)
30-44	1.11 (1.07–1.15)	1.12 (1.09–1.15)	1.12 (1.09–1.16)	1.09 (1.05–1.13)
45-59	1.06 (1.03–1.08)	1.03 (1.01–1.06)	1.05 (1.03–1.07)	1.03 (1.00–1.06)
≥60	1 (reference)	1 (reference)	1 (reference)	1 (reference)
P (interaction)*	<0.001		0.02	
Pneumonia				
<15	2.92 (1.79–4.76)	3.62 (2.79–4.69)	5.13 (3.61–7.29)	2.72 (2.02–3.67)
15-29	2.22 (1.78–2.78)	2.06 (1.85–2.29)	2.27 (1.99–2.58)	1.92 (1.66–2.21)
30-44	1.54 (1.33–1.78)	1.37 (1.27–1.48)	1.41 (1.30–1.54)	1.39 (1.24–1.55)
45-59	1.05 (0.94–1.17)	1.02 (0.95–1.10)	1.02 (0.95–1.10)	1.06 (0.95–1.17)
≥60	1 (reference)	1 (reference)	1 (reference)	1 (reference)
P (interaction)*	0.56		0.03	

eGFR, estimated glomerular filtration rate (ml/min/1.73m^2); IRR incidence rate ratio; 95% CI 95% confidence interval

* P (interaction) is a likelihood ratio test with the null hypothesis that there is no interaction between the two relevant variables

Table E 2 Effect of age and proteinuria with infection incidence among difference categories of estimated glomerular filtration rate (eGFR)

eGFR	Age (years)		History of proteinuria	
	<75	≥75	No	Yes
	IRR (95% CI)	IRR (95% CI)	IRR (95% CI)	IRR (95% CI)
LRTI				
<15	1 (reference)	1.07 (0.88–1.30)	1 (reference)	0.83 (0.68–1.00)
15-29	1 (reference)	1.28 (1.18–1.38)	1 (reference)	1.01 (0.95–1.08)
30-44	1 (reference)	1.10 (1.05–1.15)	1 (reference)	1.06 (1.02–1.10)
45-59	1 (reference)	1.07 (1.04–1.10)	1 (reference)	1.07 (1.04–1.11)
≥60	1 (reference)	1.09 (1.07–1.11)	1 (reference)	1.09 (1.07–1.12)
Pneumonia				
<15	1 (reference)	2.74 (1.59–4.72)	1 (reference)	0.68 (1.43–1.06)
15-29	1 (reference)	2.05 (1.62–2.59)	1 (reference)	1.08 (0.91–1.28)
30-44	1 (reference)	1.96 (1.70–2.28)	1 (reference)	1.26 (1.12–1.41)
45-59	1 (reference)	2.16 (1.93–2.41)	1 (reference)	1.32 (1.20–1.46)
≥60	1 (reference)	2.21 (2.06–2.38)	1 (reference)	1.28 1.18–1.38)

eGFR, estimated glomerular filtration rate (ml/min/1.73m^2); IRR incidence rate ratio; 95% CI 95% confidence interval

Table E 3 Sensitivity analysis of the association between markers of chronic kidney disease and infection incidence limited to patients with HES linkage

