
# Can Routinely Collected Primary Care Data Determine the Prevalence of Chronic Kidney Disease And Predict Clinical Events In Patients With Stages 3 of the Disease?

Dr Poorva Jain MBChB MRCP

# ABSTRACT

## Introduction

Chronic Kidney Disease (CKD) is common and is associated with cardiovascular morbidity and mortality. Few previous studies have assessed the 'true' prevalence of CKD in the general population and the predictors of morbidity and mortality in patients with stages 3-5 CKD

## Methods

Using data from the Health Improvement Network, a large United Kingdom general practice database, the prevalence of stages 1-5 CKD was ascertained using both single and duplicate blood tests between 2005 and 2009. The prevalence of stages 3-5 ascertained from two blood results was compared to the CKD prevalence on the practice register determined by Quality Outcome Frameworks Read codes in 2009. The management of patients on the register, and those with CKD not on the register was compared. Cox proportional hazard models using routinely collected primary care data as co-variables were used identify potential predictors of i) all-cause mortality and ii) the composite of cardiovascular disease and all-cause mortality. Multiple imputations were used to allow for uncertainty in missing values such as blood pressure and body mass index in the general practice data sets.

## Results

The prevalence of 'true' stages 1-5 CKD, i.e. where the chronicity of CKD is taken into account, was 5.01% and much less than estimates from previous studies. Using two laboratory eGFRs reported at least seven days apart, the prevalence of stages 3-5 CKD was 4.7% in 2009. Over two percent of the population who had stages 3-5 CKD were not on their primary care practice register and 2.9% of the population were apparently misdiagnosed

with CKD. Patients with undiagnosed CKD tended to be younger and have less co-morbidity. Patients with CKD not on the practice register were associated with worse management than those on the practice register.

The Cox proportional hazards models demonstrated that increasing age and co-morbidity were associated with worse outcomes, however continuous variables such as blood pressure, body mass index, haemoglobin and cholesterol were associated with an inverse J shaped relationship with log relative hazard ratio. Treatment with angiotensin blockade, beta-blockade, lipid lowering agents, and other antihypertensives was associated with improved outcomes, however blood thinning agents and diuretics were associated with worse outcomes. African Caribbean and Indian Sub-Continental ethnicity, though largely unreported, was also associated with better outcomes.

## Conclusions

Stages 3-5 chronic kidney disease are common in UK primary care but not as common as previously reported. Many practices under record or mis-label CKD which impacts upon management. Redirection of resources to identify patients better could ensure better outcomes with a similar workload as more people would be removed than join disease registers. In these analyses some predictors of mortality and the composite of all cause mortality and cardiovascular disease are inconsistent with previous reports and this requires further investigation especially when considering blood pressure management.

## ACKNOWLEDGEMENTS

My extreme gratitude to Professor Richard McManus, Professor Melanie Calvert and Professor Paul Cockwell for helping to conceive this project, acquiring funding from the National Institute of Health Research and most of all for fantastic supervision and encouragement.

I would like to thank Professor Nicholas Freemantle, Dr Matthew Richardson, Dr Richard Riley and Ms Kym Snell for invaluable statistical advice. My additional thanks to CSD Medical Research UK for providing the THIN dataset.

Finally I have to thank my parents who supported and encouraged me throughout this thesis. My father as a GP was an excellent sounding board and my mother educated me about points about QOF.

Last but not least I would like to thank my sister without whom this thesis would never have come to completion who didn't sleep until my thesis was completed.

## RESPONSIBILITIES

All chapters in this thesis were written by P Jain and reviewed by M Calvert, RJ McManus and P Cockwell. RJ McManus was responsible for thesis concept and succeeding in receiving funding as part of his National Institute of Health research fellowship grant. The thesis research protocol was developed by P Jain, M Calvert, RJ McManus and P Cockwell.

Chapters 2 contributed to a poster presentation (American Society of Nephrology 2011. Chapters 2-3 contributed to a clinical paper published in PLoS one where P Jain was responsible for the clinical analysis and preparation of the paper. RJ McManus and P Cockwell involved in the clinical interpretation of the paper and M Calvert in the research methodology. All authors contributed and approved the final version of the paper.

## TABLE OF CONTENTS


Abstract ..... 2

Introduction ..... 2
Methods ..... 2
Results ..... 2
Conclusions ..... 3

Acknowledgements ..... 4

Responsibilities ..... 5

List of Tables ..... 10

List of Figures ..... 13

Abbreviations ..... 15

Chapter 1. INTRODUCTION TO THE THESIS ..... 17

1.1. What Is Chronic Kidney Disease? ..... 17
1.2. Markers of Renal Damage: How is Kidney Function Measured? ..... 19
1.2.1. Glomerular filtration rates and surrogate markers ..... 19
1.2.2. Measuring GFR ..... 20
1.2.3. Creatinine based estimation ..... 20
1.2.4. Creatinine measurement ..... 21
1.2.5. MDRD equation ..... 22
1.2.6. The Mayo Quadratic formula ..... 25
1.2.7. The Chronic Kidney Disease Epidemiology Collaboration (CKD EPI) formula ..... 25
1.2.8. The impact of creatinine measurement on eGFR ..... 26
1.2.9. What is proteinuria and how is it measured? ..... 27
1.3. The Current Classification of CKD: Evidence Base ..... 31
1.4. The Prevalence of Stages 1-5 CKD ..... 34
1.5. CKD: What Are The Mortality And Morbidity Implications? ..... 35
1.5.1. Cardiovascular disease and mortality ..... 35
1.5.2. Renal Outcomes ..... 35
1.5.3. Economic Impact of CKD ..... 38
1.6. Risk Factors for Cardiovascular Disease and Mortality ..... 39
1.7. Why is Primary Care the Ideal Forum for CKD Research? ..... 41
1.7.1. Quality Outcomes Framework ..... 43
1.8. Large Primary Care Databases and their Utility in Research ..... 44
1.8.1. The strength and limitations of THIN ..... 45
1.9. Thesis Aims and Objectives ..... 49
1.9.1. Thesis Aims ..... 49
1.9.2. Thesis Objectives and Organisation ..... 49
1.10. Executive Summary ..... 50

Chapter 2. The prevalence of CKD in primary care ..... 51

2.1. The Prevalence of CKD: A review of the existing literature ..... 52

2.1.1. The prevalence of CKD in the UK....56
2.1.2. The variation in prevalence....56
2.2. Research Questions: What is the prevalence of CKD in primary care?....63
2.3. Methods....65
2.3.1. Preparing the database: first dataset....65
2.3.2. Examining renal function....67
2.3.3. The prevalence of stages 3-5 Chronic Kidney Disease....68
2.3.4. Defining Proteinuria....71
2.3.5. Prevalence of CKD stages 1-2....72
2.4. Results....73
2.4.1. Stages 3-5 CKD: Identifying CKD in the cohort and Raw data distribution....74
2.4.2. Prevalence of stages 3-5 CKD using different equations and definitions....82
2.4.3. Prevalence of stages 1-2 CKD: ACR and Urine Dip Data....89
2.5. Discussion. ....96
2.5.1. Summary of key findings for stages 3-5 CKD....96
2.5.2. Interpretation of results and comparison to existing literature....97
2.5.3. Limitations and Strengths....102
2.5.4. Summary of Results for Stages 1-2 CKD....104
2.5.5. Interpretation of Results and Comparison to existing literature....104
2.5.6. Strengths and Limitations....105
2.5.7. Executive Summary....106
Chapter 3. Do primary care practices correctly diagnose stages 3-5 CKD compared to objective measures and how does this impact on patient care?....107
3.1. Introduction....108
3.1.1. The UK Quality Outcomes Framework....110
3.2. Research Questions....115
3.3. Methods....116
3.3.1. Study Population....116
3.3.2. Estimating prevalence of CKD 3-5 on primary care practice register between 2005 to 2009....116
3.3.3. Estimating the proportion of patients with CKD (stages 3-5) Read codes had biochemical evidence of CKD in the QOF analysis period of 2008-2009....116
3.3.4. Demographic and comorbid characteristics of these groups....117
3.3.5. Characteristics of CKD patients that were excluded from the QOF CKD register 118
3.3.6. What are the differences between Uncoded and Miscoded patients – a mixed model....119
3.3.7. Comparison of QOF indicators between patients with labelled and confirmed CKD ....119
3.4. Results....121
3.4.1. Who is on the Practice CKD register?....121
3.4.2. QOF indicators: Comparison of Labelled and Confirmed CKD....130
3.5. Discussion....135
3.5.1. Summary of Key Findings....135
3.5.2. Interpretation of results and comparison to existing literature....136
3.5.3. Strengths and limitations....138
3.6. Executive Summary....139

Chapter 4. What factors predict all-cause mortality in stage 3 CKD patients? A Prognostic Model ..... 140
4.1. Introduction ..... 141
4.1.1. Cardiovascular disease and mortality in CKD patients ..... 141
4.2. Rationale in modeling for death in patients with stage 3 CKD ..... 153
4.3. Methods ..... 154
4.3.1. Model development: A review of the literature ..... 154
4.3.2. Step by Step discussion of the models. ..... 162
4.4. Results ..... 171
4.4.1. Step 1: All cause mortality outcomes and description ..... 171
4.4.2. Step 2: Missing data ..... 174
4.4.3. Step 3: Multiple imputation ..... 174
4.4.4. Step 4. Results of analyses where the relationship between continuous co-variable and time to outcome was assumed linear. ..... 176
4.4.5. Step 5: Analysis of transformed variables. ..... 179
*4.4.6.* Step 6. Selection of the model with the most consistent transformations i.e. the alternative model ..... 185
4.4.7. Step 7. Frailty Models ..... 190
4.4.8. Step 8. Is model fit improved after incorporating mfp transformation and frailty terms? ..... 196
4.4.9. Step 9. Testing the proportional hazards assumption ..... 197
4.4.10. Step 10. Complete Case Analysis ..... 197
4.5. Discussion ..... 201
4.5.1. Summary Of Results ..... 201
4.5.2. Interpretation of Results ..... 205
4.5.3. Comparison to existing literature ..... 205
4.5.4. Limitations and Strengths ..... 213
4.6. Executive Summary ..... 218
Chapter 5. Can routinely collected primary care data predict cardiovascular disease or all cause mortality in stage 3 CKD patients? ..... 220
5.1. Methods ..... 221
5.1.1. Model Cohort ..... 221
5.1.2. Outcome Measures: ..... 221
5.2. Results ..... 224
5.2.1. Step 1. Outcomes experienced by the Cohort ..... 226
5.2.2. Step 2. Variables with missing data ..... 226
5.2.3. Step 3. Missing data analysis and subsequent imputation ..... 226
5.2.4. Step 4. Analysis of co-variables where the relationship between continuous co-variables and time to outcome is assumed to be linear. ..... 228
5.2.5. Step 5: Analysis of transformed variables. ..... 231
5.2.6. Step 6. The ('alternative') model with the most consistent transformation ..... 237
5.2.7. Step 7: Frailty Models ..... 242
5.2.9. Step 8. Is model fit improved after incorporating mfp transformation and frailty terms? ..... 248
5.2.10. Step 9: Testing the proportional hazards assumption ..... 248
5.2.11. Step 10. Complete Case Analysis ..... 249

5.3. Discussion ..... 253
5.3.1. Summary of Results ..... 253
5.3.2. Limitations and Strengths ..... 257
5.4. Executive Summary ..... 258
Chapter 6. Overall Discussion ..... 259
6.1. Summary of Key Findings ..... 259
6.2. Clinical implications of the results reported in this thesis ..... 262
6.2.1. Prevalence of CKD ..... 262
6.2.2. Recognition and management of patients with stages 3-5 CKD in primary care 264
6.2.3. Prognostic Models ..... 266
6.3. Strengths and Limitations ..... 268
6.4. Future Research ..... 268
6.5. Clinical Implications and overall conclusions ..... 271
Bibliography ..... 272
Appendix A. Details of Models for Chapter 4 ..... 292
Appendix B. Scaled Schoefeld Residuals for Chapter 4 ..... 319
Appendix C Details of Models for Chapter 5 ..... 336
Appendix D. Scaled Schoefeld Residuals for Chapter 5 ..... 362

## LIST OF TABLES

Table 1-1. Stages of Chronic Kidney defined by the National Kidney Foundation/ Kidney Outcomes Quality Initiative (N/KDOQI) guidelines .....17
Table 1-2. Quantification of proteinuria: comparison of different methods .....28
Table 1-3. Current NKF KDOQI CKD staging system and alternate system of CKD risk categories.....32
Table 1-4. Prevalence variation of Stages CKD 1-5 adapted from Chapter 2.1 [8;49-91] .....34
Table 1-5. Traditional and Non traditional risk factors for CVD and Mortality in CKD: a comparison with the general population. .....40
Table 1-6. Example of routinely collected database studies in the United Kingdom with emphasis on kidney disease and cardiovascular disease .....42
Table 1-7. Description of THIN data components. .....47
Table 2-1. Summary of CKD prevalence studies from North America .....54
Table 2-2. Summary of CKD prevalence studies from Europe.....58
Table 2-3. Comparison of CKD prevalence studies from Australia and Asia .....60
Table 2-4. Preparation of database to determine CKD prevalence.....66
Table 2-5. Prevalence of Stages 3-5 CKD using two blood results either lab eGFR or calculated eGFR using by Non IDMS MDRD (95% confidence intervals in brackets) .....84
Table 2-6. Prevalence of CKD stages 1-2 using either ACR or Protein dipstick .....94
Table 2-7. Summary of prevalence of Stages 3-5 CKD in 2009 from the thesis .....95
Table 2-8. Methods of creatinine analysis using UK National External Quality Assessment Service based laboratories[31] .....97
Table 3-1. Countries with disease registers and involvement in payment for performance 108
Table 3-2. QOF indicators for CKD in 2008/2009[140] ..... 109
Table 3-3. QOF indicators for diabetes mellitus specific to CKD patients in 2008/2009 ..... 110
Table 3-4. Demographics of CKD patients between 1st January 2008 to 1st April 2009 (The proportions in brackets are crude percentages) ..... 122
Table 3-5. Multivariate logistic regression model for significant predictors for exclusion from the CKD QOF register: Uncoded vs. appropriately coded ..... 124
Table 3-6. Multivariate logistic regression model for significant predictors for exclusion from the CKD QOF register: Uncoded vs. Miscoded ..... 126
Table 3-7. CKD Management according to QOF standards between 1st January 2008 to 1st April 2009 (Proportions quoted are of the total in that group)..... 129
Table 3-8. Management of Diabetic CKD patients ..... 131
Table 4-1. List of co-variables in the analysis for all cause mortality ..... 161
Table 4-2. Demographics and clinical features of the whole for all cause mortality cohort (n= 109 017) ..... 170
Table 4-3. Variables with imputed data through imputation for all cause mortality model 172
Table 4-4. All cause mortality: combined coefficients from Cox proportional hazards model and linear risk is assumed of numerical terms ..... 175
Table 4-5. AIC for each analysed imputed dataset for analysis in Table 4-4..... 176
Table 4-6. Suggested transformation of variables from mfp algorithm in R for each imputation for mortality model..... 176
Table 4-7. AIC for each analysed imputed dataset: comparing non linear and linear terms for mortality model ..... 181
Table 4-8. A comparison between the original model with mfp and alternative model (no frailty term)..... 182

Table 4-9. The combined coefficients using the 'alternative' models (No frailty term) ..... 184
Table 4-10. Combined coefficients for all cause mortality model where linear risk is assumed of numerical terms with practice location as frailty term ..... 188
Table 4-11. AIC for each model for all cause mortality model with and without frailty term ..... 189
Table 4-12. AIC for each model for all cause mortality with original suggested fractional polynomial with frailty term ..... 190
Table 4-13. A comparison between the original model with mfp and the best model with a frailty term ..... 191
Table 4-14. Combined all cause model for each imputation using the alternative model incorporating a frailty term of practice location ..... 193
Table 4-15. Comparison of the AIC for the original all cause mortality model without frailty/mfp and then models with mfp and frailty ..... 194
Table 4-16. Demographics of the whole all cause mortality cohort and cohort with complete data ..... 196
Table 4-17. Cox model with frailty term and fractional polynomials complete case analysis for all cause mortality model ..... 198
Table 5-1. Demographics of the composite model cohort (if not stated in brackets then proportion of data complete) ..... 223
Table 5-2. Imputed datasets for composite model: The distribution of imputed variables . 225
Table 5-3. Combined coefficients from Cox proportional hazards model for composite outcome where linear risk of numerical variables is assumed. (no frailty term) ..... 227
Table 5-4. AIC for each imputed dataset for Table 5-3 ..... 229
Table 5-5. Suggested transformation of variables from mfp algorithm in R for each imputation for composite model ..... 229
Table 5-6. AIC for each model with original suggested fractional polynomial for composite model (no frailty term) ..... 234
Table 5-7. A comparison between the original mode for composite outcome with mfp and the 1st and 2nd best model (no frailty term) ..... 236
Table 5-8. Combined model for each Imputation for first alternative model for all Imputations. (No frailty term) ..... 237
Table 5-9. Combined coefficients from composite outcome model where linear risk of numerical variables is assumed with practice location as frailty term ..... 241
Table 5-10. AIC for each model for Table 5-9 ..... 242
Table 5-11. AIC for each model with original suggested fractional polynomial with frailty term ..... 243
Table 5-12. A comparison between the ordinal model with mfp and the 1st and 2nd best model with frailty term for composite outcome ..... 244
Table 5-13. Combined model for composite outcome for each imputation using the 1st best model incorporating a frailty term of practice location ..... 245
Table 5-14. Comparison of the AIC for the original model without frailty/mfp and models with mfp and frailty. The value in brackets is the difference between the current column and the previous column ..... 246
Table 5-15. Demographics of the whole cohort and complete case cohort for composite outcome analysis ..... 247
Table 5-16. Variables with missing data and their mean, median value or proportion for the composite outcome analysis cohort ..... 248

Table 5-17. Cox model for composite outcome of complete case analysis with frailty term and fractional polynomials ..... 249

## LIST OF FIGURES

Figure 1-1. A summary of kidney dysfunction .....19
Figure 1-2. Relation of estimated GFR to measured GFR in the participants of the MDRD study.....23
Figure 1-3. Composite Ranking for Relative Risks by glomerular filtration rate (GFR) and Albuminuria .....33
Figure 1-4. Risk of mortality split by eGFR and proteinuria levels from CKD Prognosis Consortium.....37
Figure 1-5. Comparison of the structure of The Health Improvement Network (THIN) population on 1 July 2008 with the Office for National Statistics (ONS) mid-2008 UK population estimate.....47
Figure 2-1. Distribution of Serum creatinine .....76
Figure 2-2. Distribution of calculated eGFR.....76
Figure 2-3. Histogram distribution of laboratory reported MDRD eGFR .....77
Figure 2-4. Histogram distribution of eGFR calculated by MDRD equation.....77
Figure 2-5. Laboratory MDRD eGFR versus Calculated EGFR using non IDMS MDRD equation .....79
Figure 2-6. Laboratory MDRD eGFR versus Calculated EGFR using non IDMS MDRD equation where both below 60 ml/min/1.73m$^2$.....79
Figure 2-7. Bland Altman Plot for laboratory eGFR and MDRD calculated eGFR (Non IDMS) .80
Figure 2-8. Calculated MDRD (non IDMS) eGFR versus Mayo Quadratic and CKD EPI derived eGFRs in Men .....81
Figure 2-9. Calculated MDRD (non IDMS) eGFR versus Mayo Quadratic and CKD EPI derived eGFRs in Women.....82
Figure 2-10. How the cohorts were defined for the principal analysis. .....84
Figure 2-11. CKD prevalence by Age and Gender using two blood results either two lab eGFRs or two calculated eGFRs using the the Non IDMS MDRD equation .....86
Figure 2-12. The prevalence of CKD using two blood results according to equation or directly reported lab eGFRs. The principal analysis is in Orange and the secondary analysis in dashed lines .....88
Figure 2-13. The prevalence of stages 3-5 CKD using a single blood result according to equation or two lab eGFRs .....89
Figure 2-14. Schematic of urine ACR/Prot Data and Derivation of CKD 1-2 .....91
Figure 2-15. Histogram of Albumin Creatinine Ratio.....92
Figure 2-16. Box and Whisker plot of ACR versus urine dip .....93
Figure 2-17. ROC curve for urine dip to diagnose albuminuria .....94
Figure 3-1. Flowchart of how patients were defined with CKD for QOF analysis cohort..... 120
Figure 3-2. Proportion of patients with biochemical CKD versus classified CKD ..... 121
Figure 3-3. Forest plot of odds ratios: uncoded CKD versus appropriately coded CKD..... 126
Figure 3-4. Forest plot of odds ratios: uncoded CKD versus miscoded CKD ..... 128
Figure 4-1. Summary of all cause mortality model..... 168
Figure 4-2. Derivation of THIN CKD cohort for all cause mortality model..... 169
Figure 4-3. Age versus Log relative Hazard Ratio for each imputation for mortality model (without frailty term) ..... 178
Figure 4-4. Log relative Hazard Ratio versus Systolic BP for each imputation for mortality model (without frailty term)..... 179

Figure 4-5. Haemoglobin versus Log relative Hazard Ratio for each imputation (without frailty term) ..... 180
Figure 4-6. Body Mass Index versus log relative hazards ratio for Imps 1-5 (no frailty term) for mortality model ..... 181
Figure 4-7. Cholesterol versus log relative hazards ratio for Imps 1-5 (no frailty term) for mortality model ..... 181
Figure 4-8. Age versus log relative hazard ratio (no frailty term) in model using most consistent mfp for all cause mortality ..... 186
Figure 4-9. Most consistent multiple fractional polynomials for death model (without frailty): Systolic blood pressure, Hb, BMI and cholesterol versus log relative hazard ratio ..... 187
Figure 4-10. Iterations for Models predicting death ..... 202
Figure 4-11. Forest plot of hazard rations for predictors of death ..... 203
Figure 5-1. Summary of Cardiovascular disease/all cause mortality model. ..... 223
Figure 5-2. Derivation of THIN CKD cohort for composite outcome ..... 224
Figure 5-3. Age versus Log relative Hazard Ratio for each imputation (without frailty term) ..... 232
Figure 5-4. Log relative Hazard Ratio versus Systolic BP for each imputation (without frailty term) ..... 233
Figure 5-5. Haemoglobin versus Log relative Hazard Ratio for each imputation (without frailty term) ..... 234
Figure 5-6. Body Mass Index versus log relative hazards ratio for Imps 1 to 5 (no frailty term) ..... 235
Figure 5-7. Multiple fractional polynomials for cardiovascular disease/ death model without frailty. Systolic blood pressure, age BMI and Hb versus log relative hazard ratio ..... 241
Figure 5-8. Age versus log relative hazard ratio in complete case analysis ..... 252
Figure 5-9. Body mass Index versus log relative hazard ratio in complete case analysis ..... 252
Figure 5-10. Iterations for Models predicting cardiovascular disease and or Death ..... 254
Figure 5-11. Forest plot of hazard ratios for covariables predicting CVD/death ..... 255

# ABBREVIATIONS

| | |
|---|---|
| ACR | Albumin Creatinine Ratio |
| AIC | Akaike Information Criteria |
| APKD | Adult Polycystic Kidney Disease |
| BMI | Body Mass Index |
| BP | Blood Pressure |
| CCA | Complete Case Analysis |
| CHD | Coronary Heart Disease |
| CKD | Chronic Kidney Disease |
| CKD EPI | Chronic Kidney Disease Epidemiology Collaboration |
| CPHM | Cox Proportional Hazards Model |
| CPRD nee GPRD | Clinical Practice Research Datalink |
| CVA | Cerebrovascular Accident or Stroke |
| CVD | Cardiovascular Disease |
| DM | Diabetes Mellitus |
| eGFR | Estimated Glomerular Filtration Rate |
| EPR | Electronic Patient Record |
| ESRD | End Stage Renal Disease |
| FP | Fractional Polynomial |
| GFR | Glomerular Filtration Rate |
| GP | General Practitioner |
| GPRD nee CPRD | General Practice Research Database |
| Hb | Haemoglobin |
| HF | Heart Failure |
| HSE | Health Survey Of England |
| IDMS | Isotopic Dilution Mass Spectroscopy |
| Imp(s) | Imputation(s) |
| KDIGO | Kidney Disease Improving Global Outcomes Initiative |
| KDOQI | Kidney Disease Outcomes Quality Initiative |
| MA | Microalbuminuria |
| MDRD | Modification Of Diet In Renal Disease |
| mfp | Multiple Fractional Polynomial |

| | |
|---|---|
| MFP | Multiple Fractional Polynomial program in R |
| N/KDOQI | The National Kidney Foundation Kidney Disease Outcomes Quality Initiative |
| NICE | National Institute Of Clinical Excellence |
| NKDEP | National Kidney Disease Education Program |
| NKF | National Kidney Foundation |
| NSAIDS | Non Steroidal Anti-Inflammatory Drugs |
| PCR | Protein Creatinine Ratio |
| PVD | Peripheral Vascular Disease |
| QOF | Quality Outcomes Framework |
| RAAS | Renin Angiotensin Aldosterone System |
| RRT | Renal Replacement Therapy |
| SIGN | Scottish Intercollegiate Network |
| THIN | The Health Improvement Network |

# CHAPTER 1. INTRODUCTION TO THE THESIS

## 1.1.What Is Chronic Kidney Disease?

Chronic Kidney Disease (CKD) is an umbrella term for kidney damage defined as a reduced glomerular filtration rate (GFR), a measure of renal function, and/or having evidence of structural damage.[1] CKD is typically subdivided into a series of stages defined by GFR and/or such structural damage. (Table 1-1)

**Table 1-1. Stages of Chronic Kidney defined by the National Kidney Foundation/ Kidney Outcomes Quality Initiative (N/KDOQI) guidelines**

| Stage | Glomerular Filtration Rate in ml/min/1.73 $m^2$ | Description |
|---|---|---|
| **1** | 90+ | Normal renal function but has evidence of i.e. urinary abnormalities of proteinuria and or haematuria ii. Structural abnormalities iii. Genetic Renal trait |
| **2** | 60-89 | Mildly reduced kidney function but evidence of the abnormalities in stage 1 |
| **3a or 3a P*** | 45-59 | Moderately reduced kidney function |
| **3b or 3b P*** | 30-44 | |
| **4 or 4P*** | 15-29 | Severely reduced kidney function |
| **5 or 5P*** | <15 or on dialysis | Very severe, or End-stage Renal Disease May require dialysis |

* Denotes additional staging where P is added if the patient has proteinuria

As will be discussed further in this chapter CKD is highly prevalent in the general population (though the exact proportion remains unclear) and the mortality and the morbidity is very high. This chapter will describe how renal function is measured and in greater detail how chronic kidney disease is staged. This chapter will then summarise the prevalence of CKD, the impact of CKD on the population and then summarise risk factors for mortality and morbidity in patients with CKD. This chapter will finally discuss why primary care is suited to

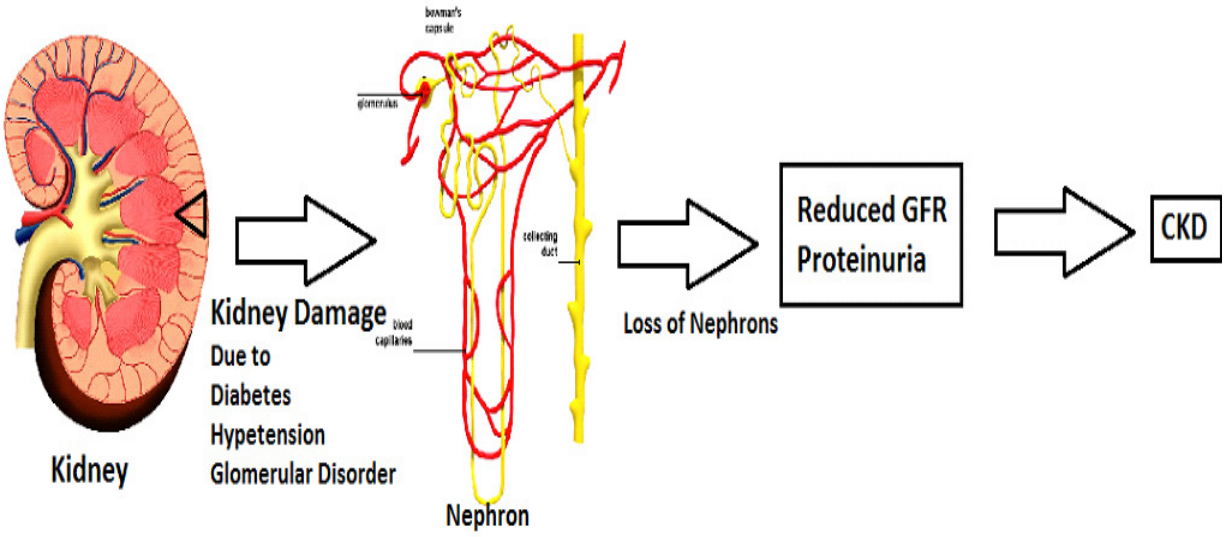
for research in patients with moderate kidney dysfunction and the aims and objectives of this thesis.

## 1.2.Markers of Renal Damage: How is Kidney Function Measured?

The following section details differing methods for determining glomerular filtration rate and/or urine protein excretion (a key marker of structural renal damage). It is split into two sections.

1. Measurement of glomerular filtration rate
2. Measurement of urine protein section

**Figure 1-1. A summary of kidney dysfunction**




### 1.2.1. Glomerular filtration rates and surrogate markers

Glomerular filtration rate (GFR) is defined as the volume of fluid filtered from the glomerular capillaries in a specified period of time. The GFR, expressed in ml/min (for an assumed body surface area of 1.73m$^2$), is higher in men than women who are age matched and declines with age.[2] Stages 3-5 CKD are defined by a reduction in GFR. This is intuitive as kidney damage will cause a reduction in the number of nephrons and as a consequence the GFR will

decrease. However in patients with kidney damage there are often compensatory mechanisms whereby the individual nephron filtration rate increases and hence there is little change in the GFR. In fact this compensatory mechanism, known as 'hyperfiltration,' can lead to elevated GFRs such as in patients with diabetic nephropathy.[3]

Because so much of this thesis depends on understanding the effects and influences of the measurement of GFR and of renal damage, the next sections consider these issues in detail.

### 1.2.2.Measuring GFR

GFR can be measured by the clearance of a continually infused exogenously administered substance such as inulin, which is freely filtered but not absorbed at the glomerulus.[4] However this measurement is invasive for the patient. The current gold standard for measured GFR is from clearance of single injection of $^{125}$I – labelled iothalamate or another radio labelled substances such as Ethylenediaminetetraacetic acid (EDTA), which corresponds closely with inulin clearance. However, though not as invasive as inulin clearance, these tests are expensive and time consuming and may not have any additional advantage when compared to estimated GFR (eGFR).[5]

### 1.2.3.Creatinine based estimation

In practice the measurement of the clearance of an endogenous substance called creatinine has been utilised for GFR estimation.[6] Creatinine is a breakdown product of creatinine phosphate in muscle. This substance is not protein bound, not metabolised by the kidney, is freely filtered through the glomerulus and is easy to measure.[7] Until the advent of directly reported eGFRs, serum or plasma creatinine was the most widespread method of determining kidney dysfunction and clinicians would use absolute values or trends to determine decline or improvement in renal function. However serum creatinine is affected by varying muscle mass and is hence affected by age, gender and race. Therefore the same

serum creatinine value in different individuals may reflect very different glomerular filtration rates.[8] A young black male with a serum creatinine of 150 µmol/l may not have renal dysfunction but an elderly white woman with the same result may have stage 3 CKD. Furthermore, creatinine is secreted by the tubule and hence serum levels are affected by medications that may interfere with this and diet, i.e. intake of meat will also affect levels.[7]

**1.2.4.Creatinine measurement**

Creatinine measurement originated in 1886 using the Jaffe reaction.[7] Current laboratory methods still use the basis of this reaction but other substances such as glucose and bilirubin molecules can be mistaken for creatinine. These are so called 'non creatinine chromagens'.[9] An American initiative to improve the accuracy of creatinine measure suggested that laboratories implement creatinine analysis that is traceable to Isotopic Dilution Mass Spectroscopy (IDMS).[9] IDMS is the gold standard of measuring creatinine (non creatinine chromagens are not measured) but is prohibitively expensive. However assays exist that are aligned or traceable to IDMS and these are more accurate at measuring the serum creatinine as they are less likely to mistake non creatinine chromagens. These are advantageous compared to analyses performed using non IDMS methods which are less specific and will therefore report higher levels of creatinine. This difference can be as high as 20%.[9] Additionally, there is considerable variability between non IDMS methods making comparability between laboratories difficult. Finally, serum creatinine measurement can be inaccurate when levels are low especially when under 88.4 µmol/litre.[9]

Another method of assessing kidney function is through the measurement of creatinine clearance. As creatinine is freely filtered at the level of the glomerulus, creatinine clearance can be calculated using serum creatinine and the creatinine concentration of a timed urine

collection over 24 hours.[7] The formula is given in Box 1-1. However creatinine clearance is cumbersome and requires accurate urine collections which patients find difficult.[10;11]

**Box 1-1. Creatinine clearance measurement**

For creatinine in mmol/l

Creatinine clearance in ml/min $\frac{Urine\ Creatinine\ concentration \times Urine\ flow\ rate}{Serum\ Creatinine}$

The urine flow rate is calculated from the volume of urine produced in the time period collected

Investigators have therefore sought to estimate GFR from simpler methods that are less patient intensive. Cockcroft and Gault found a linear correlation between mean serum creatinine and age and developed a formula (CG) which estimates creatinine clearance from the serum creatinine and uses surrogate markers such age, body weight and sex to adjust for differences in endogenous creatinine production (Box 1-2).[12] The study was carried out using only 249 men and applied an arbitrary adjustment factor of 0.85 to the creatinine clearance for women to account for reduced body mass without verifying this formula in this group. Creatinine clearance is overestimated as 15% of creatinine is secreted in the distal tubule.

