Predicting Cardiovascular Disease Risk in Chronic Kidney Disease

Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

Rupert William Leslie Major, MBChB MRes

College of Life Sciences

Department of Health Sciences

University of Leicester

September 2019

Abstract

This thesis is titled 'Predicting Cardiovascular Disease Risk in Chronic Kidney Disease' by Dr Rupert Major.

Chronic kidney disease (CKD) is a long term condition in which glomerular filtration is reduced and/or proteinuria occurs. Cardiovascular risk factors in CKD are different to the general population and overall risk is higher too. Therefore, risk prediction tools for cardiovascular disease require specific validation in CKD.

The Leicester City and County CKD (LCC) cohort of 17,248 anonymised individuals with CKD from 44 general practices was established. Cardiovascular events were identified from general practice and hospital records, and 2,072 cardiovascular events occurred during five years of follow-up.

Risk factors for cardiovascular events in CKD were identified in a systematic review and a second systematic review updated a previous systematic review of risk prediction tools for cardiovascular events in CKD. Albumin, haemoglobin and phosphate were identified as risk factors to be consider for risk prediction tools in addition to factors included in general population risk prediction tools. Seven CKD-specific and six general population risk prediction models were identified. All models were developed using the Cox proportional hazards ('Cox') model.

The LCC cohort was used to externally validate these models. Discrimination was worse and calibration suggested overprediction of risk in all models. The latter worsened as predicted risk increased. Some calibration improvement was achieved through Cox model baseline risk function re-estimation. There was no significant risk prediction improvement by including the variables identified in the systematic review. Sensitivity analysis suggested that the Cox model's censoring at random assumption may have been violated in the risk prediction models due to the competing risk from death.

Risk prediction models for cardiovascular events in people with CKD require improvement and updating to optimise risk prediction accuracy. Alternative methods, such as multi-state models, should be considered in future model development.

Acknowledgements

There are many people to thank, in so many different ways, who made this thesis possible. Firstly, Kidney Research UK has been a wonderful charity to be funded by over the last three years. So thanks must go to all the people who have ran, baked, cycled, shaken buckets, walked, bequeathed and all the other multitude of activities that raise money day in and day out for the charity. I hope I have produced something to repay your generosity of time and money. The people who work for Kidney Research UK have also been thoroughly supportive and shown a genuine interest in what I am up to, so thanks must also go to Sandra Currie *et al*. I would also like to thank all of the GP practices in the three Leicester CCGs who, despite busy clinical workloads, agreed to participate in the study.

My two supervisors, Professors Laura Gray and Nigel Brunskill have provided support and encouragement, not just for this thesis, but for nearly the last decade. Their different backgrounds and perspectives, have given me a very well rounded supervisee experience. I look forward to the next decade!

As you will read in the coming pages, a large part of this thesis focuses on the process of anonymised data collection, linking and analysis. Dr David Shepherd has been a local and national pioneer in primary care informatics, long before such terms ill-defined terms as 'big data' were conceived, and this would quite simple not have been possible without him. I hope he stays involved in this work even after he hangs up his stethoscope in the next few months.

There are many individuals who have assisted with parts of the thesis and I have thanked them in the appropriate chapters. Their contributions have considerably improved the methodological soundness, for particularly the systematic reviews. The individuals who have taught on the University of Leicester's Medical Statistics MSc course have helped enormously with their clear, concise explanation of many statistical concepts and have helped me overcome my 'equationophobia'. The two prognosis courses, hosted by Keele University, have helped me immeasurably in consolidating many of the key concepts presented in this thesis and I thank the faculty for their support. Importantly, it has allowed me to meet many likeminded clinicians who also want to see an improvement in clinical applicability of prognostic research.

My family has been endlessly supportive of my bid to be an 'eternal student', and I hope the following pages make you feel proud. My parents have always encouraged me to follow my dreams, even if it is not the most direct route and there are dead ends along the way. To Sarah and our atypical family, thanks for your support in what has been the most fulfilling three years of my personal and professional life. I promise never to use my 10% discount at the cinema again, at least not when you're with me!

Finally, my mother's descriptions of some of the Nobel Laureates she worked with in her formative Cambridge years has ensured that I remain humble and respectful of those who have come before us. I therefore dedicate this thesis to Professor Doug Altman, who although I never met in person, and only heard speak once, is one of the many giants' shoulders upon which this thesis stands.

'To maximise the benefit to society, you need to not just do research,

but do it well.'

Professor Doug Altman

(1948-2018)

Table of Contents

Table List

Figure List

List of Appendices

Chapter 1 – none

Chapter 2

2.1 - Research Ethics Committee Approval Letter

2.2 - Practice Information Leaflet

2.3 – Study Consent Form

Chapter 3 – none

Chapter 4

4.1 – Definitions of Co-morbidities

4.2 – CV Event Codes

Chapter 5

5.1 – OVID Medline Literature Search Strategy

5.2 – Embase Literature Search Strategy

5.3 - Risk of Bias Summary

5.4 – Potential Risk Factors from Systematic Review

Chapter 6

6.1 – Search Strategy from Systematic Review

Chapter 7 – none

Chapter 8 – none

Chapter 9 – none

Table of Abbreviations

ACR – albumin creatinine ratio

AHA – American Heart Association

CI – confidence interval

CKD – chronic kidney disease

CV - cardiovascular

eGFR – estimated glomerular filtration rate

ESRD – endstage renal disease

E:O – expected:observed ratio

EPI – Chronic Kidney Disease Epidemiology Collaboration

FMI – fraction of missing information

GFR – glomerular filtration rate

GP – general practice

HDL – high density lipoprotein

HES – Hospital Episode Statistics

ICD-10 – International Classification of Diseases, 10th Revision

IHD – ischaemic heart disease

IMPAKT - IMproving Patient care and Awareness of Kidney disease progression Together

IQR – interquartile range

LCC – Leicester City and County Chronic Kidney Disease Cohort

LDL – low density lipoprotein cholesterol

LP – linear predictor

KDIGO – Kidney Disease Improving Global Outcomes

MAR – missing at random

MCAR – missing completing at random

MDRD – Modification of Diet in Renal Disease

MI – multiple imputation

MINAP – Myocardial Infarction National Audit Project

MNAR – missing not at random

NHS – National Health Service

NICE – The National Institute for Health and Care Excellence

OE – observed versus expected event plot

ONS – Office of National Statistics

PCR – protein creatinine ratio

SD – standard deviation

TRIPOD - Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis

UK – United Kingdom

Chapter 1

1. Introduction

This chapter introduces the clinical background, key statistical concepts and rationale for this thesis. Firstly, the condition of chronic kidney disease (CKD) and its basic epidemiology and pathophysiology is described. Cardiovascular (CV) disease is then introduced with a focus on why the pathological processes may be different in people with CKD compared to the general population. The next section provides a general explanation of risk prediction tools and their potential role in clinical practice. The penultimate section describes what was already known about CV risk prediction tools in CKD prior to this thesis. Finally, given this context, the overall aims of the thesis are presented.

1.1   Chronic Kidney Disease

CKD has emerged over the last twenty years as a leading cause of non-communicable premature morbidity and mortality, particularly in high income countries where it is associated with the increasing prevalence of diabetes mellitus and hypertension (1-5). Reduced estimated glomerular filtration rate (eGFR) and increasing amounts of protein in an individual's urine has been shown to have a continuous independent relationship with increased risk of cardiovascular (CV), endstage renal failure (ESRD) and all-cause mortality events (6).

CKD rarely manifests as a symptomatic condition until the advanced stages where renal replacement, through the need for dialysis and transplant, is required (3). Diagnosis of CKD is therefore usually through serum creatinine measurement and/or proteinuria measurement, often for other reasons, typically for the management of related co-morbidities such as diabetes mellitus or hypertension (7).

The variation of serum creatinine by age, gender, ethnicity and body composition, meant that until 1999 routine estimation of glomerular filtration rate (GFR) was difficult to implement on a population-wide scale. In addition, despite being first describe in 1940 (8), direct measures of glomerular filtration rate through the use of inulin or radio-labelled isotopes are time consuming and not practical for regular measurement in the same patient (9). The initial introduction of the Modification of Diet in Renal Disease (MDRD) eGFR formula, which utilised the four variables, age, gender, serum creatinine and ethnicity ('White' compared to 'African American') meant that automated pathology laboratory calculation of eGFR was possible (10). MDRD eGFR was further improved by Chronic Kidney Disease Epidemiology Collaboration (EPI) eGFR (11). Using the same variables, EPI eGFR has now been implemented into routine clinical care in the United Kingdom (UK). Other methods for estimation of GFR are possible, including the measurement of Cystatin-C and other eGFR formulae (12-15), but as they have yet to be implemented into routine primary care of CKD in the UK, will not be considered further in this thesis.

Measurement of proteinuria, mainly through 24 hour urine collection, also suffered from the same limitations and inconvenience to the patient. Again, it was not until spot measurements of proteinuria quantification such as albumin creatinine ratio

(ACR) and proteinuria creatinine ratio (PCR) were validated to replace 24 hour urine collection that quantification of proteinuria on a population scale was possible (16-18).

Given these advances in the measurements of eGFR and proteinuria, standardised diagnostic criteria for CKD were developed, initially in the early 2000s (19), and then refined a decade later (6). The criteria required either or both of the following criteria to be present for at least three months:

- eGFR <60 ml/min/1.73m^2

- a marker of kidney damage

    o albuminuria (ACR ≥ 30mg/mmol)

    o urine sediment abnormality (typically the presence of haematuria)

    o urine or serum electrolyte abnormality related to a kidney tubule disorder

    o kidney structural imaging abnormality

    o histological evidence of kidney pathology

    o presence of a kidney transplant

'Kidney Disease: Improving Global Outcomes' (KDIGO) developed these criteria and in the UK The National Institute for Health and Care Excellence (NICE) implemented them into routine clinical practice (6, 7).

1.1.1 Classification

Following the introduction of routine eGFR measurement, the staging of CKD was suggested by KDIGO and subsequently adopted by NICE. Classification has been refined over the last ten to fifteen years and currently is based on staging of the condition using both eGFR and proteinuria quantification, typically through ACR (6, 19). Any eGFR measurements below <60 ml/min/1.73m^2 equate to stages 3a, 3b, 4 and 5, regardless of any other kidney damage markers set out above. Stages 1 and 2 relate to eGFR ≥60 ml/min/1.73m^2 in the presence of one of these markers. The eGFR stage is then suffixed by one of three ACR stage. Table 1.1 reproduces details of the KDIGO classification system.

Category		Description	
eGFR	**ml/min/1.73m^2**		
1	≥90	Normal or high	
2	60-89	Mildly decreased	
3a	45-59	Mildly to moderately decreased	
3b	30-44	Moderately to severely decreased	
4	15-29	Severely decreased	
5	<15	Kidney failure	
Albuminuria	**24 hour excretion (mg)**	**ACR (mg/mmol)**	**Description**
A1	<30	<3	Normal to mildly increased
A2	30-300	3-30	Moderately increased
A3	>300	>30	Severely increased

Table 1.1: KDIGO CKD Classification – reproduced from KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease (6).

1.1.2 Epidemiology

Based on the above criteria the prevalence of CKD worldwide may be up to 15% of the adult population (3). In Europe, this prevalence varies substantially and may be dependent on measurement criteria and inclusion criteria for CKD in available studies (20).

In the UK in 2012, the age-standardised estimated prevalence of CKD stages 3 to 5 (eGFR <60 ml/min/1.73m^2) had been estimated from a number of sources as shown in Table 1.2. The Quality and Outcomes Framework, based on primary care coding of CKD estimated a prevalence of 4.3% in adults (≥18 years of age), but substantial limitations of accurate diagnosis based on coding were suspected and later confirmed by the CKD National Audit (21). The Health Survey for England (HSE) 2010 report provided a national representative sample and estimated a prevalence in individuals ≥16 years of age of 6% and 7% in males and females respectively (22). The New Opportunities for Early Renal Intervention by Computerised Assessment ('NEOERICA') study suggested an overall adult prevalence of 8.5% but may have been limited by a non-nationally representative sample in relation to ethnicity and socioeconomics (23). The Quality Improvement in CKD ('QICKD') study estimated a *whole* population 5.4% prevalence with approximately twice the prevalence in females than males (24). Importantly, QICKD used the formal definition of eGFR <60 ml/min/1.73m^2 more than three months apart.

Source	Denominator	Male Prevalence	Female Prevalence	Overall Prevalence	Estimated CKD Stage 3 to 5
QOF	18+	-	-	4.3%	1,817,871
HSE	16+	6%	7%	6.5%	2,710,575
QICKD	Whole population	3.5%	7.3%	5.4%	2,817,104
NEOERICA	18+	5.8%	10.6%	8.5%	3,640,321

Table 1.2: Estimated prevalence and population numbers, CKD stages 3-5 based on UK CKD studies (25).

In 2017, the National CKD Audit published its findings (21). It highlighted the limitations in CKD coding for estimation of prevalence of CKD, both in terms of individuals not coded with CKD who had CKD and those incorrectly coded with CKD. Again, the focus was on those with eGFR <60 ml/min/1.73m^2 rather than the second of the two criteria set out above. The National CKD Audit also highlighted that CKD was a disease of older individuals with more than a third of individuals over the age of 83 having CKD based on the KDIGO criteria, compared to <1% of those aged under 50 years as shown in Figure 1.1.
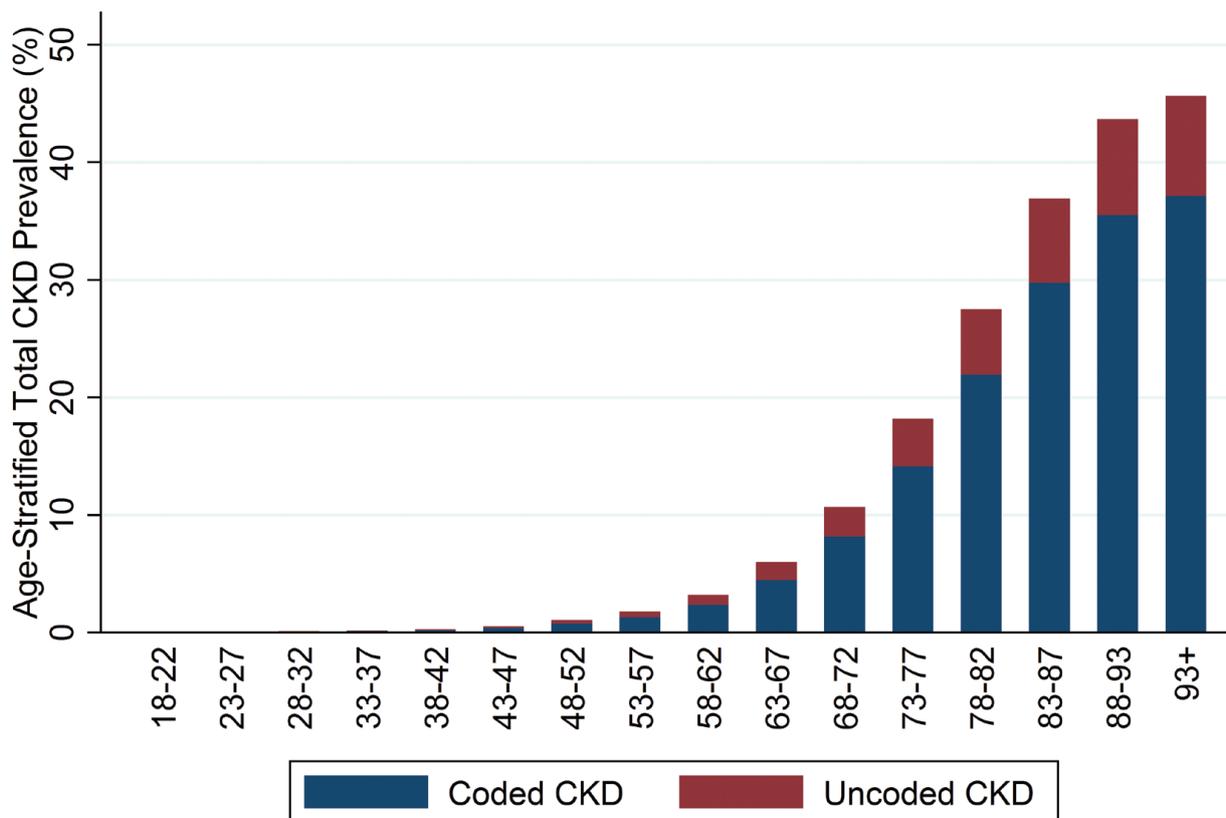
Figure 1.1: Prevalence of CKD by age groups. From Nitsch D, Caplin B, Hull S, Wheeler DC. The National CKD Audit and quality improvement programme in primary care: first national CKD audit report 2017.

Therefore, given this prevalence and association with other co-morbidities, the overall impact of CKD on population health and quality of life in the UK and internationally is high (3). From a financial perspective the cost is also high and expected to rise in line with the ageing population. The most recent estimates for England, no estimates are available for the whole UK, from 2009-10 is that CKD accounts for 1.3% of the National Health Service's (NHS) healthcare budget, of which 13% is related to the excess myocardial infarctions and strokes associated with CKD. For NHS England this equated to approximately £175 million of the overall £1.5 billion spent on CKD in 2009-10.

1.1.3 Pathophysiology

The aetiology of CKD can be broadly split into primary and secondary causes. The latter is most commonly from DM and HTN, particularly in high income countries. DM probably accounts for between a third and half of all CKD (3). The exact contribution of HTN is probably more difficult to assess because of its role as both a cause and consequence of CKD (26). Nevertheless, HTN is probably present in at least three quarters of individuals with CKD (27).

Perhaps the most controversial area of CKD is the role of ageing in the development and progression of CKD (28-30). Given the substantial increase in prevalence of CKD in older individuals as shown in Figure 1.1, there is the suggestion that CKD is 'just' the consequence of 'ageing' kidneys (28-30).

There is a plethora of rare primary causes of CKD, which when combined account for approximately 10-20% of all CKD in the UK (3). Causes do vary globally with glomerulonephritis, infections such as human immunodeficiency virus and hepatitides, and environmental poisoning causing a higher proportion in lower and middle income countries, compared to high income countries such as the UK (3). Despite this, unknown aetiology remains prevalent too and may account for up to a further 20% of CKD cases in the UK (3). Other secondary causes include renovascular disease related to systemic atherosclerosis and 'post-renal' causes. In combination these two causes contribute to a further 10% of cases (3). Post-renal

causes include structural abnormalities such as obstructive uropathy and recurrent infection (3).

1.2   Cardiovascular Disease in Chronic Kidney Disease

According to the World Health Organisation, CV disease is the leading cause of global death, accounting for just under a third of all death (31). Four fifths of these deaths relate to myocardial infarction and strokes (31). Other CV related causes of death and disability include heart failure, cardiac arrhythmias, peripheral vascular disease and thromboembolic disease (31). For the remainder of the thesis CV disease events will refer to acute coronary syndrome and cerebrovascular disease. The former includes 'unstable' angina and myocardial infarction of any type. Cerebrovascular disease includes ischaemic stroke and transient ischaemic attack.

Once CKD reaches stage 3a or above, the increased CV risk is thought to be of a similar magnitude to diabetes mellitus or established coronary heart disease (32). CV disease in CKD is associated with the 'traditional' pathological process of atherosclerosis as well as 'non-traditional' pathology (33, 34). The former leads to greater risk because of the association with other CV risk factors such as DM, HTN and dyslipidaemia. The latter relates to the potential pathophysiology consequences of CKD, sometimes referred to as uraemia-related factors. Their role increases as CKD advances. Whilst uraemia itself may contribute, it is a broad term that refers to many other clinical measurable and unmeasurable factors rather than just raised serum urea. In general terms, these factors are grouped in to arteriosclerosis, cardiomyopathy and inflammation. However, there is much interaction between these traditional and non-traditional factors meaning separation as a singular distinct pathological mechanism is not possible.
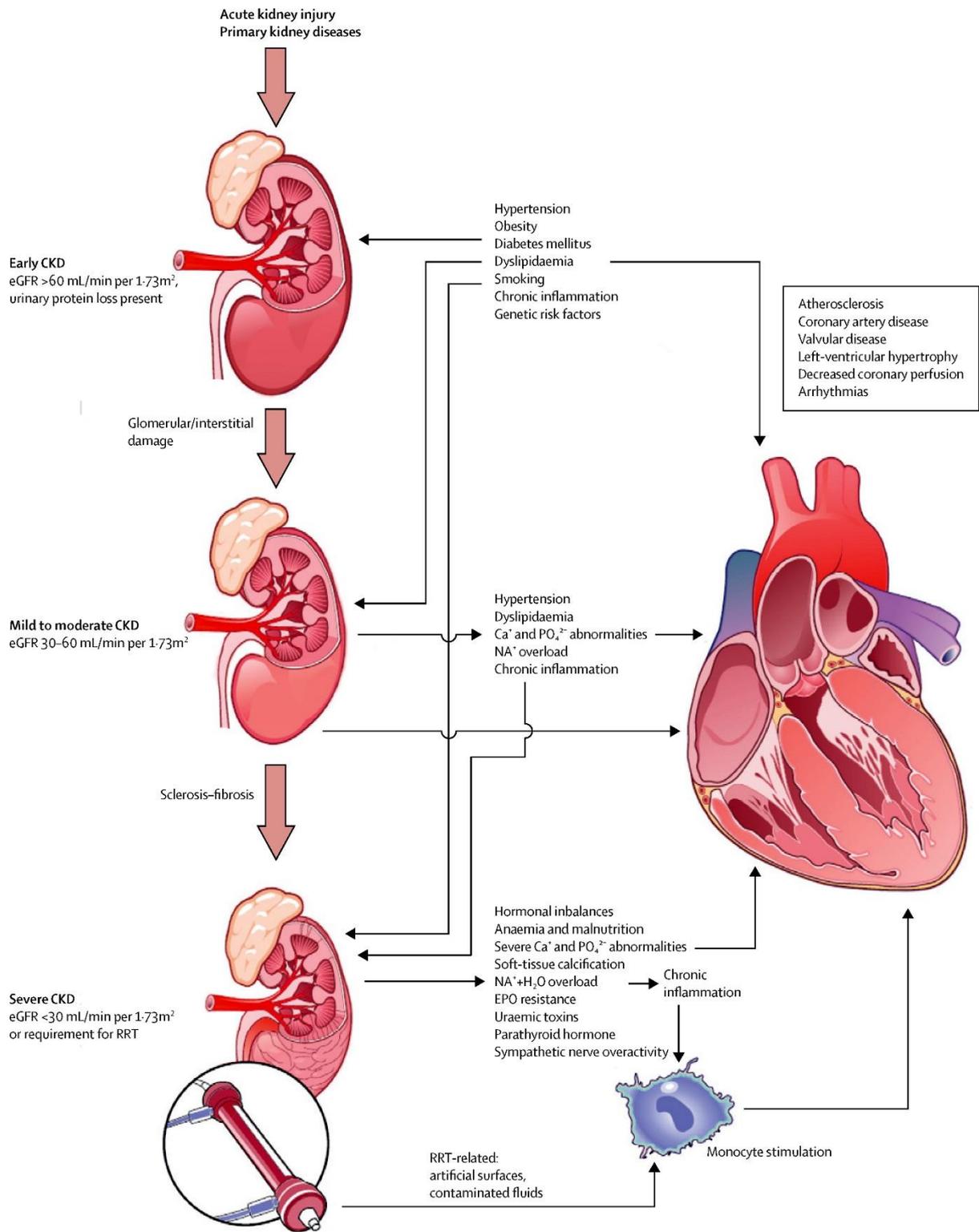
Figure 1.2: Pathophysiological Interactions between Kidney and Heart in Chronic Kidney Disease. reproduced from Gansevoort *et al* 'Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention' (35).

1.2.1 Atherosclerosis

The 'traditional' pathology of CV disease is described by the term atherosclerosis. In broad terms, this describes the narrowing of an artery due to the development of artheromatous 'plaque'. Plaques are multifarious including epithelial cells, connective tissue, foam (fat) cells, lipids and immune cells such as macrophages (36).

As the plaque grows, due to factors including ageing, it can reduce blood flow and hence reduce tissue perfusion (37, 38). This can lead to chronic clinical symptoms such as exertional chest pain in the case of a coronary artery plaque. Acutely, the plaque can rupture causing embolism of part of the plaque and thrombosis around the plaque. This can lead to total occlusion of the vessel and infarction distal to the thrombosis site. Depending on whether the site is coronary artery or cerebral artery this will lead to either a myocardial infarction or stroke respectively.
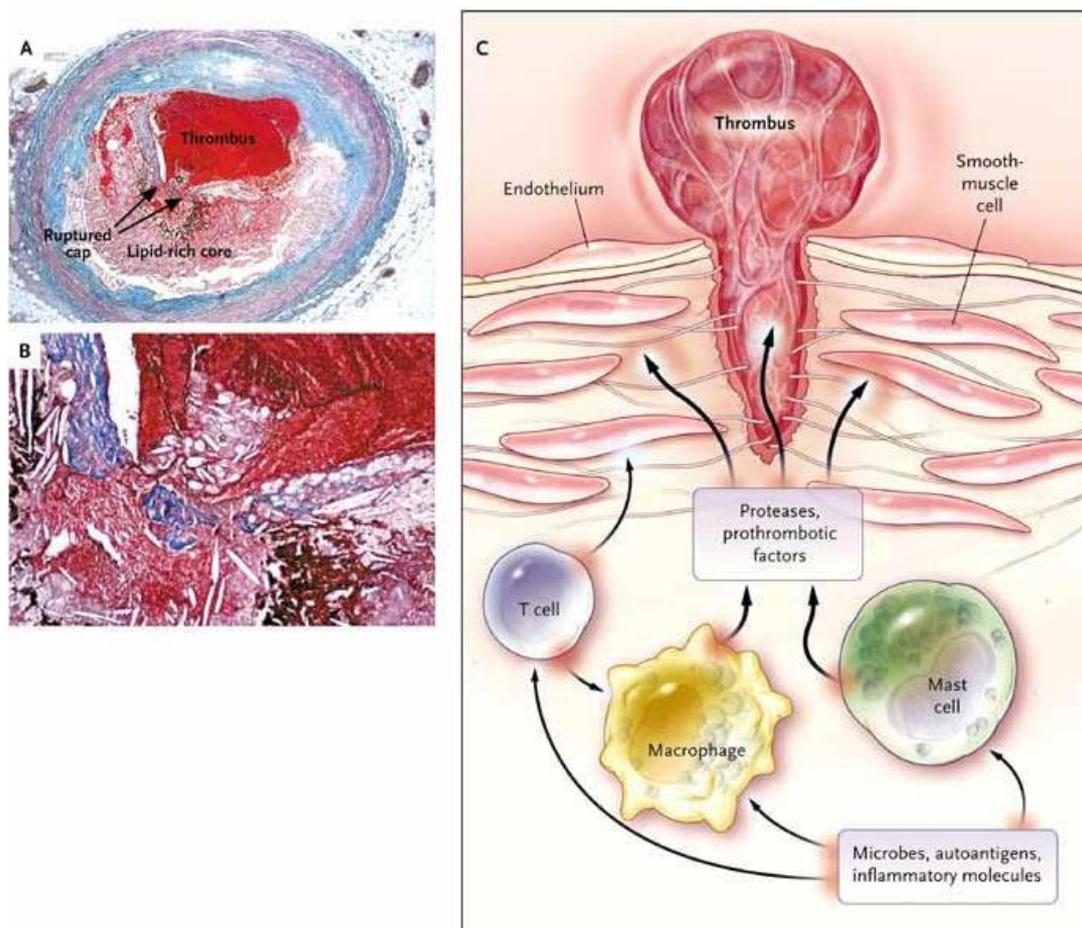
Figure 1.3: Atherosclerotic Lesion in a Human Artery. Reproduced from Hansson GK 'Inflammation, Atherosclerosis, and Coronary Artery Disease (36).

1.2.2 Arteriosclerosis

Arteriosclerosis is the process of reduced vascular wall compliance due to arterial stiffness (39). In the general population it is associated with increasing age and the presence of hypertension (40), both of which are also associated with CKD. However, its prevalence and severity is associated with worsening of eGFR independent of age and hypertension, suggesting that the process of arteriosclerosis is mediated through other factors associated with CKD (40). These factors are likely to include deranged calcium and phosphate metabolism and inflammation leading to fibrosis (33, 34). The former likely leads to microcalcification of the vasculature and hence reduced compliance (40). This process leads to potential reduced vascular flow to end organs and also contributes to worsening HTN and cardiomyopathy, most commonly through left ventricular hypertrophy (40).

1.2.3 Cardiomyopathy

Cardiomyopathy in CKD occurs through many pathways but ultimately likely explains the higher rates of sudden cardiac death and heart failure related deaths seen throughout the spectrum of CKD (41). Concentric left ventricular hypertrophy may be present in up to half of individuals with stage 4 CKD (42). Arrhythmias, such as atrial fibrillation, are also more common in CKD than the general population and may further explain the higher rates of sudden cardiac death (43). The process of cardiomyopathy in CKD is likely to be related to the high prevalence of HTN in the CKD population (44). However, this does not explain all of the difference as those without HTN also more commonly have features associated with cardiomyopathy. Other interacting factors may include arteriosclerosis, as explained above, anaemia and electrolyte abnormalities. The loss of endogenous erythropoietin production in CKD had been thought to contribute to cardiomyopathy simply through anaemia and chronic reduced oxygen delivery to cardiac tissue. More recently though, the lack of stimulation of cardiac erythropoietin receptors due to this deficiency has been shown to lead to cardiomyopathy driven by cardiac apoptosis and fibrosis (45, 46).

1.2.4 Inflammation

Inflammation is associated with atherosclerosis in the general population, but the process is more prominent in those with CKD (47). It is thought to decrease the stability of atherosclerotic plaques and thus increase their risk of rupture and subsequent acute pathology. This was initially suggested through the observational evidence of the association of increased C-reactive protein with CV events (48, 49). The evidence for a causative pathway was further secured through the reduction of inflammation, through HMG-CoA reductase inhibition medications, more commonly referred to 'statins', leading to greater benefit in individuals with raised inflammatory markers compared to those with lower values (50, 51). In addition to this process other mechanisms are thought to contribute but their exact nature is unclear. Reduced renal clearance of pro-inflammatory mediates in more advanced CKD may lead to higher systemic inflammation (47). However, markers of inflammation are raised across CKD stages 3 to 5 compared to controls without CKD, suggesting that reduced clearance is unlikely to be the only mechanistic pathway (47).

Therefore, as well as increase the absolute risk of CV events in CKD, the combination and interplay of atherosclerosis, arteriosclerosis, cardiomyopathy and inflammation may also mean that a different risk profile is present in CKD. This may mean that clinical risk prediction tools require updating and altering specifically for a CKD population.

1.3 Non-Cardiovascular Risk

In addition to the increased CV risk described throughout Section 1.2, CKD is also associated with increased risk of ESRD and all-cause mortality (6). Whilst in relative terms, the risk of ESRD is raised throughout all stages of CKD, in absolute terms the risk does not exceed CV mortality risk until an individual has an eGFR <30 ml/min/1.73m^2 and some degree of proteinuria, i.e. CKD stage 4A2 or worse (6). Some of the increased all-cause mortality risk is related to CV morbidity, but it does not explain all of this increased risk.

Non-CV causes of death that have been associated with CKD include cancer and infectious disease (52-54). Cancer has been associated with both increased morbidity and mortality in CKD compared to those without CKD (52, 54). This is true both for urological tract cancers, such as a renal cell carcinoma whether nephrectomy may be the cure, thus leading to an immediate halving of functional nephrons, and in other non-urological cancers. In addition, established cancer treatment in individuals may differ because of the exclusion of individuals with CKD and so outcomes may also be worse. This may be due to many clinical trials excluding individuals with CKD and therefore, no regulatory licensing of treatments occurs in this group.

Infectious disease, are more common in individuals with CKD than in those with CKD. Both chronic infectious diseases such as HIV and tuberculosis and acute infectious illnesses have been associated with worse outcomes in CKD (55-57). Again, this may related to the underlying pathophysiology of CKD but also the more limited treatment options that may be available in acute and chronic infections.

Other chronic conditions have been associated with CKD including chronic obstructive pulmonary disease and rheumatoid arthritis (58, 59). It is this plethora of communicable and non-communicable diseases that suggests the multi-morbidity associated with CKD leads to an increased risk of all-cause mortality above that which the excess CV risk is attributable to. Therefore, any potential assessment of risk in CKD must account for both the likely different risk factors for both CV and non-CV conditions.

1.4   Clinical Risk Prediction Tools

Prognosis in clinical medicine refers to an event or outcome happening in the future given some information, typically a condition or test result, available at the present time. Published terminology describing tools to predict prognosis, or the risk of future events, is confusing and includes the use of names such as prognostic index, risk prediction models or simply prediction models (60).

For the purposes of this thesis 'risk factor' will be used throughout to mean a measurable variable at the start of the study that is associated with a future CV disease event during the study's follow-up. Secondly, mathematical and statistical workings that combine these risk factors to predict future risk will be referred to as a 'risk prediction model'. Thirdly, the production of predicted percentages and/or probabilities of an event occurring will be described as a 'risk prediction tool', and will also include the use of this prediction in the clinical setting.

The study of prognosis, risk factors and risk prediction models can be grouped under the umbrella term of 'prognostic research' (61). This can have many benefits to healthcare including informing individual patient and clinician choice, developing and planning healthcare policy and delivery, and future research, including clinical trial planning (60).

Risk prediction models, as with any medical intervention such as a new medication, should complete three stages of development in order to show they are clinically effective and appropriately improve outcomes for patients (60). These stages are rigorous model development, validation of the model in other datasets and implementation. However, unlike the process in drug development and monitoring, this process is often performed poorly or not at all (62, 63).

Model development, is most commonly performed of the three stages, perhaps because it is seen as the 'easiest' of them (60). In this thesis the data is 'censored'. Outcomes will not occur in all inividuals before the end of the follow-up period. Survival analysis is the broad statistical term for methods used in censored data and the Cox proportional hazards model ('Cox model') is the most commonly used method (64). Combination of risk factors is used to derive a prediction of risk

('predicted probability') and then compared to what has actually occurred in the study population ('observed probability'). The process of development and its limitations are discussed in Chapter 2.

There are two different types of model testing, also known as validation. Internal validation is the most commonly performed and relies on the use of the same data that developed the model (60). There are two further subgroups of internal validation, apparent and resampling. Apparent simply occurs where the available data is used to develop the models. Resampling reuses the data used for model development to test the model. These methods include bootstrapping and random splitting of the sample. However, because the data are being repeatedly used, model performance metrics may be overly optimistic and therefore there is a risk of overstating the model's effectiveness, often referred to as 'statistical overfitting' (60).

The second type of validation is external validation, where another dataset is used to test the performance of the risk prediction model. This is often performed by the same individuals who developed the model, but should ideally be performed by another independent group (60). By using another dataset, the performance, including the risk of model overfitting, can be tested. Through this process the performance can also be assessed at different timepoints ('temporal validation') and in different locations ('geographical validation'). Ideally, but infrequently performed, this process can be repeated in many different cohorts to further confirm, or refute, the risk prediction model's ability to provide clinically useful information in different settings. Part of external validation includes the process of updating a model. There is perhaps a subtle difference between updating a model and development of a new model *per se* (60). The former will likely only see an adjustment to the calibration of the model and/or the addition of new variables, typically biomarkers that may not have been available when the model was first developed. The latter will see a change to risk factors and their contribution to the model, which is likely to see it regarded as a new risk prediction model altogether.

Finally, risk prediction model implementation requires study and assessment to see how it impacts clinical decision making (60). This step is rarely performed, particularly in a formal clinical trial setting. It has been suggested that this may be

because risk prediction tools are often not considered as a type of health technology requiring assessment in a clinical trial (60).

Further definitions and statistical methods for developing and assessing risk prediction tools are described in Chapter 2.

1.5   Cardiovascular Risk Prediction

1.5.1   General Population Background

More than 350 CV risk prediction models have been developed for assessment of CV disease risk in the general population (65). However, many of these were developed using poor methodology or have not been externally validated, and therefore are not used in routine clinical practice. The few that are used in clinical practice are well embedded in routine care, particularly in UK primary care. This is because they are based on routinely collected risk factors for CV disease and can be automated into electronic patient records. The focus of CV risk prediction has been in those who have not previously had a CV event i.e. for primary prevention of CV disease (66-71). In addition, those tools are largely used for patient counselling in relation to the initiation of cholesterol lowering medications such as statins (70, 72). Despite increased CV risk with reduced eGFR and/or the presence of albuminuria, general population CV risk scores either do not adjust for CKD or use a uniform adjustment for advanced CKD only (73).

In the UK, the tool currently recommended by NICE for CV risk prediction is QRisk2 (69, 72). This is used to predict 10 year risk of CV disease and currently if risk is greater than 10% then a 'full formal risk assessment' with the patient regarding statin commencement is recommended. QRisk2 includes medical record coded CKD stages 4 and 5 as a binary risk factor included in the risk prediction model. During the course of the thesis QRisk3 was developed and published, which extended the definition to CKD stages 3a and 3b, but again treated CKD as a single uniform adjustment across all stages (67).

This plethora of CV risk tools has meant that the focus of future development has now begun to focus on external validation and appropriate adjustment of models for specific populations (65).

1.5.2 Cardiovascular Risk Prediction in Chronic Kidney Disease

CKD populations, with well-established independent increased CV risk but potentially different risk factors as set out in Section 1.2, provide an excellent example of where adjustment to, or development of novel, risk prediction models may be warranted. In the absence of this, general population risk prediction models need to be assessed in CKD populations. The addition of eGFR and ACR to general population risk prediction models has be shown to improve model performance (74). In addition, the inclusion of risk factors associated with arteriosclerosis and reduced vascular compliance, such as calcium and phosphate, in CKD-specific CV risk prediction models, may be beneficial if model performance is shown to be suboptimal in CKD populations.

CV risk prediction in CKD might lead to different risk prediction and intervention compared to the current use of models for primary prevention of CV disease in the general population. Whilst statins (20mg atorvastatin once daily) are currently recommended by NICE for all individuals with CKD (75), more tailored treatment based on risk prediction should be considered. For instance, higher statin doses, consistent with intensity used in secondary prevention, could be considered for those at higher CV risk. Other treatments, such as lower blood pressure targets could also be considered for those at higher risk, where the additional benefit of lower targets may outweigh the potential risk of iatrogenic harm such as hypotensive related falls, particularly in older, frailer individuals. Equally, non-pharmacological interventions could also be targeted at individuals with predicted higher risk. These might include dietary, lifestyle and exercise interventions. The simple knowledge of knowing a quantifiable risk of a condition may also be enough for a clinician to initiate a conversation around future plans for a patient's care. Tools such as 'heart age' may also give the individual an intuitive measure of their risk (67).

From a public health perspective, specific CV risk prediction in CKD may aid the design of clinical trials through risk stratification of treatment or inclusion of an individual into the study. In addition, specific risk prediction may also help healthcare organisations adjust for differences in populations when assessing differences on outcomes across organisations (60).

In 2013 Tangri *et al* published a systematic review of risk prediction models in CKD for the outcomes of CV disease, ESRD and all-cause mortality (73). Three cohorts, producing a total of six models, had developed CV risk prediction models specifically for CKD populations, but they were identified to have significant methodological weaknesses. These limitations included limited assessment of the models' prediction performance and no external validation of the models. These issues contributed to the authors perceiving the models to lack clinical utility. In addition, these models were developed in secondary care, despite the majority of CV risk management in CKD being performed in primary care. Their conclusions were "further development of models for cardiovascular events….is needed" (73).

1.6   Thesis Aims and Objectives

The overall aim of this thesis was to assess the clinical utility of CV risk prediction tools in CKD populations. In order to achieve this the specific objectives were as follows:

- To develop a CKD cohort (Chapters 3 and 4)
- To perform a systematic review to establish CV risk factors in CKD populations (Chapter 5)
- To identify newly developed CV risk prediction tools for CKD populations since the systematic review by Tangri *et al* (Chapter 6)
- To assess CV risk prediction tools using a CKD cohort (Chapter 7)
- To assess the need for adjustment of risk prediction models in order to aid and improve their implementation into clinical practice. (Chapter 8)

In addition to these objectives, Chapter 2 provides the methodology background, including statistical methods, for the thesis and Chapter 9 summarises the findings and discusses the implications for clinical practice in relation to CKD care.

Chapter 2

2. Development of Dataset

This chapter gives an overview of the methods used to develop the cohort for the thesis. It describes the ethical approval received for the study. This is followed by a description of the 'IMproving Patient care and Awareness of Kidney disease progression Together' (IMPAKT) tool and how it anonymised data for the dataset. A description of the use of medical record coding in primary care is then provided. An overview of the secondary care data sources, Hospital Episode Statistics (HES) and Myocardial Ischaemia National Audit Project (MINAP), is then given. Finally, event identification in the merge datasets is presented.

2.1. Introduction

In the UK, clinical research must be reviewed by the Health Research Authority, part of the NHS, before it can take place. The review process is broadly based on the Declaration of Helsinki, first adopted in 1964 and most recently updated in 2013 (76). In the case of this thesis, the most pertinent parts of the declaration are privacy and confidentiality of personal information, protection of the health and rights of patients, and the legal and regulatory standards in the UK.

Privacy and confidentiality of patient information is vital for the Leicester City and County (LCC) study particularly as practice level, but not individual patient consent was given. Even though data were anonymised, there may still have been the potential for re-identification of individuals even if a name was not attached to an individual patient record. This is why very specific processes for patient anonymisation and prevention of re-identification were developed for the study, as will be described.

Protection of the health and rights of patients is linked to the issue of privacy and confidentiality. In order for large scale epidemiological studies to take place in the UK and elsewhere, patients and the general public need to be reassured that this is protected, otherwise health, particularly psychological, may be detrimentally affected.

Finally, clinical research must follow the legal and regulatory standards set out in UK law to ensure that what takes place is appropriate and legal. These standards have been designed to protect individual patients and the general public from unethical research. Again, they are particularly applicable when individual consent has not been received as they act on behalf of all of us to protect our rights and privacy.

2.2. Ethics Review

Prior to the commencement of practice recruitment and data collection for the study, full ethical approval for the development of the LCC cohort was received from the Health Research Authority (Integrated Research Application System, 'IRAS' identifier 197145, Research Ethics Committee, 'REC' reference 16/EM/0315) and the approval letter is presented in Appendix 2.1. As anonymised data were extracted, practice level consent was received from GP surgeries and individual patient consent was not sought. A copy of the practice information leaflet and consent form are attached in Appendices 2.2 and 2.3. Data were extracted from GP electronic records using the IMPAKT tool.

2.3. Data Extraction Tool for Cohort Creation

IMPAKT is a web-based CKD management and audit software tool (77). It was initially developed as a quality improvement and audit tool for CKD. IMPAKT uses Morbidity Information Query and Export Syntax, also known as 'MIQUEST', search methodology to analyse primary care clinical electronic medical records (78). It is compatible with all major primary care medical records systems including SystmOne and Egton Medical Information Systems.

The programme was developed and is maintained and updated by Dr David Shepherd, University of Leicester and Saffron Group Practice. A modified version of IMPAKT was developed specifically to create the LCC cohort for this thesis. Data are identified through the use of Read codes, medical record codes used nationally and maintained by the NHS Centre for Coding and Classification (79). The data associated with the Read codes such as the date the code was entered in the system and a value, such as serum creatinine, were also extracted.

In practical terms, all practices within the three Leicestershire based Clinical Commissioning Groups (CCGs), were invited by letter to consider participating in the study. If agreement was received then a visit to the practice was arranged, where formal consent for the study was discussed and completed. After consent was received, the IMPAKT tool, run through an automated Mircrosoft Excel program, or 'macro', was downloaded from the study's secure website. The macro created

specific MIQUEST searches to be run on the practice's clinical system. These searches were then completed under the supervision of a member of the practice's staff. The individual MIQUEST search results were then amalgamated by IMPAKT into a single, anonymised comma-separated values file for extraction from the practice. Depending on the size of the practice, this process took between 15 and 60 minutes to complete.

2.4.   Dataset Anonymisation and Security

Anonymised datasets from each practice were exported for secure storage to the University of Leicester Clinical Trials Unit. Individuals were identified through a one-way hashed version of the NHS number, or 'hNHS'. Any data leaving a practice were identified through this 40 character alpha-numeric identifier. NHS numbers did not leave practice sites. The hashing process was performed using a 'Secure Hash Algorithm 1' embedded in the IMPAKT tool. Figure 2.1 provides an example of the process. At no point did the data processors have direct access to the algorithm and the owner of the encryption key, Dr David Shepherd, had no access to the research data. Therefore, re-identification of individuals, even inadvertently was not possible by myself or any other member of the research team.

NHS Number: **1234567890**

one-way hashing algorithm embedded in IMPAKT

Y18IF2BAJMGHY57C659AFC6369XYB0AC79B56XY8
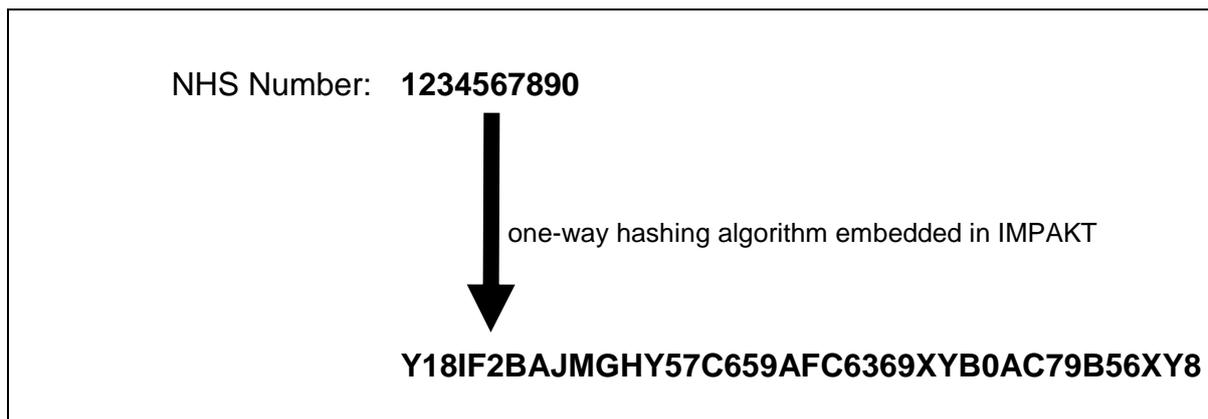
Figure 2.1: An example of a one-way hashed NHS number from the IMPAKT tool.

In addition, pseudo-identifiers such as date of birth and postcode, were adjusted to reduce the risk of re-identification. Date of birth was converted to age in years and deprivation score was calculated based on postcode and the former was extracted by the IMPAKT tool. All data were transferred using password protected encrypted USB sticks with FIPS 140-2 256 bit encryption, an approved international standard for data encryption. Data were transferred to Leicester Clinical Trials Unit for secure storage.

2.4.1. The Leicester City and County Chronic Kidney Disease Cohort

The IMPAKT tool was used to collect CKD related data for practices for the LCC cohort. IMPAKT allows for the retrospective collection of data. To create a cohort with five years of follow-up, the start date of the cohort was set at 1st November 2011 with a fixed end date of 1st November 2016. Full informed consent was taken from each practice.

At the beginning of the cohort's development, the use of EPI eGFR was recommended for diagnosis of CKD (11). The threshold for diagnosis was two or more eGFR <60 ml/min/1.73m^2 more than 90 days apart regardless of the formula used (6). University Hospitals of Leicester continued to report the MDRD formula at the beginning of the study (10). Based on analyses of conversions between formulae, a threshold of MDRD <65 ml/min/1.73m^2 equates to an EPI eGFR <60ml/min/1.73m^2 in all individuals regardless of the demographic details utilised in eGFR formulae. Therefore, the IMPAKT tool's threshold for data extraction was

altered to MDRD <65 ml/min/1.73m^2 to detect all individuals with a CKD diagnosis when eGFRs were recalculated via the EPI eGFR formula. Throughout the study time period University Hospitals of Leicester Biochemistry department used the Jaffe method for measurement of serum creatinine (80).

2.5.   External Data Sources and Linkage

In addition to primary care Read code identified events, two external data sources were utilised to identify additional CV events. University Hospitals of Leicester's data were extracted and anonymised. This data related to all inpatient stays from HES and MINAP data where either the admission reason was CV event related or an inpatient CV event occurred.

HES data are collected by all hospital trusts in England, including University Hospitals of Leicester, on behalf of NHS Digital. It records all inpatient admissions, accident and emergency attendances and outpatient appointments. Data are based on clinical coding of medical notes entries and for the purpose of this thesis ICD-10 codes from HES data were used.

MINAP is a national audit of the quality of management myocardial infarctions in England and Wales. It was setup in 1999 and was performed within University Hospitals of Leicester throughout the follow-up period of the LCC cohort. Data for all individuals admitted with suspected acute coronary syndrome are recorded with a final definitive discharge diagnosis made by the cardiology team responsible for the patients care.

Data were merged and agreement was checked by comparing non-modifiable factors of age, gender and ethnicity code. All 7,005 potential records matched between the LCC cohort and the secondary care datasets for gender. 5,320 (76.0%) of ages, exactly matched between datasets. One (0.01%) age difference was two years and the remaining age matches (24.0%) were all within one year of each other. For anonymisation purposes, dates of birth were not extracted for any of the datasets, and so these minor differences probably reflect differences in rounding of ages in the datasets.

2.6. Use of primary care medical records in research

Electronic medical records have been established in primary care for the last two decades. Their use has also been embedded in UK medical research. National standardised Read codes allow the identification of events through these records (79). The two most commonly used electronic medical record systems in UK primary care are SystmOne and Egton Medical Information Systems, both are compatible with MIQUEST search methodology to allow for efficient searches of medical records.

2.7.   Event Identification

Outcomes were assessed based on the time to the first event after the index date and multiple events for the same individual were not considered. Codes used for all outcomes in the different data sources are described and discussed further in Chapter 4.

Herrett *et al* suggested that primary care records were the single most complete source of non-fatal myocardial events and identified 75% of events, as shown in Figure 2.2 (81). The additional HES data alone would contribute a further 17% and MINAP data alone 15%. The LCC cohort was not linked to Office for National Statistics (ONS) death registry outcome data. Herrett *et al*'s data would suggest that the lack of this data would underestimate fatal and non-fatal myocardial infarction events by 3.0% (crude incidence of 242.6 per 100,000 person-years versus 250.0) as shown in Figure 2.3.
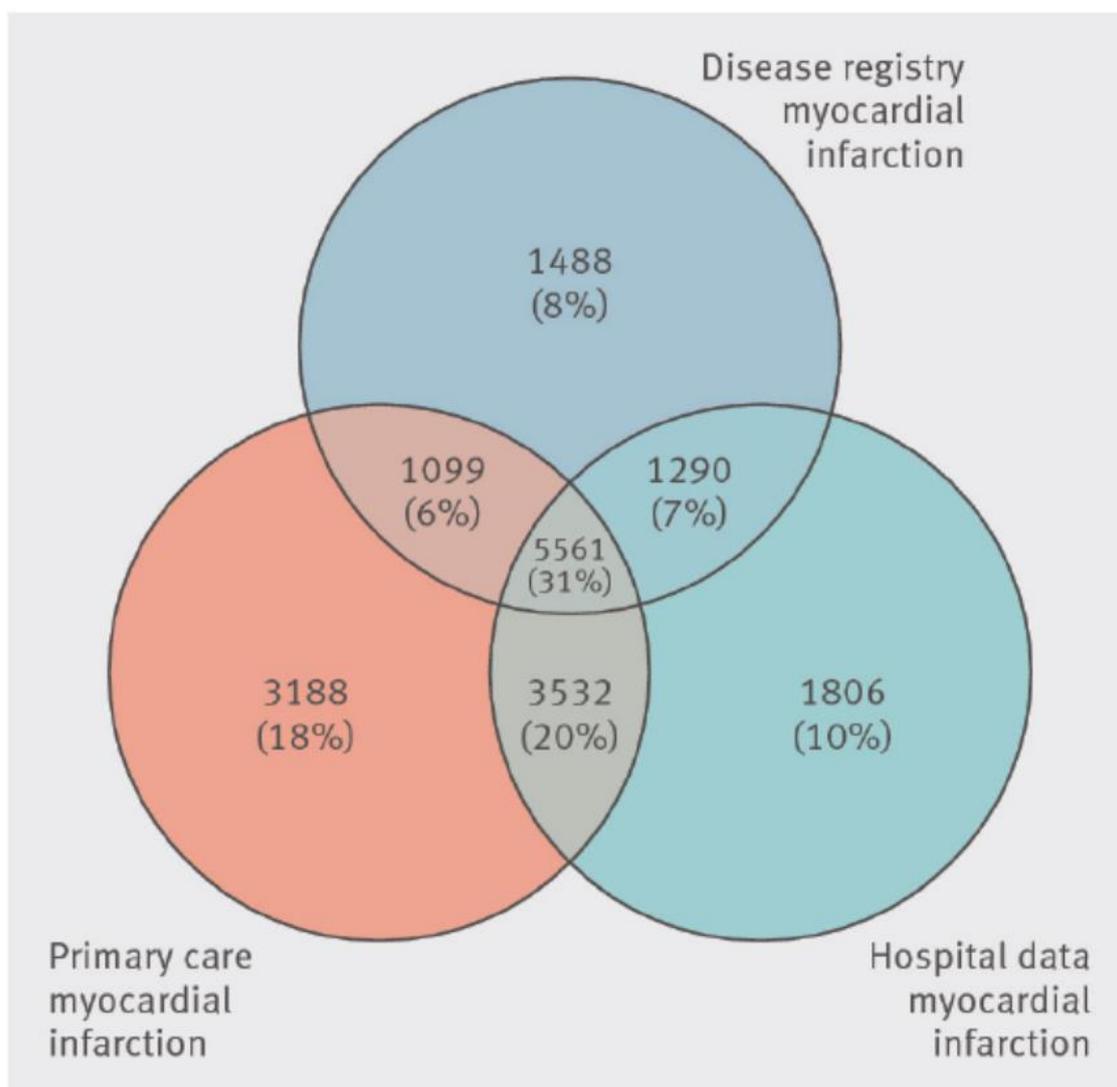
Figure 2.2: Distribution of Non-fatal Myocardial Infarction across National Databases. Figure reproduced from Herrett et al 'Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study' (81).
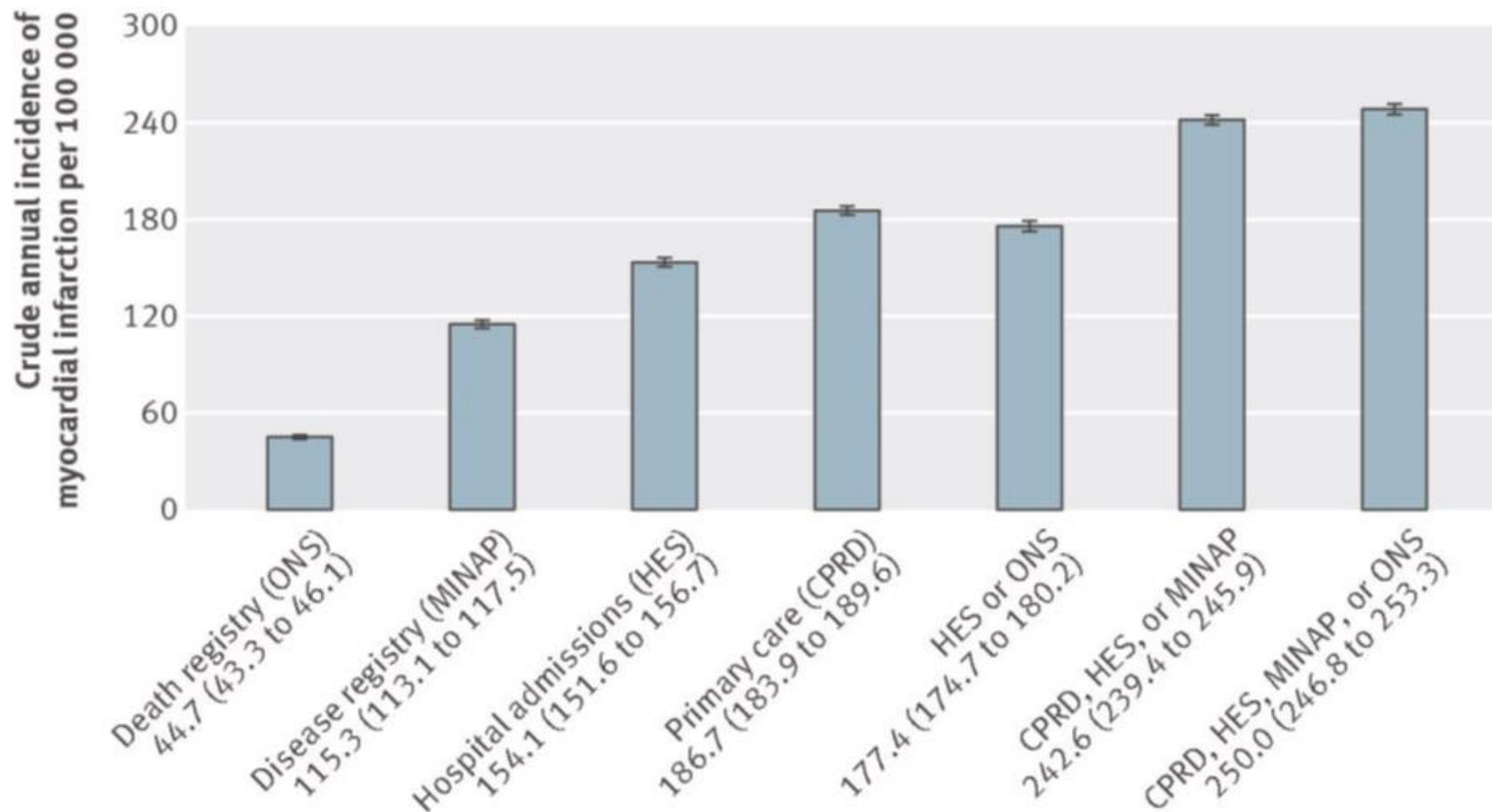
Figure 2.3: Crude Annual Incidence of Fatal and Non-fatal Myocardial Infarctions across different National Databases. Figure reproduced from Herrett *et al* 'Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study' (81).

2.8. Conclusions

The Declaration of Helsinki underpins the process of ensuring that clinical research performed involving humans is ethical and protects their rights. In the case of anonymised epidemiological research this is particularly important as individual patient consent for data is not received. This means the legal and regulatory framework, overseen by the HRA in the UK is particularly pertinent in the case of the LCC cohort developed for this thesis. The anonymisation of individuals for LCC has been ensured through a robust process and the risk of re-identification has been minimised. This process though has continued to allow for data linkage to other datasets to enhance the quality of identification of CV events for the rest of the thesis. The importance of this will be described and discussed further in the next and later chapters.

2.9. Summary

This chapter has given an overview of the methods used to develop the cohort for the thesis. The IMPAKT tool is a data extraction tool designed for CKD clinical audit and anonymised research data collection. In order to explain the methodology of how the data were collected, an overview of medical record coding in both primary and secondary care has been given. The role and rationale of using and linking outcome codes between primary and secondary care has been explained. The methods for how the data were collected has now been set out and the next chapter moves on to describe the specific methods used to analyse the collected data for risk prediction model assessment.