	Infections n	Person years	Minimally adjusted rate ratio (95% CI)[1]	Fully adjusted rate ratio (95% CI) [2]	P value [3]
Lower respiratory tract infection (LRTI) n=113,085					
eGFR					
<15	342	1,387	1.84 (1.61–2.10)	1.45 (1.28–1.66)	<0.0001
15-29	3,155	14,594	1.45 (1.38–1.53)	1.20 (1.14–1.26)	
30-44	10,036	56,525	1.22 (1.18–1.26)	1.09 (1.06–1.12)	
45-59	18,166	123,818	1.06 (1.03–1.08)	1.12 (0.99–1.04)	
60+	38,162	288,786	1 (reference)	1 (reference)	
Proteinuria					
Yes	19,053	118,201	1.13 (1.11–1.16)	1.08 (1.05–1.11)	<0.0001
No	50,808	366,908	1 (reference)	1 (reference)	
Total	69,861	485,109			
Pneumonia n=113,105					
eGFR					
<15	75	1,404	5.26 (3.98–6.95)	3.61 (2.75–4.75)	<0.0001
15-29	474	14,784	2.46 (2.18–2.78)	1.84 (1.63–2.07)	
30-44	1,085	57,164	1.45 (1.33–1.57)	1.21 (1.11–1.32)	
45-59	1,406	125,061	1.00 (0.93–1.07)	0.94 (0.87–1.01)	
60+	2,593	291,457	1 (reference)	1 (reference)	
Proteinuria					
Yes	1,890	119,452	1.39 (1.30–1.49)	1.28 (1.20–1.37)	<0.0001
No	3,743	370,418	1 (reference)	1 (reference)	
Total	5,633	489,871			
Sepsis n=113,106					
eGFR					
<15	19	1,406	6.55 (3.91–10.98)	4.82 (2.88–8.08)	<0.0001
15-29	117	14,796	3.71 (2.95–4.68)	2.78 (2.20–3.53)	
30-44	229	57,204	1.82 (1.52–2.16)	1.52 (1.27–1.82)	
45-59	287	125,116	1.51 (0.99–1.34)	1.07 (0.92–1.25)	
60+	499	291,568	1 (reference)	1 (reference)	
Proteinuria					
Yes	380	119,518	1.29 (1.13–1.48)	1.20 (1.04–1.37)	0.01
No	771	370,572	1 (reference)	1 (reference)	
Total	1,151	490,090			

HES, Hospital Episode Statistics; 95% CI, 95% confidence interval; eGFR, estimated glomerular filtration rate (ml/min/1.73m^2)

1. Adjusted for age, sex, practice IMD, and calendar year pre/post 2004

2. Adjusted for age, sex, practice IMD, calendar year pre/post 2004, comorbidities (updated), smoking (baseline), and characteristics of diabetes (baseline)

3. Likelihood ratio test

Table E 4 Sensitivity analysis of the association between markers of chronic kidney disease and infection incidence limited to first occurrence of each infection only

	Infections n	Person years	Minimally adjusted rate ratio (95% CI)[1]	Fully adjusted rate ratio (95% CI) [2]	P value [3]
Lower respiratory tract infection (LRTI) n=191,672					
eGFR					
<15	269	1,752	2.06 (1.79–2.38)	1.64 (1.42–1.88)	<0.0001
15-29	2,185	17,473	1.56 (1.48–1.65)	1.26 (1.19–1.32)	
30-44	7,311	71,741	1.23 (1.20–1.27)	1.08 (1.05–1.11)	
45-59	14,335	160,949	1.08 (1.05–1.10)	1.02 (1.00–1.04)	
60+	31,976	390,543	1 (reference)	1 (reference)	
Proteinuria					
Yes	13,649	149,771	1.15 (1.12–1.18)	1.09 (1.07–1.12)	<0.0001
No	42,427	492,687	1 (reference)	1 (reference)	
Total	56,076	642,457			
Pneumonia n= 191,706					
eGFR					
<15	90	2,488	3.68 (2.94–4.51)	3.17 (2.50–4.01)	<0.0001
15-29	586	24,690	2.15 (1.95–2.38)	1.77 (1.59–1.96)	
30-44	1,365	95,568	1.29 (1.29–1.48)	1.20 (1.12–1.29)	
45-59	1,795	204,579	1.01 (0.95–1.07)	0.95 (0.90–1.01)	
60+	3,259	479,906	1 (reference)	1 (reference)	
Proteinuria					
Yes	2,365	201,310	1.32 (1.25–1.40)	1.29 (1.21–1.36)	<0.0001
No	4,730	605,921	1 (reference)	1 (reference)	
Total	7,095	807,231			
Sepsis n= 191,708					
eGFR					
<15	39	2,535	7.58 (5.20–11.04)	6.27 (4.25–9.27)	<0.0001
15-29	180	25,195	3.30 (2.71–4.02)	2.67 (2.19–3.26)	
30-44	372	96,908	1.80 (1.57–2.07)	1.54 (1.34–1.77)	
45-59	78	206,365	1.19 (1.06–1.34)	1.11 (0.99–1.26)	
60+	833	483,174	1 (reference)	1 (reference)	
Proteinuria					
Yes	681	203,904	1.45 (1.30–1.62)	1.38 (1.24–1.55)	<0.0001
No	1,221	610,272	1 (reference)	1 (reference)	
Total	1,902	814,176			