**Box 1-2. Cockcroft Gault estimation of creatinine clearance**

Cockcroft Gault formula – This calculates the creatinine clearance which is estimation of glomerular filtration rate

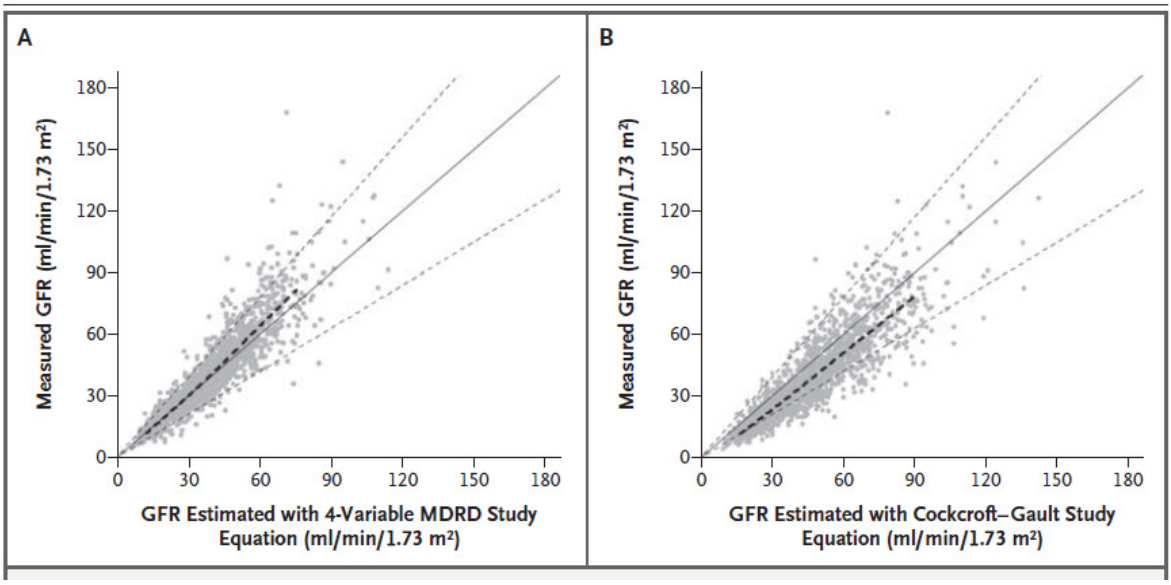
For creatinine in mmol/l

Creatinine clearance in ml/min = $\frac{(140-age)\times(weight\ in\ kg)\times(\ 1.23\ if\ male\ or\ 1.04\ if\ female)}{Serum\ Creatine}$

### 1.2.5.MDRD equation

The investigators of the Modification of Diet in Renal Disease (MDRD) study of CKD patients developed a new formula to estimate GFR.[13] In their study, 1628 patients had their GFR assessed by measuring the clearance of $^{125}I$ – labelled iothalamate serum. Additionally they measured the serum creatinine and a 24 urine collection for urinary creatinine. They then calculated the measured creatinine clearance and Cockcroft Gault clearance. Using other demographic and biochemical variables collected they developed mathematical models and created a 6 variable formula that demonstrated better precision and accuracy at estimating GFR in comparison with the CG formula (Figure 1-2).[13;14]

**Figure 1-2. Relation of estimated GFR to measured GFR in the participants of the MDRD study.**




**Each point represents the baseline measurement. The solid line represents the line of identity. The bold dashed line represents the fitted line with smoothing splines function plotted for the the 2. and 97.5 percentile of estimated EGFR. The thin dashed lines repesent the difference of ± 30 percent between estimated and measured GFR**

The same investigators then further validated a simplified equation (4 variable MDRD equation in Box 1-3) that only required the serum creatinine, age, gender and black or non-black ethnicity.[15]

**Box 1-3. The MDRD Equation to calculate eGFR**

Modification of Diet in Renal Disease formulae.

This estimated the glomerular filtration rate (eGFR) for creatinine in mmol/l

For serum creatinines not analyzed by the standardized creatinine assay traceable to IDMS methods

Estimated Glomerular filtration in mls/min/173m$^2$

$$= 186 * \times \left(\frac{\text{Creatinine (mmol/litre)}}{88.4}\right)^{-1.154} \times \text{Age}^{-0.203} \times (0.742 \text{ if female}) \times (1.21 \text{ if black})$$

***175 when IDMS method used**

However several factors must be considered when interpreting eGFR using the MDRD equation. This equation has not been validated in certain ethnic groups. Adjustments to the formula have to be made for patients in Japan, China and Korea as these patients may have been under-represented in the MDRD cohort.[16-18] Though there were sufficient numbers of African Americans in this cohort, they are unlikely to represent all black patients. In South African blacks, the MDRD eGFR, with correction for ethnicity, overestimated GFR and the equation for white people corresponded more closely with actual renal function.[19] Additionally it is likely that the MDRD eGFR underestimates eGFR when the eGFR is above 60 ml/min/1.73m$^2$.[20] Only patients with CKD were included in the MDRD cohort and studies in living kidney donors show the MDRD equation underestimates GFR.[21] Additionally both the Cockcroft Gault and the MDRD equation perform less well in obese individuals. It is difficult to comment how applicable this formula is to the UK population as estimated eGFR has not been compared to measured GFR in the UK. This may be particulary relevent to the large minority population of Indian-subcontinentals who were mostly absent in the MDRD study.[19]

### 1.2.6.The Mayo Quadratic formula

In order to more accurately describe renal function in healthy individuals as well as those with CKD, investigators modelled and developed the Mayo Quadratic Formula(Box 1-4).[22] Using 900 individuals (of whom a third had CKD and the rest were healthy), who had their renal function tested formally using isotopic methods, the investigators found better agreement with this formula in their cohort then the MDRD equation. In another study performed in secondary care patients classified with CKD using the Mayo equation were more likely to develop cardiovascular disease then when classified by the MDRD equation suggesting better risk prognostication if the Mayo formula is used.[23] However as will be discussed later in this section, this formula was developed using a different creatinine assay to the MDRD and the MDRD equation may not have been accurately used. In subsequent studies the MDRD equation has performed better in general populations and healthy individuals.[24;25]

**Box 1-4. The Mayo Quadratic formula**

For creatinine in mg/dl

Estimated Glomerular filtration in mls/min/1.73m$^2$ = $\exp\left(1.911 + \frac{5.249}{Serum\ Creatinine} - \frac{2.114}{Serum\ Creatinine^2} - 0.00686 \times Age - 0.205(if\ female)\right)$

If Serum creatinine < 0.8mg/dl then assume creatinine is 0.8

### 1.2.7.The Chronic Kidney Disease Epidemiology Collaboration (CKD EPI) formula

The same group that developed the MDRD equation have subsequently developed another set of equations (Box 1-5) aiming to improve accuracy in the black ethnicity and at higher GFRs >60mls/min/m$^2$.[26] The investigators developed their equations from ten multinational studies which included patients with measured GFR and then validated this

against patients with measured GFRs from a further 16 studies. This cohort was much more diverse and included patients who were healthy and had a higher proportion of black patients (31-33% versus 12% in the MDRD study). The new equation (CKD-EPI) had better performance in the elderly, patients with eGFRs above 60 ml/min/1.73m$^2$, obese individuals and transplant patients.[27;28] However the eGFR was more likely to be accurate with the MDRD equation at lower eGFRS.[29]

**Box 1-5. The CKD EPI formulae**

| |
|---|
| If Female:<br>eGFR= $144^{\$} \times \frac{\text{Creatinine}^{a}}{61.9} \times 0.993^{\text{Age}}$<br>If serum creatinine (Scr) <= 61.9 mmol/l than a = $-0.329$<br>If Scr > 61.9 mmol/l than a= -1.209<br>$^{\$}$If black race multiply by 166 instead of 144 |
| If male<br>eGFR= $141^{£} \times \frac{\text{Creatinine}^{-a}}{79.6} \times 0.993^{\text{Age}}$<br>If serum creatinine (Scr) <=79.6 mmol/l than a=-0.411<br>If serum creatinine (Scr) > 79.6 mmol/l than a=--1.209<br>$^{£}$ Multiply by 163 instead of 141 if Black race<br>Estimated Glomerular filtration in mls/min/1.73m$^2$ |

### 1.2.8.The impact of creatinine measurement on eGFR

As mentioned above, creatinine values vary due to laboratory methods and these differences can have a big impact upon eGFR measurement. Prior to the advocacy of IDMS standardisation as discussed above, considerable variation existed between non IDMS methods. The original MDRD equation was derived using creatinines analysed using the Beckman Analysis (Non IDMS method). However if the creatinine is measured by a different assay and was not 'calibrated' to the Beckman method then at a GFR of 60 ml/min/1.73m$^2$ there could be up to a 10% difference in the eGFR.[30] As the GFR increases beyond 60ml/min/1.73m$^2$ this percentage error increases and consequently some laboratories report the eGFR simply as above 60 ml/min/1.73m$^2$ rather than an actual figure.[30] This can

lead to difficulty in classifying stages 1-2 CKD. In the UK, laboratories report this 'un-calibrated' serum creatinine, then apply an adjustment factor to the creatinine to make it equivalent to Beckman method and then calculate the eGFR using the MDRD equation.[31]

In order to counter the variability in serum creatinine assay, a laboratory working group was devised by National Kidney Disease Education Program (NKDEP).[9] The NKDEP recommended in 2006 that creatinine is measured by methods traceable to IDMS and that laboratories should aim to reduce the error to less than 10% between separate measurements of eGFR.[9] Laboratory methods aligned to IDMS allow calibration and standardisation of serum creatinine measurement, however serum creatinines appear lower and the MDRD equation had to be altered to reflect this (Box 1-3).[32]

In the United Kingdom although the Department of Health recommends IDMS aligned methods the uptake across laboratories has been not been universal: even by 2009 only 50-60% of laboratories were using IDMS aligned methods.[31] The CKD EPI equation has been validated using only IDMS aligned laboratory methods. The variation in laboratory methods may be why the MAYO equation was not applicable to other populations. The variation in formula to estimate eGFR may lead to differing estimates in the prevalence of CKD and this is discussed and explored in section 1.4 and Chapter 2.

### 1.2.9.What is proteinuria and how is it measured?

Proteinuria, is excess protein excretion into urine and is a marker for renal progression and cardiovascular disease.[33] In a normal kidney, only small amounts of albumin (the main protein in the body) are excreted in the urine. In diseased kidneys, (e.g. those found in patients with diabetes), the glomerular barrier is damaged and greater amounts of albumin

are filtered into the glomerulus, which the nephron is unable to absorb downstream and therefore albumin passes into the urine resulting in albuminuria.[34]

Even moderate levels of urine albumin excretion i.e. 'high' albuminuria (albumin excretion of 30-300mg per day or urine albumin to creatinine ratio of 3-30mg/mmol) are a sensitive marker for glomerular pathology prior to reduction of glomerular filtration rate. Therefore the Kidney Disease Improving Global Outcomes (KDIGO) guidelines use this threshold to diagnose proteinuria and classify those with stages 1-2 CKD patients (if they have normal range GFRs). These guidelines should not to be confused with N/KDOQI guidelines above. The differentiation between different levels of albuminuria/proteinuria are show in Table 1-2.[33] The different methods of measuring albuminuria are discussed later.

**Table 1-2. Quantification of proteinuria: comparison of different methods**

| | **Microalbuminuria** | **Macroalbuminuria/Significant proteinuria** | **Overt Proteinuria** |
|---|---|---|---|
| **24 hour Urine Collection of Protein (mg/24 hours)** | N/A | 500-999 | 1000+ |
| **Protein Creatinine Ratio in mg/mmol** | N/A | 50-99+ | 100+ |
| **24 Urine Collection of Albumin(mg/24 hours)** | 30 - 299 mg | >300mg | NA |
| **Albumin Creatinine Ratio in mg/mmol** | 3.5 in women<br>2.5 in men | 30-69 | 70 |
| **Urine Dip test** | Trace/1+ | >1+ | NA |
| **Definition in contemporary literature[35]** | High | Very High | |

#### *1.2.9.1.Methods for measuring urine protein excretion*

#### *1.2.9.2.Laboratory Methods*

The gold standard for quantifying proteinuria is 24 hour collection of urine for protein or albumin levels.[33] However this method is expensive, patients find this cumbersome and difficult to comply with and subsequently many samples are inadequate.[10;11] Spot urine protein creatinine ratio (urine PCR) is effective at ruling out proteinuria, correlates highly with 24 hour collections but is not as sensitive.[36-38] Additionally exercise and fever can affect levels of urine PCR but these are the trade-offs for an easier test.[39] In practice spot urine albumin:creatinine ratio (Urine ACR) is used: high albuminuria is a marker of kidney damage and predictive of morbidity and mortality and cannot be detected by Urine PCR.[33]

#### *1.2.9.3.Urine dipsticks: Advantages and Disadvantages*

Urine dipsticks are cheap, easy to use and give an immediate result for multiple analytes including protein, blood, nitrates and leucocytes. They semi-quantitatively analyse amounts of albumin in urine and give results of negative, trace, 1+, 2+ and 3+ for the presence of protein (Table 1-2). However their use in clinical practice is problematic and has several major criticisms. Firstly the results of dipstick tests are user dependent as the colour on the dipstick is compared to a colour chart. Secondly, urine albumin excretion is dependent upon urine volume and dipsticks do not account for this and erroneous results are possible. Consequently studies report that they have variable sensitivities and specificities.[40;41] More recently, automated urine dip testing has become common place and consequently dipstick results may be more accurate.[33]

In a screening study of 10944 Australians aged over 25, where individuals had concomitant urine dipstick using Bayer Multistick (automated result) and Urine ACRs, a dipstick threshold

of 1+ detected an ACR above 3.4mg/mmol with a sensitivity of 58% and the specificity of 95%. To detect an ACR of 33.9 mg/mmol the sensitivity was 99% and the specificity was only 93%. This indicates that urine dip has poor sensitivity at detecting albuminuria but if a dipstick is positive then the patient is likely to have albuminuria.[42]

The equivalent results from urine PCR, ACR, 24 urine protein excretion and urine dipstick are shown in Table 1-2 and are adapted from the KDIGO report.[33]

## 1.3.The Current Classification of CKD: Evidence Base

The current classification of CKD in the United Kingdom according to NICE (National Institute for Clinical Excellence) is based on the N/KDOQI guidelines (Table 1-1).[1] The N/KDOQI guidelines have been updated and adopted internationally with the guidelines being produced by the Kidney Disease Improving Global Outcomes (KDIGO) in 2012.[43] These original N/KDOQI guidelines were developed after a systematic review was undertaken, and although eGFR below 15 ml/min/1.73m$^2$ was strongly associated with poor outcomes such as requiring dialysis or cardiovascular events, there were few studies on the relationship between stages 1-4 CKD and clinical outcomes.[1]

To classify patients with stages 3-5 CKD requires two eGFRS under 60 mls/min/1.73m$^2$ measured on two occasions sufficiently apart to rule out a temporary reduction in GFR (eg from acute kidney injury); the impact of this on the prevalence will be discussed later.(Chapter 2.1) This definition has been operationalised by KDIGO as *at least* three months apart and adopted by the UK National Institute of Clinical Excellence (NICE) guidelines for CKD.[1;44] However, the UK payment for performance system called Quality Outcomes Framework (QOF) defined stages 3-5 CKD by using two eGFRs under 60 mls/min/1.73m$^2$ *within* 3 months(for QOF 2008/2009).[44;45]

The definition of proteinuria is interesting; the original guidelines state a definition of proteinuria with a PCR of 23mg/mmol, the UK and Scottish guidelines (SIGN) define proteinuria as a PCR of 70mg/mmol and 100mg/mmol respectively.[44;46] The recent KDIGO guidelines suggest even lower thresholds of ACR of 3.4mg/mmol as proteinuria but note that as low levels of albuminuria can be transient, it should be confirmed over a period of 3 months (the original NKDOQI guidelines stated that this be above 2 weeks).[47] For CKD

stages 3-5, The UK and Scottish guidelines suggest that if patients have proteinuria, a 'p' suffix should be used e.g. CKD stage 3aP if the patient has proteinuria.[44;46]

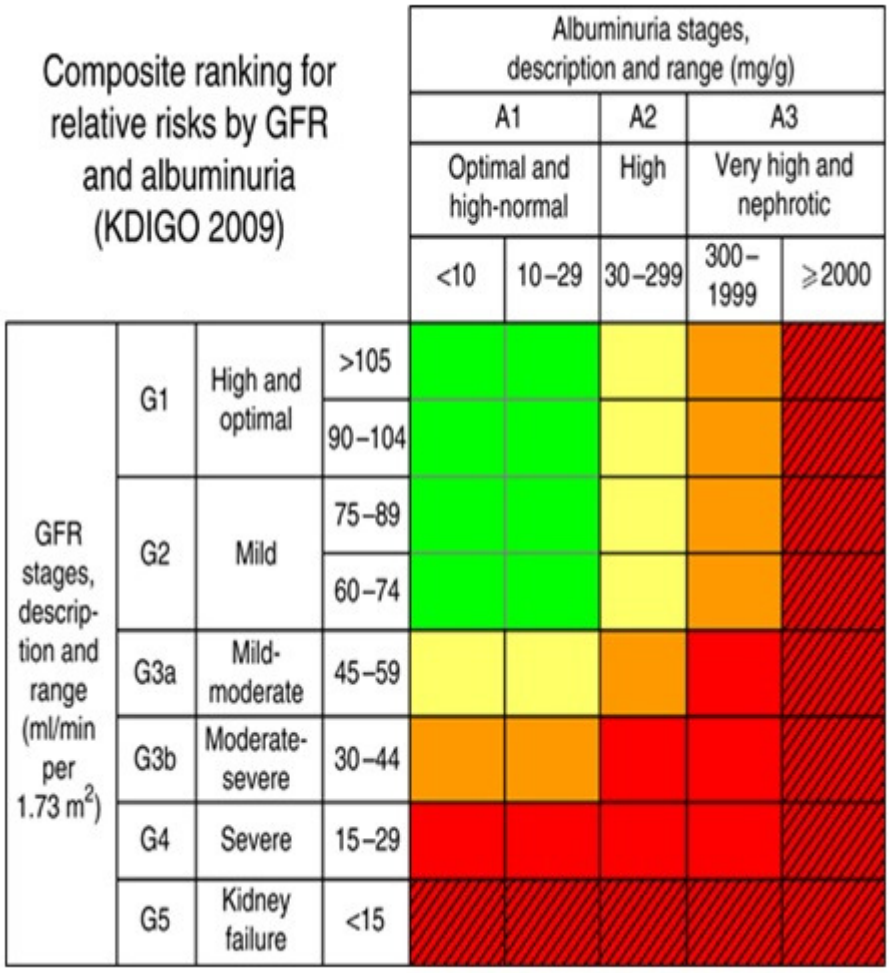
The relationship between eGFR and increasing albuminuria and clinical outcomes is independent and this is not accounted for in the current classification. In the CKD Prognosis Consortium, patients with an eGFR 45-60 (CKD stage 3a) but no albuminuria had a HR for death 1.3 times to that of the reference group, but a patient with an eGFR of 60-75 ml/min/1.72m$^2$ and an ACR above 30 mg/mmol (CKD Stage 2) had a HR of 2.7.[43] Based on these observations Tonelli et al proposed a classification system based on actual risk of developing renal sequelae (Table 1-3),[35] this system resulted in some patients being reclassified to lower risk category.[35]

**Table 1-3. Current NKF KDOQI CKD staging system and alternate system of CKD risk categories**

<table>
<tr><td colspan="4">Current NKF KDOQI System</td></tr>
<tr><td rowspan="2">GFR ml/min/1.73m$^2$</td><td colspan="3">Albuminuria</td></tr>
<tr><td>Normal</td><td>High</td><td>High</td></tr>
<tr><td>90</td><td rowspan="2">Stage 0<br>(no CKD)</td><td colspan="2">Stage 1</td></tr>
<tr><td>60-89.9</td><td colspan="2">Stage 2</td></tr>
<tr><td>45-59.9</td><td colspan="3" rowspan="2">Stage 3</td></tr>
<tr><td>30-44.9</td></tr>
<tr><td>15-29.9</td><td colspan="3">Stage 4</td></tr>
<tr><td colspan="4">Alternate System</td></tr>
<tr><td></td><td colspan="3">Albuminuria</td></tr>
<tr><td>GFR ml/min/1.73m$^2$</td><td>Normal</td><td>High</td><td>Heavy</td></tr>
<tr><td>90</td><td rowspan="2">Risk category 0<br>(no CKD)</td><td>Risk category<br>1</td><td>Risk category<br>3</td></tr>
<tr><td>60-89.9</td><td rowspan="2">Risk category<br>1</td><td rowspan="2">Risk category<br>2</td><td rowspan="3">Risk category<br>4</td></tr>
<tr><td>45-59.9</td></tr>
<tr><td>30-44.9</td><td>Risk category<br>2</td><td>Risk category<br>3</td></tr>
<tr><td>15-29.9</td><td>Risk category<br>3</td><td colspan="2"></td></tr>
</table>

KDIGO using the CKD Prognosis Consortium data, has suggested in their 2012 guidelines that patients should be graded according to their level of albuminuria as well as eGFR and have introduced the use of a heat map to allow the clinician to predict risk(Figure 1-3).[43;47]

**Figure 1-3. Composite Ranking for Relative Risks by glomerular filtration rate (GFR) and Albuminuria**




(Kidney Disease: Improving Global Outcomes (KDIGO) 2009). The risk for morbidity is from lowest to highest by colour is Green, Yellow, Amber, Red and Red with hashes.

## 1.4.The Prevalence of Stages 1-5 CKD

The prevalence of stages CKD 1-5 will be discussed in length in chapter 2.1 but is introduced here. There is considerable variation in how creatinine is measured (hence variation in eGFR result) and the methods of estimating eGFR i.e. MDRD and CKD EPI as demonstrated in Chapter 1.2. This inconsistency in eGFR leads to changeable estimations of stages 3-5 CKD prevalence of between 1%-13.6%. Over half these patients have stages 3, discussed and developed further in chapter 2.1.[48] To counter this unpredictability in eGFR, all CKD guidelines state that stages CKD 3-5 should be confirmed using two eGFRs below 60 ml/min/1.73m$^2$. However only one prevalence study has attempted to correct for chronicity of stages 3-5 CKD i.e. estimate the prevalence of CKD using two blood results and this was 6.76%.[49] This study used two blood tests where available and a single blood test where two blood tests were not available to define stages 3-5 CKD.[49] This leads to the important question of what is the prevalence of stages 3-5 CKD as the mortality and morbidity is very high and this disease state requires considerable resources.(Chapter 1.5)

The prevalence of stages 1-2 CKD have generally been ascertained on excess urine protein secretion and this, like eGFR estimates, depends upon testing in the first place and also what methodology is used to define proteinuria.

**Table 1-4. Prevalence variation of Stages CKD 1-5 adapted from Chapter 2.1 [8;49-91]**

| Method | Prevalence |
|---|---|
| Stages 3-5 CKD using MDRD formula<br>Single blood result | 1.0 - 13.6% |
| Stages 3-5 CKD using MDRD formula<br>Two blood results | 6.8%<br>(Adapted from the QICKD paper)[49] |
| Stages 3-5 CKD using CKD EPI formula<br>Single blood result | 4.4 – 6.0 % |
| Stages 1-2 CKD | 1.6- 12.1% |

## 1.5.CKD: What Are The Mortality And Morbidity Implications?

### 1.5.1.Cardiovascular disease and mortality

The commonest CKD stage is Stage 3 (30-50% of all patients with CKD).[48] However renal sequelae such as reaching stage 5 CKD and or requiring dialysis are uncommon in this group. In a cohort of nearly 28 000 patients in the US, only 1% of patients with stage 3 CKD developed ESRD after 5 years follow up.[92] However in the same study, despite the relatively low risk of ESRD, the mortality rate was extremely high at 24.5% for stage 3 CKD compared to 10.2% mortality in those without CKD.[92] .

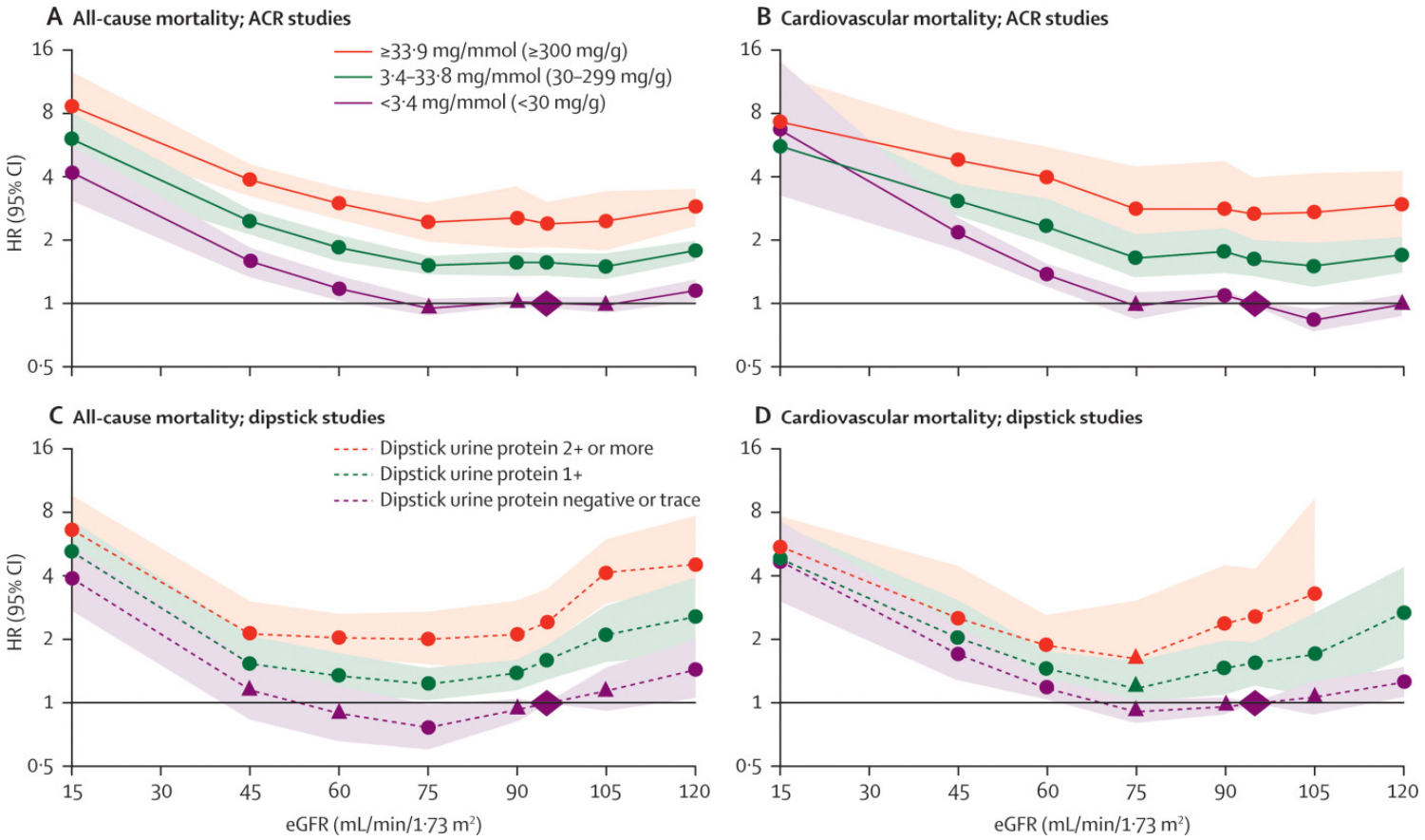
Subsequently the CKD Prognosis Consortium conducted several meta-analyses to examine the impact of CKD in general, high risk and kidney disease populations.[34;43;93-95] The largest study consisted of 1 234 182 individuals included in 23 studies of the general population.[93] In a multivariable analysis, the hazard ratio for all-cause mortality was 1.18 (95% CI 1.05-1.32) for patients with stage 3a CKD (a cubic spline knot at an eGFR of 60 ml/min/1/73m$^2$) in comparison to patients with normal renal function (a cubic spline knot at 95 ml/min/1/73m$^2$).[93] The risk increased as the GFR fell (Figure 1-4). This risk could be further stratified by the level of proteinuria with increased levels of proteinuria being associated with a stepwise increase in mortality and cardiovascular disease (Figure 1-4). Even small amounts of albumininuria were associated with increased morbidity and mortality. A large proportions of deaths in the study were attributable to cardiovascular disease (20%, 9637/45 584) [93].

### 1.5.2.Renal Outcomes

In two further meta-analyses, a stepwise increase in albuminuria and decreased eGFR was associated with increased likelihood of requiring renal replacement therapy, developing

progressive renal disease and acute kidney injury in the general population, high risk and kidney disease cohorts.[95] In the general population, patients with CKD stage 3a (HR 9.6 95% CI 7.0-13.2) were 10 times less likely to develop ESRD compared to Stage 3b (HR 98.1, 61-8-156), in comparison with the reference group (those with an average eGFR of 95ml/min/1.73m$^2$). The risk of developing ESRD rose exponentially to 573 times more likely for patients with Stage 4 compared to patients with the reference group.[95]

**Figure 1-4. Risk of mortality split by eGFR and proteinuria levels from CKD Prognosis Consortium**



### 1.5.3.Economic Impact of CKD

The healthcare cost of CKD will discussed in this subsection but it is worth highlighting that CKD is not only associated with these costs. A review by Perico et al, suggests that renal/urinary tract diseases were associated with 830,000 deaths per year world wide and 1 886 700 disability adjusted life years in 2001.[96]

Although the risk of ESRD is not high, where it occurs, these patients require renal replacement therapy (RRT, either dialysis or transplantation) and this is expensive.[97] Renal replacement programmes costs from developed countries varied from 2% (Australian Healthcare expenditure on RRT in 2004) to 6.7% (United States Medicare Budget in 2006).[98;99] In the UK in 2009 there were 49 808 patients on RRT and of these 25 790 patients were on dialysis.[100] The annual cost of dialysis alone was over £780 million comprising 1.6% of the annual National Health Service (NHS) budget of £48 billion in 2009.[97;100] Such costs are highly disproportionate since patients with ESRD only comprise of small proportion of the population.

The true economic impact of all stages of CKD is likely to be much higher. In United States, in the Medicare health care systems (patents aged over 65 and or long term conditions), care for CKD patients accounts for nearly 27% of the costs.[101] In a UK model, investigators modelled the cost of CKD stages 3-5 based on RRT costs, primary care costs, outpatient attendances and inpatient costs between 2009-2010 and found the cost to be £1.45 billion pounds, approximately 3% of NHS budget.[97]

In summary these studies highlight the large public health burden of CKD and as will be discussed further in this chapter, this disease is not uncommon and therefore the risk factors for cardiovascular disease and mortality require further discussion and exploration.

## 1.6.Risk Factors for Cardiovascular Disease and Mortality

As described earlier in this chapter, incident cardiovascular disease and mortality is abundant in patients with chronic kidney disease. However as will be described in further detail in chapter 4, cardiovascular risk factors such as diabetes, hypertension, hypercholesterolaemia, obesity, and atrial fibrillation are associated with the development of CKD.[102-108] This is a paradox when considering risk factors for cardiovascular disease in patients with CKD.[109] For example smoking is associated with increased cardiovascular disease, diabetes and cancers. However smoking is implicated in the development of CKD.[103]

As shown in Table 1.5, there is a complex relationship between risk factors and subsequent mortality and morbidity. Many of the continuous risk factors have a non linear relationship such as BP, BMI and cholesterol. This may be a result of reverse causation as the CKD disease state modifies the relationship between the risk factor and the outcome.

Additionally most of these studies are not CKD specific and are either post hoc or sub group analyses.[110;111]

Additionally there are CKD specific risk factors such as low haemoglobin, bone mineral disorders and proteinuria.[112;113] The complexity of these risk factors and the lack of available evidence make it difficult to prognosticate risk in cohorts associated with high rates of death and cardiovascular disease.