Chapter 3

3   Methods for Risk Model Assessment

This Chapter provides an overview of the methods used to assess the performance of the risk prediction models identified. Firstly, survival analysis, and its most commonly used model, The Cox proportional hazards model, is introduced. Secondly, the methods for statistical assessment of risk prediction models of discrimination and calibration are presented. Finally, the approach of the thesis to missing data and the methods used are described. Throughout this thesis Stata version 15.0 was used for data analysis.

3.1. Introduction

In its simplest form risk prediction is about 'using today's information to try to say what will happen tomorrow'. If you would like to know whether it will rain this afternoon, the most basic risk prediction tool would be, and probably used by humans for millennia, to look up to the sky for clouds this morning. Of course, weather prediction has moved on from this primitive assessment but nevertheless the principles remain the same albeit hopefully with better accuracy.

By 'today' it is meant any information available today rather than collected today. It is unlikely, particularly in routinely collected data, to have all information, such as blood test, that has been taken today. By tomorrow, it is meant a pre-defined, and in medical research clinically relevant, time span. This will vary from condition to condition and will partly depend on disease incidence and severity. In this thesis the time horizon for CV events is assessed in terms of years.

Risk prediction relies on a number of processes that must each be performed robustly to produce reliable 'todays' and 'tomorrows', ultimately so that reproducible, accurate predictions are delivered. These collection processes for the LCC cohort are explained in Chapter 4 so are not discussed further here, with the exception of the issue of missing data.

Once data collection is completed, and adjustments for missing data have been made, statistical models are produced which give predictions. As follow-up occurs across a time period, it may vary from individual to individual and perhaps most importantly not all individuals will end up having the outcome in question. This therefore relies on statistical analysis of time-to-event, more commonly called 'survival analysis'. This statistical process is introduced in this chapter, along with its close association with the Cox proportional hazards model. The focus of this part of the methods is then how the model can be used to devise a risk prediction tool. This will then lead to how risk prediction tools are assessed using the concepts of 'discrimination' and 'calibration'. Discrimination is the ability of a risk prediction tool to distinguish those who experience an event from those who do not. Calibration is concerned with

establishing whether what was predicted by the risk prediction tool matches reality. As will be explained in coming chapters, the focus of the thesis is on assessing pre-existing risk prediction tools, or 'external validation'. Therefore this part of the chapter will be in relation to these methods from the perspective of external validation.

3.2. Missing Data

As LCC utilises routinely collected clinical data, missing data will be present. The proportion of missing data is likely to vary significantly by variable, with demographics like gender and age likely to have close to no missing data and other variables higher levels. Clinical tests that are performed less frequently, such as serum bicarbonate, will likely have missing data for a majority of individuals. The hypothesised reason for missing data is also likely to vary between variables. All individuals should have an age or gender and therefore if missing this is likely to be due to administrative errors within the clinical records. Whereas, a missing serum bicarbonate could be an administrative error but is more likely to be related to a clinical decision that the test was not required.

3.2.1. Missingness Mechanism and Types of Missing Data

The complexity of how data become missing is referred to as the 'missingness mechanism'. Whilst looking at patterns of missing data can aid reasons for it being missing, it is only within the context of clinical knowledge that the reason for it being missing can be considered. Missing mechanisms were first described and categorised more than 40 years ago (82).They can be considered in two ways, the probability of being missing and the difference, or similarities, between individuals with and without missing data (83). Rubin categorised data into 'missing completely at random' (MCAR), 'missing at random' (MAR) and 'missing not at random' (MNAR) (83).

If the probability of a missing value is unrelated to both the missing value and any other variable in the data than it can be categorised as MCAR. For instance, if a serum bicarbonate sample was dropped by a technician in the laboratory then this is likely to be MCAR. If data is MAR a value's probability of being missing is not related to its value after other variables have been considered. Serum bicarbonate is rarely taken without other biochemical tests, so if a serum sodium sample has not been taken then it is likely that serum bicarbonate will also be missing, thus the absence of serum sodium, and other electrolytes may explain the absence of serum bicarbonate. MNAR occurs when variables not in the dataset explain the reason for missing data. In the

cases of bicarbonate, individuals with chronic obstructive pulmonary disease may have it checked, but this co-morbidity may not be recorded in the dataset.

As can be seen by the serum bicarbonate example, the type of missing data for the same variable can vary within the same dataset. Therefore, rather than speculating about the type of missing data, the purpose of statistical methods to address missing data are to explore inferences made under the different categories of missing data.

3.2.1. Approaches to Handling Missing Data

Many methods have been developed and suggested for handling missing data. For this thesis two approaches will be undertaken and compared in sensitivity analysis; complete case analysis and multiple imputation. Other methods such as simple imputation of population mean or the regression mean are prone to increasing bias and will therefore not be considered (83).

3.2.1.1.   Complete Case Analysis

Complete case analysis, whilst prone to increasing bias and reducing efficient use of data, is the most commonly used approach in data analysis (84, 85). Two main approaches can be taken, casewise and pairwise deletion.

Casewise deletion removes an individual from the analysis if any missing data are present. This may be more acceptable where the vast majority of individuals will have complete data. When using routinely collected data, where across multiple variables the number of excluded individuals will be high, large percentages of individuals will be excluded, particularly where less commonly recorded variables, such as serum bicarbonate, are included in the analysis. Provided data is MCAR, unbiased results will be given but precision will be substantially decreased. Other types of missing data will lead to bias estimates and again reduced precision.

Pairwise deletion deletes values on a by variable basis instead of the data for the whole individual. This allows for investigation of relationships between

variables where the data are available. Similarly, MCAR data will produce unbiased results with reduced precision, albeit to a lesser degree than casewise deletion. MAR and MNAR will suffer from the same limitations as casewise deletion.

3.2.1.2.   Imputation Methods

In general terms, all imputation methods utilise the available data to create a distribution for the missing data. Two main 'principled' methods exist, maximum likelihood and random imputation. Both aim not to replace the missing value directly but to draw from an estimation of the variables distribution based on the present values within the dataset. The process to estimate the distribution is the main difference between these two methods.

Maximum likelihood aims to estimate the missing value that would lead to the greatest probability of the observed values having occurred (86). This relies on an iterative approach, most commonly using the expectation-maximisation algorithm, to estimate the missing values from the distributions parameters. This is then used to update the distribution until convergence is achieved. This method however is limited in its use in Cox models and with large amounts of missing data across multiple variables because its computational requirements increase exponentially. Maximum likelihood will therefore not be considered further as a method (86, 87).

Random imputation is based on estimates of the mean from multi-variable regression analysis of the available data (83). Rather than utilising a simple mean from the data, is accounts for variation in the process. This is calculated through the observed values' and predicted values' differences, otherwise known as the residuals (83). The missing value is calculated from the distribution of the residuals. This is repeated for all variables to produce a 'complete' dataset. However, because each value is drawn from a distribution, repeating the exercise again will produce different missing values. Hence, this process is repeated to produce multiple imputed complete datasets thus helping to account for variation between the single datasets. These complete datasets are then individually analysed using the selected statistical techniques.

The mean of the results is then calculated, but importantly so too is the variance from the variance within and between each complete dataset's results (83). This process of random imputation across multiple datasets and consolidation of results will be referred to as 'multiple imputation' (MI).

3.2.2. Methods for Multiple Imputation

A number of considerations must be made when applying a MI model. The nature of the variables need to be considered from two different perspectives, the characteristics of the variable to be imputed and what is used to estimate the imputed values (88). In addition, the number of imputations needs to be considered, which will relate to the proportion of missing data for the variable (88).

3.2.2.1.   Characteristics of Imputed Variable

The imputed variable can be of any nature and the methods need to be adjusted accordingly (88). The next chapter will show that data for categorical variables were complete, primarily because most categorical data were medical code related. As individuals are very rarely coded for the absence of a disease in medical records, in this thesis, if an individual was not coded with a condition then it was assumed not to be present. Therefore, only continuous variables required imputation and categorical variables, either ordered or unordered, were not considered further for the description of methods.

Approaches for semi-continuous variables, where many values are equal, often zero in value, were also considered. For external validation of the models, what is being imputed will be dictated by the model initially presented. In models where variables have been categorised for use in the model, the continuous variables will be imputed before the imputed values are then categorised. Where ratios of two variables are to be included in the model, then the ratio should be imputed. If the ratio is based on imputed values of the individual components of the ratio then the results may lead to extreme values of the ratio and significantly affect risk prediction if inappropriately used (68, 89).

3.2.2.2.   Assessment of Non-missing Values of Imputed Variables

The first consideration for the imputed values is its distribution. The use of Rubin's rule to aggregate the imputed datasets relies on the assumption of the

variable being normally distributed (88). Where a normal distribution is not present, a number of approaches can be considered within the MI process.

Firstly, an approach that occurs in other types of statistical modelling is the transformation of a variable to a form with a distribution closer to normal (88). Values are then imputed based on the transformed variable before reverse transformation back to the original units. Secondly, predictive mean matching, which is more suitable when the normality assumption is inappropriate. This can occur when transformation techniques do not produce a normal distribution and/or when the relationship to other variables is non-linear. It is also only appropriate when no extrapolation outside of the observed range of values is required (90). Thirdly, ordinal logistic regression can also be considered (88). Ordinal logistic regression considers the variable as ordinal, as opposed to continuous, therefore allowing the use of the proportional odds assumption. Again, no extrapolation outside of the observed values can be performed. Each unique value requires the creation of a level within the model, and the number of levels may be restricted by the statistical software.

Where semi-continuous values are present, predictive mean matching or ordinal logistic regression can be used. An alternative is to create a binary variables as an indicator for whether the value is zero and a separate continuous value for other (positive) values (88, 91). Therefore, all variables require assessment of their distribution. This was performed using histograms and overlayed normal distribution probability ('P-P'), and quantile ('Q-Q') plots.

3.2.2.3.   Variable Selection for MI Model

Variable selection and form for the MI model is based on a number of factors and may vary between the risk prediction models being validated. Firstly, variables must have no missing values, i.e. MI should not be performed based on other incomplete variables. Secondly, variables included in the analysis model must be included in the MI model. Clearly, this may vary depending on the model being tested, but for external validation is dictated by the development model (92). Thirdly, the model's outcome variable needs to be included otherwise significantly biased estimates can occur (93). Fourthly,

specific to survival analysis, a measure of time, censoring and the cumulative hazard function needs to be included in the model, although the precise best form, particularly for time, is unclear (94-97).

In addition to these variables, predictors of the presence of missing values as well as their actually values should be considered if the MAR assumption is to remain true. Suggestions have been made that this should therefore include any variable that 'significantly predicts' the presence of an incomplete variable (88).

The form the variable takes also needs to be considered, but for the use of MI in external validation this is partially dictated by the model under assessment. For model development this will dependent on the presence of non-linear relationships and potential interaction terms specified in the model (88). However, the exact choice is unclear and sensitivity analysis for model terms will be required to 'find an imputation that is both congenial and a good representation of the data' (88).

3.2.2.4.   Number of Imputation Cycles

Rubin's original description of MI and the subsequent decade of research suggested that three to five imputation cycles were 'adequate' to achieve 'efficiency' of the process (83, 98). 'Efficiency' refers to loss of the true variance in the estimation compared to an infinite number of imputation cycles (92). This relative efficiency is related to the fractional of missing information (FMI) which in itself is based on the between and within-imputation variance:

$$FMI = B/(W + B) \qquad\qquad (3.1)$$

where B = between-imputations variance and W = within-imputation variance.

B and W are both estimates, and therefore so is FMI i.e. the number of imputations to limit the loss to a certain level is also an estimate. The relative loss of efficiency for a number of imputations, m, can be estimated through:

$$FMI/m \qquad (3.2)$$

So from equation 3.2 for FMI up to 0.25, five imputations will limit loss to 5% and 20 imputations to 1.25%. This approach though does depend on what is regarded as an 'acceptable' loss of power, and will vary by research question. Further development of the methodology in the last 15 years has stressed the importance of not only the efficiency of the process, but also its reproducibility (99). As FMI is likely to be less than the fraction of missing cases a general 'rule of thumb' has been proposed as (88):

$$m = FMI \times 100 \approx c \times 100 \qquad (3.3)$$

where c = proportion of incomplete cases.

Two separate simulations studies have suggested this creates re-producible results provided FMI < 0.50 (88, 100).

3.2.3. Summary of Multiple Imputation Model Selection

Based on the above methodology and assessment of the cohort, the following general principles will be used for the MI models throughout the thesis:

1.   When non-normally distributed variables require imputation

    1.2.   Transformation to non-skewed values

    1.3.   Imputation of above missing values

    1.4.   Re-transformation back to original values

2.   Variables to be selected in the imputation model

    2.1.   All variables in the model to be tested

    2.2.   The outcome variable

    2.3.   Time indicators

        2.3.1.   Time variable

        2.3.2.   Transformed time variable

        2.3.3.   Indicator of censoring

    2.4.   Any other variable that predicts the absence of the variable to be imputed ($p < 0.05$ for prediction)

3.   Number of imputation cycle

    3.1.   Proportion of incomplete cases to be calculated

    3.2.   Imputation cycles to be 100 times the above proportion to nearest five cycles

3.3. Survival Analysis

The statistical approach to assessing the time taken for an event to occur has varied nomenclature including 'time-to-event analysis', 'failure time analysis' and 'event history analysis'. Throughout this thesis this approach will be referred to as 'survival analysis' and the primary events of interest will be combined CV events. CV events will not be observed in all individuals for three main reasons. Firstly, a CV event may not have occurred before the end of the study. Secondly, the individual may die of a non-CV cause. Thirdly, the individual may be 'lost' to follow-up, most likely because they have moved GP practice before the end of the study time period. These are 'censored' individuals and cannot be simply ignored from the study due to their incomplete data. Therefore, individuals are only considered at risk when they are under observation within the study. For the dataset developed for and used in this thesis, the date of the patient record searches was, and therefore entry into the study was fixed at 1st November 2011. If the individual was not censored then the study follow-up finished on 1st November 2016, when the searches for the end of the study were set.

3.3.1. The Survival and Failure Functions

The survival function, S(t), records the proportion of individuals who have not had a CV event as a function of time (t). For t = 0, S(t) = 1 as no individual has had a study relevant CV event before follow-up begins (101). Similarly, the failure function, f(t), is the probability density function of failure and for t = 0, f(t) = 0 as no individual has started follow-up and therefore failed. The rate of decline of S(t) is dependent on the risk of an individual having a CV event at a specific time, which is referred to as the hazard function, h(t). The proportion of individuals at risk at a time increment will be the denominator for the h(t). This will exclude individuals who have been censored during the time period in question, either because they have experienced the event or have been lost to follow-up. In its simplest form, S(t) can be non-parametrically displayed as a plot of the time against the proportion of the population surviving, this is most commonly displayed as a Kaplan-Meier plot (102). This demonstrates that the rate of S(t) decline is dependent on the hazard rate, or instantaneous failure rate. h(t) is related to S(t) but represents the failure rate given that an individual has survived up to time t, in other words the conditional failure rate. Its value can lie between zero, when essentially there is no risk of the event and infinity when in that time period failure is certain. Hazard rates can change over time, or not change at all during the follow-up period and this rate is determined by the population in question and the disease process (103).

3.3.2. The Cumulative Hazard Function

The cumulative hazard function, H(t), is the core concept in survival analysis and again it is related to the two previously described functions. It sums the total risk accumulated up until time, t. These three functions can be related to each other as follows (103):

$$H(t) = -\ln\{S(t)]$$

(3.4)

$$h(t) = \frac{f(t)}{S(t)} \qquad\qquad (3.5)$$

$$f(t) = h(t)\exp\{-H(t)\} \qquad\qquad (3.6)$$

3.3.3. The Cox Proportional Hazards Model

Whilst other models can be for survival analysis, the focus of this thesis will be the Cox model (64) because this is the model utilised to create and validate CV risk prediction models in the existing literature identified by Tangri *et al* and updated in Chapter 6 (73). Therefore external validation of these models will also utilise it.

Overall, the model assumes that its co-variables multiply in their effect on the baseline hazard function i.e. if two risk factors that both double risk are present, then overall risk is four times higher than the baseline hazard. Any multiplicity of effect also holds over time, i.e. if a factor halves the rate then this will be true whether it is a second, day, year or millennium into the follow-up period. The Cox model does not allow for an estimation of baseline hazard from the data available. This is a general statistical strength of the model, but in a prediction setting is a weakness because the baseline hazard must be estimated.

Mathematically for individual 'i' the hazard function from the Cox model is expressed as (64):

$$h_i(t|x_i) = h_0(t)\exp(x_i\beta) \qquad\qquad (3.7)$$

Where x_i represents an individual's values of the variables in the model and β is the vector of these variables' regression coefficients. In its simplest form this is the sum of the products of each value of the variable in the model for person *i* and the value of β for that variable, or $x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3$......... . $h_0(t)$ is the baseline hazard function and does not take a specific shape i.e. the variables are assumed to have the same effect across time given the baseline value and

the coefficient value. $h_0(t)$ is essentially the hazard function when all variables in the model are zero, in other words when all value of x_i are zero and therefore $x_i\beta$ is also zero meaning exp(0)=1. It is important to note that 'baseline' does not refer to lowest risk but merely a reference point. This concept will be expanded upon later in the chapter when risk prediction is discussed in more detail.

The Cox model makes two main assumption of the data, non-informative censoring and proportional hazards (64). Methods for testing these assumptions are fully explained in Chapter 8 (103). Non-informative censoring suggests that the probability of an event occurring is unrelated to the probability of censoring. For this thesis, most individuals are censored at the date of data collection from the practice and this is therefore unlikely to be related to the risk of a CV event. The second situation to consider is when an individual has left their GP practice. In order for the non-informative censoring assumption to be violated, individuals would have to systematically be leaving practices and registering at other practices knowing that they would be more likely to have a CV event, again this is unlikely. Finally, death is a censoring event in the context of the standard Cox model as observation of the individual is terminated at this point. However, it differs from the previous two examples of censoring as it is known for certain an individual will not experience an event in the future. In previous cases the individual may still experience an event in the future but we just do not know when. The impact of this assumption is considered throughout the thesis and discussed in detail in Chapter 8.

The second assumption, proportional hazards, is that if covariables are constant then so is the multiplicative effect of the model (64). i.e. if a risk factor doubles risk then it doubles risk throughout the follow-up period regardless of when risk is considered within the model's timeframe. The proportional hazards assumption can be assessed through differing methods (103).

Statistical tests can be performed to test for evidence against the null hypothesis of proportional hazards being present. The global and perhaps most powerful test is the link test (103). This suggest that if the Cox model is correctly specified then, except through chance, no further independent variables associated with risk will exist. If this is not the case then a 'link' error

will occur. This is tested through taking $x_i\beta$, and adding its square to the model. If the model has no link error, i.e. the model is correctly specified then the β co-efficient for the squared term should have no evidence for being different from one i.e. the null hypothesis is that the squared term equals 0. This interaction term is referred to as hat^2. This does not specify the part of the model that violates the assumption, but whether evidence exists or not, hence it being referred to as a global test.

As proportional hazards suggest that there is no change over time, it can be also be assessed through the use of interaction terms between the variables of interest and time, or transformations of time such as ln(time). For a multivariable risk prediction model this can be assessed for the overall sum of the β co-efficients and if evidence is found for violation repeated for each individuals variable in the model.

Graphically, the cumulative survival function can be plotted against time for different groups and if they do not cross then proportionality can be assumed. Further testing of the data will add information in addition to the proportional hazards assumption. The log-log plot is the plot of log cumulative hazard versus log survival time, this also provides information in relation to hazard time. The vertical distance between the two lines on this plot is an approximation of the hazard ratios over time. Therefore in order for the proportional hazards assumption to hold, the vertical distance over time must be constant i.e. the lines must appear parallel. These methods apply to comparisons of single variables, but are less applicable in multivariable analysis such as in risk prediction models. This can be tested through assessing the residual values of the β co-efficients of the model against time. If the proportional hazards assumption holds then the residual values should be consistent across time i.e. the line on the plot will be horizontal.

3.3.4. Alternatives to the Cox Proportional Hazards Model

For decades after it was first fully described in 1972 by Sir David Cox, the Cox Model was almost ubiquitously associated with survival analysis and related

predictions (64). More recently, flexible parametric models have been described and used in survival analysis (104). Flexible parametric models have a number of advantages over the Cox model, but for risk prediction and hence the purposes of this thesis there are probably two main benefits. Firstly, the assumption of proportional hazards does not need to be met and therefore there can be greater confidence that better estimations are being made from the data. Secondly, it directly models the hazard and cumulative hazard functions rather than providing an estimation as the Cox model does. In the latter, the baseline hazard is estimated non-parametrically though the Kaplan-Meier estimation and therefore leads to 'steps' in the estimation. This may be more of a problem at later time points where the number of individuals at risk may be small. By modelling a smoothed hazard function, flexible parametric model can be used to model the baseline survival function directly.

3.4. Risk Prediction

Risk prediction models, in both the area this thesis covers and in most others, are dominated by the Cox model (60). All predictions are made in relation to the baseline risk. Importantly, 'baseline' does not refer to lowest risk, but as previously noted, when all variables in the model are zero in value. This may cause confusion in a number of situation. If we consider a model that predicts ESRD in adults with CKD that includes eGFR and age, then the baseline risk refers to the risk of a person aged 0 years, with 0 ml/min/1.73m^2 of kidney function. This imaginary person has essentially been born without any kidney function, and yet we are using them as our reference point for the rest of the model. Instead, values are often 'centred' before modelling takes place. This allows for more clinically relevant baseline risk prediction. The baseline risk might now refer to a 70 year old, with eGFR of 45 ml/min/1.73m^2. If we now try to predict risk in a 65 year old, the age part of the risk prediction now relates to predicting risk with a values of minus five years, or more intuitively, five years younger than the mean. The same is true of binary and categorical values where we select arbitrary categorisation of what zero represents. For instance,

being female could be the reference category and given a value of zero, so the model is now describing the risk of being male using female as a reference.

The total value of all these variables multiplied by their β co-efficients, or $x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3...$, is referred to by many terms in the literature, including 'prognostic index', but for this thesis will be referred to as the 'linear predictor' (LP). From using the value of LP and the baseline survival at time t we can then calculate predicted risk for individual i at time t based on the variables' values. t can be varied and baseline risk calculated at a different time, for instance two instead of five years to predict survival at this timepoint. Throughout survival has been predicted at time point t, if we wish to calculate risk then the survival value is subtracted from one to give this value at timepoint t.

Once predictions of risk have been made through modelling a new risk prediction equation, or from an existing model, statistical testing can be performed to see if these predictions match reality.

3.5. Assessment of the Validity of Risk Prediction Models

Validation of risk prediction models firstly aims to test the performance of the statistical model in separating individuals who reach the outcome from those who do not. Secondly, it assesses how well the predictions of the model, 'predicted risk', match what actually happens to individuals, 'observed risk'. These two aims are referred to respectively as 'discrimination' and 'calibration' of the model. Most risk prediction model assessment metrics were originally developed for binary outcomes associated with event sensitivity and specificity but have been adapted for the purposes of survival analysis (60). As the Cox model does not specify a baseline risk function and is often not presented in published risk prediction models, difficulties can occur in estimating event probabilities and hence externally validating risk prediction models.

3.5.1. Discrimination

Discrimination is the ability of the model to group individuals into those where the event has occurred and those where it has not. The optimum statistical method for assessing discrimination is an ongoing debate and different approaches may be preferred depending on the overall aim of assessing discrimination (60, 105). The most widely used metric for assessment of discrimination is the concordance-index, first developed for use in survival analysis by Harrell (106). Harrell's concordance-index is referred to in the literature by many names including 'C-statistic', 'Harrell's C' and 'C-index'. For consistency, throughout this thesis it will be referred to as the C-statistic.

3.5.1.1. The C-Statistic

The C-statistic is synonymous with the area under the Receiver Operating Characteristic, often simply referred to as a ROC curve where a model assesses a binary outcome with a fixed and known follow-up period. ROC curves plot '1 – (model specificity)' on the x-axis versus model 'sensitivity' on the y-axis. In simplified terms, the false positive rate versus the true positive rate at different cut-off points for the model.

However, in the case of survival analysis where a percentage prediction is produced by the model, rather than a binary outcome, it measures the proportion pf comparisons where the model predicts a higher percentage chance of an event in the true events cases versus the true non-event cases.

The C-statistic assesses all pairings of individuals and whether an event is experienced or not. This is calculated through:

$$C = ((E + T/2))/P \qquad (3.5)$$

Where, E = number of correct orderings of pairs, T = number of tied predictions, P = number of pairs.

The C-statistic is limited to a maximum of one, as the number of correctly ordered pairings (E) can never be greater than the total number (P). Its minimum value is zero, where all rankings are incorrect. However, if ranking of pairs was randomly allocated, say by tossing a coin, then it would be expected that half of the time rankings would be correct i.e. when the C-statistic equals 0.50. Therefore, in order for a risk prediction model to be of use in aiding decision making, clinical or otherwise, the value must be greater than 0.50. In practice during the thesis, this means the lower 95% confidence interval being greater than 0.50, so that we can reject the null hypothesis that the risk prediction model is of no benefit in risk prediction of CV events in CKD. However, whilst this would confirm a statistical effect, whether it is clinically useful will depend on the nature of the disease and how the clinical risk prediction tool might change decision making.

The following example and calculations demonstrate the calculation of the C-statistic using a fictitious cohort of five individuals shown in the Table 3.1.

Identity	Predicted Risk	Predicted Rank	Observed Event	E	T	P
U	50%	1	Yes	3		3
V	40%	2	No	-	-	-
X	20%	=4	Yes	1	1	3
Y	20%	=4	No	-	-	-
Z	10%	5	No	-	-	-
				Totals		
				4	**1**	**6**

Table 3.1: Example of the how to calculate the C-statistic from an example cohort. Abbreviations are as set out in formula for C-statistic above.

The two individuals who experience events (U, X) are compared with those who do not experience events (V, Y, Z). This results in a total of six pairings. U has the highest predicted risk so is ranked first; so E=3 for U. X is only ranked higher than Z, is tied with Y and outranked by V; so for X E=1 and T=1. Base on the totals and using the formula:

$$C = ((4 + 1/2))/6 = 0.75 \tag{3.6}$$

Therefore, in simple terms it suggests that the predicted risk in this example correctly ranks the individual who experiences an event above those who do not experience an event in 75% of pairings. The C-statistic therefore gives a measure of the *overall* performance of a model. However, for clinical risk prediction the patient, or the clinician, is unlikely to want to know if their predicted risk is likely to correctly rank them at a higher or lower risk than their the previous patient who has been seen in clinic; the patient wants to know their *individual* predicted risk and how that compares to real, observed risk.

3.5.2. Calibration

Calibration compares predicted outcomes with observed outcomes. It can be assessed through a number of methods, and as with discrimination, no method is universally agreed on as optimum. For the purposes of this thesis, three methods will be assessed, one, 'observed versus predicted event plots' ('OE plots'), provides a specific assessment of calibration based on risk groups. The other two, 'calibration slope' and 'expected event to observed event ratio' provide overall information about calibration and are more easily explained once OE plots are understood.

3.5.2.1.   Observed versus Predicted Event Plots

OE plots show predicted risk from the risk model on the x-axis and observed risk on the y-axis. An example is given in Figure 3.1. A reference line is shown at 45°, purple line in Figure 3.1, to show where predicted risk is equal to observed risk i.e. where groups for a perfectly calibrated model should lie on the graph.

If a group lies away from the reference line then it can be said to be miscalibrated. If it lies above the reference line, then observed risk is greater than predicted risk and therefore risk has been underestimated, this is shown in Figure 3.2 as a red triangle. If it is below the reference line then the opposite is true and risk is overpredicted, green triangle in Figure 3.2.
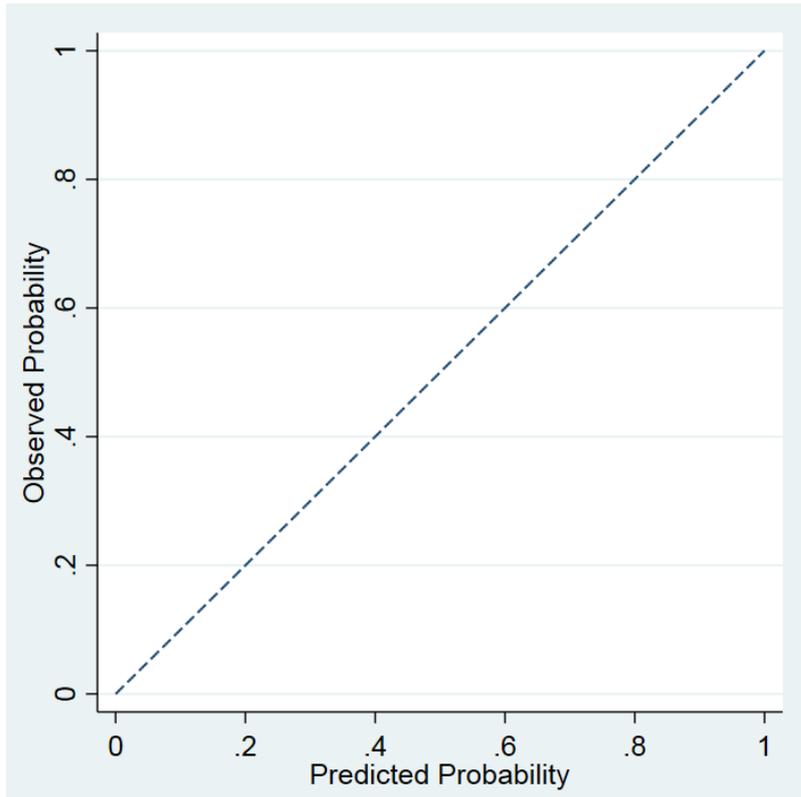
Figure 3.1: An OE plot showing the 45° reference line of 'perfect' calibration.
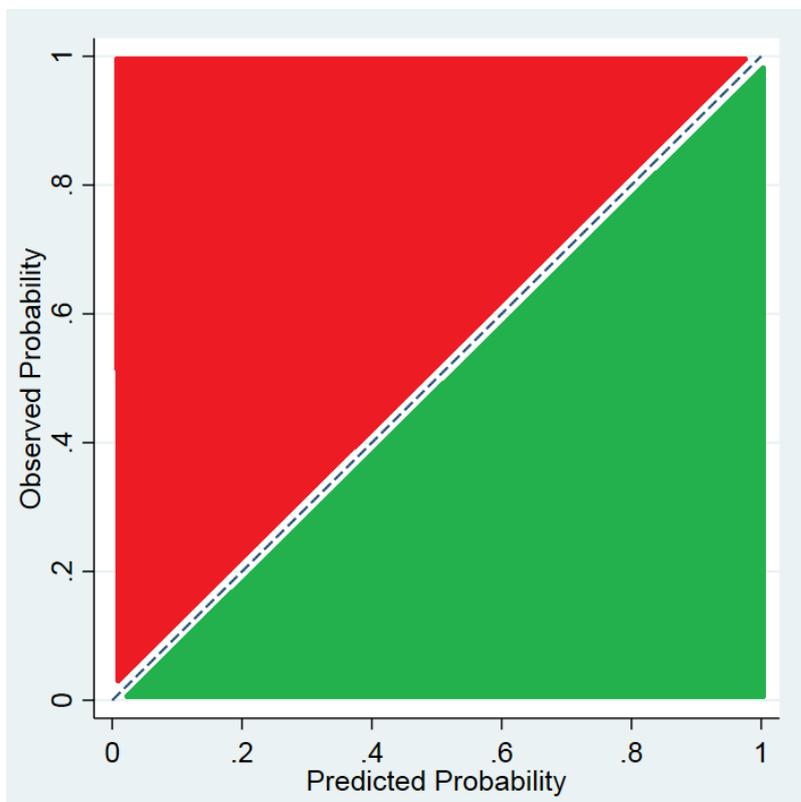
Figure 3.2: An OE plot showing areas of underprediction of risk (red triangle) and overprediction of risk (green triangle).

The vertical distance of the group to the reference line represents the amount by which the risk score is miscalibrated. The distance can vary between risk groups i.e. across the spectrum of risk. Risk can be underpredicted in lower risk groups and over predicted in higher risk groups, as is shown in Figure 3.3.
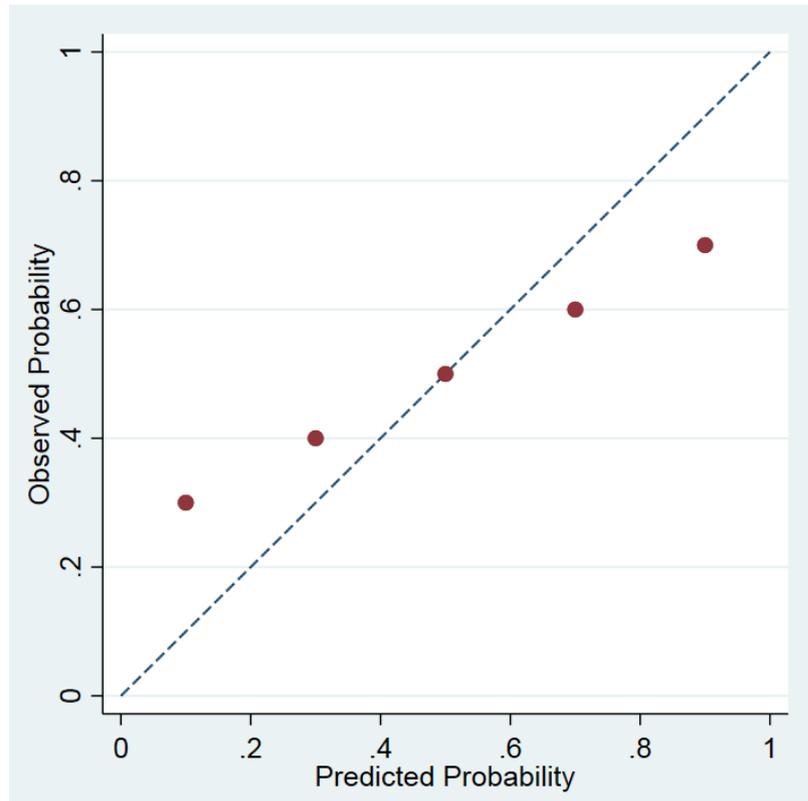
Figure 3.3: An OE plot showing underprediction of risk in lower risk groups and overprediction of risk in higher risk groups.

The optimum number of risk groups and on what basis to use to create them is unclear from the literature (60). This can potentially lead to the visual appearance of calibration, or miscalibration, in an OE plots being subject to how the groups have been defined. Therefore for the thesis, risk group sensitivity analysis was performed. The main results for OE plots are shown with groups split into declines based on predicted risk. An alternative to grouping is to use smoothing techniques to produce a 'best fit' result of calibration. As observed outcomes in this thesis are binary, this could be performed using a Loess algorithm. However, whether this is appropriate in the context of risk prediction is unclear. Loess based plots can be formally assessed through the Hosmer-Lemeshow X^2 and through other methods by Nam and D'Agostino. These metrics provide overall 'calibration-in-the-large' for the model.

3.5.2.2. Calibration Slope

A calibration slope assesses overall agreement between observed and expected events. A slope of one suggests that predicted risk in the model is equal to observed risk. A slope less than one suggests that predictions closer to an observed risk nearing one are too high and vice versa as prediction approach an observed risk of zero. A slope of greater than one indicates that the risk is narrower than predicted. As it produces a single, summary figure for the gradient, miscalibration in certain parts of the risk spectrum, typically at extreme ends may not be detected. For example, Figure 3.4 shows an OE plot of a well calibrated model with a calibration likely to be close to one, but it clearly is poorly calibrated in the lowest and highest deciles of risk.
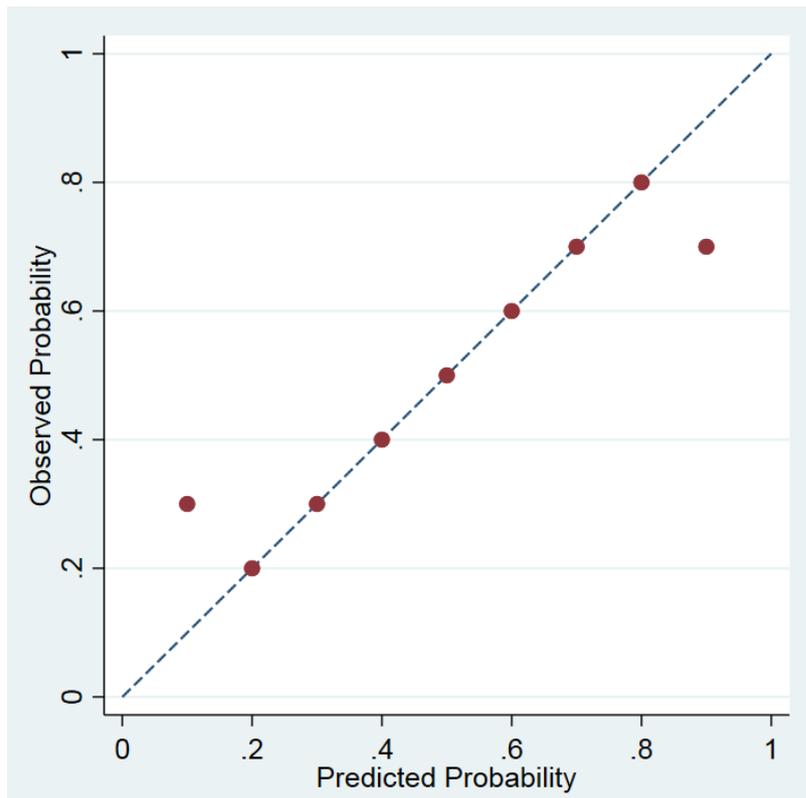
Figure 3.4: An OE plot showing a well calibrated model for central deciles, but poor calibration in the highest and lowest deciles. The calibration slope and E:O results for this example would likely suggest excellent, generally calibration.

3.5.2.3.   Expected Event to Observed Event Ratio

The ratio of expected to observed events ('E:O') gives an overview of systematic calibration of the model. If the ratio is greater than one, i.e. expected events are greater than observed events, then overall risk is overestimated by the model. If less than one then risk is underestimated. Again, miscalibration may still be present even if the E:O is close to one. Figures 3.3 and 3.4 would both likely have an E:O close to one, but as previously discussed are not well calibrated at the extreme risk groups.

3.5.3. Reclassification Indices

Risk prediction tools may be used to split individuals into strata to inform clinical decisions. An examples would be the suggestion of the use of a greater than 10% ten year risk of CV disease being used to counselling a patient on starting

a statin medication. Risk prediction models can therefore also be assessed to see how accurately they make these reclassifications. There are various indices proposed to evaluate reclassification with perhaps the most intuitive being a simple table of predicted strata before and after adjustments. The net reclassification index and integrated discrimination improvement are perhaps the most commonly used in clinical biomarker research (105).

Net reclassification index assesses how many individuals with and without an event are reclassified to either high or lower risk strata (105). It then compares the performance between the original risk prediction model and the new model to see if classification has been improved overall. Importantly though, the net reclassification index is subject to potential bias unless the strata are predefined, typically using clinical relevant thresholds as described for the statin example.

The integrated discrimination improvement metric considers the benefit of the new model based on the risk as a continuous variable as opposed to categorical reclassification (105). It is calculated as follows:

$$IDI_{events} = \frac{\sum p(events_{new})}{n_{events}} - \frac{\sum p(events_{ref})}{n_{events}}$$

$$IDI_{nonevents} = \frac{\sum p(events_{ref})}{n_{nonevents}} - \frac{\sum p(events_{new})}{n_{nonevents}}$$

$$IDI = IDI_{events} - IDI_{nonevents}$$

Where, IDI = integrated discrimination improvement, new = updated risk predication model, ref = reference risk prediction model.

The advantage of this approach is that an updated risk prediction model is credited more for higher increases in risk prediction in those who have an event. Equally, by reducing risk prediction in those who do not go onto to have an event the metric result further improves. The opposite is also true for the

magnitude of the penalty given to an updated model for incorrect predictions. The value of the integrated discrimination improvement for a model will be dependent on the event rate in the cohort and hence may be difficult to interpret, particularly when making comparisons between different samples (105).

3.6.   Statistical Analysis Plan for External Validation

All models were assessed for potential external validation. Risk prediction models were eligible for assessment if similar CV outcomes were presented in the model as were available in the LCC cohort, i.e. acute coronary syndrome and cerebrovascular disease events. If CV mortality was the only risk prediction model outcome then the model was not assessed. Variables in the risk prediction model also had to be available in the validation dataset. For models eligible on these criteria, core descriptors for the model, including model metrics were collated. An overall assessment of the 'Level' of information published was also completed (107). Level one information refers to the publication of regression coefficients only. Level two includes level one plus risk groups and associated Kaplan-Meier plots. Level three includes the previous information levels plus the baseline hazard or survival function. The different information levels allow different model metrics to be calculated from the validation dataset as further detailed in Table 3.2. Only level three data allows for full assessment of calibration of the model. Where data were not published for the identified models, authors were contacted via email for further information.

Method	Aspect Assessed	Information Level		
		1	2	3
Measures of Discrimination	D	●	●	●
Kaplan-Meier curves for risk groups	D,C		●	●
Logrank/Cox tests between risk groups	D		●	●
Hazard ratios across risk groups	D		●	●
Calibration	C			●

Table 3.2: Level of Information Required to Perform External Validation of Risk Prediction Models. D – discrimination, C – calibration. Edited version of Table 3 from 'Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC medical research methodology. 2013 Dec;13(1):33'.

Multiple imputation of missing values of variables in models was performed as previously described. Imputed values were then used to calculate predicted risk before assessment of models was performed. The mean of C across the imputed datasets was used to assess general discrimination for all models. Calibration was assessed where level 3 data was presented by calculating the expected versus observed event ratio and using OE plots. The mean values across imputed datasets were again used for these metrics. All risk calculation were based on the full baseline risk reported for a model, in some cases up to 15 decimal places, but for simplicity in the text are referred to using four decimal places.

Firstly, all assessments were made using every individual in the cohort of the LCC cohort ('Whole Cohort'). Secondly, a subgroup of LCC was tested that matched the inclusion and exclusion criteria, and outcomes of the cohort used in the risk prediction model's development ('Cohort Specific'). For example, if a risk score had been developed for primary prevention of CV disease then only individuals without a history of CV disease were included in the cohort specific model assessment.

3.7. Conclusions

Robust, transparent methods are an integral part of performing ethical high quality research. The description set out this chapter have explained the methods that will be used for the rest of this thesis. Missing data is a universal issue within all clinical research, and its frequency is higher in epidemiological studies using routinely collected information. Despite this, it often goes unacknowledged in published clinical studies. Two approaches were considered in this thesis, complete case analysis and MI. The latter involves the creation of datasets with missing values replaced by those from the distribution of the related available data. Statistical analysis is then performed for each dataset before amalgamation of results into a single result. The evidence based process that LCC underwent for its dataset imputation was also described. Again, this is rarely presented in clinical studies.

Survival analysis is the statistical process of describing the process of time until an event occurs. It includes a number of different functions, each of which describes a different element of the analysis including both cumulative events over time and the instantaneous risk an individual may be exposed to. The censoring of individuals, where individuals do not experience an event and do not complete follow-up, is the major difference in the analysis process for survival analysis. Until recently, the Cox model was almost omnipresent in its association with estimation of hazards associated with survival analysis. As will be described in coming chapters it is used to describe the probability attributed to different independent risk factors and to develop risk prediction models. Its two central tenets, proportional hazards and non-informative censoring, were also described.

Risk prediction models use information available at baseline to try to predict an outcome in the future. In the case of the Cox model, this use the baseline risk and the sum of the β co-efficients calculated from the model. The former relates to the predicted risk when the sum of the β values equals zero, and does not relate to the lowest risk as is sometimes assumed. The individual values for the β co-efficients are from the Cox model used to develop the risk prediction

model. This reliance of the Cox model and its associated assumptions will be evaluated and discussed in later chapters.

Once a model has been developed, its performance is assessed in two general categories, discrimination and calibration. These process are also used in the evaluation of models that have already been developed, external validation. Discrimination describes the ability of the model to correctly rank individuals' risk and for this thesis will be assessed through C, which can be thought of as the proportion of pairs of individuals whose ranking is correctly predicted by the model. Calibration describes whether predictions of risk probability match what actually occurs in reality. As with discrimination, there are a number of general metrics to assess this but as will be described in coming chapters this thesis has predominantly used OE plots. These graphically describe how group predicted risk matches to observed risk. They have some limitations but are probably more easily interpretable than the more global metrics and can identify specific spectrums of risk where miscalibration may have occurred.

The plan for external validation of risk prediction models was also described. The extent to which a model can be externally validated is reliant on the data presented in publication. Whilst most studies present β co-efficient values, or Level One data, full model specifying Level Three data are often lacking. Therefore, the plan for external validation will be dependent on the availability of appropriate data. Where appropriate though, the aim was to fully externally validate a model using the described methods for discrimination and calibration. As will be seen in coming chapters, this was not always possible.

3.8. Summary

This chapter has given an overview of the methods used to assess risk prediction models. It has also discussed why missing data needs to be assessed and how it will be approached for the remainder of the thesis. The assessment of models will be based on the level of data previously presented for the model, plus any additional data provided by authors. Discrimination, or the ability of the model to correctly separate individuals who have an event from those that do not, will be assessed primarily using Harrell's C-statistic. Where level 3 data, including the presentation of a baseline risk for the model, is available calibration will be assessed through the plotting of OE event graphs. Now that these methods have been explained, the cohort used to assess the risk prediction models will be introduced, including the baseline characteristics and the CV events that occurred during follow-up.

Chapter 4

4. The Leicester City and County CKD Cohort

This chapter describes the LCC cohort which was developed for this thesis and used to validate the risk prediction models identified. Firstly, an overview of the baseline characteristics of the cohort is presented. Secondly, missing data for baseline characteristics is assessed. Finally, CV outcomes across the primary and secondary care data sources are described and their concordance assessed.

The following presentations have been made in relation to this chapter:

- 'Prevalence of Co-Morbidities by Ethnicity in a UK Primary Care CKD Cohort', American Society of Nephrology 2017. – poster (108)

- 'Heart Attack and Stroke Risk Prediction in Kidney Disease', Kidney Research UK Fellows' Day 2018 – oral (109)

4.1. Introduction

The previous two chapters have set out the methods for the LCC cohort to ensure that a secure, ethical and high quality cohort was developed for this thesis. The increased CV risk associated with CKD is well established (2, 32), but most CKD cohorts included in these studies were based in secondary care and therefore may have different characteristics to primary care (44). Individuals in primary care are less likely to have CKD that is at risk of progression to ESRD, and the mainstay of their management will be in relation to CV risk.

The prevalence of CKD is well established through national studies of CKD as set out in Chapter 1 (25). Reproducibility of results, as will be described in the systematic review in Chapter 5, is also an important issue and therefore definitions of conditions, including CKD, require standardisation across studies to reduce heterogeneity. This was achieved through the use of CKD Prognosis Consortium definitions for identification of co-morbidities (44). As described in the missing data plan, where data were not available for biochemical measures it underwent assessment of its normality distribution, transformation where appropriate and then MI. Event identification is an important issue that may limit the findings of CV epidemiological studies, therefore event matching across data sources was assessed in detail.

This chapter therefore aims to describe the LCC cohort, with particular emphasis on how its development specifically addressed the central aim of the thesis in relation to validation of CV risk prediction models in CKD.

4.2.   Overview of LCC Cohort

One hundred and three practices were invited to participate in the study, 44 (42.7%) agreed to participate and the records of 277,248 registered patients were assessed for eligibility for the LCC cohort. Twenty-three (52.3%) of the 44 practices were recruited from Leicester City CCG and were classified as urban, the remainder were classified as rural. Individuals with a medical history at baseline of a kidney transplant or dialysis were excluded. 17,248 registered individuals (6.2%) had two or more EPI eGFRs <60 ml/min/1.73m^2 >90 days apart before the start of the follow-up period on 1st November 2011 and were included in the cohort. Individuals in the cohort therefore had prevalent CKD and no new individuals were added to the cohort after the beginning of the follow-up period, even if the inclusion criteria were met.
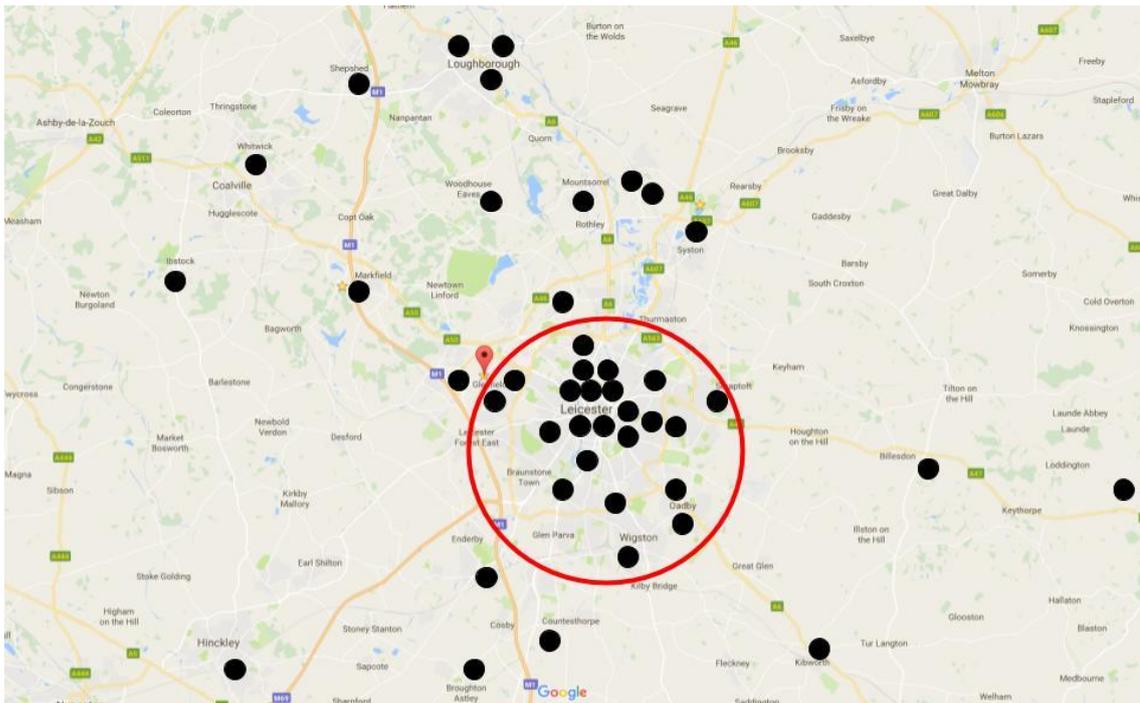
Figure 4.2: Graphical distribution of practices recruited to the LCC Cohort. Each black dot represents the approximate geographical position of a recruited practice. Practice within the red circle were those recruited from Leicester City CCG.

Prevalence of CKD increased substantial with age reaching more than 50% in those over the age of 80 years. Table 4.1 describes prevalence for the LCC cohort.

Age Group	Whole Cohort		
	Population (n)	CKD (n)	Prevalence (%)
18-29	53,112	7	0.01%
30-39	50,431	32	0.06%
40-49	48,734	173	0.35%
50-59	47,076	616	1.31%
60-69	37,882	2,609	6.89%
70-79	24,906	5,941	23.85%
80+	15,107	7,870	52.10%
Total	**277,248**	**17,248**	**6.22%**

Table 4.1: Prevalence of CKD by Age Group in LCC Cohort.

Table 4.2 describes the full baseline descriptors of the LCC cohort, all definitions for co-morbidities are listed in Appendix 4.1 and were consistent with those used by the CKD Prognosis Consortium (44). Ethnicity code was available for 14,883 (86.3%) individuals. 12,560 (72.8%) individuals were White and 1,928 (11.2%) were South Asian. Prevalence of CKD was 8.0% in White individuals and increased with age. Mean eGFR at baseline was 48.5 ml/min/1.73m^2 (SD 9.9) with median eGFR 52 ml/min/1.73m^2 (IQR 43 to 56). Median ACR was 2.9 mg/mmol (IQR 0.8 to 7.3). KDIGO CKD stage data are shown in Table 4.3. The majority of the cohort consisted of early stages of CKD, with 95% of the cohort having CKD stage 3a or 3b, and approximately two-thirds having either grade A1 ACR staging or no assessment of proteinuria.

4.3. Missing Data

The level of missing data was assessed for all variables in the cohort. As inclusion criteria for the cohort was based on EPI eGFR, which relies on age, gender, serum creatinine and black ethnicity, data were complete for the first three of these four variables. An individual deprivation score was available for 16,885 (97.9%) individuals. A practice based deprivation score was calculated

for the remaining 362 individuals. Missing data were present for all other variables including all blood results, with the exception of serum creatinine as described above, urine protein measurements and blood pressure measurements. The amount of missing data for each variable is described in Table 4.2.

Variable	Whole Cohort		Missing Data	
	Mean*/n	SD/IQR/%	n	%
Age	77.4	10.0	0	0.0
Female	10,353	60.0%	0	0.0
White Ethnicity	12,560	72.8%	4,687	13.7
ACR, mg/mmol	10.6	35.4	5,614	32.6
Median ACR, mg/mmol	2.9	0.8 to 7.3	5,614	32.6
EPI eGFR, ml/min/1.73m^2	48.5	9.9	0	0.0
Median EPI eGFR, ml/min/1.73m^2	52	43 to 56	0	0.0
Diabetes Mellitus	5,592	32.4%	-	-
Hypertension	15,674	90.9%	-	-
Systolic BP, mmHg	134.8	16.4	195	1.1
Diastolic BP, mmHg	74.0	9.9	104	0.6
History of Ischaemic Heart Disease	4,384	25.4%	-	-
History of Heart Failure	1,690	9.8%	-	-
History of Cerebrovascular Disease	2,127	12.3%	-	-
Albumin, g/L	43.0	3.0	535	3.1
Bicarbonate, mmol/L	26.5	3.8	16,244	94.2
Calcium, mmol/L	2.32	0.11	9,649	55.9
HbA1c, %	7.3	1.3	264	4.7
Haemoglobin, g/L	129.1	15.5	2,010	11.7
HDL, mmol/L	1.4	0.4	1,214	7.0
LDL, mmol/L	2.4	0.9	2,009	11.6
Phosphate, mmol/L	1.09	0.20	7,358	42.7
TC, mmol/L	4.6	1.1	763	4.4
Mean Deprivation Decile	5.7	3.0	362	2.1
Median Deprivation Decile	6	3 to 8	362	2.1
Statin	11,023	63.9	-	-
ACE inhibitor or ARB	10,928	63.4	-	-
Aspirin	8,044	46.6	-	-

Table 4.2: Baseline characteristics of the LCC cohort stratified by ethnicity. *For continuous variables, values are for mean and standard deviation (SD) unless otherwise stated. "Missing value refers to number of individuals with diabetes without known type of diabetes. HbA1c refers to individuals with known diabetes mellitus only. ACE – angiotensin converting enzyme, ARB – angiotensin II receptor blocker, BP – blood pressure, HDL – high density lipoprotein cholesterol, LDL – low density lipoprotein cholesterol, TC – total cholesterol.

	ACR Stage				
EPI Stage	**Missing**	**1**	**2**	**3**	**Total**
3a	4,657	4,266	3,169	270	12,362
%	27.0	24.7	18.4	1.6	71.7
3b	801	1,386	1,486	250	3,923
%	4.6	8.0	8.6	1.5	22.7
4	140	189	367	173	869
%	0.8	1.1	2.1	1.0	5.0
5	17	5	27	45	94
%	0.1	0.03	0.2	0.3	0.5
Total	5,615	5,846	5,049	738	17,248
%	32.6	33.9	29.3	4.3	

Table 4.3 KDIGO CKD stages based on categorisation of EPI eGFR and ACR.

4.3.1. Assessment of Distribution of Missing Values

Figure 4.1 shows histograms, with overlayed normal distribution plots, for serum albumin, total cholesterol, high density lipoprotein cholesterol (HDL), low density lipoprotein cholesterol (LDL), haemoglobin, phosphate, calcium, glycated haemoglobin (HbA1c), urine ACR and systolic and diastolic blood pressure. Figures 4.2 and 4.3 show standardised normal probability ('P-P') and quantile ('Q-Q') plots respectively for these variables.

These figures suggested that ACR was the variable that visually deviated most from a normal distribution, with a positive skew (median 3 mg/mmol, mean 10.2 mg/mmol). The P-P plots suggested that HbA1c may centrally deviate from normal, although to a lesser degree than ACR. All other variables, based on the P-P plots appeared to follow normality centrally. The Q-Q plots showed peripheral deviation from the normal distribution, to varying degrees, across all variables. D'Agnostino's χ^2, both with and without Royston's adjustment, suggested that the null hypothesis of a normal distribution should be rejected for all variables assessed (p<0.01 for all variables) (110). These statistical results may relate to the large sample size, meaning that even a minor deviation from normality would result in a 'significant' test result.