95% CI, 95% confidence interval; eGFR, estimated glomerular filtration rate (ml/min/1.73m^2)

1. Adjusted for age, sex, practice IMD, and calendar year pre/post 2004

2. Adjusted for age, sex, practice IMD, calendar year pre/post 2004, comorbidities (updated), smoking (baseline), and characteristics of diabetes (baseline)

3. Likelihood ratio test

Table E 5 Sensitivity analysis of the association between markers of chronic kidney disease and infection incidence limited to time from 1 April 2004 onwards

	Infections n	Person years	Minimally adjusted rate ratio (95% CI)[1]	Fully adjusted rate ratio (95% CI) [2]	P[3]
Lower respiratory tract infection (LRTI) n=173,152					
eGFR					
<15	461	1,935	1.71 (1.52–1.92)	1.44 (1.29–1.61)	<0.0001
15-29	3,952	19,375	1.35 (1.52–1.92)	1.17 (1.12–1.22)	
30-44	12,345	73,044	1.16 (1.13–1.42)	1.05 (1.03–1.08)	
45-59	21,785	151,508	1.05 (1.03–1.07)	1.02 (1.00–1.04)	
60+	49,265	373,352	1 (reference)	1 (reference)	
Proteinuria					
Yes	26,562	171,338	1.12 (1.09–1.14)	1.06 (1.0–1.09)	<0.0001
No	61,246	447,877	1 (reference)	1 (reference)	
Total	87,808	619,215			
Pneumonia n=173,181					
eGFR					
<15	74	1,965	3.82 (2.01–.03)	2.83 (2.16–3.70)	<0.0001
15-29	487	19,642	2.07 (1.84–2.32)	1.63 (1.45–1.83)	
30-44	1,156	73,897	1.33 (1.22–1.44)	1.13 (1.05–1.23)	
45-59	1,469	153,084	0.95 (0.89–1.02)	0.91 (0.85–0.97)	
60+	2,943	376,967	1 (reference)	1 (reference)	
Proteinuria					
Yes	2,259	173,216	1.40 (1.32–1.49)	1.29 (1.21–1.37)	<0.0001
No	3,870	452,339	1 (reference)	1 (reference)	
Total	6,129	625,555			
Sepsis n=173,185					
eGFR					
<15	34	1,966	7.87 (5.30–11.69)	5.99 (4.03–8.90)	<0.0001
15-29	143	19,657	3.06 (2.50–3.75)	2.38 (1.94–2.93)	
30-44	303	73,939	1.75 (1.51–2.04)	1.49 (1.28–1.74)	
45-59	380	153,143	1.16 (1.02–1.33)	1.10 (0.96–1.25)	
60+	702	377,089	1 (reference)	1 (reference)	
Proteinuria					
Yes	606	173,294	1.43 (1.28–1.60)	1. 32 (1.18–1.48)	<0.0001
No	956	452,499	1 (reference)	1 (reference)	
Total	1,562	625,793			

95% CI, 95% confidence interval; eGFR, estimated glomerular filtration rate (ml/min/1.73m^2)

1. Adjusted for age, sex, practice IMD, and calendar year pre/post 2004

2. Adjusted for age, sex, practice IMD, calendar year pre/post 2004, comorbidities (updated), smoking (baseline), and characteristics of diabetes (baseline)

3. Likelihood ratio test