**Table 1-5. Traditional and Non traditional risk factors for CVD and Mortality in CKD: a comparison with the general population.**

| Traditional CVD Risk factors | Effect in general population | Effect in CKD population |
|---|---|---|
| Blood pressure (BP) & Antihypertensive drugs | Increasing BP associated with increased mortality and treatment with all classes beneficial[114;115] | U shaped relationship with mortality. Angiotensin blockers demonstrate greatest efficacy with questionable benefits from other agents.[116-118] |
| Anti platelet agents | Secondary prevention of CVD[119] | As general population but with increased bleeding[120] |
| Cholesterol reduction & Statins | Raised cholesterol and reduction with statins beneficial[121;122] | Raised cholesterol has U shaped relationship with CVD risk but statins beneficial[121;123;124] |
| Smoking | Increased risk[113] | Same as general population[113] |
| Atrial Fibrillation | Increased risk[125] | Same as general population[126] |
| Body Mass Index | Increased risk[125] | U shaped relationship[127;128] |
| Ethnic Group | Increased risk with black and Asian population[125] | Unknown. Asian and black population with ESRD have increased survival[129] |
| Deprivation index – Townsend Quintiles | Increased risk with lower deprivation[130] | Same as general population[130] |
| CKD SPECIFIC RISK FACTORS | | |
| Proteinuria | NA | Increased risk with high and very high albuminuria[93] |
| Haemoglobin | NA | Increased risk with low Hb but also treatment with erythropoietin[131;132] |
| Bone Mineral Disorder | NA | Increased risk with Bone mineral disorders[112;113] |

## 1.7.Why is Primary Care the Ideal Forum for CKD Research?

As will be discussed in the introduction in chapter 3, NICE guidelines suggest that stage 3 CKD should be managed in primary care unless patients are hypertensive or have progressive renal dysfunction.[44] This is the largest group of CKD patients and may constitute up to 5% of the population.[133] Primary care is suited to epidemiological research in CKD because primary care physicians are the point of access for all patient in the UK.[134] The majority of the population (98%) are registered with a General Practice and most primary care providers have complete electronic records.[134;135]

Electronic patient record (EPR) systems have existed for many years.[136] These contain demographic information, clinical information about consultations, and prescriptions and locally recorded medical information such as weight or blood pressure. Additionally they are linked to local pathology data such as blood and radiology test results.

Pharmaco-epidemiological information within the EPR is generally accurate and the diagnoses recorded, such as cardiovascular disease, have been shown correspond with diagnoses made in secondary care.[137;138] Table 1-6 shows examples of research conducted using primary care databases and their utility in research.

**Table 1-6. Example of routinely collected database studies in the United Kingdom with emphasis on kidney disease and cardiovascular disease**

| Title | Database | Description |
|---|---|---|
| **Q Risk Cardiovascular risk Calculator 2007[138;139]** | Q Research/ THIN | A cohort study in which patients were free of cardiovascular disease and diabetes. The investigators modelled risk factors in 1.28 million patients aged 35-74 to determine cardiovascular disease. This model was externally validated by analysing the THIN cohort [140] |
| **Q Risk 2 Cardiovascular risk calculator 2008[124]** | Q Research/ THIN | A similar cohort study as above where they CVD outcomes were analysed in 2.3 million patients in primary care. A better calibrated cardiovascular risk prediction tool was developed then the Framingham equation.[141] |
| **Q Kidney 2010[142]** | Q Research/ THIN | CKD disease predicting model derived from over 1.5 million patients aged 35-79. External validation from analysis of the THIN database. Had missing creatinine data[143] |
| **Patients with Hypertension: A Population-Based Case-Control Study [144]** | THIN | Case control study in hypertensive patients with new incident Gout (n= 24 768) compared with matched controls (n=50 000). Found that calcium channel blockers and losartan were independently associated with Gout |
| **Suicide related Events in patients treated with anti-epileptic drugs[145]** | THIN | Case control study where patients with epilepsy, depression and bipolar disorder were studied to determine whether anti-epileptics were associated with increased suicide related events |
| **Use of Antihypertensive Medications and Mortality of Patients With Autosomal Dominant Polycystic Kidney Disease: A Population-Based Study[146]** | CPRD | 2085 patients with polycystic kidney disease. They found that the number of patients with anti-hypertensives increased by 2008 and there was a decrease in mortality |
| **Risk of cardiovascular disease and all cause mortality among patients with type 2 diabetes prescribed oral anti-diabetes drugs: retrospective cohort study using UK general practice research database.[147]** | CPRD | 915 21 patients with diabetes. Patients on sulphonyuria monotherapy compared to metformin were found to have increased mortality. Pioglitazone was found to be associated with less risk compared to rosiglitazone |
| **Exposure to oral bisphosphinates and risk of oesophageal cancer[148]** | CPRD | Case control study examining bisphosphonate use and association with oesophageal cancer. The study found no association with bisphosphinates |

However despite primary care being a sensible setting for CKD research, there are a lack of specific CKD prognostic studies in primary care. QRESEARCH incorporates CKD as a variable into their risk prediction model but do not delineate risk by CKD stage.[125] Another primary care CKD study models the development of stage 5 CKD in patients with CKD.[139]. However as discussed before stage 5 CKD is relatively uncommon sequelae of CKD in comparison cardiovascular disease.[95] Therefore there is gap in cardiovascular research in CKD patients in primary care despite this being suitable setting for such research

### 1.7.1. Quality Outcomes Framework

Management of CKD in primary care is incentivised by the Quality Outcomes Framework.[140] This specifies that primary care physicians should produce a CKD register. However the proportion of patients on this register is less than national estimates.[140] This suggests that patients may not be recognised with CKD. This may affect subsequent management of patients with CKD as in other diseases, those not QOF disease register had worse management.[141] In a secondary care setting patients not identified with CKD were less likely to have monitoring for CKD complications.[142] Therefore primary care should be evaluated for CKD recognition and management of patients identified.

## 1.8.Large Primary Care Databases and their Utility in Research

The primary care resource chosen for use in this thesis was the 'Health Improvement Network' (THIN) and its use in epidemiological research is described in the next section.

Medical research databases in the United Kingdom began in 1987 when Value Added Medical Practice system Ltd (VAMP) was conceived.[136] This was initially an electronic practice system set up by Dr Alan Dean in the UK, to manage his practice and then subsequently was taken up by other practices. VAMP consisted of two components, the practice data computer system and practice research database. Funding to support the system was obtained by selling data to pharmaceutical companies. In 1993 the VAMP research database was no longer profitable and VAMP was split into the clinical system and research database. The clinical system was sold to Reuters [becoming what is now known as "Vision"] and the research system sold to the UK Government and labelled as the General Practice Research Database (GPRD) for use in non-profit research.[143] This system was eventually managed by the Medicines Healthcare Regulatory Authority (MHRA) and has recently been relabelled as Clinical Practice Research Datalink. (CPRD) CPRD now receives information from other sources other than VAMP and Vision.[143]

At the time of the development of CPRD, a separate licence for the research data was maintained by Epidemiology and Pharmacology Information Core (EPIC UK). In 2002, following the expiry of the original licence, EPIC (UK) set up the Health Improvement Network (THIN) in collaboration with In Practice Systems ltd. (INPS).[136] This practice research database received information from participating practices using Vision software (owned by INPS). Due to the common origin of data source for CPRD and THIN half of the practices that contribute data to CPRD also contribute data to THIN. CPRD is the largest primary care research database followed by QRESEARCH and then THIN.[136;143]

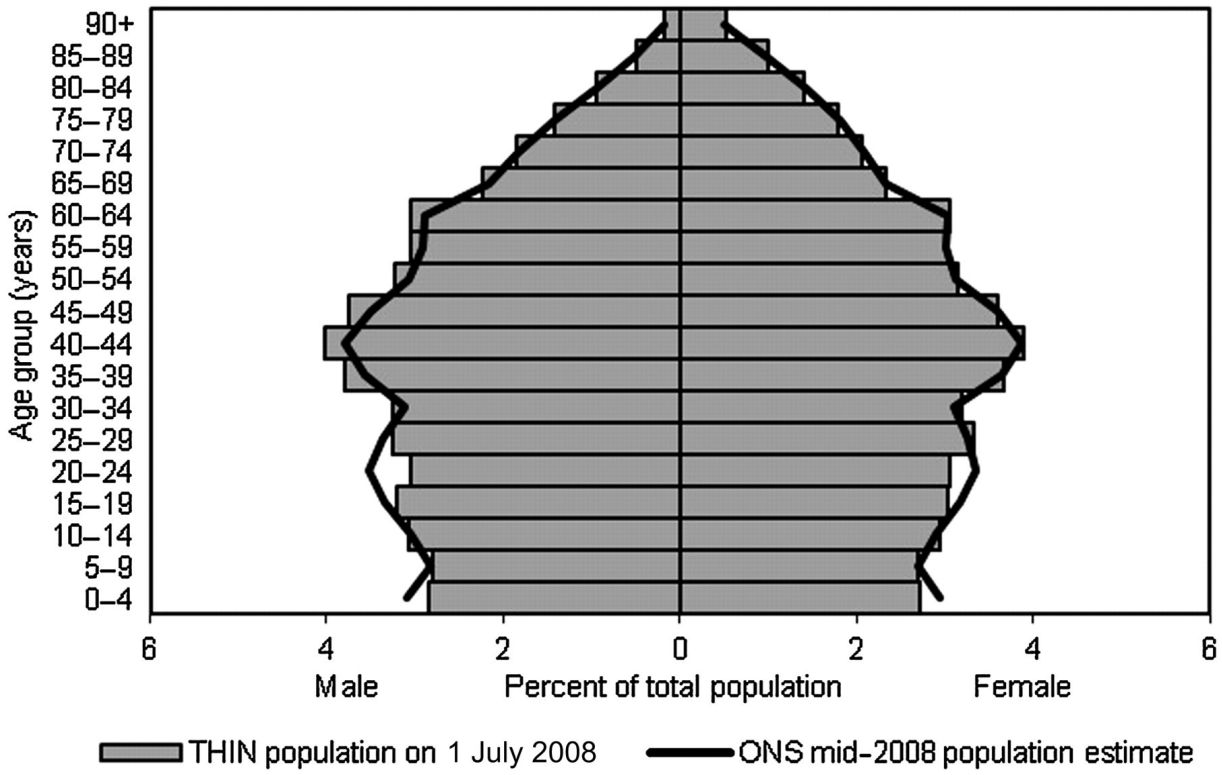
THIN consists of anonymised primary care records from general practices in the United Kingdom: in November 2009 there were just over six million patients in the database from 426 practices comprising approximately 6% of the United Kingdom population.[136] Data include historical records from 1980 when VAMP began and has been collected prospectively by VISION practices since 2002. Routinely collected data is entered into the patient's record forming the general practice clinical record and is automatically anonymised and sent to THIN. The data has pre-approved ethics for research with a scientific committee overseeing new applications for use of the data. The data consists of five components which are described in Table 5. All data files have common elements of a unique practice identifier and unique patient identifier in the practice.[136]

### 1.8.1.The strength and limitations of THIN

#### *1.8.1.1.Strengths*

The THIN database has been used in a number of research studies (Table 2). THIN covers 6% of the population and is broadly representative in age and sex of the UK general population(Figure 1-5).[144] As THIN is an unselected patient population it is more likely to represent a 'real-world' cohort where the impact of different treatments and guidelines can be evaluated.[136]

**Figure 1-5. Comparison of the structure of The Health Improvement Network (THIN) population on 1 July 2008 with the Office for National Statistics (ONS) mid-2008 UK population estimate.**




The consultation patterns and prescription rates documented in THIN are similar to national estimates from the general household survey (an annual survey of private households by the Office of National Statistics) and Department of Health.[145] Research using the THIN database shows that known cardiovascular risk factors are associated with cardiovascular disease and these results were similar to those from GPRD.[146] The THIN database demonstrates the association of HMG Co-A reductase inhibitors (statins) with the reduction in vascular outcomes similar to reductions observed in the Heart Protection Study clinical trial.[147] THIN prevalence and incidence of smoking and coronary heart disease mirror national estimates and estimates from other trials.[144;148;149]

#### *1.8.1.2.Limitations*

THIN has a number of limitations. The mortality rate is 5 % less than national averages probably reflecting reduced deprivation in practice areas contributing patient records to THIN such as the south of England.[144;150] In THIN, 22.7% of people are within the most affluent quintile according to the national Townsend score and 12.2% are in the least affluent quintile suggesting deprived areas are under-represented.[150] THIN is not useful at examining personal patient's characteristics such as non-compliance and over the counter medication because this information is not recorded. Like any of the research databases there is poor linkage to secondary care.[144;151]

#### *1.8.1.3.Using THIN database for research – key considerations and data components:*

After a patient registers with a practice, their medical records may take another 6 months to arrive and be entered electronically. Therefore only data after this 6 month period is likely to be accurate; especially for older data captured before GP to GP data transfer was available.[138;144;146;148;152;153] Additionally data is only likely to be accurate in patient's records after the acceptable mortality reporting (AMR) date. The AMR date is a date when the practice's reported mortality rate was analogous to the standardised mortality rate of the practice since this is likely to represent when the practice information was accurately computerised. This date not only represents from when death reporting was likely to be accurate, therefore reducing immortal time bias, but from when the practice was likely to report all events.[150] The data components of THIN are shown in Table 1-7.[136]

**Table 1-7. Description of THIN data components.**

| Data Component | Description |
|---|---|
| **Patient File** | This includes the practice identifier and patient identifier, the date of registration at the practice, year of birth, date of deregistration and the date of death if applicable. |
| **ADDITIONAL HEALTH DATA(AHD)** | Includes all vaccinations, laboratory results, patient measurements and information about smoking, and death. |
| **THERAPY DATA** | This includes all issued prescriptions which include the date, the mode of administration, the dosage, quantity and how frequently the prescription is taken. |
| **POST CODE VARIABLE INDICTORS DATA** | This includes the patient's post code linked socio-economic data which is represented in a national quintile for the Townsend Index of Deprivation (this is based on unemployment, car ownership, type of housing and overcrowding), ethnicity and environmental indices.[154] |
| **MEDICAL DATA** | This includes the medical diagnoses and other consultation data, which are entered in the form of Read codes.<br>These are a system of hierarchal codes that specify types of diseases, symptoms, diagnoses or other medical information from consultations or letters. These codes allow some standardisation of diagnosis across medical practices and allow ease of extraction for medical diagnosis for research and audit purposes. However they are not truly hierarchal and have had gaps in coding that have been remedied with annual updates. Read codes (Version 2) are accompanied by the date of the diagnosis and location of diagnosis. [155] |

## 1.9.Thesis Aims and Objectives

Having introduced the thesis and discussed some of the gaps in our knowledge of CKD patient outcomes, this final section of my first chapter states the overall aims and objectives. Subsequent chapters (2-5) include a specific introduction, methods, results and conclusion:

### 1.9.1. Thesis Aims

The overall aim of the thesis was to evaluate the prevalence and management of CKD in UK general practice and identify potentially modifiable risk factors that can be used to inform individual patient care and UK health policy.

### 1.9.2.Thesis Objectives and Organisation

The thesis had a number of specific objectives:

1) **What is the prevalence of CKD in primary care in the UK?** This is described in Chapter 2.
2) **How accurately is CKD recorded in the Quality Outcomes Framework and how does this affect management of care?** This is described in Chapter 3.
3) **What routinely collected primary care data predicts all-cause mortality in stages 3-5 CKD patients?** This is described in Chapter 4.
4) **What routinely collected primary care data predicts the composite outcome of cardiovascular disease or all-cause mortality in stages 3-5 CKD patients?** This is described in Chapter 5.

## 1.10.Executive Summary

- CKD is defined by NKF/KDOQI guidelines
- It is based upon GFR and urine protein excretion
- CKD is associated with high rates of morbidity and mortality
- Estimates of CKD prevalence are variable and will be discussed further in Chapter 2
- Primary care is ideal setting to conduct research
- The THIN database is a useful tool for pharmaco-epidemiological research in primary care.

# CHAPTER 2. THE PREVALENCE OF CKD IN PRIMARY CARE

The purpose of chapter 2 is to address the first objective of the thesis: To provide a robust estimate of the prevalence of CKD in UK primary care. The introduction to this chapter reviews existing prevalence data and considers the strengths and limitations of these estimates as applied to the UK population. The methods for estimating prevalence using the THIN database are described and findings presented. These findings are then discussed in the context of previous prevalence estimates.

## 2.1.The Prevalence of CKD: A review of the existing literature

This section summarises prevalence estimates for CKD in the literature. A systematic review published in 2008 described the prevalence of stages 3-5 CKD from populations greater than 1000 patients.[77] Zhang et al included studies that defined CKD Stages 3-5 CKD were the creatinine clearance (using Cockcroft Gault eGFR) or eGFR (using the non IDMS MDRD equation) was less than 60 ml/min. They found 26 studies meeting inclusion criteria from around the world. Note this study didn't examine the prevalence of CKD stages 1-2 or the prevalence using CKD-EPI formula.

Table 2-1 - Table 2-3 shows prevalence of stages 1-5 CKD by continent and were adapted from this systematic review to include more contemporary studies till 2012, studies including stages 1-2 CKD and studies defining CKD using the CKD-EPI formula. The prevalence according to Cockcroft Gault is not reported as the MDRD equation has superseded Cockcroft Gault in accuracy.[13]

In all studies the prevalence of CKD was higher in women (Table 2-1 - Table 2-3). The studies that only included populations over the age of 40 are underlined. The prevalence was variable between studies and continents. There were no studies from Africa and this may have been due to the exclusion of studies with less than 1000. The majority of patients had stage 3a CKD (data not shown).

In most of the North American studies, patients had been prospectively screened (Table 2-1). The prevalence of CKD stages 1-2 varied from 4.4% to 4.9% [50;52] and CKD stages 3-5 varied from 1.0% to 8.6%.[50;51] CKD prevalence was higher in patients aged over 45 varying from 23.4% - 43.5%.[54;55] However the high prevalence in this age group was due to inclusion of the REGARDS study.[54] This may reflect the inclusion of a relatively high proportion of participants from ethnic minorities.

All the European studies (Table 2-2) were screening studies apart from those in UK which will be discussed later (2.1.1). In contrast to the North American studies, stages 1-2 CKD prevalence varied considerably from 1.6% to 13.2%. Studies at the lower estimation of prevalence were likely due to the use of dipsticks to define proteinuria which as discussed earlier is an insensitive method (Chapter 1.2).[59] Studies with higher estimates were likely due to using an ACR corresponding high albuminuria to define stages CKD 1-2.[65] The prevalence of CKD stages 3-5 varied from 4.7 to 8.1% [61;62] and as described previously in North American studies prevalence was higher in older populations 19.9% to 35.8%.[68;70]

In Australian and Asian studies (Table 2-3) stages 1-2 CKD prevalence varied from 2.4 to 12.1% [71;77] and stages 3-5 CKD prevalence varied from 1.7 to 13.6%.[89] Older populations had higher prevalence of CKD and there was marked variation in CKD prevalence in countries such as China, Thailand and Taiwan.

**Table 2-1. Summary of CKD prevalence studies from North America**

| Paper | Participants | Number | Year | Prevalence using MDRD equation |
|---|---|---|---|---|
| **Kramer et al<br>Dallas Heart Study[50]<br>(2000-2002)** | Prospective study: Community survey in Dallas in unselected patients aged 30-65 | 2660 | 2005 | CKD stages 1-2 Overall 4.9%[1]<br>CKD stages 3-5 Overall 1.0% |
| **Fox et al<br>Framingham Offspring study [51]<br>(1995-1998)** | Prospective Study: Children of the Framingham Heart Study enrolled in 1971. Mean age is 59 | 3047 | 2006 | CKD stages 3-5 Overall 8.6% |
| **Coresh<br>NHANES<br>Between 1988-1994 and 1999-2004[52]** | Prospective study: Patients who were surveyed as part of the survey aged over 20 in USA | 1988-1994<br>15 488<br>1999-2004<br>13 233 | 2007 | 1988-1994<br>CKD 1-2[2] Overall 4.4%<br>CKD 3-4 Overall 5.6%<br>1999-2004<br>CKD 1-2 Overall 5.0%<br>CKD 3-4Overall 8.1% |
| **Shankar et al [53]<br>(1988-1990)** | Prospective study: Wisconsin of 43-86 year olds | 4898 | 2006 | CKD 3-5 Overall 6.6% |
| **McClellan[54]<br>REGARDS study<br>(2000)** | Prospective study : Representative US population aged 45 | 20667 | 2006 | CKD 3-5 -Overall 43.3%<br>Men 38.9% Women 47.5%<br>White 49.9% Black 33.7% |
| **Paper** | Participants | Number | Year | Prevalence using MDRD equation |
| **Manjunath et al[55]** | Prospective study: patients aged over 65 | 4893 | 2003 | CKD 3-5 - Overall 23.4% |

[1] CKD 1-2 defined as microalbuminuria as ACR above 1.9mg/mmol in Men and 2.8mg/mmol in Women

[2] CKD 1-2 defined as microalbuminuria or more i.e. ACR above 3.4mg/mmol. Note this study corrected for persistence. They estimated from earlier studies that albuminuria was persistent in 51% if GFR was above 90 mmol/l/1.73m$^2$ and if 75% if GFR between 60 -89 mmol/l/1.73m$^2$

| | | | | |
|---|---|---|---|---|
| **Cardiovascular Health Study between 1989 and 1990** | | | | Men 25% Women 22.2% |
| **Garg et al [56] Between 2001-2002** | Retrospective study: Elderly patients in long term facilities aged over 65 in Ontaria, Canada | 9931 | 2004 | CKD stage 4 - Overall 35.7% Men 27.1% Women 38.8% |
| **Hemmelgarn et al [57] Between 2001-2003** | Retrospective study: Community based cohort in Ontaria, Canada in patients aged over 66 | 10184 | 2006 | CKD stages 3-5 - Overall 35.4% Men 32% Women 38.2% |

### 2.1.1.The prevalence of CKD in the UK

In the United Kingdom and Ireland (in italics in Table 2-2), with the exception of Health Survey for England study (HSE), included studies using retrospective databases.[49;63;66;67;69;133] Only the HSE used screening to establish the prevalence of stages 1-2 CKD (7.5%). Stage 3-5 CKD prevalence varied from 4.9 to 8.5%[66;133], and was more common in older individuals.[69] Only one study from the UK (shaded in grey in Table 2-2), partially attempted to correct for chronicity of CKD and this estimated the prevalence of stages 3-5 as 6.8% (using the MDRD equation).[49] This study used laboratory blood results extracted from primary care databases from various centres from the UK. Individuals were identified as having stages 3-5 CKD if they two blood results 3 months apart and had eGFR (calculated or laboratory) persistently below 60ml/min/1.73m$^2$ as per the KDOQI guideline. However individuals would also be labelled as having stages 3-5 CKD if they had lab eGFR or calculated eGFR, on the basis of a single blood result, below 60ml/min/1.73m$^2$.

In studies reporting prevalence of stages 3-5 CKD using both the MDRD and CKD EPI formula, the prevalence was lower using the CKD EPI formula.[49;66;67;80;87] This was likely due to that equation resulting in the eGFR being higher.

### 2.1.2.The variation in prevalence

The prevalence between studies and countries is highly variable. For stages 1-2 CKD this is likely due to whether urine dipstick or ACR was used to determine proteinuria. Urine dip associated estimates where likely to be lower as this method is insensitive at determining proteinuria. Prevalence will obviously be higher in groups where patients are older due to aging being related to reduced eGFR. This is illustrated by the Dallas

study where the patients were under 65 and the prevalence was only 1%.[50] CKD prevalence could also vary between countries especially for the MDRD equation which hasn't been validated outside the United States.[13] Additionally variability in the United States could be due to sampling groups with more ethnic minorities.[54]

**Table 2-2. Summary of CKD prevalence studies from Europe**

| Paper | Participants | Number | Year | Prevalence - MDRD | Prevalence – CKD EPI |
|---|---|---|---|---|---|
| **De Zeeuw et al [58]<br>PREVEND stud<br>Between 1997-1998** | A prospective study: General population screening study in Groningen, Netherlands | 8459 | 2005 | CKD 1-2 Overall 11.8%<br>CKD 3-5 Overall 5.7% | N/A |
| **Viktorsdottir et al<br>Reykjavik Heart Study[59]<br>Betweem 1967-1996** | Prospective study: of aged 33-85 from Reykjavik 1967-1996 | 19256 | 2005 | CKD1-2[3] Overall 1.6%<br>Men 2.39%, Women 0.89%<br>CKD 3-5 Overall 7.2%<br>Men 3.7%, Women 10.9% | N/A |
| **Cirillo et al<br>Gubbio studies [60]** | Prospective study: Population based study in 18-95 year olds in Central Italy | 4574 | 2006 | CKD 3-5 Overall 6.4%<br>Men 6.6% Woman 6.2% | N/A |
| **Nitsch al[61]<br>SALPADIA study 2002** | Prospective study: aged over 18 in Switzerland in 2002-2003 | 6317 | 2006 | CKD 3-5 Overall 8.1%<br>Men 4.5%, Women 11.5% | N/A |
| **Hallan et al<br>HUNT 2[62]<br>Between 1995-1997** | Prospective study: Population screening study<br>Aged over 20 in Norway | 65181 | 2006 | CKD 1-2[4] - Overall 5.9%<br>CKD 3-5 - Overall 4.7%<br>Men 3.6%, Women 5.7%<br>Over 70 - 18.6% | N/A |
| ***Stevens et al[8]<br>NEOERICA study<br>Between 1998- 2003*** | *Retrospective cross-sectional study from UK GP databases in Kent, Manchester and Surrey* | *130226* | *2007* | *CKD 3-5 Overall 8.5%<br>Women 10.8% Men 5.8%* | *N/A* |

[3] CKD 1-2 defined as 1+ on Albustix

[4] CKD 1-2 defined as persistent albuminuria i.e. ACR above 1.9mg/mmol and 2.8mg/mmol in men and women respectively on 2 out of 3 samples

| Paper | Participants | Number | Year | Prevalence - MDRD | Prevalence – CKD EPI |
|---|---|---|---|---|---|
| ***Health Survey of England 2009 [63]*** | *Prospective General Health screening survey* | *3261* | *2010* | *CKD 1-2 Overall: 7.5% Men 9%; Women 6% CKD 3-5 Overall 6.0% Men 5%; Women 7%* | *N/A* |
| **Vinhas et al[64] PREVADIAB study Between 2008-2009** | Prospective study:Nationally represented sample of people aged over 18 in Portugal | 5167 | 2011 | CKD 3-5 Overall 6.1% | N/A |
| **Suleymanlar et al [65] CREDIT study** | Prospective study: Screening study in Turkey in over 18 | 10748 | 2011 | CKD 1-2[5] Overall 10.6% CKD 3-5 Overall 4.7% | N/A |
| ***Carter et al[66] Between 2009-2010*** | *Retrospective study: primary care patients from East Kent, UK* | *174448* | *2011* | *CKD 3-5 Overall 4.9%* | *CKD 3-5 Overall 4.4%* |
| ***Gifford et al[67] 2004 and 2009*** | *Retrospective study: CKD prevalence extracted from Laboratory database in Scotland. Lowest creatinine used.* | *293880* | *2011* | *CKD 3-5 Overall 5.6%* | *CKD 3-5 Overall 4.94%* |
| ***De Lusignan et al[49] QICKD study 2009*** | *Retrospective cross-sectional study from UK GP databases from five areas* | *744 216* | *2011* | *CKD 3-5 Overall 6.76%* | *CKD 3-5 Overall - 6.0%* |
| **Wasen et al [68]** | Prospective study: Cross-sectional survey aged 64-100 from Finland | 1246 | 2004 | CKD 3-5 Overall 35.8% | N/A |
| ***Glynn et al[69]*** | *Retrospective study Cross-sectional study primary care aged over 50* | *2602* | *2009* | *CKD 3-5 Overall 16.7%* | *N/A* |
| **Gambaro et al[70] INCIPE study** | Prospective study: Randomly selected patients from North Eastern Italy screened for CKD aged over 40 in 2006 | 6200 | 2010 | CKD 1- 2 Overall 13.2% CKD 3-4 Overall 6.7% | N/A |

[5] CKD 1-2 defined as patients with ACR above 3.4mg/mmol or 24 urine collection of albumin 30mg/day

**Table 2-3. Comparison of CKD prevalence studies from Australia and Asia**

| Paper | Participants | Number | Year | Prevalence - MDRD | Prevalence - CKD EPI |
|---|---|---|---|---|---|
| **Chadban et al Ausdiab[71] (1999-2000)** | Prospective study: Cross-sectional study of Australian population aged over 25 | 11247 | 2004 | CKD 1-2[6] Overall 2.4%<br>CKD 3-5 Overall 11.2%<br>Men 9.3% Women 13.0% | N/A |
| **Domrongkitchaiporn et al[72] (1985-1997)** | Prospective study: Cross-sectional studies employees in Thailand aged 35-55 | 3499 | 2005 | CKD 3-5 Overall 6.8% | N/A |
| **Chen et al InterAsia[73] (2000-2001)** | Prospective study: Cross-sectional study of Chinese patients in aged 35-74 | 15540 | 2005 | CKD 3-5 Overall 2.5%<br>Men 1.3%, Women 3.8% | N/A |
| **Wen et al[74] (1994-2006)** | Prospective study: Cross-sectional study of participants taking part of a screening program in Taiwan of aged 20 or over | 462294 | 2005 | CKD 1-2 Overall 4.81%[7]<br>CKD 3-5 Overall 7.12% | N/A |
| **Hsu et al[75] (2001)** | Prospective cross-sectional study of a population screened in Taiwan in 2001 | 6001 | 2006 | CKD 3-5 Overall 6.9% | N/A |
| **Perkovic et al [156] (2000)** | Prospective study: InterAsia screening program in Thailand aged over 35 | 7909 | 2007 | CKD 3-4 Overall 13.6% | N/A |
| **Kuo et al [76] (1995)** | Retrospective study: Cross-sectional study of Patients from national health insurance enrollees in Taiwan | 200000 | 2007 | Coded CKD 1-5 - 9.83% | N/A |

[6] CKD 1-2 defined as ACR greater than above 2.5mg/mmol in Men and 3.5mg/mmol in Women

[7] CKD 1-2 defined as microalbuminuria as trace on dipstick. . Note this study corrected for persistence. They estimated from earlier studies that albuminuria was persistent in 51% if GFR was above 90 mmol/l/1.73m$^2$ and if 75% if GFR between 60 -89 mmol/l/1.73m$^2$

| Paper | Participants | Number | Year | Prevalence - MDRD | Prevalence CKD EPI |
|---|---|---|---|---|---|
| **Zhang et al[77]** | Prospective study: Cross Study of 13925 adults in Beijing | 13925 | 2008 | CKD 1-2 Overall 12.1%[8]<br>CKD 3-5 Overall 1.8% | N/A |
| **Chen et al[78] (2006)** | Prospective study: Cross-sectional study from Shanghai | 2596 | 2009 | CKD 1-2 Overall 6.3%[9]<br>CKD 3-5 Overall 5.8% | N/A |
| **Ong –Ajyooth et al[79] (2004)** | Prospective study: Cross sectional survey of over 15 in Thailand | 3117 | 2009 | CKD 3-5 Overall 8.45% | N/A |
| **Varma et al[80] (2008-2009)** | Prospective study: Indian government employees aged over 18 screened | 3398 | 2010 | CKD 1-2[10] Overall 10.0%<br>CKD 3 Overall 3.0% | CKD 1-2 Overall 10.1%<br>CKD 3 Overall 2.1% |
| **Ingsathit et al[81] SEEK study (2007-2008)** | Prospective study: Cross-sectional study of Screened population in Thailand | 3459 | 2010 | CKD 1-2 Overall 6.2%<br>CKD 3-4 Overall 8.2% | N/A |
| **Sabanayagam et al [82]** | Prospective study: Screening study of patients aged 24-95 in Singapore of a multi – ethnic cohort | 4499 | 2010 | CKD 1-2 Overall 10%<br>CKD 3-5 Overall 5.5%<br>11.4% in Chinese<br>18.6% in Malays<br>17.6 in Indians | N/A |
| **Jiang et al[83] (2006-2007)** | Prospective study: Cross-sectional study of Rural Chinese population over 30 | 5105 | 2010 | CKD 1-5 Overall 15.2% | N/A |

[8] CKD 1-2 defined as microalbuminuria as ACR above 1.9mg/mmol in Men and 2.8mg/mmol in Women. Note this study corrected for persistence. They estimated from earlier studies that albuminuria was persistent in 51% if GFR was above 90 mmol/l/1.73m$^2$ and if 75% if GFR between 60 -89 mmol/l/1.73m$^2$

[9] CKD 1-2 defined as those with positive micro-albumin creatinine dipstick and then subsequent ACR greater than above 2.5mg/mmol in Men and 3.5mg/mmol in Women

[10] CKD 1-2 defined as renal damage as either Microalbuminuria (ACR of 3.4mg/mmol) on dipstick test, haematuria and leucocyturia.

| Paper | Participants | Number | Year | Prevalence - MDRD | Prevalence CKD EPI |
|---|---|---|---|---|---|
| **Jang et al[84] KHANES study (2005)** | Prospective Screening Study in Korea aged over 20 | 5136 | 2010 | CKD 3-5 Overall 6.8% | N/A |
| **Lee et al [85] (2007)** | Prospective study: Screening study in Korea aged over 20 | 8400 | 2010 | CKD 3-5 Overall 7.2% | N/A |
| **Sharma et al[86] (2007)** | Prospective study in over 18: Cross sectional study in Nepal(n=8398), Mongolia(n=997) and China(n=1999) | 11394 | 2010 | CKD 3-5 Overall 7.3% | N/A |
| **White et al[87] AUSDIAB (1999-2000)** | Prospective study: Cross sectional study of Australian aged over 25 | 11427 | 2010 | CKD 1-2 Overall 5.6%<br>CKD 3-5 Overall 7.8% | CKD 1-2 Overall 5.7%<br>CKD 3-5 Overall 5.8% |
| **Cepoi et al [88] (2007-2008)** | Prospective study: Romanian cross-sectional health screening study | 60969 | 2010 | CKD 3-5 Overall 6.7% | CKD 3-5 Overall 7.3% |
| **Zhang et al[89] (2007-2010)** | Prospective study: Screening study of nationally representative Chinese adults | 47204 | 2012 | CKD 1-2 Overall 9.1%<br>CKD 3-5 Overall 1.7% | N/A |
| **Li et al[90]** | Chinese prospective study: Urban Cross-sectional survey in patients over 40 | 2310 | 2005 | CKD 1-2 Overall 8.4%<br>CKD 3-5 Overall 4.9%<br>Men 4.8% Women 5.0% | N/A |
| **Nimomya [91] (1988)** | Japanese prospective study: of patients aged 40 with no cardiovascular disease | 2634 | 2005 | CKD 3-5 Overall 10.3%<br>Men 5.3% Women 6.2% | N/A |

## 2.2.Research Questions: What is the prevalence of CKD in primary care?

In section 2.1 the prevalence of CKD varied from 1%-14% for stages 3-5 and 2-12% for stages 1-2 CKD. Stages 1-2 CKD are defined generally using urine protein excretion and defining stages CKD 1-2 is difficult in primary care as very few patients are likely to have a urine ACR measured but many patients have urine dipsticks for proteinuria. There was very limited evidence on the prevalence 'true' of stages CKD 3-5 based on two tests of renal function. Additionally the new CKD-EPI formula may be more accurate in determining eGFR but the prevalence changes from when the GFR is estimated by the MDRD equation.(Chapter 1.2)

As CKD is major risk factor for cardiovascular disease and mortality in the population and associated with high economic burden (Chapter 1.5),it is important to determine the accurate prevalence of CKD using two blood results (to account for variation in GFR measurement) to allow these high risks groups to be targeted for intervention and predict resource requirement for this population. The latter has implications for whole NHS budget. Estimates from a large scale cohort such as THIN could allow the burden of CKD to be truly ascertained.