All variables were assessed for evidence of semi-continuity. Albumin had the single most frequent unique value across all of the variables. 2,288 individuals, 14.7% of available values, had a value of 43 g/L. ACR was the only variable which included a value of zero and this applied to only 47 individuals (0.43%). Therefore, approaches for MI in semi-continuous variables were not considered further for the current analysis. Due to evidence of their non-normal distribution, the values for ACR and HbA1c were log transformed, and reassessed for normality. Transformation was performed using both a simple log transformation and the Stata 'lnskew0' command to create a new variable with 'zero skew' (111). The latter produces this through taking the log values of the original value minus a fixed values 'k', where k is set to minimise the skew of the new variable. Results for the transformed variables are shown in Figures 4.4, 4.5 and 4.6. Transformation of values by either method suggested a normal

distribution for central values and higher values. The zero skew method suggested an improved proximity to normality at lower values.
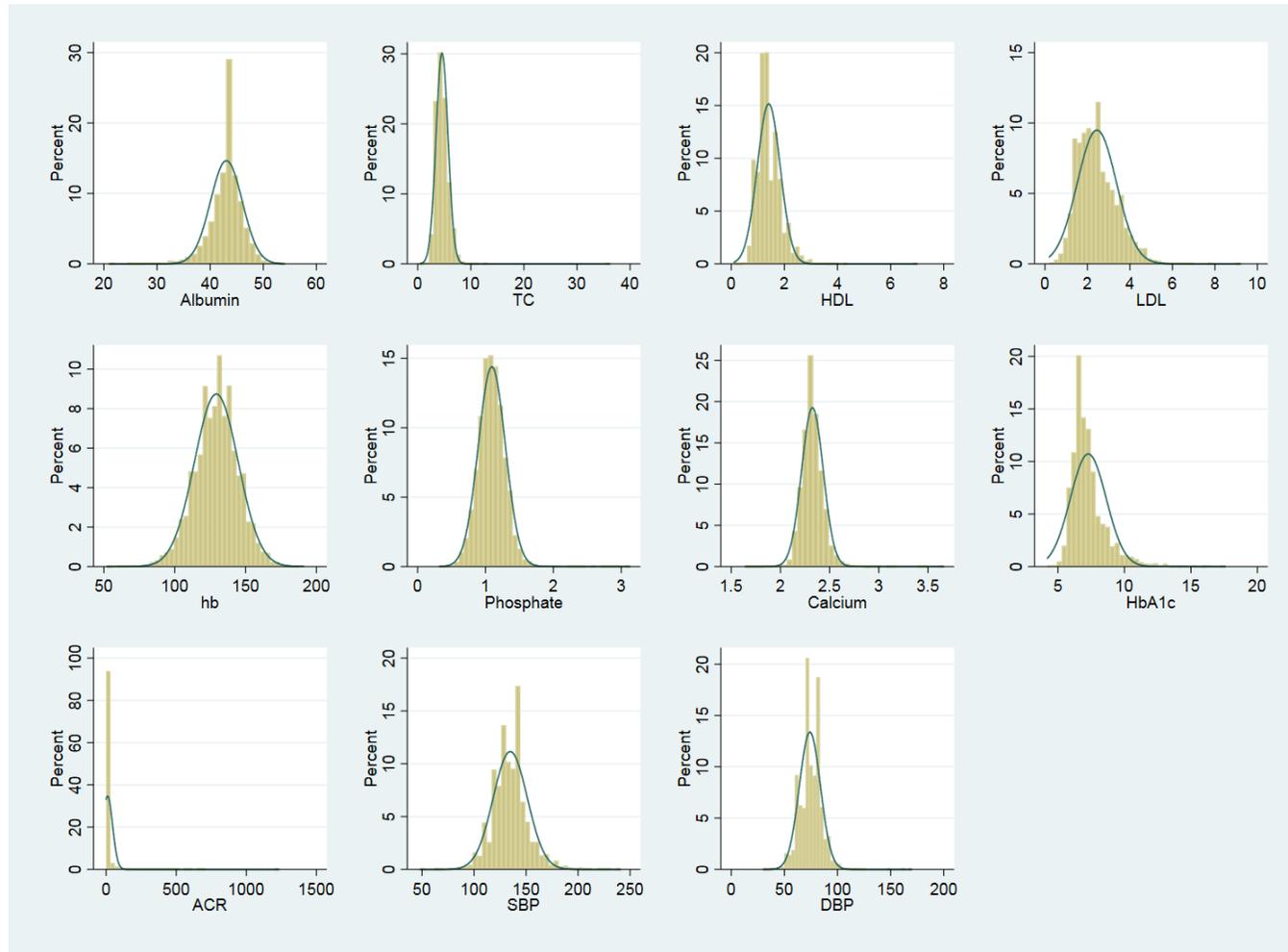
Figure 4.1: Histograms of variables with overlayed normal distributions. TC – total cholesterol, HDL – high density lipoprotein cholesterol, LDL – low density lipoprotein cholesterol, Hb – haemoglobin, HbA1c – percentage glycosylated haemoglobin, ACR – urine albumin creatinine ratio, SBP – systolic blood pressure, DBP – diastolic blood pressure.
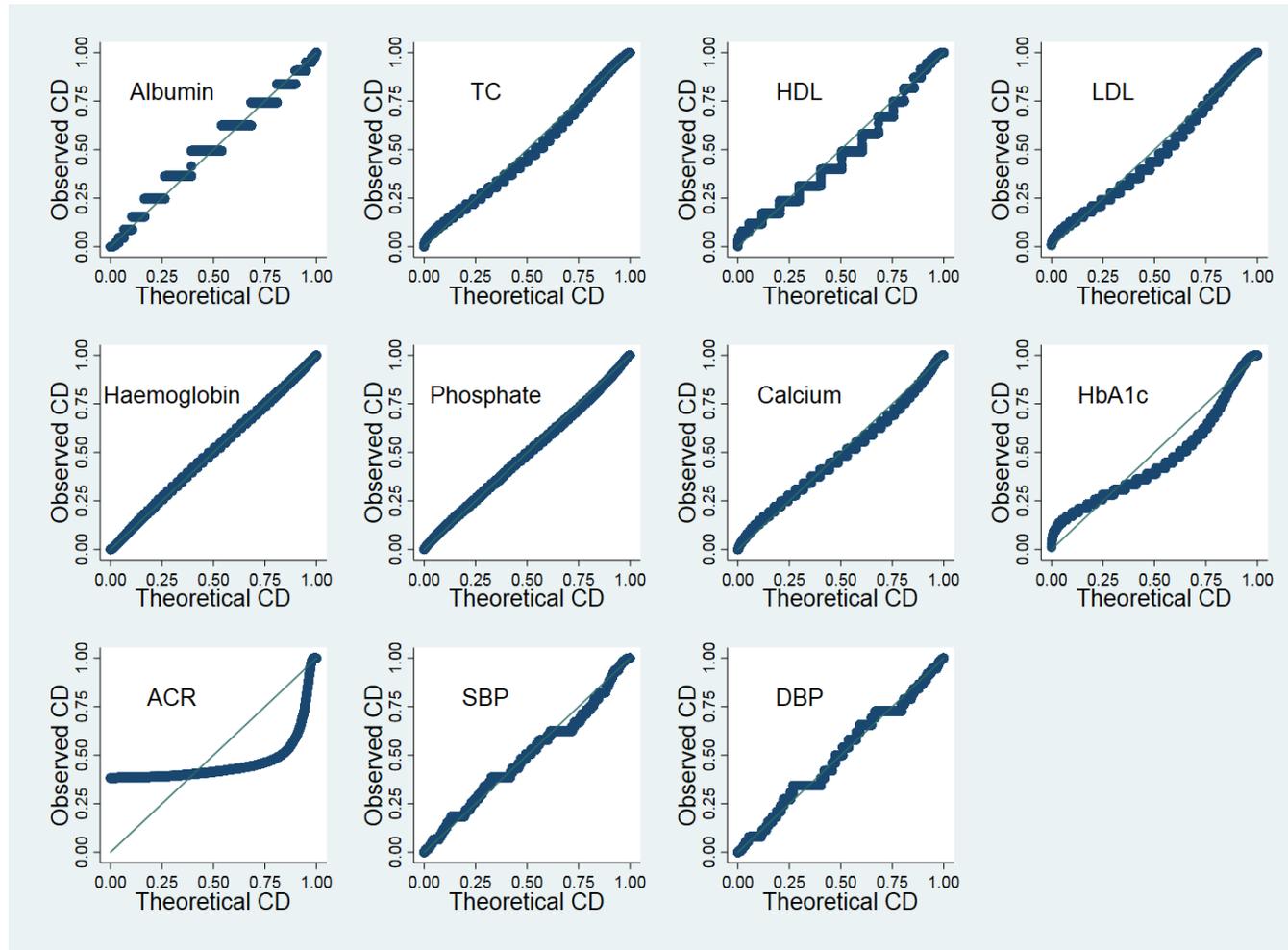
Figure 4.2: Standardised normal probability ('P-P') for the variables. CD – cumulative distribution, TC – total cholesterol, HDL – high density lipoprotein cholesterol, LDL – low density lipoprotein cholesterol, Hb – haemoglobin, HbA1c – percentage glycosylated haemoglobin, ACR – urine albumin creatinine ratio, SBP – systolic blood pressure, DBP – diastolic blood pressure.
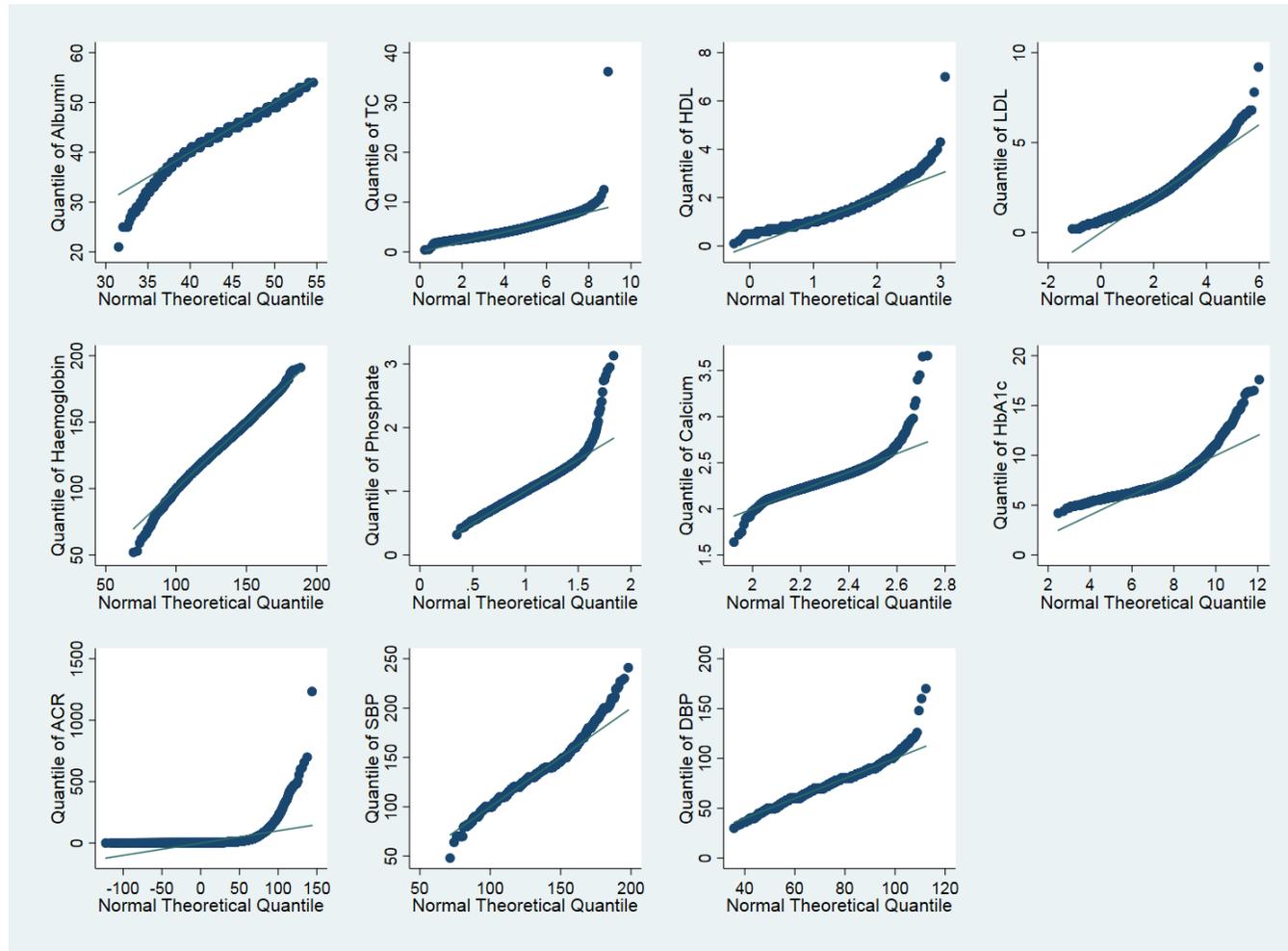
Figure 4.3: Quantile ('Q-Q') plots for the variables. TC – total cholesterol, HDL – high density lipoprotein cholesterol, LDL – low density lipoprotein cholesterol, Hb – haemoglobin, HbA1c – percentage glycosylated haemoglobin, ACR – urine albumin creatinine ratio, SBP – systolic blood pressure, DBP – diastolic blood pressure.
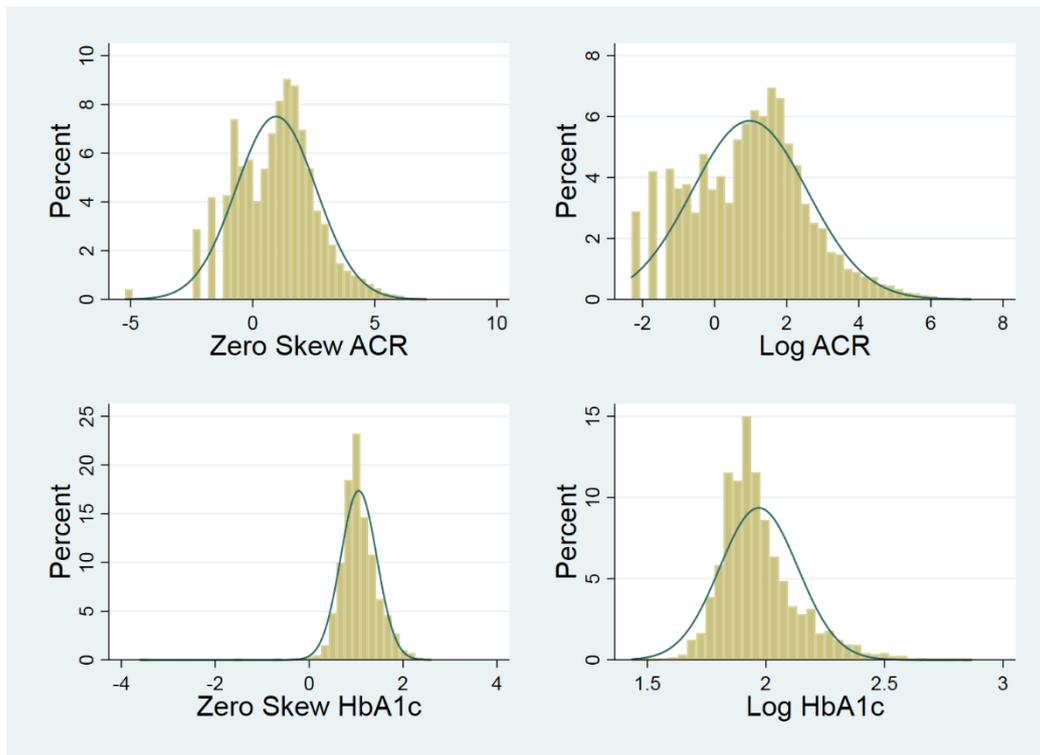
Figure 4.4: Histograms, with overlayed normal distributions, of transformed ACR and HbA1c values. HbA1c – percentage glycosylated haemoglobin, ACR – urine albumin creatinine ratio.
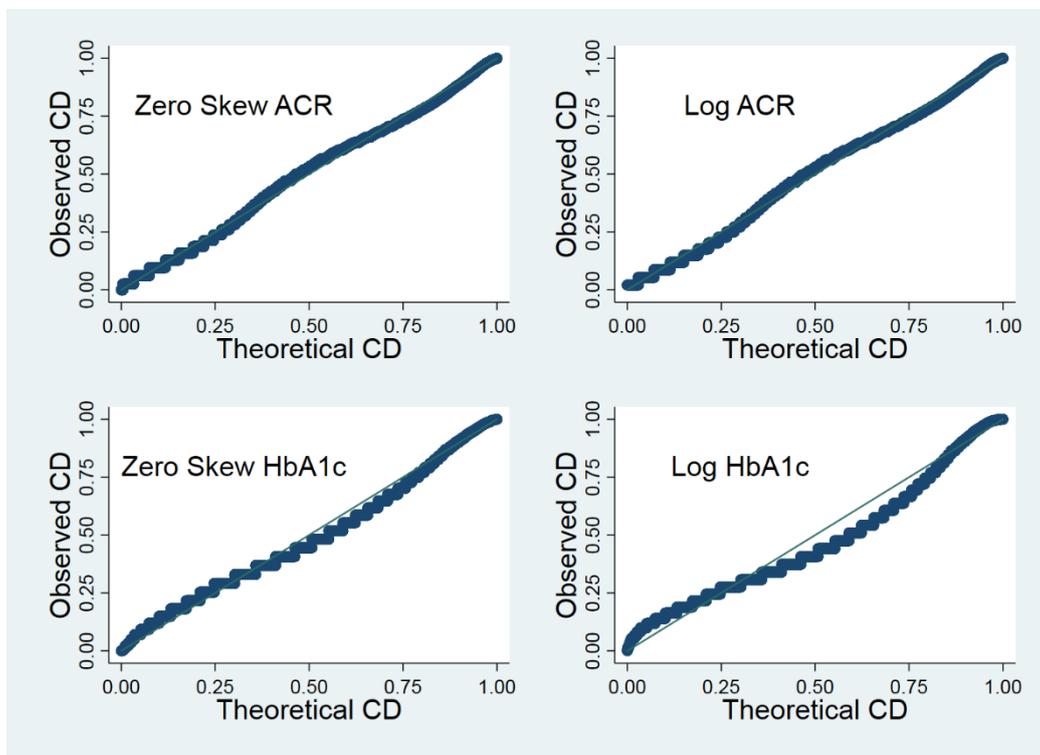
Figure 4.5: P-P plots of transformed ACR and HbA1c values. CD – cumulative distribution, HbA1c – percentage glycosylated haemoglobin, ACR – urine albumin creatinine ratio.
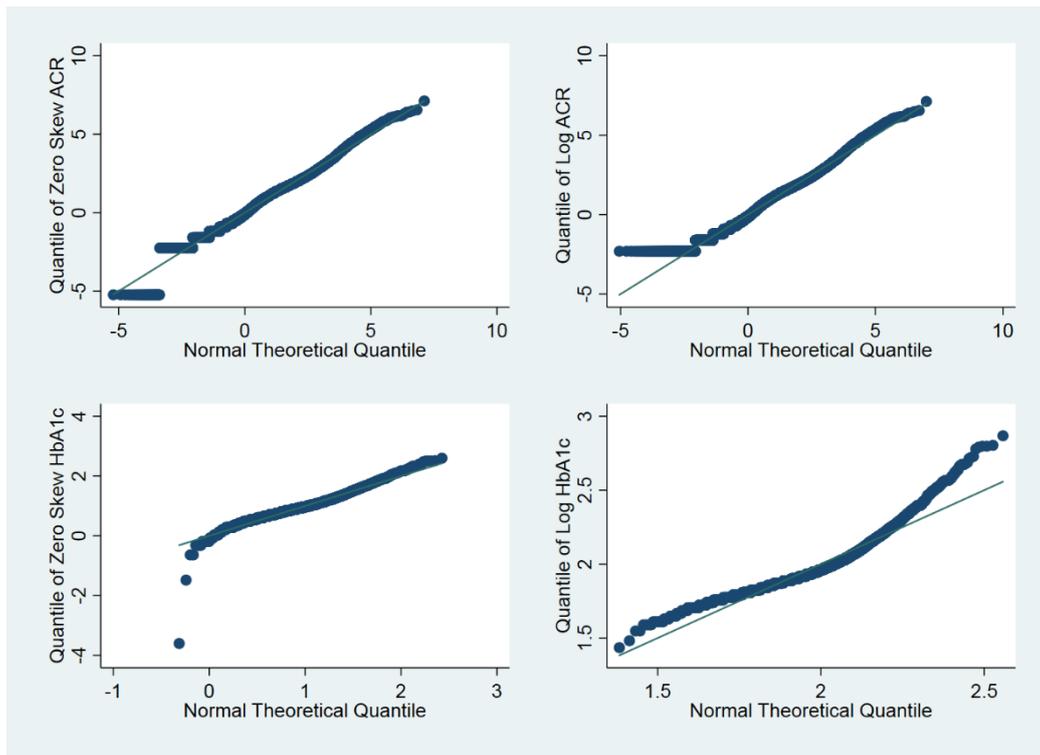
Figure 4.6: Q-Q plot of transformed ACR and HbA1c. HbA1c – percentage glycosylated haemoglobin, ACR – urine albumin creatinine ratio.

4.4. Outcomes for LCC

During a median follow-up of 5.0 years (IQR 3.3 to 5.0), 5,109 (29.6%) deaths and 2,072 (12.0%) CV events occurred. Mean follow-up was 4.0 years (SD 1.5 years). 8,752 (54.5%) individuals completed the full follow-up period without having a CV event or being censored and therefore had follow-up of 5.0 years. Table 4.4 summarises the outcomes. 1,035 out of 11,410 (9.1%) individuals with no known ischaemic heart disease or cerebrovascular disease at baseline had a CV event during follow-up.

Cohort	Outcomes	n	%	Per 1000 PYs (95%CI)
Whole	Deaths	5,109	29.6	73.4 (71.4 to 75.4)
n=17,248	CV	2,072	12.0	31.1 (29.8 to 32.5)
Baseline CV Disease	Deaths	2,262	38.8	102.6 (98.5 to 107.0)
n=5,838	CV	1,037	17.8	50.8 (47.8 to 54.0)
No Baseline CV Disease	Deaths	2,847	25.0	59.8 (57.7 to 62.0)
n=11,410	CV	1,035	9.1	22.4 (21.1 to 23.8)

Table 4.4: Summary of Outcomes for the LCC Cohort. PY – person-years.

4.4.1. Background of Comparison of Cardiovascular Events between Data Sources

The occurrence of events and their agreement between data sources was compared. Herrett *et al,* has previously compared events identification between primary and secondary care resources for myocardial infarction (81). Using, Clinical Practice Research Datalink (CPRD) as the GP data source comparisons were made with HES, MINAP and ONS mortality data between 1st January 2003 and 31st March 2009. In their study, events were considered the same event if they occurred within 30 days of each other in a different data source. 31.0% of events occurred in all three data sources. LCC is not currently linked to ONS death registry data. Herrett *et al*'s data suggested a crude annual incidence per 100,000 individuals of 250.0 with ONS data and 242.6 without ONS data. Therefore, lack of this data would suggest an underestimate of

myocardial infarction events by 3.0%. The previously presented Figure 2.3 details further information in relation to event identification from this study.

LCC does not have linkage to national event data. Currently, there is no literature as to the proportion of individuals admitted to their local hospital with a confirmed myocardial infarction or stroke. LCC population is of an average older age and possibly less likely to be transitory within the country. In addition, the use of regional data including that from the regional tertiary cardiology centre at University Hospitals of Leicester, may reduce the risk of event identification loss.

4.4.2. Overview of Codes

CV disease includes any individual with a Read code diagnosis of previous ischaemic heart disease, stable angina, cerebrovascular accident, transient ischaemic attack or heart failure. Whilst no definite list of Read codes for CV event identification exists, the utilised lists of Read, International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) and MINAP codes were identified through previous publications in the area of CV event prediction (69, 81, 112-114). The full list of CV event codes used are listed in Appendix 4.2.

4.4.3. Cardiovascular Outcomes by Data Source

As previously stated, CV outcomes for the LCC cohort were identified from three sources, GP records (Read codes), HES and MINAP. The latter two were provided by University Hospital of Leicester and were based on CV related ICD-10 codes. These codes are also listed in Appendix 4.2. All outcome codes were reviewed and categorised. Duplicates coded events for the same individual on the same date were removed from the dataset. For ischaemic heart disease related events categorisation occurred into groups as:

1. Myocardial infarction

2. Unstable angina

3. Stable angina

4. Unknown type of ischaemic heart disease (IHD) event

For cerebrovascular events, categorisation occurred as:

1. Ischaemic stroke

2. Haemorrhagic stroke

3. Transient ischaemia attack

4. Unknown type of stroke event

For external validation of risk prediction models, outcomes were used that matched the outcome presented in the original published score. For presentation of the baseline data for the LCC cohort, the categories above are presented. Acute peripheral vascular disease events, coronary revascularisation and acute decompensated heart failure events were not included as CV outcomes for the cohort.

4.4.4. General Practice Identified Events

3,641 IHD related and 2,121 atherosclerotic events were coded in the GP records. 633 (3.7%) individuals had a myocardial infarction coded and 716 (4.2%) an atherosclerotic stroke. 39 (0.2%) individuals had both a myocardial infarction and stroke coded during follow-up, therefore 1,310 (7.6%) had a myocardial infarction or atherosclerotic stroke. The mean and median time to first myocardial infarction event was 2.3 years (SD 1.5) and 2.3 years (IQR 1.0 to 3.7) respectively. For time to first atherosclerotic event mean and median times were 2.2 years (SD 1.5) and 2.0 years (IQR 0.9 to 3.5). Table 4.5 describes these findings.

All Events	n	%
Myocardial Infarction	952	26.2
Unstable Angina	72	2.0
Stable Angina	892	24.5
Unknown IHD Type	1,725	47.4
Total IHD Events	**3,641**	
Ischaemic Stroke	1,300	61.3
Haemorrhagic Stroke	89	4.2
Transient Ischaemic Attack	436	20.6
Unknown Stroke Type	296	14.0
Total Stroke Events	**2,121**	
Individuals with Event		
Myocardial Infarction	633	30.7
Unstable Angina	57	2.8
Stable Angina	437	21.2
Unknown IHD Type	937	45.4
Individuals with any IHD Event	**1,760**	
Ischaemic Stroke	716	59.2
Haemorrhagic Stroke	45	3.7
Transient Ischaemic Attack	280	23.2
Unknown Stroke Type	168	13.9
Individuals with any Cerebrovascular Event	**1,209**	

Table 4.5: Total number of events (top half) and number of individuals (bottom half) by Ischemic Heart Disease (IHD) and stroke categories identified for the cohort form GP records.

4.4.5. Hospital Episode Statistics Events

9,731 IHD related and 2,281 atherosclerotic events were coded in HES records. 620 (3.6%) individuals had a myocardial infarction coded and 889 (5.2%) an atherosclerotic stroke. 79 (0.5%) individuals had both a myocardial infarction and stroke coded during follow-up, therefore 1,430 (8.3%) had a myocardial infarction or atherosclerotic stroke. The mean and median time to first myocardial infarction event was 839 days (SD 512) and 803 days (IQR 400 to 1254) respectively. For time to first ischaemic stroke event mean and median times were 969 days (SD 520) and 987 days (IQR 540 to 1411). These findings are displayed in Table 4.6.

All Events	n	%
Myocardial Infarction	714	7.3
Unstable Angina	168	1.7
Unknown IHD Type	8,849	90.9
Total IHD Events	**9,731**	
Ischaemic Stroke	1,056	46.3
Haemorrhagic Stroke	155	6.8
Unknown Stroke Type	1,070	46.9
Total Stroke Events	**2,281**	
Individuals with Event		
Myocardial Infarction	620	16.2
Unstable Angina	149	3.9
Unknown IHD Type	3,055	79.9
Individuals with IHD Event	**3,824**	
Ischaemic Stroke	889	49.8
Haemorrhagic Stroke	131	7.3
Unknown Stroke Type	768	42.9
Individuals with Stroke Event	**1,788**	

Table 4.6: Total number of events (top half) and number of individuals (bottom half) by ischaemic heart disease (IHD) and stroke categories identified for the cohort form HES records.

4.4.6. Myocardial Infarction National Audit Project Events

218 MINAP events were coded for individuals in the cohort, of which 217 (99.5%) were myocardial infarction events. These events occurred in 208 different individuals in the cohort and represented 23.3% of all myocardial infarction events. Mean time to first myocardial infarction events was 863 days (SD 521) and median time was 867 days (IQR 389 to 1301). MINAP does not record events for stroke outcomes.

4.4.7. Comparison of Number of Myocardial Infarction Events

During the follow-up period, 893 myocardial infarction events were identified across the three data sources. 140 (15.7%) of events were identified in all three sources and 465 (52.1%) were identified in only one source. Between GP and HES records, 882 (98.8%) events were identified. Figure 4.7 displays the events in a Venn diagram.
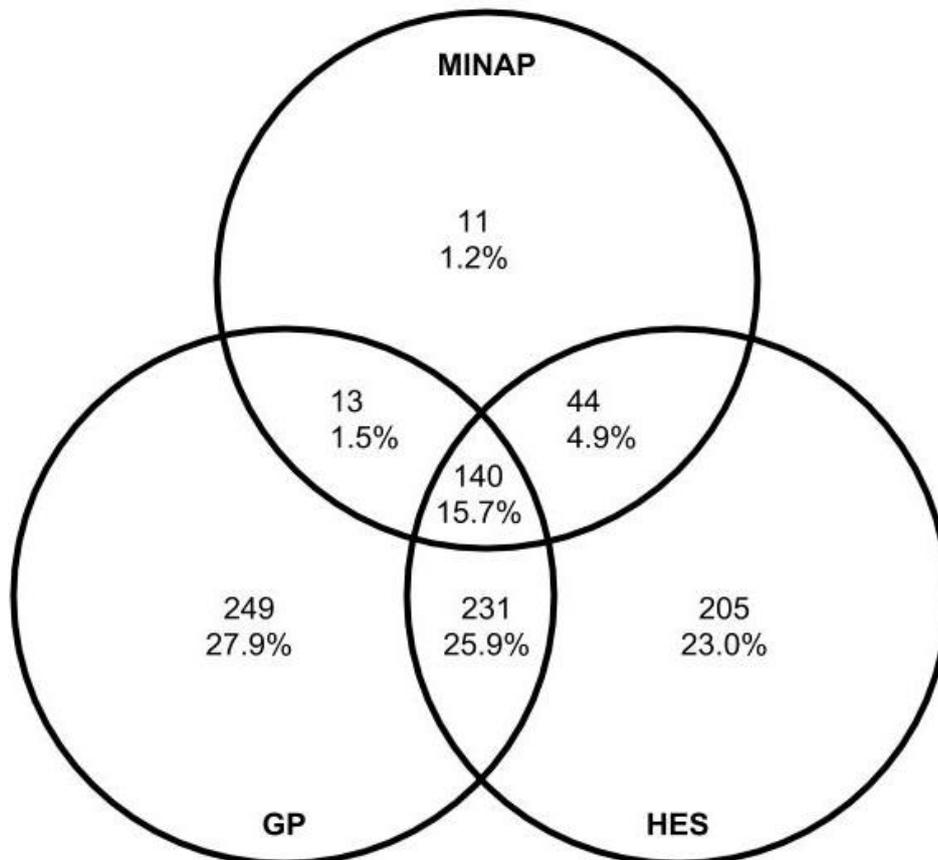
Figure 4.7: Venn diagram of the 893 myocardial infarction events identified across the three available data sources.

4.4.8. Comparison of Number of Stroke Events

1,308 stroke events were identified, of which 297 (22.7%) of events were identified by both GP and HES records. A Venn diagram of stroke events is shown in Figure 4.8.
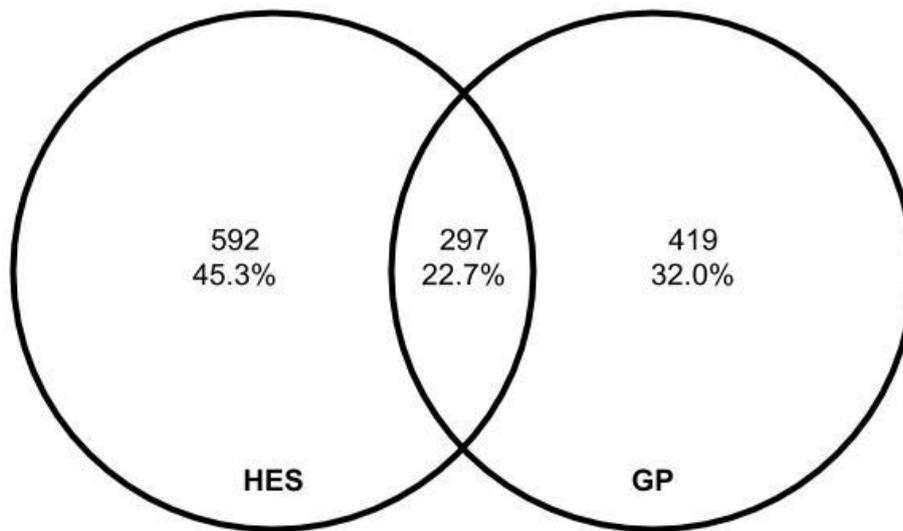
Figure 4.8: Venn diagram of the 1,308 stroke events identified in GP and HES records.

Differences in dates were similar between the different data sources for both myocardial infarction and stroke. For the former, the largest mean difference was 20 days between MINAP and GP events. For myocardial infarction, 583 out of 708 (82.3%) dates comparable across the three combinations of data sources were within 30 days of each other. For stroke, the mean difference was 49 days (SD 321) with a median of 13 days (IQR 0 to 38 days). 171 (59.8%) events' dates were within 30 days of each other. Table 4.7 details the differences between dates for myocardial infarction and stroke events between the different date sources.

	Myocardial Infarction			Stroke
	MINAP-HES	MINAP-GP	HES-GP	HES-GP
n	184	153	371	297
Mean difference	-2	20	11	49
SD	88	163	147	321
Median difference	-14	0	8	13
IQR	-22 to 0	-6 to 0	0 to 20	0 to 38
Within 30 days	155 (84.2%)	133 (86.9%)	295 (79.5%)	171 (59.8%)

Table 4.7: Difference between dates for myocardial infarction and stroke events between data sources.

4.4.9. Limitations of Event Identification

4.4.9.1.   Myocardial Infarction Event Identification

Herrett *et al*'s data suggested that 13,380 out 17,964 (74.5%) non-fatal myocardial infarction events were identified in GP records. Using this as a basis, this section aims to establish an estimation of the CV events that may not have been detected due to the LCC cohort having local, instead of national, HES and MINAP data. Table 4.7 summarises the results for this section.

893 myocardial infarction events occurred during follow-up in the LCC cohort, of these 633 were identified from GP records (70.9%). If it is assumed that the proportion of events identified in GP records is the same as in Herrett *el al* (81), then, to the nearest whole event, the number of expected events for GP, HES and MINAP would be:

$$633 \times \left(\frac{1}{0.745}\right) \approx 850 \qquad (4.1)$$

Therefore, potentially more events have been recorded in GP records than were expected. However, it is not possible that more events were recorded locally in HES and MINAP than nationally for the LCC cohort as the UHL data represents a subgroup of the national dataset. Nevertheless it is reassuring that a potential significant deficit has not been identified through this analysis.

In order to further test these assumptions the crude incidents for myocardial infarction based on the earlier described Figure 2.3 were tested. This was also used to test the potential number of events not identified through the absence of ONS data in LCC. Again, the assumption was made that the proportion of events identified in GP records from Herrett *et al* as a percentage of all events is the same proportion in LCC (81). Therefore, 186.7 out 250.0 (74.7%) were identified from GP records. Again, this is similar to the percentage (74.5%) identified form GP records in LCC. In relation to ONS data, from Figure 2.3 it is estimated that 7.4 out of 250.0 (3.0%) events were detected in this data source but not in the other three data sources. Therefore it might be expected with

ONS data included that, to the nearest whole myocardial infarction event, the total number of events would be:

$$893 \times \left(\frac{1}{0.970}\right) \approx 920 \qquad (4.1)$$

Therefore, Based on the 95% CI for crude event incidence the percentage of missed events would be 1.6% to 4.2%, or an estimated total of 908 to 933 events.

Event Identified	Herrett *et al*		LCC	
	n	%	n	%
All Myocardial Infarction (Herrett non-fatal)	17,964	-	893	-
GP Myocardial Infarction (Herrett non-fatal)	13,380	74.5	633	70.9
Expected Events*	-	-	850	-
Possible Excess Events	-	-	43	4.8
	Incidence			
All Myocardial Infarction	250.0	-	-	-
All Myocardial Infarction, except ONS	242.6	97.0	-	-
Expected Events with ONS*	-	-	920	3.0
GP Myocardial Infarction	186.7	74.7	-	-
Expected Events*	-	-	847	-
Possible Excess Events	-	-	46	5.2

Table 4.8: Expected Myocardial Events across Data Sources. 'Expected events' assume that the proportion of events identified in GP records only in LCC is the same as in Herrett *et al*. All calculations for number of events are rounded to the nearest whole event.

4.4.9.2. Stroke Event Identification

A similar approach was made to try to assess the identification of stroke events in the context of local versus national data. A similar study to Herrett *et al* is underway to assess accuracy of coding across databases for stroke events but results have not yet been published (115).

Therefore, to compare the LCC data to national data, an assumption was made that the distribution of events was similar to myocardial infarction events for GP and HES data. Table 4.8 shows the results of this extrapolation. As Table 4.8 shows, the distribution of events between data sources was more evenly distributed. Due to this, in the extrapolation, there was a possible 425 excess events identified using the LCC methods. The results for the ONS data estimation were limited by the lack of a crude incidence figure for GP and HES data from Herrett et al. If again it is assumed that 3.0% of stroke events were missing for stroke through a lack of ONS data then this would equate to 40 events.

Event Identified	Herrett *et al*		LCC	
	n	%	n	%
GP & HES Stroke (Herrett MI)	16,476	-	1,308	-
GP Stroke (Herrett non-fatal MI)	13,380	81.2	716	54.7
Expected Events*	-	-	882	-
Possible Excess Events	-	-	-426	32.6
	Incidence			
GP, HES & ONS MI	250.0	-	-	-
GP, HES MI	242.6	97.0	-	-
Expected Events with ONS*	-	-	1348	3.0
GP MI	186.7	74.7	-	-
Expected Events*	-	-	959	-
Possible Excess Events	-	-	349	26.7

Table 4.9: Expected Stroke Events across Data Sources. 'Expected events' assume that the proportion of events identified in GP records only in LCC is the same as in Herrett *et al*. All calculations for number of events are rounded to the nearest whole event. GP – general practice, HES – Hospital Episode Statistics, MI – myocardial infarction, ONS – Office of National Statistics

4.5. Summary of Event Identification

A summary flowchart of the number of myocardial infarction and atherosclerotic events and individuals identified from the three different data sources is presented in Figure 4.9.
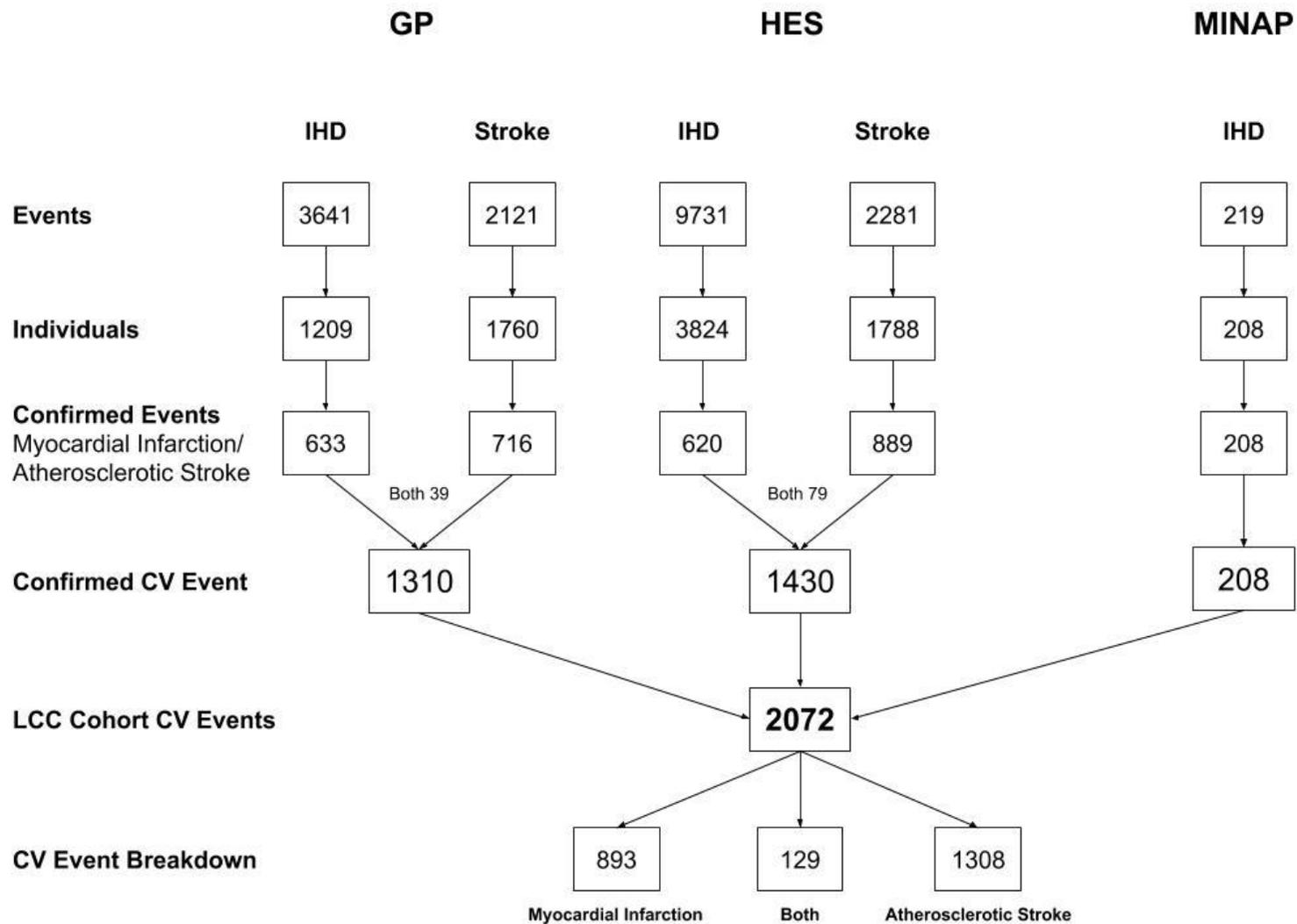
Figure 4.9: Summary flowchart of CV event identification from the three different data sources for the LCC cohort.

4.6. Conclusions

The LCC cohort is an anonymised primary care CKD based in the three Leicester GP Clinical Commissioning Groups with full ethical approval received from the Health Research Authority. It includes 17,248 individuals with the standard KDIGO CKD definition of CKD stage 3a or more advanced. Using event outcomes linked to University Hospitals of Leicester data for HES and MINAP, as well as GP events, 2,072 individuals with CV events were identified across a median and mean follow-up of five and four years respectively. Identification of co-morbidities were also based on CKD Prognosis Consortium definitions ensuring comparison and reproducibility with their results was possible (44).

The cohort consists of an older populations with early stage CKD, both in relation to eGFR and proteinuria, and a high prevalence of co-morbidities. Approximately three-quarters of the cohort were of white ethnicity and there was a wide range of deprivation levels. Proportions of missing data were generally low except for some biochemical data such as bicarbonate and calcium. Where missing data were present multiple imputation was possible with the transformation of two variables, HbA1c and ACR, required for non-normal distributions.

The use of multiple sources for event identification and the missing data methods are strengths of the LCC cohort. However, as with any study, there are a number of limitations. Whilst CKD Prognosis Consortium definitions were used, if the data had not been correctly analysed at source, i.e. in the GP practice, then issues such as incorrect coding may occur. This includes lack of standardisation of recorded diagnoses, perhaps even from clinician to clinician and measurements in the data, such as how blood pressure was recorded in the clinical setting.

Further, the well-established biases of systematically collected observational studies may be enhanced by the use of routinely collected data (116). This may particularly be the case in relation to confounding by indication, whereby the most unwell individuals attend more frequently to healthcare services and thus their data is more likely to feature in routinely collected data sources (117).

As analysed in section 4.4.9, the limitations of LCC being not linked to national CV event databases, but local ones, may be an unquantifiable confounder for the risk prediction model assessment. This analysis has suggested that 3.0% of events may not have been identified through the lack of ONS data (81). HES and MINAP events occurring in non-University Hospitals of Leicester probably did not contribute a significant number of events that were not detected in GP records. Limitations of data in this way has not been assessed in the literature. LCC GP outcome data was probably more complete than that assesse in Herrett *et* al and this may be related to changes in patterns of event recording in GP records. Herrett *et al* studied data between 2003 and 2009, whereas the LCC follow-up data was between 2011 and 2016 (81). During this time accurate coding of data became a more prominent issue for both GP and secondary care. Some of this was linked to financial remunerations linked to accurate coding of data and disease registries linked to the Quality Outcomes Framework in primary care (118).

In conclusion for this chapter, LCC represents a large primary CKD dataset developed specifically for this thesis. It represents a socioeconomically diverse population, mainly with early stage CKD and related co-morbidities. The LCC cohort most likely has a sufficient number of CV events, identified from both primary and secondary care sources, to allow robust external validation of CV risk prediction models in CKD to be performed (119).

4.7.　Summary

This chapter described the LCC cohort which has been developed for the specific purpose of validating CV risk prediction models in a CKD population. An overview of the characteristics of the 17,248 individuals is given and describes a population with earlier CKD stages, through both eGFR and proteinuria criteria. The majority of individuals were of seventy years of age or older and pre-existing co-morbidities were highly prevalent. Missing data for the cohort was described and the imputation process for the external validation of models was also explained. Over 2000 CV events occurred during a median follow-up of four years, which is more than sufficient for the external validation of risk prediction models (119). These events were recorded across the primary and secondary data sources and are probably of a similar distribution to those previously describe by Herrett *et al* (81). Now that the cohort has been described, the evidence base for CV risk factors within CKD is described in the next chapter by a systematic review and meta-analysis of the topic. This established the risk factors that might be considered in CKD risk prediction tools and possible variables to augment existing CV tools.

Chapter 5

5.   Cardiovascular Disease Risk Factors in Chronic Kidney Disease: a Systematic Review and Meta-analysis

This chapter summarises the rationale, methods, results and conclusions for the first systematic review for this thesis, 'Cardiovascular Disease Risk Factors in Chronic Kidney Disease: a Systematic Review and Meta-analysis'. Firstly, it identifies risk factors from studies with adjusted analysis that are associated with CV disease events in CKD. Secondly, it performs meta-analysis of these risk factors.

This chapter has been published as follows:

5.1. Introduction

The first step in the process of prognosis research, prior to development of clinical risk prediction models, is the identification of risk factors for a condition (61). As described in Chapter 1, a 'risk factor' refers to a measurable variable available before the beginning of a study that is associated with a future outcome of interest. Often the focus of published literature is in relation to 'novel' biomarkers, but it is important to remember that even variables recorded in routine clinical practice, such as blood pressure and cholesterol, remain important biomarker risk factors whose roles may be different in CKD than in the general population. Therefore, one of the aims of the chapter was to confirm the role of traditional routinely collected risk factors in CV risk prediction in CKD and then to identify non-traditional factors that may be related to the processes of arteriosclerosis, cardiomyopathy and inflammation as described in Chapter 1.

A new risk factor is only clinically useful if it adds predictive performance to a model beyond currently utilised standard risk factors, i.e. once a model has been adjusted for said factors, so additional risk factors must be novel and routinely collected in clinical care. Therefore, assessment of these factors is crucial before risk prediction models can be rationally optimised. Specific validation in CKD is warranted because the relative role of atherosclerosis in CV outcomes diminishes, and is replaced by non-traditional CV risk factors. These uraemia-related risk factors may have an increasingly important role with advancing CKD (33). This may warrant inclusion of risk factors such as calcium and phosphate (34) related to arteriosclerosis and reduced vascular compliance, in CKD-specific CV risk prediction models. Equally, consideration of risk factors associated with cardiomyopathy, such as echocardiographic evidence of left ventricular dysfunction or systemic inflammation may also be justified (33). Thus other novel routinely collected risk factors require consideration for validation of CV risk prediction models in CKD.

In summary, the aim of this systematic review was to identify routinely collected risk factors with potential value in CV risk prediction in CKD beyond those already included in existing CV risk prediction models. This new information would be used to inform the development of risk prediction models as will be

described in coming chapters. Even if a risk factor is established as being related to risk of an event, its role in improving risk prediction is not always straightforward. So, this chapter aimed to establish the existing literature for risk factors in CKD. However, this did not necessarily mean that when incorporated into a model guaranteed statistical or clinical improvement in risk prediction would be seen with these factors. Nevertheless, establishment of this evidence base could be used to guide the methods for future chapters.

5.2. Methods

5.2.1. Search Strategy

Ovid MEDLINE and Embase were searched using a pre-defined and International Prospective Register of Systematic Reviews (PROSPERO) registered systematic review and meta-analysis protocol, PROSPERO ID - 2016:CRD42016036187 (120). Search strategies for the systematic review are available in Appendices 5.1 and 5.2 of the thesis. The inclusion criteria were observational cohort studies and secondary analyses of randomised controlled trials in adult (≥18 years of age) with either CKD stage 3a or worse (any eGFR formula <60 ml/min/1.73m^2) or proteinuria based on standard definitions. The search was limited to English language manuscripts. General population studies with subgroup analysis presenting results for CKD groups were also included. Studies including individuals with ESRD, either receiving maintenance dialysis or with a renal transplant, were excluded. Studies of outcomes after acute kidney injury were also excluded. Pre-existing CV disease or the use of CV prevention medications such as aspirin or statins were not exclusion criteria. The minimum follow-period was six months. A formal definition of CKD using a standardised eGFR formula was first established in 1999 (10), therefore the search range was restricted from this date until 20th October 2017.

The primary outcome was a composite of CV disease events which included acute coronary syndrome, congestive cardiac failure and ischaemic stroke. Composite CV outcomes including CV-specific mortality were included unless CV events were grouped with all-cause mortality and/or renal related outcomes.

5.2.2. Assessment of Papers

The title and abstracts of all studies identified by the literature search were assessed. The full text of any abstract meeting the inclusion criteria was then reviewed. Data were extracted using a standardised extraction form which included a risk of bias assessment based on the 'Quality in Prognostic Studies' tool (121). Confounders adjusted for in each model were also extracted. The data extraction form was modified and optimised after data collection from three manuscripts had been performed. High risk of bias was not used as a reason for excluding a study. Where missing data in relation to a cohort's characteristics or model were not published, the corresponding author for the cohort was contacted via email.

5.2.3. Assessment of Risk Factors

As previously stated, 'risk factor' will be used throughout to mean a measurable variable at the start of a study that is associated with a future CV disease event during the study's follow-up. Any variable was considered as a candidate risk factor if it was collected at or prior to the start point of the observational period for the study. In addition, factors were only included if they were likely to be routinely collected as part of standard primary care clinical practice. Whether a variable was routinely collected was assessed by myself and at least one other clinician (see acknowledgements). Where there was disagreement regarding a variable's inclusion, it was discussed between myself and the other two assessors until a consensus was reached. For all other stages of the methods, assessment was performed independently myself and at least one other assessor. Where discrepancies occurred, results were compared until a consensus was reached. If no consensus was achievable, a further assessor was consulted to make a final decision.

5.2.4. Meta-analysis

Data for the risk factors were extracted in the form of HR and 95% CI for the primary outcome. A random effects model using the Mantel-Haenszel method

was used to pool data as heterogeneity was expected to be present (122). Data were meta-analysed where more than one study reported results for the same risk factor. Heterogeneity was assessed using the I^2 statistics (122). Subgroup analysis was considered by CKD stage including both eGFR and proteinuria. Due to the limited clinical applicability and bias of univariable analysis of risk factors, only results from studies where multivariable adjustment for traditional CV risk factors were considered further. Models were then assessed for the number of 'core' risk factors they adjusted for. Core risk factors included age, gender, ethnicity, body mass index, smoking, diabetes mellitus, hypertension, CV disease and dyslipidemia. These risk factors are all included in general population risk prediction tools or have a firmly established association with CV disease risk (32, 69, 70, 123). In addition, because of their additive benefit to CV risk prediction models, eGFR and proteinuria measurements were also included as core adjustment co-variables (74). Where the same study had published results for a risk factor in more than one manuscript the manuscript with the most complete data was used. If the data were the same, the results from the most recent publication were used. Where more than one model was presented in the same publication, the model with the greatest number of core risk factors included was used.

5.2.4.1.   Standardisation of Variables

HR, their associated upper (U) and lower (L) 95% CI and p-values were extracted. Where these were not available, event rates were used. There was no restriction on the number of risk factors that one study could supply to the analysis. Categorical risk factors were standardised to the same reference category and continuous variables to the same units. For example, the gender risk factor was presented as the risk for being male. Therefore, if risk was presented for being female, data was converted as follows:

$$HR(male) = \frac{1}{HR(female)} \tag{5.1}$$

$$UCI(male) = \frac{1}{LCI(female)} \tag{5.2}$$

$$LCI(male) = \frac{1}{UCI(female)} \tag{5.3}$$

Where different units were reported for the same variable, those units reported in the majority of studies were used, and the minority studies' results were converted to the same units. In the case of continuous variables, units were standardised to the same unit. For example, to convert hemoglobin (Hb) from g/l to g/dl, a conversion factor of 10:

$$HR\left(Hb\left[\frac{g}{dl}\right]\right) = HR(Hb[\frac{g}{l}])^{10} \tag{5.4}$$

$$UCI\left(Hb\left[\frac{g}{dl}\right]\right) = UCI(Hb[\frac{g}{l}])^{10} \tag{5.5}$$

$$LCI\left(Hb\left[\frac{g}{dl}\right]\right) = LCI(Hb[\frac{g}{l}])^{10} \tag{5.6}$$

5.3. Results

5.3.1. Systematic Review

Three thousand two hundred and thirty-two abstracts were reviewed. Figure 5.1 shows the screening process, including the number of cohorts and risk factors identified, and reasons for any exclusion. Twenty-one cohorts were included in the systematic review (124-144). Fourteen (66.7%) studies were observational cohort studies with recruitment from nephrology outpatient settings and the others were randomised controlled trials. Six authors provided additional, unpublished data (124-129). Overall, a total of 27,465 individuals were included in these studies, representing a cumulative total of 100,838 person-years. Table 5.1 summarises the characteristics of the cohorts contributing to the systematic review

5.3.2. Risk of Bias Assessment

The risk of bias using the 'Quality in Prognostic Studies' tool (121) was medium to high in all studies. The results are summarised in Appendix 5.3. In addition to the observational nature of the studies as a source of bias, other factors relating to study participant inclusion and exclusion, assessment of outcomes, reporting of missing data and statistical methods were considered. Six cohorts (28.6%) were recruited from a single-centre. CV outcomes were broadly similar but 15 studies (71.4%) did not blind their outcome assessors. Seven cohorts (33.3%) reported no information in relation to missing data. No study pre-specified or registered their published analysis plan.
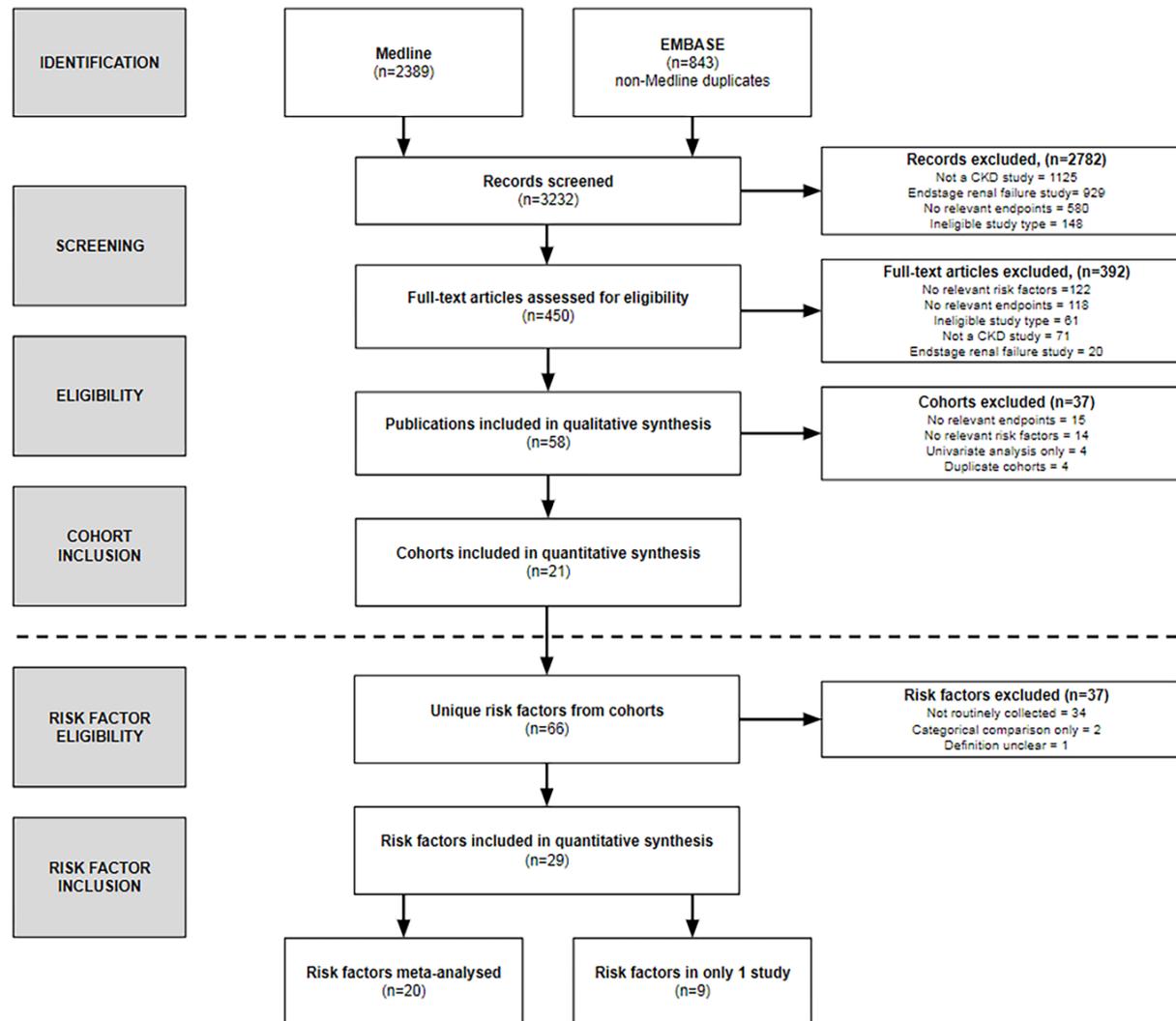
Figure 5.1: Flowchart showing the number of cohorts and risk factors identified, screened and included in the systematic review.

Study Name	Year	Journal	Study Type	Cohort Size	Mean/median follow-up (months)	Mean/median age, years	Male %	White %	Black %	Other ethnicity %	GFR Measurement	eGFR	urine	CVD %	DM %	HTN %
AASK	2006	AJKD	RCT	1094	49	55	61.2	0	100	0	125-iothalamate	46	proteinuria 0.31mg/mg	51.6	0	100
Ankara	2014	CJASN	Cohort	403	38	53.2	56.5	-	-	-	MDRD	~20% in each CKD category	1.61 g/day	13.4	22.6	15.9
CanPREDDICT	2016	Kidney International	Cohort	2529	36	68.2	62.5	88.7	-	-	MDRD	28.0	ACR 16.3 mg/mmol	33.5@	48.2	26.5$
CARE FOR HOMe	2014	CJASN	Cohort	444	31	65	60	99.8	-	0.2	MDRD	45+-16	proteinuria 37 mg/g	30.0	38	37.2^
CREATE	2010	Current Medical Research & Opinion	RCT	291	24	59.9	48.8	-	-	-	CG	-	-	93.5	-	90.4
CRIC	2013	AJKD	Cohort	3904	47	58.2	54.8	45.5	41.8	12.7	CRIC-GFR	44.8	1.07 g/day	33.4	48.5	86.1
CRISIS	2015	Nephrology	Cohort	463	46	63.8	61.8	96	-	-	MDRD	29.4	0.49 g/L protein	29.4	31.3	13.0$
Digitalis	2010	Circulation: Heart Failure	RCT	1974	57	68	65.6	89.2	-	10.8% 'non-white'	MDRD	47	-	100	50	60.2
Fujita	2013	Heart and Vessels	Cohort	404	33	67	63.6	-	-	-	MDRD	24.1	351 mg/g Cr	33.2	37.6	73.5^
Genoa	2016	CJASN	Cohort	445	71	64.1	62.0	100	0	0	MDRD	39.9	0.4 g/d	22.0	19.1	100
ICKD	2013	CJASN	Cohort	3303	36	63.5	57.8	-	-	-	MDRD and EPI-CKD	23.4 (EPI-CKD)	PCR 1118.3 mg/g	26.4	44.6	67.1
Kaohsiung	2013	Nephron Clinical Practice	Cohort	356	25	66.3	73	-	-	-	EPI-CKD	% stage given	dipstick	11.8	58.4	83.7
Kyushu	2014	Hypertension Research	RCT	320	30	72	68.1	0	0	100% Japanese	Japanese equation	18.4	1.5 g/day	19.0	51	94
Leuven	2015	Kidney International	Cohort	476	57	64	54.6	98.0	-	2.0% 'non-Caucasian'	EPI-CKD	34	0.27 g/day	27.7	18.1	70.7^
Madrid	2010	CJASN	RCT	113	23	71.6	64.6	100	0	0	MDRD	40.1	35.5 mg/d albuminuria	23.0	21	80^
MAURO	2015	CJASN	Cohort	755	31	62	60	100	0	0	MDRD	36	0.6 milligram/24 hours	29.0	35	92
Naples	2013	JACC	Cohort	436	57	65	58.3	100	0	0	MDRD	42.9	0.31g/day	30.5	36.5	72.9
OSERCE-2	2015	CJASN	Cohort	742	35	66	65	99	0	1	MDRD	27.3	proteinuria 106 mg/g	11.0	66	94
Pravastatin	2005	JASN	RCT	4670	64	62.3	21.3	>90	-	-	MDRD	56.7	dipstick	75.3	12.2	48.2
RRI	2012	NDT	Cohort	305	32	59.5	50.5	78.4	17.7	3.9	MDRD,CG	28.2	ACR 192.0 (2-9259)	36.7	30.8	88.9
TREAT	2016	Journal of Human Hypertension	RCT	4038	29	68	42.7	63.6	20.2	16.1	MDRD	33	PCR 0.39 g/g	36.5"	100	92.4

Table 5.1: Summary of 16 cohorts contributing data to systematic review. Year refers to year of publication of study. '-' refers to data not presented. ^figure based on proportion on RAAS blocker, for the Madrid cohort also 29.2% on CCB and 63.7% on diuretics. $refers to percentage with hypertensive nephropathy as cause of CKD. "refers to number with coronary heart disease, 17.6% had cerebrovascular disease. @refers to proportion with ischaemic heart disease. Journals: AJKD – American Journal of Kidney Disease, CJASN – Clinical Journal of the American Society of Nephrology, JACC – Journal of the American College of Cardiology, JASN – Journal of the American Society of Nephrology. GFR measurement: CG - Cockcroft-Gault, EPI - Chronic Kidney Disease Epidemiology Collaboration, MDRD - The Modification of Diet in Renal Disease

5.3.3. Risk Factor Assessment

Sixty-six potential risk factors for CV events were identified and are shown in Appendix 5.4. Twenty-nine of these were deemed to be routinely collected and were therefore included in the systematic review. Nine risk factors were only reported in one study and therefore the data on 20 risk factors reported in multiple studies were pooled to produce a single estimate. The confounders which were adjusted for in all the included models are shown in Table 5.2. Age was corrected for in 20 out of 21 models (95.2%) and was the most frequently adjusted for variable. Diabetes mellitus was corrected for in 17 out of 19 models (89.5%) making it the co-morbidity most frequently corrected for. Ethnicity was included in four models, six models had no published ethnicity data and eleven cohorts had a population with a single ethnicity making up more than 90% of the population. Seventeen (81.0%) studies corrected for eGFR and eleven (52.4%) for proteinuria. Three studies (14.3%) adjusted for all established core CV risk factors.