This chapter will address the first objective of the thesis to assess the prevalence of CKD stages 1-5 in UK general practice using the THIN database.

Specifically the chapter will address the following research questions:

1. *What is the prevalence of stages 3-5 CKD from two blood results seven days apart? Stages 3-5 CKD will be defined by either two lab eGFRs or two calculated (using the non IDMS formula) eGFRs (using the non IDMS formula) below 60 ml/min/1.73m$^2$.*
2. *Does the prevalence of CKD vary using different equations for eGFR calculated by the serum creatinine and laboratory reported eGFR? This is the secondary analysis.*
3. *What is the prevalence of stages 3-5 CKD using a single blood result? This will be the sensitivity analysis.*
4. *Can urine dip data and albumin creatinine ratio be used to define the prevalence of CKD stages 1 and 2?*

## 2.3.Methods

This section highlights the methods used to: 1) prepare the database for analysis and examine the raw data; 2) compare laboratory eGFR and calculated eGFR; 3) ascertain the prevalence of CKD stages 3-5 using a single or two blood results; 4) compare urine ACR to urine dipstick and to determine the prevalence of stages 1-2 CKD.

### 2.3.1.Preparing the database: first dataset

All analyses were undertaken in SAS statistical software (SAS v9.2, SAS Institute, Cary, NC, USA) and R 2.15.1 software version (http://www.r-project.org/). The following data were received in a THIN data file containing data until November 2009:

1. Clinical Data
    a. Serum creatinine data (numerical variable)
    b. Laboratory reported estimated glomerular filtration rates (numerical variable)
    c. Urine albumin creatinine ratios (ACR) (numerical variable)
    d. Urine micro albumin data (numerical variable)
    e. Urine dip for protein data (categorical variable)
2. Patient File – This contained the patient demographics such as year of birth and gender and registration status
3. AMR File – This contained the date the practice attained Acceptable mortality reporting indicating that data entry was accurate.[150]
4. Midyear Counts. These consisted of the total number of patients split by age and for all the years since the creation of THIN.

#### *2.3.1.1.Preparing and examining the dataset*

A summary of the steps taken to prepare the database for analysis is shown in Table 2-4. Histograms were constructed for numerical variables using the SAS procedure (Proc Univariate). These included plots of creatinine; laboratory reported eGFR, urine dip data, ACR, weight, height and urine microalbumin data.

Table 2-4. Preparation of database to determine CKD prevalence

| **Database preparation** | **Method and Rationale** |
|---|---|
| **Removal of entries with missing event date** | |
| **Any additional Health Data with null values removed** | Data with null values coded as 0, 0000000, and 0.000000 were removed. 0 values are removed as cannot be distinguished from nonsense values |
| **Dataset merged with patient file and then AMR file** | |
| **Any patients with AHD values occurring before the age of 18, 6 months prior to registration and prior to AMR date removed** | Data, 6 months pre registration is removed as when a patient is newly registered there is a lag between the arrival of their notes from their last primary care provider.[148]<br>Data removed pre AMR date is discussed before. [150] |
| **Only patients with permanent or temporary registration retained** | |
| **Data split into constituent variables files** | |

### 2.3.2.Examining renal function

#### *2.3.2.1.Calculating eGFR: Does laboratory eGFR correspond with calculated eGFR?*

The MDRD (both Non IDMS and IDMS formula), the CKD EPI and Mayo quadratic formula were used to calculate the eGFR as detailed in Box 1-3-Box 1-5. Lab eGFR was compared to the non IDMS calculated eGFR using Bland-Altman plots. (This will be relevant in Chapter 4 to 5 where only calculated eGFRs were used to determine CKD) There was considerable variability in the method by which creatinine was measured across the study period; scatter plots and Bland Altman plots were used to compare the consistency of calculated and lab eGFRs in patients with creatinine and lab based eGFR data recorded on the same day.[109] These comparisons were made only when the eGFR was below 60 as lab eGFRs are only likely to be reported directly below this threshold and many eGFRs would have been reported as 'above 60'.[31] Additionally plots of CKD EPI versus the non MDRD equation were drawn.

### 2.3.3.The prevalence of stages 3-5 Chronic Kidney Disease

#### *2.3.3.1. The principal analysis*

The primary question is what the is prevalence of stages 3-5 CKD? There is considerable variability in eGFR (Chapter 1.2) and therefore all the guidelines state that patients should have a persistent reduction in eGFR over 3 months. In the QICKD study the prevalence was derived from patients with one or two blood test results and the prevalence may not reflect the actual prevalence and therefore the true prevalence of CKD needs to be ascertained. The principal analysis was to identify prevalence using two consecutive blood results.

In the UK, the Quality and Outcomes Framework defines stages 3-5 CKD as two eGFRs below 60mls/min/m2 *within* 3 months and two blood results seven days apart were taken as a pragmatic definition to capture all those with two blood results as it was likely that much less of the population would have two blood tests 3 months apart and there is good agreement between two blood results.[49;157]

The prevalence of CKD was ascertained for the years 2005-2009. For the year in question, the population included for this analysis had to have the following: be alive and registered prior to July; two consecutive laboratory eGFRs seven days apart or if these were not available two consecutive calculated eGFRs (using the non IDMS equation in Box 1-3) from serum creatinine. The proportion of patients with stages 3a, 3b, 4 and 5, in the year in question, were determined using the higher eGFR from the two blood results.

#### *2.3.3.2.Secondary analysis*

In the secondary analysis the prevalence was ascertained using two calculated eGFRs seven days apart where the eGFR was calculated using the non IDMS MDRD equation, IDMS MDRD equation, CKD-EPI equation and the Mayo Quadratic equation (Box 1-3-Box 1-5). This prevalence was also extracted from those with two consecutive laboratory reported eGFRS.

A separate sensitivity analysis was carried out using the single blood test results to determine prevalence using either calculated eGFR or laboratory eGFR from the latest blood result before July in each year in question and in addition the observed time difference between blood results was calculated.

As the THIN population is older than the general population, it was necessary to age and gender standardise the prevalence.[144] To calculate this, for all CKD stages for each formula, the patients were split by gender and age and these groups were weighted using United Kingdom Office for National statistics population estimates for each year in question (Box 2-1).[135] This is known as direct standardisation.[158] Following this, 95% confidence intervals for proportions were constructed (Box 2-1).[158]

**Box 2-1. Crude Prevalence, Age and Gender Standardisation, 95% Confidence Intervals**

Calculating crude prevalence

$$Crude\ prevalence = \frac{total\ number\ of\ cases\ at\ a\ given\ time\ i.e\ patients\ with\ CKD}{total\ population\ at\ that\ time}$$

Age and gender standardisation

$$\rho^{l} = \frac{\sum N_i\, p_i}{N_i}$$

Where $\rho^{l}$ is age and gender standardised prevalence, N is the population (in this case the Office of National Statistics yearly age and gender estimates) for the age group i p is the prevalence of the condition in the cohort in question in age group i.

Calculating 95% Confidence intervals for proportions e.g. prevalence rates

Lower Confidence interval = $p - 1.96 \times \sqrt{\frac{p\,(1-p)}{N}}$

Upper Confidence interval = $p + 1.96 \times \sqrt{\frac{p\,(1-p)}{N}}$

Where p is the proportion and N is the population.

### 2.3.4.Defining Proteinuria

#### *2.3.4.1.Urine dip distribution and comparison with ACR in detecting microalbuminuria*

High albuminuria was used as a definition point over very high albuminuria, as even small amounts of albuminuria are predictive of cardiovascular risk and progression of renal disease (Chapter 1.5).[93] In primary care, patients are more likely to have urine dip data to estimate urine protein.[33] This is a semi-quantitative method that has poor negative predictive value but reasonable positive predictive value to detect microalbuminuria.[42] For the purpose of using urine dip data in the cardiovascular disease or all cause mortality model (Chapter 4 & 5), how adequately the urine dip result predicted microalbuminuria was determined. Urine dip data were transformed to the universal scale of nil, trace, 1+, 2+, 3+ and 4+ and frequency distribution of this was plotted in a bar graph. The ACR file was merged with the urine dip file where both ACR and urine dip occurred on the same event date. A box and whisker plot was constructed to examine a trend between urine dip data and ACR spread as well as Receiver Operating Characteristic curve.[109]

### 2.3.5.Prevalence of CKD stages 1-2

Stages 1-2 CKD are defined by evidence of kidney damage (Chapter 1.3). This can either be the presence of persistent albuminuria and or persistent haematuria, or presence of structural kidney disease such as Adult Polycystic Kidney Disease (APKD).[1] As haematuria and evidence of APKD were not available, CKD 1-2 was staged using only urine protein excretion. The KDIGO definition of stages 1-2 CKD was used i.e. persistent "high" albuminuria (A urine ACR of 3 or above). This definition was used as even small amounts are associated with mortality.[93] By combining urine dip values that corresponded to high albuminuria (Table 1-2) and the standard definitions for high albuminuria from ACR, patients were coded as having high albuminuria or not. As with eGFR, albuminuria had to be persistent as it can be transient or due to measurement error and therefore needs confirmation on second measurement.[33] Patients with "high" albuminuria on two occasions at least one week apart, who were still registered by the 1st of July in the years 2005-2009 were included in the prevalence calculations. The crude prevalence was calculated for each year as detailed above (Box 2-1). Prevalence of stages 1-2 CKD were then age and gender standardised as above. Following this 95% confidence intervals for proportions were constructed.[158]

## 2.4.Results

The results for prevalence of CKD in UK general practice are split into two sections:1) The prevalence of CKD stages 3-5 and 2) the prevalence of CKD stages 1-2.

In the first section, the raw data used to identify stages CKD 3-5 are presented. Initially this comprises laboratory reported eGFR (lab eGFR). The section then further explores whether investigator calculated eGFR, using reported serum creatinine and the non IDMS MDRD equation, was comparable to laboratory reported eGFR. Furthermore, how the different methods of calculating eGFR such as the Mayo Quadratic or CKD EPI formula compared to the MDRD eGFR. This section then describes the prevalence of CKD 3-5 using the principal analysis, the secondary analysis and the sensitivity analysis. The 2$^{nd}$ section details the distribution of ACR and how accurate Urine Dip determines albuminuria. The prevalence of CKD stages 1-2 is then reported.

### 2.4.1. Stages 3-5 CKD: Identifying CKD in the cohort and Raw data distribution

Patients with CKD 3- 5 were identified using either creatinine data or laboratory eGFRs.

The baseline cohort consisted of 6 581 419 patients.

- In total 1 982 568 patients (54.7% female) had a reported serum creatinine with a mean age of 62.9 (± 16.4) years.
- 1 296 783 patients had a reported laboratory eGFR. The mean age was 62.9 (± 16.9) years and 55.0% were female.
- There were 688 518 patients who had a serum creatinine but no lab eGFR.

The bar histograms for creatinine and laboratory eGFR are shown in Figure 2-1 to Figure 2-4. The medium serum creatinine was 87 (range 40 to 1000) mmol/litre and serum creatinine was slightly positively skewed. The histogram for calculated eGFR using the non IDMS MDRD equation showed a slight positive skew approaching a normal distribution (Figure 2-2) and the median calculated eGFR was 71.0 (range 3.4 to 254.0) mmol/litre/1.73m$^2$. The mean lab eGFR was 70.5 (Standard deviation ± 93.8) mmol/litre/1.73m$^2$ and lab eGFR appeared normally distributed but there were large peaks where eGFR was equal to 60 and 90 ml/litre/1.73m$^2$. These two peaks accounted for 30% of the data with the larger peak at 60 ml/litre/1.73m$^2$. Laboratories reported eGFRs at either the 60 or 90 ml/litre/1.73m$^2$ thresholds rather than actual values above these thresholds (Figure 2-3)

**Figure 2-1. Distribution of Serum creatinine**

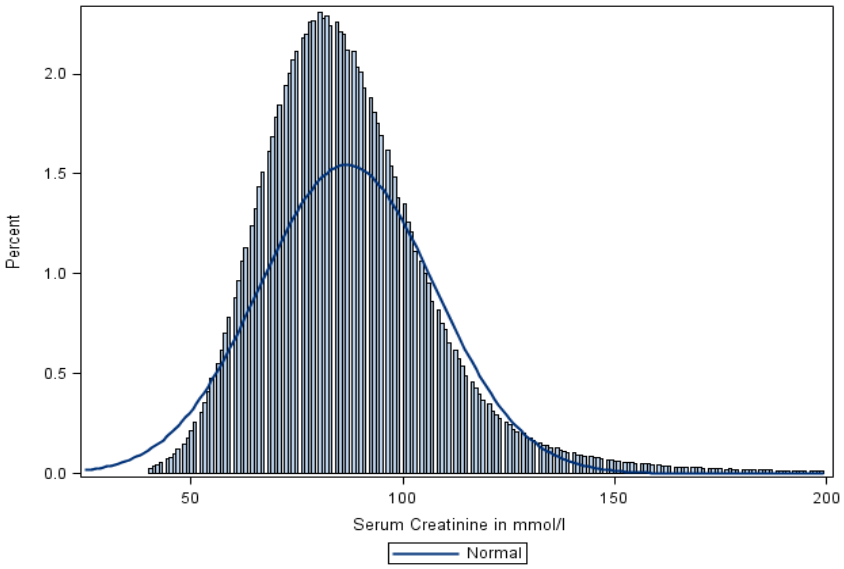



**Figure 2-2. Distribution of calculated eGFR**

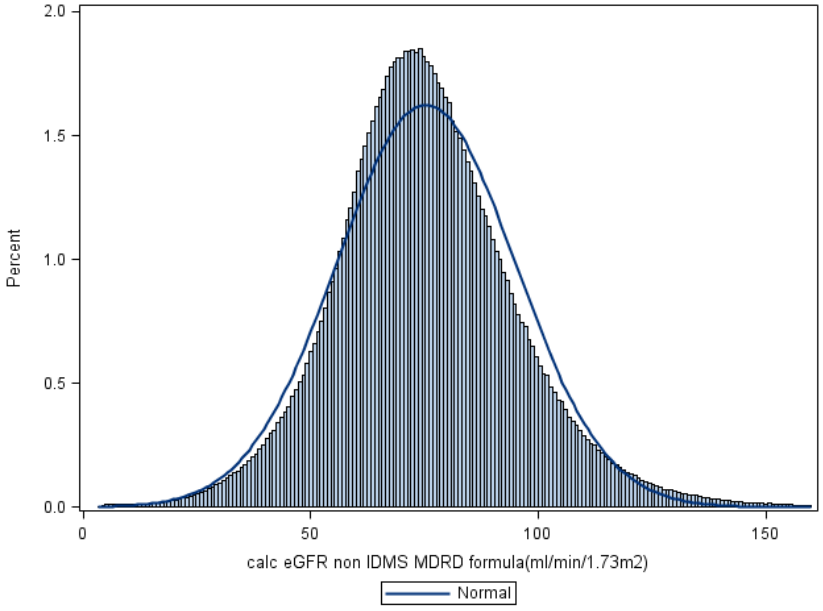


**Figure 2-3. Histogram distribution of laboratory reported MDRD eGFR**

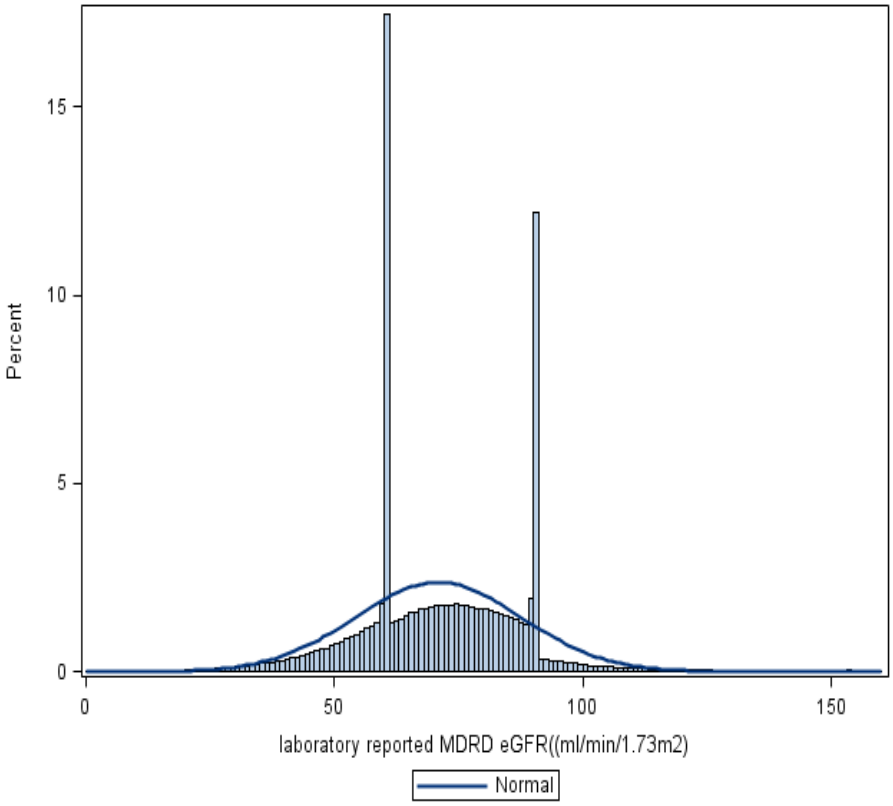
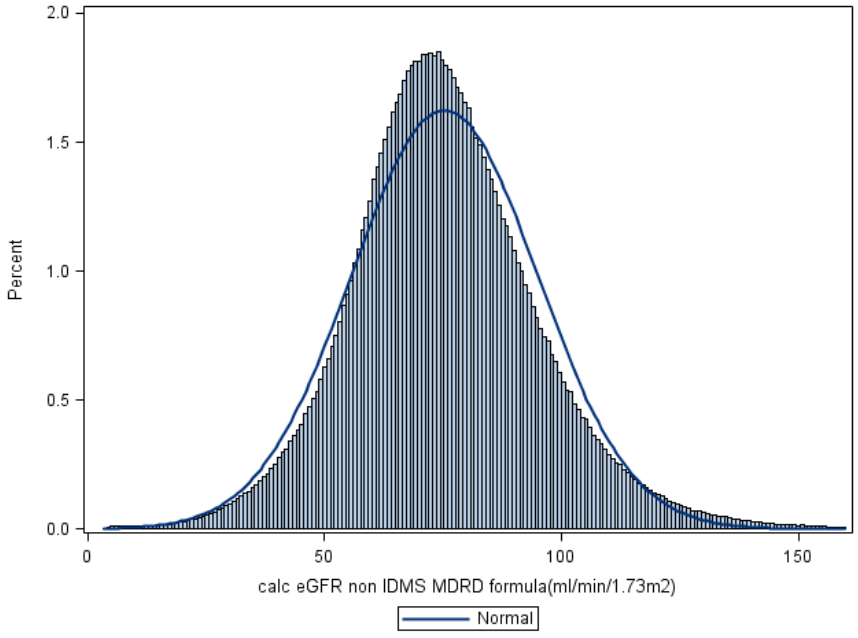



**Figure 2-4. Histogram distribution of eGFR calculated by MDRD equation**



#### *2.4.1.1.Stages 3-5 CKD: Comparing different equations*

##### 2.4.1.1.1.Comparing eGFRS: Laboratory eGFR versus non IDMS MDRD calculated eGFR

In patients with both a serum creatinine and lab eGFR reported on the same blood results, scatter plots of lab eGFR showed a correlation between lab eGFR and calc eGFR (using the non IDMS equation) but there were vertical lines at where the lab eGFR equaled 15, 60, 90, 120, 150 and 180 ml/litre/1.73m$^2$ (Figure 2-5). This was due to laboratories reporting the eGFR thresholds as an absolute value; for example if eGFR was above 60 ml/litre/1.73m$^2$ it was reported as 60 ml/litre/1.73m$^2$. In a plot where lab eGFR and calc eGFR were less than 60 ml/litre/1.73m$^2$ the relationship again appeared linear (Figure 2-6). In a Bland-Altman plot (Figure 2-7), comparing calc eGFR versus lab eGFR, where both values were under 60 ml/min/1.73m$^2$, the majority of data points lay within two standard deviations of the difference. There was a diagonal line emerging where the mean was 15 ml/litre/1.73m$^2$and this was likely to represent where an eGFR of 15 ml/litre/1.73m$^2$or below was reported at 15 ml/litre/1.73m$^2$ by laboratories.

**Figure 2-5. Laboratory MDRD eGFR versus Calculated EGFR using non IDMS MDRD equation**

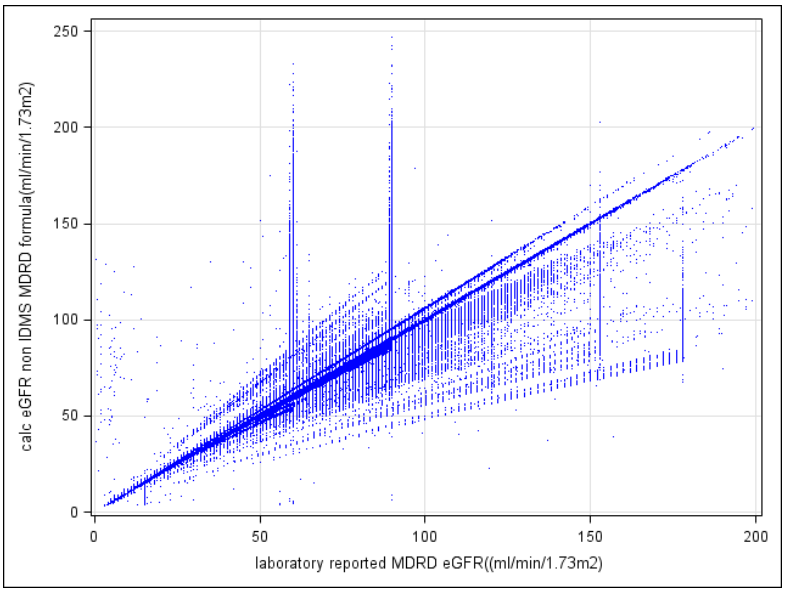



**Figure 2-6. Laboratory MDRD eGFR versus Calculated EGFR using non IDMS MDRD equation where both below 60 ml/min/1.73m$^2$**

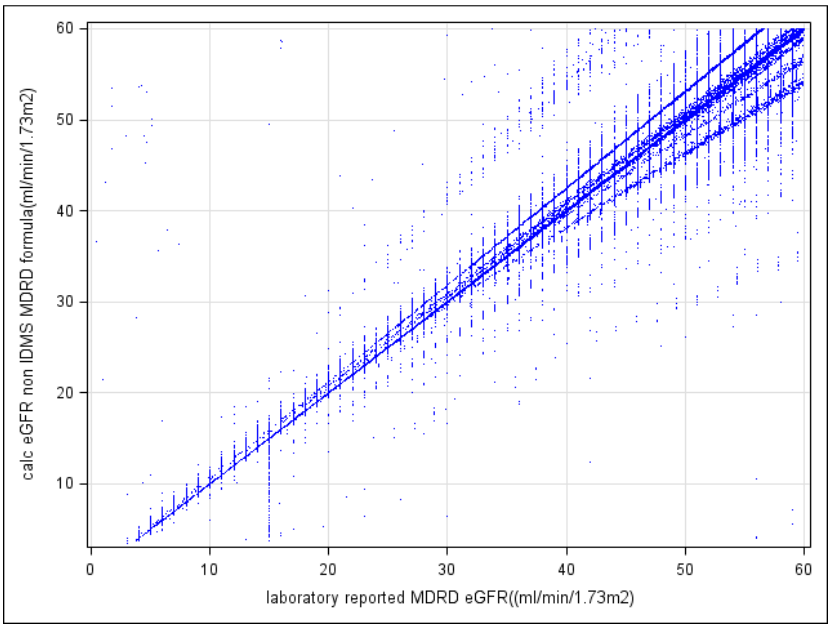
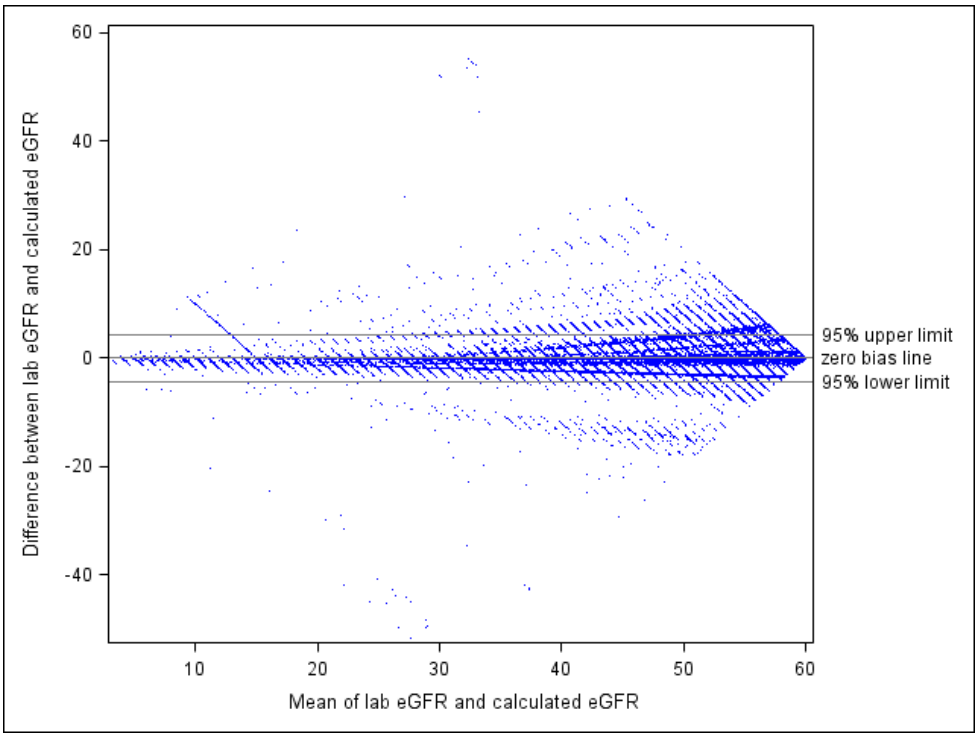


**Figure 2-7. Bland Altman Plot for laboratory eGFR and MDRD calculated eGFR (Non IDMS)**




#### 2.4.1.1.2.Comparing eGFRS: Non IDMS MDRD versus CKD EPI and Mayo Quadratic eGFRS

When comparing the different equations for calculating eGFR (Box 3-6 in Methods), the MDRD eGFR (non IDMS equation) was compared to Mayo Quadratic eGFR and CKD EPI eGFR ( CKD EPI eGFR and the MDRD eGFR. (Figure 2-8 to Figure 2-9) in scatter plots split by gender. The Mayo plots initially showed a positive linear relationship between Mayo eGFR and MDRD eGFR. However the Mayo eGFR was higher at the same MDRD eGFR. For example when the MDRD eGFR was 70 ml/min/1.73m$^2$ the Mayo eGFR was 100 ml/min/1.73m$^2$ in men. Above a MDRD eGFR of 75 ml/min/1.73m$^2$ in men and 60 ml/min/1.73m$^2$ in women, the Mayo eGFR increased less sharply. There was a similar relationship between the CKD EPI eGFR and the MDRD eGFR.