Study Name	Age	Gender	Ethnicity	Diabetes	HTN	CVD	Lipids	BMI	Smoking	eGFR	Proteinuria	Total
AASK	●	●	N/A	N/A	N/A		●			●	●	5
Ankara	●	●		●	●				●	●		6
CARE FOR HOMe	●	●	N/A	●		●				●	●	6
CanPREDDICT	●			●	●	●				●		5
CREATE	●	●		●	●	●						5
CRIC	●	●	●	●	●	●^	●	●	●	●	●	11
CRISIS	●	●	N/A	●	●	●			●	●*		6
Digitalis	●	●	●	●	●	N/A		●				6
Fujita	●	●		●		●				●	●	6
Genoa	●	●	N/A	●	●	●	●			●	●	8
ICKD	●	●		●	●	●	●	●	●	●	●	10
Kaohsiung				●	●	●				●		4
Kyushu	●		N/A		●	●		●		●		5
Leuven	●	●	N/A		●	●				●	●	6
Madrid	●		N/A	●		●				●		4
MAURO	●	●	N/A	●	●		●	●	●	●	●	9
Naples	●	●	N/A	●	●	●		●		●	●	8
OSERCE-2	●		N/A	●	●	●	●		●	●		7
Pravastatin	●		N/A	●	●	●	●		●			6
RRI	●	●	●	●	●	●	●	●	●	●	●	11
TREAT	●	●	●	N/A		●					●	5
Total	**95.2%**	**71.4%**	**40.0%**	**89.5%**	**80.0%**	**85.0%**	**38.1%**	**33.3%**	**38.1%**	**81.0%**	**52.4%**	

Table 5.2: Summary of inclusion of established CV risk factors in multi-variable models included in systematic review. 'Lipids' includes correction for using any measure of serum lipids and/or use of lipid lowering medications. N/A indicates that the model could not include the variable because 100% of study individuals were in this category, for example AASK-RCT was a study of 100% African Americans with hypertension. Where this occurred the variable was not included for percentage calculations. Notes - *corrected for serum creatinine. HTN – hypertension, CVD – cardiovascular disease

5.3.4. Meta-analysis

Data for the extracted risk factors are shown in Table 5.3.

5.3.4.1.   Traditional Risk Factors

Traditional risk factors of male gender, increasing age, smoking, established CV disease, diabetes mellitus and increasing total cholesterol were all associated with statistically significant increased risk of a CV event. Systolic and diastolic blood pressures were not associated with increased CV event risk.

5.3.4.2.   Non-traditional Risk Factors

The forest plots for the non-traditional risk factors of albumin, haemoglobin, phosphate and urate are shown in Figures 5.2, 5.3, 5.4 and 5.5. Non-traditional risk factors associated with increased risk of CV events were albumin (pooled HR 0.62 per g/dL increase, 95% CI 0.52-0.75, p<0.001), haemoglobin (pooled HR 0.90 per g/dL increase, 95% CI 0.86-0.95, p<0.001), phosphate (pooled HR 1.20 per mg/dL increase, 95% CI 1.08-1.33, p=0.005) and urate (pooled HR 1.07 per mg/dL increase, 95% CI 1.02-1.12, p=0.004). Left ventricular hypertrophy on echocardiogram (pooled HR 1.78, 95% CI 1.35-2.35, p<0.001) was also found to be associated with an increased risk of a CV event. Serum urea nitrogen, sodium and pulmonary hypertension on echocardiogram were all statistically significant but only present in one study each. Calcium, bicarbonate and parathyroid hormone were not associated with altered risk in the single studies in which they were included.

5.3.4.3.   Heterogeneity

Heterogeneity varied substantially between variables and is shown in Table 5.3. Of the potential novel risk factors for incorporation in to risk prediction models albumin (I^2=66.4%), urate (I^2=78.3%) and left ventricular hypertrophy (I^2=72.1%) showed substantial levels of heterogeneity. Based on the pre-specified protocol, subgroup analyses to explore heterogeneity were

considered for eGFR and proteinuria stages. These sub-analyses, and other post hoc analyses based on core cohort characteristics in Table 5.1, did not explain the heterogeneity for albumin. For urate and left ventricular hypertrophy, exploration of heterogeneity was limited by the inclusion of only two studies in the systematic review.

Variable	Units	Studies	Pooled HR	95% CI	p-value for HR	I² (%)
Male	-	9	1.451	1.220-1.726	<0.001	0.0
Age	per year	12	1.031	1.025-1.038	<0.001	58.6
Smoker	-	5	1.433	1.149-1.787	0.001	3.3
Body mass index	per kg/m2	3	0.994	0.964-1.025	0.7	23.0
Cardiovascular disease	-	11	2.391	2.061-2.773	<0.001	68.1
Ischemic heart disease	-	5	2.406	1.870-3.096	<0.001	43.2
Congestive heart failure	-	3	1.325	0.989-1.774	0.06	0.0-
Peripheral vascular disease	-	1	2.49	1.10-5.63	0.03	-
Diabetes mellitus	-	14	1.454	1.338-1.579	<0.001	73.5
Systolic blood pressure	per mmHg	8	1.002	0.999-1.004	0.17	77.8
Diastolic blood pressure	per mmHg	3	0.999	0.993-1.005	0.67	0.0
Mean arterial pressure	per 10 mmHg	1	1.14	1.03-1.27	0.01	-
Pulse pressure	per mmHg	3	1.002	0.998-1.005	0.38	58.7
Left ventricular hypertrophy	-	2	1.78	1.354-2.351	<0.001	72.1
Pulmonary hypertension	-	1	1.23	1.00-1.52	0.04	-
Albumin	per g/dL	7	0.624	0.519-0.749	<0.001	66.4
Bicarbonate	per mEq/L	1	0.99	0.95-1.03	0.6	-
Cholesterol to HDL ratio	ratio	1	1.03	0.998-1.065	0.07	-
Calcium	per mg/dL	1	0.846	0.503-1.422	0.5	-
Haemoglobin	per g/dL	8	0.901	0.856-0.948	<0.001	0.0
HDL Cholesterol	per mg/dL	1	0.998	0.992-1.003	0.5	-
LDL Cholesterol	per mg/dL	2	1.001	0.999-1.003	0.2	0.0
Non-HDL Cholesterol	per mg/dL	2	1.001	1.000-1.003	0.04	70.4
Parathyroid hormone	per pg/mL	1	1.00	0.99-1.00	1.00	-
Phosphate	per mg/dL	7	1.198	1.084-1.325	<0.001	0.0
Sodium	per mmol/L	1	0.954	0.919-0.990	0.01	-
Total cholesterol	per mg/dL	3	1.001	1.000-1.002	0.01	65.8
Urate	per mg/dL	2	1.068	1.021-1.117	0.004	78.3
Urea nitrogen	per 5mg/dL	1	1.14	1.02-1.29	0.03	-

Table 5.3: Results for routinely collected risk factors for combined CV events. Abbreviations: HDL – high density lipoprotein, HR – hazard ratio, LDL – low density lipoprotein. For categorical variables, the comparator was absence of the condition, except for gender, where the comparator was being female. Results are given to 3 decimal places, unless data were only available from a single study that published results to 2 decimal places.
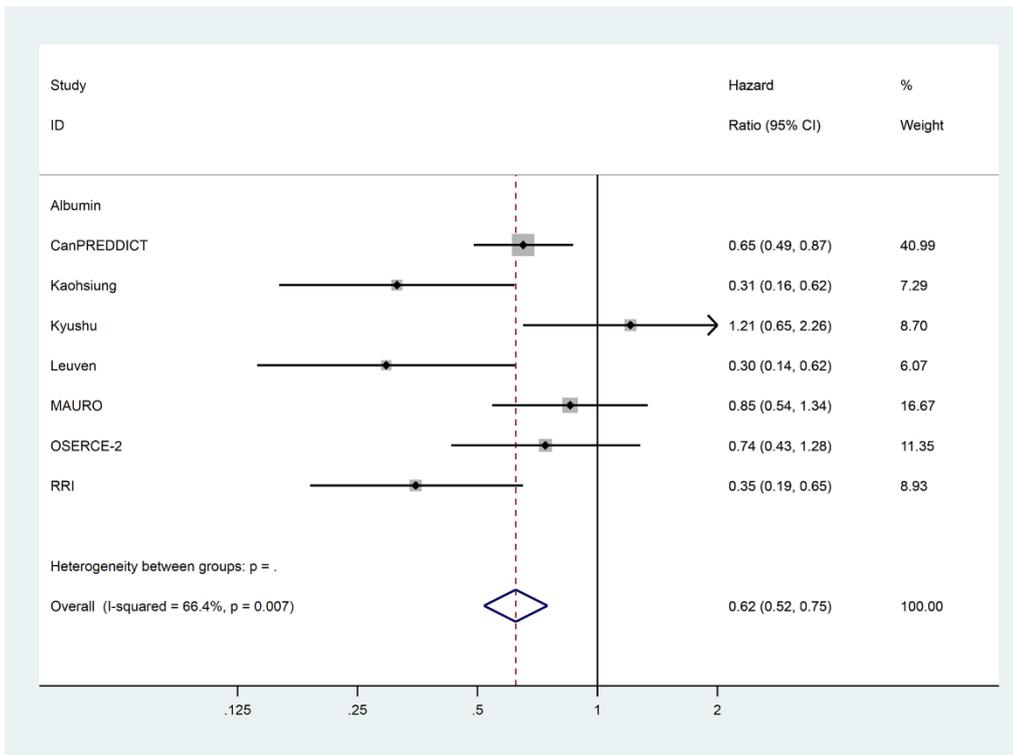
Figure 5.2: Forest plot for cardiovascular events of pooled hazard ratio for albumin per g/dL.
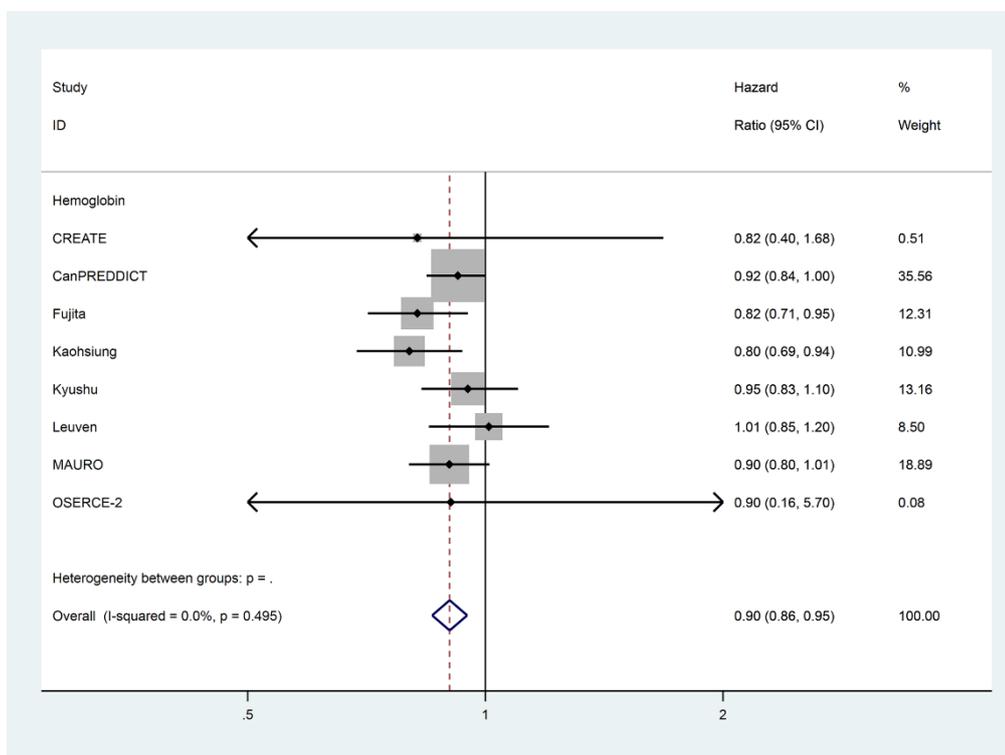
Figure 5.3: Forest plot for cardiovascular events of pooled hazard ratio for haemoglobin per g/dL.
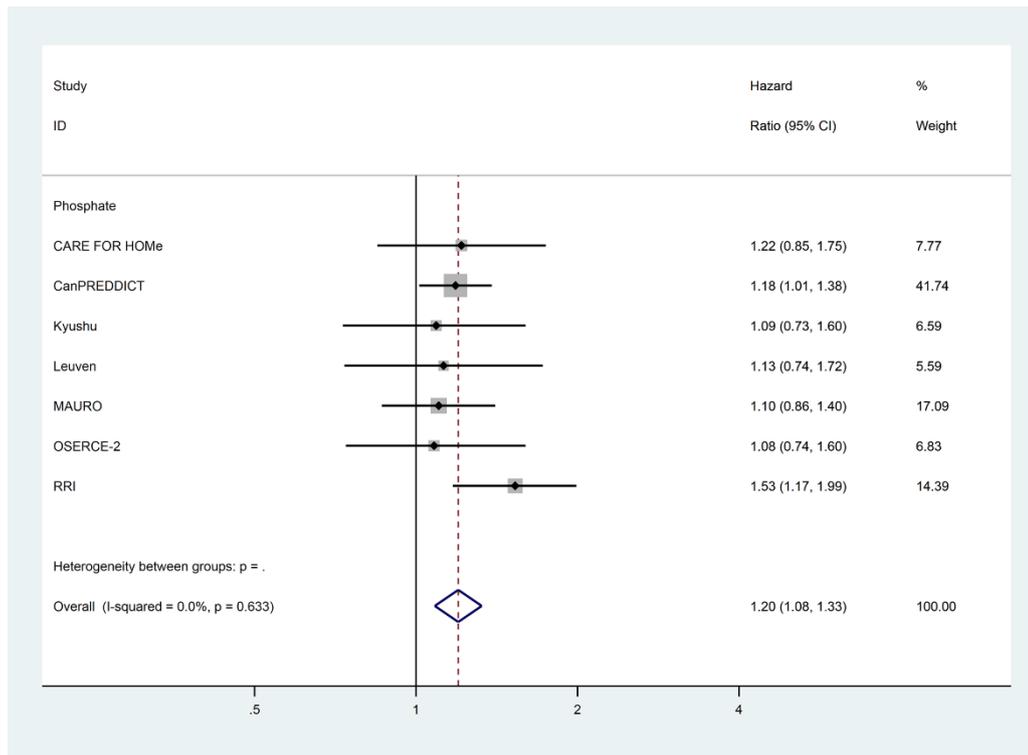
Figure 5.4: Forest plot for cardiovascular events of pooled hazard ratio for phosphate per mg/dL.
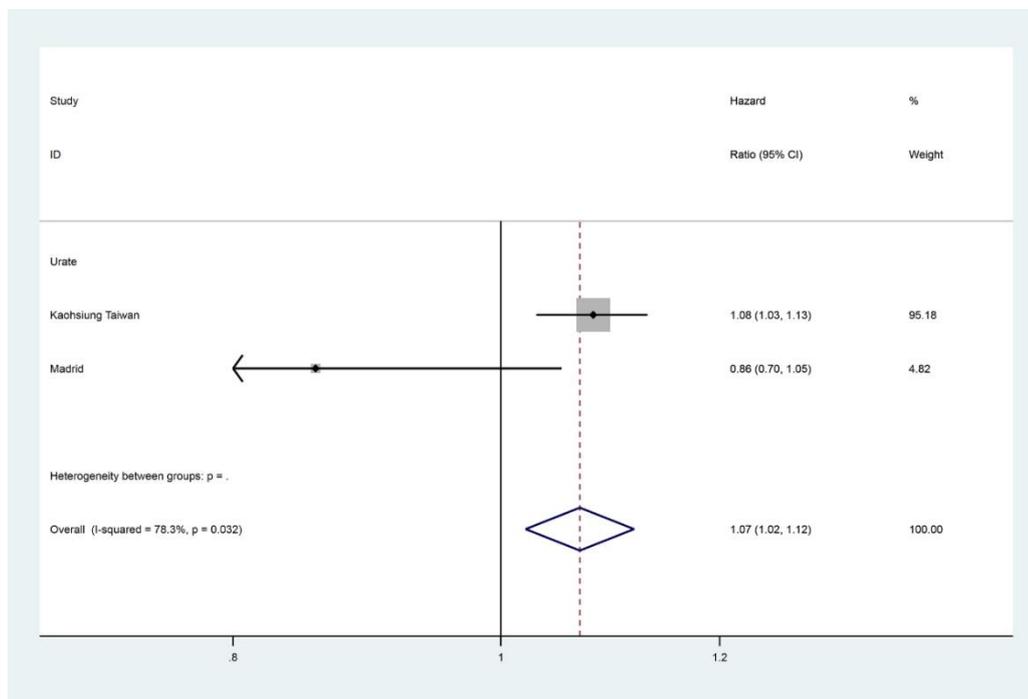
Figure 5.5: Forest plot for cardiovascular events of pooled hazard ratio for the urate per mg/dL.

5.4.   Conclusions

The systematic review presented in this chapter uses a pre-defined and registered protocol and presents the association between routinely collected risk factors and CV disease events in individuals with CKD. The results confirm that most traditional atherosclerotic related risk factors confer risk in CKD populations. These include age, gender, smoking, established CV disease and diabetes mellitus, all of which were statistically significant risk factors that are incorporated in general population risk prediction models and/or are established risk factors.

Risk factors associated with the non-traditional pathological processes described in Chapter 1, such as uraemia-related arteriosclerosis and cardiomyopathy were also identified by the systematic review. Of these risk factors, albumin, haemoglobin and phosphate were included in at least four studies and had a statistically significant pooled hazard ratio for CV events. Other non-traditional risk factors had small quantities of evidence and may need further investigation. These include risk factors associated with cardiomyopathy, such as left ventricular hypertrophy, urate, and those associated with both cardiomyopathy and arteriosclerosis including calcium, parathyroid hormone and urea nitrogen.

The results of some risk factors were more difficult to interpret. Systolic and diastolic blood pressures were not statistically significant in their association with CV events. However, mean arterial pressure was statistically significant in the single study in which it was considered. Previous studies, including individual participant meta-analysis, have suggested that the relationship of blood pressure with mortality and CV events in CKD is non-linear and may be due to uremic related myocardial and vascular remodelling (27, 145, 146). The limited availability of study-level data, and therefore the opportunity to study non-linear relationships of blood pressure to CV events in CKD, makes it difficult to draw a firm conclusion. The 'Blood Pressure Lowering Treatment Trialists' Collaboration' identified that blood pressure lowering in CKD is probably beneficial but was unable to identify a clear target (147). Recent CKD subanalysis of SPRINT, a large general population randomised controlled trial

in hypertension, suggested a possible reduction of CV events with more intensive systolic blood pressure control of <120mmHg versus <140mmHg (HR 0.81, 95% CI 0.63 to 1.05) (148). In summary, the evidence presented in this chapter in relationship to blood pressure is probably insufficient to alter the conclusions of IPD and SPRINT based findings.

Similarly, lipid measurements, including total cholesterol and low density lipoprotein cholesterol, did not have a clear relationship. A previous study of myocardial infarction events has suggested a weaker association with low density lipoprotein cholesterol as CKD advances (149). Similarly, the association of body mass index with CV events was unclear. It was not possible to assess the risk associated with ethnicity as most studies did not present data that could be utilised in models, often because ethnicity was completely, or nearly, homogenous.

Data for eleven risk factors were only included in one study each, of which four had statistically significant association with CV disease events. Therefore, replication of these findings for peripheral vascular disease, pulmonary hypertension, mean arterial pressure and serum urea nitrogen in other CKD populations is required.

Heterogeneity between studies limits the interpretation of the results of the meta-analyses presented in this chapter, particularly in the observational studies (150-152). Further, poor reporting of individual studies makes comparison of results difficult (153, 154). Innumerable evidence exists to encourage the avoidance of categorisation of continuous variables, but despite this many studies were limited by this weakness (155-157). Further common limitations include the lack of pre-registered protocols for analysis, the sole reliance on the Cox model, including limited explanation of whether its assumptions were met. The use of published values for variables, often only to two decimal places, also meant that rounding errors may be been introduced into the summation of results in the meta-analyses. This was particularly the case where a single unit of change was associated with a small change in risk. An example of this is blood pressure where often results were presented for change in a single mmHg of blood pressure and rounding a HR to two decimal

places such as 1.01 could equate to a true risk of either 1.00500 or 1.01499, or a 50% misrepresentation of relative risk.

The ideal method for selecting and combining studies, particularly between observational and randomised controlled trial data, is uncertain. The methods in this chapter try to limit this by only including results where at least some adjustment for traditional CV risk factors and CKD severity was considered. Disappointingly, despite often being published in high ranking kidney journals, not all studies adjusted for CKD staging, either via eGFR or proteinuria, which has strong evidence for its independent association with CV risk. This attempt to reduce heterogeneity may have been at the cost of reduced power, via exclusion of some cohort's results, in the meta-analysis. This approach also ensured that the results of the reported risk factors reflect the additional prognostic information above already established risk factors. Whilst individual patient data meta-analysis is the 'gold standard', the additional unpublished data from six studies used in the current study may have reduced bias in the meta-analysis results.

Despite the conservative approach taken to study exclusion, heterogeneity was substantial (122) for nine risk factors. Two characteristics of the cohorts and their analysis may explain this. Firstly, the difference in variable standardisation between studies' models may contribute to heterogeneity. Secondly, cohorts varied in the typical stage of CKD, measured through both eGFR and proteinuria, represented and this may have further increased heterogeneity.

The relatively small number of studies identified by the systematic review reflects its specific pre-specified inclusion criteria. This specificity relates to the outcome inclusion criteria of composite cardiovascular events including CV specific mortality but excluding all-cause mortality and renal related events. Prominent CKD related studies were identified by the literature review but excluded based on the inclusion criteria and/or the nature of the risk factors presented. Appendix 5.4 provides further information in relation to these studies and their exclusion.

5.5. Summary

This chapter presented the findings from a pre-registered systematic review and meta-analysis of CV risk factors in CKD. Its findings suggest that traditional general population CV risk factors are also risk factors in CKD. It also suggests that the routinely collected risk factors of serum albumin, haemoglobin and phosphate should be considered as risk factors for use in CV risk prediction models in CKD populations. Whether these risk factors are already included in models will be established in the next chapter through an updated systematic review of CV risk prediction models in CKD.

Chapter 6

6. Systematic Review of CV Prediction Models for Patients With Chronic Kidney Disease

This chapter describes an update performed to a systematic review of risk prediction models in CKD previously published by Tangri *et al* in 2012 (73). It summarises the previously identified CKD specific models and those newly identified by the literature review. General population models used in clinical practice and/or recommended by guideline authors for CV disease management in the UK, the rest of Europe and North America are also identified and summarised. Finally, I describe which models could be externally validated in the LCC cohort and the reason why some models could not be considered.

6.1. Introduction

As was noted in Chapter 1, CV risk prediction models are well established tools in the assessment of CV risk in the general population (69-71). Models for use with those with CKD are less well incorporated into routine clinical care, despite the probable different risk profile in this group of patients. Specific CKD CV models have not been integrated into clinical care for many reasons, including the methodological limitations of their development and the use of non-routinely collected risk factors for CV disease (73). Similarly, in the general population there are over 300 CV risk prediction models, many of which have not, at least on a large scale, been implemented into routine clinical care (65).

With this in mind, the aim of this chapter was to update a systematic review published in 2012 to identify new CV risk prediction models in CKD developed since its publication (73). Given the numerous general population models, it also aimed to establish a core group of risk prediction tools that are used in clinical practice, because of their recommendation in UK and international guidelines in the area. These identified models would then be assessed in external validation by the LCC cohort in the subsequent chapters of this thesis.

6.2. Methods

An update to the systematic review of CV risk prediction tools in CKD previously published by Tangri *et al* was performed (73). The lead author was contacted and the search strategy was kindly provided to me. The previous study had performed a literature review between 1966 and November 2012. Therefore, Ovid MEDLINE was searched from the beginning of 2012 to 27th April 2017. The literature review search strategy is described in Appendix 6.1.

The title and abstracts of all studies identified by the literature search were assessed. The inclusion criteria were as previous described by Tangri *et al,* 'longitudinal cohort studies involving at least 100 participants with CKD who were not treated with dialysis and had not had a kidney transplant at baseline and if they had at least one year of follow-up' (73). Pre-existing CV disease and the use of CV prevention medications were not an exclusion criteria in Tangri *et al*'s original literature search (73). Outcomes were CV events including, if applicable, acute heart failure events. If tools included a non-CV composite outcomes such as CV events and all-cause mortality they were excluded. The full text of any abstract meeting the inclusion criteria was then reviewed in full. All newly identified models, and the previously identified models, had their development cohorts' core characteristics extracted including size of cohort, definition of variables and the number of CV events extracted. The risk prediction models had the HRs and/or β coefficients for variables included in the model extracted, this included any information about a model's interaction terms. Risk prediction timeframe, survival methods used and baseline survival in relation to models were also extracted. The methods for model development were recorded as well as the level of data presented as described in Table 3.2 (107). The metrics used to present results of model fit, discrimination, calibration and the type of validation, if performed, were also collated. Information in relation to the inclusion and exclusion for both the cohort and, if different, the model were also reviewed and extracted.

Once this information had been collected, assessment was made on the likelihood of the LCC cohort data being suitable to perform external validation of the identified models. In addition, where level three data, i.e. the values of the β

co-efficients and the associated baseline survival function, for a study was not presented in the paper, the authors of the study were contacted in order to obtain further information in relation to the model. All stages of the literature review, data extraction and bias assessment were performed by myself and at least one other reviewer.

A review of UK and international CV risk management guidelines was also undertaken. The most recent versions of UK, European and North American guidelines were reviewed (70-72). Superseded versions of guidelines were not reviewed because it would be unlikely that a risk prediction tool would be used in routine clinical practice if it was no longer recommended in current guidelines. These models again underwent the same process of review as for the CKD specific models with development cohort characteristics, model descriptors and their eligibility for external validation assessed.

6.3. Results

6.3.1. Literature Review for CKD Models

Three thousand one hundred and thirty-three abstracts were reviewed. Figure 6.1 shows the screening process, including the number of publications and risk prediction models identified, and reasons for any exclusion. The full-text of 35 papers was reviewed. Six publications, including seven models, were identified (74, 135, 158-161). Therefore, including the publications and models identified by Tangri *et al*, nine publications including 15 models were assessed for potential external validation (144, 161, 162). Three models presented CV mortality outcomes only (158, 163) and five models included variables, including experimental biomarkers, that were not available in the LCC cohort (74, 144, 159, 160). Therefore, seven separate CKD-specific models from three CKD cohorts were considered for external validation (135, 161, 162). There was heterogeneity of exclusion criteria between the three cohorts. Two excluded individuals with CV disease at baseline. The Weiner cohort (162) excluded individuals 75 years of age or older but the other two cohorts had no age-based exclusion criteria. No model specifically excluded individuals who were on a specific medications, such as a statin.

The models' characteristics are shown in Table 6.1. Similarly to Tangri *et al*'s earlier findings, presentation of models was suboptimal. None of the new models presented all core model metrics of discrimination, calibration, model fit, reclassification and external validation. Level one data, i.e. the values of the β co-efficients for the variables in the respective models, were presented in all seven models. Four of these models used a baseline hazard function from another study, the Framingham Heart Study, and did not specify a re-calibrated baseline function for their own model. Variables included in the models are shown in Table 6.2. All models included age, gender, smoking, diabetes mellitus and hypertension. eGFR was included in one model but proteinuria was not included in any model. Variables assessing body habitus such as body mass index or waist circumference were not included in any models.
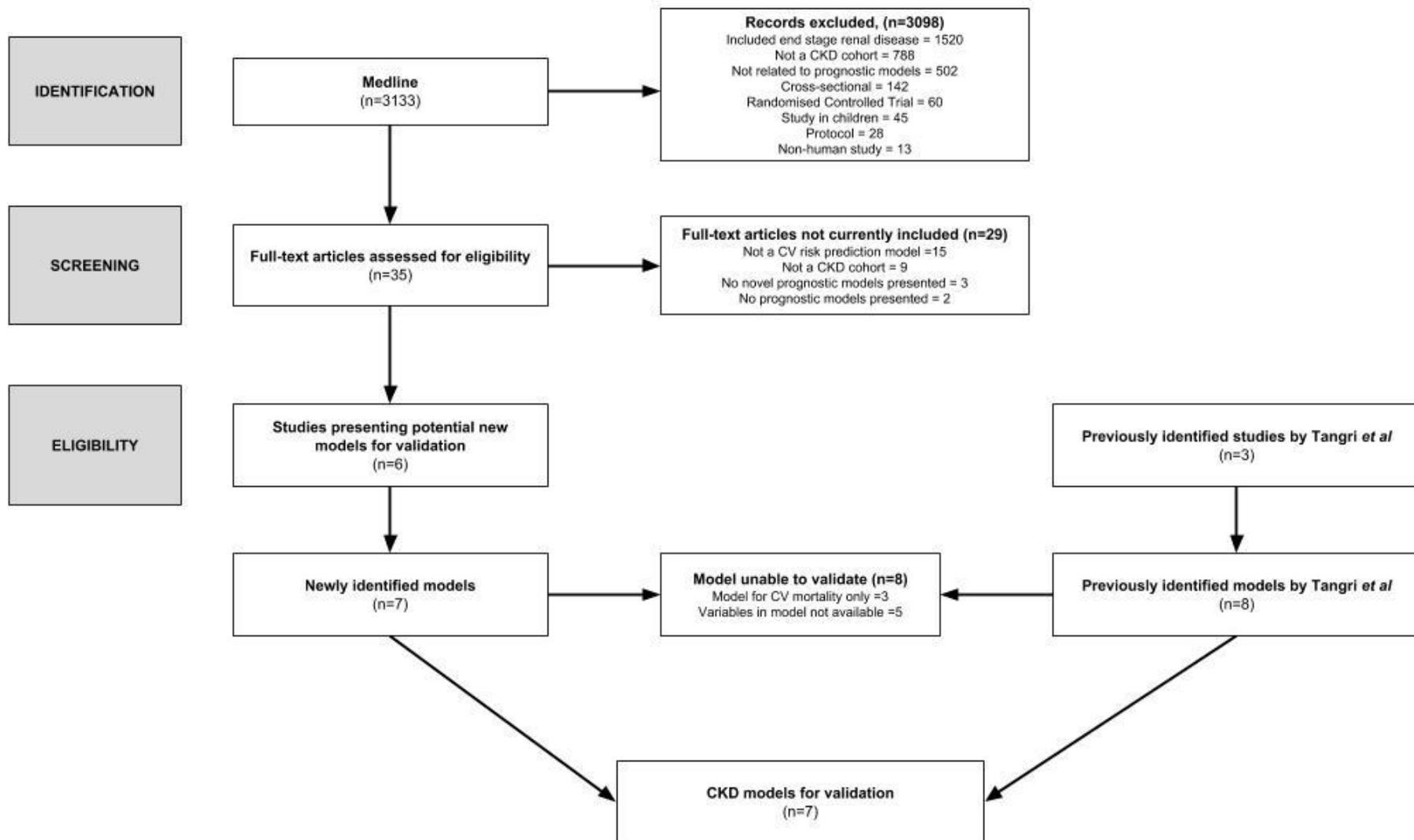
Figure 6.1: Literature Review and Assessment of CKD Risk Prediction Models.

6.3.2. General Population Models

Based on NICE and international guidelines, four main general population risk prediction models were identified for potential external validation. The risk prediction models recommended in their current guidelines by The American Heart Association combined cohort equation ('AHA') (70), The European Society of Cardiology ('SCORE') (71) and National Institute of Health and Clinical Excellence (England and Wales) ('QRisk2') were considered for external validation (72). Since the latter's publication of its latest recommendation, an update of QRisk2 has been published, QRisk3 (67), and therefore, although not included in the guidelines, this was considered for assessment too. All general population models produced specific models for males and females and these were therefore assessed separately. AHA also produced separate models for non-Hispanic black populations, however, because the cohort used for assessment only included 14 events from the 123 black individuals included, these models were not assessed. SCORE only assesses the risk of CV related mortality only and therefore this model was not considered further.

Therefore the male and female models for each of QRisk2, QRisk3 and AHA were assessed, meaning that six general population models were externally validated. All six models presented Level 3 data. Summary of the general population models are presented in Tables 6.1 and 6.2.

The general population cohorts were significantly larger than the CKD cohorts both in relation to the cohort size and the number of events during follow-up. All three were primary prevention cohorts and included age-based exclusion criteria, albeit at different cut-off points. QRisk2 and QRisk3 excluded individuals who were already receiving a statin medication. None of the general population models included continuous kidney function measurements such as eGFR or proteinuria in their model. A binary variable for the presence of CKD stages 4 or 5 was included in the QRisk2 model, and extended to include CKD stages 3a or 3b for QRisk3. Body mass index, and associated interactions terms, were included in the QRisk2 and QRisk3 models, but no body habitus

variable was included in AHA. All three general population models included interaction terms with age and other variables.

Name	Year	Cohort		Model Timeframe (years)	Baseline Exclusions	Variables in Model	Baseline Hazard Presented	Model Validation	Data Level
		n	Events						
CKD									
Weiner FH (162)	2007	934	65	5	75 years and older, coronary heart disease	6	Not presented	Internal	1
Weiner CBF (162)	2007	934	65	5	75 years and older, coronary heart disease	6	Not presented	Internal	1
Matushita (161)	2014	940	336	10	CHD, stroke, HF, ethnicities other than white or black	8	Yes, from main Framingham paper	None	1
Alderson (135)	2015	463	108	3.8[+]	None	13	Not presented	None	1
General Population									
AHA (70)	2013	20,338	2,161	10	<40 or >79 years of age, nonfatal myocardial infarction, stroke, HF, percutaneous coronary intervention, coronary artery bypass surgery, or atrial fibrillation	7	Female 0.9665 Male 0.9144	Internal 10x10 cross-validation technique and external validation	3
QRisk2 (69)	2008	1,535,583[*]	96,709[*]	10	<35 or >74 years of age, cardiovascular disease, statin use, no deprivation score	14	Yes	Random split by practice. 2/3 derivation, 1/3 validation	3
QRisk3 (67)	2017	7,889,803	363,565	10	<25 or >84 years of age, cardiovascular disease, statin use, no deprivation score	20	Yes	Random split by practice. 3/4 derivation, 1/4 validation	3

Table 6.1: Summary Characteristics of Risk Prediction Models for External Validation. Based on findings of Tangri et al and updated literature review. + - median follow-up, * - refers to derivation cohort.

Model	Age	Gender	Smoking	CVD	DM	HTN	eGFR	Proteinuria	TC	HDL	Other
Weiner FH (162)	●	●*	●	N/A	●	●	-	-	●	●	-
Weiner CBF (162)	●	●*	●	N/A	●	●	-	-	●	●	-
Matushita (161)	●	●	●	N/A	●	●	-	-	●	●	BNP
Alderson (135)	●	●	●	●	●	●	●	-	-	-	Serum phosphate, calcium, albumin, haemoglobin, PTH
AHA (70)	●	●*	●	N/A	●	●	-	-	●	●	Interaction terms include – age*TC, age*HDL, age*SBP, age*smoking
QRisk2 (69)	●	●*	●	N/A	●	●	-	-	●	●	BMI, family history of CHD, deprivation, treated hypertension, rheumatoid arthritis, CKD, AF
QRisk3 (67)	●	●*	●	N/A	●	●	-	-	●	●	BMI, family history of CHD, deprivation, treated hypertension, rheumatoid arthritis, CKD, AF, SBP variability, migraine, corticosteroid use, systemic lupus erythematosus, 2nd generation "atypical" antipsychotic use, severe mental illness, erectile dysfunction

Table 6.2: Variables included in identified models. N/A – no individuals with CVD in cohorts. * - separate models with same variables presented for male and female cohorts. FH – Framingham model, CBF – 'Cox Best' Framingham model, TC – total cholesterol, HDL – high density lipoprotein, BNP – brain natriuretic peptide, PTH – parathyroid hormone.

6.4.   Conclusions

In this chapter I have presented results of an updated systematic literature review previously performed by Tangri *et al.* Six CV risk prediction models, from three different studies, had been previously identified, but all had significant methodological weaknesses and were unlikely to be able to be implemented into routine clinical practice (73). The updated findings from this chapter identify a further six studies, each presenting one model. Again, the models continued to have significant methodological flaws and/or utilised experimental biomarkers, meaning that they were unlikely to be usable on a population-wide scale as routine clinical prediction aides. In total, seven CKD-specific risk prediction models were identified as suitable for external validation by the LCC cohort.

The review of general population guidelines for CV risk management identified four risk prediction models, male and female models for AHA and QRisk2 (69, 70). In addition, during the course of the review a materially significant update to QRisk2 was published, QRisk3 (67). While QRisk3 is not included, in the current NICE guidelines it was considered eligible as it was identified in a 'surveillance report' for consideration for incorporation into the guideline update (164). Four cohorts were assessed but one, SCORE, only assessed CV mortality and therefore was not considered further for external validation by LCC. Only one CKD specific model, Alderson, included a quantifier of the severity CKD, eGFR, in their model. Both QRisk2 and QRisk3 included medically coded CKD in their models, but both included it as a binary variable with no quantification, either through eGFR or ACR, included. In addition, the accuracy of medically coded CKD may be incorrect in up to a quarter of individuals, both in relation to false positives and false negatives compared to biochemical CKD diagnoses (21).

The current literature review represents an update from the previous literature review and therefore the robustness of the current review relies on Tangri *et al*'s review methods and findings. Again, all searches were duplicated by a second individual and comparison of results probably improved the reproducibility of the findings. The fact that the updated model of QRisk3 was

identified through the literature review was also reassuring in relation to its thoroughness. An extended period of overlap of 11 months, between the current update starting and the previous literature review ending ensured that it would be unlikely that an eligible model was missed between the searches.

There are some limitations to the current review and these mostly relate to the nature of the systematic review as an update as opposed to a new search *per se*. Only one medical literature database, OVID Medline, was searched for the purpose of the review. In addition, neither the work by Tangri *et al* nor this update registered pre-defined protocols. For the former, this may relate to the fact that prospective registries of systematic reviews such as PROSPERO had not been conceived when the review was originally performed (165). Finally, performance of different timeframe periods of the review by two different groups of researchers may have led to some differences in findings between groups. However, the process of literature searches for systematic reviews will always have a degree of heterogeneity between them and the methods set out above and in the previous chapter have aimed to minimise these issues.

6.5. Summary

The first aim of this chapter was to update the systematic review of CV risk prediction models previously published by Tangri *et al* in 2012 (73). The second aim was to identify general population models recommended for use in routine clinical practice. The process for the chapter also set out to establish which models could be externally validated using the LCC cohort. A total of three CKD cohorts described seven CKD specific models (135, 161, 162) and three general population cohorts presented a further six models (67, 69, 70). Five out of the six cohorts published models for male and female groups within their datasets. In the following chapter, the external validity of these risk predictions models is further assessed in the LCC cohort.

Chapter 7

7. External Validation of CKD and General Population CV Disease Risk Prediction Models

This chapter performs external validation of CKD and general population models identified in the previous chapter using the LCC cohort. Firstly, model discrimination is assessed using both complete case and the imputed datasets. Secondly, where baseline risk ('level three') data were available from previously published model calibration was performed. Finally, results for simple re-calibration of the baseline risk are presented for each model where calibration was performed. Due to the number of models identified, detailed results, including sensitivity analysis, for the QRisk3 model are presented with summary results for the other models. Results are also presented for the whole LCC cohort and the 'cohort specific' subcohort which would have been included under the inclusion criteria of the model's original development cohort.

The following presentation has been given in relation to this chapter:

- 'Heart Attack and Stroke Risk Prediction in Kidney Disease', Kidney Research UK Fellows' Day 2018 – oral presentation

7.1. Introduction

In order to assess a risk prediction model, ideally prior to clinical implementation, it should be assessed in other independent datasets in a process referred to as external validation. The proportion of developed models undergoing external validation is low in CV risk prediction (65), although there are notable exceptions (166). The 'Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis' (TRIPOD) statement describes a checklist for reporting risk prediction models, including external validation and its suggestions are followed in the coming chapter (167).

The LCC cohort was described in detail Chapter 4 and only the details relevant to external validation will be repeated in this chapter. Assessment of prediction was performed for gender specific models individually and in the Whole Cohort and Cohort Specific groups. Therefore specific cohort details will be set out in relation to this. The LP, or amount to which individual and population risk varies from the baseline risk, is described. Risk is compared between Whole Cohort and Cohort Specific groups, where the latter only includes individuals with the same inclusion criteria as the original development model. Also, where available, risk is compared to the original values of the LP in the development cohort. This allows for an overall comparison of risk between the development dataset and the LCC cohort. Discrimination of the risk prediction models, for both MI dataset results and complete case analysis, is then described and its difference to the originally reported C statistic compared. Where level three data were available the performance of calibration through OE plots, as described in Chapter 3, is then presented. Re-calibration is also performed where miscalibration has occurred and the change in baseline risk from development to the LCC cohort is described.

7.2. Methods

As previously described in section 3.6, models were assessed for external validation if similar CV outcomes and variables were available in the LCC cohort, i.e. for outcomes acute coronary syndrome and cerebrovascular disease events. As previous described in Table 3.2, the 'level' of data in the original model dictated the level of external validation that could be performed. Level one information refers to where only regression coefficients are available and level two where regression coefficients, risk groups and associated Kaplan-Meier plots are presented for the model. Level three information, which allows full assessment of both the discrimination and calibration of the model to take place, is only possible where level two data and the baseline hazard or survival function are available for the model (107).

Multiple imputation of missing values of variables in models was performed as previously described. Imputed values were then used to calculate predicted risk before assessment of models was performed. The mean of C across the imputed datasets was used to assess general discrimination for all models. Calibration was assessed where level 3 data was presented by calculating the expected versus observed event ratio and using OE plots. The mean values across imputed datasets were again used for these metrics. All risk calculation was based on the full baseline risk reported for a model. In some cases this was described to up to 15 decimal places in the original publication but for simplicity in the thesis four decimal places are given.

All assessments were made using every individual in the cohort ('Whole Cohort') of LCC and the subgroup of LCC that matched the inclusion and exclusion criteria of the model ('Cohort Specific') characteristics and outcomes of the cohort used in the risk prediction model's development. For example, if a risk prediction model had been developed for primary prevention of CV disease then only individuals without a history of CV disease were included in the cohort specific model assessment. Equally, if use of a medication, such as a statin, was an exclusion criteria in the risk prediction model's development cohort then it was also an exclusion criteria in the Cohort Specific group.

This approach of using a separate Cohort Specific group to test the models allowed the population with CKD similar to the original development cohorts group to be tested and validated. However, by also testing the Whole Cohort group it also allowed general testing of the validity of the models in all CKD, regardless of pre-existing co-morbidities, age and medication use.

As previously described, the variables collected for the LCC were selected based on their inclusion in general population risk prediction models, including QRisk2 (69). QRisk3 was published after data had been collected for the LCC cohort and it was not possible to retrospectively collect some new variables included in QRisk3 that had not been previously included in QRisk3 (67).

Eight new variables are included in the QRisk3 risk prediction model. One of these is a continuous variable, systolic blood pressure variability based on its standard deviation, and this data was available in the LCC cohort. Two categorical variables, an expanded definition of CKD to include stages 3a and 3b, and the use of corticosteroid medications, were also available in the LCC cohort and therefore could be included in the assessment. The remaining six categorical variables, migraine, systemic lupus erythematosus, atypical anti-psychotic use, severe mental illness and erectile dysfunction, were not available in the LCC cohort. Their prevalence in the QRisk3 cohort varied from approximately 6% for severe mental illness to less than 0.1% for systemic lupus erythematosus. For the external validation assessment of QRisk3, it was assumed that no individual in the cohort had any of these conditions/medications.

7.3. Results

7.3.1. Summary of LCC Cohort

The LCC cohort is described in detail in Chapter 4, but as suggested by the TRIPOD statement, the cohort's descriptors relevant to external validation are briefly re-described.

Forty-four GP practices participated in the LCC cohort representing 277,248 registered adults. Individuals with a medical history at baseline of a kidney transplant or dialysis were excluded. 17,248 individuals (6.2%) had two or more EPI eGFRs <60 ml/min/1.73m2 >90 days apart and were included in the cohort. 10,353 (60.0%) individuals were female, mean age was 77.4 years (SD10.0), 12,560 (72.8%) were of white ethnicity, 11,410 (66.2%) had no known history of CV disease at baseline. 12,362 (71.7%) individuals has CKD stage 3a (EPI eGFR 45-59 ml/min/1.73m^2) and 5,846 (33.9%) ACR stage A1 (ACR<3mg/mmol). 5,592 (32.4%) of the cohort had diabetes mellitus and 15,674 (90.9%) hypertension. Follow-up commenced on 1st November 2011 and finished on 1st November 2016. During a median follow-up of 5.0 (IQR 3.3 to 5.0) years 2,072 (12.0%) CV events occurred.

7.3.2. Description of Risk Prediction

The mean values for predicted risk and the LP were calculated for the models with Level 3 information. LP were calculated for all model using the imputed datasets and on a complete case analysis. Table 7.1 describes these results for five year predicted risk. Compared to published data available for QRisk2, QRisk3 and AHA, both LCC Whole Cohort and Cohort Specific groups had substantially higher predicted mean risk than the original models. For instance, the absolute difference in mean predicted risk between QRisk2 Female development and LCC was 13.7% (95% CI 13.2 to 14.1, p<0.001). No comparative published data was available for QRisk3. Mean LP for the whole cohort group was higher in LCC Whole Cohort than in Cohort Specific groups for all models. For example, the β co-efficient 0.309 (95% CI 0.296 to 0.322, p<0.001) higher in the Whole Cohort compared to the Cohort Specific group. This was also reflected in mean predicted risk being higher in Whole Cohort groups compared to Cohort Specific groups.

Model	Development		Whole Cohort				Cohort Specific			
	5 Year Predicted Risk	95%CI	5 Year Predicted Risk	95% CI	Mean LP	95% CI	5 Year Predicted Risk	95% CI	Mean LP	95% CI
QRisk3 Female	-	-	0.1969	0.1948 to 0.1990	3.533	3.520 to 3.546	0.1522	0.1484 to 0.1559	3.224	3.193 to 3.255
QRisk3 Male	-	-	0.2392	0.2367 to 0.2417	3.083	3.070 to 3.096	0.2101	0.2047 to 0.2155	2.914	2.879 to 2.950
QRisk2 Female	0.0311*	0.0301 to 0.0321*	0.2268	0.2238 to 0.2298	3.717	3.701 to 3.733	0.1676	0.1630 to 0.1723	3.357	3.321 to 3.392
QRisk2 Male	0.0471*	0.0444 to 0.0497*	0.2542	0.2513 to 0.2571	3.203	3.188 to 3.218	0.2144	0.2084 to 0.2205	2.971	2.230 to 3.012
AHA Female	0.0168	-	0.2796	0.2752 to 0.2840	2.603	2.580 to 2.627	0.1237	0.1204 to 0.1270	1.765	1.731 to 1.799
AHA Male	0.0428	-	0.2383	0.2350 to 0.2417	1.631	1.611 to 1.651	0.1624	0.1580 to 0.1669	1.220	1.184 to 1.255

Table 7.1: Summary Mean Predicted 5 Year Risk and LP values for CKD and General Population CV prediction models for Whole Cohort and Cohort Specific cohorts. *Figures based on mean predicted risk from external validation by Collins and Altman (166).

7.3.3. Discrimination

Overall discrimination for imputed and complete case analysis for the identified CKD and general population models is shown in Table 7.2. In general, all models had worse discrimination in the current cohort than in their original development models.

The Matsushita Framingham based model had the best discrimination of the CKD based models. However, overall CKD based models performed worse than general population models both in the current Whole Cohort and Cohort Specific group. Results were similar for complete case analysis.

In relation to the female general population models, the QRisk3 Female model had the best discrimination. For male population models, the AHA model had the highest c-statistic but it was not significantly different to the other male general population models. Therefore, the QRisk3 Female and QRisk3 Male models were selected for sensitivity analysis of their discrimination and for assessment of calibration. Table 7.3 shows the results of the sensitivity analysis for discrimination for QRisk3. Generally, model discrimination performed better in lower age groups, with the caveat of smaller sizes limiting their interpretation. All models had better discrimination in the female cohort compared to the male cohort. Discrimination was similar for the separate outcomes of myocardial infarction only, stroke only, death only and the composite outcome of either a cardiovascular event or death.

Model	Original Performance		MI Results				Complete Case Results			
			Whole Cohort		Cohort Specific		Whole Cohort		Cohort Specific	
	C	95% CI	C	95% CI	C	95% CI	C	95% CI	C	95% CI
CKD										
Alderson	0.873	0.825 to 0.921	0.652	0.641 to 0.664	n/a	n/a	0.587	0.575 to 0.600	n/a	n/a
Weiner Female Framingham	0.77	not reported	0.553	0.537 to 0.569	0.549	0.531 to 0.568	0.563	0.545 to 0.580	0.567	0.535 to 0.598
Weiner Male Framingham	0.62	not reported	0.569	0.554 to 0.585	0.577	0.556 to 0.599	0.586	0.567 to 0.605	0.606	0.570 to 0.641
Weiner Female Best Cox	0.82	not reported	0.542	0.525 to 0.559	0.542	0.523 to 0.562	0.540	0.522 to 0.558	0.549	0.517 to 0.582
Weiner Male Best Cox	0.72	not reported	0.564	0.546 to 0.583	0.560	0.533 to 0.586	0.560	0.540 to 0.579	0.557	0.520 to 0.593
Matsushita Female Framingham	0.679	not reported	0.601	0.585 to 0.617	0.610	0.586 to 0.633	0.570	0.581 to 0.614	0.611	0.586 to 0.636
Matsushita Male Framingham	0.679	not reported	0.591	0.572 to 0.610	0.645	0.613 to 0.677	0.588	0.569 to 0.607	0.640	0.606 to 0.674
General Population										
QRisk3 Female	0.880	0.879 to 0.882	0.666	0.651 to 0.681	0.684	0.647 to 0.720	0.625	0.606 to 0.644	0.637	0.574 to 0.699
QRisk3 Male	0.858	0.857 to 0.860	0.618	0.599 to 0.637	0.628	0.573 to 0.683	0.676	0.660 to 0.691	0.680	0.639 to 0.720
QRisk2 Female	0.817	0.814 to 0.820	0.656	0.641 to 0.671	0.670	0.633 to 0.707	0.608	0.582 to 0.634	0.640	0.542 to 0.739
QRisk2 Male	0.792	0.789 to 0.794	0.625	0.606 to 0.643	0.634	0.579 to 0.688	0.664	0.644 to 0.683	0.695	0.643 to 0.747
AHA Female	0.806	not reported	0.665	0.650 to 0.680	0.645	0.607 to 0.684	0.616	0.597 to 0.635	0.663	0.611 to 0.714
AHA Male	0.746	not reported	0.621	0.603 to 0.639	0.667	0.617 to 0.717	0.665	0.650 to 0.680	0.645	0.607 to 0.684

Table 7.2: Summary C-statistics for MI and Complete Case Results for CKD and General Population CV prediction models. C-statistics and 95% CI presented are the mean of c-statistic results for each imputed cycle. C – Harrell's concordance statistic, CI – confidence interval.

	QRisk3 Female				QRisk3 Male			
	Whole Cohort		Cohort Specific		Whole Cohort		Cohort Specific	
	C	95% CI	C	95% CI	C	95% CI	C	95% CI
All CVD events	0.684	0.647 to 0.720	0.666	0.651 to 0.681	0.628	0.573 to 0.683	0.618	0.599 to 0.637
Age Group								
<59	0.712	0.583 to 0.840	0.792	0.586 to 0.998	0.591	0.476 to 0.705	0.571	0.269 to 0.873
60-69	0.696	0.625 to 0.767	0.578	0.373 to 0.783	0.591	0.529 to 0.652	0.455	0.306 to 0.605
70-79	0.581	0.550 to 0.613	0.559	0.483 to 0.635	0.602	0.571 to 0.634	0.589	0.495 to 0.684
80+	0.580	0.560 to 0.601	0.566	0.505 to 0.627	0.554	0.526 to 0.582	0.568	0.486 to 0.651
Outcomes								
Myocardial Infarction Only	0.666	0.643 to 0.690	0.668	0.603 to 0.734	0.611	0.584 to 0.638	0.562	0.462 to 0.661
Stroke Only	0.668	0.650 to 0.685	0.688	0.646 to 0.730	0.621	0.597 to 0.646	0.672	0.604 to 0.741
Death Only	0.701	0.691 to 0.710	0.675	0.651 to 0.699	0.655	0.643 to 0.666	0.651	0.617 to 0.685
Myocardial infarction, Stroke & Death	0.687	0.678 to 0.696	0.677	0.656 to 0.698	0.64	0.629 to 0.651	0.642	0.611 to 0.674

Table 7.3: Sensitivity Analysis for C-Statistics for QRisk3 Models. C and 95% CI presented are the mean of C results for each imputed cycle. *change in C based on recalculated risk of models with the addition of log transformed eGFR and ACR as additional continuous predictor variables. C – Harrell's concordance statistic, CI – confidence interval.

7.3.4. Calibration

7.3.4.1.   QRisk3 Female

For the current analysis the calibration of the QRisk3 Female model was initially assessed, based on its good performance in relation to discrimination, the contemporary nature of its development in relation to other models and its development in a UK primary care population (67). Calibration was assessed using the originally published 10-year CV event survival estimate for the QRisk3 Female of 0.9889 (67). To initially estimate five year risk, baseline risk was halved to produce a new survival estimate of 0.9944. Calibration for this initial model, as shown in Figure 7.1, substantially overestimated the risk of a cardiovascular event for both Whole Cohort and Cohort Specific groups.