**Figure 2-8. Calculated MDRD (non IDMS) eGFR versus Mayo Quadratic and CKD EPI derived eGFRs in Men**

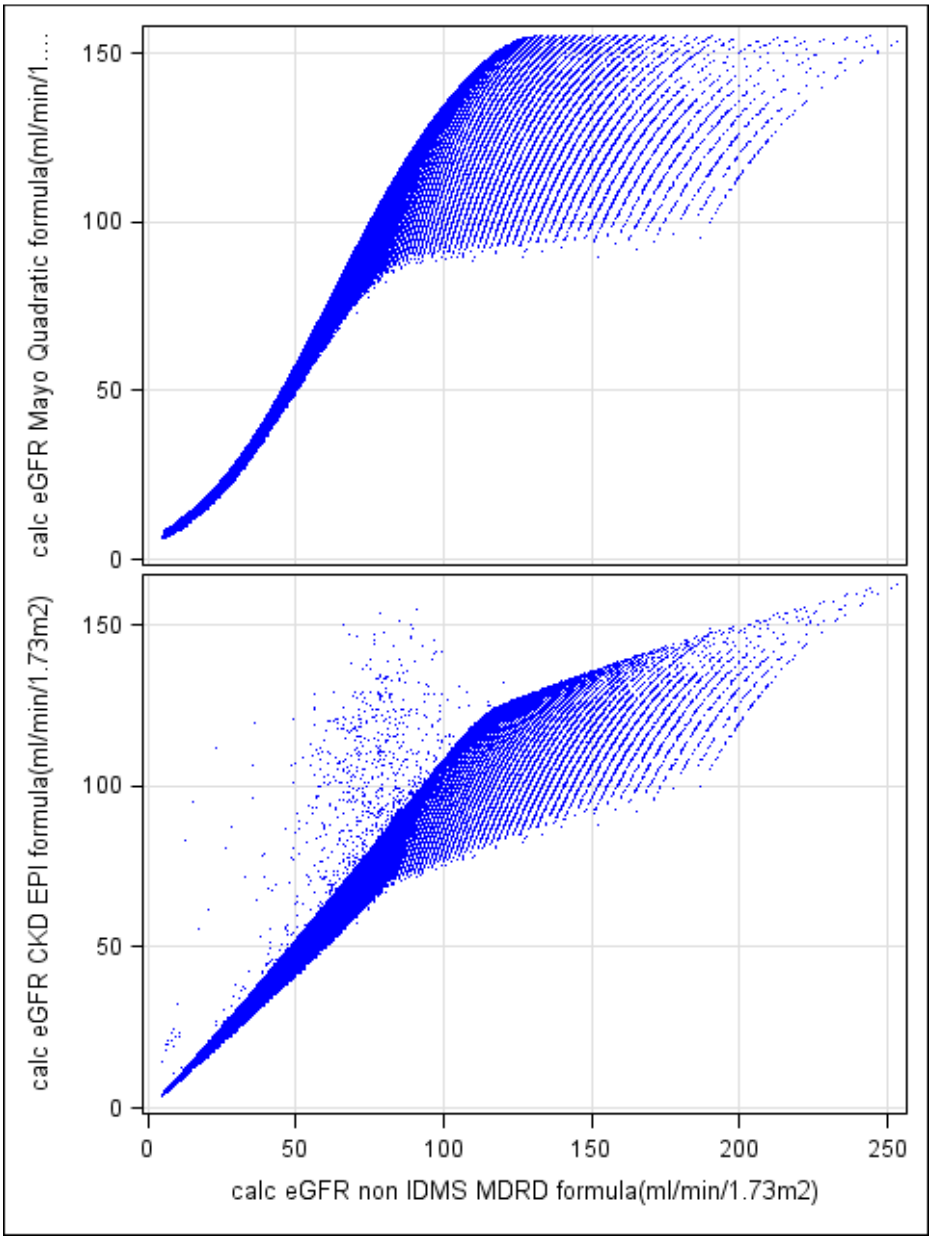
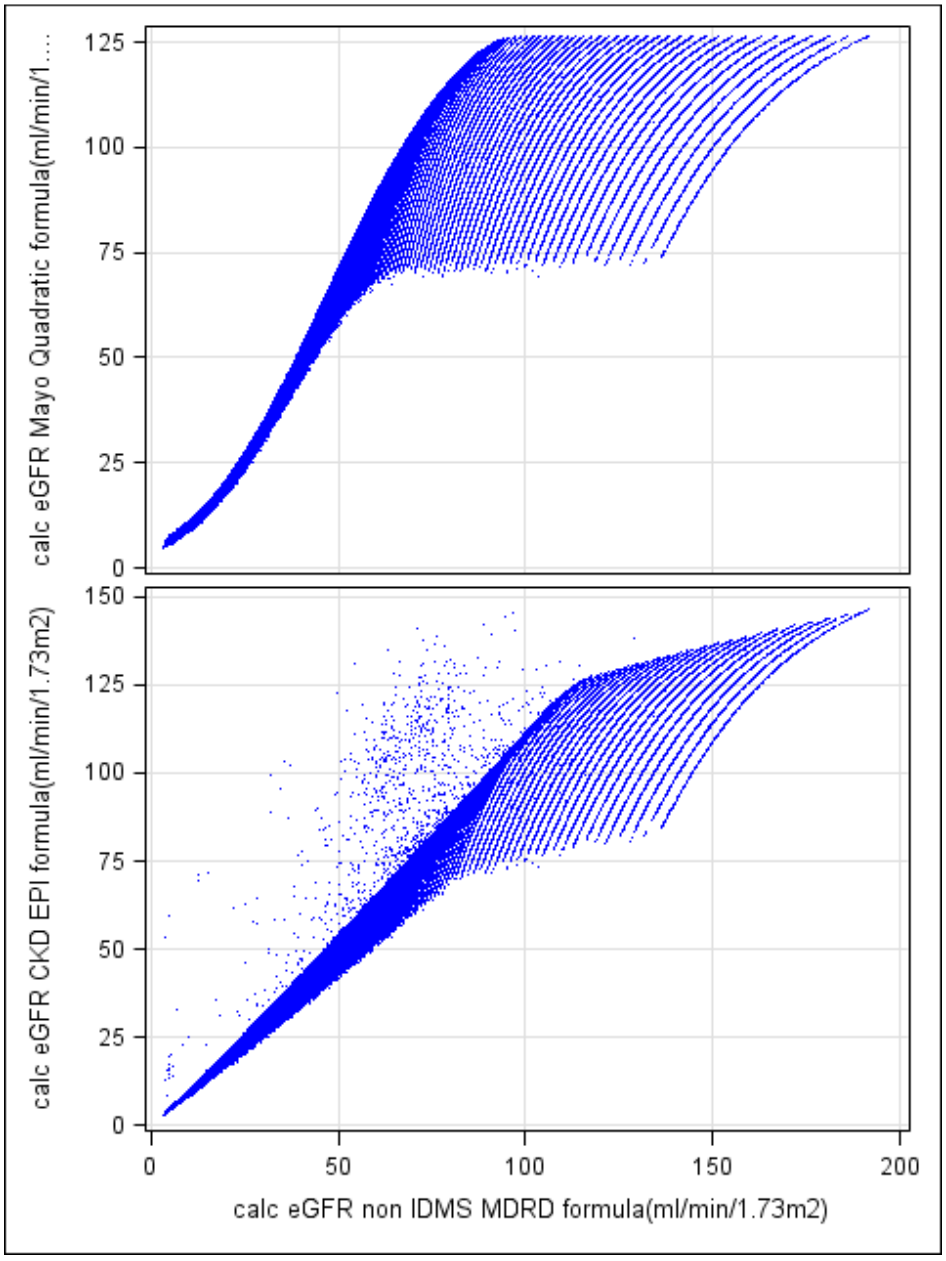


**Figure 2-9. Calculated MDRD (non IDMS) eGFR versus Mayo Quadratic and CKD EPI derived eGFRs in Women**



### 2.4.2.Prevalence of stages 3-5 CKD using different equations and definitions

The following section details how the prevalence varies when Stages 3-5 CKD are staged using:

- Principal analysis – Prevalence ascertained from two blood results - either lab or calculated eGFR (if no eGFR reported) using the Non IDMS MDRD equation
- Secondary analysis – Prevalence ascertained from two blood results
    - Two consecutive calculated eGFRs using Non IDMS MDRD, IDMS MDRD equation, CKD EPI equation and Mayo Quadratic equation
    - Two consecutive laboratory eGFRS

#### 2.4.2.1.1.: Principal analysis: CKD staged using two blood results either lab eGFR or calculated eGFR

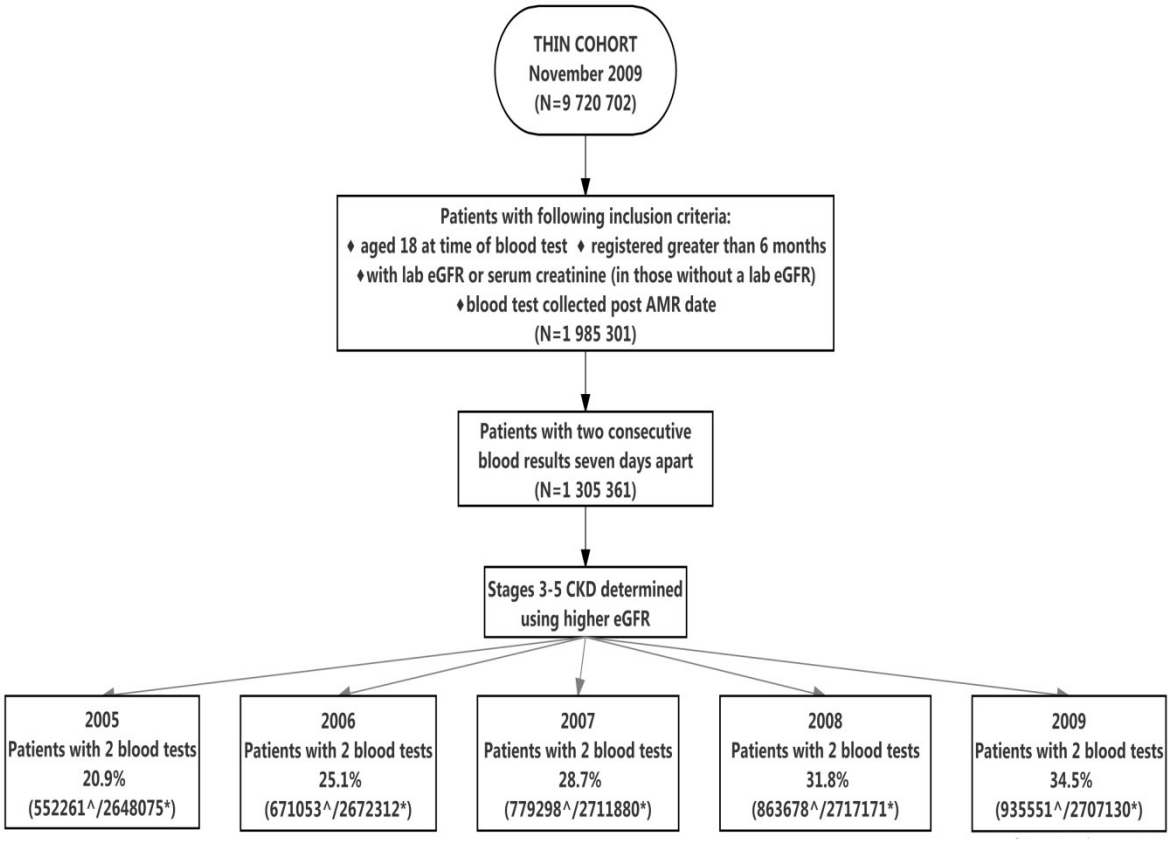
How the cohort was derived from the original THIN cohort for the principal analysis is shown in Figure 2-10. The proportion of patients with two serum creatinines seven days apart during the study period rose from 20.9% in 2005 to 34.5% in 2009. The median time apart from each laboratory report was 288 days (Interquartile range 126 to 457 days) and by 2009 this was 339 days (153 to 579 days). The following proportions in text will be reported as age and gender standardised unless otherwise stated. The age and gender standardised prevalence of Stages 3-5 CKD derived from two lab eGFRs or two calculated eGFRs(where lab eGFR not available), using the non IDMS MDRD equation, was 4.27% (95% CI 4.25 to 4.29) in 2005, rose to 5.19% (5.16 to 5.20) by 2007, then fell in 2008 to 5.13% (5.10 to 5.14) and continued to fall by 2009 to

4.78% (4.76 to 4.79) (Table 2-5). Regardless of year the commonest CKD stage was CKD 3a, with prevalence of 3.56 % (3.54 – 3.57) in 2009 (Table 2-5).

### *CKD prevalence by age and gender*

When deriving prevalence by age group and gender, two things were apparent, patients over 60 and women were more likely to have CKD. In women aged over 75 over 30% of the population had stages 3-5 CKD (defined using eGFR calculated using two serum creatinines and the Non IDMS MDRD equation) (Figure 2-11) and in men this was over 20%.

**Figure 2-10. How the cohorts were defined for the principal analysis.**



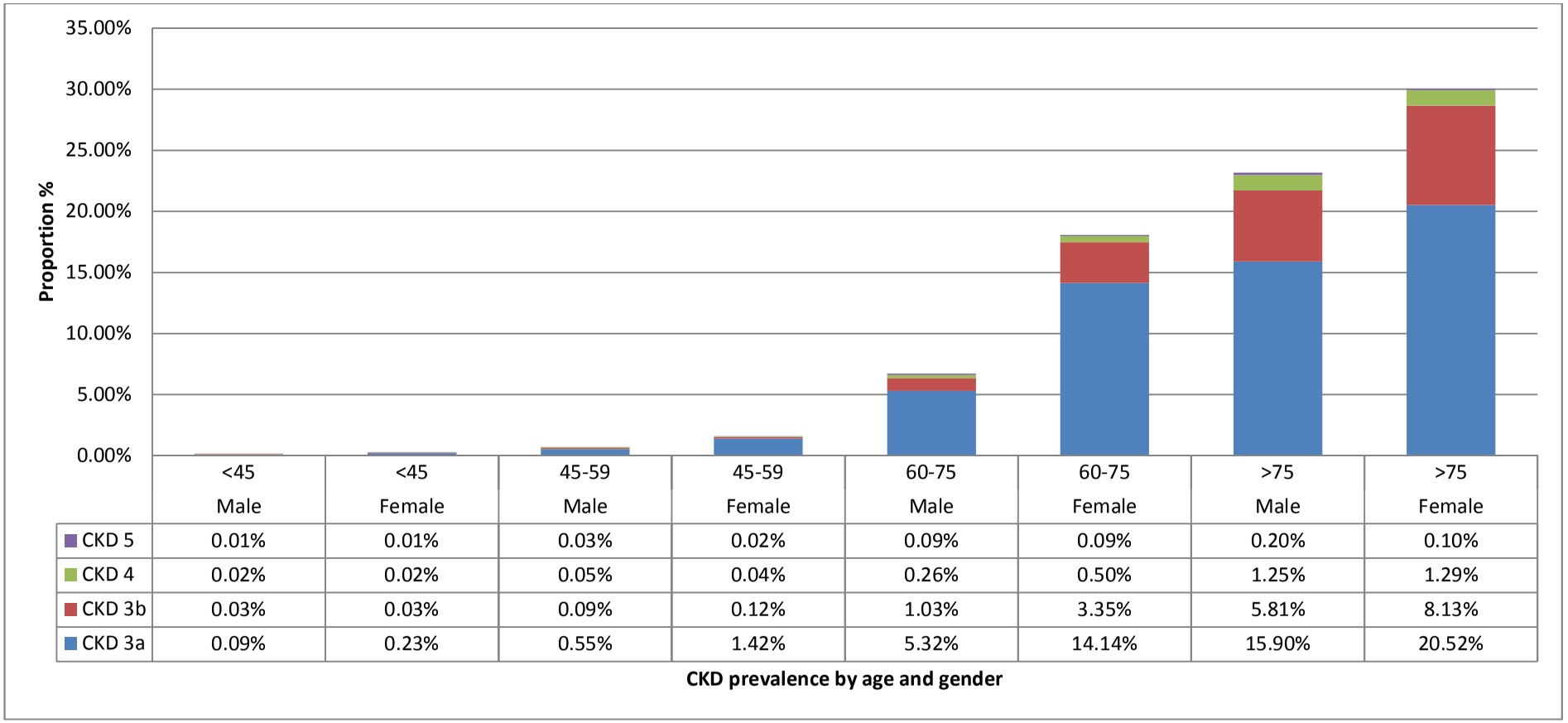

^ indicated the patients who were alive and still registered at the practice by July of the year in question.

* Patients aged over 18, alive and registered at the practice in July of the year in question

**Table 2-5. Prevalence of Stages 3-5 CKD using two blood results either lab eGFR or calculated eGFR using by Non IDMS MDRD (95% confidence intervals in brackets)**

| Year (Denominator population) | eGFR > 60 ml/min/1.73m$^2$ | CKD 3a | CKD 3b | CKD 4 | CKD 5 | Total CKD 3-5 |
|---|---|---|---|---|---|---|
| 2005 **(2648075)** | 16.06% (16.03-16.09) | 3.21% (3.19-3.22) | 0.87% (0.86-0.87) | 0.16% (0.15-0.16) | 0.03% (0.03-0.03) | 4.27% (4.25-4.29) |
| 2006 **(2672312)** | 19.71% (19.67-19.74) | 3.61% (3.59-3.62) | 0.96% (0.95-0.97) | 0.18% (0.17-0.18) | 0.04% (0.03-0.03) | 4.79% (4.76-4.80) |
| 2007 **(2711880)** | 22.82% (22.78-22.85) | 3.89% (3.87-3.90) | 1.06% (1.04-1.06) | 0.20% (0.19-0.19) | 0.04% (0.03-0.04) | 5.19% (5.16-5.20) |
| 2008 **(2717171)** | 26.01% (25.97-26.04) | 3.84% (3.82-3.85) | 1.04% (1.03-1.05) | 0.20% (0.19-0.20) | 0.04% (0.03-0.04) | 5.13% (5.10-5.14) |
| 2009 **(2707130)** | 29.02% (28.98-29.06) | 3.56% (3.54-3.57) | 0.98% (0.97-0.99) | 0.20% (0.19-0.19) | 0.04% (0.04-0.04) | 4.78% (4.76-4.79) |

**Figure 2-11. CKD prevalence by Age and Gender using two blood results either two lab eGFRs or two calculated eGFRs using the the Non IDMS MDRD equation**




| | <45 Male | <45 Female | 45-59 Male | 45-59 Female | 60-75 Male | 60-75 Female | >75 Male | >75 Female |
|---|---|---|---|---|---|---|---|---|
| CKD 5 | 0.01% | 0.01% | 0.03% | 0.02% | 0.09% | 0.09% | 0.20% | 0.10% |
| CKD 4 | 0.02% | 0.02% | 0.05% | 0.04% | 0.26% | 0.50% | 1.25% | 1.29% |
| CKD 3b | 0.03% | 0.03% | 0.09% | 0.12% | 1.03% | 3.35% | 5.81% | 8.13% |
| CKD 3a | 0.09% | 0.23% | 0.55% | 1.42% | 5.32% | 14.14% | 15.90% | 20.52% |

**CKD prevalence by age and gender**

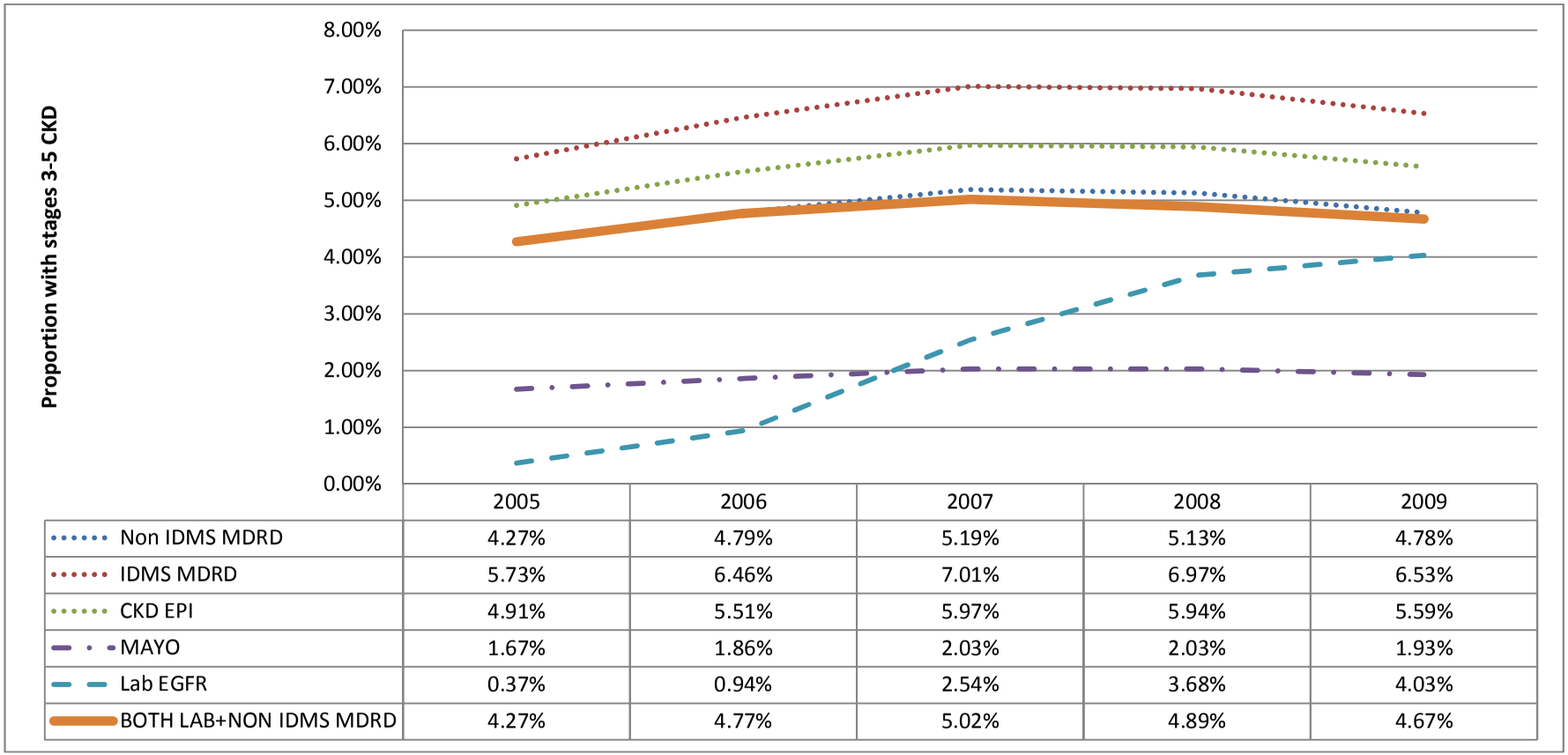
#### 2.4.2.1.2.Secondary Analysis: CKD staged using two blood results

The prevalence of CKD using two calculated eGFR using the non IDMS equation was higher than the principal analysis. (4.78 vs 4.67%). The prevalence of CKD was even higher using the CKD EPI equation and the IDMS MDRD equation (Figure 2-12). The prevalence of stages 3-5 CKD using these two equations in 2009 was 5.59% (5.56-5.60) and 6.53% (6.50-6.54)(Figure 2-12). CKD stage 3a was the commonest CKD stage regardless of equation.

The proportion of patients identified with CKD stages 3-5 based on the Mayo quadratic equation was considerably less and varied from 1.67% to 2.03% depending on the year CKD stage 3a was the commonest CKD stage regardless of equation, followed by CKD stage 3a, 4 and 5. (Figure 2-12)

The proportion of patients with two laboratory eGFRs rose from 2.4% to 21.43% between 2005 and 2009 (crude proportions) and the prevalence of Stages 3-5 CKD rose from 0.37% to 4.03%.(Figure 2-12)

**Figure 2-12. The prevalence of CKD using two blood results according to equation or directly reported lab eGFRs. The principal analysis is in Orange and the secondary analysis in dashed lines**

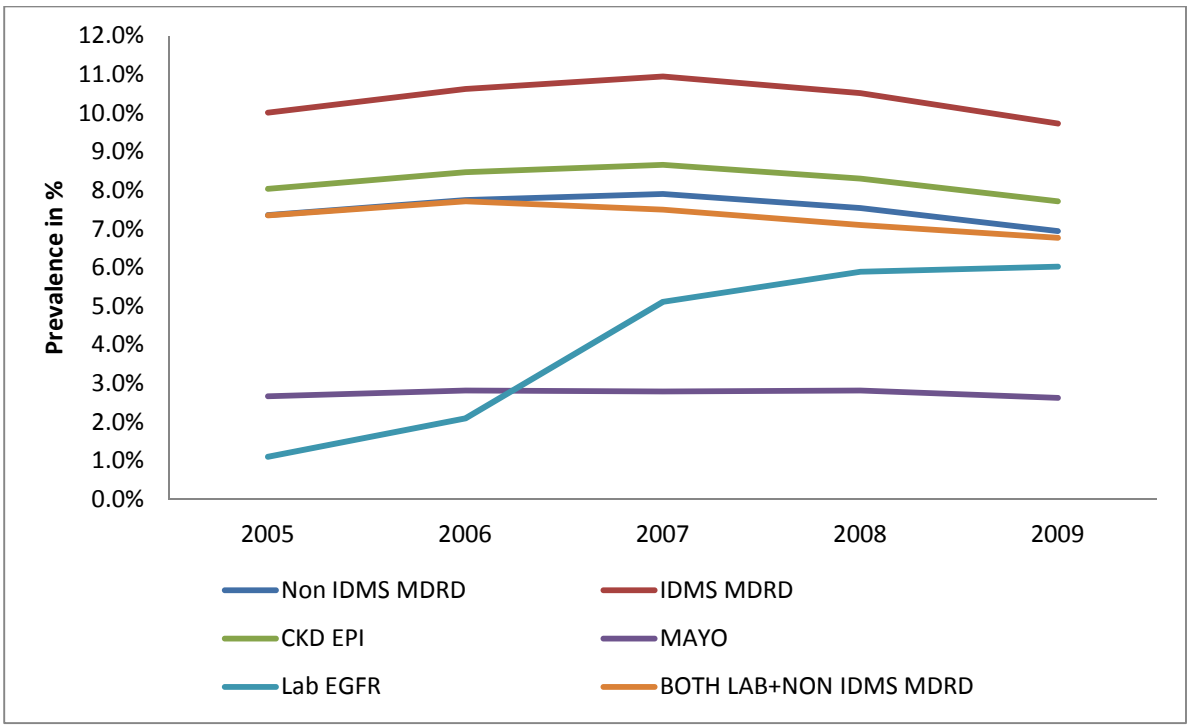



| | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|
| Non IDMS MDRD | 4.27% | 4.79% | 5.19% | 5.13% | 4.78% |
| IDMS MDRD | 5.73% | 6.46% | 7.01% | 6.97% | 6.53% |
| CKD EPI | 4.91% | 5.51% | 5.97% | 5.94% | 5.59% |
| MAYO | 1.67% | 1.86% | 2.03% | 2.03% | 1.93% |
| Lab EGFR | 0.37% | 0.94% | 2.54% | 3.68% | 4.03% |
| BOTH LAB+NON IDMS MDRD | 4.27% | 4.77% | 5.02% | 4.89% | 4.67% |

#### 2.4.2.1.3.Sensitivity analysis :Prevalence based on a single blood results

The proportion of patients with two serum creatinines rose from 36.9% in 2005 to 50.8% in 2009 (crude proportions). The prevalence results followed similar trends to those observed estimated using two blood test; CKD 3-5 prevalence rose between 2005 to 2007 and then fell in 2009, using eGFR calculated by serum creatinine. In comparison with patients staged with CKD using two blood test result the prevalence was much higher. The 2009 prevalence estimates according to the different formulae are shown in Figure 2-13. In summary the 2009 prevalence was estimated at 6.94% (95% CI 6.91 to 6.96) using the Non IDMS MDRD eGFR; and 9.73% (9.70 to 9.75) using the IDMS MDRD eGFR and 7.71% (7.68 to 7.73) using the CKD EPI MDRD equation.

**Figure 2-13. The prevalence of stages 3-5 CKD using a single blood result according to equation or two lab eGFRs**



Similar to prevalence results using two blood tests, the majority of patients had CKD stage 3a, followed by CKD stage 3b, 4 and 5. (data not shown). The CKD prevalence using the Mayo formula was approximately 3% (Figure 2-13). CKD 3-5 prevalence defined by lab eGFR was 7.71% (7.69% to 7.73%) in 2009 (Figure 2-13). In patients with either 2 eGFRs or 2 calculated eGFRs using the non IDMS MDRD equation, CKD prevalence was 6.77% (Figure 2-13).

### 2.4.3.Prevalence of stages 1-2 CKD: ACR and Urine Dip Data

In whole THIN cohort, 173 107 patients had a urine ACR and 752 993 patients had a urine dip. (Figure 2-14). In patients with an ACR, 94 542 patients had two ACRs 7 days apart and 22 000 patients had persistent microalbuminuria. In 378 665 patients with two urine dips 7 days apart 17 162 had a urine dip of trace or above.

Urine ACR was positively skewed and the median ACR was 1.5 mmol/mmol (range 0 to 5000). (Figure 2-15) In patients with urine dips, the majority of results were no proteinuria (85.7%), followed by trace proteinuria (6.0%) and then 1+ proteinuria (4.0%).

**Figure 2-14. Schematic of urine ACR/Prot Data and Derivation of CKD 1-2**

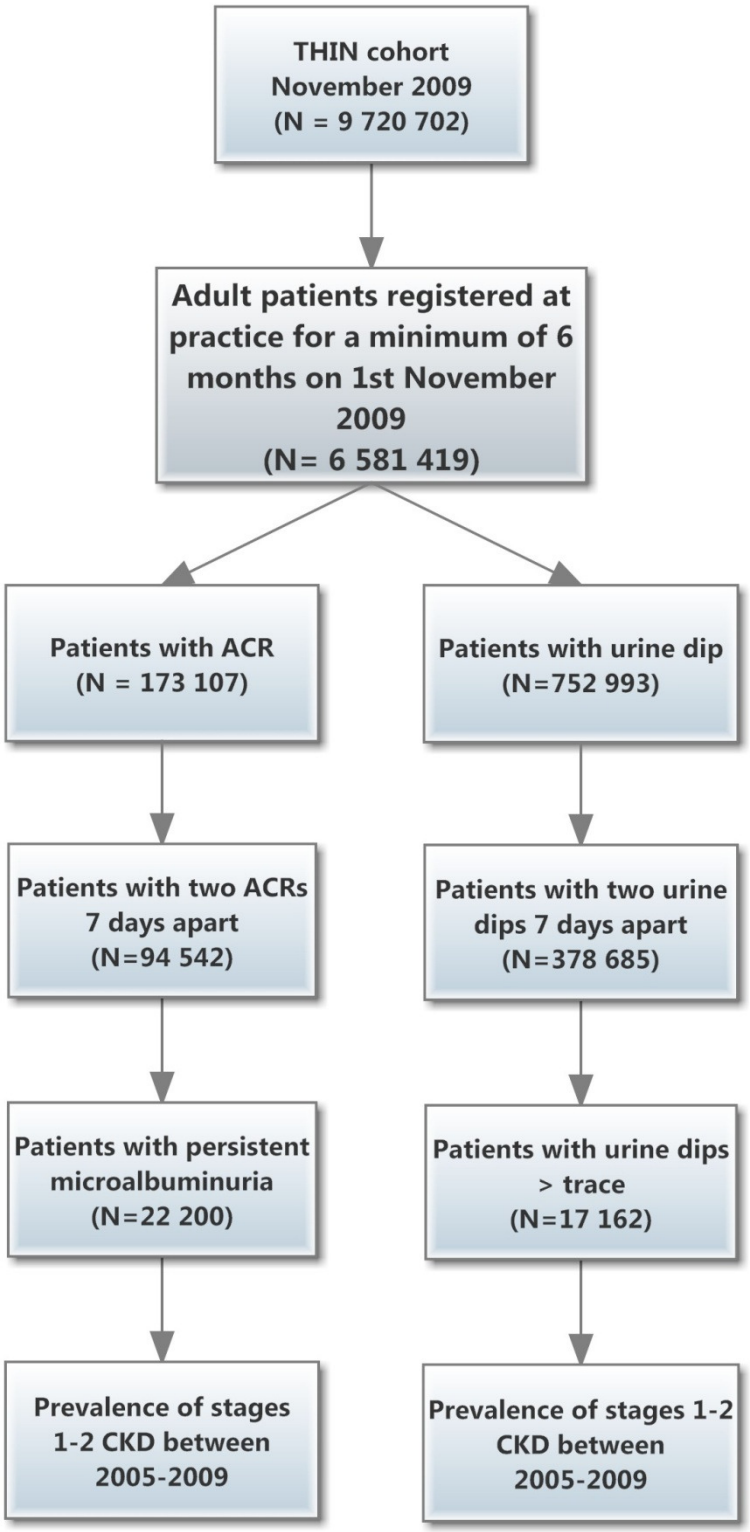
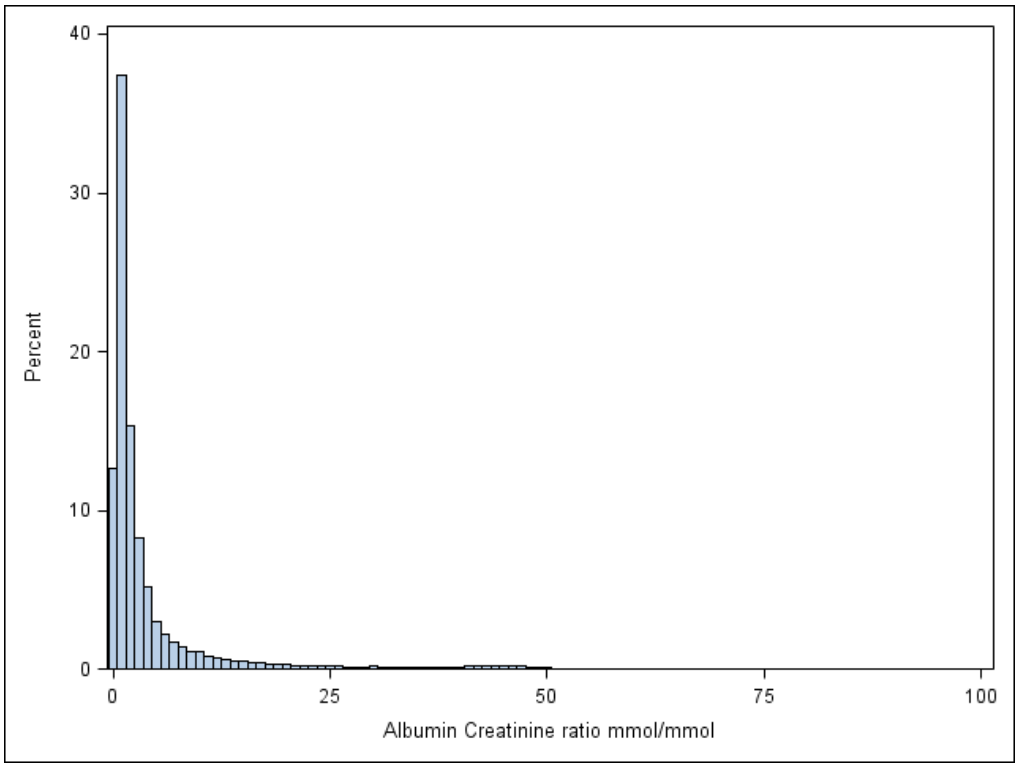