The baseline survival value was calculated from the mean of the survival re-estimates at 5 years for each imputed dataset. Each re-estimate of survival was based on the Cox model results using the off-set beta-coefficient for the imputed dataset. For Whole Cohort the five year survival estimate was 0.9960, or a 28.6% increase. The Cohort Specific had a re-calibrated baseline risk of 0.9969, representing a 44.6% increase in the baseline survival estimate. These re-estimated baseline risks, as shown in Figure 7.2, led to improved calibration for each decile for both Whole Cohort and Cohort Specific groups for QRisk3 Female.
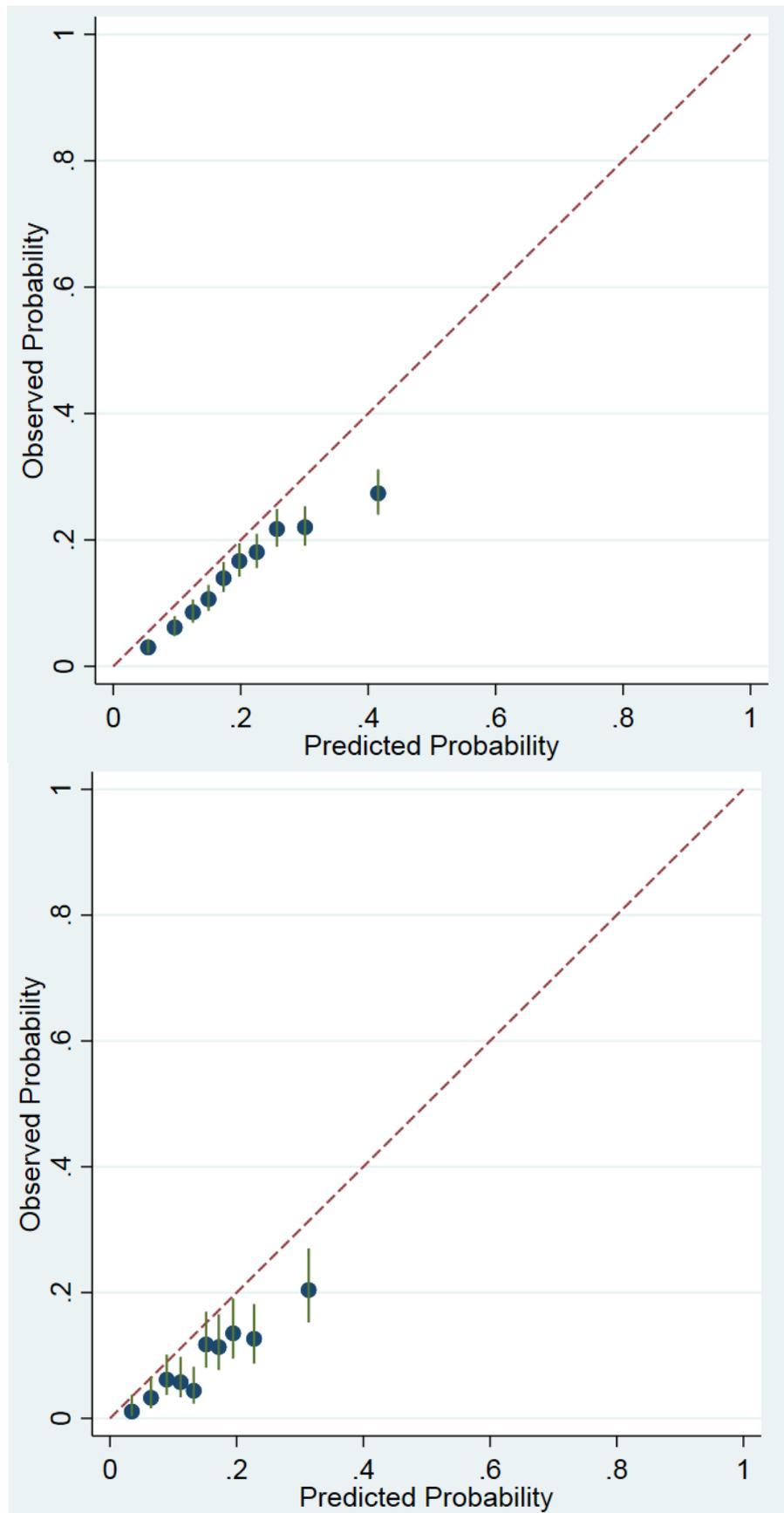
Figure 7.1: Calibration plot for **Female QRisk3** for a cardiovascular event for Whole Cohort (top) and Cohort Specific (bottom). Groups split in to deciles and risk based on original baseline risk as published in QRisk3.
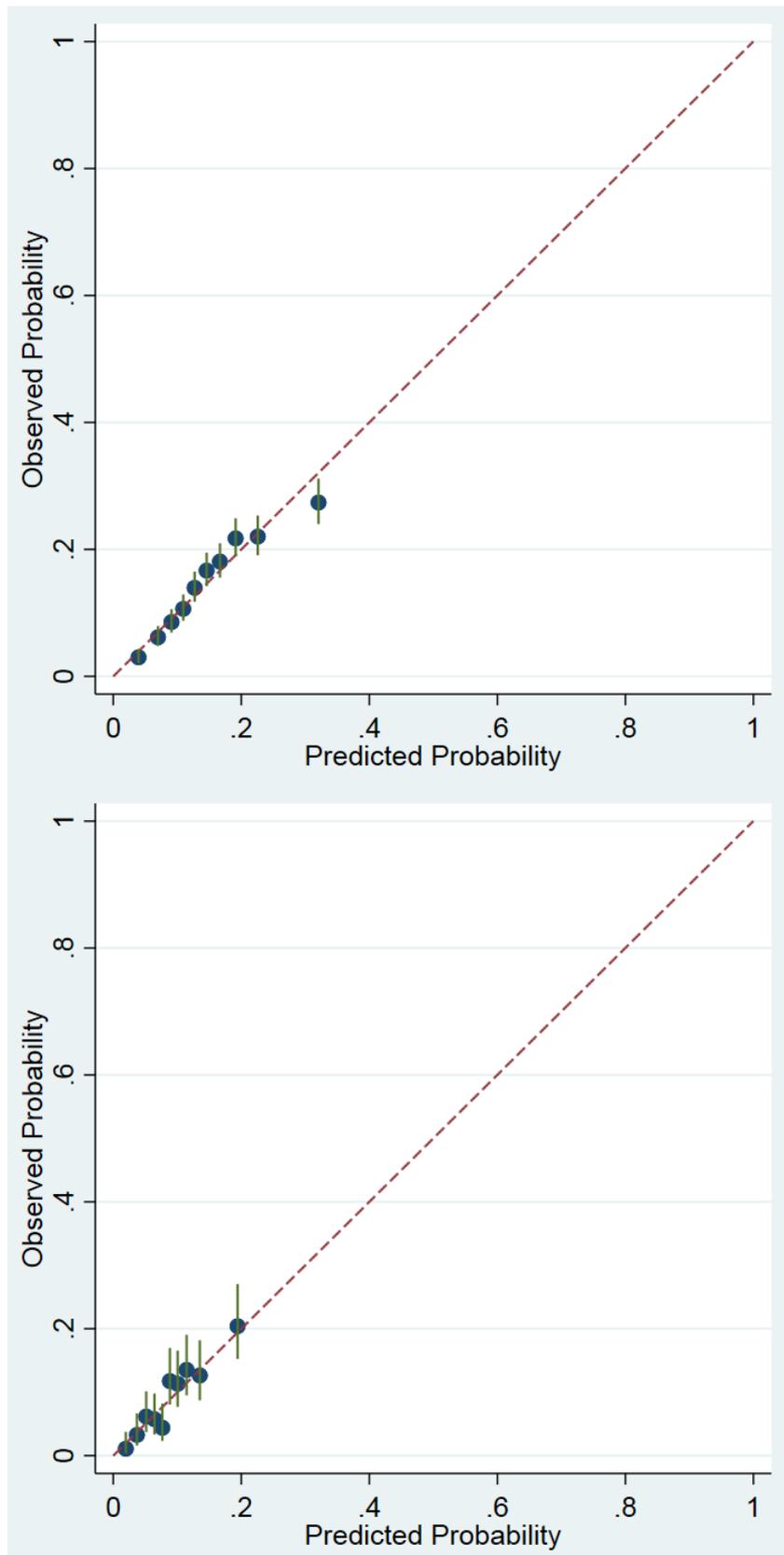
Figure 7.2: Calibration plot for recalibrated **Female QRisk3** risk equation for a cardiovascular event for Whole Cohort (top) and Cohort Specific (bottom). Baseline 5 year risk for Whole Cohort=0.99602109, Baseline 5 year risk for Cohort Specific=0.9968673. Groups split in to deciles.

7.3.4.2. Other Models

Calibration for the other models, QRisk3 Male, QRisk2 Female, QRisk2 Male, AHA Female and AHA Male, was also performed and the results are shown in Figures 7.3 to 7.7 inclusive. Similar methods were used to adjust the baseline risk for five year prediction. All models again showed similar over prediction of risk based on their individual original baseline risk estimations. All models had improvement in prediction when they were recalibrated. For the latter, some models continued to display overestimation of risk in higher risk deciles.
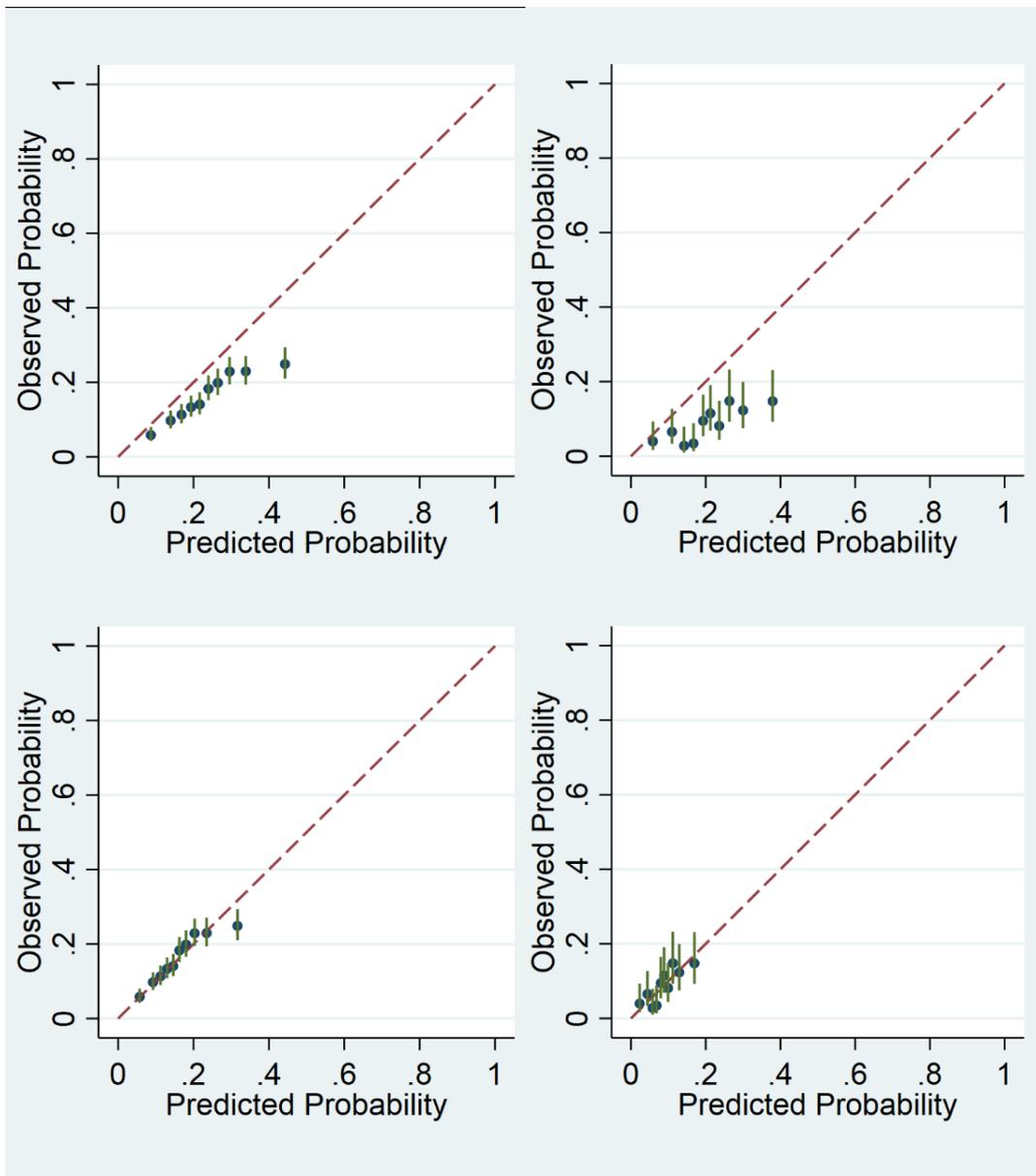
Figure 7.3: Calibration plots for **QRisk3 Male** for Whole Cohort (left column) and Cohort Specific (right column). Top row – original risk prediction model, bottom row – re-calibrated risk prediction model.
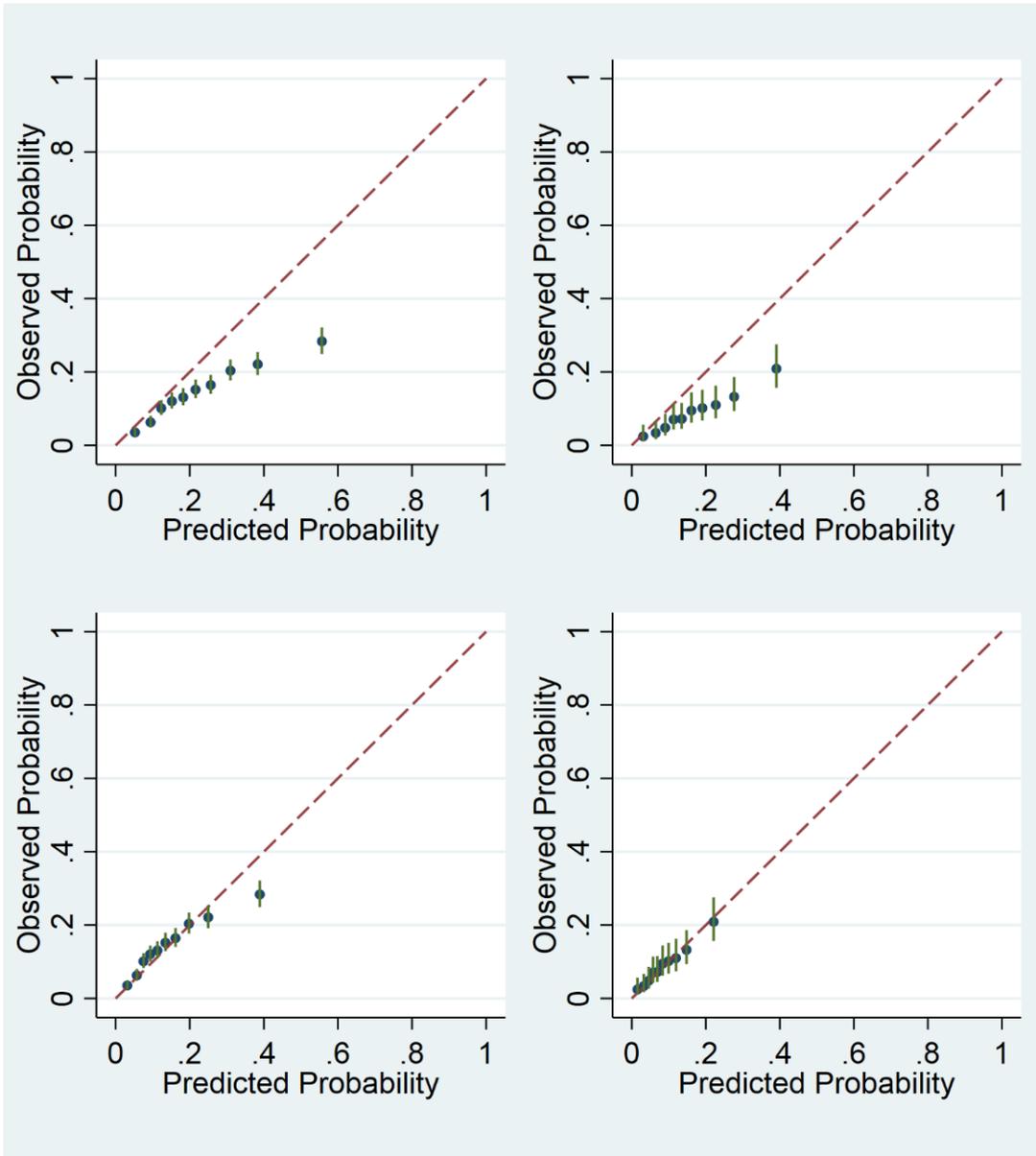
Figure 7.4: Calibration plots for **QRisk2 Female** for Whole Cohort (left column) and Cohort Specific (right column). Top row – original risk prediction model, bottom row – re-calibrated risk prediction model.
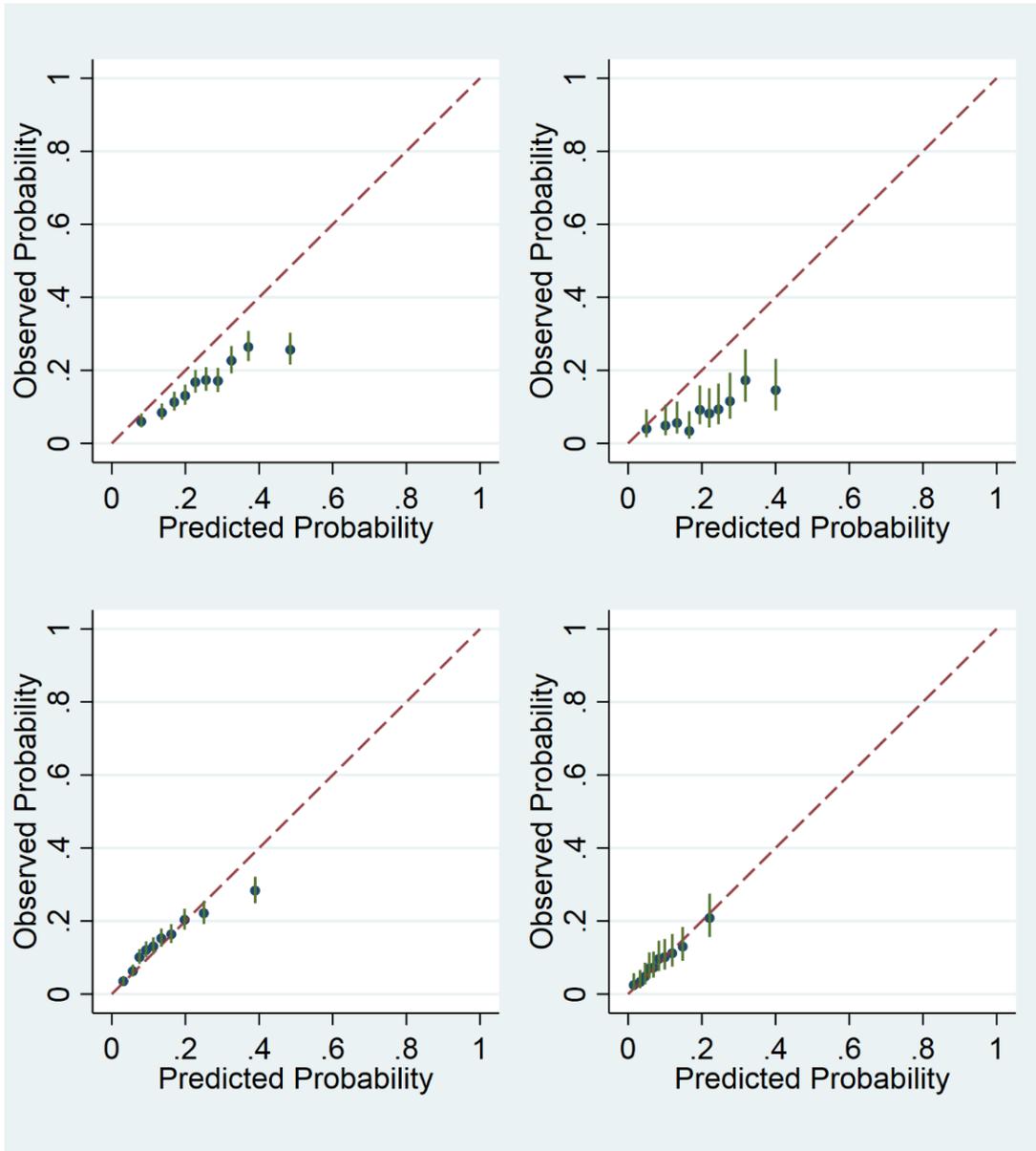
Figure 7.5: Calibration plots for **QRisk2 Male** for Whole Cohort (left column) and Cohort Specific (right column). Top row – original risk prediction model, bottom row – re-calibrated risk prediction model.
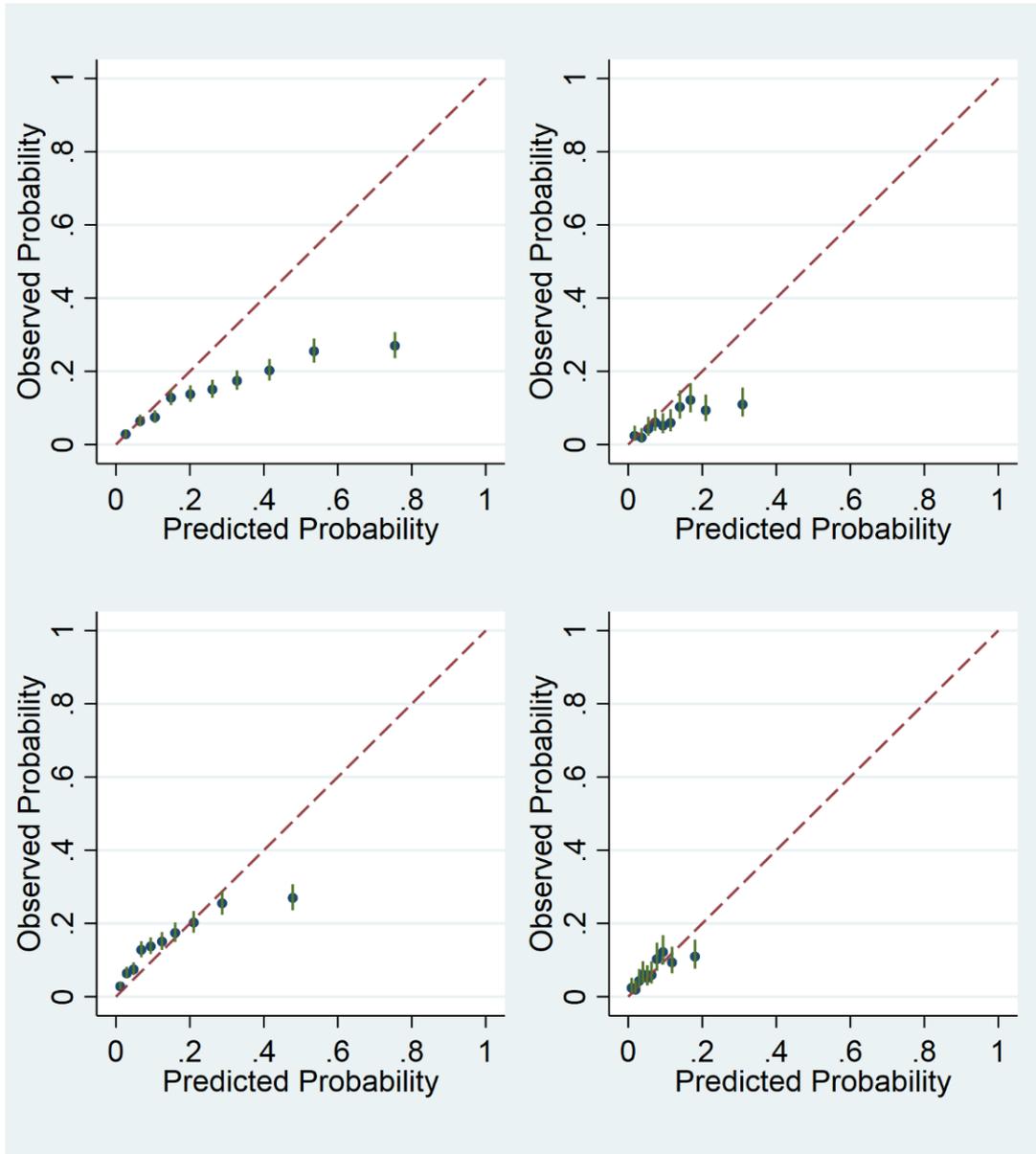
Figure 7.6: Calibration plots for **AHA Female** Whole Cohort (left column) and Cohort Specific (right column). Top row – original risk prediction model, bottom row – re-calibrated risk prediction model.
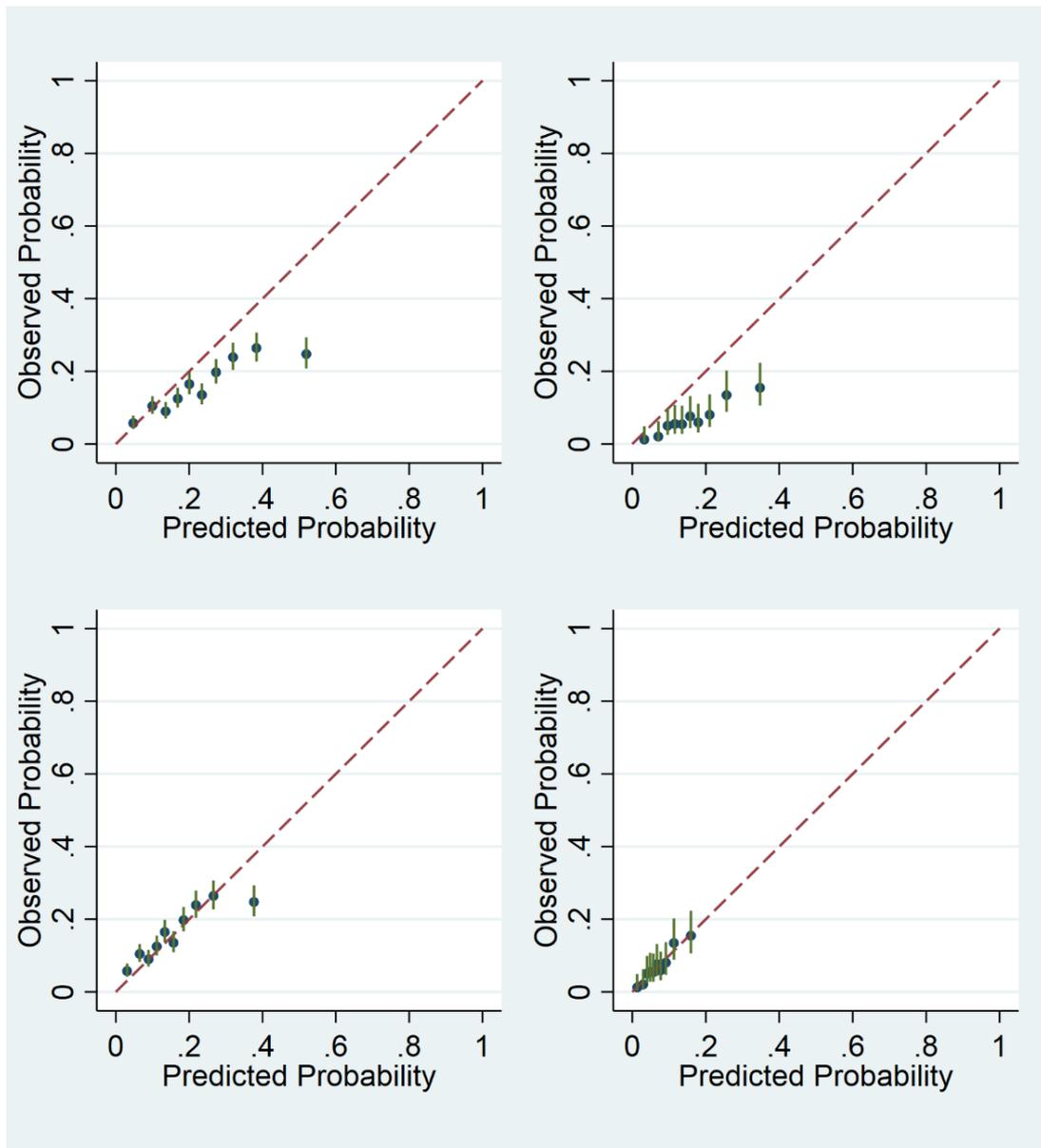
Figure 7.7: Calibration plots for **AHA Male** for Whole Cohort (left column) and Cohort Specific (right column). Top row – original risk prediction model, bottom row – re-calibrated risk prediction model.

7.4. Conclusions

Despite the plethora of CV risk prediction models external validation is rarely performed (65). Results of this process for CKD specific and general population models in the LCC cohort are presented in this chapter. The vast majority of models were split by gender. Whilst the Whole Cohort for male and female groups remained large with more than five hundred events each, the Cohort Specific groups were smaller, reflecting the exclusion of individuals with many of the co-morbidities associated with CKD. However, the groups probably remained adequate in size to allow for appropriate external validation of these identified models (119).

Predicted risk was probably higher, based on the mean LP values, in the LCC cohort compared to the development models' cohorts. However, the limited presentation of this information in models meant this could not be tested for all of them. This is likely to reflect a more co-morbid population in the LCC cohort compared to general population groups.

Discrimination was worse, as suggested by a lower value of C, in the LCC cohort than in the original development cohorts. This is a common finding in external validation and does not necessarily invalidate the use of the risk prediction model in clinical care (168). No CKD specific model substantially outperformed the general population models. Again, this may reflect that the CKD-specific models were developed in secondary care CKD groups and not in primary care as the LCC cohort was. All models had better discrimination in women than in men in the LCC cohort. This finding was consistent with the original development models. This general finding may be explained by a number of differences in CV disease pathophysiology in men compared to women. As an example, pre-menopausal women are at low risk of CV events (169), suggesting that all these women would have low predicted risk and actual events. This would likely lead to a substantial increase in the number of correct rankings, 'E' in equation 3.5, thus increasing the value of the C-statistic. This broad menopausal 'cut-off' for CV risk is not present in men and so the effect on increasing the C-statistic is less pronounced. These findings justifies

the development and validation of separate CV risk prediction models for the different genders.

As general population models showed equal or superior performance to CKD-specific models, calibration was assessed in detail in the former. Although LP suggested mean predicted risk was probably higher than the general population, calibration indicated that this reflected an overprediction of observed risk. This re-calibration saw a 28.6% increase in baseline risk for the whole cohort and 44.6% for the Cohort Specific group. The reasons for this change will be analysed, explored and discussed in the next chapter.

Overall, re-calibrated models showed improved graphical calibration despite ongoing sub-optimal discrimination compared to the original models. This probably suggested that although calibration was improved overall by adjustment of the baseline risk there continued to be ongoing sub-optimal calibration. This was most evident in the highest risk groups, where many of the 95% confidence intervals for the highest three deciles did not all cross the 45° line. This suggested mis-calibration was ongoing even after the adjustment of the five year baseline risk.

7.5. Summary

This chapter presents results of the external validation of risk prediction models identified in Chapter 6 in a CKD population, the LCC cohort. It assesses both models developed in CKD and general populations. The external validation assesses overall model performance through assessment of LP, discrimination and calibration. Analysis was performed using both complete case analysis and in the MI datasets. In addition, the models' performance in the whole LCC cohort, Whole Cohort, and in a population, Cohort Specific, matched to the original development datasets inclusion and exclusions criteria was assessed.

Generally, predicted risk was higher in the Whole Cohort groups than in the Cohort Specific subgroups. Discrimination was worse in the LCC cohort than in the development cohorts. No CKD specific model substantially outperformed the general population models assessed. General population models, with particular emphasis on QRisk3 Female, were calibrated using OE plots. Overall, risk was overpredicted by the tested models. This was more the case in the Cohort Specific than in the Whole Cohort group. Re-calibration of the baseline risk was able to adequately adjust the predicted risk to improve calibration, except for in some of the models highest predicted risk deciles. Mean change in baseline risk for QRisk3 Female was 28.6% for the Whole Cohort and 44.6% for the Cohort Specific subgroup.

As previously suggested, the LCC cohort probably represents a higher risk prediction group than the general population. However, this predicted risk probably is an overestimation of reality. The potential reasons for this are investigated in the next chapter.

Chapter 8

8.   Risk Prediction Model Adjustment

This chapter identifies and investigates the hypotheses regarding why CV risk prediction models may have performed less well in the LCC cohort compared to their original development cohorts. Firstly, I analysed whether assumptions of the current models were met, and if not, what could be done to adjust models for these issues. Secondly, the effect of outliers within the dataset is explored. Thirdly, I investigated whether differences between the cohorts may account for the differences. In the penultimate section I considered alterations to the models, based on some of the evidence base established in earlier chapters. Finally, I considered alternative approaches including adjustments to the Cox model. All results are again presented in full detail for the QRisk3 Female and QRisk3 Male models.

8.1. Introduction

Chapter 7 described the external validation of CV risk prediction models in the LCC cohort and showed the performance, measured through discrimination and calibration, was suboptimal in a CKD population. This chapter will explore the possible reason for these results, with a focus on the dataset and the models tested including their assumptions and limitations.

Firstly, I discuss the limitations of the data, as initially described in Chapter 4, and expand to include new issues raised by the external validation analysis. These limitations were noted in Chapter 4, particularly in relation to the routinely collected nature of the data and the limitations of how the CV outcomes were identified in relation to the linked secondary care sources. This will lead to exploration of the role of baseline risk calculations and re-calculations, as well as assessment of outliers in the dataset and their influence. Subsequently, this will lead to an analysis of the assumptions of the Cox model including its two main tenets, proportional hazards and non-informative censoring. If there is evidence for violation then the whole risk prediction model(s) may be invalid.

I then consider eGFR, ACR (74) and the variables identified in Chapter 5 as predictors of risk in the LCC cohort. Initially, their association with risk is tested in unadjusted analysis, then in adjusted analysis, before they were finally added to the risk prediction models. Their influence on risk prediction, through changes in discrimination and calibration, is assessed and then discussed. Subsequently, I investigate the assumptions of the conversion of baseline risk from a 10 to five year risk prediction timeframe. This is followed by assessment of other outcomes, including the role of 'competing risk'.

The models identified so far have considered the risk of CV events in isolation. Therefore they assess the risk of a CV event given that there is no risk of other events, such as death, that would terminate follow-up and prevent a CV event occurring. This simplification is likely to have an effect on any population but particularly those with a high risk of death in the follow-up period, such as in LCC. Finally alternatives outcomes, including competing risk, for the risk prediction model are assessed. The difference in the CKD population compared to the general population in relation to both increased CV and all-cause

mortality may account for some of this phenomenon and this will be investigated. The simplest adjustment of outcomes through prediction of a composite endpoint of CV events and mortality is first considered. Competing risk of death and its effect on CV event risk is also analysed. How and what effect it might have is addressed and alternative available approaches are then considered.

As no CKD cohort risk prediction model substantially outperformed general population tools in relation to discrimination and no baseline risks were reported, the CKD models were not considered further in this chapter. The results are therefore for the general population models, with a particular focus on the models most relevant to UK primary care, QRisk3 Female and QRisk3 Male.

8.2. Model Assumptions

All the models identified and tested used the Cox model to develop their respective risk prediction tools. As discussed in more detail in Chapter 2, the two main assumptions of this type of model are proportional hazards across time and non-informative censoring of individuals. In addition, the major assumption of the previous chapter for predicting risk was the estimation of the baseline risk for five year risk prediction. Each of these issues will be explored in turn.

8.2.1. Proportional Hazards

The proportional hazards assumption was tested using the methods described in Chapter 2. Results for all of the non-graphical tests described in this section are shown in Table 8.1. The 'link test' was used across the six models to screen models for violation of the proportional hazards assumption. The mean values for the coefficients for hat^2 across the imputed datasets for both the Whole Cohort and Cohort Specific groups were calculated. The coefficient values for QRisk3 Female suggested a possible difference from zero in the Whole Cohort (hat^2 co-efficient -0.297, 95% CI -0.456 to -0.138, p<0.001) but probably not in the Cohort Specific subgroup (hat^2 co-efficient -0.327, 95% CI -0.650 to -0.005, p=0.047). The group who would have been excluded from the original development model, i.e. Whole Cohort with the Cohort Specific individuals excluded, had evidence of violation (hat^2 co-efficient -0.309, 95% CI -0.531 to -0.086, p=0.007). The related model, QRisk2 Female, also showed some potential evidence of divergence from proportional hazards (hat^2 co-efficient -0.201, 95% CI -0.326 to -0.076, p=0.002). QRisk3 Male, and the other models, showed no such evidence of such an effect.

8.2.1.1. Statistical Assessments

Time varying interactions were tested for all models using the model's beta-coefficient interacting with time and log(time). There was limited evidence for interaction with log(time) in QRisk3 Female (p=0.030), which may have been accounted for due to multiple testing as Bonferroni adjusted p-values were non-significant. The strongest evidence of an interaction was in AHA Male (timeβ 0.111, 95% CI 0.044 to 0.179, p=0.001 and for log(time)β 0.114, 95% 0.035 to 0.192, p=0.004). As there was evidence for potential violation of proportional hazards in both hat^2 and in time interactions for QRisk3 Female, this model was selected for further investigation into this possible breach.

Model	hat^2 Coefficient	95% CI	p-value	Time Interaction	95% CI	p-value	Ln(Time) Interaction	95% CI	p-value
QRisk3 F	-0.297	-0.456 to -0.138	<0.001	0.029	-0.023 to 0.081	0.277	0.079	0.008 to 0.150	0.030
QRisk3 M	-0.162	-0.431 to 0.108	0.241	0.053	-0.043 to 0.148	0.280	0.079	-0.039 to 0.198	0.188
QRisk2 F	-0.201	-0.326 to -0.076	0.002	0.015	-0.038 to 0.069	0.542	0.053	-0.016 to 0.123	0.130
QRisk2 M	-0.061	-0.276 to 0.155	0.581	0.097	0.009 to 0.185	0.031	0.127	0.021 to 0.232	0.018
AHA F	-0.312	-0.879 to 0.256	0.282	-0.011	-0.049 to 0.027	0.568	0.011	-0.039 to 0.062	0.669
AHA M	0.019	-0.203 to 0.240	0.869	0.111	0.044 to 0.179	0.001	0.114	0.035 to 0.192	0.004

Table 8.1: Results of Tests of Proportional Hazard Assumptions for Whole Cohort for all Risk Prediction Models. F –female, M- male.

8.2.1.2.   Graphical Assessment

For the graphical assessments, I visually assessed the plots in all imputed datasets. For the purposes of brevity, results are presented for one dataset only. A different dataset from separate imputations is presented for each type of graphical assessment to demonstrate that results were not dependent on choice of the imputed dataset.

Schoenfeld residuals were then calculated and plotted against time for QRisk3 Female. These graphs were then visually inspected and evidence for non-zero slope gradient was tested. There was no evidence of deviation from the proportional hazards assumption using this method. An example plot of the residuals for each model using dataset eight in shown in Figure 8.1. The Chi^2 for the global test of proportional hazards were not significant for all models.

Log-log plots of survival were then made for each model and each imputed dataset. In QRisk3 Female, when follow-up time was less than log(-2) years (equivalent to less than two months) there was some evidence of violation proportional hazards, but data were sparse. When follow-up time increased beyond log(-2) years, there was no visual evidence of divergence from proportional hazards. Figure 8.2 shows these plots for QRisk3 Female using imputed dataset four as an example and different groups based on predicted risk. There was no significant difference in results between different datasets and using difference numbers of categories for risk.

Kaplan Meier observed survival was then plotted with the Cox model's predicted survival. Again, there was no visual evidence of major deviation from the proportional hazards assumption using this method, even with varying number of risk groups. In the highest risk group, lowest groups in the plots, there was perhaps some minor deviation from predicted hazard between two and four years of follow-up. Example plots for QRisk3 Female using dataset five are shown in Figure 8.3.
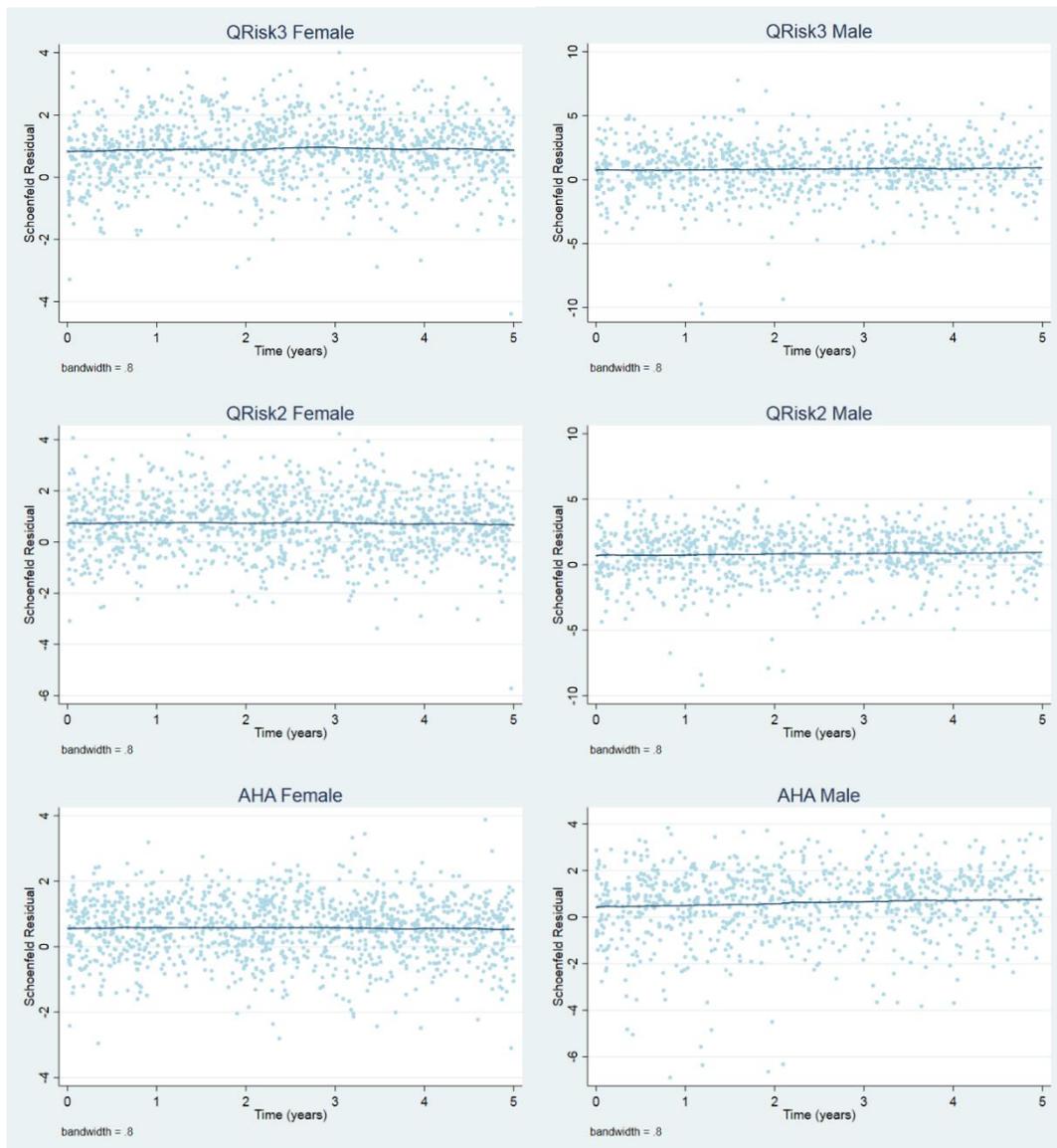
Figure 8.1: Schoenfeld Residuals plotted against Time for the Whole Cohort for all Risk Prediction Models using imputed Dataset 8 except AHA Male where only five datasets were impute and Dataset 5 is plotted. **Note difference in y-axis scales.**
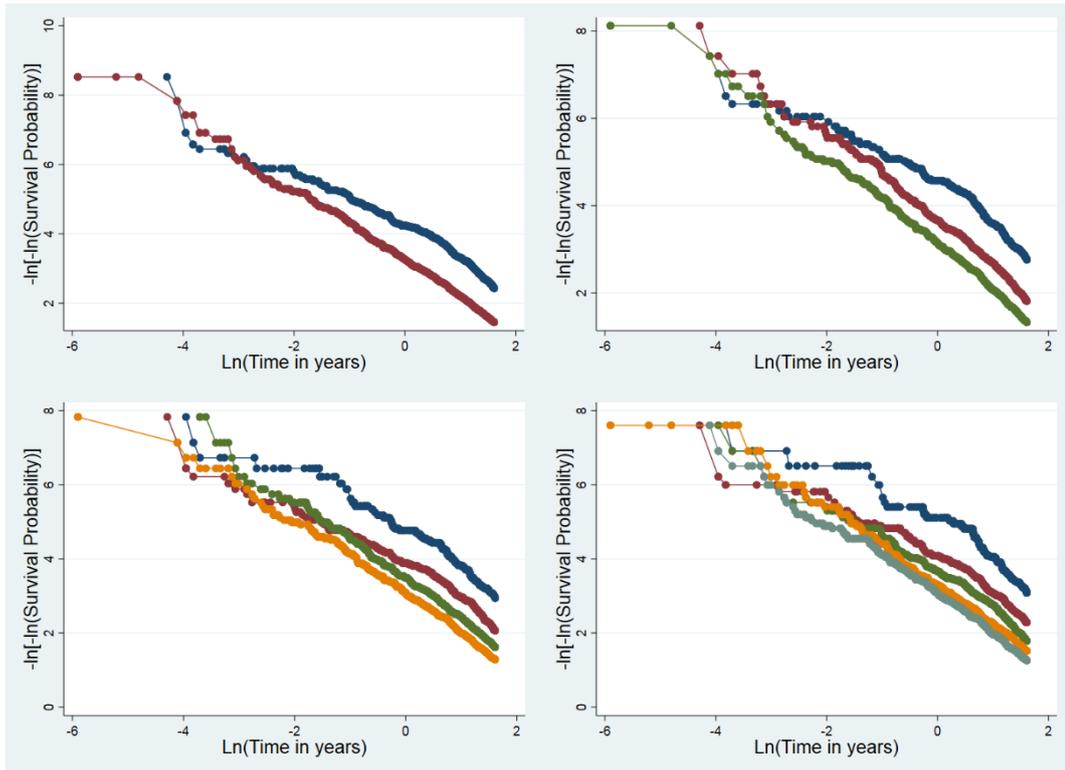
Figure 8.2: Log-log Plots of Survival for QRisk3 Female using different numbers of quantiles and Dataset 4 as an Example. **Note difference in y-axis scales.**
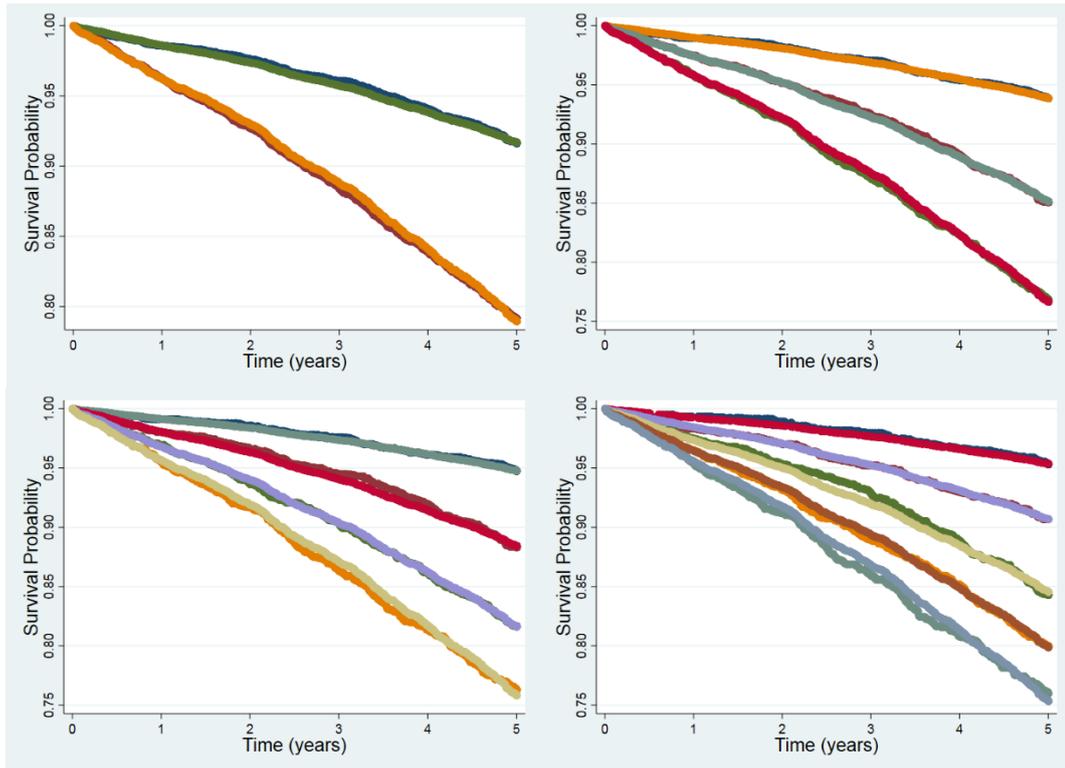
Figure 8.3: Graphs of Kaplan-Meier observed survival versus Cox Survival Functions for QRisk3 Female using different size quantiles Model using Imputed Dataset Five as an example. **Note difference in y-axis scales.**

8.2.2. Non-informative Censoring

Non-informative censoring of individuals in the LCC cohort would occur if individuals who were censored during follow-up were equally as likely to have experienced an event in the unobserved follow-up period as those who remained in the study. There are a number of potential reasons why an individual may have been censored prior to the end of the study, including death and moving away from the GP practice the individual was registered at to another practice that was not within the study. Assessment of non-informative censoring would require data beyond the censoring date of an individual, meaning that they would no longer be censored at that date. Therefore investigation of the non-informative censoring assumption is difficult. The literature in relation to formally assessing these assumptions is sparse and mainly focuses on clinical trial settings where 'progression free survival' of cancer is considered and may not have occurred at the exact time of the study's follow-up visit (170-172).

Formal assessment of methods to assess the non-informative censoring assumption are difficult to apply when a model has already been specified, such as in the case of external validation. However, it is possible to assess how many individuals could potentially be affected by this phenomenon and to categorise the potential reasons for censoring. In addition, the predicted risk of these individual with early censoring can be assessed to see if they differ from individuals censored at the end of the study.

More individuals were censored because they had died than because they had left the practice in both QRisk3 Female Whole Cohort (absolute difference 14.0%, 95% CI 13.0% to 15.0%, p<0.001) and Cohort Specific (absolute difference 5.7%, 95% CI 3.9% to 7.5%, p<0.001). In general terms, individuals who were censored because of death were likely to be more similar in predicted risk and age to those who were not censored, i.e. experienced a CV event during follow-up. Likewise, there were similarities in relation to these two variables when those who were censored at the end of the study were compared with those censored because they deregistered from the practice. Similar findings were found for QRisk3 Male for all the same groups of

censoring and for mean β co-efficient values and age. Table 8.2 details these results.

Therefore, if individuals censored because they died are older and at higher predicted risk than those censored for other reasons then it suggests that the assumption of non-informative censoring in the Cox model, used to develop the risk prediction models, has been violated.

Censoring	n (%)	Mean β (95% CI)	Difference in β	p-value for Difference	Mean Age	Linear Regression Co-efficient	p-value for Co-efficient
QRisk3 Female							
Whole Cohort							
End of study	5,714 (56.7%)	3.35 (3.33 to 3.36)	-	-	75.0 (74.7 to 75.2)	-	-
Death before CV Event	2,301 (22.9%)	3.88 (3.86 to 3.91)	0.54	<0.001	83.7 (83.4 to 84.1)	8.78	<0.001
Deregistered from practice	895 (8.9%)	3.52 (3.48 to 3.57)	0.18	<0.001	77.8 (77.1 to 78.4)	2.73	<0.001
Non-censored	1,161 (11.5%)	3.80 (3.77 to 3.83)	0.46	<0.001	81.8 (81.4 to 82.3)	6.87	<0.001
Cohort Specific							
End of study	1,688 (67.5%)	3.07 (3.04 to 3.11)	-	-	71.9 (71.4 to 72.4)	-	-
Death before CV Event	385 (15.4%)	3.55 (3.48 to 3.61)	0.47	<0.001	78.1 (77.3 to 78.8)	6.26	<0.001
Deregistered from practice	242 (9.7%)	3.16 (3.07 to 3.26)	0.09	0.065	72.7 (71.3 to 74.1)	0.69	0.278
Non-censored	186 (7.4)	3.56 (3.48 to 3.63)	0.48	<0.001	78.3 (77.4 to 79.2)	6.39	<0.001
QRisk3 Male							
Whole Cohort							
End of study	3,596 (53.8%)	2.93 (2.91 to 2.95)	-	-	73.4 (73.1 to 73.7)	-	-
Death before CV Event	1,723 (25.8%)	3.30 (3.28 to 3.32)	0.37	<0.001	81.5 (81.1 to 81.9)	8.08	<0.001
Deregistered from practice	508 (7.6%)	3.05 (3.00 to 3.11)	0.12	<0.001	75.5 (74.6 to 76.3)	1.89	<0.001
Non-censored	863 (12.9%)	3.23 (3.20 to 3.26)	0.29	<0.001	78.7 (78.2 to 79.4)	5.37	<0.001
Cohort Specific							
End of study	882 (64.7%)	2.78 (2.74 to 2.83)	-	-	71.4 (70.8 to 72.1)	-	-
Death before CV Event	246 (18.0%)	3.14 (3.09 to 3.19)	0.36	<0.001	78.3 (77.4 to 79.2)	6.83	<0.001
Deregistered from practice	136 (10.0%)	2.84 (2.71 to 2.96)	0.05	0.378	71.8 (69.7 to 73.9)	-0.03	0.974
Non-censored	100 (7.3%)	3.09 (2.98 to 3.20)	0.31	<0.001	76.7 (75.0 to 78.5)	5.33	<0.001

Table 8.2: Difference in Censored Individuals for QRisk3 Female for Predicted Risk (β) and Age by Reason for Censoring. 'Difference' uses those censored at the end of the study as the reference category. Means refer to mean values across all imputed datasets. Non-censored refers to individuals experiencing a CV event.

8.3.   Baseline Risk Re-estimation

For the purposes of the estimation of risk over a five year period the 10 year baseline hazard was adjusted as set out in Section 7.2.5. The initial assumption when assessing calibration was that the change in risk was directly proportional to the amount of time i.e. that the risk for five years was exactly half of that of the risk for 10 years. This section performs sensitivity analysis for this assumption.

Firstly, an exponential method is used, as set out in equations 8.1, 8.2 and 8.3 below. Secondly, the role of event identification is consider and its potential influence on baseline risk. This includes an estimation of the proportion of events identified in data sources not linked to LCC that would be required in order for the models to be well calibrated. This part of the analysis utilises the data previously presented by Herrett *et al* (81).

8.3.1. Exponential Method for Re-estimation

Instead of a simple halving of risk, using an exponential calculation of risk is based on:

$$R_0(x_1) = R_0(x_2)^{x_1/x_2} \qquad\qquad (8.1)$$

Where R_0 is the baseline risk, x_1 is the new time period for risk calculation and x_2 is the old time period. For this sensitivity analysis where risk is change from 10 to five years this equates to:

$$R_0(5) = R_0(10)^{5/10} \qquad\qquad (8.2)$$

Or more simply:

$$R_5(x_1) = \sqrt[2]{R_{10}(x_2)} \qquad\qquad (8.3)$$

Table 8.3 shows the recalculated mean risks and differences using this method for risk calculation. As only the baseline risk has changed and βs remain the same, the ranking of individuals does not change and therefore C does not change. Figure 8.4 shows a histogram of the change in risk for a single dataset of QRisk3 Female and QRisk3 Male. There was a bigger difference in change in mean predicted risk difference for male models than for female models. There were no clinically significant changes between imputed datasets of models assessed. When calibration plots were visually inspected there were no substantial differences for risk calculated through these two different baseline risk estimator methods. Figure 8.5 shows a comparison of QRisk3 Female and QRisk3 Male. There was no significant difference between changes in risk or calibration plots between Whole Cohort and Cohort Specific groups.

Risk Model	Original Baseline R_{10}	Linear Calculated R_5	Exponential Calculated R_5	Percentage Risk Difference	Mean (95%CI) of Calculated Risk Difference
QRisk2 Female	0.9897	0.9949	0.99490	-0.048	-0.048 to -0.048
QRisk2 Male	0.9788	0.9894	0.98934	-0.112	-0.113 to -0.111
QRisk3 Female	0.9889	0.9944	0.99442	-0.048	-0.048 to -0.047
QRisk3 Male	0.9773	0.9886	0.98857	-0.116	-0.117 to -0.115
AHA Female	0.9665	0.9833	0.98311	-0.170	-0.172 to -0.168
AHA Male	0.9144	0.9572	0.95624	-0.442	-0.447 to -0.437

Table 8.3: Difference in Baseline Risk and Calculated Risk by using Exponential Calculation for Baseline Risk Conversion. Baseline risks are shown to 4 decimal places, but all calculations were made using as many decimal places as provided by the published model.
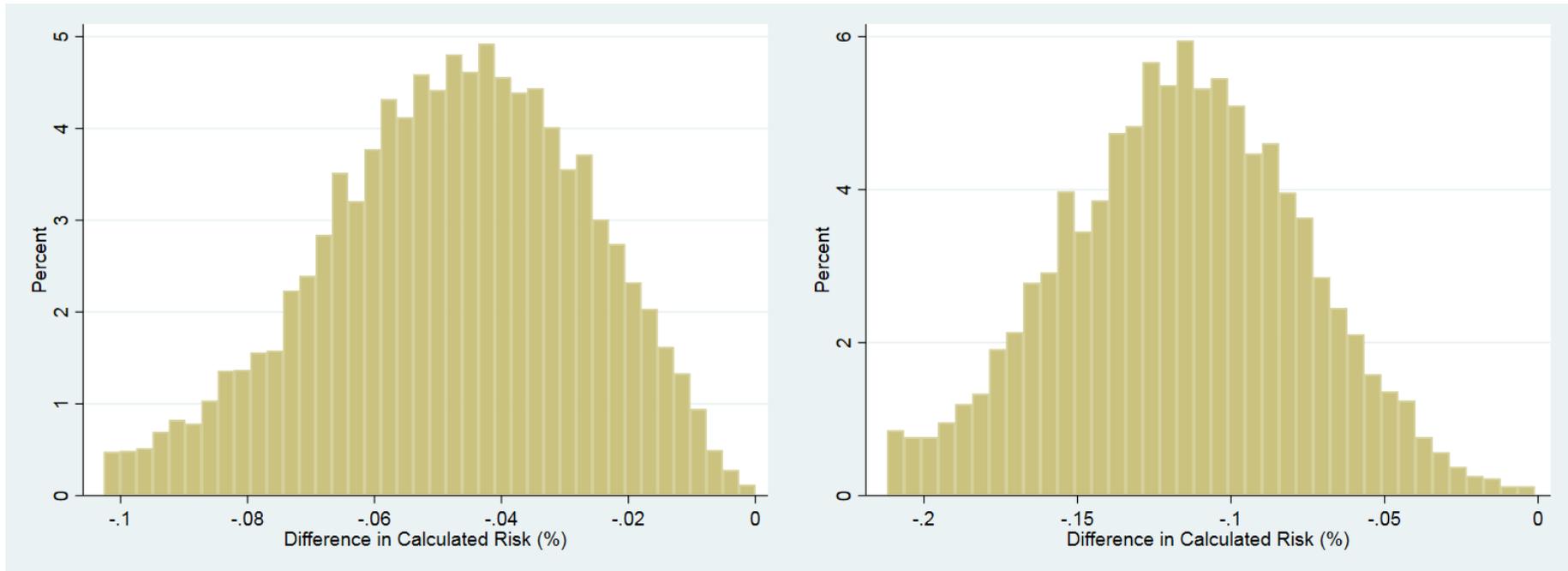
Figure 8.4: Difference in Calculated Risk Percentage between originally calculated 5 Year Risk and 5 Year Risk using the exponential methods for baseline risk. Results shown are for Dataset 3 for QRisk3 Female (left) and QRisk3 Male (right). **Note difference in x and y-axes' scales.**
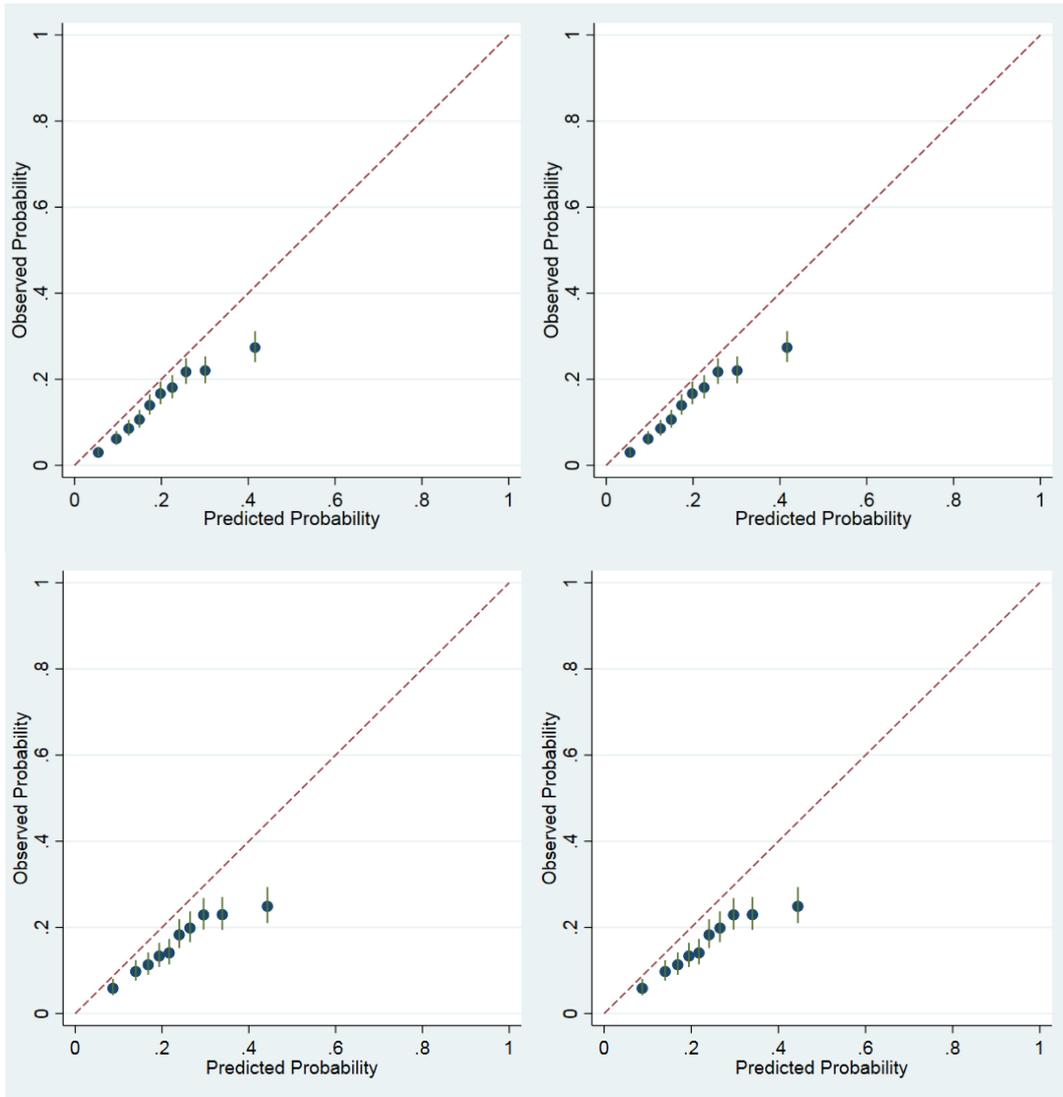
Figure 8.5: Whole Cohort Calibration Plots for QRisk3 Female (top row) and QRisk3 Male (bottom row) using the Original Baseline Risk Calculation (left of figure) and the Exponential Baseline Risk Calculation (right of figure).

8.3.2. Sensitivity Analysis for Baseline Risk Re-estimation

As a reminder, compared to the QRisk2 and QRisk3 models, LCC was not linked to ONS data for CV mortality but did have linked local MINAP data. The former may have accounted for 3.0% of fatal and non-fatal myocardial infarctions and the latter 8.0% of fatal and non-fatal myocardial infarctions. As the HES and MINAP data are locally sourced and not nationally, a secondary care event occurring outside of University Hospitals of Leicester and not recorded in the primary care notes, would have not been identified for the study.

Sensitivity analysis was performed to calculate the proportion of additional events in the LCC cohort that would be required, i.e. not currently identified, in order for the original model to be well calibrated:

$$extra\ events\ required\ \% = \frac{(mean\ predicted\ risk\ \%) - (mean\ observed\ risk\ \%)}{mean\ observed\ risk\ \%} \qquad (8.4)$$

Where $h_0(orig)$ refers to the original baseline hazard, $h_0(recal)$ to the recalibrated baseline hazard.

As can be seen for Table 8.4, the estimated percentage of events required to have been missed during event identification would have to be substantially higher than the data from Herrett *et al* would suggest may have been missed (81). Therefore, it is unlikely that misidentification of events is the sole reason for poor calibration in the external validation of the model in the LCC cohort.

Model	Mean Predicted Risk	Mean Observed Risk	Difference in Mean Risk	95% CI	Events Required	95% CI
Whole Cohort						
QRisk3 Female	17.5	14.7	2.8	1.9 to 3.6	18.7	12.9 to 24.3
QRisk3 Male	21.9	16.8	5.1	3.9 to 6.2	30.2	23.3 to 36.6
Cohort Specific						
QRisk3 Female	12.7	8.7	4.0	2.7 to 5.2	46.5	31.2 to 59.8
QRisk3 Male	18.4	8.8	9.6	7.7 to 11.1	109.4	88.0 to 127.1

Table 8.4: Comparison of predict and observed risk. Events required refers to the additional percentage of events that would be required for the mean observed risk to match the mean predicted risk. All figures are for percentage risk over 5 years.

8.4. Influence of Outliers

The influence of outliers within the dataset was considered next. Firstly, the difference of β co-efficients when an individual was removed was assessed ('DFBETA'). This represents an approximation of the difference in the model's estimated parameter β for the whole dataset versus an estimate of β for every dataset with a single observation removed in turn. As the computational requirements for this is high, for instance when the size and the number of imputed datasets is accounted for in this thesis over a quarter of a million β values per model would require calculation for some models. DFBETA is an approximation of this value and is calculated through the product of the variance-covariance matrix of β and the efficient score residuals. IT represents the change in the value of the β co-efficient if the particular individual was removed from the dataset.

Secondly, likelihood displacement values were calculated. This compares the difference in the model's value for the log likelihood when an individual is removed as opposed to the full dataset. The difference in the two values will be small if the value removed is not influential on the two vectors of the coefficient. The maximum value of the likelihood displacement value (LMax) calculates the largest value of displacement of β's vector caused by, again, the deletion of a single value from the dataset.

These three methods are assessed through visual inspection of their values plotted against time. There may be some subjectivity to their interpretation. Figure 8.6 shows the results for DFBETA, likelihood displacement and LMax values for QRisk3 Female and QRisk3 Male.

Four individuals identified from these plots as potential outliers for QRisk3 Female. They were evaluated for their reason for influence and considered for possible removal from the dataset. Two individuals had the two highest values of β co-efficient, which was attributable to a combination of both high level of blood pressure variability and extreme low values of HDL cholesterol, which subsequently greatly increased the value of the TC:HDL ratio, a variable used in all QRisk models. This issue has been previously noted by the authors of the original QRisk model when the original model was published where imputation

of individual components of ratios led to some extreme values of TC:HDL cholesterol. Another had a value of β in the lowest ten values and experienced a CV event ten days before the end of the study, when a high number of individuals were censored. The final individual had a CV event within 10 days of the follow-up period starting and had a β value in the lowest 1%. Similarly, in QRisk3 Male up to ten individuals were visually noted to be significant outliers across the plots. All had had a CV event during follow-up and most within two years of the study commencing. In addition, all had extreme values β which was typically explained by extreme values of blood pressure variability, HDL and/or age.

Given that these individuals represented in the region of one in a thousand individuals in the cohort, they were unlikely to influence the overall results of the external validation in relation to discrimination and calibration. Therefore, exploratory analysis was performed in one QRisk3 Female dataset and no influence on the values of C or calibration of the model were found (data not shown).
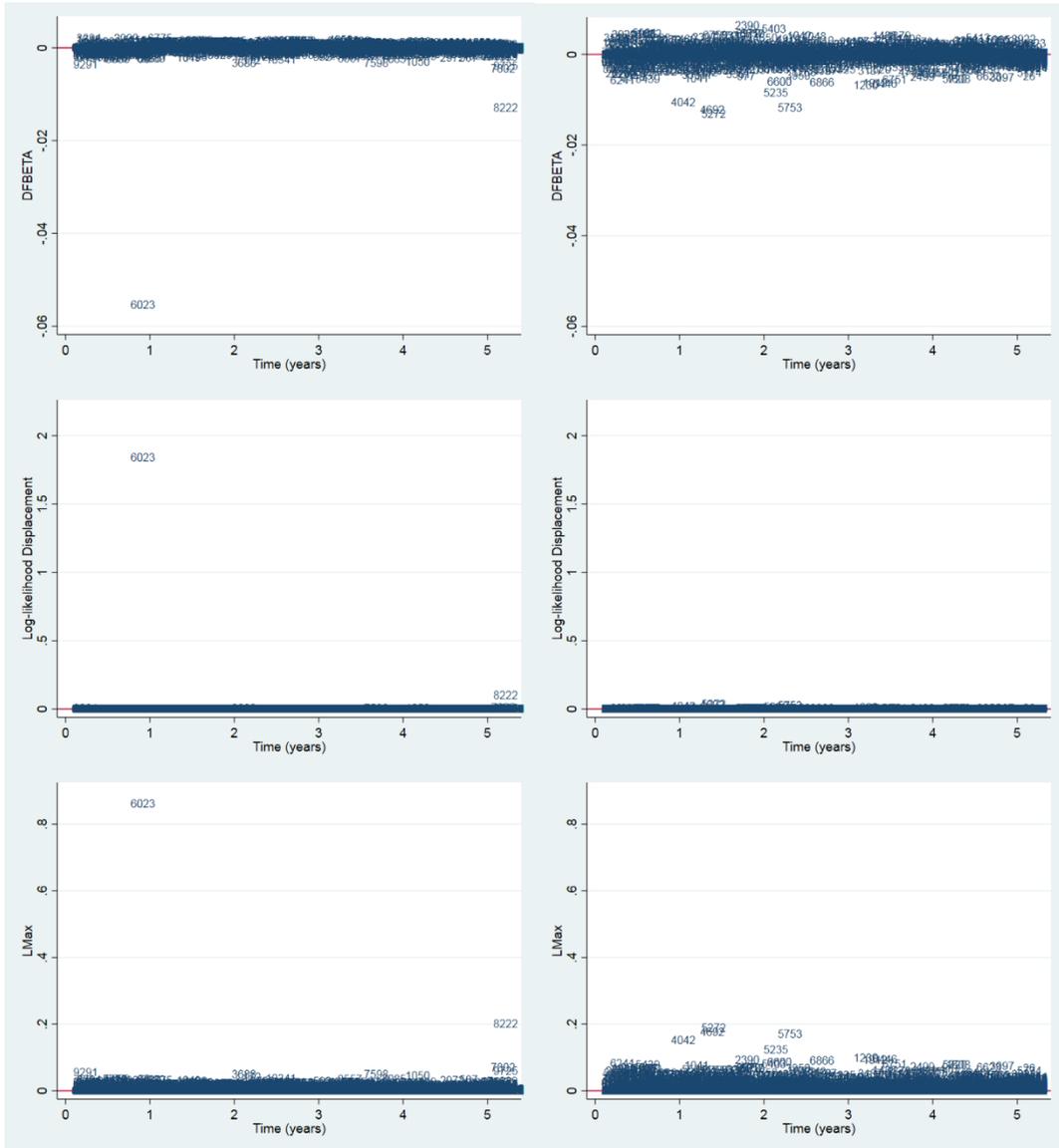
Figure 8.6: DFBETA, Log-likelihood and LMax for Individuals in Female (left column) and Male (right column) QRisk3 Female. Model using Imputed Dataset Five as an example. All graphs underwent jittering to prevent over plotting so exact positioning on graphs, particularly for time may not accurately represent their values.

8.5. Addition of Additional Co-variables

The addition of further co-variables is now considered in order to improve the performance of the models. Firstly, the co-variables were selected based on previous individual patient data meta-analysis evidence that the kidney markers of eGFR and urine ACR improved prediction. Secondly, the evidence collated in Chapter 5, meant that the additional markers of serum albumin, haemoglobin and phosphate were tested. eGFR and ACR were non-normally distributed variables and were log transformed and centred around their mean before being used for analysis. Initially their prediction of CV events in unadjusted analysis was tested, then individually added to each risk prediction tool, followed by the addition of both eGFR and ACR, and finally the addition of all five risk co-variables. Tables 8.4 and 8.5 summarises the results for QRisk3 Female and QRisk3 Male for the analysis in the following three sections.

8.5.1. Unadjusted Analysis

In unadjusted analysis all five variables were statistically significant in their association with CV event prediction (p<0.001 for all variables) in QRisk3 Female Whole Cohort. In the Cohort Specific subcohort only eGFR and albumin were associated with statistically significant increased risk (p<0.01 for both). Similar results were found for QRisk3 Male, except that ACR was the only statistically significant variable in the Cohort Specific group.

8.5.2. Adjusted Analysis

In QRisk3 Female Whole Cohort, all variables, except for haemoglobin (HR 0.997, 95% CI 0.992 to 1.001, p=0.180), were associated with risk when added to a model including β sum from the original risk prediction model. In the Cohort Specific group, only log(eGFR) was associated with increased risk of CV events (HR 0.338, 95% CI 0.188 to 0.605, p=0.180). For QRisk3 Male Whole Cohort, similar patterns to the female model were seen except that phosphate, instead of haemoglobin, was the only variable not associated with risk (HR 1.375, 95%

CI 0.888 to 2.129, p=0.153). In the Cohort Specific group, only log(ACR) was associated with risk (HR 1.234, 95% CI 1.061 to 1.435, p=0.006).

All five variables were then added to the model with β sum. Again, haemoglobin (HR 1.003, 95% CI 0.999 to 1.008, p=0.163) was the only variable not statistically significant in its relationship with CV events in QRisk3 Female Whole Cohort. In the cohort specific group, only log(eGFR) was associated with risk (HR 0.327, 95% CI 0.177 to 0.604, p<0.001). In analysis of QRisk3 Male Whole Cohort, only log(ACR) (HR 1.136, 95% CI 1.083 to 1.192, p<0.001) and albumin (HR 0.972, 95% CI 0.949 to 0.996, p=0.022) were associated with risk. In the Cohort Specific group, log(ACR) (HR 1.292, 95% CI 1.095 to 1.526, p=0.003) was the only variable associated with risk. Throughout the adjusted analysis β sum remained associated with risk in both models and in Whole Cohort and Cohort Specific analysis.

Altered Q3Risk3 FEMALE Model	Unadjusted			Model β with Variable			Model β with 5 Variables		
	HR	95% CI	p-value	HR	95% CI	p-value	HR	95% CI	p-value
Whole Cohort									
Ln(ACR)	1.165	1.117 to 1.214	<0.001	1.125	1.077 to 1.175	<0.001	1.098	1.052 to 1.145	<0.001
Ln(eGFR)	0.404	0.337 to 0.483	<0.001	0.512	0.420 to 0.625	<0.001	0.616	0.498 to 0.762	<0.001
Albumin	0.922	0.904 to 0.940	<0.001	0.947	0.929 to 0.966	<0.001	0.951	0.932 to 0.971	<0.001
Haemoglobin	0.991	0.987 to 0.996	<0.001	0.997	0.992 to 1.001	0.180	1.003	0.999 to 1.008	0.163
Phosphate	2.013	1.414 to 2.867	<0.001	1.965	1.357 to 2.845	<0.001	1.777	1.212 to 2.606	0.003
Ln(ACR) with Ln(eGFR)*	1.127	1.081 to 1.175	<0.001	1.104	1.057 to 1.152	<0.001	-	-	-
Ln(eGFR) with Ln(ACR)*	0.478	0.395 to 0.579	<0.001	0.577	0.470 to 0.709	<0.001	-	-	-
Cohort Specific									
Ln(ACR)	1.020	0.926 to 1.125	0.683	1.051	0.955 to 1.156	0.308	0.991	0.906 to 1.084	0.839
Ln(eGFR)	0.270	0.157 to 0.463	<0.001	0.338	0.188 to 0.605	<0.001	0.327	0.177 to 0.604	<0.001
Albumin	0.930	0.884 to 0.979	0.005	0.953	0.904 to 1.004	0.068	0.948	0.899 to 1.000	0.051
Haemoglobin	0.999	0.988 to 1.010	0.844	1.003	0.992 to 1.015	0.570	1.011	0.999 to 1.022	0.078
Phosphate	1.810	0.780 to 4.199	0.167	1.847	0.756 to 4.512	0.178	1.750	0.712 to 4.305	0.222
Ln(ACR) with Ln(eGFR)*	1.013	0.927 to 1.106	0.780	0.997	0.909 to 1.093	0.944	-	-	-
Ln(eGFR) with Ln(ACR)*	0.276	0.157 to 0.486	<0.001	0.336	0.184 to 0.612	<0.001	-	-	-

Table 8.5: Unadjusted and adjusted hazard ratios for variables using the QRisk3 Female model. 'Model β' columns refer to the hazard ratio for the variable when an adjusted model was used with the value of β for an individual from QRisk3 Female was included. Unadjusted for this line refers to adjustment for the 'with' second kidney marker.

Altered Q3Risk MALE Model	Unadjusted			Model β with Variable			Model β with 5 Variables		
	HR	95% CI	p-value	HR	95% CI	p-value	HR	95% CI	p-value
Whole Cohort									
Ln(ACR)	1.174	1.123 to 1.227	<0.001	1.148	1.098 to 1.200	<0.001	1.136	1.083 to 1.192	<0.001
Ln(eGFR)	0.624	0.503 to 0.774	<0.001	0.723	0.572 to 0.912	0.006	1.062	0.808 to 1.397	0.666
Albumin	0.945	0.924 to 0.967	<0.001	0.957	0.935 to 0.980	<0.001	0.972	0.949 to 0.996	0.022
Haemoglobin	0.989	0.985 to 0.993	<0.001	0.994	0.989 to 0.998	0.004	0.996	0.991 to 1.001	0.145
Phosphate	1.583	1.039 to 2.410	0.038	1.375	0.888 to 2.129	0.153	1.174	0.735 to 1.876	0.500
Ln(ACR) with Ln(eGFR)*	1.162	1.108 to 1.219	<0.001	1.144	1.091 to 1.200	<0.001	-	-	-
Ln(eGFR) with Ln(ACR)*	0.843	0.662 to 1.073	0.164	0.942	0.730 to 1.122	0.645	-	-	-
Cohort Specific									
Ln(ACR)	1.255	1.086 to 1.449	0.002	1.234	1.061 to 1.435	0.006	1.292	1.095 to 1.526	0.003
Ln(eGFR)	1.416	0.552 to 3.634	0.470	1.832	0.632 to 5.311	0.265	2.878	0.892 to 9.283	0.077
Albumin	1.001	0.933 to 1.073	0.986	1.023	0.951 to 1.100	0.547	1.027	0.951 to 1.109	0.496
Haemoglobin	0.999	0.986 to 1.012	0.842	1.003	0.990 to 1.017	0.607	1.002	0.987 to 1.017	0.805
Phosphate	1.206	0.364 to 3.995	0.759	1.228	0.393 to 3.834	0.724	1.372	0.401 to 4.692	0.613
Ln(ACR) with Ln(eGFR)*	1.301	1.112 to 1.523	0.001	1.285	1.090 to 1.514	0.003	-	-	-
Ln(eGFR) with Ln(ACR)*	2.413	0.863 to 6.750	0.093	2.988	0.968 to 9.220	0.057	-	-	-

Table 8.6: Unadjusted and adjusted hazard ratios for variables using the QRisk3 Male model. 'Model β' columns refer to the hazard ratio for the variable when an adjusted model was used with the value of β for an individual from QRisk3 Male was included. Unadjusted for this line refers to adjustment for the 'with' second kidney marker.

8.5.3. Change in Discrimination

The change in C ('ΔC') by adding variables to the original model of QRisk3 Female was then assessed for both Whole Cohort and Cohort Specific groups. Table 8.6 presents these findings. The point mean estimate for all ΔC were positive i.e. C was improved with addition of the variables. The addition of ACR and eGFR have been previously shown to improve CV risk prediction in individual participant data meta-analysis (74), so ΔC was first tested with the additional of these two kidney markers. Neither were able to show a statistically significant improvement in ΔC in either Whole Cohort or Cohort Specific, but the magnitude of ΔC was of similar size to the previously published work. The addition of phosphate and albumin to kidney markers in the model led to a further increase in ΔC that was statistically significant in Whole Cohort but not Cohort Specific. Additional small improvement was gained from adding haemoglobin to the model. For QRisk3 Male, there were again similar improvements in the point estimates for ΔC to those previously reported, but none were statistically significant.

Updated Model	ΔC	95% CI	p-value
Q3 FEMALE			
Whole Cohort			
Ln(eGFR) & Ln(ACR)	0.0167	0.0021 to 0.0313	0.114
Plus Phosphate & Albumin	0.0227	0.0081 to 0.0374	0.040
Plus Haemoglobin	0.0234	0.0087 to 0.0381	0.035
Cohort Specific			
Ln(eGFR) & Ln(ACR)	0.0144	-0.0225 to 0.0513	0.344
Plus Phosphate & Albumin	0.0172	-0.0197 to 0.0541	0.324
Plus Haemoglobin	0.0211	-0.0165 to 0.0586	0.293
Q3 MALE			
Whole Cohort			
Ln(eGFR) & Ln(ACR)	0.0140	-0.0047 to 0.0326	0.231
Plus Phosphate & Albumin	0.0158	-0.0030 to 0.0346	0.200
Plus Haemoglobin	0.0158	-0.0030 to 0.0346	0.199
Cohort Specific			
Ln(eGFR) & Ln(ACR)	0.0366	-0.0242 to 0.0973	0.272
Plus Phosphate & Albumin	0.0384	-0.0218 to 0.0986	0.260
Plus Haemoglobin	0.0391	-0.0287 to 0.0990	0.256

Table 8.7: Change in C-Statistic with Addition of New Variables for QRisk3 Female and QRisk3 Male: 'ΔC' refers to increase in C when for the new variable the product of the β co-efficient and its value is added to the β value calculated from the original model. p-value refers to interaction between QRisk3 Female model and updated model.

8.5.4. Change in Calibration

In order to assess the change in calibration between the adjusted models calibration plots were plotted and compared. Figures 8.7, 8.8, 8.9 and 8.10 display these plots. There was no substantial improvement in risk prediction in either QRisk3 Female or QRisk3 Male and no difference for Whole Cohort or Cohort Specific groups. All models continued to overpredict risk, more so in higher deciles of risk. There was some increased spread of predicted risk, with again higher risk deciles increasing their mean risk predictions more than lower risk groups. For QRisk3 Female, the plots of the deciles became visually more linear, suggesting that there may have been minor improvements in risk prediction ranking in keeping with the improved C-statistic. This did not occur to such a degree in the QRisk3 Male adjusted models. Other than some of the lowest risk groups, most 95% CI did not include the 45° line and therefore, there was no evidence for good calibration of the models.
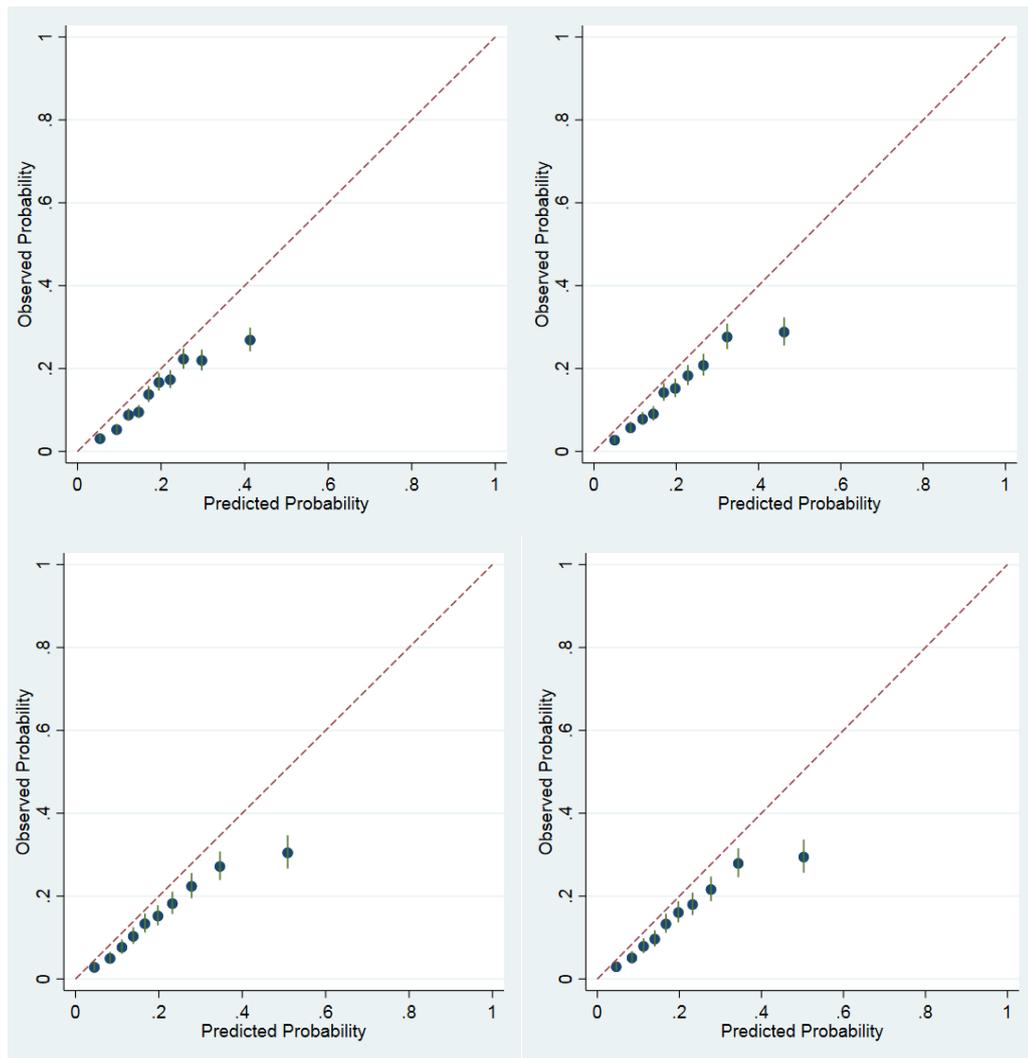
Figure 8.7: Calibration Plots for Whole Cohort original QRisk3 Female model (top left), with addition of eGFR and ACR (top right), eGFR, ACR, Albumin and Phosphate (bottom left) and eGFR, ACR, Albumin, Phosphate and haemoglobin (bottom right).
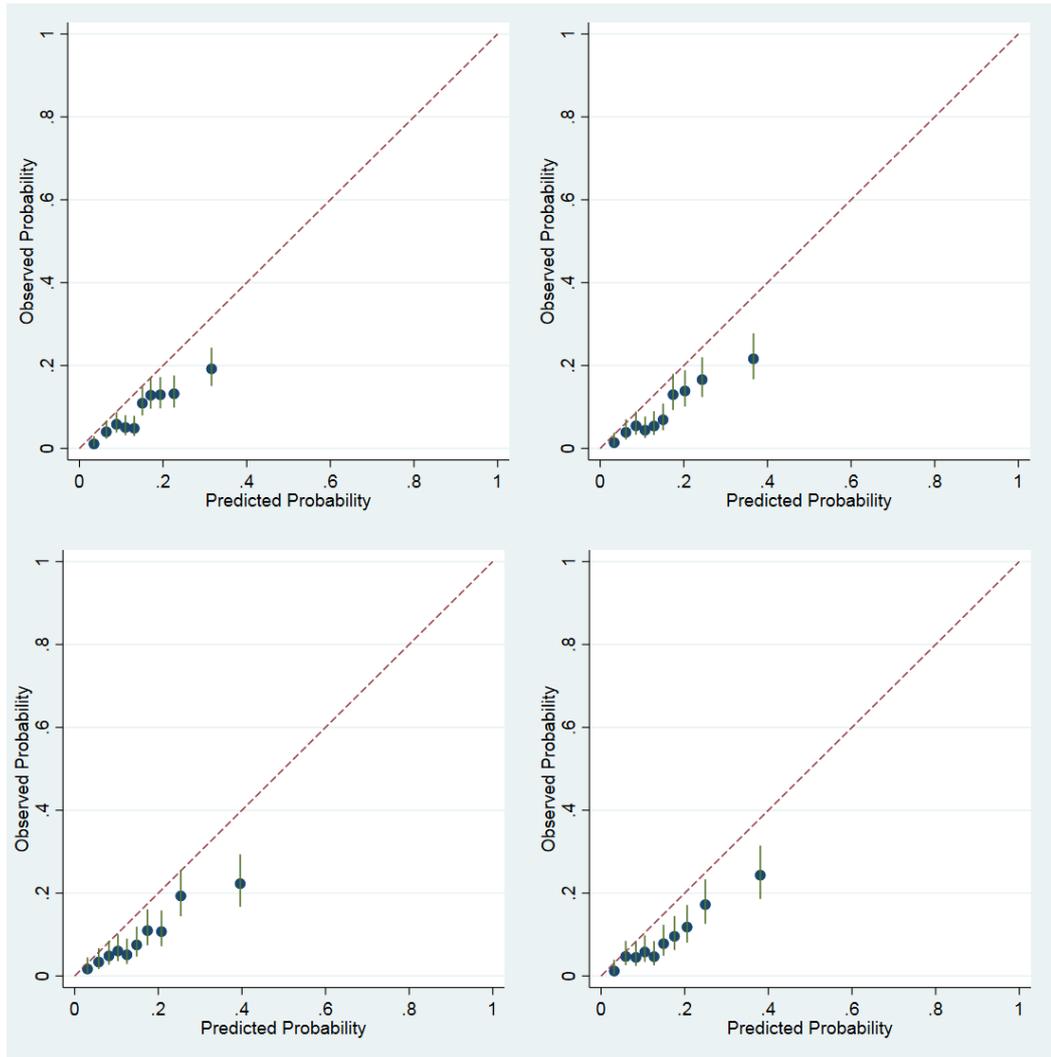
Figure 8.8: Calibration Plots for Cohort Specific original QRisk3 Female model (top left), with addition of eGFR and ACR (top right), eGFR, ACR, Albumin and Phosphate (bottom left) and eGFR, ACR, Albumin, Phosphate and haemoglobin (bottom right).
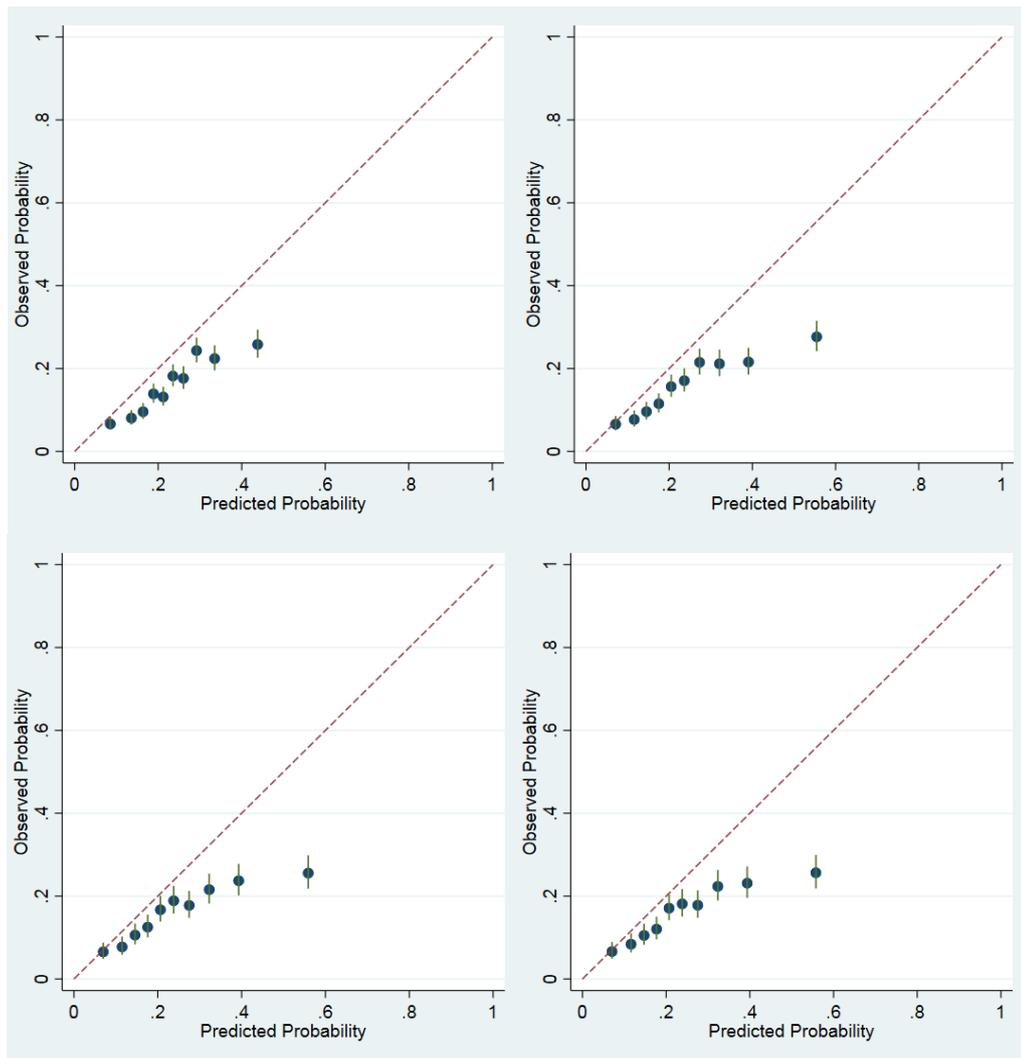
Figure 8.9: Calibration Plots for Whole Cohort original QRisk3 Male model (top left), with addition of eGFR and ACR (top right), eGFR, ACR, Albumin and Phosphate (bottom left) and eGFR, ACR, Albumin, Phosphate and haemoglobin (bottom right).
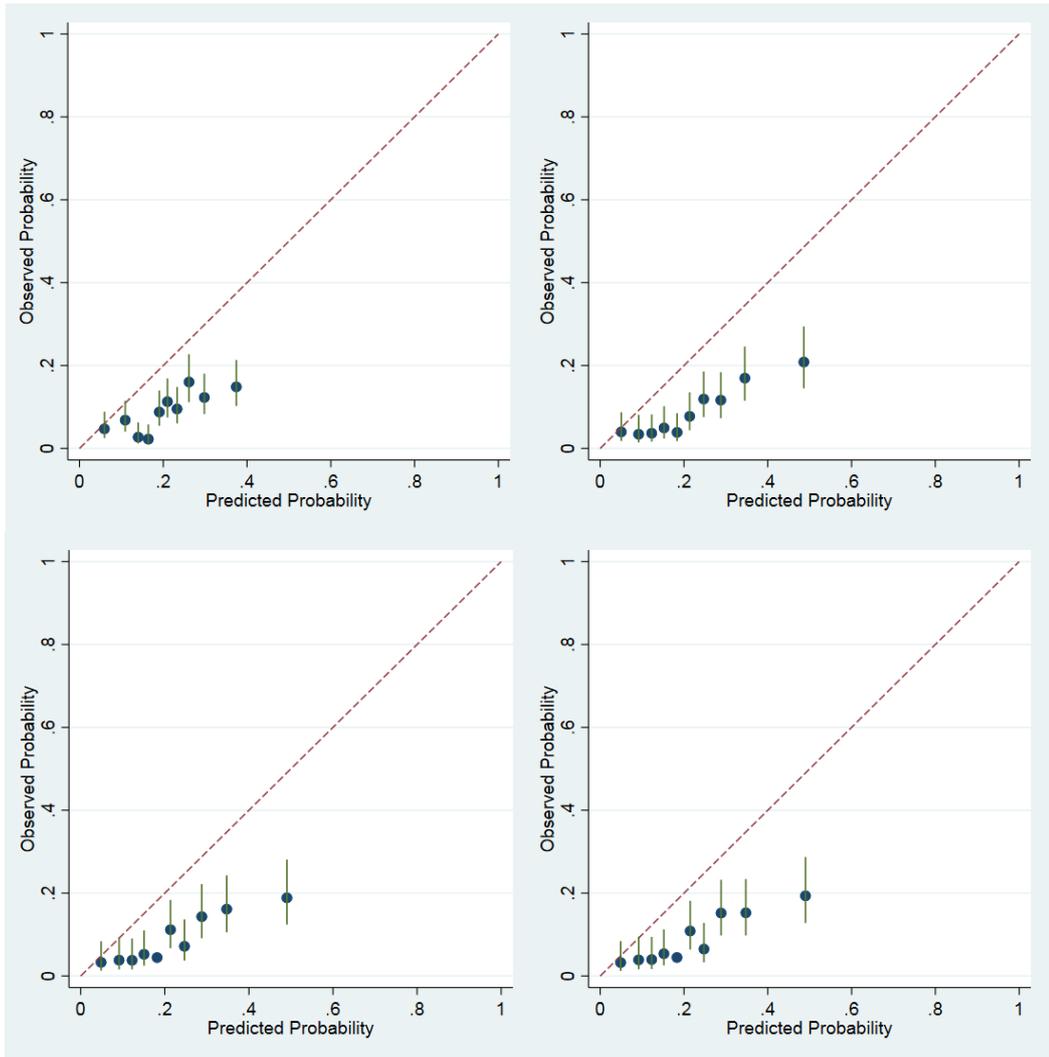
Figure 8.10: Calibration Plots for Cohort Specific original QRisk3 Male model (top left), with addition of eGFR and ACR (top right), eGFR, ACR, Albumin and Phosphate (bottom left) and eGFR, ACR, Albumin, Phosphate and haemoglobin (bottom right).

8.6.   Alternative Prediction Approaches

Alternative approaches to risk prediction were considered next. Firstly, predicting the composite outcome of CV event and all-cause mortality was assessed. The role of death as a competing risk was then considered. This part was considered in two stages. Firstly, how adjusted incidence for CV events were altered in the presence of death as a competing risk. This initially included adjustments to the Cox model's estimation of baseline risk calculated through the Kaplan-Meier estimator. Secondly, recent developments of assessment of multiple outcomes, including multi-state models was considered. This included how the recently developed model for advanced CKD developed by the CKD Prognosis Consortium might be adjusted and adapted to include the less advanced forms of CKD which predominate in the LCC cohort.

8.6.1. Composite Outcomes

The composite outcome of time to the first of either death or CV event was tested in the QRisk3 Female and QRisk3 Male models. Out of a population of 10,353 in QRisk3 Female, there was an additional 2,369 (22.9%) mortality event in those who did not have a CV event. Within Cohort Specific for QRisk3 Female, 397 (15.5%) additional events were included in the outcome using the composite method. The use of the composite outcome of CV and death led to no significant change in discrimination with C similar for the composite outcome as it was for the original outcome of CV events. When calibration plots were assessed the composite outcome led to significant underestimation of risk in both Whole Cohort and Cohort Specific, but the magnitude was less in the latter. Calibration plots for the composite outcome are shown in Figure 8.11.
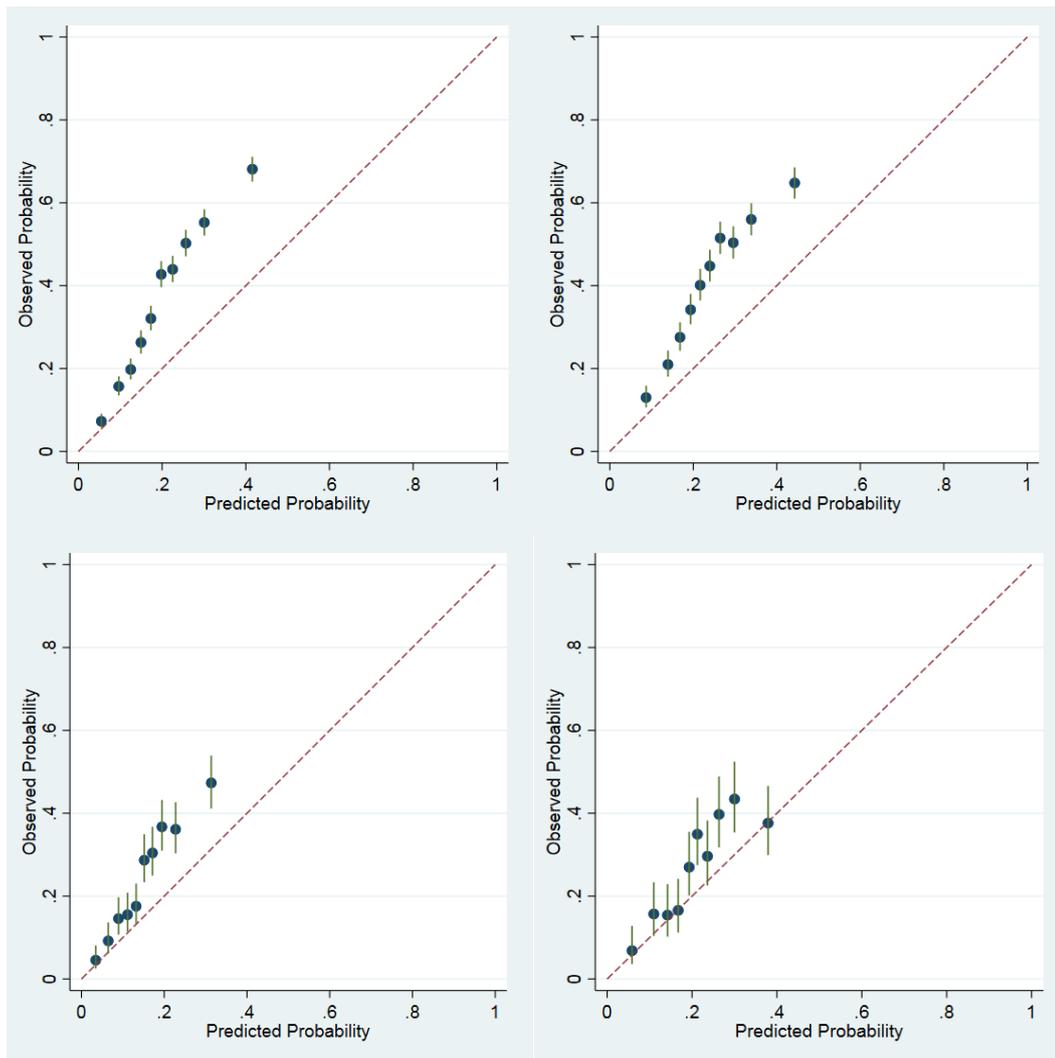
Figure 8.11: Calibration plot for QRisk3 Female and QRisk3 Male for the composite endpoint of a cardiovascular event or death for Whole Cohort (top) and Cohort Specific (bottom). Groups split in to deciles and risk based on original baseline risk as published in QRisk3.

8.6.2. Competing Risk

So far in this thesis, risk of outcomes have been considered in isolation. The sole purpose of the risk prediction tools assessed have been to quantify the risk of a CV event within a follow-up period. As was described in Table 4.4, the event rate for death was more than twice that of CV events during follow-up (73.4 versus 31.1 per 1000 person-years). As explained earlier in this chapter, death is a censoring event, an individual has no more follow-up beyond this period and cannot experience a CV event. Contrast this with individuals who have left the practice and study follow-up, who can have a CV event in the future. However, Cox model treats these individuals in the same way, despite the fact that they may be very different, as was shown in Table 8.2, in their characteristics. Therefore how the Cox model models this difference is of importance to how risk is assessed.

When an individual is at risk of an event, such as death, that precludes them from having a second different event, then this is described as a competing risk. In the current analysis, death is counted as an 'absorbing' state, the person can never continue follow-up and experience the other event, a CV outcome. However, in a full model of competing risk CV events are considered a 'transition' state as after a CV event occurs an individual can still go on to the absorbing state, i.e. die.

Under the current framework of the models that were externally validated in Chapter 7, net survival, has been the way risk has been described. The issue with this description is that it produces results from a theoretical world where an individual is only at risk of the event of interest. So, for the LCC cohort we are predicting risk for individuals who cannot die from a non-CV event. As can be seen from the event rates described above, probably related to the older age of the cohort, this precludes a significant number of the cohort and may prevent accurate individual risk prediction in the real world.

Instead, the concept of relative survival using cause-specific hazard ratios aims to address this issue and provide quantification of risk of an event in overall context. Relative survival is more commonly used in cancer risk prediction, where specific survival from the cancer is of interest in isolation i.e excluding

other potential causes of death (173). A similar approach is desirable in CV risk prediction particularly in the context of a high level of competing risk from non-CV events, as is the case in the LCC cohort. A relative survival approach would result in both different hazard ratios and baseline risk for the cause-specific CV event risk compared to the Cox model's results. However, because external validation was performed of models in Chapter 7 only the baseline risk can be altered if this alternative approach to risk prediction is taken. If beta coefficients were adjusted too then a different *de novo* model would have been developed and external validation would no longer be taking place (107). This alternative way to model the baseline risk is known as the cumulative incident function. The cumulative incidence function models the cumulative failure rate over time for a specific causes and accounts for the effect of other competing events. By modelling individual curves for each cause, a sum of the individual causes can be made. These can be 'stacked' upon each other to provide a visual example of how total risk accumulates across time from the individual outcomes of interest.

Figures 8.12 and 8.13 displays the stacked cumulative incidence for CV events and death over time for QRisk3 Female Whole Cohort and Cohort Specific. The y-axes for the cumulative incidence are scaled differently for the purpose of displaying the relative difference in event accumulation rather than the absolute accumulation. The relative accumulation of the two events was similar for Whole Cohort and Cohort Specific. 75 years of age, the median to the nearest five years for the Cohort Specific QRisk3 Female, was used to divide the cohort for the figures. Again, the relative contribution of cumulative incidence was probably similar.
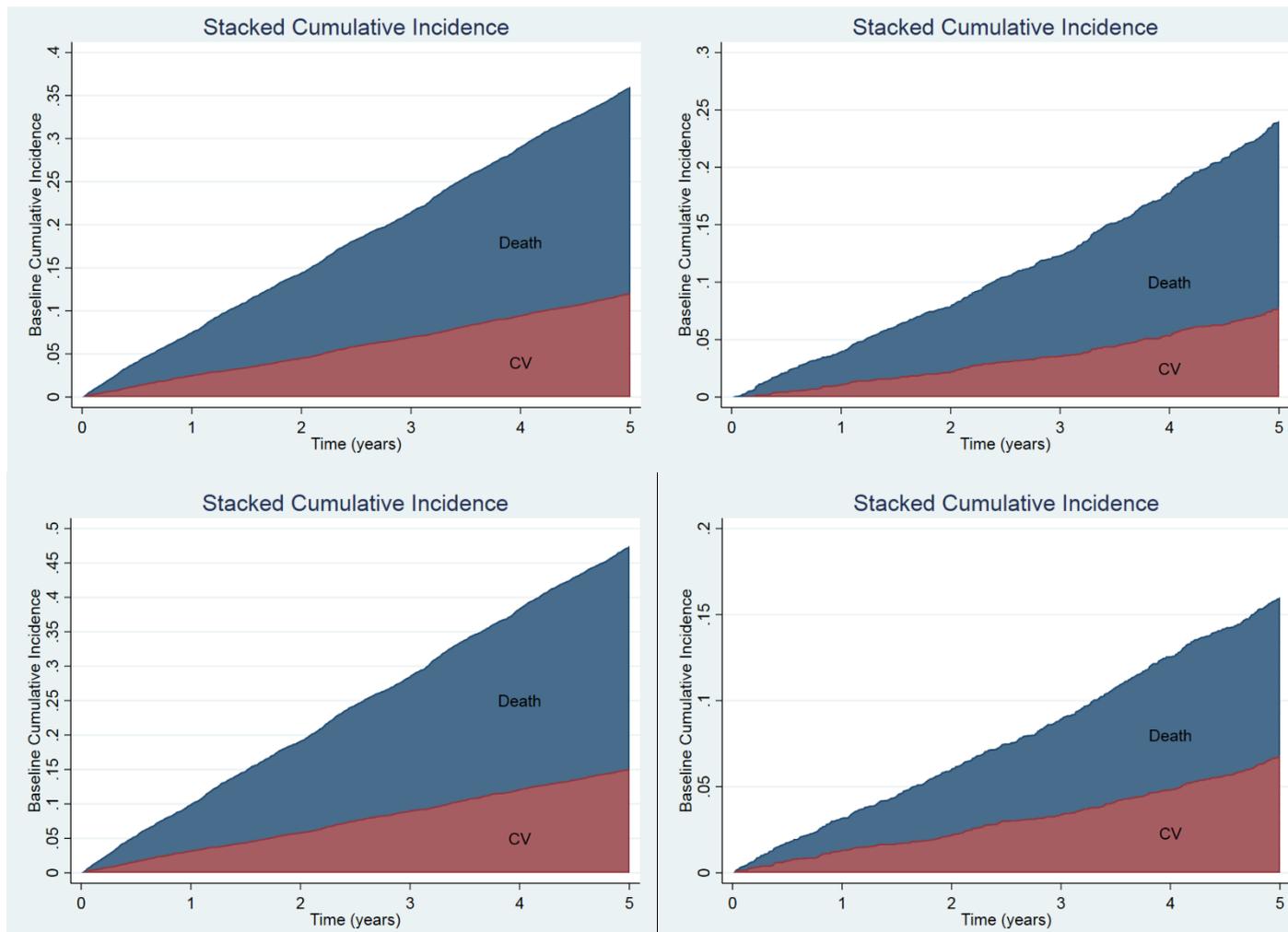
Figure 8.12: Stacked Cumulative Incidence for QRisk3 Female. Clockwise from top left, Whole Cohort, Cohort Specific, Whole Cohort under 75 years of age, Whole Cohort over 75 years of age. **Note difference in y-axis scales.**
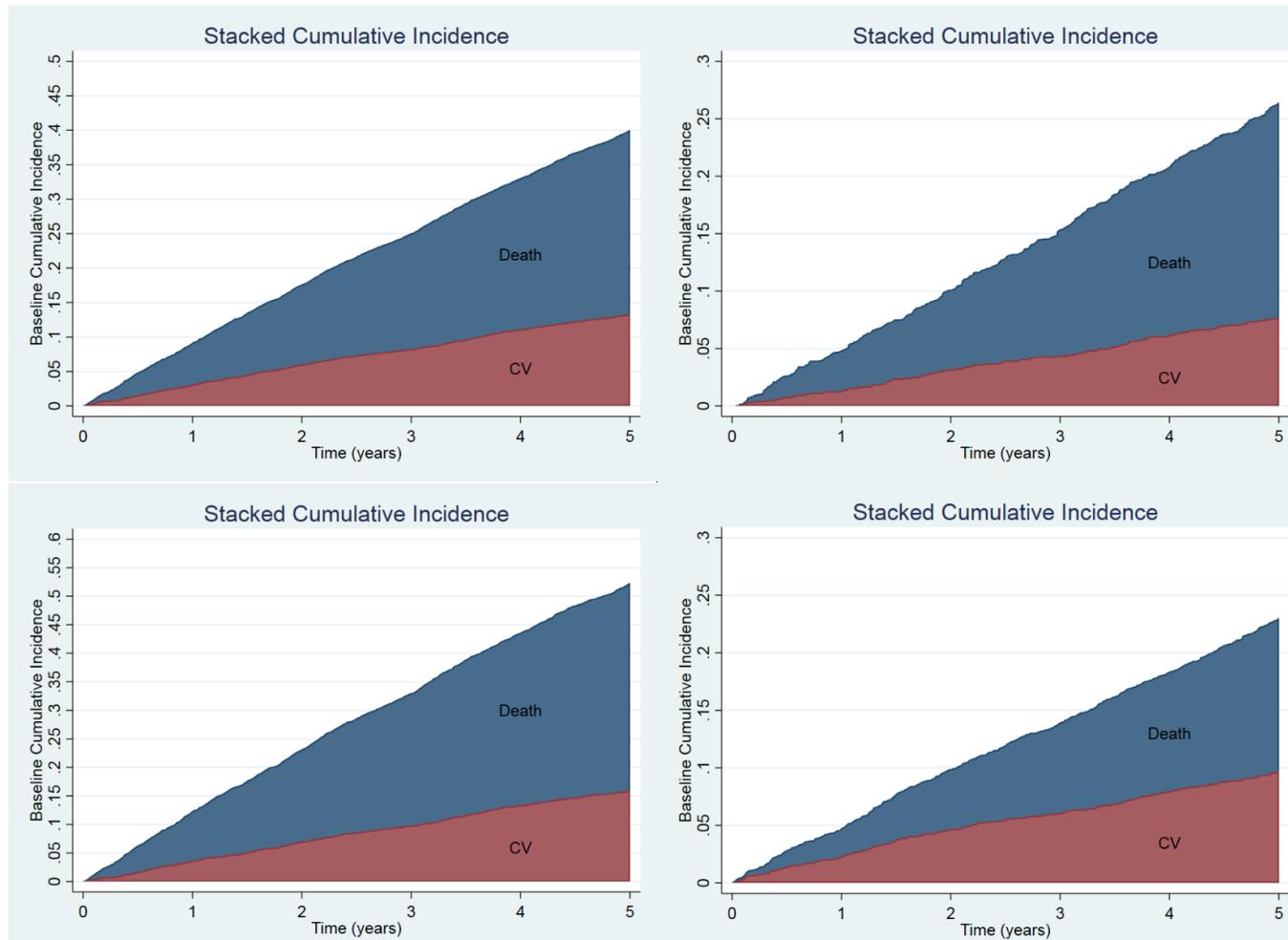
Figure 8.13: Stacked Cumulative Incidence for QRisk3 Male. Clockwise from top left, Whole Cohort, Cohort Specific, Whole Cohort under 75 years of age, Whole Cohort over 75 years of age. **Note difference in y-axis scales.**

8.6.3. Analysis of Adjusted Incidence

The cumulative incidences for CV and death events suggest that the baseline risk estimation may be affected by the presence of competing risk. In exploratory analysis, the adjusted incidence of CV events using the cumulative incidence function was utilised to change the baseline risk for all six models. This was based on the percentage change from the Kaplan-Meier estimated risk to the cumulative incident function percentage for each decile. This was then used to adjust the model's baseline risk for each decile.

Table 8.7 presents the changes in baseline risk calculated from these methods. The increase in baseline survival was greatest in those at highest predicted risk as they were also the deciles who had the greatest competing risk of death. The highest four deciles for both QRisk3 Female and QRisk3 Male had over one percent mean increase in survival i.e. calculated mean risk was lower in these groups due to the competing risk. Calibration was improved by a marginal amount, but still continued to mainly overpredict risk. Some of the QRisk3 Female upper 95% CI for some deciles crossed the 45° line suggesting the appropriate calibration could not be ruled out from the dataset. The calibrations plots for these findings are shown in Figure 8.14.

Risk Decile	Re-estimated Baseline	Mean (95% CI) % Δ in Predicted Risk
Q3 Female **Baseline 0.99444**		
1	0.99445	0.014 (0.014 to 0.015)
2	0.99446	0.043 (0.043 to 0.044)
3	0.99451	0.170 (0.170 to 0.171)
4	0.99453	0.223 (0.222 to 0.224)
5	0.99456	0.356 (0.355 to 0.358
6	0.99466	0.707 (0.705 to 0.709)
7	0.99474	1.071 (1.069 to 1.073)
8	0.99470	1.059 (1.057 to 1.062)
9	0.99480	1.637 (1.632 to 1.642)
10	0.99491	2.651 (2.632 to 2.670)
Q3 Male **Baseline 0.98863**		
1	0.98866	0.022 (0.022 to 0.023)
2	0.98869	0.061 (0.061 to 0.062)
3	0.98880	0.224 (0.222 to 0.225)
4	0.98882	0.284 (0.282 to 0.285)
5	0.98889	0.439 (0.436 to 0.442)
6	0.98909	0.838 (0.835 to 0.840)
7	0.98924	1.231 (1.229 to 1.234)
8	0.98917	1.192 (1.189 to 1.194)
9	0.98937	1.803 (1.797 to 1.809)
10	0.98960	2.791 (2.774 to 2.808)

Table 8.8: QRisk3 Female and QRisk3 Male Re-estimated 5 Year Baseline Risk and Mean Change in Percentage Predicted Risk. Re-estimation based on Cumulative Incidence Function adjusted baseline risk by decile.
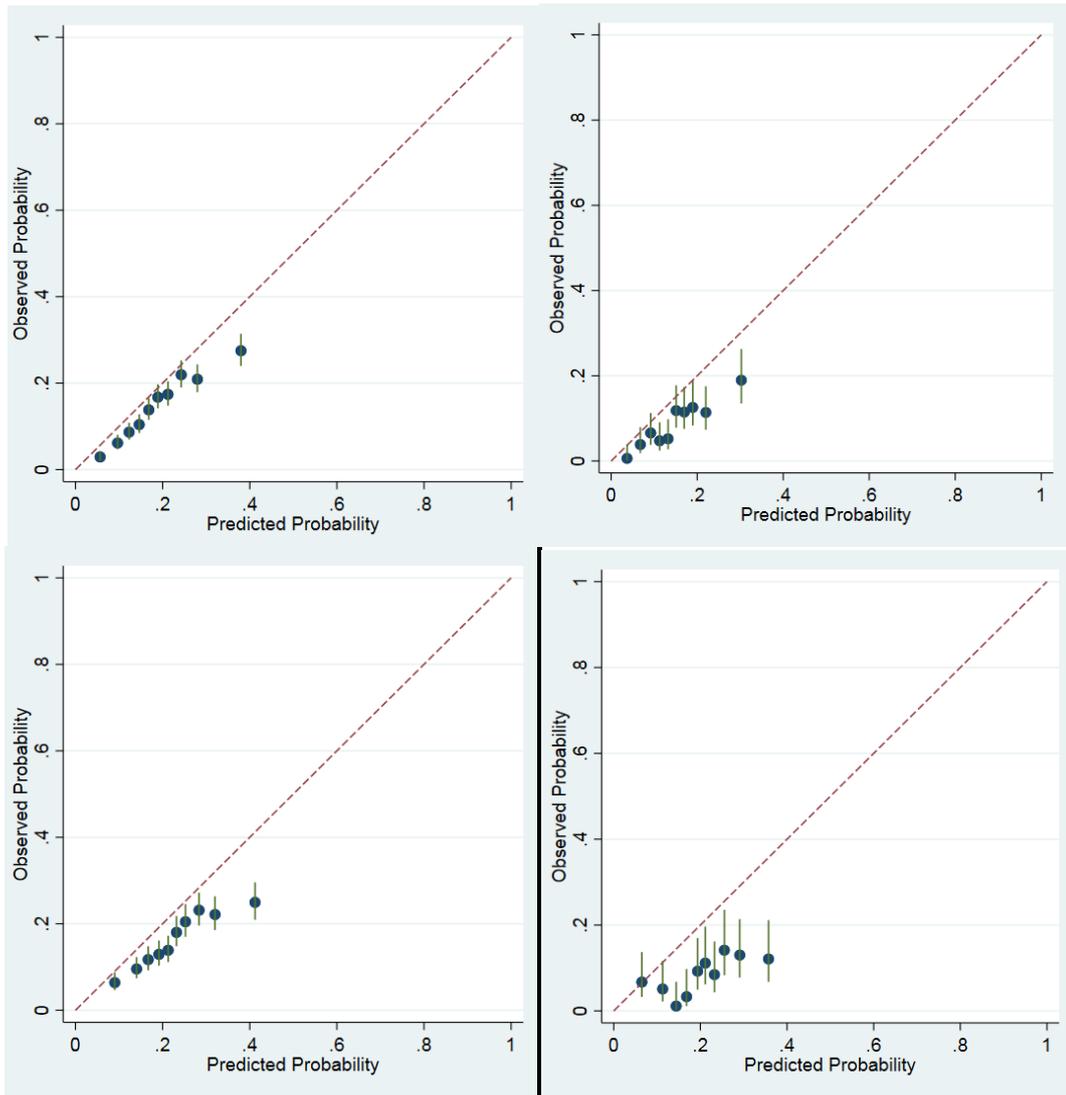
Figure 8.14: Calibration Plots Using Re-estimated Baseline Risks from Cumulative Incidence Function Adjusted Baseline Risk. Top row QRisk3 Female, bottom row QRisk3 Male, left Whole Cohort, right Cohort Specific.

8.6.4. Multi-State Models

Despite these updates and improvements to the models, they would still only predict CV risk in isolation and provide no information on other event risks. The interest to the patient and clinician would be in simultaneous prediction of these competing risks. For example, a 10% risk of a CV event risk might be managed differently if the risk of death in the same period was 2% or 30%. Whilst there are directly modifiable risk factors for CV disease such as hyperlipidaemia, the general risk of death has fewer modifiable factors with diminishing returns as the probability over a long enough time span always reaches one.