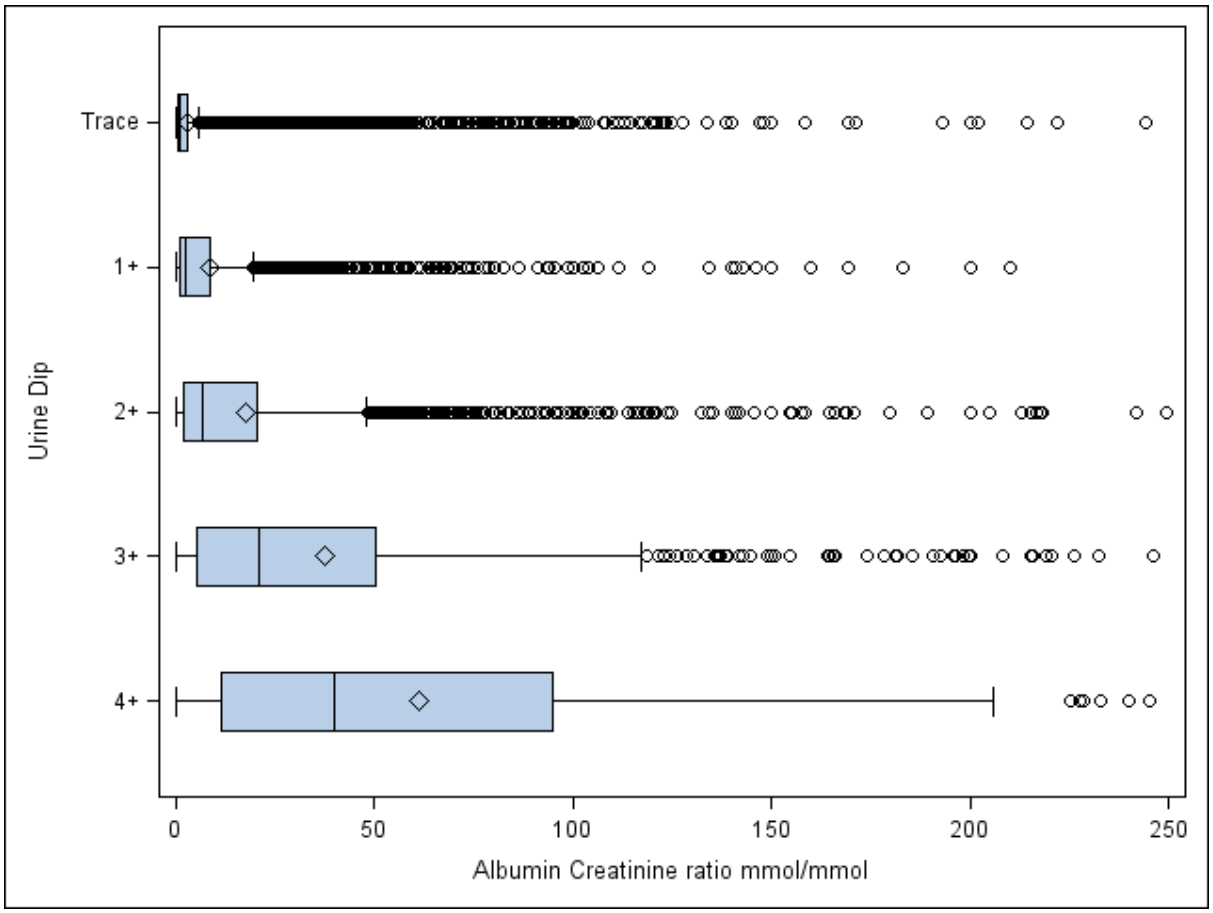
**Figure 2-15. Histogram of Albumin Creatinine Ratio**




### *2.4.3.1.The accuracy of urine dip*

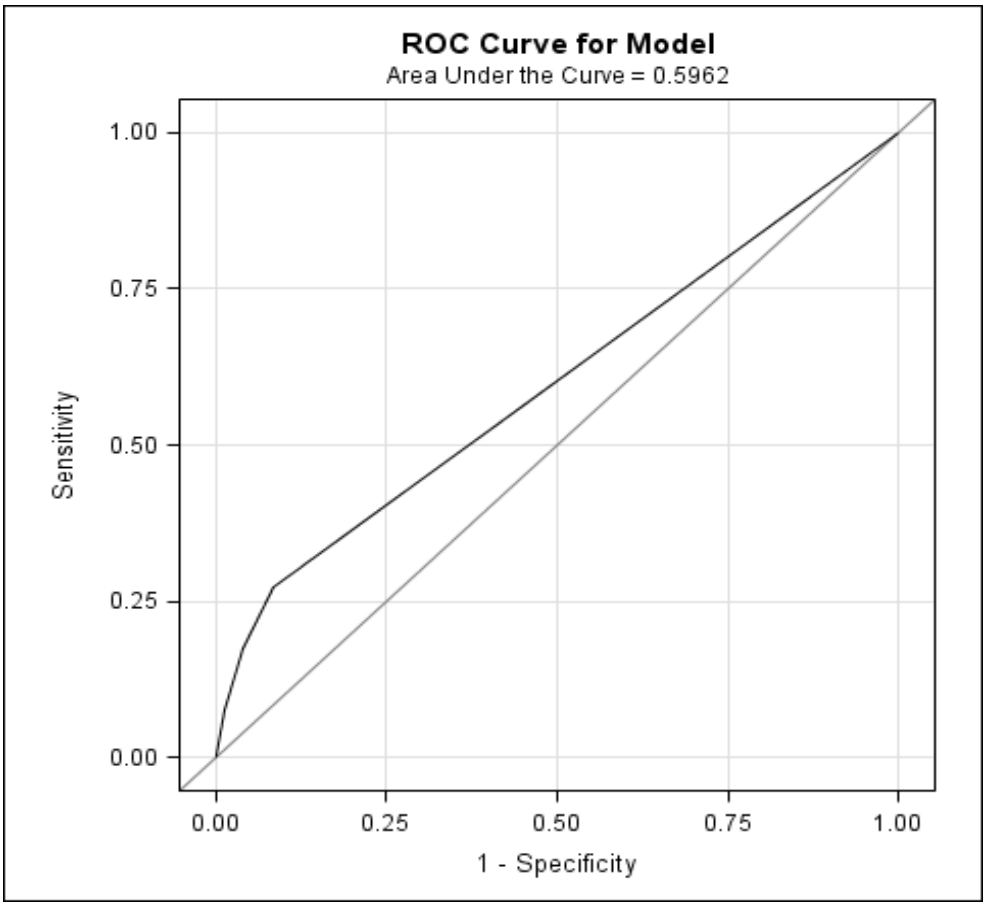
There were 58 216 instances where patients had a urine dip and reported urine ACR on the same day. There was a trend for increasing albuminuria with increasing quantification on urine dip (Figure 2-16).

**Figure 2-16. Box and Whisker plot of ACR versus urine dip**




If the urine dip was defined as trace proteinuria or above compared to ACR then the sensitivity of detecting albuminuria was 27.1%, the specificity was 91.6%, the positive predictive value was 41.7% and the negative predictive value was 85.0%. If defining albuminuria 1+ or above then sensitivity fell to 17.1%, the specificity rose to 96.3%, the positive predictive value rose to 50.4% and the negative predictive value fell to 84%. The ROC curves demonstrate this (detailed in Figure 2-17) and therefore trace proteinuria was used to define proteinuria.

**Figure 2-17. ROC curve for urine dip to diagnose albuminuria**



### *2.4.3.2. Prevalence of stages 1-2 CKD*

The proportion of patients with two consecutive ACRs rose from 5.51% to 8.26% (crude proportions) between 2005 and 2009. The proportion of patients with two urine dips greater than 14 days apart rose from 5.59% to 8.34% between 2005 to 2009. The age – standardised prevalence of CKD stages 1-2 rose from 0.04% (95% CI 0.041 to 0.046) defined by albuminuria on ACR to 0.32% (0.312 to 0.325) between 2005 and 2009 (Table 5). The age – standardised prevalence of CKD stages 1-2 rose from 0.21 % (95% CI 0.20 to 0.21) defined by urine dip of trace or above 0.34% (0.33 to 0.35) between 2005 and 2009 (Table 2-6).

**Table 2-6. Prevalence of CKD stages 1-2 using either ACR or Protein dipstick**

| | Prevalence % (95% CI in brackets) | |
|---|---|---|
| Year (Denominator population) | Using Urine ACR | Using Urine Protein dipstick |
| 2005 **(2648075)** | 0.04(0.04-0.05) | 0.21(0.20-0.21) |
| 2006 **(2672312)** | 0.08(0.08-0.09) | 0.23(0.23-0.23) |
| 2007 **(2711880)** | 0.13(0.13-0.14) | 0.27(0.26-0.27) |
| 2008 **(2717171)** | 0.24(0.23-0.24) | 0.30(0.29-0.31) |
| 2009 **(2707130)** | 0.32(0.31-0.33) | 0.34(0.33-0.35) |

## 2.5.Discussion.

This section will discuss the results for prevalence of CKD using different combinations of reported results and different methods of estimating eGFR. Overall combining CKD 3-5 with stages 1 and 2, the combined prevalence was 5.01%. Because of the different methodologies employed in defining CKD 3-5 and CKD 1-2, they will considered separately.

### 2.5.1.Summary of key findings for stages 3-5 CKD

#### *2.5.1.1.The Prevalence of stages 3-5 chronic kidney disease*

In the principal analysis the prevalence of stages 3-5 was 4.67%. In the secondary and sensitivity analysis, prevalence varied according to available data (for lab eGFR), the equation used to derive eGFR (and therefore staging) and whether two blood results were available. The table below summarises the prevalence of stages 3-5 CKD using a single blood result or two consecutive blood results by equation used to derive eGFR and lab eGFR in 2009. The numbers are age and gender standardised.

**Table 2-7. Summary of prevalence of Stages 3-5 CKD in 2009 from the thesis**

| Formula or Lab eGFR | Two blood results | A single blood result – sensitivity analysis |
|---|---|---|
| **Lab eGFR/non IDMS MDRD** | **4.67%** | 6.77% |
| **Non IDMS MDRD** | 4.78% | 6.94% |
| **IDMS MDRD** | 6.53% | 9.73% |
| **CKD EPI** | 5.59% | 7.71% |
| **Mayo Quadratic** | 1.93% | 2.63% |
| **Lab eGFR** | 4.03% | 6.02% |

The commonest CKD stage was stage 3a CKD regardless of how eGFR was calculated. For example in 2009 using two consecutive lab eGFRS the prevalence of CKD stage 3a

was 3.56%. This accounted for 76% of people with stages 3-5 CKD. Stages 3-5 CKD were more common in patients aged over 60 and women.

### 2.5.2.Interpretation of results and comparison to existing literature

#### *2.5.2.1.Prevalence of stages 3-5 CKD*

When comparing the lab eGFR and the eGFR calculated by the non IDMS MDRD formula, even though the lab and the method analysed was unknown the two tests showed reasonable agreement when the eGFR was below 60 ml/min/1.73m$^2$. This suggests that, despite the method of measuring creatinine varied between laboratories, diagnosing stages CKD 3-5 with a calculated eGFR using the Non IDMS MDRD equation was reasonably accurate.

This was supported by fact that the difference between the proportion of patients with CKD 3-5 using both lab eGFRS/calculated eGFR and only Non IDMS MDRD calculated eGFRS was similar. To further illustrate this point, by 2009, 931 262 patients had two serum creatinines and of these 645 852 had had a reported eGFR. If prevalence was defined by a calculated eGFR (Non IDMS MDRD) then the prevalence would be 4.78%. However the prevalence fell slightly to 4.67% when using both lab eGFR and calculated eGFR. This suggests a marginal understimation of calculated eGFR over lab eGFR.

CKD prevalence varied over time using the non IDMS MDRD equation. It is likely the observed increase and subsequent decrease in prevalence over the study period is due to changes in the creatinine analysis method used in laboratories providing blood results to practices in THIN.[31] Post 2006, to enable greater accuracy and generalisability of creatinine results, more laboratories switched to IDMS traceable methods of creatinine analysis.(1.2.8)[9] A creatinine generated by non IDMS methods was likely to be higher if the same blood test was analysed by IDMS methods (hence the change in coefficient in MDRD equation). So, as more laboratories switched to IDMS methods, serum creatinines would be lower (the lower the creatinine the higher the GFR) and hence the prevalence fell.[31]

**Table 2-8. Methods of creatinine analysis using UK National External Quality Assessment Service based laboratories[31]**

| | **Proportion of laboratories using this method** | | |
|---|---|---|---|
| **Method of Measuring Creatinine** | April 2006 | April 2009 | December 2010 |
| **Dry slide** | 10.2 | 9.6 | 8.1 |
| **Endpoint Jaffe** | 3.5 | 2.4 | 1.3 |
| **Enzymatic*** | 0.8 | 3.9 | 9.4 |
| **Compensated Kinetic Jaffe*** | 20.4 | 38.6 | 57.6 |
| **Traditional Kinetic Jaffe** | 52.9 | 37.1 | 20.7 |
| **O'Leary** | 10.6 | 7.5 | 2.9 |

*Traceable to IDMS

This is also the reason why the prevalence in general was higher when Stages 3-5 CKD are derived from IDMS calculated eGFR. The prevalence of CKD was likely to be more

accurate where CKD was staged by Non IDMS eGFR between 2005-2008; whereas the IDMS eGFR staged CKD would be more accurate in the later years i.e. 2009 onwards because the method of creatinine analysis was likely traceable to IMDS.

#### *2.5.2.2.MDRD Equation single blood result*

Using a single blood test, the prevalence derived by the non IDMS MDRD formula 6.94%. This was higher than results estimated from previous screening studies (NHANES -5.6 %, HUNT – 4.7%, HSE 6.0 %) but lower than single studies using primary care estimates such as NEOERICA (8.5%).[32;52;62] The reason that this was lower than the NEOERICA is because the NEOERICA study examined period prevalence rather than point prevalence, i.e. patients were diagnosed with CKD 3-5 if they had an eGFR below 60 between 1998 and 2003 and did not account for fluctuation in eGFR between years. The reason for these differences may be that the characteristics differ of patients who have been screened. People who participated in voluntary screening programmes or studies are likely to be co-operative and healthier patients. Interestingly the proportion of CKD 3-5 defined using a single lab eGFR and/or calculated eGFR in this study (6.77%) was similar to the prevalence observed in the UK based QICKD study (6.7%).[49] QICKD defined prevalence of CKD 3-5 using two consecutive lab eGFRs, two consecutive calculated eGFRS, a single lab eGFR and single calculated eGFR combined. They preferentially used lab reported eGFR and two results where available.

#### *2.5.2.3.Prevalence based on two blood results and MDRD equation*

Perhaps unsurprisingly, prevalence estimates based on two blood results were much lower than those based on a single blood test. Creatinine shows inherent fluctuation and may increase temporarily if the patient suffers acute kidney injury or has started

on angiotensin blockade. This will temporarily lower the eGFR. Serum creatinines show better calibration on more than one occasion.[32] This is why the N/KDOQI guidelines recommend confirmation of CKD in two blood tests three months apart. In the analysis in this thesis, a minimum difference of 7 days was specified but the majority of patients had serum creatinines taken more than 90 days apart (in 2009 this was 339 days (IQR 153-579 days). As mentioned previously using two blood results is the best method to define CKD but with closer time points. (Chapter 2)

The results showed that Stage 3 CKD was the most commonest stage and that the prevalence was higher in women and older patients. These findings are consistent with the literature.[48]

#### *2.5.2.4.Prevalence: other equations*

Prevalence of CKD 3-5 was also defined using a variety of other equations. The CKD EPI based CKD prevalence is higher than Non IDMS prevalence but lower than IMDS based prevalence. The CKD EPI formula was designed for serum creatinines analysed using IDMS methods and therefore non IDMS serum creatinines would be higher and inflate the prevalence of CKD. However CKD EPI formula tries to correct for under estimation of eGFR when using the (IDMS) MDRD equation and prevalence is therefore lower than the IDMS derived estimates.[27;49;66;67;80] Currently all laboratories use IDMS methods and therefore CKD EPI derived prevalence will not be confounded by

laboratory method. The risk of death and ESRD may be better predicted in those diagnosed with CKD by CKD EPI formula.[159]

The prevalence derived using Mayo quadratic based equation was much lower. The Mayo quadratic formula was derived using serum creatinine analysed by a different assay from the MDRD and CPD EPI studies. It therefore requires a correction to make the creatinine comparable to serum creatinines analysed in the MDRD study.(Chapter 1.2)

### 2.5.3.Limitations and Strengths

The biggest limitation in calculating the eGFR is that the method of creatinine analysis used in the lab contributing the result was unknown and furthermore that the ethnicity of many affected patients was unknown. The method of creatinine analysis may lead to inaccuracy when calculating eGFR: for example if the creatinine was analysed by IDMS methods and then the eGFR was calculated by Non IDMS MDRD equations, the prevalence of CKD will appear lower as the eGFR will be lower than expected.[9] The converse is true the other way round and this is why the prevalence dips after 2007. However the calculated eGFR using non IDMS formula and the lab eGFR were broadly comparable and this method of defining CKD will have a high specificity but still at least 80% sensitivity and is the optimal way to define CKD from a cohort such as this.[157] The combined two result prevalence of CKD derived from the lab eGFR and non IDMS MDRD equation is likely to be closest to the population

prevalence of Stages 3-5 CKD. This is provided that adults without 2 results do not have CKD.

It is difficult to interpret the findings of prevalence using the non MDRD equations. The prevalence defined using CKD EPI is likely to inflate prevalence especially early in the study period years when the implementation of IDMS aligned creatinine analysis was low.

A further limitation is that the proportion of patients with black ethnicity was unknown. In such patients the eGFR would have been under-estimated (black patients have higher muscle mass, therefore they have higher creatinines in comparison to white population). This population would have been misdiagnosed with CKD. However the proportion of those with black ethnicity in the UK at the time was only 3% and therefore this would have limited impact on the whole population prevalence.[135] The QICKD study found little difference in prevalence when eventually accounting for ethnicity.[49]

Other drawbacks to this work include the usual limitations of retrospective studies. Though THIN data were collected prospectively, the data were analysed retrospectively. Patients who have had blood tests are not a randomly selected, which means they are less representative of the whole population: those likely to have serum creatinines are older and have more co-morbidity in comparison the general population.[66] Despite these caveats, this cohort is more likely to represent a 'real world' population than other secondary care based estimates and patients with higher

co-morbidity are more likely to have CKD.[133] Furthermore NICE advises that health care practitioners target high risk populations for CKD testing and screening.[44] Even if a large screening study were undertaken, it is likely that patients included in this study would not necessarily have participated and therefore this represents one of the largest samples from which the prevalence of stages 3-5 has been defined. Apart from the QICKD study this is the only study to examine the chronicity of CKD in primary care.[49]

### 2.5.4.Summary of Results for Stages 1-2 CKD

To derive the prevalence of stages 1-2 CKD, the accuracy of urine dip to detect high albuminuria was examined. When using the threshold of trace proteinuria or above on dipstick, the sensitivity was 27.1% and specificity was 91.6%. When using the threshold of 1+ proteinuria or above on dipstick there was a marginal increase in specificity (96.3%) at a cost of sensitivity (17.1%). Therefore high albuminuria was defined as trace proteinuria or above. The proportion with two urine dips, 7 days apart was 8.26% and two urine ACRS 7 days apart was 5.51%. Stages 1-2 CKD were rare and the age and gender standardised prevalence were 0.34% and 0.32% using urine dip and urine ACR respectively in 2009.

### 2.5.5.Interpretation of Results and Comparison to existing literature

Urine dip is insensitive but specific in detecting high albuminuria.[42] Therefore if a patient consistently has trace proteinuria they are likely to have high albuminuria levels. This study confirms the findings from the AUSDIAB study that urine dip was

insensitive but highly specific at detecting high albuminuria.[42] In this study population very few patients had two consecutive urine dip results and this may reflect under ascertainment of CKD as in the screening studies a significant larger proportion of patients had stages 1-2 CKD.[62]

### 2.5.6.Strengths and Limitations

The major limitation to this part of the thesis is that very few people had the gold standard of urine ACR and although many patients had urine dips, this is a relatively insensitive method to detect albuminuria.[42] Additionally stages 1-2 CKD was defined using high albuminuria as opposed very high albuminuria which is required by the NKDOQI, NICE, and SIGN guidelines. However the international KDIGO guidelines state that the high albuminuria threshold that should be used defined stages 1-2 CKD and patients with high albuminuria are at high risk of cardiovascular morbidity and mortality and therefore the KDIGO definition was adopted to define stages 1-2 CKD. Areas of future research will be discussed in chapter 6.[93]

### 2.5.7.Executive Summary

- CKD is defined by reduced GFR or kidney damage i.e. proteinuria
- eGFR can be estimated from creatinine based equations such as the MDRD or CKD-EPI formula
- There can be considerable variation in creatinine measurement and this coupled with the different formulae can lead to a variation in eGFR and CKD
- CKD prevalence is very variable between countries and study groups.
- Overall this study found a combined prevalence of stages 1-5 CKD of 5.01%
- In the primary analysis the prevalence of stages 3-5 CKD was 4.67% using two blood tests
- CKD 3-5 prevalence varied according to the equation used to derive CKD.
- CKD 3-5 was more common in women and the elderly.
- Urine dip was insensitive but specific at determining microalbuminuria.
- CKD 1-2 prevalence was relatively low was about 0.32-0.33% in 2009.

# CHAPTER 3. DO PRIMARY CARE PRACTICES CORRECTLY DIAGNOSE STAGES 3-5 CKD COMPARED TO OBJECTIVE MEASURES AND HOW DOES THIS IMPACT ON PATIENT CARE?

## 3.1.Introduction

This chapter describes current guidance for the management of CKD in primary care from an international and UK perspective. Specifically it describes the UK Quality Outcomes Framework for CKD which has driven primary care management in the UK. The chapter explores the accuracy of diagnosis based on QOF reporting compared to eGFR results. The management of CKD Stages 3-5, including those recognised, unrecognised and mislabelled, is reported and the implications for UK general practice discussed.

In response to the public health problem of CKD, a variety of international and national guidelines for diagnosis and management of CKD have been produced.[1;43;160;161] Guidelines for CKD originate from work by the National Kidney Foundation Kidney Disease Outcomes Quality Initiative (N/KDOQI) [1] The guidelines call for early recognition of CKD so that earlier intervention can be implemented i.e. treatment of cardiovascular risk factors. They specify management guidelines, including interventions for blood pressure, cholesterol and anaemia.[1]

The KDIGO guidelines were developed after an international conference in order to compare (and unify) different international guidelines in 2003 to improve the care and outcomes of CKD patients.[43] The guidelines have recently been updated for blood pressure, CKD classification and management. Interestingly they define proteinuria as "high albuminuria" (Table 1-2) when diagnosing CKD stages 1-2, this differs from current UK NICE guidance but may be adopted in the future.[43;44]

In the UK as a response to the high prevalence and poor outcomes of CKD, the Department of Health set up a specific National Service Framework (NSF) for dialysis and transplantation in 2004. CKD, acute renal failure and end of life care were added subsequently to the framework.[160] The NICE guidelines for the diagnosis and management of CKD were published with consultation with various stakeholders such as the Royal Colleges of Physicians, GPs and Pathologists plus charities including Diabetes UK and Kidney Research UK in 2008.[44]

In brief, NICE specify guidelines for diagnosing and classifying CKD with emphasis on using IDMS aligned methods to report eGFR and ACR to define proteinuria.[44] They specify pathways for the diagnosis and monitoring of renal function in patients with and without diabetes and regard patients with an ACR above 30 mg/mmol as having proteinuria in non diabetics. They also specify management pathways for patients with or without diabetes. In patients with diabetic CKD, high albuminuria and hypertension, Angiotensin blockers are recommended whilst in patients with non-diabetic CKD, Angiotensin blockers are recommended in patients with hypertension and an ACR above 30 mg/mmol. Patients with non-diabetic CKD and an ACR below 30 should be offered standard antihypertensive regimes. The guidance also specifies referral to specialist (secondary care) if patients have stage 4 or 5 CKD, non Diabetic CKD and ACR above 70, declining eGFR, poorly controlled hypertension and rare diseases.[44]

How these guidelines have impacted upon the management of CKD is difficult to estimate. In UK, health care is free at the point of access and primary care is the first

point where patients consult before being referred to additional services. Since 2006, as primary care physicians started to receive direct measures of renal function from laboratories for their patients, the diagnosis of CKD increased as did referrals to secondary care.[162] Stable CKD 3 patients without high levels of proteinuria and recalcitrant hypertension can be managed in primary care.[44] However, this is a huge undertaking as the majority of CKD patients newly diagnosed are in this category and specialist care has further discharged patients back to primary care.[163]

### 3.1.1.The UK Quality Outcomes Framework

The QOF is a payment for performance (P4P) scheme first introduced in the General Medical Services Contract for primary care physicians in the UK in 2004.[164] In a review of QOF and the selection of indicators it difficult to determine whether QOF aims to improve quality of care or implement basic standards.[164] The QOF consists of four domains which are clinical management, organisation of clinical services, patient experience and additional services. In the clinical domain, in 2009/2010 there were 20 diseases or health management areas covered and in total 86 management targets or indicators.[140] This accounted for up to 25% of the revenue that primary care physicians received from the NHS if they achieved targets in 70% of relevant patients.[45] Practices are remunerated in a scaled payment dependant on whether they achieve a certain proportion of the targets in their denominator population. This system was almost universally adopted by general practitioners by 2006.[165]

QOF targets are specific to common conditions, have impact on morbidity and mortality, are identifiable and applicable to primary care.[164] Therefore it is not

surprising that there are clinical indicators for chronic disease management such as diabetes, asthma, hypertension and CKD. When managing these patients according to QOF, the primary care practice is asked to first identify patients with such a disease on the chronic disease register and then meet management targets. These QOF points are identified through coding in the primary care electronic patient records i.e. Read codes which are described further in the methods section.

Chronic Disease registers not unique to the UK. European countries such as Sweden and Denmark have an extensive number of disease registers, for example Sweden had over 90 such registries (Table 3-1).[166] These have been shown to improve health and outcomes in both countries.[166] In the United States and Canada such national registers have only been developed in the last few years. For example in 2006 only 52% of health care providers instituted P4P related disease registers.[167-169](Table 25)

**Table 3-1. Countries with disease registers and involvement in payment for performance**

| Country | Disease register | Related to Payment for Performance |
|---|---|---|
| **Australia** | Diabetes<br>Hypertension<br>Chronic Kidney Disease<br>Cancer<br>Cardiovascular Disease<br>COPD | Yes[40] |
| **United States America** | Asthma<br>Diabetes<br>Hypertension<br>Chronic Kidney Disease<br>Arthritis | Yes part of Medicare / Medicade called the Physician Quality Reporting System<br>Also adopted by other Health care organisations[170] |

| | | |
|---|---|---|
| **Canada** | Cardiovascular Heart Disease<br>Diabetes<br>Hypertension | Not at a national level[171] |
| **Finland** | Cardiovascular Disease | Yes[172] |
| **Sweden** | Multiple Registers<br>Heart Failure<br>Diabetes<br>Cancer<br>Stroke<br>COPD<br>and another 70 conditions | Yes multiple initiatives which incorporate use of registers[173] |

The UK QOF targets for CKD for the years 2008-2009 are described

Table 3-2, but there are additional indicators that pertain to CKD such as diabetes

(Table 3-3) and hypertension both of which are highly prevalent in CKD patients.[140]

**Table 3-2. QOF indicators for CKD in 2008/2009[140]**

| Indictor | Points |
|---|---|
| **Records** | |
| CKD 1: The practice can produce a register of patients aged 18 years and over with CKD (US National Kidney Foundation: Stage 3 to 5 CKD) | 6 |
| **Initial management** | |
| CKD 2: The percentage of patients on the CKD register whose notes have a record of blood pressure in the previous 15 months | 6 |
| **Ongoing management** | |
| CKD 3: The percentage of patients on the CKD register in whom the last blood pressure reading, measured in the previous 15 months, is 140/85 or less | 11 |
| CKD 5: The percentage of patients on the CKD register with hypertension and proteinuria who are treated with an angiotensin converting enzyme inhibitor (ACE-I) or angiotensin receptor blocker (ARB) (unless a contraindication or side effects are recorded) | 4 |

**Table 3-3. QOF indicators for diabetes mellitus specific to CKD patients in 2008/2009**

| Indicators | Points |
|---|---|
| Ongoing management | |
| DM 2. The percentage of patients with diabetes whose notes record BMI in the previous 15 months | 3 |
| DM 5. The percentage of patients with diabetes who have a record of HbA1c or equivalent in the previous 15 months | 3 |
| DM 20. The percentage of patients with diabetes in whom the last HbA1c is 7.5 or less (or equivalent test/reference range depending on local laboratory) in the previous 15 months | 17 |
| DM 7. The percentage of patients with diabetes in whom the last HbA1c is 10 or less (or equivalent test/reference range depending on local laboratory) in the previous 15 months | 11 |
| DM 11. The percentage of patients with diabetes who have a record of the blood pressure in the previous 15 months | 3 |
| DM 12. The percentage of patients with diabetes in whom the last blood pressure is 145/85 or less | 18 |
| DM 13. The percentage of patients with diabetes who have a record of micro-albuminuria testing in the previous 15 months (exception reporting for patients with proteinuria) | 3 |
| DM 15. The percentage of patients with diabetes with a diagnosis of proteinuria or micro-albuminuria who are treated with ACE inhibitors (or A2 antagonists) | 3 |
| DM 16. The percentage of patients with diabetes who 40-90% have a record of total cholesterol in the previous 15 months | 3 |
| DM 17. The percentage of patients with diabetes whose 40-70% last measured total cholesterol within the previous 15 months is 5mmol/l or less | 3 |

The advent of QOF has had advantages and disadvantages. QOF allows management of patients to universal standards, regardless of gender, ethnicity or socioecomnic background and evidence suggests that there is a reduction in health inequalities in some but not all of these areas.[174;175] There is also evidence to suggest improvement in indicators, such as the management of blood pressure in patients with hypertension, cardiovascular disease and diabetes [165] and glycaemic control and cholesterol in patients with diabetes.[141;176-178] The evidence for translation of achievement of short term goals to long term benefit is limited to two economic analyses, however both studies suggest a reduction in hospital admissions related to stroke and diabetes.[179;180] Potential arguments against the QOF are that universal care moves away from patient led care, evidenced based care should be standard and not incentivised, and care may not improve in non-incentivised areas.[165;181]

For patients with CKD, limited evidence suggests that blood pressure management improved after the introduction of QOF but there have been no large published studies. Additionally recognition of CKD may not be accurate as in 2009 the crude prevalence of CKD was 4.2% according to QOF which substantially differs from previous UK prevalence estimates. This suggests that practices do not identify all patients with CKD. Given the evidence above that patients in other diseases are not appropriately managed if not recognised by the practice QOF disease registers, it is important to analyse QOF performance in patients with CKD.[45;63].

## 3.2.Research Questions

This chapter will address the following research questions:

1. *Do primary care practices correctly identify those people with stages 3-5 CKD on the QOF primary care register?*

   This is ascertained by what proportion of patients with a QOF Read code for CKD had biochemical evidence of CKD.(Chapter 3.4.1)

2. *What is the burden of co-morbidity of those with CKD based on an e-GFR diagnosis and those on the primary care register and did they differ in a univariate analysis?(Chapter 3.4.1.1)*
3. *What factors lead to exclusion from the primary care CKD register in a multivariate analysis? (Chapter 3.4.1.2)*
4. *Does achievement of QOF indicators differ between CKD patients excluded from the practice register to those on the register?* (Chapter 3.4.2)
5. *Does achievement of Diabetes QOF indicators differ between CKD patients with diabetes excluded from the practice register to those on the register? (Chapter 3.4.2.1)*

## 3.3. Methods

### 3.3.1.Study Population

### 3.3.2.Estimating prevalence of CKD 3-5 on primary care practice register between 2005 to 2009

The prevalence of stages 3-5 CKD identified on practice registers was calculated by extracting the proportion of patients with a QOF business code for CKD for the QOF year ( on the 1st April ) in question between 2005 and 2009. This period corresponds to the time when the QOF would be analysed nationally in each practice. QOF business codes are specific read codes that indicate diagnosis of a disease.[45] Crude prevalences were then age and gender standardised and 95% confidence intervals were calculated. (Box 1-1)

### 3.3.3.Estimating the proportion of patients with CKD (stages 3-5) Read codes had biochemical evidence of CKD in the QOF analysis period of 2008-2009

CKD was defined using laboratory eGFR. Patients had to demonstrate two eGFRs below 60 ml/min/1.73m$^2$, at least seven days apart, to be diagnosed with CKD stage 3-5 before April 2009. Patients had to be aged 18 or above at the time of CKD diagnosis. The following groups were defined and the proportions in each calculated:

- *Confirmed CKD:* sustained biochemical evidence of stages 3-5 CKD, i.e. laboratory eGFR under 60 mls/min/1.73m$^2$ on the last two consecutive eGFRs before 1st April 2009 that were at least seven days apart.
- *Labelled CKD:* patients recorded by the practice as having stages 3-5 CKD by a Read code defined according to the UK QOF Business Rules.[140]

- *Appropriately coded:* patients with both a relevant Read Code for CKD *and* biochemical evidence of an eGFR under 60 mls/min/1.73m2 In the last two consecutive eGFRs before 1st April 2009 that were at least seven days apart.
- *Uncoded CKD:* patients with confirmed CKD (i.e. biochemical evidence as per definition above) but no relevant CKD Read code entered into their records.
- *Miscoded CKD:* patients with a relevant Read code for stages 3-5 CKD but no biochemical evidence of stages 3-5 CKD (using above definition).