In the case of simultaneous event prediction in CKD, the events of interest would probably include death, CV events and ESRD. This has recently been modelled in advanced CKD (eGFR <30 ml/min/1.73m^2) by the CKD Prognosis Consortium. This tool uses the variables of age, sex race (white or black), eGFR, ACR, systolic blood pressure, previous history of CV disease, diabetes mellitus and smoking history to give two and four year risk of CV, ESRD and all-cause mortality events. This estimation includes probability of transition through the eight combinations of these events, or potential pathways. Risk of no event is also calculated. For instance, it can provide the percentage risk of any CV event, as well as the risk within the next four years of a CV event followed by death. An example of the tool and its related output is shown in Figure 8.15.

Timing of clinical outcomes in CKD with severely decreased GFR

Age (30-85 yrs)	80	⌄
Sex	Male	⌄
Race (White or Black)	White	⌄
eGFR (mL/min/1.73m²)	29	⌄
Systolic Blood Pressure (mmHg)	140	⌄
History of Cardiovascular Disease	No	⌄
Diabetes	Diabetes	⌄
Urine Albumin to Creatinine (mg/mmol) Click to change between mg/g and mg/mmol	1.1	⌄
Smoking History	Not Current Smoker	⌄

Risk at 4 Years ⌄

Outcome	Risk
Any Kidney Replacement Therapy (KRT a.k.a. ESRD)	3.7%
Any Cardiovascular Disease (CVD)	15.6%
Any Death	24.2%

● CKD G4+ - no event	64.4%
● KRT Only	1.9%
● KRT After CVD	0.1%
● CVD After KRT	0.3%
● Death After KRT	1.1%
● Death After KRT and CVD	0.3%
● Death After CVD	5.8%
● Death Only	17.0%
● CVD Only	9.1%

Figure 8.15: An example of risk prediction by the CKD Prognosis Consortium's 'Low GFR Events' tool. The clinical characteristics used for the example are the same as those used for the earlier example in the chapter of predicted risk for QRisk3.

8.7.   Conclusions

This chapter has explored the potential reasons why models may not perform as well in the LCC, and potentially wider CKD groups, compared to the general population. Results of testing these potential reasons are summarised in Table 8.9.

Reason	Summary Result	Possible Impact
Proportional hazards assumption	No strong evidence for violation of proportional hazards either from statistical testing or graphically	Minimal
Non-informative censoring	Formal statistical testing methodology difficult to perform as model already specified. ~25% of individuals were censored due to death as opposed to moving out of practice in study (loss to follow-up). Individuals who died were more likely to have higher predicted risk than those who left the practice for another reason	May have impacted the baseline risk calculation
Baseline risk re-estimation	<0.5% change in baseline risk for all models when using the exponential method	Minimal
Event identification from linked data	Percentage of events missed by methodology in thesis would have needed to be substantial higher than data from Herrett et al would suggest	Some, but very unlikely to account for all of the difference
Addition of new co-variables	Addition of new variables were associated with statistically significant HRs, but no clinically meaningful change in discrimination or calibration	Minimal
Composite outcome including mortality	Substantial underprediction of risk	Substantial, but would lead to ongoing mis-calibration in other direction
Cumulative incidence to model baseline risk	Improvement in calibration, but model continued to overestimate risk	Some, reduced degree of mis-calibration in all models

Table 8.9: Summary of tests performed to assess hypotheses for model underperformance.

Firstly, the dataset was discussed as a possible reason for this finding. This was initially highlighted in Chapter 4 in relation to the identification of outcomes and routine nature of the data collected. It is unlikely that non-identification of CV outcomes from secondary care sources is the sole major reason for the overestimation of risk. Approximately four in ten CV events would have had to have been missed in order for the risk prediction models tested to have been well calibrated. Even using the upper 95% CI, a quarter of events would have needed to have been missed. It is important to note that the QRisk tools, may have themselves missed events through a lack of linkage to MINAP, therefore this may cancel out the LCC's lack of ONS data.

Next the assumptions of the assessed models were considered. There was limited evidence for deviation from the proportional hazard assumptions in all models. Perhaps the greatest deviation was in the QRisk3 Female model, where there was minor evidence of non-proportional hazards in the highest risk groups. However, given the number of multiple comparisons, it is not possible to rule out that this was simply artefact of the data. Non-informative censoring is harder to assess, particularly in observational data, due to the lack of suitable methodologies. A simple comparison of characteristics by reason for censoring suggested that those who experienced a CV events and individuals who were censored due to death, were similar in relation to predicted risk and age. Similarly, those who completed follow-up and those who were censored as they had left the practice during the study follow-up were also similar. This suggests that the censoring process within the study may indeed have been informative about the possible outcomes of individuals.

The methods used to convert baseline risk from 10 to five years was also considered. For all models there were minor changes to mean risk when an exponential method was used to recalculate this figure, instead of a proportional method. There was no evidence that this improved calibration and as ranking was unchanged, there was no difference in discrimination.

The role of outliers within the dataset was also explored. Based on visual inspection of subjective methods to identify significant outliers, less than one in a thousand individuals were identified and they had no major influence on the

findings. These outliers were more likely to have extreme values of the β sum for the model, typically due to extreme values of blood pressure variability, age and cholesterol constituents. They also had events either very early on in follow-up or extremely late.

Based on the evidence collated in Chapter 5 and from the CKD Prognosis Consortium, the addition of eGFR, ACR, albumin, haemoglobin and phosphate was tested for the models. Generally, all were associated with increased risk in unadjusted analysis. In adjusted analysis, first with β sum, then with all the other variables, only the kidney markers were consistently associated with increased risk. In relation to performance of the model with these additions, there was some minor improvement in discrimination seen, but no substantial correction of the miscalibration.

Finally, alternatives approaches to prediction were considered, and where possible assessed. The first method was to predict a composite output of CV events and death, but this led to no major improvement in discrimination and reversal of the direction of miscalibration of risk to underestimation. The concept of competing risk was then introduced and alternative methods to predict the baseline risk were considered. These alternative approaches accounted for the high incidence of death during follow-up and corrected risk predictions for this issues. These exploratory approaches using an adjusted baseline risk from the cumulative incidence function led to minor improvements in calibration.

However, this improvement is perhaps not the most methodological sound approach to risk prediction in this group as it does not include quantifiable prediction of risk for death, or other events such as ESRD. The recently published CKD Prognosis Consortium's 'Low GFR Events' tool begins to address this clinical problem. However, with its restriction to eGFR<30 ml/min/1.73m^2 it is unlikely to be applicable to many individuals in the LCC cohort, who are not already under the care of secondary care renal services. Nevertheless, this concept of simultaneous risk prediction of multiple events should be expanded to include a wider range of individuals with earlier stages of CKD.

8.8. Summary

The results of Chapter 7 suggested that general population CV risk prediction models overestimated risk in CKD populations. This chapter explored some of the reasons why this may have occurred. LCC has robustly identified CV events and even in a worst case scenario of missed events this would not account for the miscalibration seen. The cohort also showed no major violation of the proportional hazards of the Cox model used in development of the six risk predictions models. However, there was evidence that censored individuals varied between the two main reasons censoring: death and de-registration from the practice. In both groups the characteristics were different in relation to age and predicted risk based on the β value. This suggests there may have violation of the non-informative censoring assumption. This hypothesis led to investigation of the role of death in risk prediction within the cohort. Using a cause specific approach, i.e. when death was accounted for, the observed risk of a CV event was lower. This suggests that current standard Cox models may overestimate predicted risk of a CV event. Exploratory analysis of adjusting baseline risk in the Cox models did not lead to substantial improvement in models' calibration. Further work, to complement and extend that by the CKD Prognosis Consortium, is required in relation to multi-variate and multi-state models in primary care CKD populations.

Chapter 9

9.   Thesis Discussion

This chapter aims to summarise the thesis and consolidate the conclusions for each individual chapter. The overall strengths and weaknesses of the thesis are then discussed, followed by how the findings may influence clinical practice and health resource planning. Finally, there is a summary of the future work required in this research area.

9.1. Introduction

CKD is independently associated with increased risk of adverse outcomes, including CV events (2, 32). Existing CV risk prediction tools have been developed using flawed methodology and are therefore not suitable for clinical use (73). Equally, general population CV risk prediction tools have not been specifically externally validated in CKD populations. CV risk factors may be different in CKD compared to the general population in two ways. Firstly, traditional risk factors may have different magnitudes and directions of effect in CKD. Secondly, non-traditional risk factors related to arteriosclerosis, cardiomyopathy and inflammation are likely to have a role in increasing CV disease risk in CKD (33).

Whilst methods for survival analysis, including the Cox model, have been available for more than 50 years (64), it is only more recently that they have been applied to prognostic research in the form of risk prediction models (60). However, risk prediction models rarely progress to inform routine clinical care decisions because of their methodology limitations (60, 73).

Therefore, as set out in this thesis, the specific role of CV risk prediction in CKD warrants further investigation and testing in order to aid its fuller integration into primary care management of individuals with CKD.

9.2. Thesis Findings Summary

This thesis has aimed to establish risk factors for and prediction of CV events in individuals with CKD. Chapter 4 described the primary care CKD cohort developed for this thesis, The Leicester City and County Chronic Kidney Disease (LCC) Cohort. Forty-four out of 106 (41.5%) practices participated in the study representing a patient population of over 275,000 adults. Using the full KDIGO definition, 17,248 (6.2%) individuals had CKD. The mean age was 77 years, three-fifths were female and just over a half had CKD 3a with either A1 or no recorded data for albuminuria. Less than 1% of the general population under the age of 50 years had CKD but more than 50% aged over 80 had a biochemical diagnosis of CKD. CV related co-morbidities were common with just under a third having diabetes mellitus, more than nine in ten having hypertension and approximately a quarter having established CV disease. During follow-up, two outcomes were collected through both primary and secondary care records, all-cause mortality and CV events. Death was the most common event, with just under three in ten of the cohort dying. CV events occurred in approximately one in eight of the cohort.

Chapter 5 described a systematic review of CV risk factors in CKD cohorts. Most of the papers described secondary care cohorts with more advanced CKD, identified through low eGFRs and significant proteinuria. After adjustment for establish traditional CV risk factors, a number of non-traditional risk factors were associated with CV disease events. Serum albumin, haemoglobin and phosphate were the most commonly reported non-traditional risk factors and each was associated, in adjusted analysis, with a statistically significant change in CV event risk. Other risk factors that may be associated with CKD related mechanisms for CV disease such as serum calcium, urate and left ventricular hypertrophy had limited evidence for their association.

Chapter 6 described an update to the systematic review into risk prediction tools in CKD as previously published by Tangri *et al* (73). There was a limited number of new publications for CV disease event prediction identified. There was also no identified external validation of existing CV prediction previously identified by Tangri *et al*. Therefore, there is no reason to differ from their final

conclusions that 'further development of models for cardiovascular events [in CKD]…is needed'.

The LCC cohort was then used to externally validate existing CV disease risk prediction tools, both those used in clinical practice in the general population and those specifically developed in CKD, as identified in Chapter 6. Due to the routine nature of the data, multiple random imputation was used to account for missing data and compared to results from complete cases analysis. C-statistic and OE plots were the primary measures to assess discrimination and calibration respectively. The C-statistic was lower in all CKD and general population models. The Weiner model was the least well performing model of all the CKD models, with even the best performing CKD model having a C in the region of 0.65. General population models, performed slightly better with the QRisk models probably performing better in the female population and the AHA perhaps performing better in males. Results were similar whether exclusion criteria according to the development model, the Cohort Specific group, or the whole LCC cohort was analysed. Similarly, complete case analysis had comparable performance to multiple random imputation analysis.

Assessment of calibration was limited by the limited availability of the baseline risk function for the models. For Framingham based CKD models, the baseline risk function was utilised. General population risk prediction models were recalibrated to assess five year, instead of 10 year, risk. All models substantially *over*estimated risk. For QRisk3 Female, 5-year year baseline survival was adjusted from 0.9944 to 0.9969, a 44.6% increase, which led to good model calibration. A similar magnitude of re-calibration of baseline risk was required for all other general population models to provide good calibration.

Chapter 8 described model adjustments made to improve risk prediction using QRisk3 Female as an example. Firstly, methods to improve discrimination were tested. Markers of CKD severity previously identified as being associated with increased risk of CV events, eGFR and ACR, were assessed (174). The addition of these improved the C-statistic by a similar magnitude to what has been previously described (74). Following on from the systematic review findings in Chapter 5, the other biochemical risk factors associated with CV

disease in CKD, serum albumin, haemoglobin and phosphate were then tested. All were associated with statistically significant increases in risk in unadjusted analysis both in Whole Cohort and Cohort Specific groups. In adjusted analysis including the β coefficient for the model, all variables remained associated with risk in Whole Cohort except for haemoglobin. In the smaller Cohort Specific group, only eGFR remained significantly associated with risk. Results were similar in a full adjusted model using β coefficients and the other identified variables. Patterns of results were similar in QRisk3 Male, except phosphate instead of haemoglobin was not statistically significant in Whole Cohort and ACR was associated with risk in the Cohort Specific group.

Performance of the QRisk3 Female and QRisk3 Male models were then assessed. When the kidney markers, eGFR and ACR, were added to the model there were similar magnitudes of change in change in C as to those previously reported (74). However, the changes were not statistically significant. When the three identified risk factors identified in Chapter 4 were also added to the altered risk prediction model, there were further improvements in C but not to a statistically significant level. Calibration showed similar small improvements with the addition of these risk factors, but risk continued to be overpredicted.

Alternative risk prediction approaches and methods were then considered, particularly in relation to the high number of deaths in the cohort and its possible role as a competing risk. A change to the models to predict a composite outcome of CV events or death, led to no change in discrimination and continuing large amounts of model miscalibration, except for now underprediction of risk by the models. Cause specific risk was then assessed, through stacked cumulative incident plots and compared to risk plots using the standard approaches used for the original development of the models. Adjusting the baseline risk to account for this different approach did lead to some improvement, but not enough to adjust for the full amount of miscalibration. Multi-state models were then introduced and identified as a potential area for improvement in risk prediction. Potential adjustments to the CKD Prognosis Consortium's recently published multi-state model (175) to include individual likely to be under primary care CKD were then discussed.

9.3. Strengths

The strengths of this thesis are again considered chapter-by-chapter. The LCC cohort was developed in a number of ways to address known limitations of observational studies in CKD. KDIGO definitions for CKD stage 3a, or more advanced, ensured that individuals were appropriately identified. The conversion to EPI eGFR, instead of MDRD eGFR most commonly used in routine clinical practice at the beginning of thesis, ensured that again a suitable population was identified. Geographical and socioeconomic variation was ensured through the inclusion of even split of practice from urban and rural areas. Event identification often varies in epidemiological studies. To address this, comparative methods were used for both identifying baseline characteristics and follow-up events. Baseline characteristic definitions were consistent with definition used by the CKD Prognosis Consortium, probably the largest and most influential CKD based epidemiology collaboration in the world (44). This ensured that any results would be comparable and consistent with their findings. For outcomes, event identification and quality was ensured through the use of both primary and secondary care records. The former on its own may identify only three-quarters of acute myocardial infarction events (81). The addition of the latter, through HES and MINAP data for follow-up, ensured that a similar proportion of acute myocardial infarction events were identified through these sources. The proportion of acute stroke events identified by HES, but not in primary care records, was much higher at nearly half. This may reflect the higher levels of coding for non-specific cerebrovascular events, particularly in secondary care data. The identification of events for all-cause mortality and ESRD also benefits from the combination of primary and secondary care events. Given these methods, over two thousands CV events were identified in the whole cohort, which is probably a sufficient number to externally validation a risk prediction tool (119).

The systematic review to identify CV risk in Chapter 5 aimed to utilise recommended guidance as set out by the Cochrane collaboration (122). This chapter's strengths include its pre-registered protocol, independent duplication of literature searches and independent data extraction. In addition, by only including data from studies where known general population CV risk factors

were adjusted for, the additional benefit in prediction of these novel CKD risk factors could be assessed. This may have reduced inter-study heterogeneity. The second systemic review used similar techniques, with the exception of the pre-registered protocol as the search protocol was already published in the original review, and therefore the findings are also likely to be robust.

Methods for assessing the risk prediction tools and presentation of results, was consistent with suggested guideline from the 'The PROGnosis RESearch Strategy' (PROGRESS) partnership's TRIPOD statement (167). All results for external validation underwent model assessment using both the imputed datasets and complete case analysis. This ensured that the maximum statistical power from the datasets were gained using the former, but that results were comparable using both methods. In addition, further sensitivity analysis was performed using both the Whole Cohort LCC cohort and a sub-cohort, Cohort Specific, matched to the original inclusion and exclusion criteria in the development cohort of the risk prediction tool. Results were again consistent across these different two methods.

Discrimination was assessed using the C-statistic a reliable indicator of a risk prediction tool's overall ability to separate those who experience an event from those that do not. Calibration was primarily assessed through predicted versus observed probability plots, rather than measures of general calibration such as the ratio of observed versus predicted events (O/E) or calibration-in-the-large. This allowed for the individual assessment of different strata of risk. The use of a small number of metrics meant that the performance between models could be directly compared. For instance, the superior results for discrimination in the general population models compared to the CKD models meant that the latter could be disregarded for further testing and adjustment, such as through the addition of new co-variables to the model.

When models were recalibrated, the baseline risk was reduced for QRisk3 Female by 28.3% and 44.1% for the Whole Cohort and Cohort Specific subgroup respectively. Similarly, the reductions were 35.2% and 61.1% for QRisk3 Male. When sensitivity analysis was performed to calculate the additional events required for the mean observed risk to match the predicted

mean risk, the minimum percentage was 18.7% (95% CI 12.9% to 24.3%) for QRisk3 Female Whole Cohort. Other models had even higher percentages. This suggests that missed events were unlikely to be the sole reason for model miscalibration. This is far in excess of the 3.0% of events that the data from Herrett *et al* might suggest were missed without the presence of ONS data (81). This again acts as reassurance that the methods used to develop the LCC cohort, and in particular the CV event identification, were robust.

In addition to the identification of events compared to Herrett *et al*, Chapter 8 thoroughly explored other potential reasons for the suboptimal performance of the risk prediction models in external validation, and therefore provides another major strength of the thesis. Further investigations included improvement of the models' performance by new co-variables, studying the impact of outlier values of the sum of the β co-efficients, different methods to recalculate the baseline risk function for the re-calibrated models and the validity of the assumptions of the Cox model, proportional hazards and non-informative censoring.

There was no evidence that the proportional hazards assumption of the Cox model had been violated, but, whilst more difficult to formally assess, there was evidence that the proportion of individuals censored due to non-CV death may have impacted the non-informative censoring assumption. This was explored further through calculating the cumulative incidences for CV events and non-CV death. These analyses provided further evidence for informative censoring having occurred. Exploratory analyses to adjust the baseline risk function for different levels of risk were performed but did not provide a suitable solution. Therefore a further strength of this thesis is provided by this extensive subanalysis; that there is unlikely to be a simple solution to this problem using the Cox model and the development and validation of multi-state models is likely to be needed.

In summary, the strength of the thesis are the high quality methods used for the systematic reviews, the robust data collection methods for the cohort and the statistical approaches and sensitivity analyses used for the external validation of risk prediction tools.

9.4. Limitations

In assessing the limitations of the thesis, both limitations of the methods as well as how they have been applied will be considered. The systematic review of risk factors relied on the use of observational data, as is common in all risk factor research, to identify and assess risk factors. Observational data are likely to be limited by their inability to differentiate an effect being causal or associative. Although methods were used to limit the effect of heterogeneity, such as only including results from models adjusted for traditional CV risk factors, it remained a prominent problem for nine of the risk factors. Adjustment for eGFR and proteinuria in presentation of results varied between studies, despite three-quarters of the studies being published in kidney journals. Reporting of data in the original study, as has been noted in other areas of systematic reviews, also remained an issue (153, 154). This problem was only partially addressed through provision of additional data from studies' authors. With the publication of guidance from the PROGRESS consortium, it is hoped that this issue will improve in the future (60, 61, 167, 176, 177).

All studies, both in the systematic review and the validated models, used the Cox model. There was limited evidence of the testing of the two main assumptions of the model, proportional hazards and non-informative censoring. There was also no noted sensitivity analysis for the use of models, including testing for competing risk for mortality or any other outcome.

The LCC cohort is limited by a number of general methodology issues. Firstly, its observational nature means that again only association, and not causality, of the risk factors can be established. Whilst it has many similarities to national studies, in relation to CKD prevalence and severity, it is geographically limited by being a study based in one area. However, rural and urban locations were well represented as was the full spectrum of deprivation. Secondly, the routine clinical basis for the data collection limits the findings, especially in relation to the possibility of confounding by indication, or the more unwell a person is the more likely they are to visit their doctor, have tests, and therefore be recorded in the dataset.

The role of a reduced eGFR in elderly individuals, has been widely debated almost ever since the concept of eGFR was first clinically used in the early 2000s (28, 30, 178). The mean age of the LCC cohort and the approximately 50% with CKD3a with either A1 or no recorded proteinuria suggests that there could be a substantial proportion of the cohort with 'ageing' kidneys instead of true 'CKD'. The current thesis cannot answer this debate, but perhaps should be seen as a further indication to use risk prediction tools in the elderly with caution.

Whilst the LCC cohort's data linkage between primary and secondary care is a significant strength of the current thesis, some limitations to this process must be acknowledge. LCC does not have data linkage to Office of National Statistics CV mortality event data, this may account for missing up to 3% of all acute myocardial infarction events (81). The number of events missed for cerebrovascular events is less clear and an ongoing area of research (115). In addition, the data linkage from secondary care is not national HES or MINAP data. Therefore, if an individual had a secondary care event outside of University Hospitals of Leicester's three hospitals which was also not recorded in the primary care notes, then it would have not been detected. It is likely though that all studies have missing data of this nature, with independent outcome committees in randomised controlled trials suffering the least and observational data such as LCC and QRisk3 (67), which did not have access to MINAP data, the most.

The assessment tools for the process of external validation also have limitations that must be considered. Whilst PROGRESS and its related TRIPOD statement suggest methods, the 'gold standard' for assessment of risk prediction tools is not universally agreed, as is suggested by the number of methods available, their advantages and disadvantages, and the academic debate that surrounds them (105). For discrimination, the C-statistic is likely to underperform in external validation of secondary subpopulations compared to its results in the original development cohort. An example of this is in QRisk3 itself where the C-statistic performance worsened in age subgroups, particularly as age increased, and also in the coding identified CKD subgroup (67).

Calibration was also limited by three factors, no available baseline risk for the CKD models, the adjustment of the prediction interval for general populations' baseline risk and the potential for missing events. For the first factor, no additional data was provided by corresponding authors for the CKD models. The second limitations meant that baseline hazard had to be adjusted from ten years to five years, meaning it had to be estimated for this interval. Initially, this was on the basis of the estimated five year risk being half of the risk at ten years. It is unclear what the correct method for adjusting this is, but if near linear risk is assumed then this is likely to be a near approximation to reality.

The methods for developing LCC had been completed, ethical approval had been received and data collection had already begun when QRisk3 was published. Given the anonymous nature of the data collection, it would have been unethical to have collected full patient records as this would have increased the chance of re-identification of individuals. Therefore, selected data extraction was performed with data relevant to CV risk and already published CV risk prediction tools. QRisk3 utilised six new risk factors to compute their risk prediction (67). Data on the use of corticosteroids was included in the original data collection tool for LCC, but coding for migraines, systemic lupus erythematosus, erectile dysfunction, atypical antipsychotic use and severe mental illness was not. This may therefore have altered the performance of QRisk3 during validation. These six new risk factors varied in prevalence from approximately 6% for severe mental illness to <0.1% for systemic lupus erythematosus. No other risk prediction tool had missing variables. In relation to calibration, as all these factors were associated with increased predicted risk of a CV event in the model, then their inclusion would have increased individual predicted risk and would only have worsened the magnitude of the model's over prediction of risk. The correct approach to account for missing, categorical, relatively uncommon variables in an external validation study is unclear. Given the prominence of QRisk2, and likely QRisk3 too in the near future (164), I felt it was important to validate this model, and note these limitations, rather than exclude it.

The issue of data linkage of the GP records to other CV outcome sources has been described in great detail in this and other chapters. The original plan for

the fellowship and thesis had been to use data from 'The Primary-Secondary Care Partnership to Improve Outcomes in Chronic Kidney Disease' study (PSP), a cluster randomised controlled trial of CKD nurse specialists in primary care. Follow-up for the trial was up to three years. An ethics extension was approved to perform observational follow-up of the 49 primary care practices included in the original study. 38 (77.6%) consented and participated in follow-up. However, substantial difficulties and delays arose accessing anonymised CV outcome data for MINAP and HES for the area's hospitals. These data for one hospital were received in November 2018 and had not yet been received from the other hospital at the time of the submission of this thesis. The original aim had been to use the data from PSP, instead of LCC, to externally validate the risk prediction models identified. LCC would then have been the external validating cohort for potentially adjusted or novel models.

The work to update the PSP was a substantial part of the first year's work for this thesis. Clearly, it was disappointing that this work was not able to be included in my thesis. After discussion with my supervisors, we agreed that the more robust and higher quality CV event outcome identification in the LCC cohort was more important than simply having a larger cohort to analyse as was the case with PSP. During reflective periods whilst writing up my thesis, this particular experience has taught me a great deal. Firstly, it has highlighted to me the challenges and issues that surround working between different datasets, in particular where multiple organisations are involved. Secondly, as described in Chapter 2, working without individual patient consent with anonymised data provides challenges in relation to linkage and maintenance of anonymity. Thirdly, responding and adapting to new challenges as they occur in a large, medium term project such as this an important skill that I have developed throughout the time of my fellowship. This will be an important attribute to use in my future in academia, clinical medicine and life in general.

More positively, the LCC has been used in collaboration with the CKD Prognosis Consortium for a number of publications include multi-state modelling in advanced CKD (175). It is also anticipated that the LCC cohort will contribute to the CKD Prognosis Consortium in the future.

In summary, the main weaknesses of this thesis are the observational nature of both the LCC cohort and the systematic review into CV risk factors, the possibility that some LCC individuals may have a process of 'ageing kidneys' rather than 'true' CKD, the possible loss of up to 10% of events due to not having national secondary care and ONS data, and the limited data in relation to baseline risk.

9.5. Potential Clinical Impact

Clinical risk prediction tools can be used in a number of settings including for direct patient information to inform decision making with a clinician, health service planning and aiding the development of future research projects. The information produced by this thesis is likely to be primarily used in partnership between patients and clinicians. Rather than a discussion around 10 year CV risk only, as set out by QRisk2, QRisk3 and the AHA tools, it is more likely to be used to discuss the relative chances of CV risk and mortality. This could allow for general planning of individualised care for a patient above and beyond the 'statin decision' that general population tools are currently utilised for. For instance, currently under QRisk3 a consultation might predict the following:

- **10 year CV risk** for a 80 year old white male, ex-smoker, with CKD stage 4, type 2 DM, treated systolic blood pressure of 140 mmHg, cholesterol to HDL ratio 4 and a body mass index of 26.4kg/m^2 – **45.6%**

Using a competing risk prediction model predicted from the LCC cohort for the same individual we might expect:

- 5 year death risk

- 5 year CV event risk

- 5 year risk of CV event followed by death

This additional information, might change the consultation from consideration of statin use, to a wider discussion around other factors for CV risk, general health frailty and possibly future end of life planning. This would particularly be the case in multi-morbid populations such as many of the individuals in the LCC cohort. If risk of non-CV mortality is not considered in conjunction with CV risk than the risk and benefits of medications such as statins, aspirin and anti-hypertensives can not be fully assessed and overuse of medications might be consider. For instance, whilst aspirin has been considered for CV primary and secondary prevention the associated risk of non-CV events, such as gastrointestinal or cerebral bleeding, is higher in individuals with CKD and so the risk-benefit balance of CV versus CV-risk would also be different (179).

However, despite risk prediction models having potential impact on clinicaldecision making, their impact either positively or negatively is rarely assessed (60). This may be possible due to a lack of recognition of them as 'health technologies' requiring regulatory overview. In addition, their adoption may be dependent on endorsement of prominent individuals in the research field, rather than good quality evidence, a suggestion that the field of prognosis still relies in '*eminence* based medicine' rather than '*evidence* based medicine'.

In order to robustly assess this, a prospective clinical impact study of the model is required, a step that is infrequently performed, including for widely clinically used models such as QRisk2 and AHA (60, 62, 63). This may be due to the perceived high cost of a full randomised controlled trial. Novel approaches such as step wedge design have been suggested to improve trial efficiency, but the introduction of new risk prediction tools, with minimal impact evidence, by guideline producers may potentially contaminate findings of results. Despite these adapted approaches, trials may require many years of follow-up and many thousands of patients to achieve sufficient power to detect a difference in 'hard' outcomes such as CV endpoints. Therefore, adapted outcomes have been suggested, including analysis of decision making (62, 63). From a nephrology point of view, an example of this might be how a prediction tool such as the KFRE effects clinicians' decision making to refer a patient to secondary care nephrology or not.

An additional consideration prior to implementation into routine clinical practice is its adoption by clinical systems such as EMIS and SystmOne into their clinician interfaces and the ability to automate the calculations for a whole practice. It is likely that manual calculations for an individual on a website, 'app' or score chart would lead to wide use of the risk prediction tool as a 'novelty' rather than widespread clinical use. In the context of the CKD Prognosis Consortium's 'low GFR events' tool, this could be used in secondary care in an individual consultation, but is unlikely to be usable as a primary care screening tool (175). An example of the data provided by the tool is shown in Figure 8.15.

These clinical software producers will usually adopt a process once it is incorporated and endorsed into national guidance, i.e. for England, NICE

guidelines. Therefore, further research into implementation to improve the validation of the tool is required to promote its inclusion in guidelines.

9.6.   Further Areas of Research

Further progress in the prediction of CV risk in CKD is required in two broad interconnected areas, methodological and clinical impact.

9.6.1. Methodology

The CKD Prognosis Consortium has recently developed a multi-state model for CV, ESRD and all-cause mortality outcomes over two and four years in individuals with an eGFR <30 ml/min/1.73m^2. Its clinical utility is probably limited in primary care by its eGFR based inclusion criteria (175). The model included the PSP-CKD cohort data described in section 9.4 for its development.

These methods address the primary limitation of the Cox model in the context of a complex disease pathway, primarily that of competing non-CV mortality and the ability to simultaneously predict multiple events. In the Cox model the risk of fatal or non-fatal CV event is calculated without considering non-CV causes of death (64). This means the individual can only ever in effect have a CV event or remain alive. This leads to an overestimation of these events as the risk of non-CV death leads to censoring of an individual and a reduction in the denominator in the risk calculation.

This may not be a significant problem in populations with low risk of other causes of death, such as in a younger population with limited mortality risk. However, in a population such as the LCC cohort death due to other causes is common and will influence the predicted risk. Using the cumulative incidence function to estimate risk accounts for this competing risk (180, 181).

Multi-state models represent an extension of competing risk which can calculate the simultaneous risk of more than one event (180). This allows the potentially complex healthcare pathways that a cohort of patients may go through to be studied and the individual events that may occur to an individual to be simultaneously studied. In turn this can allow for intuitive measures of events such as impact on life expectancy and mean survival. Whilst multi-state models have been principally been used in cancer, where predicting recurrence will likely impact on overall survival, as the CKD-PC have shown they can be

used within CKD to predict the overall risk of CV, ESRD and all-cause mortality events, as well as the risk of an event occurring before the 'absorbing' event of death occurs (175).

The two main components of a multi-state are the transition probability and transition hazard (180). The former is the probability of being in a certain state, *b*, at time *t* given that the individual was in state *a* at time *s*. It is also conditional on the previous trajectory of the individual up until time *s*. The transition hazard is based on the cause-specific hazard from the competing risk model for each transition. Once the multi-state model is fitted and adequately tested, such as the assumption of proportional hazards, transition probabilities can be calculated from the model. These can be utilised to graphically display stacked plots of probabilities or mean survival in a certain state, given a certain set of patient characteristics.

Whilst external validation and extension of the model to include all individuals with CKD is an attractive proposition for future research, the methodology required to do so needs further development and evaluation (personnel communication, Dr Michael Crowther, University of Leicester and Professor Richard Riley, Keele University). Therefore for this to occur appropriate methodological research, beyond the scope of this thesis, is required. Such developments may address whether multiple validations are required of each individual state or transition in the model, or if validation of the final state only is sufficient.A further issue, first highlighted to me when discussing risk prediction tools during study visits with primary care physicians, was the issue of updating, or 'dynamic', risk prediction. Their concern as clinicians was whether risk prediction tools, such as QRisk3, could be used to update a risk prediction once a variable had changed. In their experience this was typically when blood pressure or cholesterol had been lowered to a certain level. Equally, in the area of CKD this could be related to a change in eGFR or ACR, the latter typically through the use of angiotensin–converting-enzyme inhibitors such as ramipril. Whilst current risk prediction tools are probably used clinically in this way, the appropriateness of this is unclear. The arbitrary nature of the start point of LCC of 1st November 2011 may have led to some bias (182). QRisk3 does incorporate historical data via the use of the standard deviation of previous

blood pressure, but this is not a true dynamic model. Less than 10% of models are thought to incorporate longitudinal measures (183). One reason for this has been the computational intensive methods required meaning that in larger datasets such as QRisk3, model development has not been feasible (184). Recent work by Paige *et al*, has proposed a computational feasible 2-stage dynamic model (184). In CKD, and nephrology in general, these and other dynamic models have been proposed as methods to improve risk prediction but clinically focused research has been rarely performed (185).

9.6.2. Clinical Impact Studies

Even if a model is robustly established from a statistical perspective, there is no guarantee that it will have a beneficial clinical impact (60). Therefore this must be studied to assess how clinical decisions, and ideally hard endpoints, are affected by the risk prediction tool.

For the case of CV risk prediction in CKD, this is likely to be in the context of a primary care cluster design study at the level of each practice. Whether a stepped wedge design or pre-trial observations are included is another design consideration. However, these types of prognostic research studies are rarely performed (60). Where they have been performed, due to the limited methodological research in the area, their design and appropriate statistical powering have been inappropriate. This has reduced the chance of meaningful conclusions being drawn from the impact study (186). Studying current clinical decision making through a decision analysis study could help to understand clinician decision making and aid in the design and powering of a larger impact study (186, 187).

Assuming that a risk prediction model is thought to be ready for implementation studies, how the prediction data is presented is an important, but often overlooked topic. In the case of QRisk2 and QRisk3, the results are integrated into electronic clinical systems such as EMIS and SystmOne. This allows for automated calculation of risk predictions both in an individual consultation and for practice-level audit and quality improvement purposes, such as 'targeted' consultations for high risk individuals. In the context of impact studies how data

is presented and to whom is important. QRisk3's website gives a good example of the different ways risk can be presented that can be interpretable numerically and visually by patient and clinician alike (67). In an impact study whether the clinician is then guided by the risk prediction tool results such as a prompt stating NICE blood pressure targets for CKD (7), which are different to the non-CKD population, is also an important step in design.

Your risk of having a heart attack or stroke within the next 10 years is:

40.4%

In other words, in a crowd of 100 people with the same risk factors as you, 40 are likely to have a heart attack or stroke within the next 10 years.

Risk of
a heart attack or stroke

How does your 10-year score compare?

Your score	
Your 10-year QRISK®3 score	40.4%
The score of a healthy person with the same age, sex, and ethnicity*	31.5%
Relative risk**	1.3
Your QRISK®3 Healthy Heart Age***	> 84

Figure 9.1: An example of the different types of risk prediction tool output from QRisk3.

Further consideration for design include the level at which clustering occurs. Whilst practice level clustering may be attractive, the impact of the risk prediction tool on the different clinicians' behaviours within the practice will also need to be considered. Clinicians may also change their opinion in relation to the risk prediction tool, either positively or negatively, over time thus leading to a clinician time dependent effect. Further, if the result is shared with the patient then how this effects their decision making will also need to be considered.

In summary, clinical impact studies of risk prediction tools are vital to study how they influence decision making, both clinician and patient, and outcomes for a condition. However, they are rarely performed and when they are they present multiple methodological challenges (60, 186). Initial further work from this thesis may include studying the impact of decision making from results of the QRisk2, QRisk3, or an alternative multi-state models in CKD. This could then be followed by a full randomised cluster designed study of the risk prediction tool to assess its impact on clinical decision making and outcomes.

9.7.   Final Conclusions

This thesis has been built on three foundations. Firstly, that CKD is an independent risk factors for CV disease events. Secondly, the pathophysiology, and hence risk factors, are likely to be different in CKD compared to the general population. Thirdly, that general population risk prediction tools can accurately predict CV disease event risk in the general population.

Firstly, this thesis has suggested that in CKD the routinely collected risk factors of serum albumin, haemoglobin and phosphate have the strongest evidence of all the 'non-traditional risk factors' for predicting CV events. Secondly, a primary care cohort of CKD, LCC, with robust identification of CV event outcomes from primary and secondary care, can be used to externally validate CKD specific and general population risk prediction models. Thirdly, and perhaps the most surprisingly, given the initial hypothesis, that rather than underprediction of risk, as might be expected from CKD being an independent risk factor, overprediction occurred. This finding is likely to be due to the possible violation of the non-informative censoring assumption of the Cox model used. This is likely to have been due to the high rates of death and its related role as a competing risk in the model. Fourthly, risk prediction tools may be improved through the use competing risk models and the inclusion of additional risk factors such as eGFR, ACR and other non-traditional risk factors. Finally, implementation of these models will require further prospective studies to assess the impact on clinical decision making in CKD in relation to CV, ESRD and all-cause mortality outcomes.

References

1. Jha V, Garcia-Garcia G, Iseki K, Li Z, Naicker S, Plattner B, et al. Chronic kidney disease: global dimension and perspectives. The Lancet. 2013;382(9888):260-72.

2. Go AS, Chertow GM, Fan D, McCulloch CE, Hsu C-y. Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization. New England Journal of Medicine. 2004;351(13):1296-305.

3. Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. The Lancet. 2017;389(10075):1238-52.

4. Levey AS, Coresh J. Chronic kidney disease. The lancet. 2012;379(9811):165-80.

5. Levey AS, De Jong PE, Coresh J, Nahas ME, Astor BC, Matsushita K, et al. The definition, classification, and prognosis of chronic kidney disease: a KDIGO Controversies Conference report. Kidney international. 2011;80(1):17-28.

6. Group KDIGOKCW. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. Kidney inter, Suppl. 2013;3:1-150.

7. Excellence NIfHaC. Chronic kidney disease in adults: assessment and management 2014 [cited 2018 29th October]. Available from: https://www.nice.org.uk/guidance/cg182.

8. Smith HW, Chasis H, Goldring W, Ranges HA. Glomerular dynamics in the normal human kidney. The Journal of clinical investigation. 1940;19(5):751-64.

9. Kelly JT, Palmer SC, Wai SN, Ruospo M, Carrero J-J, Campbell KL, et al. Healthy Dietary Patterns and Risk of Mortality and ESRD in CKD: A Meta-Analysis of Cohort Studies. Clinical journal of the American Society of Nephrology : CJASN. 2017;12(2):272-9.

10. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Annals of internal medicine. 1999;130(6):461-70.

11. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Annals of internal medicine. 2009;150(9):604-12.

12. Inker LA, Eckfeldt J, Levey AS, Leiendecker-Foster C, Rynders G, Manzi J, et al. Expressing the CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) cystatin C equations for estimating GFR with standardized serum cystatin C values. American journal of kidney diseases. 2011;58(4):682-4.

13. Stevens LA, Coresh J, Schmid CH, Feldman HI, Froissart M, Kusek J, et al. Estimating GFR using serum cystatin C alone and in combination with serum creatinine: a pooled analysis of 3,418 individuals with CKD. American journal of kidney diseases. 2008;51(3):395-406.

14. Inker LA, Schmid CH, Tighiouart H, Eckfeldt JH, Feldman HI, Greene T, et al. Estimating glomerular filtration rate from serum creatinine and cystatin C. New England Journal of Medicine. 2012;367(1):20-9.

15. Schaeffner ES, Ebert N, Delanaye P, Frei U, Gaedeke J, Jakob O, et al. Two novel equations to estimate kidney function in persons aged 70 years or older. Annals of internal medicine. 2012;157(7):471-81.

16. Heerspink HJL, Gansevoort RT, Brenner BM, Cooper ME, Parving HH, Shahinfar S, et al. Comparison of different measures of urinary protein excretion for prediction of renal events. Journal of the American Society of Nephrology. 2010:ASN. 2010010063.

17. Guy M, Borzomato JK, Newall RG, Kalra PA, Price CP. Protein and albumin-to-creatinine ratios in random urines accurately predict 24 h protein and albumin loss in patients with kidney disease. Annals of clinical biochemistry. 2009;46(6):468-76.

18. Gansevoort RT, Brinkman J, Bakker SJ, De Jong PE, de Zeeuw D. Evaluation of measures of urinary albumin excretion. American journal of epidemiology. 2006;164(8):725-7.

19. Levey AS, Coresh J, Bolton K, Culleton B, Harvey KS, Ikizler TA, et al. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. American Journal of Kidney Diseases. 2002;39(2 SUPPL. 1).

20. Zoccali C, Kramer A, Jager KJ. Epidemiology of CKD in Europe: an uncertain scenario. Nephrology Dialysis Transplantation. 2010;25(6):1731-3.

21.   Partnership HQI. National Chronic Kidney Disease Audit 2017 [cited 2018 27th October]. Available from: https://www.lshtm.ac.uk/media/9951.

22.   Roth M, Roderick P, Mindell J. Kidney disease and renal function. Health survey for England. 2010;1.

23.   Stevens P, O'donoghue D, De Lusignan S, Van Vlymen J, Klebe B, Middleton R, et al. Chronic kidney disease management in the United Kingdom: NEOERICA project results. Kidney international. 2007;72(1):92-9.

24.   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 Clinical Practice. 2011;117(3):c213-c24.

25.   Kerr M. Chronic kidney disease in England: the human and financial cost. Insight Health Economics Ltd. 2012.

26.   Schiffrin EL, Lipman ML, Mann JF. Chronic kidney disease: effects on the cardiovascular system. Circulation. 2007;116(1):85-97.

27.   Mahmoodi BK, Matsushita K, Woodward M, Blankestijn PJ, Cirillo M, Ohkubo T, et al. Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without hypertension: a meta-analysis. The Lancet. 2012;380(9854):1649-61.

28.   Glassock RJ, Winearls C. An epidemic of chronic kidney disease: fact or fiction? : Oxford University Press; 2008.

29.   Moynihan R, Glassock R, Doust J. Chronic kidney disease controversy: how expanding definitions are unnecessarily labelling many people as diseased. Bmj. 2013;347:f4298.

30.   Winearls CG, Glassock RJ. Classification of chronic kidney disease in the elderly: pitfalls and errors. Nephron Clinical Practice. 2011;119(Suppl. 1):c2-c4.

31.   Organisation WH. Cardiovascular disease [cited 2018 29th October]. Available from: https://www.who.int/cardiovascular_diseases/en/.

32.   Tonelli M, Muntner P, Lloyd A, Manns BJ, Klarenbach S, Pannu N, et al. Risk of coronary events in people with chronic kidney disease compared with those with diabetes: a population-level cohort study. The Lancet. 2012;380(9844):807-14.

33.   Menon V, Gul A, Sarnak MJ. Cardiovascular risk factors in chronic kidney disease. Kidney international. 2005;68(4):1413-8.

34.   Covic A, Kothawala P, Bernal M, Robbins S, Chalian A, Goldsmith D. Systematic review of the evidence underlying the association between mineral metabolism disturbances and risk of all-cause mortality, cardiovascular mortality and cardiovascular events in chronic kidney disease. Nephrology Dialysis Transplantation. 2009;24(5):1506-23.

35.   Gansevoort RT, Correa-Rotter R, Hemmelgarn BR, Jafar TH, Heerspink HJL, Mann JF, et al. Chronic kidney disease and cardiovascular risk: epidemiology, mechanisms, and prevention. The Lancet. 2013;382(9889):339-52.

36.   Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. New England Journal of Medicine. 2005;352(16):1685-95.

37.   Libby P, Ridker PM, Hansson GK. Progress and challenges in translating the biology of atherosclerosis. Nature. 2011;473(7347):317.

38.   Faxon DP, Fuster V, Libby P, Beckman JA, Hiatt WR, Thompson RW, et al. Atherosclerotic vascular disease conference: writing group III: pathophysiology. Circulation. 2004;109(21):2617-25.

39.   Fishbein MC, Fishbein GA. Arteriosclerosis: facts and fancy. Cardiovascular Pathology. 2015;24(6):335-42.

40.   Williams B. Arterial ageing: from the shadows to centre stage in cardiovascular medicine. European Heart Journal Supplements. 2017;19(suppl_B):B1-B3.

41.   Edwards NC, Hirth A, Ferro CJ, Townend JN, Steeds RP. Subclinical abnormalities of left ventricular myocardial deformation in early-stage chronic kidney disease: the precursor of uremic cardiomyopathy? Journal of the American Society of Echocardiography. 2008;21(12):1293-8.

42.   Cai Q-Z, Lu X-Z, Lu Y, Wang AY-M. Longitudinal changes of cardiac structure and function in CKD (CASCADE study). Journal of the American Society of Nephrology. 2014:ASN. 2013080899.

43.   Shamseddin MK, Parfrey PS. Sudden cardiac death in chronic kidney disease: epidemiology and prevention. Nature Reviews Nephrology. 2011;7(3):145.

44.   Matsushita K, Ballew SH, Astor BC, Jong PEd, Gansevoort RT, Hemmelgarn BR, et al. Cohort profile: the chronic kidney disease prognosis consortium. International journal of epidemiology. 2012;42(6):1660-8.

45.   Smith KJ, Bleyer AJ, Little WC, Sane DC. The cardiovascular effects of erythropoietin. Cardiovascular Research. 2003;59(3):538-48.

46.   Sanchis-Gomar F, Garcia-Gimenez JL, Pareja-Galeano H, Romagnoli M, Perez-Quilis C, Lippi G. Erythropoietin and the heart: physiological effects and the therapeutic perspective. International journal of cardiology. 2014;171(2):116-25.

47.   Oberg BP, McMenamin E, Lucas F, McMonagle E, Morrow J, Ikizler T, et al. Increased prevalence of oxidant stress and inflammation in patients with moderate to severe chronic kidney disease. Kidney international. 2004;65(3):1009-16.

48.   Ridker PM, Rifai N, Clearfield M, Downs JR, Weis SE, Miles JS, et al. Measurement of C-reactive protein for the targeting of statin therapy in the primary prevention of acute coronary events. New England Journal of Medicine. 2001;344(26):1959-65.

49.   Ridker PM, Rifai N, Pfeffer MA, Sacks FM, Moye LA, Goldman S, et al. Inflammation, pravastatin, and the risk of coronary events after myocardial infarction in patients with average cholesterol levels. Circulation. 1998;98(9):839-44.

50.   Ridker PM, Danielson E, Fonseca FA, Genest J, Gotto Jr AM, Kastelein JJ, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. New England Journal of Medicine. 2008;359(21):2195.

51.   Ridker PM, MacFadyen J, Cressman M, Glynn RJ. Efficacy of rosuvastatin among men and women with moderate chronic kidney disease and elevated high-sensitivity C-reactive protein: a secondary analysis from the JUPITER (Justification for the Use of Statins in Prevention–an Intervention Trial Evaluating Rosuvastatin) trial. Journal of the American College of Cardiology. 2010;55(12):1266-73.

52.   Iff S, Craig JC, Turner R, Chapman JR, Wang JJ, Mitchell P, et al. Reduced estimated GFR and cancer mortality. American Journal of Kidney Diseases. 2014;63(1):23-30.

53.   Lowrance WT, Ordoñez J, Udaltsova N, Russo P, Go AS. CKD and the risk of incident cancer. Journal of the American Society of Nephrology. 2014;25(10):2327-34.

54.   Na SY, Sung JY, Chang JH, Kim S, Lee HH, Park YH, et al. Chronic kidney disease in cancer patients: an independent predictor of cancer-specific mortality. American journal of nephrology. 2011;33(2):121-30.

55.   James MT, Quan H, Tonelli M, Manns BJ, Faris P, Laupland KB, et al. CKD and risk of hospitalization and death with pneumonia. American Journal of Kidney Diseases. 2009;54(1):24-32.

56.   Ibrahim F, Hamzah L, Jones R, Nitsch D, Sabin C, Post FA, et al. Baseline kidney function as predictor of mortality and kidney disease progression in HIV-positive patients. American journal of kidney diseases. 2012;60(4):539-47.

57.   Romanowski K, Clark EG, Levin A, Cook VJ, Johnston JC. Tuberculosis and chronic kidney disease: an emerging global syndemic. Kidney international. 2016;90(1):34-40.

58.   Incalzi RA, Corsonello A, Pedone C, Battaglia S, Paglino G, Bellia V, et al. Chronic renal failure: a neglected comorbidity of COPD. Chest. 2010;137(4):831-7.

59.   Chiu H-Y, Huang H-L, Li C-H, Chen H-A, Yeh C-L, Chiu S-H, et al. Increased risk of chronic kidney disease in rheumatoid arthritis associated with cardiovascular complications–a national population-based cohort study. PLoS One. 2015;10(9):e0136508.

60.   Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS medicine. 2013;10(2):e1001381.

61.   Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, et al. Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. PLoS medicine. 2013;10(2):e1001380.

62.   Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. Annals of internal medicine. 2006;144(3):201-9.

63.   Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. Bmj. 2009;338:b606.

64.   Cox DR. Regression models and life-tables. Breakthroughs in statistics: Springer; 1992. p. 527-41.

65.   Damen JA, Hooft L, Schuit E, Debray TP, Collins GS, Tzoulaki I, et al. Prediction models for cardiovascular disease risk in the general population: systematic review. bmj. 2016;353:i2416.

66.   Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation. 1998;97(18):1837-47.

67.   Hippisley-Cox J, Coupland C, Brindle P. Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study. bmj. 2017;357:j2099.

68.   Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. Bmj. 2007;335(7611):136.

69.   Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. Bmj. 2008;336(7659):1475-82.

70.   Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Journal of the American College of Cardiology. 2014;63(25 Part B):2935-59.

71.   Conroy RM, on behalf of the Spg, Pyörälä K, on behalf of the Spg, Fitzgerald AP, on behalf of the Spg, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. European Heart Journal. 2003;24(11):987-1003.

72.   Excellence NIfHaC. Cardiovascular disease: risk assessment and reduction, including lipid modification. 2016.

73.   Tangri N, Kitsios GD, Inker LA, Griffith J, Naimark DM, Walker S, et al. Risk prediction models for patients with chronic kidney disease: a systematic review. Annals of internal medicine. 2013;158(8):596-603.

74.   Matsushita K, Coresh J, Sang Y, Chalmers J, Fox C, Guallar E, et al. Estimated glomerular filtration rate and albuminuria for prediction of cardiovascular outcomes: a collaborative meta-analysis of individual participant data. The lancet Diabetes & endocrinology. 2015;3(7):514-25.

75.   Excellence NIoHaC. Chronic kidney disease in adults: assessment and management (CG182). 2015.

76.   Association WM. WMA Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects 2013 [cited 2018 30th October]. Available from: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/.

77.   Together IPcaAoKdp. [cited 2018 27th October]. Available from: http://impakt.org.uk/.

78.   Digital N. [cited 2018 27th October]. Available from: https://digital.nhs.uk/services/nhais/morbidity-information-query-and-export-syntax-miquest.

79.   Digital N. Clinical Classifications Service [cited 2018 27th October]. Available from: https://isd.digital.nhs.uk/trud3/user/guest/group/61/home.

80.   Personal E-mail Correspondence, Dr Virginia Lee, Consultant Clinical Biochemist, University Hospital of Leicester. 2016.

81.   Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. Bmj. 2013;346:f2350.

82.   Rubin DB, editor Multiple imputations in sample surveys-a phenomenological Bayesian approach to nonresponse. Proceedings of the survey research methods section of the American Statistical Association; 1978: American Statistical Association.

83.   Rubin R. Multiple imputation for nonresponse in surveys (J Wiley & Sons, New York, NY). 1987.

84.   Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. Clinical trials. 2004;1(4):368-76.

85.   Eekhout I, de Boer MR, Twisk JW, de Vet HC, Heymans MW. Brief Report: Missing Data: A Systematic Review of How They Are Reported and Handled. Epidemiology. 2012:729-32.

86.   Schafer JL. Multiple imputation: a primer. Statistical methods in medical research. 1999;8(1):3-15.

87.   Dong Y, Peng C-YJ. Principled missing data methods for researchers. SpringerPlus. 2013;2(1):222.

88.   White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Statistics in medicine. 2011;30(4):377-99.

89.   Carlin J, Sterne A, White I, Royston P, Kenward M, Wood A. Multiple imputation needs to be used with care and reported in detail.

90.   Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. BMC medical research methodology. 2014;14(1):75.

91.   Burton A, Billingham LJ, Bryan S. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. Clinical Trials. 2007;4(2):154-61.

92.   Schafer JL. Analysis of incomplete multivariate data: CRC press; 1997.

93.   Moons KG, Donders RA, Stijnen T, Harrell FE. Using the outcome for imputation of missing predictor values was preferred. Journal of clinical epidemiology. 2006;59(10):1092-101.

94.   White IR, Royston P. Imputing missing covariate values for the Cox model. Statistics in medicine. 2009;28(15):1982-98.

95.   Barzi F, Woodward M. Imputations of missing values in practice: results from imputations of serum cholesterol in 28 cohort studies. American journal of epidemiology. 2004;160(1):34-45.

96.   Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Statistics in medicine. 1999;18(6):681-94.

97.   Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. Journal of clinical epidemiology. 2003;56(1):28-37.

98.   Carpenter J, Kenward M. Multiple imputation and its application: John Wiley & Sons; 2012.

99.   Royston P. Multiple imputation of missing values. Stata journal. 2004;4(3):227-41.

100.   Bodner TE. What improves with increased missing data imputations? Structural Equation Modeling. 2008;15(4):651-75.

101.   Cleves M, Gould W, Gould WW, Gutierrez R, Marchenko Y. An introduction to survival analysis using Stata: Stata press; 2008.

102.   Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. Journal of the American statistical association. 1958;53(282):457-81.

103.   Cleves M, Gould WW, Marchenko Y. An introduction to survival analysis using Stata. 3rd Edition ed: Stata press; 2016.

104.   Lambert P. STPM2: Stata module to estimate flexible parametric survival models. 2018.

105.   Cook NR. Quantifying the added value of new biomarkers: how and how not. Diagnostic and Prognostic Research. 2018;2(1):14.

106.   Harrell Jr FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. Jama. 1982;247(18):2543-6.

107.   Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. BMC medical research methodology. 2013;13(1):33.

108.   Major R, Xu G, Gray L, Brunskill N. Prevalence of Co-Morbidities by Ethnicity in a UK Primary Care CKD Cohort. American Society of Nephrology; New Orleans, United States of America2017.

109.   UK KR. Fellows Day 2018 Meeting Report. 2018:4.

110.   D'agostino RB, Belanger A, D'Agostino Jr RB. A suggestion for using powerful and informative tests of normality. The American Statistician. 1990;44(4):316-21.

111.   Corporation S.

112.   Organisation WH. International Statistical Classification of Diseases and Related Health Problems 10th Revision. 2010.

113. Digital N. Hospital Episode Statistics Data Dictionary.

114. Springate DA, Kontopantelis E, Ashcroft DM, Olier I, Parisi R, Chamapiwa E, et al. ClinicalCodes: an online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. PloS one. 2014;9(6):e99825.

115. Datalink CPR. An investigation of the completeness and accuracy of the recording of stroke events (including by subtype) in primary (CPRD) and secondary care databases (HES and ONS). 2016.

116. Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JP. Routinely collected data and comparative effectiveness evidence: promises and limitations. Canadian Medical Association Journal. 2016;188(8):E158-E64.

117. Kyriacou DN, Lewis RJ. Confounding by indication in clinical research. Jama. 2016;316(17):1818-9.

118. Roland M, Guthrie B. Quality and Outcomes Framework: what have we learnt? BMJ. 2016;354:i4060.

119. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Statistics in medicine. 2016;35(2):214-26.

120. PROSPERO. Cardiovascular disease risk factors in chronic kidney disease: a systematic review and meta-analysis. [cited 2018 27th October]. Available from: http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD420160361 87. .

121. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Annals of internal medicine. 2013;158(4):280-6.

122. Green S, Higgins J. Cochrane handbook for systematic reviews of interventions. Version; 2005.

123. Lu Y, Hajifathalian K, Ezzati M, Woodward M, Rimm EB, Danaei G. Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1· 8 million participants. Elsevier; 2014.

124. Seiler S, Rogacev KS, Roth HJ, Shafein P, Emrich I, Neuhaus S, et al. Associations of FGF-23 and sKlotho with cardiovascular outcomes among

patients with CKD stages 2–4. Clinical Journal of the American Society of Nephrology. 2014:CJN. 07870713.

125. Chen S-C, Hung C-C, Tsai Y-C, Huang J-C, Kuo M-C, Lee J-J, et al. Association of cholesterol levels with mortality and cardiovascular events among patients with CKD and different amounts of proteinuria. Clinical Journal of the American Society of Nephrology. 2013:CJN. 02350213.

126. Yoshitomi R, Nakayama M, Ura Y, Kuma K, Nishimoto H, Fukui A, et al. Ankle-brachial blood pressure index predicts cardiovascular events and mortality in Japanese patients with chronic kidney disease not on dialysis. Hypertension Research. 2014;37(12):1050.

127. Meijers B, Poesen R, Claes K, Dietrich R, Bammens B, Sprangers B, et al. Soluble urokinase receptor is a biomarker of cardiovascular disease in chronic kidney disease. Kidney international. 2015;87(1):210-6.

128. Goicoechea M, de Vinuesa SG, Verdalles U, Ruiz-Caro C, Ampuero J, Rincón A, et al. Effect of allopurinol in chronic kidney disease progression and cardiovascular risk. Clinical Journal of the American Society of Nephrology. 2010:CJN. 01580210.

129. Spoto B, Mattace-Raso F, Sijbrands E, Leonardis D, Testa A, Pisano A, et al. Association of IL-6 and a functional polymorphism in the IL-6 gene with cardiovascular events in patients with CKD. Clinical Journal of the American Society of Nephrology. 2015;10(2):232-40.

130. McMullan CJ, Bakris GL, Phillips RA, Forman JP. Association of BP variability with mortality among African Americans with CKD. Clinical Journal of the American Society of Nephrology. 2013:CJN. 10131012.

131. Yilmaz MI, Solak Y, Saglam M, Cayci T, Acikel C, Unal HU, et al. The relationship between IL-10 levels and cardiovascular events in patients with CKD. Clinical Journal of the American Society of Nephrology. 2014:CJN. 08660813.

132. Kim RB, Morse BL, Djurdjev O, Tang M, Muirhead N, Barrett B, et al. Advanced chronic kidney disease populations have elevated trimethylamine N-oxide levels associated with increased cardiovascular events. Kidney international. 2016;89(5):1144-52.

133. Locatelli F, Eckardt K-U, Macdougall IC, Tsakiris D, Clyne N, Burger H-U, et al. Value of N-terminal brain natriuretic peptide as a prognostic marker in

patients with CKD: results from the CREATE study. Current medical research and opinion. 2010;26(11):2543-52.

134. Dobre M, Yang W, Chen J, Drawz P, Hamm LL, Horwitz E, et al. Association of serum bicarbonate with risk of renal and cardiovascular outcomes in CKD: a report from the Chronic Renal Insufficiency Cohort (CRIC) study. American Journal of Kidney Diseases. 2013;62(4):670-8.