Note that these definitions were applied by year hence if a patient had CKD previously but no longer had CKD in 2008-2009 then they would be defined as miscoded. Note patients with biochemical CKD were defined by two blood results not a single blood result.

### 3.3.4.Demographic and comorbid characteristics of these groups

From the November 2009 THIN dataset, as of 1st April 2009, the following were ascertained for the general population: confirmed CKD, labelled CKD, appropriately coded, uncoded CKD and miscoded CKD groups.

- Basic demographics such as age, gender and race
- Deprivation index: latest Townsend score as defined in Chapter 2[154]
- Smoking Status

- The following co-morbid conditions:
- Cardiovascular Disease (Ischaemic Heart Disease, Stroke and Peripheral Vascular Disease)
- Diabetes Mellitus
- Hypertension
- Hypercholesterolaemia

The co-morbid conditions were extracted from Read codes for the specific conditions using QOF business codes for the disease definitions, from previous definitions and direct examination of the Read codes based on clinical experience.[45;182]

### 3.3.5.Characteristics of CKD patients that were excluded from the QOF CKD register

It was important to recognise the factors that may predict the inclusion of individuals onto the QOF register as inclusion on the register was likely to result in better management.[141] Identifying the predictors for inclusion on the CKD register would help us to target any populations in whom the CKD diagnosis was missed.

Patients with confirmed CKD (based on 2 e-GFR results) were included in this analysis and the outcome was exclusion from the QOF register, i.e. a binary outcome. The co-variables for this analysis were age, gender, race, Townsend Quintile, co-morbid conditions as detailed in the previous section and smoking status. The co-variables were entered into a mixed model in a backward stepwise selection process with $\alpha = 0.05$ as criteria for model inclusion using the SASV9.2 PROC GLIMMX with a logit link and practice location as random effects term.[183] Non linear functional forms were

considered for age (log transformation).[184;185] The final model contained the significant co- variables. Practices were included as a random effects term, as practice management and the characteristics are likely to be different and represent clustering which may not be captured in the other co-variables.[183]

### 3.3.6.What are the differences between Uncoded and Miscoded patients – a mixed model

It was of interest to determine the differences between miscoded and uncoded CKD patients. The same methodology as above was used. The outcome in this model was exclusion from the CKD QOF register. The co variables were the same as above with the exception of CKD stage which is not applicable to miscoded CKD patients. Practice location was fitted as random effects term for the reasons described above. The same model selection procedures as above were used.

### 3.3.7.Comparison of QOF indicators between patients with labelled and confirmed CKD

As discussed in the introduction to this chapter, it is possible that patients with a particular disease who were not on a QOF disease register may have worse management than those on the disease register.[141] The management of patients with appropriately coded CKD on the QOF register was therefore compared to those patients with uncoded CKD. In further exploratory analyses, the management of uncoded CKD patients was compared with miscoded patients to assess the influence of coding.

Patients with labelled CKD and confirmed CKD with the appropriate QOF indicators were included in the analysis. The QOF indicators have been described earlier in this

chapter. The proportion of patients with labelled CKD, confirmed CKD, appropriately coded CKD, miscoded CKD and uncoded CKD with following indicators prior to April 2009 were calculated:

- The proportion who had a blood pressure in the last 18 months
- The proportion who had a blood pressure ≤ 145/85 in the last 18 months
- The proportion who had a urine ACR in the last 18 months
- The proportion who had a urine ACR >30 units in the last 18 months
- Those proportion coded with hypertension, an ACR above 30 units and on an ACE inhibitor in last 18 months

The proportions of achieved indicators were compared between the uncoded and appropriately coded CKD patients and the uncoded and miscoded CKD groups using Pearson Chi-Square tests.[109] Cholesterol and blood pressure results were extracted and were compared using student's T test if normally distributed or using the Mann Whitney Test if not.

As a significant proportion of patients with CKD have diabetes, the appropriate Diabetes QOF indicators were compared between the same groups using the same statistical methods as above.

## 3.4. Results

### 3.4.1.Who is on the Practice CKD register?

The baseline population for analysis consisted of 2 701 730 patients aged over 18, alive and registered at their current practice for at least six months on the 1st April 2009 of whom 631 905 patients (23.4%) had two lab eGFR seven days apart.

The flowchart in figure 3-1 demonstrates how patients were identified by biochemical criteria or having 'CKD Read codes' specific to QOF between 2008/2009 (Figure 3-1).

There were 108 911 (4.0%) patients with stages 3 to 5 CKD (confirmed CKD) and the majority of these had CKD stage 3a (79 505, 73%), followed by CKD stage 3b (24 018, 22%), then CKD stage 4 (5010, 4.6%) and lastly CKD stage 5 (872, 0.8%).(Figure 3-2)

Figure 3-1. Flowchart of how patients were defined with CKD for QOF analysis cohort

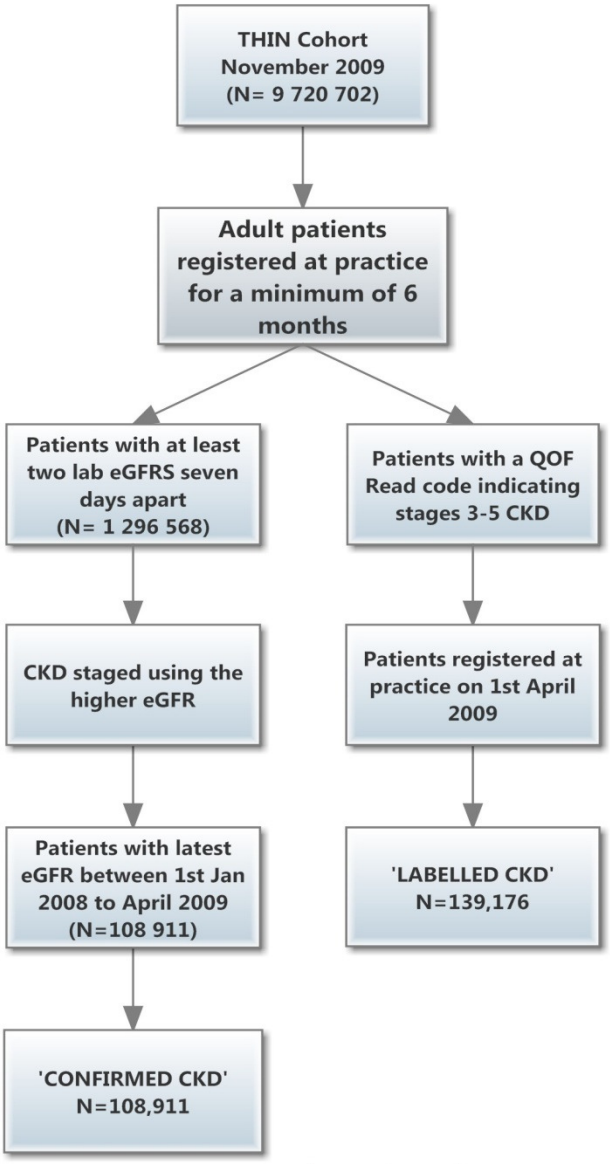
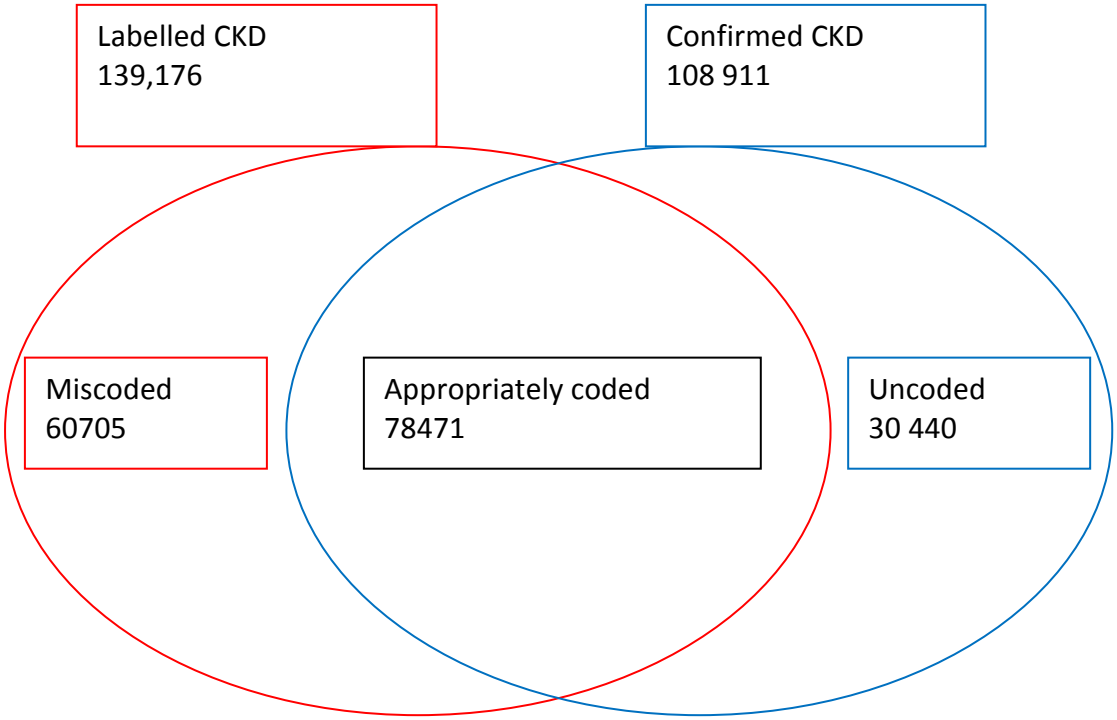


**Figure 3-2. Proportion of patients with biochemical CKD versus classified CKD**




In 2009, 139 176 (5.1%) appeared on practice registers (labelled CKD) (Figure 19). Of those with biochemically confirmed CKD, 78 471/108 911 (72%) were labelled with a Read code for CKD 3-5. A further 60 705/139 176 (44%) of labelled CKD patients had a Read code but no biochemical evidence of CKD (i.e. were miscoded) and 30 440/108 911 (28%) fulfilled biochemical criteria for CKD but were not on a practice CKD register (uncoded CKD).

#### *3.4.1.1.Characteristics of confirmed and labelled CKD patients*

The demographic and clinical characteristics of the general population compared to labelled or confirmed CKD patients are shown in Table 3-4. Patients with either labelled CKD or confirmed CKD in comparison with the general population were likely to be older, female, Caucasian, smoke and have more comorbidity. Uncoded patients in comparison with appropriately CKD patients were more likely younger, female, non-diabetic, non-hypertensive, have no cardiovascular disease and be non-smokers ($p < 0.01$). Uncoded CKD patients compared with miscoded patients were more likely to be older, women, be non-diabetic, hypertensive, have cardiovascular disease and smoke ($p < 0.01$).

**Table 3-4. Demographics of CKD patients between 1st January 2008 to 1st April 2009 (The proportions in brackets are crude percentages)**

| | Total Population | Patients with biochemical evidence of CKD | Practice CKD Register (with or without biochemical evidence)[b] | Patients on CKD register with biochemical evidence [c] | Patients on CKD Register but no biochemical evidence of Stages 3-5 CKD [d] | Patients with biochemical evidence of CKD not on Practice Register[e] |
|---|---|---|---|---|---|---|
| **True CKD status** | Baseline Population | Confirmed CKD | Labelled CKD | Appropriately coded CKD | Miscoded CKD | Uncoded CKD |
| **Number (crude %)** | 2 707 130 (100) | 108 911 (4.0) | 139 176 (5.1) | 78 471 (2.9) | 60 705 (2.2) | 30 440 (1.1) |
| **Age in years (median (Range))** | 47(19-114) | 79(19-106) | 76 (19-106) | 78 (19-106) | 71(19-105) | 76(19-105) |
| **Female Gender (%)** | 1378 422 (51) | 67 352 (62) | 82 949 (60) | 47 666 (60) | 35 283 (58) | 19 686 (65) |
| **Ethnic Group[f]** | | | | | | |
| **Black** | 11 329 (0.40) | 528 (0.48) | 592 (0.43) | 324 (0.41) | 268 (0.44) | 204 (0.67) |
| **White** | 355121 (13) | 25 381 (23) | 29 820 (21.4) | 18 764 (23) | 11 056 (18) | 6617 (21) |
| **South Asian (%)** | 20 032 (70) | 594 (0.54) | 1051 (0.76) | 468 (0.59) | 583 (0.96) | 126 (0.41) |
| **Diabetes Mellitus[g]** | 174 622(6) | 26 985 (25) | 33 083(23) | 20 970(26) | 12 113(20) | 6015(19) |

| True CKD status | Population | Confirmed CKD | Labelled CKD | Appropriate CKD | Miscoded CKD | Uncoded CKD |
|---|---|---|---|---|---|---|
| **Hypertension** | 413 235 (16) | 63 917 (59) | 73 037 (52) | 48 587 (62) | 24 450 (40) | 15 330 (50) |
| **CHD*[g]** | 158 701 (6) | 31 234 (29) | 35 471 (25) | 24 254 (31) | 11 217 (18) | 6980 (23) |
| **PVD[g]** | 35 033 (1) | 8330 (8) | 9011 (6) | 6656 (8.5) | 2355 (4) | 1674 (5.5) |
| **Stroke[g]** | 67 075 (3) | 14 339 (13) | 15 370 (11) | 11 020 (13.9) | 4350 (7) | 3319 (11) |
| **Hyper-cholesterolaemia** | 95766 (4%) | 12959 (12%) | 15298 (11%) | 9949 (13%) | 5349 (9%) | 3010 (10%) |
| **Ever Smoked** | 446 034 (17) | 20 805 (19) | 24 301 (17) | 15 296 (19) | 9005 (15) | 5509 (18) |

Crude preparations in brackets

a. Patients with two consecutive eGFRS under 60 at least seven days apart
b. Patients with a QOF business rule Read code for stages 3-5 CKD
c. Patients with 2 consecutive eGFRS <60 at least 7 days apart and a QOF business rule Read code
d. Patients with a QOF business rule Read code for CKD but no sustained eGFRs below 60 coded in 2008-2009
e. Patients with 2 consecutive eGFRS < 60 at least 7 days apart but no QOF Read code for CKD
f. Where reported as not all patients have Ethnic group reported.
g. Diabetes Mellitus – Type 1&2, CHD = Coronary Heart Disease, PVD = peripheral vascular disease

All demographic clinical features were either coded or recorded in the dataset prior to the 1st April 2009

* When comparing groups d and e or c and e, p <0.0001 using Pearson chi-square test

### *3.4.1.2.Multivariable analysis of predictors of exclusion and inclusion on Practice CKD register*

#### 3.4.1.2.1.Uncoded versus appropriately coded

In the multivariable analyses, the independent variables associated with exclusion from the practice register i.e. the odds of being uncoded patients versus appropriately coded) were younger age, female gender, and reduced co-morbidity (Table 3-5 and Figure 3-3)

**Table 3-5. Multivariate logistic regression model for significant predictors for exclusion from the CKD QOF register: Uncoded vs. appropriately coded**

| Risk Factor | Odds Ratio of Uncoded CKD compared to appropriately Coded CKD<br>Odds Ratio (95 % CI) |
|---|---|
| **Age**[I] | 0.991(0.990-0.993) |
| **Female sex**[II] | 1.20(1.16-1.24) |
| **CKD stage**<br>**3a** [III]<br>**3b**<br>**4**<br>**5** | <br>1*<br>0.37(0.35-0.38)<br>0.24(0.22-0.27)<br>0.24(0.19-0.31) |
| **Coronary Heart Disease**[iv] | 0.81(0.76-0.86)* |
| **Hypertension**[iv] | 0.61(0.59-0.63)* |
| **Diabetes Mellitus** | 0.72(0.69-0.75)* |
| **Cardiovascular Disease (composite of Coronary Heart Disease, Peripheral Vascular Disease and Stroke)** | 0.83(0.78-0.88)* |
| **Peripheral Vascular Disease**[iv] | 0.88(0.82-0.95)* |
| **Hypercholesterolaemia**[iv] | 0.80(0.76-0.84)* |

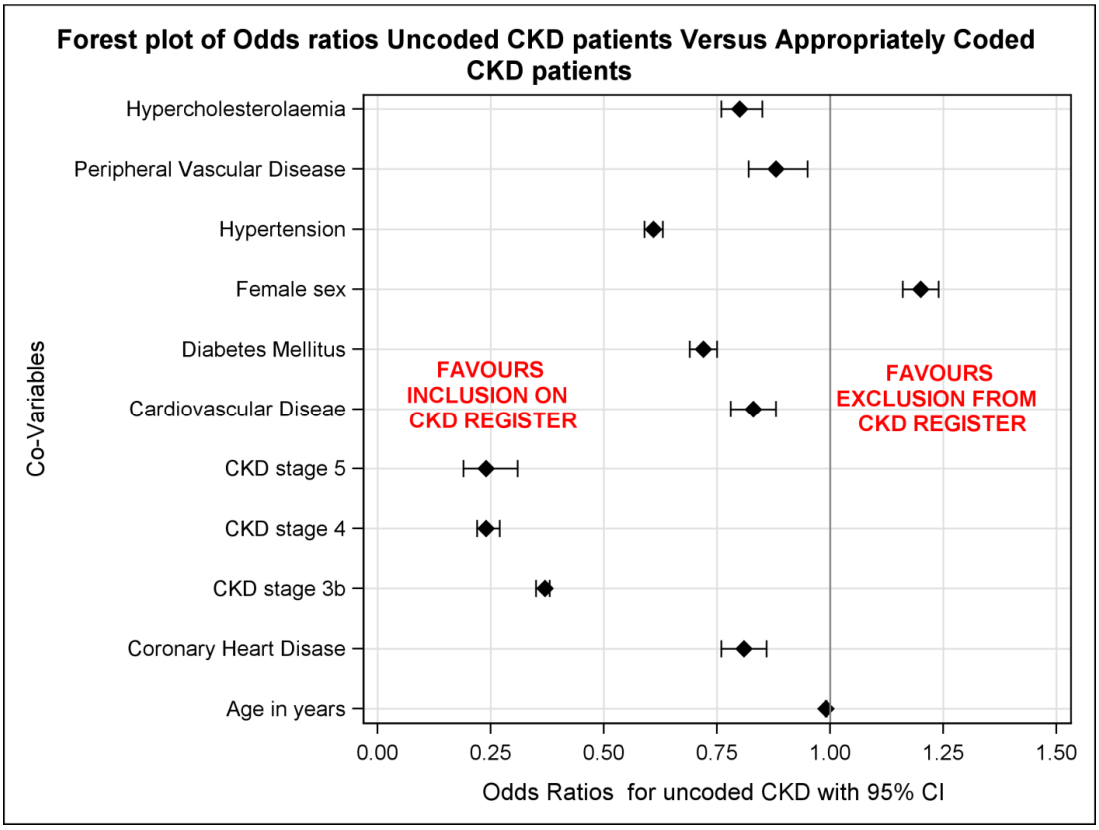
i. For an increase in years from the mean age

ii. In comparison with male gender

iii. Indicates the reference indicator

iv. The presence of the disease in comparison to those without it * $p < 0.0001$

**Figure 3-3. Forest plot of odds ratios: uncoded CKD versus appropriately coded CKD**




Note stage 3b, 4, 5 CKD are compared to stage 3a CKD

#### 3.4.1.2.2.Uncoded versus miscoded

Uncoded patients were more likely to be older, have hypertension, have cardiovascular disease and to smoke compared to miscoded patients. They were less likely to have diabetes, coronary heart disease and hypercholesterolaemia. (Table 3-6 and Figure 3-4)

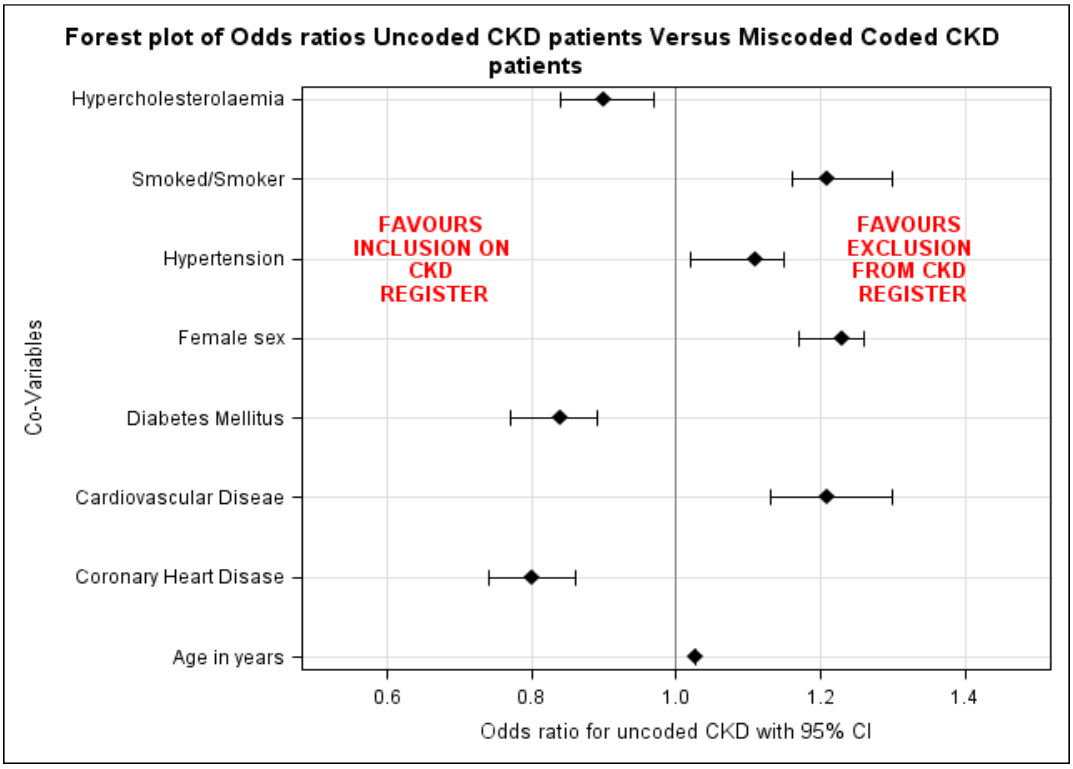
**Table 3-6. Multivariate logistic regression model for significant predictors for exclusion from the CKD QOF register: Uncoded vs. Miscoded**

| Risk Factor | Odds Ratio of Uncoded CKD compared to miscoded CKD<br>Odds Ratio (95% CI) |
|---|---|
| **Age**[i] | 1.027(1.026-1.028)* |
| **Female sex**[ii] | 1.23(1.17-1.26)* |
| **Coronary Heart Disease**[iv] | 0.80(0.74-0.86)* |
| **Hypertension**[iv] | 1.11(1.02-1.15)* |
| **Diabetes mellitus** | 0.84(0.77-0.89)* |
| **Cardiovascular Disease** | 1.21(1.13-1.30)* |
| **Peripheral Vascular Disease**[iv] | NS |
| **Hypercholesterolaemia**[iv] | 0.90(0.84-0.97)* |
| **Smoking**[iv] | 1.21(1.16-1.30)* |

* $p < 0.0001$

i. For an increase in years from the mean age
ii. In comparison with male gender
iii. Indicates the reference indicator
iv. The presence of the disease in comparison to those without it. The disease was ascertained by having a Read code for the disease

**Figure 3-4. Forest plot of odds ratios: uncoded CKD versus miscoded CKD**




### 3.4.2.QOF indicators: Comparison of Labelled and Confirmed CKD

The management of CKD patients according to QOF is shown in Table 3-7. There was a trend for the confirmed CKD group of patients to have worse management compared to those with labelled CKD.

There was significant underachievement in QOF outcomes in the uncoded CKD group versus the coded CKD group where CKD patients not on the practice CKD register were less likely to have a recorded blood pressure, a blood pressure on target or have a recorded ACR. Similarly uncoded patients were significantly less likely to achieve CKD QOF indicators in comparison to miscoded CKD patients.

The mean blood pressure was significantly different in patients in with uncoded CKD (mean systolic 136.2 mmHg, 95% CI 136.0 to 136.4 mm Hg/mean diastolic 75.8mmHg 95% CI 75.7 to 75.9 mmHg, $p < 0.0001$ ) compared to those appropriately coded (mean systolic 136.0 mmHg, (95% CI 135.9 to 136.1 mmHg/mean diastolic 75.0 mmHg 95% CI 75.0 to 75.1 mmHg, $p < 0.001$). The cholesterol was likely to be lower in the uncoded CKD group versus the coded CKD group (median cholesterol 5.0 vs. 5.1 mmol/l, $p < 0.005$). Mean systolic blood pressure, in mmHg, was significantly higher in patients with uncoded CKD (136.2 (95% CI 136.0 to 136.4)/75.8 (75.7 to 75.9) compared to miscoded CKD 134.5 (134.3 to 134.6)/ 76.3 (75.0 to75.1), $p <0.0001$.

Cholesterol was measured in 92% of individuals regardless of group. There was no difference in serum cholesterol level between miscoded and uncoded CKD (median cholesterol 4.7 vs. 4.7 mmol/l, $p =0.9455$). All patients with proteinuria as defined by an ACR ≥30 mg/mmol and with a Read code for hypertension (n=1018) were on angiotensin blockade regardless of CKD coding (Table 3-7).

**Table 3-7. CKD Management according to QOF standards between 1st January 2008 to 1st April 2009 (Proportions quoted are of the total in that group)**

| | Patients with biochemical evidence of CKD[a] | Practice CKD Register (with or without biochemical evidence)[b] | Patients on CKD register with biochemical evidence [c] | Practice CKD Register but no biochemical evidence of Stages 3-5 CKD [d] | Patients with biochemical evidence of CKD not on Practice Register[e] |
|---|---|---|---|---|---|
| **True CKD status** | Confirmed CKD[a] | Labelled CKD[b] | AppropriateCKD[c] | Miscoded CKD[d] | Uncoded CKD[e] |
| **Total Number** | 108 911 | 139 176 | 78 471 | 60 705 | 30 440 |
| **Proportion with Blood pressure in last 15 months** | 104 213(96%) | 132 343(95%) | 76 602(98%)[1] | 55 741(92%)[2] | 27 611(91%) |
| **Patients whose last BP is less than 140/85 in last 15 months** | 57 169 (52%) | 73 852 (53%) | 42986 (55%)[1] | 30866 (51%)[2] | 14 183 (47%) |
| **Patients with CKD who have an ACR** | 19 483(18%) | 25 445(18%) | 15 806(20%)[1] | 9639(16%)[2] | 3677(12%) |
| **Patients with CKD who have proteinuria** | 1358(1.2%) | 1474(0.9%) | 1153(1%)[1] | 320(1%)[2] | 205(1%) |
| **Patients with Hypertension and proteinuria (ACR > 30) on angiotensin blockade** | 1018(100%) | 1073(100%) | 883(100%)[1] | 190(100%)[2] | 135(100%) |

(a) Patients with two consecutive eGFRS under 60 at least seven days apart (b) Patients with a QOF business Rule Read code for CKD (c) Patients with two consecutive eGFRS under 60 at least seven days apart and a QOF business rule Read code (d) Patients with a QOF business Rule Read code for CKD but no sustained eGFRs below 60 (e) Patients with two consecutive eGFRS under 60 at least seven days apart but no QOF business Rule Read code for CKD

1. When comparing groups c and e, $p < 0.0001$ 2.When comparing groups d and e, $p < 0.0001^{+}$ QOF business rules look back 15 months hence time period

#### *3.4.2.1.Comparison of Diabetic QOF indicators in CKD patients*

In exploratory analyses of the subgroup of patients with diabetes and CKD (Table 3-8), similar patterns were observed where patients in the confirmed CKD group were less likely to have their BMI, HbA1c, and ACR measured in comparison to the labelled CKD group. These differences were highly significant when comparing the coded CKD and uncoded CKD group with the exception of blood pressure management. Similarly CKD management was better in miscoded patients versus uncoded patients with the exception of attaining the lower blood pressure target of 135/85 mmHg. There was a trend for better management in patient labelled and therefore on the CKD register compared to those with confirmed CKD.

**Table 3-8. Management of Diabetic CKD patients**

| True CKD status | Labelled CKD[a] | Confirmed CKD[b] | Appropriate CKD[c] | Miscoded CKD[d] | Uncoded CKD[e] |
|---|---|---|---|---|---|
| **Number** | 33083 | 26985 | 20970 | 12113 | 6015 |
| **Patients with a BMI in the last 15 months** | 30287(91%) | 24301(90%) | 19209(91%)[1] | 11258(93%)[4] | 5272(88%) |
| **Patients with a HbA1c in the last 15 months** | 29558(89%) | 23953(88%) | 18710(89%)[1] | 10848(90%)[4] | 5243(87%) |
| **Patients with a HbA1c ≤ 10** | 28119(85%) | 22728 (84%) | 17790(85%)[1] | 10329(85%)[4] | 4938(82%) |
| **Patients with a HbA1c ≤ 7** | 15734(48%) | 12673(47%) | 9915(47%)[2] | 5819(48%)[4] | 2758(46%) |
| **Patients with ACR in the last 15 months** | 19624(59%) | 15397(57%) | 12180(58%)[1] | 7444(61%)[4] | 3217(53%) |
| **Patients with a cholesterol ≤ 5** | 25932(78%) | 20896(77%) | 16464(79%)[1] | 9471(78%)[4] | 4432(73%) |
| **Patients with a BP ≤ 140/85** | 19152(60%) | 15455(57%) | 12125(58%)[1] | 7027(58%)[4] | 3330(55%) |
| **Patients with a BP ≤ 130/85** | 10215(31%) | 8289(31%) | 6469(31%)[3] | 3746(31%)[5] | 1820(30%) |

(a)Patients with a QOF business Rule Read code for Chronic Kidney Disease (b)Patients with two consecutive eGFRS under 60 at least seven days apart (c) Patients with two consecutive eGFRS under 60 at least seven days apart and a QOF business rule Readcode (d)Patients with a QOF business Rule Read code for Chronic Kidney Disease but no sustained eGFRs below 60 (e)Patients with two consecutive eGFRS under 60 at least seven days apart but no QOF business Rule Readcode for CKD

1. When comparing groups cand e, p <0.001
2. When comparing groups c and e, p =0.05
3. When comparing groups c and e, p =0.38
4. When comparing groups d and e, p <0.001
5. When comparing groups dand e, p =0.35

## 3.5.Discussion

### 3.5.1.Summary of Key Findings

The age and gender standardised prevalence of labelled CKD (i.e. a QOF Read code for stages CKD 3-5) rose from 0.11% in 2005 to 5.05% in 2009. In 2009 out of 139 176 many labelled CKD patients, 78 471 (56.0%) of patients were appropriately coded and 60 705 (44%) were miscoded as having CKD. In 2009, 4% (108 911) of the total population had actual or 'confirmed' CKD (based on 2 lab e-GFRs) and of these 28% (30 440) were uncoded i.e. not labelled with CKD by a Read Code on the computer system of the practice.

Patients with CKD were less likely to be appropriately coded if they were younger, female and had less co-morbidity such as hypertension, diabetes mellitus, cardiovascular disease or hypercholesterolaemia. When comparing patients with uncoded CKD to miscoded CKD patients, uncoded patients were more likely to be older, female, have hypertension, cardiovascular disease and smoke. However they were less likely to have diabetes, coronary heart disease or hypercholesterolaemia.

There were significant differences in the management of patients according to their coding: CKD patients who were appropriately coded were more likely to have their blood pressure checked, have their blood pressure on target and have an ACR compared to uncoded patients. Similarly management was better in patients with miscoded CKD versus uncoded CKD. In patients with diabetes and CKD, again those uncoded for CKD had worse management and were less likely to have a BMI or ACR recorded, have a HbA1c of less than 7 and 10 %, and have blood pressure to target in

comparison to CKD patients with a Read code (whether or not appropriate). The only exception was having a lower blood pressure target of 130/85 mmHg which was similar amongst all groups.