135. Alderson HV, Ritchie JP, Middleton R, Larsson A, Larsson TE, Kalra PA. FGF-23 and Osteoprotegerin but not Fetuin-A are associated with death and enhance risk prediction in non-dialysis chronic kidney disease stages 3–5. Nephrology. 2016;21(7):566-73.

136. Bowling CB, Pitt B, Ahmed MI, Aban IB, Sanders PW, Mujib M, et al. Hypokalemia and outcomes in patients with chronic heart failure and chronic kidney disease: findings from propensity-matched studies. Circulation: Heart Failure. 2010:CIRCHEARTFAILURE. 109.899526.

137. Hasegawa M, Ishii J, Kitagawa F, Takahashi K, Hayashi H, Koide S, et al. Urinary neutrophil gelatinase-associated lipocalin as a predictor of cardiovascular events in patients with chronic kidney disease. Heart and vessels. 2015;30(1):81-8.

138. Paoletti E, De Nicola L, Gabbai FB, Chiodini P, Ravera M, Pieracci L, et al. Associations of left ventricular hypertrophy and geometry with adverse outcomes in patients with CKD and hypertension. Clinical Journal of the American Society of Nephrology. 2016;11(2):271-9.

139. Chen S-C, Chang J-M, Tsai Y-C, Huang J-C, Chen L-I, Su H-M, et al. Ratio of transmitral E-wave velocity to early diastole mitral annulus velocity with cardiovascular and renal outcomes in chronic kidney disease. Nephron Clinical Practice. 2013;123(1-2):52-60.

140. De Nicola L, Gabbai FB, Agarwal R, Chiodini P, Borrelli S, Bellizzi V, et al. Prevalence and prognostic role of resistant hypertension in chronic kidney disease patients. Journal of the American College of Cardiology. 2013;61(24):2461-7.

141. Górriz JL, Molina P, Cerverón MJ, Vila R, Bover J, Nieto J, et al. Vascular calcification in patients with nondialysis CKD over 3 years. Clinical Journal of the American Society of Nephrology. 2015:CJN. 07450714.

142. Tonelli M, Keech A, Shepherd J, Sacks F, Tonkin A, Packard C, et al. Effect of pravastatin in people with diabetes and chronic kidney disease. Journal of the American Society of Nephrology. 2005;16(12):3748-54.

143. Theilade S, Claggett B, Hansen T, Skali H, Lewis E, Solomon SD, et al. Pulse pressure is not an independent predictor of outcome in type 2 diabetes patients with chronic kidney disease and anemia—the Trial to Reduce Cardiovascular Events with Aranesp Therapy (TREAT). Journal of human hypertension. 2016;30(1):46.

144. McMurray JJ, Uno H, Jarolim P, Desai AS, De Zeeuw D, Eckardt K-U, et al. Predictors of fatal and nonfatal cardiovascular events in patients with type 2 diabetes mellitus, chronic kidney disease, and anemia: an analysis of the Trial to Reduce cardiovascular Events with Aranesp (darbepoetin-alfa) Therapy (TREAT). American heart journal. 2011;162(4):748-55. e3.

145. Weiss JW, Johnson ES, Petrik A, Smith DH, Yang X, Thorp ML. Systolic blood pressure and mortality among older community-dwelling adults with CKD. American Journal of Kidney Diseases. 2010;56(6):1062-71.

146. Kovesdy CP, Trivedi BK, Kalantar-Zadeh K, Anderson JE. Association of low blood pressure with increased mortality in patients with moderate to severe chronic kidney disease. Nephrology Dialysis Transplantation. 2006;21(5):1257-62.

147. Trialists'Collaboration BPLT. Blood pressure lowering and major cardiovascular events in people with and without chronic kidney disease: meta-analysis of randomised controlled trials. Bmj. 2013;347:f5680.

148. Cheung AK, Rahman M, Reboussin DM, Craven TE, Greene T, Kimmel PL, et al. Effects of intensive BP control in CKD. Journal of the American Society of Nephrology. 2017;28(9):2812-23.

149. Tonelli M, Muntner P, Lloyd A, Manns B, Klarenbach S, Pannu N, et al. Association between LDL-C and risk of myocardial infarction in CKD. Journal of the American Society of Nephrology. 2013;24(6):979-86.

150. Sutton AJ, Higgins JP. Recent developments in meta-analysis. Statistics in medicine. 2008;27(5):625-50.

151. Fleiss JL, Gross AJ. Meta-analysis in epidemiology, with special reference to studies of the association between exposure to environmental

tobacco smoke and lung cancer: a critique. Journal of clinical epidemiology. 1991;44(2):127-39.

152.   Shrier I, Boivin J-F, Steele RJ, Platt RW, Furlan A, Kakuma R, et al. Should Meta-Analyses of Interventions Include Observational Studies in Addition to Randomized Controlled Trials? A Critical Examination of Underlying Principles. American Journal of Epidemiology. 2007;166(10):1203-9.

153.   Hemingway H, Philipson P, Chen R, Fitzpatrick NK, Damant J, Shipley M, et al. Evaluating the quality of research into a single prognostic biomarker: a systematic review and meta-analysis of 83 studies of C-reactive protein in stable coronary artery disease. PLoS medicine. 2010;7(6):e1000286-e.

154.   Riley RD, Abrams KR, Sutton AJ, Lambert PC, Jones DR, Heney D, et al. Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. British journal of cancer. 2003;88(8):1191-8.

155.   Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med. 2006;25(1):127-41.

156.   Laber EB, Shedden K. Statistical Significance and the Dichotomization of Evidence: The Relevance of the ASA Statement on Statistical Significance and p-values for Statisticians. J Am Stat Assoc. 2017;112(519):902-4.

157.   Altman DG, Lausen B, Sauerbrei W, Schumacher M. Dangers of Using "Optimal" Cutpoints in the Evaluation of Prognostic Factors. JNCI: Journal of the National Cancer Institute. 1994;86(11):829-35.

158.   Wang, Guo, Li, Hu, Zhao, Ma, et al. Predictive value of ankle-brachial index to all-cause mortality and cardiovascular mortality in Chinese patients with chronic kidney disease. Vasa. 2012;41(3):205-13.

159.   Mehta NN, the Chronic Renal Insufficiency Cohort Study I, Matthews GJ, the Chronic Renal Insufficiency Cohort Study I, Krishnamoorthy P, the Chronic Renal Insufficiency Cohort Study I, et al. Higher plasma CXCL12 levels predict incident myocardial infarction and death in chronic kidney disease: findings from the Chronic Renal Insufficiency Cohort study. European Heart Journal. 2014;35(31):2115-22.

160.   Liabeuf S, Desjardins L, Diouf M, Temmar M, Renard C, Choukroun G, et al. The Addition of Vascular Calcification Scores to Traditional Risk Factors

Improves Cardiovascular Risk Assessment in Patients with Chronic Kidney Disease. PLOS ONE. 2015;10(7):e0131707.

161.   Matsushita K, Sang Y, Ballew SH, Astor BC, Hoogeveen RC, Solomon SD, et al. Cardiac and kidney markers for cardiovascular prediction in individuals with chronic kidney disease: the Atherosclerosis Risk in Communities study. Arterioscler Thromb Vasc Biol. 2014;34(8):1770-7.

162.   Weiner DE, Tighiouart H, Elsayed EF, Griffith JL, Salem DN, Levey AS, et al. The Framingham Predictive Instrument in Chronic Kidney Disease. Journal of the American College of Cardiology. 2007;50(3):217.

163.   Shlipak MG, Fried LF, Cushman M, et al. Cardiovascular mortality risk in chronic kidney disease: Comparison of traditional and novel risk factors. JAMA. 2005;293(14):1737-45.

164.   Excellence NIfHaC. Surveillance report 2018 – Cardiovascular disease: risk assessment and reduction, including lipid modification (2014) NICE guideline CG181
https://www.nice.org.uk/guidance/cg181/resources/surveillance-report-2018-cardiovascular-disease-risk-assessment-and-reduction-including-lipid-modification-2014-nice-guideline-cg181-4724759773/chapter/Surveillance-decision?tab=evidence2018 [cited 2018 27th October].

165.   International Prospective Register of Systematic Reviews (PROSPERO) [cited 2018 27th October]. Available from: https://www.crd.york.ac.uk/prospero/.

166.   Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. Bmj. 2010;340:c2442.

167.   Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMC medicine. 2015;13(1):1.

168.   Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. Journal of clinical epidemiology. 2015;68(1):25-34.

169.   Rosano G, Vitale C, Marazzi G, Volterrani M. Menopause and cardiovascular disease: the evidence. Climacteric. 2007;10(sup1):19-24.

170. Campigotto F, Weller E. Impact of informative censoring on the Kaplan-Meier estimate of progression-free survival in phase II clinical trials. Journal of Clinical Oncology. 2014;32(27):3068.

171. Leung K-M, Elashoff RM, Afifi AA. CENSORING ISSUES IN SURVIVAL ANALYSIS. Annual Review of Public Health. 1997;18(1):83-104.

172. Zhang J, Heitjan DF. Nonignorable censoring in randomized clinical trials. Clinical Trials. 2005;2(6):488-96.

173. Dickman PW, Coviello E. Estimating and modeling relative survival. The Stata Journal. 2015;15(1):186-215.

174. Consortium CKDP. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. The Lancet. 2010;375(9731):2073-81.

175. Grams ME, Sang Y, Ballew SH, Carrero JJ, Djurdjev O, Heerspink HJ, et al. Predicting timing of clinical outcomes in patients with chronic kidney disease and severely decreased glomerular filtration rate. Kidney international. 2018;93(6):1442-51.

176. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. Bmj. 2013;346:e5595.

177. Hingorani AD, van der Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. Bmj. 2013;346:e5793.

178. Kiberd B. The chronic kidney disease epidemic: stepping back and looking forward. Journal of the American Society of Nephrology. 2006;17(11):2967-73.

179. Major RW, Oozeerally I, Dawson S, Riddleston H, Gray LJ, Brunskill NJ. Aspirin and cardiovascular primary prevention in non-endstage chronic kidney disease: A meta-analysis. Atherosclerosis. 2016;251:177-82.

180. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. Statistics in medicine. 2007;26(11):2389-430.

181. Austin PC, Lee DS, Fine JP. Introduction to the analysis of survival data in the presence of competing risks. Circulation. 2016;133(6):601-9.

182. Dafni U. Landmark analysis at the 25-year landmark point. Circulation: Cardiovascular Quality and Outcomes. 2011;4(3):363-71.

183.	Goldstein BA, Navar AM, Pencina MJ, Ioannidis J. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. Journal of the American Medical Informatics Association. 2017;24(1):198-208.

184.	Paige E, Barrett J, Stevens D, Keogh RH, Sweeting MJ, Nazareth I, et al. Landmark Models for Optimizing the Use of Repeated Measurements of Risk Factors in Electronic Health Records to Predict Future Disease Risk. American journal of epidemiology. 2018:kwy018.

185.	Brankovic M, Kardys I, Hoorn EJ, Baart S, Boersma E, Rizopoulos D. Personalized dynamic risk assessment in nephrology is a next step in prognostic research. Kidney international. 2018.

186.	Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KG. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. Diagnostic and Prognostic Research. 2018;2(1):11.

187.	Portela M, Pronovost P, Woodcock T, Carter P, Dixon-Woods M. How to study improvement interventions: a brief overview of possible study types. BMJ Qual Saf. 2015; 24: 325–36.

Appendices

2.1 Research Ethics Committee Approval Letter

NHS
Health Research Authority

East Midlands - Leicester Central Research Ethics Committee

The Old Chapel
Royal Standard Place
Nottingham
NG1 6FS

> Please note: This is the favourable opinion of the REC only and does not allow you to start your study at NHS sites in England until you receive HRA Approval

15 August 2016

Professor Nigel Brunskill
Professor of Renal Medicine
University of Leicester
John Walls Renal Unit
Leicester General Hospital
Gwendolen Road, Leicester
LE5 4PW

Dear Professor Brunskill,

Study title:	The Leicester City and County Chronic Kidney Disease Cohort
REC reference:	16/EM/0315
Protocol number:	UNOLE 0576
IRAS project ID:	197145

The Research Ethics Committee reviewed the above application at the meeting held on 05 August 2016. Thank you to Dr Rupert Major for attending to discuss the applications.

We plan to publish your research summary wording for the above study on the HRA website, together with your contact details. Publication will be no earlier than three months from the date of this favourable opinion letter. The expectation is that this information will be published for all studies that receive an ethical opinion but should you wish to provide a substitute contact point, wish to make a request to defer, or require further information, please contact the REC Manager, Ellen Swainston, NRESCommittee.EastMidlands-LeicesterCentral@nhs.net. Under very limited circumstances (e.g. for student research which has received an unfavourable opinion), it may be possible to grant an exemption to the publication of the study.

Ethical opinion

The members of the Committee present gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below. .

Conditions of the favourable opinion

The REC favourable opinion is subject to the following conditions being met prior to the start of the study.

1) The REC name is to be included under the section "Who has reviewed the study?" in the participant information sheet.
2) All documents are to be proof read for typographical and grammatical errors.
3) The PCT's are to be removed from the IRAS form and replaced with the correct CCG's

You should notify the REC once all conditions have been met (except for site approvals from host organisations) and provide copies of any revised documentation with updated version numbers. Revised documents should be submitted to the REC electronically from IRAS. The REC will acknowledge receipt and provide a final list of the approved documentation for the study, which you can make available to host organisations to facilitate their permission for the study. Failure to provide the final versions to the REC may cause delay in obtaining permissions.

Management permission must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements. Each NHS organisation must confirm through the signing of agreements and/or other documents that it has given permission for the research to proceed (except where explicitly specified otherwise).

Guidance on applying for HRA Approval (England)/ NHS permission for research is available in the Integrated Research Application System, at www.hra.nhs.uk *or at* http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of management permissions from host organisations.

Registration of Clinical Trials

All clinical trials (defined as the first four categories on the IRAS filter page) must be registered on a publically accessible database. This should be before the first participant is recruited but no later than 6 weeks after recruitment of the first participant.

There is no requirement to separately notify the REC but you should do so at the earliest opportunity e.g. when submitting an amendment. We will audit the registration details as part of the annual progress reporting process.

To ensure transparency in research, we strongly recommend that all research is registered but for non-clinical trials this is not currently mandatory.

If a sponsor wishes to request a deferral for study registration within the required timeframe, they should contact hra.studyregistration@nhs.net. The expectation is that all clinical trials will be registered, however, in exceptional circumstances non registration may be permissible with prior agreement from the HRA. Guidance on where to register is provided on the HRA website.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Ethical review of research sites

NHS Sites

The favourable opinion applies to all NHS sites taking part in the study taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Summary of discussion at the meeting

The committee complemented the applicant on a well written application.

Recruitment arrangements and access to health information, and fair participant selection

The committee questioned who the contemporary cohort of participants was in the study and how long they would have had CKD, as it was not clear from the information provided. *The applicant explained that the cohort would have been recognised at the point of time they would go into the practice. They would see who see who had been diagnosed with CKD five years previously and follow those patients up for a further five years.*

The committee commented that they had noticed information included in the application referring to Northamptonshire and asked what their involvement in the study was. *It was clarified that it was part of a completed trial which was separate to this trial but would be using similar methods from that study to collect data on this.*

The committee asked if that particular study had been published and peer reviewed. *The applicant confirmed that cross sectional data had been published and that the clinical data was currently being actioned.*

Care and protection of research participants; respect for potential and enrolled participants' welfare and dignity

The committee commented that there had been no information included in the application as to who would have access to patient data. *The applicant clarified that clinical trial unit co-investigators would have access to that information. It was also said that the administration team and regulatory authorities would have access as part of their audit.*

Informed consent process and the adequacy and completeness of participant information

The committee noted that the REC name had not been included in the participant information sheet.

Other general comments

The committee commented that the application was to be checked for typographical and grammatical errors.

It was noted that the insurance certificate for the study had expired and commented that it should be renewed before the start of the study.

It had been noted that part C of the IRAS form referred to PCT's instead of CCG's and asked why that was. The committee agreed that the information was to be updated. *The applicant explained that the IRAS system had pulled through an old address from the address book on the system; however, the details could be updated to reflect the three CCG's.*

The committee had queried why the Chief Investigator had not booked the application in as part of the proportionate review process. *The applicant stated that they had considered doing that but did not think that the study met the criteria to go through the review.*

Approved documents

The documents reviewed and approved at the meeting were:

Document	Version	Date
Evidence of Sponsor insurance or indemnity (non NHS Sponsors only) [Sponsor Insurance]		14 June 2016
IRAS Application Form [IRAS_Form_05072016]		05 July 2016
IRAS Checklist XML [Checklist_05072016]		05 July 2016
Letter from funder [Funder Confirmation]		27 April 2015
Letters of invitation to participant [Practice Invitation Letter]	1.0	10 June 2016
Other [Evidence of Sponsor insurance or indemnity (non-NHS Sponsors only)]		
Other [Evidence of Sponsor insurance or indemnity (non-NHS Sponsors only)]		
Other [Funder Confirmation]		27 April 2015
Participant consent form [Consent Form]	1.0	10 June 2016
Participant information sheet (PIS) [Practice Information Leaflet]	1.0	10 June 2016
Participant information sheet (PIS) [Expression of Interest Form]	1.0	10 June 2016
Research protocol or project proposal [Study Protocol]	1.0	10 June 2016
Summary CV for Chief Investigator (CI) [Professor Brunskill CV]		30 June 2016

Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the attached sheet.

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

<u>Reporting requirements</u>

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The HRA website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

User Feedback

The Health Research Authority is continually striving to provide a high quality service to all applicants and sponsors. You are invited to give your view of the service you have received and the application procedure. If you wish to make your views known please use the feedback form available on the HRA website: http://www.hra.nhs.uk/about-the-hra/governance/quality-assurance/

HRA Training

We are pleased to welcome researchers and R&D staff at our training days – see details at http://www.hra.nhs.uk/hra-training/

16/EM/0315	Please quote this number on all correspondence

With the Committee's best wishes for the success of this project.

Yours sincerely,

Mr Ken Willis
Chair

E-mail: nrescommittee.eastmidlands-leicestercentral@nhs.net

Enclosures: *List of names and professions of members who were present at the meeting and those who submitted written comments*

              "After ethical review – guidance for researchers"

Copy to: *Dr Diane Delahooke*
              Ms Debbie Wall, Research and Development Lead

2.2 Practice Information Leaflet

UNIVERSITY OF LEICESTER

University Hospitals of Leicester NHS
NHS Trust

Kidney Research UK
Funding research to save lives

NHS
National Institute for
Health Research

IMPAKT

Practice Information Leaflet
The Leicester City & County Chronic Kidney Disease Cohort (LCC-CKD)

IRAS Number: 197145

Chief Investigator: Professor Nigel Brunskill

Summary

Chronic Kidney Disease (CKD) is a common condition in adults and is associated with increased risk of cardiovascular disease (CV) events. Its epidemiology is poorly understood in both a primary care setting and in those of non-white ethnicity. This project proposes creating an observational cohort from anonymised primary care patient records to study the epidemiology of CKD, including risk factors associated with CV disease.

Background and Purpose of Study

CKD is a common condition affecting 6-8% of the adult population. It is often associated with hypertension, diabetes mellitus and CV disease. Individuals with CKD are at higher risk of death, CV events and endstage renal failure (requiring dialysis or transplantation) compared to those without CKD. CKD and these associated complications are also more common in non-white ethnic groups.

Large primary care epidemiological studies of CKD and its complications are lacking. Most available evidence relates to findings from clinical trials in secondary care, or other selected research populations. CV risk stratification is also poorly understood in CKD. Currently, general population CV risk tools are thought to underestimate the risk of CV events in CKD.

Your practice is invited to participate in the establishment of a CKD cohort to address two main research aims:

1. To establish the factors influencing the natural history of CKD and related complications in a large, multi-ethnic primary care based CKD cohort
2. To validate the accuracy of a CV risk score in individuals with CKD

A retrospective cohort will be created with up to 5 years of follow-up data from the date of data extraction. This will be formed from the electronic clinical records of participating practices by identifying individuals with CKD on the basis of serum creatinine and urinary protein results (dipstick, ACR and PCR).

What is involved in participating?

Your practice's involvement in the study would involve a single visit by a researcher. The visit typically lasts less than one hour but during this time the researcher would require supervision by a member of your staff. A practice staff member would need to run a standard 'MIQUEST' search on your clinical system. The MIQUEST search can be set up and run either before or during the researcher's visit to your practice. The study's data extraction tool, IMPAKT (www.impakt.org.uk), would then produce an anonymised version of the data to transfer to the University of Leicester's Clinical Trials Unit for secure digital storage. It is unlikely that your practice will incur any direct costs in relation to this project.

Who will have access to the data?

Your practice's anonymised data will be principally accessed by the research team. The data will be stored securely in the Leicester Clinical Trials Unit where access is password protected and limited to the research team. In order to comply with research data management regulations, the data may be reviewed by University of Leicester (in its role as the study's sponsor), the NHS or other regulatory bodies. Their access would be only for audit and monitoring purposes. The researchers may, in the future, collaborate with other external research consortia and share redacted data from your practice. Neither your patients nor your practice would be identifiable should this occur. You would be contacted separately regarding this and, of course, would be free to opt out of this process.

What are the potential benefits to my practice?

As well as actively involving your practice in clinical research, the researchers can provide, if requested, summary results in relation to the management of CKD in your practice to support quality improvement. This will include reference to both NICE guidance and the latest QOF standards for CKD and, because of their high prevalence in CKD, diabetes mellitus and hypertension.

What are the risks of participating in the study?

The main risk in relation to this study is in relation to information governance. No patient identifiable data will be removed from your practice. The researcher will not have access to patient identifiable information from your clinical system unless their role includes clinical work with your practice. All researchers involved in the study have undergone research training including specific information governance training. The anonymised data removed from your practice will be stored on password protected encrypted digital storage devices.

What if I am harmed by the study?

It is very unlikely that there would be any harm incurred by taking part in this type of research study. However, if you wish to complain or have any concerns about the way you have been approached or treated in connection with the study, you should ask to speak to Professor Nigel Brunskill who will do his best to answer your questions. He can be contacted by email (njb18@le.ac.uk) or phone (0116 258 8043).

In the event that something does go wrong and you are harmed during the research and this is due to someone's negligence then you may have grounds for a legal action for compensation against University of Leicester. You may, however, have to pay your legal costs.

Can we withdraw our consent?

Yes. If you decided to withdraw from the study please contact Professor Nigel Brunskill, the chief investigator. Data related to your practice can be removed from the cohort and digitally destroyed.

Who is the study funded by?

The study is funded by Kidney Research UK. The development of the data extraction tool, IMPAKT (www.impakt.org.uk), was funded by the National Institute of Health Research.

Who has reviewed this study?

The study has been reviewed by University of Leicester, Leicester City Clinical Commissioning Group (CCG), West Leicestershire CCG, East Leicestershire and Rutland CCG and Leicester Central Research Ethics Committee.

Do the researchers involved have any conflicts of interest?

IMPAKT was created by and updated for this project by one of the co-investigators, Dr David Shepherd, a Leicester GP. IMPAKT and its intellectual property rights are owned by Dr Shepherd and University of Leicester. The co-investigators have no relevant commercial conflicts of interest to declare.

If my practice would like to take part, what should we do next?

Please complete the enclosed declaration of interest form and return it in the stamp addressed envelope. Once we have received this, you will then be contact to arrange a date for a researcher to visit your practice to answer any further questions you may have, complete the consent process and the data extraction.

If we decide not to take part what will happen?

Participating is entirely voluntary and we understand not all practices will wish to be involved. If this is the case, please could you still complete the enclosed form and return it so that we know not to contact you again.

We are not sure whether the study is right for our practice, who can we contact to discuss this further?

Please contact either Professor Nigel Brunskill or Dr Rupert Major (rwlm2@le.ac.uk or 07510 303 405) to discuss the project further. Alternatively, return the enclosed form requesting more information and we will contact you to discuss the project in more detail.

Whatever your decision is, we would like to thank you for taking the time to read about our research project.

2.3 LCC Cohort Study Consent Form

University Hospitals of Leicester **NHS** NHS Trust

UNIVERSITY OF LEICESTER

Kidney Research UK Funding research to save lives

NHS National Institute for Health Research

IMPAKT

Practice Consent Form

The Leicester City & County Chronic Kidney Disease Cohort (LCC-CKD)

Please initial box

IRAS Number: 197145
Chief Investigator: Professor Nigel Brunskill

1. I confirm that I have read the Practice Information Leaflet (dated 23rd August 2016, Version 1.1) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.

2. I understand that my practice's participation is voluntary and that we are free to withdraw at any time without giving any reason and without affecting my practice's legal rights.

3. I understand that anonymised data in relation to my practice's patients will be collected and transferred for secure storage at Leicester Clinical Trials Unit.

4. In addition to research use, I understand any data provided may be reviewed by appropriate regulatory authorities for the purposed of audit and regulation.

5. I understand that the information collected may be used to support other research in the future, and may be shared anonymously with other researchers.

6. I confirm I am an authorised research representative of the general practice named below and that the practice agrees to take part in the above study.

Name and Address of Practice

_____

_____

_____

_____

_____   _____   _____
Name of Authorised Representative   Date   Signature

_____   _____   _____
Name of Person taking consent   Date   Signature

4.1 Definitions of Co-morbidities

History of ischaemic heart disease	Positive history of coded myocardial infarction, bypass grafting, percutaneous coronary intervention
History of heart failure	Positive history of coded heart failure
History of stroke	Positive history of coded stroke
Diabetes mellitus	HbA1c ≥6.5%, or coded fasting glucose ≥7.0 mmol/L, or any glucose ≥11.1 mmol/L, or use of glucose lowering drugs, or coded diagnosis of diabetes mellitus
Diabetes type	Type 1 or Type 2 diabetes mellitus based on code
Hypertension	Systolic blood pressure ≥140 mmHg, diastolic blood pressure ≥90 mmHg, use of anti-hypertensive drugs, or coded diagnosis of hypertension

4.2 CV Event Codes

Primary Care Read Codes

IMPAKT searched for and extracted the following codes for CV related events:

Cerebrovascular disease

> G6%, X00D%,XE0VJ%,Xa00%, Xa0kZ, XaB%, XaJgQ, XaQbK,

Ischaemic heart disease

> G3%, Gyu3%, X200%, XaIf1, XaJX0, Xa0YL, XE0Uh, XE2aA

Where '%' refers to any sub-code of the one listed.

Within the above identified codes the following codes for specific CV events were identified:

Atherosclerotic stroke

G63, G631-1, G641-1, G64z-2, G64z1-1, G6770, G630, G631, G631-, G631., G632, G63y0, G63z., G64, G64-1, G64-2, G64-3, G64.., G640, G640., G6400, G641, G6410, G64z, G64z-, G64z., G64z0, G64z1, G64z2, G64z3, G64z4, G65, G65-1, G65-2, G65-3, G65.., G650, G650., G650-1, G651, G651., G6510, G65y., G66, G66-1, G66-2, G66-3, G66-9, G66.., G660, G660., G661, G662, G662., G663, G663., G664, G664., G665., G665, G667, G667., G668, G668., G671, G6711, G676, Gyu63, Gyu64, G6773, X00D1, X00D3, X00D5, X00D6, X00D7, X00D8, X00D9, X00DA, X00DB, X00DG, XE0VJ, Xa00I, Xa00J, Xa00K, Xa00M, Xa00R, Xa0kZ, XaB4Z, XaBEC, XaBED, XaJgQ, XaQbK

Myocardial infarction

G30, G30-4, G30-5, G30-7, G30-9, G30.., G300, G300., G301, G301., G3010, G3011, G301z, G30-2, G302, G302., G303, G303., G304, G304., G305, G305., G306., G307, G307., G3070, G3071, G308, G308., G30X0, G30y., G30y2, G30yz, G30z, G30z., G31, G310, G32, G32-1, G32-2, G32.., G33z5, G35.., G350, G351, G351., G353., G35X., G38.., Gyu34, X200d, X200L, X200M, X200P, X200X, X200A, X200E, X200K, X200V, XaIf1, XaJX0, Xa0YL, XE0Uh, XE2aA

Hospital Episode Statistics

ICD-10 admission and in hospital events coded as:

Atherosclerotic stroke   I63.0-I63.6

Myocardial infarction   I21–I23, and I25.6

MINAP

Initial diagnosis identified as:

> "Definite myocardial infarction" or "Acute coronary syndrome"

Or, ECG determining treatment coded as:

> "ST segment elevation"

Or, Discharge diagnosis identified as:

> "Myocardial infarction (ST elevation)", "Threatened MI", "Myocardial Infarction (unconfirmed)", "Acute coronary syndrome (troponin positive) nSTEMI", "Acute coronary syndrome (troponin negative)" or "Acute coronary syndrome (troponin positive)"

5.1 OVID Medline Literature Search Strategy for 1st Systematic Review

Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R), 1946 to 20th October 2017

1   exp RENAL INSUFFICIENCY, CHRONIC/
2   "chronic renal".ti,ab.
3   "chronic kidney".ti,ab.
4   "ckd".ti,ab.
5   exp CARDIOVASCULAR DISEASES/
6   "cardi*".ti,ab.
7   "heart*".ti,ab.
8   "coronary*".ti,ab.
9   "myocard*".ti,ab.
10   "ischem*".ti,ab.
11   "ischaem*".ti,ab.
12   "stroke*".ti,ab.
13   "cerebrovasc*".ti,ab.
14   "cerebral vascular*".ti,ab.
15   "cohort*".ti,ab.
16   exp Cohort Studies/
17   "random*".ti,ab.
18   "rct*".ti,ab.
19   exp Randomized Controlled Trials as Topic/
20   "risk*".tw.
21   "prognosis*".tw.
22   "predict*".tw.
23   "associat*".tw.
24   "factor*".tw.
25   20 or 21 or 22 or 23 or 24
26   5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14
27   15 or 16 or 17 or 18 or 19
28   1 or 2 or 3 or 4
29   25 and 26 and 27 and 28
30   limit 29 to yr="1999 -Current"
31   limit 30 to (english language and humans and "all adult (19 plus years)")
32   limit 31 to (clinical study or clinical trial, all or clinical trial or meta analysis or observational study or pragmatic clinical trial or randomized controlled trial or systematic reviews or validation studies)

5.2 Embase Literature Search Strategy for 1st Systematic Review

exp RENAL INSUFFICIENCY, CHRONIC/ OR "chronic renal".ti,ab. OR "chronic kidney".ti,ab. OR "ckd".ti,ab.

AND

exp CARDIOVASCULAR DISEASES/ OR "cardi*".ti,ab. OR "heart*".ti,ab. OR "coronary*".ti,ab. OR "myocard*".ti,ab. OR "ischem*".ti,ab. OR "ischaem*".ti,ab. OR "stroke*".ti,ab. OR "cerebrovasc*".ti,ab. OR "cerebral vascular*".ti,ab.

AND

"cohort*".ti,ab. OR exp Cohort Studies/ OR "random*".ti,ab. OR "rct*".ti,ab. OR exp Randomized Controlled Trials as Topic/

AND

"risk*".tw. OR "prognosis*".tw. OR "predict*".tw. OR "associat*".tw. OR "factor*".tw.

limit to yr="1999 -Current"

limit to (english language and humans and "all adult (19 plus years)")

limit to (clinical study or clinical trial, all or clinical trial or meta analysis or observational study or pragmatic clinical trial or randomized controlled trial or systematic reviews or validation studies)

5.3 Risk of Bias Assessment for 1st Systematic Review

Study	Study participation	Study exclusion/attrition	Outcome measurement	Missing data	Statistical Analyses and Reporting
AASK-RCT	African Americans with hypertensive nephrosclerosis and a GFR of 20–65 ml/min per 1.73 m² recruited from February 1995 to September 1998.	2802 screened, 1094 enrolled, 186 excluded for this analysis - due to incomplete BP measurements (168), deaths within first 12 months (15), 'censored' (3). If ESRD reached then patient also censored (153).	Cardiovascular death, cardiovascular revascularization, nonfatal myocardial infarction (MI), hospitalization for heart failure, or stroke.	168 excluded due to incomplete BP data. Amlodipine arm terminated early September 2000. 7 lost to follow-up. 20 deaths after ESRD reached excluded.	Trial was a 3x2 factorial design with BP target and 1 of 3 possible anti-hypertensives. Cox proportional hazards regression for analysis included in current systematic review.
Ankara	Patients referred to outpatient renal unit between March 2006 to February 2011.	1276 eligible, 873 excluded due to 'conditions that may influence endothelial dysfunction', 65 excluded due to acute infection or unwillingness to participate.	Fatal and non-fatal stroke and myocardial infarction. Recorded by telephone contacts and routine outpatient clinic visits.	80 lost to follow-up, 73 withdrew consent. No other mention of missing data	Cox proportional hazards models. Does not report full multivariable model confounding factors.
CanPREDDICT	Multi-centre prospective cohort study. Recruitment dates not specified. Patients with CKD. eGFR between 15 to 45 ml/min/1.73m².	2529 included in final analysis. Number screened not given in current paper. Exclusion criteria - organ transplant recipient, life expectancy less than 12 months, acute vasculitis	Ischemic cardiovascular events 'were adjudicated and defined as myocardial infarction, unstable angina, ischemic stroke, coronary revascularization, new onset of coronary heart disease (proven by cardiac catheterization), amputation due to peripheral vascular disease, peripheral artery bypass, and gangrene.' Unclear if assessors 'a cardiologist, a nephrologist, and a neurologist' were blinded to patient characteristics for this particular study.	15 individuals excluded as TMAO results not available	Multivariable cox proportional hazards regression
CARE FOR HOMe	Single outpatient renal recruitment from September 2008 to November 2012. Patients with CKD stages 2-4, eGFR 15-89 ml/min per 1.73m2. Patients with CKD stage 2 had 'one or more markers of kidney damage, including albuminuria and/or plasma creatinine/cystatin C above references values'.	Systemic immunosuppressive medication and those with concomitant human immunodeficiency virus infection, clinical apparent infections (defined as CRP levels above 50 mg/l, and/or requiring systemic antibiotic therapy), active cancer disease, malignant hematological disorders, and/or acute renal failure (defined as increase of plasma creatinine >50% within four weeks) were excluded from study participation. Moreover, we excluded allograft recipients, pregnant women, and subjects <18 years of age'	Atherosclerotic events/death, which comprises acute myocardial infarction (defined as a rise in troponin T above the 99th percentile of the reference limit accompanied by symptoms of ischemia and/or electrocardiographic changes indicating new ischemia) , surgical or interventional coronary/cerebrovascular/peripheral arterial revascularization, stroke (defined as rapidly developing clinical symptoms or signs of focal [or at times global] disturbance of cerebral function lasting .24 hours [unless interrupted by surgery] or leading to death, with no apparent cause other than of vascular origin), amputation above the ankle'.	None lost to follow-up.	Multivariable Cox regression.
CREATE	605 enrolled, 476 completed study (roughly equal withdrawal between groups). 291 included in this substudy. Multi-centre RCT of individuals with Cockcroft–Gault GFR of 15–35 mL/min , Hb 11.0-12.5 g/dL. Systolic/diastolic blood pressure <170/95 mmHg. Recruited between July 2000 and November 2002. Clinical trial followed-up until November 2004.	Anticipated need for renal replacement therapy within 6 months, advanced cardiovascular disease (as defined by a diagnosis of clinically significant valvular disease, congestive heart failure, myocardial infarction, unstable angina, or stroke within the preceding 3 months), non-renal causes of anemia, receipt of blood transfusions within the preceding 3 months, a serum ferritin level of less than 50 ng per milliliter, a C-reactive protein level exceeding 15 mg per liter, and previous treatment with erythropoietin.'	Sudden death, myocardial infarction, acute heart failure, stroke, transient ischemic attack, angina pectoris resulting in hospitalization for 24 hours or more or prolongation of hospitalization, complication of peripheral vascular disease (amputation or necrosis), or cardiac arrhythmia resulting in hospitalization for 24 hours or more.'	Not stated.	Cox regression models. States 'pre-defined' sub-analysis but unclear how.
CRIC	'3,939 participants with CKD stages 2-4 who enrolled in the Chronic Renal Insufficiency Cohort (CRIC) between June 2003 and December 2008'. Age 21 to 74 years. 3,904 individuals included in current analysis (see 'missing data' column). 7 centres from USA. Follow-up until June 2009, death or	'inability to consent, institutionalization, enrolment in other studies, pregnancy, New York Heart Association classes III-IV heart failure, human immunodeficiency virus (HIV) infection, cirrhosis, myeloma, polycystic kidney disease, renal cancer, recent chemotherapy or immunosuppressive therapy, organ transplantation, or prior treatment with dialysis for at least 1 month.	'Definite or probable myocardial infarction, stroke, or peripheral arterial disease'. Outcome adjudicated by 'blinded reviewers'.	'Thirty-five participants had missing serum bicarbonate levels at baseline and were excluded from this study'. 'Approximately 9.4% of participants had missing covariable information and were excluded'. Individuals with missing data were 'not significantly	'Multivariable Cox proportional hazards models were used'. 'Death was treated as a censoring event when it was not part of the outcome'. Quadratic splines used to explore nonlinearity, but for cardiovascular events a linear relationship was found. Proportional hazard model assumption tested through use of Martingale residuals. 'interactions by race, diabetes, eGFR, and proteinuria

	voluntary withdrawal from study.			different at baseline'. No specific information on voluntary withdrawal from study in current study.	for death and renal outcomes and by diuretic use for cardiovascular outcomes'. Sensitivity analysis performed. Database 'locked' for analysis at 30th June 2009.
CRISIS	Prospective cohort study. Patients recruited between 1st October 2002 and 31st October 2009. Patients with CKD stages 3-5. eGFR >10 ml/min/1.73m²and <60 ml/min/1.73m².	1316 patients enrolled in CRISIS cohort. 470 patients included in analysis with OPG, FGF-23 and Fetuin-A measurements. 463 patients included in final analysis due to 7 with incomplete baseline data.	'Non-fatal stroke or myocardial infarction, coronary angiogram plus angioplasty or stenting and coronary artery bypass graft surgery. Heart failure defined as left ventricular ejection fraction ≤50% or diastolic dysfunction on echocardiogram or a clinical diagnosis of heart failure with no other alternative cause for symptoms.'	7 patients with incomplete baseline data but not included in final analysis. No other mention of incomplete data or loss to follow up.	Multivariable cox proportional hazards regression.
Digitalis	Multi-centre RCT of digoxin in chronic systolic and diastolic HF in normal sinus rhythm in 302 centres from 1991 to 1993. eGFR < 60 ml/min per 1.73m2. 7788 patients screened, 2793 in CKD in analysis for this study.	Serum creatinine >2.5 mg/dL, potassium ≥5 mEq/L,	Cardiovascular, and HF hospitalizations.'	Vital status data were complete for 99% of patients during 57 months of follow-up.' 931 had no baseline potassium, 579 excluded due to potassium K ≥5 mEq/L, unclear how many had serum creatinine >2.5 mg/dL	Propensity score matching. Matched Cox regression analysis.
Fujita	404 patients in a 'prospective cohort' of individuals with eGFR lower than 60 ml/min/1.73 m² recruited from February 2009 to September 2010.	Individuals with a 'history of hospitalization for acute coronary syndrome, worsening heart failure (New York Heart Association functional class III or IV), stroke, aortic dissection, or aortic aneurysm within 6 months before enrolment; fever or a sign of infectious diseases.'	CV death, acute coronary syndrome, hospitalization for worsening heart failure, stroke, and dissection of aorta. The diagnosis of acute coronary syndrome was based on the current guidelines. Heart failure was diagnosed based on the Framingham diagnostic criteria. Stroke was defined as clinical signs of focal or global disturbance of cerebral function caused by cerebrovascular damage. Aortic dissection was diagnosed based on computed tomography. The presence of CV events was determined independently by physicians who were blinded.'	No further details on missing data.	Cox proportional hazards regression analysis. All variables with P<0.05 by univariable analysis were included in the adjusted model. Not a pre-specified analysis.
Genoa	Multi-centre (2 centres) prospective cohort study. Patients recruited between 1st January 1999 and 31st December 2003. Patients included with non dialysis CKD < 60 ml/min/1.73m,² or GFR 60-90ml/min/1.73m² with proteinuria > 0.3 g/24hr twice with and interval ≥ 3 months.	693 patients screened. 445 patients included in the final analysis. Patients excluded due to follow up <3 months, no CKD, refusal of ABPM or echocardiography, wall motion abnormalities, inadequate ABPM measurement, office and 24 ABP < 130/80mmHg without antihypertensive treatment, change of antihypertensive treatment prior to study, severe valvular or ischemic heart disease, poor quality echocardiography, atrial fibrillation or bundle branch block, acute GFR change.	Cardiovascular events defined as myocardial infarction, congestive heart failure, stroke, revascularization, peripheral vascular disease, and nontraumatic amputation. Unclear from paper if adjudication team blinded	74 patients of those screened lost to follow up. 67 of those screened with inadequate or absent ABPM or echocardiography data.	Multivariable cox proportionalhazards regression
ICKD	2 centres prospective cohort recruited from a 'medical centre' and a 'regional hospital'. KDIGO CKD stages 1–5. Recruited between 11/11/2002 and 31/5/2009.	Acute kidney injury 'defined as more than a 50% decrease in eGFR in 3 months'. 356 stages 1-2 CKD patients were excluded.	Cardiovascular events defined as acute coronary syndrome, acute cerebrovascular disease, congestive heart failure, and peripheral arterial occlusion disease and death by aforementioned causes.'	90 patients lost to follow up. Possible underestimation of cardiovascular events due to patients receiving renal replacement therapy outside of their hospital and subsequently developing cardiovascular events.	Cox proportional hazard model
Kaohsiung	3749 consecutively enrolled outpatient pre-dialysis patients with CKD 3-5 recruited from 11/11/2002 to 31/5/2009. Follow-up until	3303 in final analysis. 446 excluded (356 patients CKD stages 1-2, 90 <3 months follow-up). 'significant mitral valve disease, atrial fibrillation, or inadequate image visualization' also exclusion criteria.	Hospitalization for acute coronary syndrome (Deyo's modified Charlson score, ICD-9 - 410.x–412.x), acute cerebrovascular disease (ICD-9 - 430.x–438.x), congestive heart failure (ICD-9 - 428.x), and peripheral arterial occlusion disease (ICD-9	90 patients lost to follow-up within 3 months. 193 patients included had more than 20% missing data.	Cox proportional hazard model. Significant variables (<0.05) in univariable analysis were selected for multivariable forward analysis. Not pre-specified analysis

	31/5/2010. 'Most' referred from local medical units or other specialists in the two recruiting hospitals for 'their impairment or progression of renal function'.		443.9, 441.x, 785.4, V43.4, procedure 38.48) and death by aforementioned cause'. 'Ascertained by reviewing charts'		
Kyushu	Prospective observational study of 320 consecutive admitted to single centre for 'evaluation of and education about CKD' from January 2005 to September 2012. Japanese patients with CKD stages 3-5, eGFR < 60 ml/min per 1.73m^2. No discussion of how many screened for study.	Malignancy or history of PAD.	PCI/CABG for IHD, CHF, CVD (brain infarction and haemorrhage), carotid endartectomy, percutaneous transcatheter angioplasty, lower limb amputation, bypass for PAD, dissecting aneurysm for thoracic and or abdominal aorta, rupture of aneurysm, PE, sudden death'	No further details on missing data.	Cox proportional hazard model. Significant variable in univariable analysis were selected for multivariable forward analysis. Unclear if pre-specified analysis
Leuven	Prevalent CKD patients seen at a single centre nephrology outpatient clinic. Recruited between November 2005 and September 2006. Followed up until 31/12/2010. CKD stages 1 to 5.	548 individuals eligible, 49 excluded as not eligible for study. 476 out of 499 analysed in this study. 13 FSGS patients and 10 patients without suPAR measurement excluded.	'Composite of death from cardiac causes, nonlethal MI, myocardial ischemia, coronary intervention, ischemic stroke, or new-onset peripheral vascular disease'. 'Cases of unobserved sudden death were considered cardiovascular death only when other potential causes could be excluded'. ''End points were prospectively recorded and coded, blinded from clinical and biochemical data'.	'If information [regarding outcome] could not be obtained, the patient was assumed to be lost to follow-up starting from the date of the last actual visit'. Paper does not stated how many this occurred to.	Cox proportional hazard model
Madrid	135 patients from renal clinic between January and May 2007 screened. eGFR lower than 60 ml/min; stable clinical condition in terms of no hospitalizations nor cardiovascular events within the 3 months before screening; and 'stable renal function' (baseline serum creatinine had not increased by 50% in the 3 months before screening). 113 included in study.	History of allopurinol intolerance or already on allopurinol treatment, active infections or inflammatory diseases, HIV infection, chronic hepatopathy, patients receiving immunosuppression. 22 excluded as eGFR >=60. 98 completed study - 2 started HD, 9 lost to follow up, 2 deaths, 2 due to 'minor adverse events'	Myocardial infarction, coronary revascularization, or angina pectoris. Congestive heart failure (CHF) was diagnosed by x-ray examination (pulmonary edema) and echocardiogram with left ventricular dysfunction. This diagnostic was considered as the patients were symptomatic and in New York Heart Association (NYHA) class II to IV with a left ventricular ejection fraction <45%. Cerebrovascular disease was established if the patient had a history of transient ischemic attacks, whenever stroke could be verified by computer tomography or carotid artery stenosis 70% could be verified by doppler ultrasound. Peripheral vascular disease was diagnosed by intermittent claudication, stenosis of the major arteries of the lower limbs angiographically or sonographically proven, and the presence of ulcers caused for atheroesclerotic disease or by surgery was used for diagnosis. Each event was reviewed by physicians. This information always included study hospitalization records and in the case of an out-of hospital death, family members were interviewed by telephone to better ascertain the circumstances surrounding death. Clinicians blinded during assessment.'	9 lost to follow-up	Cox proportional hazard models.
MAURO	755 patient with CKD stages 2-5 from 22 southern Italy nephrology units recruited from October 2005 to September 2008. Aged 18-75 and in 'stable clinical condition'.	Rapidly evolving renal disease, kidney transplant, acute intercurrent infections or acute inflammatory processes, pregnancy, cancer, or diseases in the terminal phase.' Creatinine >1.5-4.0 mg/dL in men and >1.3-3.5 mg/dL in women. Pregnancy.	Myocardial infarction, documented by electrocardiography and biomarkers of myocardial injury; heart failure, defined as dyspnea in addition to two of the following conditions: raised jugular pressure, bi-basilar crackles, pulmonary venous hypertension, or interstitial edema on chest radiography requiring hospitalization; electrocardiography documented arrhythmia; stroke; peripheral vascular disease; and major arterial or venous thrombotic episodes'	No mention in current paper.	Cox regression model.
Naples	Multicenter prospective cohort study of consecutive patients attending 4 outpatient nephrology clinics in Italy between 1/1/2003 and 31/12/2005'. Follow-up until 31/12/12.	530 screened, 472 eligible, 436 included. Excluded if missed more than 20% of treatment. From another publication from the same cohort 'We included 459 of 530 eligible patients, reasons for exclusion were inadequate ABPM recordings (n = 35), change of antihypertensive therapy 2 weeks before the study (n = 24), atrial	Cardiovascular death or nonfatal cardiovascular event that required hospital stay (myocardial infarction, congestive heart failure, stroke, revascularization, peripheral vascular disease, and non-traumatic amputation), whichever occurred first. Hospital records were obtained to establish diagnosis'	23 patients excluded as lost to follow-up after initial visit.	Cox proportional hazards model, stratified by center.

277

		fibrillation (n = 8), and a GFR change of more than 30% (n = 4).			
OSERCE-2	Observational prospective study of 39 nephrology centers of CKD stages 3–5 enrolled. Consecutive recruitment by including the first 20 adult patients. 742 baseline patients, only those with radiological data included in analysis (77%). Study from April 2009 May 2012.	Acute renal failure, serious illness that presupposed a life expectancy of <12 months, and hospital admission during the month before inclusion.	Cardiovascular hospitalization'	20 unexplained deaths.	Cox proportional hazards model.
Pravastatin	3 RCTs (WOSCOPS, CARE & LIPID). All double blinded 40mg pravastatin versus placebo for approximately 5 years. CKD defined as eGFR <60 or 60-89.9 with at least proteinuria. Initially, 19,737 patients, 4099 (20.8%) had CKD but not DM, 571 (2.9%) had CKD and DM.	Excluded based on serum creatinine greater than certain level (WOSCOPS >1.7mg/dL, CARE >2.5mg/dL, LIPID >4.5mg/dL).	Coronary heart disease death, nonfatal MI, or coronary revascularization (coronary artery bypass grafting or percutaneous transluminal coronary angioplasty). Secondary outcome - the time to an expanded composite cardiovascular outcome (first occurrence of coronary heart disease death, nonfatal MI, coronary revascularization, or nonfatal stroke)', blinded observers.	Not reported in current combined analysis manuscript	Proportional hazards regression models.
RRI	305 subjects included for analysis with 24 hour Holter data. From the main RRI-CKD trial 627 were enrolled between 1/1/2000 and 31/12/2002 of which 408 were alive at 1/1/2003. For this sub-analysis, 149 recruited who were alive at end of main trial. An additional 199 individuals were recruited for this subanalysis. 43 declined to have Holter monitor. Follow-up until 31/12/2006.	Transient renal impairment prior to enrolment (eGFR on two occasions at least 1 month apart). Those recruited were 'significantly healthier than those who did not' (younger, higher mean eGFR, less DM, less CVD history). Nitro-glycerine patch, defibrillator, active pacing or with allergy to electrode adhesive material) were excluded.	CAD, cerebrovascular disease, peripheral arterial disease, CHF and cardiac arrest. All outcomes were ascertained on an ongoing basis by study coordinators from regular review of electronic health records, direct patient contact in clinic and periodic telephone communication.	Not reported.	Cox regression.
TREAT	4038 patients with T2DM, CKD and anaemia, previously enrolled into the TREAT randomised controlled trial. 623 sites in 24 countries, from Aug 2004 to Dec 2007.	Exclusions for 'uncontrolled HTN, previous renal transplant or scheduled for live donor transplant, current IV Abx/chemotherapy/radiotherapy, cancer, HIV, active haematological disease / bleeding, pregnancy, recent MACE/seizure/surgery (last 12 weeks)'. 4047 patients enrolled, 9 excluded (not for clinical reasons).	'MI, stroke, ESRD, and the composite of cardiovascular death, MI or hospitalization due to ischemia, heart failure or stroke.' Definitions reported in main trial manuscript. End points adjudicated by a blinded clinical end point committee.	Outcomes reported for all 4038 patients. No loss to follow-up reported in original study.	Cox proportional hazards models.

5.4 Potential Risk Factors for 1st Systematic Review

Variable	Study	Details	Included in Systematic Review	Rationale	Reference if not listed in main eferences
Adragao score	OSERCE-2	based on presence of calcification on plain radiographic films of pelvis and hands	No	Not routinely collected	-
Age	See forest plot	at baseline	Yes	-	See forest plot
Albumin	See forest plot	serum	Yes	-	See forest plot
Ankle–brachial pressure index	Kyushu, OSERCE-2	-	No	Not routinely collected	-
Baseline NT-proBNP	CREATE	serum	No	Not routinely collected	-
Beck Depression Index II	AASK - cohort	multiple-choice self-reporting questionnaire for depression	No	Not routinely collected	Fischer M et al, Kidney International; (2011) 80:670-678
Beta-2 microglobulin	CRIC	serum	No	Not routinely collected	Foster MC et al, Am J Kidney Dis. 68(1):68-76.
Beta-trace protein	CRIC	serum	No	Not routinely collected	Foster MC et al, Am J Kidney Dis. 68(1):68-76.
Bicarbonate	CRIC	serum	Yes	-	Dobre M et al, American Journal of Kidney Disease; 2013; 62(4):670-678
Body Mass Index	See forest plot	kg/m^2	Yes	-	See forest plot
Calcium	See forest plot	serum	Yes	-	See forest plot
Cardiovascular disease	See forest plot	diagnosis at baseline	Yes	-	-
Cholesterol to HDL ratio	ICKD	serum	Yes	-	-
Clinic versus ambulatory blood pressure	Minutolo *et al*,, Kushiro *et al*	comparison of clinic versus ambulatory blood pressure to categorise patients	No	Ambulatory blood pressure not routinely collected	Minutolo R et al, American Journal of Kidney Disease; 2014; 64(5):744-752 Kushiro T et al, Hypertension Research (2017) 40, 87–95
Congestive heart failure	CREATE	diagnosis at baseline	Yes	-	-
C-reactive protein	Kyushu, Madrid, Haryana, Leuven	serum	No	Not routinely collected for purpose of CV risk assessment	Haryana - Nand N et al, Journal, Indian Academy of Clinical Medicine; 2009; 10(1 & 2): 18-22
Diabetes mellitus	See forest plot	diabetes mellitus at baseline	Yes	-	See forest plot
Diabetic nephropathy	CREATE	diabetic nephropathy at baseline	No	Not routinely collected	-
Diastolic blood pressure	See forest plot	clinic based	Yes	-	See forest plot
Degree (high school)	AASK - RCT	education	No	Not routinely collected	-
Dyslipidaemia	Kyushu	multiple categories	No	Multiple categorical variables	-
Electrocardiogram	AASK - RCT	normal' versus 'abnormal'	No	Definition unclear from reference	-
Fetuin-A	CRISIS	serum	No	Not routinely collected	-
Fibroblast growth factor-23	CRISIS	serum	No	Not routinely collected	-
Gender	See forest plot	-	Yes	-	See forest plot
Haemoglobin	See forest plot	serum	Yes	-	See forest plot
High density lipoprotein	ICKD	serum	Yes	-	-
High sensitivity C-reactive protein	Fujita, Ankara	serum	No	Not routinely collected	-
Homocysteine	HOPE-2	serum	No	Not routinely collected	Mann J et al, Nephrology Dialysis Transplantation; 2008; 23: 645–653
Hydration status	Tsai *et al*	bioelectric impedence spectroscopy measured at baseline	No	Not routinely collected	Tsai Y-C et al, Clinical Journal of the American Society of Nephrology; 2015; 10: 39–46
IL-33	Ankara	serum	No	Not routinely collected	Gungor O et al, PLoS ONE 12 (6): e0178939
Income	AASK - RCT	self-reported monetary income at baseline	No	Not routinely collected	-
Ischemic Heart Disease	See forest plot	diagnosis at baseline	Yes	-	See forest plot
Kauppila score	OSERCE-2	based on presence of calcification on plain radiographic films of lumbar spine	No	Not routinely collected	-

Variable	Study	Details	Included in Systematic Review	Rationale	Reference
Left ventricular end-diastolic volume	Kaohsiung	echocardiogram	No	Not routinely collected	-
Left ventricular hypertrophy	AASK - RCT	echocardiogram	Yes	-	-
Left ventricular mass index	CREATE	echocardiogram	No	Not routinely collected	-
Low density lipoprotein	See forest plot	serum	Yes	-	See forest plot
Mean arterial pressure	AASK - RCT	clinic based	Yes	Not routinely collected	-
Metabolic Syndrome	MESA	as a whole diagnosis (not the presence of individual parts such as HTN, DM etc)	No	Not routinely collected, categorical variables	Agarwal S et al, Cardiology Research and Practice; 2012; Article ID 806102
Non-HDL cholesterol	ICKD, AASK - RCT	serum	Yes	-	-
Osteoprotegerin	CRISIS	serum	No	Not routinely collected	-
Parathyroid Hormone	CARE FOR HOMe	serum	Yes	-	-
Peripheral Vascular Disease	CREATE	diagnosis at baseline	No	Definition unclear from reference	-
Phosphate	See forest plot	serum	Yes	-	See forest plot
Potassium	Digitalis	serum	No	Unable to compare due to multiple serum categories	-
Prolactin	Heraklion	serum	No	Not routinely collected	Carrrero J et al; Clinical Journal of the American Society of Nephrology; 2012; 7: 207–215
Pseudoresistant hypertension	Naples	normal ambulatory blood pressure but raised clinic blood pressure	No	Not routinely collected	-
Pulmonary hypertension	CRIC	echocardiogram	Yes	-	-
Pulse pressure	Kyushu	clinical based	Yes	-	See forest plot
Root mean square of the successive differences of heart rate	CRIC	-	No	Not routinely collected	Drawz P et al; American Journal of Nephrology;2013;38(6): 517–528
Sodium	CanPREDDICT	serum	Yes	-	-
Sodium	CRIC	urine (24 hour collection)	No	Not routinely collected	Mills KT et al, JAMA. 2016;315(20):2200-2210
Sodium to potassium ratio	AASK - RCT	urine	No	Not routinely collected	-
Systolic blood pressure	See forest plot	clinic based	Yes	-	See forest plot
Systolic blood pressure visit-to-visit variability	AASK - RCT	-	No	Not routinely collected	-
Standard deviation of NN intervals for heart rate variability	CRIC	-	No	Not routinely collected	Drawz P et al; American Journal of Nephrology;2013;38(6): 517–528
Trimethylamine N-oxide	CanPREDDICT	serum	No	Not routinely collected	-
Urea	AASK - RCT	serum	Yes	-	-
Smoking	See forest plot	smoker at baseline versus non-smoker	Yes	-	See forest plot
ST2	Ankara	serum	No	Not routinely collected	Gungor O et al, PLoS ONE 12 (6): e0178939
Sustained hypertension	Naples	raised ambulatory blood pressure but normal clinic blood pressure	No	Not routinely collected	-
Total cholesterol	See forest plot	serum	Yes	-	See forest plot
True resistant hypertension	Naples	raised ambulatory and clinic blood pressure	No	Not routinely collected	-
Urate	See forest plot	serum	Yes	-	See forest plot
Years with hypertension	AASK - RCT	years of diagnosis at baseline	No	Not routinely collected	-

6.1 Search Strategy for 2nd Systematic Review

Search run on 27th April 2017
Database: Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) <1946 to Present>

1   chronic kidney disease.tw
2   chronic kidney disorder.tw
3   kidney failure, chronic/ or chronic kidney failure.tw
4   chronic kidney insufficiency.tw
5   chronic nephropathy.tw
6   chronic renal disease.tw
7   chronic renal failure.tw
8   renal insufficiency, chronic/ or chronic renal insufficien$.tw
9   1 or 2 or 3 or 4 or 5 or 6 or 7 or 8
10   predict*.tw
11   validat*.tw
12   develop*.tw
13   Predictive value of tests/
14   scor*.tw
15   observ*.tw
16   Observer variation/
17   ROC curve/
18   discriminat*.tw
19   c-statistic.tw
20   c statistic.tw
21   area under the curve.tw
22   area under curve.tw
23   auc.tw
24   calibration.tw
25   Algorithm/
26   10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25
27   Mortality/
28   (mortal$ or dead or death).tw
29   cardiovascular disease.tw. or Cardiovascular Diseases/
30   exp Cardiovascular Diseases/
31   cardi*.tw
32   heart.tw
33   myocard*.tw
34   ischem*.tw
35   ischaem*.tw
36   stroke.tw

37   cerebrovasc*.tw

38   cerebral vascular*.tw

39   endstage renal.tw

40   endstage kidney.tw

41   esrf.tw

42   esrd.tw

43   transplant*.tw

44   peritoneal dialysis.tw

45   hemodialysis.tw

46   haemodialysis.tw

47   27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46

48   9 and 26 and 47

49   limit 48 to (english language and humans and yr="2012 -Current")

50   limit 49 to (clinical study or clinical trial, all or clinical trial or comparative study or evaluation studies or observational study or validation studies)