### 3.5.2.Interpretation of results and comparison to existing literature

These results suggest that in 2009 there was considerable inaccuracy in the practice CKD register. A substantial proportion of patients were inappropriately placed on the register, whilst many with a CKD diagnosis were omitted from the register. These findings were consistent with another study examining the sensitivity and specificity of CKD diagnosis by Read code in THIN.[186] They found poor sensitivity, approximately 50% but good specificity of 90%. However they used non QOF Read codes for CKD to define CKD therefore coding for CKD would not lead to patients being recognised for incentivised management and therefore my work is novel.[186] Additionally 'undercoding' patients with CKD exists in other health care systems, for example in US Veteran hospitals where patients biochemical evidence of CKD may not have the coding for the latter.[187] In patients with CKD, it was the younger, 'healthier' patients who were at higher risk of being unrecognised. This may be because CKD is unexpected in this cohort. However despite these patients lacking the usual risk factors for CVD, the CKD prognosis consortium, which adjusted for comorbidity, demonstrated that younger patients with CKD were still at greater risk for all-cause mortality (HR 2.79 for patients aged 18-54 with CKD stage 3 compared to those with eGFRs 75-89 ml/min/1.73m$^2$). When comparing patients with uncoded CKD to miscoded CKD, patients with uncoded CKD were still likely to be at risk of

cardiovascular disease and mortality as they were more likely to be hypertensive and smoke.

Though not interpreted as causal, patients with uncoded CKD were less likely to achieve QOF management targets in comparison with those on the register. This could suggest that GPs are only including patients more likely to achieve QOF indicators but also could signify that uncoded patients will not be automatically recalled for health checks. Further research to better understand the reasons for these results is needed. Patients who are excluded from QOF demonstrate that GPs need to direct resources to appropriately identify patients with CKD.

Potential reasons for miscoding could be due methodological issues. It was not possible to tell when the code had lapsed and therefore practices may have amended the CKD diagnosis but I was unable to account for this in my analysis and therefore inflating the numbers of those who were miscoded. Other reasons were that practices may have calculated the GFR and determined patients with CKD by this method. However only another 0.77% of the population had CKD if calculated eGFRs were used to stage CKD 3-5 and the proportion of miscoded patients were 2.2%. Therefore hypothetically at least 1.43% of the population would have still remained uncoded. Another hypothesis is that patients are diagnosed with CKD by the practice based on a single blood result. Looking at the latest result in 2009 in patients with only one previous blood test, in the miscoded cohort, another 25.1 % of this cohort (n=15247, 0.56% of the denominator population) had a lab eGFR below 60 ml/min/1.73m$^2$.[188] These findings are consistent with evidence that the management in patients with type

2 Diabetes, excluded from the practice register was worse.[141] These results are also comparable to a US study where CKD patients without CKD (ICD) codes on hospital administrative systems were less likely to have markers for renal complications such as anaemia and renal bone disease checked.[154] Being labelled with disease may incur harm as well as benefit with loss of income reported in patients labelled with hypertension.[189] However on a positive note, the goals achieved in the UK are generally higher than some western cohorts.[190]

### 3.5.3.Strengths and limitations

It cannot be definitively proved by this thesis that CKD misclassification leads to poorer management as this was a retrospective cohort study. Patients may have better CKD classification due to unrecorded confounders that are associated with CKD misclassification and subsequent management. For example patients with hypertension are likely to receive angiotensin blockade and better CKD recognition as GPs should check patient's renal function after instituting this.

As mentioned earlier, patient information prior to registration at the practice and pre AMR date was excluded. This may have resulted in patients being misclassified as miscoded. Additionally it may have led to the loss of co-variables in the multivariable analysis. A further limitation is that patients without a Read code were assumed not to have the disease or condition. This seems a reasonable assumption given that smoking status and cardiovascular reporting in Primary care databases is reasonably accurate.[137;138;144;146] There may be unmeasured confounders that explain why management is worse in miscoded CKD patients.

This cohort had several advantages: firstly it was a large cohort representative of the general population.[136] Patients excluded from practice registers were included and therefore their management could be assessed. Finally THIN and other primary care databases lack recall bias because the data do not rely on patients or researchers to provide information.

## 3.6.Executive Summary

- Various international and national guidelines exist for CKD including NICE guidance in the UK.
- The UK QOF incentivized management for CKD during the time period included in the analysis.
- The impact of QOF coding on the management of CKD was previously unknown.
- The CKD practice register is inaccurate as many patients with CKD are undiagnosed as having CKD and many patients are misclassified as having CKD without any evidence of CKD.
- Uncoded CKD patients are likely to be younger and have less comorbid conditions in comparison with appropriately coded CKD patients.
- Uncoded CKD patients are likely to be older than miscoded CKD patients.
- Uncoded CKD patients are less likely to attain QOF targets.

# CHAPTER 4. WHAT FACTORS PREDICT ALL-CAUSE MORTALITY IN STAGE 3 CKD PATIENTS? A PROGNOSTIC MODEL

This chapter describes risk factors for mortality in patients with stages 3-5 CKD according to contemporary literature. The chapter then describes the development of a prognostic model which aims to identify potential risk factors associated with all-cause mortality using data routinely collected in primary care. The results may be used by general practitioners to target care to those at highest risk of an event.

## 4.1.Introduction

### 4.1.1.Cardiovascular disease and mortality in CKD patients

Cardiovascular (CVD) risk factors are abundant in patients with CKD because cardiovascular factors are associated with the development of CKD(reverse causality). Additionally the relationships between CVD risk factors and cardiovascular disease in patients with CKD are complex (eg raised blood pressure, cholesterol and body mass index) and may not always result in cardiovascular sequelae.[191] Furthermore renal failure results in other non-atherogenic cardiovascular abnormalities such as left ventricular diastolic dysfunction, left ventricular hypertrophy and arteriosclerosis.[192] The reasons for these include malnutrition, anaemia, hyperhomocysteinuria, bone mineral disorders and increased inflammation.[191] These risk factors also aggravate atherosclerosis. The following section will discuss the impact of traditional and non traditional risk factors on cardiovascular disease in CKD patients.

#### *4.1.1.1.Diabetes and other co-morbid conditions*

Risk factors such as hypertension, diabetes and atrial fibrillation are implicated in mortality and CVD in patients with CKD. Hypertension and blood pressure control are discussed in a separate section below. Diabetes is a known risk factor for CKD and CVD in the general population.[193] The Framingham study found that men with diabetes had double the risk of CVD compared to non-diabetics and this risk was three fold higher for women after 20 years follow-up.[194] Not only is diabetes implicated in CVD but it is directly associated with the development of CKD.

CKD due to diabetes mellitus is labelled as 'diabetic nephropathy'. Diabetic nephropathy is associated with both type 1 (T1DM) and type 2 Diabetes Mellitus (T2DM).[195] In a study of patients with T1DM, diabetic nephropathy developed in 25% of the population in a conventionally treated arm. In patients who had been treated with intensive insulin treatment, the proportion was lower at 17%.[106;107] Therefore diabetes is implicated both in the cause of CKD and the development of CKD. It is worth noting that although CKD patients with diabetes are more likely to be obese, have hypertension and lipid abnormalities, diabetes is still an independent risk factor for mortality across all stages of CKD.[196]

Atrial Fibrillation (AF) is present in up to 15% of patients with CKD and is a known risk factor for stroke.[126;197;198] Risk stratification for anticoagulation in CKD patients is complicated as they are more likely to bleed, but in a subgroup analysis of CKD patients in the SPAFIII trial, ischaemic stroke was reduced by 76% in patients with CKD, randomised to warfarin versus placebo.[199] There are, however, no specific trials of warfarin therapy in CKD patients with AF.

#### *4.1.1.2.Hypertension and proteinuria: A complicated CVD risk factor in CKD patients*

Hypertension is one of the commonest risk factors for CVD in the general population and is defined as a systolic blood pressure and diastolic blood pressure greater than 140 mmHg and 90 mmHg respectively. In the UK the Health Survey of England (HSE) in 2010 found a prevalence of hypertension of 31.5% in Men and 29.0% in women.[200] The prevalence of hypertension has risen considerably and the current prevalence of hypertension was not anticipated until 2025.[201] In a large epidemiological meta-

analysis of people with cardiovascular disease (i.e. stroke, ischaemic heart disease and other vascular disease) the risks of further cardiovascular events doubled for a rise in a systolic and diastolic blood pressure of 20 and 10 mmHg respectively.[202]

The prevalence of hypertension in CKD patients is much higher. In the NHANES screening study: between 1999 and 2004, the proportion of patients with hypertension with stages 1-5 CKD was 75-85%.[203] There is an inverse linear relationship between GFR and systolic blood pressure as demonstrated by the MDRD study.[204] Furthermore development of CKD can lead to de novo hypertension and worsening hypertension.[205]

The aetiology of hypertension in CKD patients may be different from the general population due to increased arterial stiffness,[206] salt and water retention,[207] increased activity of the renin angiotensin system due to local scarring and increased sympathetic activity.[208] Furthermore, secondary hyperparathyroidism, treatment with erythropoietin and dysregulated nocturnal blood pressure contribute to hypertension.[112;209] The relationship between hypertension in CKD and CVD is not linear but more J-shaped: individuals with a blood pressure of 120mmHg or less have an equivalent risk of death to those with a blood pressure of 180mmHg.[207;210] This may be due to reverse epidemiology where hypertension may result in greater blood flow to failing nephrons.[210]

However hypertension as a risk factor for CVD in patients with CKD and BP control is confounded by proteinuria. In next section the management of hypertension in CKD patients and the impact of proteinuria will be discussed.

The majority of antihypertensive trials that have examined cardiovascular risk reduction in CKD patients are derived from post hoc analysis of subgroups with CKD from trials in the general hypertensive population or from diabetic nephropathy trials. Unfortunately unless such subgroup analysis has been specified a priori it is difficult to ascertain a causal association between intervention and outcome. Any results are hypothesis generating and should be interpreted with caution. As evidence for diabetics and non-diabetics with CKD differs it will be discussed separately.

#### 4.1.1.2.1.Diabetic Nephropathy

Multiple trials have shown that antihypertensive treatment in diabetic nephropathy reduces renal morbidity, cardiovascular morbidity and all-cause mortality.[119;211;211-217] The Reduction Of Endpoints In NIDDM With The Angiotensin II Antagonist Losartan (RENAAL) study, where patients with diabetic nephropathy were randomised to either Losartan or placebo showed a 16% reduction in Renal Events (a composite outcome of ESRD or doubling of serum creatinine) and that an increase of blood pressure of 10mmHg led to an increase in the absolute risk of ESRD or death of 6.7% (p=0.007).[212;218] This study also found that the risk of first hospitalization for heart failure was reduced by 32 % (p=0.005), and though there was no reduction in other cardiovascular events, there was an increased risk of CVD associated with albuminuria.[219] Similar findings were observed in the Irbersartan Diabetic Nephropathy Trial (IDNT) with a 20 % reduction in renal morbidity in patients randomized to Irbersartan versus placebo as well as a reduction in heart failure and cardiovascular mortality.[116;211]

#### 4.1.1.2.2.Non Diabetic Chronic Kidney Disease

##### *Subgroup Analysis of larger Blood pressure lowering trials*

Unfortunately there are no specific randomized control trials looking at whether antihypertensive treatment reduces CVD in non-diabetic CKD patients.[220] In a post hoc subgroup analysis of the Heart Outcomes Prevention Evaluation Study (HOPE) (Ramipril vs. placebo), renal dysfunction was found to be an independent predictor of CVD.[221] In the CKD subgroup, ramipril was associated with a reduction in cardiovascular death and a reduction in all-cause mortality by almost a third. Similarly in a CKD subgroup analysis of the "Avoiding Cardiovascular Events through Combination Therapy in Patients Living with Systolic Hypertension" study there was reduction in pre-specified renal event (doubling of serum creatinine) with Bezapril and Amlodipine.

##### *The MDRD and the AASK study*

Two landmark trials have examined blood pressure reduction in non-diabetic CKD patients. In the MDRD study of 840 CKD patients (predominantly non-diabetic CKD), were randomised to either a low 'BP' or 'standard' BP. They achieved a systolic blood pressure of 125.6 and 132.7 mmHg in the low versus standard blood pressure groups respectively. There was no difference in renal sequelae in either blood pressure group apart from patients with significant proteinuria and this may be explained by higher angiotensin blockade use in the lower BP group. In a longitudinal follow up study it was suggested that there were lower rates of death and ESRD in the 'low BP' group (HR [0.68 (95% CI 0.57 -0.82)].

In the African-American Study of Kidney Disease and Hypertension (AASK), African-American patients with hypertensive kidney disease were randomised to similar blood pressure targets as the MDRD study.[222] There was no difference in renal events between the two groups. However in the subsequent longitudinal observational study, patients assigned to a lower BP had lower rates of ESRD or doubling of creatinine in those patients with proteinuria (Urine protein excretion > 300mg/day).[220] In both longitudinal studies of AASK and MDRD caution must be taken in interpretating the results as the low blood pressure and normal blood pressure group were no longer randomised.

As the studies in diabetic and non diabetic CKD show that treating hypertension is beneficial in most groups it is worth discussing what is the optimal blood pressure target in CKD patients.

#### 4.1.1.2.3.The effect of target blood pressure

In the UK, the most recent published guidelines are from the National Institute of Clinical Excellence (NICE).[161]

They specify the following targets:

Less than **140/90 mmHg** for people with non-diabetic non proteinuric CKD

Less than **130/80 mmHg** for people with diabetes or proteinuria (Albumin Creatinine Ratio above 70)

Target blood pressure in CKD is not a straightforward topic. Even in the general population, many previous trials did not achieve a low blood pressure on treatment.

For example, the mean blood pressure achieved in Systolic High blood pressure and Elderly Program, was 143/68 mmHg and in the large Hypertension and Outcomes Trial (HOT) there was no risk difference seen between those randomly assigned to a diastolic blood pressure of less than 80, 85 and 90 mmHg apart from the diabetic sub group. The blood pressures achieved in the HOT trial were 144/85, 141/83 and 140/80 respectively.[114;119]

In Diabetes, where there has been a reduction in microvascular or macrovascular events, the blood pressures achieved varied, with many higher than 130/80mmHg; for example in the UKPDS trial the target blood pressure was 154/82 mmHg achieved; the HOPE study achieved a mean systolic blood pressure of 134.5 mmHg and the RENAAL achieved a blood pressure of 146/78 mmHg in the treatment group.[212-214] In the IDNT trial, a systolic blood pressure below 120 mmHg was associated with the same cardiovascular risk as that for patients with a systolic of 180 mmHg. Similarly, a meta-analysis suggested a blood pressure of below 110 mmHg was associated with worse renal outcomes.[223] It should be noted that only a few studies have compared specific blood pressure targets and they have not shown a benefit below 130/80 mmHg in their main analysis.[119;217;222;224] A recent Cochrane review looking at target blood pressure to reduce cardiovascular and microvascular complications in the general population, evaluated 11 studies and found no benefit beyond lowering a blood pressure below 140/90 mmHg.[225]

#### 4.1.1.2.4.The impact of different classes of blood pressure medication

In diabetic nephropathy, the majority of trials have shown a reduction in both cardiovascular and renal morbidity with angiotensin blockade and predominantly angiotensin II blockers.[211;218;226;227] Similarly in non diabetic CKD the REIN studies showed that decline in renal function was reduced in those treated with Ramipril.[228] In the AASK study a risk reduction of 22% (95% CI 1-38%) of renal morbidity was seen in the Ramipril cohort.[222] These effects seem to be independent of blood pressure as confirmed by two meta-analyses evaluating cardiovascular disease and progression of renal disease.[117;229] Angiotensin blockade appears superior to calcium channel blockers and beta blockers. Dual blockade of the Renin Angiotensin system is not beneficial and may be harmful: the ONTARGET study considered cardiovascular risk reduction in high risk patients by dual blockade but found an increase in adverse events as well as lack of benefit from this.[230]

Which adjunctive therapy is best is uncertain. Adding a calcium channel blocker to a ACE inhibitor in the REIN 2 study added no benefit in the treatment of CKD patients[231] but this conflicted with the ACCOMPLISH trial that may not have had the appropriate endpoint.[220;232] Direct renin inhibition may reduce risk as it has been shown to reduce urine albumin excretion in patients with diabetic nephropathy on losartan.[233]

#### 4.1.1.2.5.Is treatment of blood pressure related to reduction in proteinuria?

Treatment of hypertension in CKD is likely to treat proteinuria as well. Even small amounts of proteinuria are associated with higher rates of morbidity and mortality. In

the collaborative meta-analysis, independent of blood pressure and eGFR patients with an ACR of 1.1mg/mmol, 3.43mg/mmol and 33.9mg/mmol were likely to have a hazard ratio for all cause mortality 1.2 (95% CI 1.15-1.26 ), 1.63 (1.50-1.77) and 2.22 (1.97-2.51) compared to an ACR of 0.6mg/mmol. As discussed before most of the trials have taken place in patients with diabetic nephropathy and in the RENAAL trial cardiovascular risk fell by 18% if albumininuria fell by 50%.[219] In the MDRD and AASK study, blood pressure reduction was only effective if patients had proteinuria though these were subgroup analysis on very small number of patients. Another confounder is that most of the blood pressure studies in diabetic and non diabetic CKD that show reduction in cardiovascular events have used angiotensin blockade which reduce proteinuria directly as well as reducing blood pressure.[117] The only study to use an alternative or additional agent in CKD patients was the negative REIN 2 study.[231]

#### *4.1.1.3.Antiplatelet agents*

Aspirin and other anti-platelet agents have been shown to prevent recurrent cardiovascular disease in the general population. However there is a risk for increased bleeding.[234] CKD patients are at increased risk of bleeding due to platelet dysfunction and therefore at greater risk of harm then benefit.[235] There are no specific randomized control trials conducted for CKD patients and the available evidence comprises of post hoc analysis of other studies. In subgroup analyses of the HOT study, patients were stratified to categories of either a GFR of greater than or equal to 60ml/min/1.73m$^2$ (n=14,977), CKD stage 3a (n=3083) and CKD stage 3b (n=536). In patients randomized to Aspirin there was a reduction in all forms of cardiovascular disease in CKD 3b patients but not CKD stage 3a, but CKD 3b patients

was a very small subgroup. In general there were increased risks of minor bleeding.[110]

Further clarification came from a meta-analysis of 9969 CKD patients. Treatment with antiplatelet agents reduced the composite outcome of fatal or non fatal myocardial infarction patients with stable cardiovascular disease but also increased the risk of bleeding. There was no reduction in all cause mortality, hospitalisation or stroke. In patients with unstable coronary disease, antiplatelet agents showed no benefit.[236] This meta-analysis did not delineate risk for specific CKD groups. Therefore further trials need to be conducted in CKD specific populations to determine whether there is a benefit across all CKD groups and whether any benefits outweigh the harm of bleeding.

#### *4.1.1.4.Hypercholesterolaemia/other lipid abnormalities*

Patients with CKD often have other lipid abnormalities apart from hypercholesterolaemia such as hypertryglyceridaemia due to decreased lipoprotein lipase activity, raised apoliprotein A levels due to reduced renal clearance and decreased HDL cholesterol.[237] CKD patients may also have reduced cholesterol due to malnutrition and that is why low cholesterol is associated with mortality i.e. reverse causality.[123;237] Though it is well established that HMG CoA-Reductase inhibitors (statins) (by reducing LDL cholesterol) reduce cardiovascular disease in the general population, only recently has a randomized clinical trial specific to CKD been conducted with adequately powered end-points to determine the impact of statins. The Study of Heart and Renal Protection (SHARP) study randomized 9270 to

simvastatin and ezetimibe or placebo.[121] Two thirds of the patients were not on dialysis and in the treatment arm there was a 0.85mmol/l reduction in LDL cholesterol and 17% reduction in the composite end point of non fatal myocardial infarction, stroke or revascularization events. However statins reduce proteinuria as well and whether it is a statin specific effect or true reduction in cholesterol is difficult to discern as lower cholesterol is not always associated with reduced morbidity and mortality.[238]

#### *4.1.1.5.Age, Sex, Ethnicity and Deprivation*

The prevalence of CKD is higher in the elderly population. (see section 7) It has been debated whether decreased GFR is a normal part of ageing. The CKD Prognosis Consortium examined the impact of different age groups on mortality and ESRD in the general population and high risk cohorts. The risk of death rose with increasing age and the number of extra deaths rose from 9.0 to 27.2 extra deaths per 1,000 person-years for patients aged 18–54 years and ≥75 years.(183) Men are more likely to die with CKD across all stages in comparison to women in the CKD prognosis consortium general population cohorts and there may be gaps in their treatment as prevalence of CKD is less in this cohort.[239] The prevalence of chronic kidney disease is variable in non Caucasians but African Americans or blacks have higher rates mortality.[54;240] South Asians and South East Asians with ESRD have lower rates of mortality in comparison to Caucasians despite age correction.[129;241]

Chronic Kidney Disease is more prevalent in more deprived areas in both secondary care or primary care populations.[63;242] Deprivation is more common in ethnic

minority groups[243] and even despite improving healthcare is still associated with increased mortality.[130]

#### *4.1.1.6.Lifestyle*

Smoking is associated with the development of CKD and like the general population is associated with cardiovascular disease and mortality in CKD patients.[112;113] Although obesity is more common in CKD patients and obesity is associated with increased prevalence of CKD, patients with higher BMI (even in morbidly obese patients) may have lower mortality rates and the cause of this is uncertain.[128;244]

#### *4.1.1.7.Non-traditional risk factors*

Reduced renal function results in reduced erythropoietin production even with moderately reduced renal function.[245] The relationship between haemoglobin may not be linear as both anaemia (defined as less than 13mg/dl in men and less than 12mg/dl in women) and very high haemoglobins are associated with cardiovascular events and mortality.[245;246] However intervention with Erythropoietin has not been beneficial in CKD patients and even harmful in some meta-analysis.[131;247] As renal function declines, bone mineral regulation becomes disordered and results in hyperphosphatemia, hypercalcemia or hypocalcaemia, decreased active Vitamin D and secondary hyperparathyroidism. Hyperphosphatemia is linked to vascular calcification, a risk factor for cardiovascular disease and in a recent meta-analysis was an independent risk factor for cardiovascular disease.[248] Phosphate binding drugs are used to reduce serum phosphate in CKD with hyperphosphatemia. In a recent study although the desired effect was achieved, phosphate binders resulted in worsening

vascular calcification but there was no change in CVD[249] Other modifications such as vitamin D supplementation and calcimimetics have uncertain effects and have not shown improvement in cardiovascular morbidity or mortality.[248;249]

## 4.2. Rationale in modeling for death in patients with stage 3 CKD

As discussed in Chapter 1.5, the mortality and morbidity is high patients with chronic kidney disease as many patients die of cardiovascular disease. The commonest stage of CKD is stage 3 and it is recommended that this cohort should be managed in Primary Care by NICE.[44;52] The previous section demonstrated that there was inconsistent evidence for that treatment of cardiovascular risk factors for patients with CKD and what demographic, clinical features and medication are associated with increased or decreased survival in patients with stage 3 CKD.

## 4.3.Methods

### 4.3.1.Model development: A review of the literature

Given the uncertainties described above, a prognostic model was generated to evaluate the relationship between non modifiable and modifiable risk factors and the composite outcome of CVD or all-cause mortality in CKD 3 patients. Modifiable risk factors included lowering blood pressure or using pharmacological agents such as aspirin or cholesterol lowering agents. The following is a review of literature of prognostic modelling that will be used in the methods.

When considering modelling risk the following have to be considered:[250]

- Model selection
- Variable selection and overfitting
- Conditional variable selection
- Whether the relationship between variable and outcome is linear
- Model fit
- Missing data
- Whether the model satisfies the assumption of the model
- Data clustering and random effects

#### *4.3.1.1.Model Selection*

A Cox proportional hazards analysis was chosen over multivariate logistic regression as patients in the dataset had variable follow up and censored data. Logistic regression was considered unsuitable since it does not allow for censoring.[251] A simple survival probability for a group of individuals or survivor function at a particular time, t, is:

$$\frac{Number\ of\ individuals\ with\ survival\ times\ \geq t}{Number\ of\ individuals\ in\ the\ dataset}$$

The above function when plotted yields a familiar survival graph known as the Kaplan Meier curve.[251] Survival analysis is unique since those individuals that do not experience the outcome of interest, are lost to follow up or alive at the end of the study period can be censored. There are several forms of censoring but in this study right censoring was used as it occurred after the individual had been entered into the study.[251] Univariate analysis such Kaplan Meier curves can be plotted for different groups in the study and compared using the log rank test; a non-parametric test. The log rank test assumes proportional hazards i.e. that the risk of events in each group is proportional throughout the study period.[109]

To evaluate the effect of multiple covariates a Cox proportional hazards model was selected. This is a semi parametric analysis and again is based on the assumption that there are proportional hazards.[109]

The hazards function $h(t)$ which can be interpolated from the survival curve, is the risk of dying or suffering an event at time, t. If there are X1 to Xp variables then this function is given by:

$$h(t) = h_0(t)\ \times \exp\left(b_1X_1 + b_2X_2 + \ \ldots\ldots + b_3X_3\right)$$

Where $h_0(t)$ is the model where there are no covariables.[251]

#### *4.3.1.2.Selection of co-variables and overfitting*

In order for any statistical model to fit the data properly it must include the appropriate explanatory co-variables. A pragmatic literature review was undertaken to identify potential items for inclusion as summarised in and described in the introduction to this chapter.

The potential for overfitting was considered during model development.[250] Overfitting is where there are models with too many degrees of freedom i.e. too many variables in the model. In general terms between 10 and 20 events per co-variable are required. This allows for less than 5% optimism (i.e. negligible overfitting).[250] Given the large number of events under consideration for this study problems of overfitting were considered negligible.

##### 4.3.1.2.1.Conditional variable selection

Conditional variable selection can take several forms but the commonest are forward conditional selection or backward conditional selection. In the former model, each variable is entered step by step and any co-variable that has a p value greater than 0.05 is discarded from the model. In a backward selection process all the co-variables are entered in the model and non-significant co-variables i.e. those with a p value less than 0.05 are excluded. This is done in a hierarchal fashion where the least significant factor is discarded and then 2nd least factor is discarded and so on until the final model contains only significant co-variables. Forward conditional selection does not

allow us to assess all the co-variables together and interactions may be missed thus backward selection was selected as the method of choice for this study.[252;253]

#### *4.3.1.3.Non-linearity of variables*

A further consideration when entering continuous co-variables into a Cox regression analysis, is whether the relationship between the co-variable and hazard ratio is linear.[184] As discussed earlier in this chapter the relationship between blood pressure or BMI and risk may be non-linear. Several strategies can be applied to deal with this issue. Firstly the co-variable can be grouped into categorical variables, however this strategy can lead to loss of information and may not display the true form of the data.[254]

Secondly data can be transformed: the simplest transformations are first degree polynomials where the variable can be raised to a particular power for example it can be squared [158], cubed or most commonly the natural log can be taken. So instead of variable X, the transformed variable in the form of $X^2$ or Log(X) is entered into the model. However the co-variable may have a turning point in the relationship with risk and this will not be modelled by simple transformation.[184] Royston et al suggest that adding another polynomial term can appropriately model a continuous variable.[184] Although this looks inherently complex, most continuous variables can be modelled using to two fractional polynomials (FP)i.e. where $X^p + X^p$ is added to the model where p is in the family (-2, -1, -0.5, 0, 0.5, 1, 2 and 3, where 0 is the terminology for log).[254] Royston and colleagues have devised a macro for SAS (MFP) that assesses which FP to include using a closed test procedure and does a stepwise comparison of

functional forms until the appropriate one is found.[255] This process was used to assess non-linearity of forms for age, BMI, blood pressure and cholesterol. The best model fit was selected as described below.

#### *4.3.1.4.Model fit*

Model fit in survival analysis is given by the likelihood function (*L*) that summarises unknown parameter information in the model.[251] When considering different models the largest likelihood function represents best model fit to the observed data. In practice the value of $-2\log L$ is used and incorporated into Akaike Information Criteria (AIC) which is based on the likelihood statistic:

$$AIC = -2\log L - \alpha q$$

Where $q$ is the unknown number of parameters, $\alpha$ is a constant usually set to between 2-6 and generally 2 is used.[256] As the log likelihood statistic has been negatively log transformed, the smaller the AIC statistic the better the model fit. If two models are similar and one has more co variables, then AIC will not be much different and will penalize the model with more variables. Using fewer variables will avoid overfitting and improve the goodness of fit.[256] Usually a reduction of 4 in AIC between models is regarded as significant.[257]

#### *4.3.1.5.Missing data*

It was anticipated that there would be missing data as this is typical in primary care as shown in previous THIN studies.[144;258] If patients with missing data are excluded (a complete case analysis) then this can result in unstable and inaccurate models.[259] Complete case analysis is associated with loss of power in the study and may result in

overfitting.[260] An additional challenge is that the missing data are generally related to patient and study characteristics, unless the missing data are genuinely independent of observed and unobserved characteristics of cohort (missing completely at random, MCAR). If data are missing due to observed characteristics then the data is deemed missing at random (MAR) and if data are missing due to unobserved characteristics then the missing data are deemed missing not at random.[260] An important assumption that is made when examining missing data is that the missing data are MAR.[261]

In this study, complete case analysis may have led to selection bias and potentially could have excluded either well patients who have little illness and therefore fewer measurements and entries into the notes and also patients who are at greater risk of non-compliance who do not engage with their clinician and therefore also have fewer entries in their electronic record. Therefore missing data had to be considered when devising an outcome model and this will be discussed below.[261]

#### 4.3.1.5.1.Missing data in multivariable models and multiple imputation

The threshold of missing data at which models become unrealistic and unpredictable is unclear but Harrell et al suggested that when more than 5% of data are missing that complete case analysis should no longer be considered.[262] Missing data in analyses can be handled by several different ways. Simple strategies such as mean or median substitution or creating a category that labels the variable as missing are inefficient and inaccurate.[260] More complex strategies such EM algorithms, single mean conditional imputations and multiple imputation can be used.[259] For the purposes of

this thesis, only multiple imputation will be discussed since this is the current method of choice.[259]

#### 4.3.1.5.2.Multiple imputation

In single imputation, any missing data is assumed to be missing at random, and is replaced with a value generated using a multivariate regression model of existing non missing patient data. This allows analysis of the whole dataset. However the values imputed may have not been drawn from a representative dataset of the true population, and although small p values may be generated, the model may be over accurate, when in reality there is more uncertainty in how the data is imputed. To overcome this limitation, substituting multiple values for each missing value, using a regression analysis which has some random variation is preferable. This is called multiple imputation Each imputed dataset is analysed separately and then the regression estimates and coefficients are combined. Generally between 5-10 datasets yield optimum efficiency.[260]

There are numerous regression models for calculating imputed variables, but as the missing data in my model were likely to be categorical as well as continuous, and the datasets were very large, then the simplest and most efficient way was to use chained equations. Here a series of univariate regression analyses were carried out to generate imputed values.[263] Although this feature is not directly available in SAS software researchers have developed a SAS, “callable” function called IVEWARE.[263] This function imputes missing values using linear, logistic, Poisson, or generalized logit regression depending on the nature of the missing variable.[263]

It was anticipated that there would be greater than 15% missing data for at least some variables from previous THIN studies and therefore multiple imputation was used.[250] Multiple imputed datasets were created, then they were analysed individually and regression coefficients combined to produce median regression coefficients.[259]

#### *4.3.1.6.Checking Cox Proportional Hazards assumption*

If the included co-variables in the model violate the proportional hazards function then the model is not valid. There are numerous methods to examine this but hazards are only proportional if the ratio of hazards are independent of time.[251] This assumption was tested graphically by examining plots of scaled Schoenfeld residuals of particular variables with survival time. If a non zero relationship was observed this the assumption was violated.[251]

#### *4.3.1.7.Frailty Model to assess for clustering*

A traditional Cox proportional model does not account for clustering of data at the practice level in the THIN dataset. A frailty model, an extension of the Cox Model, corrects for random effects i.e. unobserved heterogeneity in survival data.[264] The prognostic models in this thesis included general practices as a frailty term into the Cox proportional hazards model assuming the variable has a gamma distribution.[264] This was done in R software using the frailty package.

### 4.3.2.Step by Step discussion of the models.

#### *4.3.2.1.All cause mortality Model: Extracting the patients, co-variables and outcomes and step by step analysis plan*

##### 4.3.2.1.1.Patient Selection

Patients with CKD stage 3a and 3b were characterized by two consecutive estimated eGFRs calculated from serum creatinine at least seven days apart using the non IDMS MDRD equation[13], diagnosed between 2005 to 2008. Calculated eGFRs were used as lab eGFR reporting did not commence until 2006 and it was anticipated that a cohort based solely on lab based data would yield insufficient events for the model. Secondly two consecutive results of creatinine where the conservative estimate is taken, provide relatively good agreement with each other.[9] Patients had to have survived at three months following a diagnosis of stage CKD 3 to allow a model to be devised. Patients had to be 18 or over at time of diagnosis of CKD and registered at the practice for greater than 6 months. All data post the AMR date was considered for each practice.[150] Patients entry in the cohort was the time of their initial diagnosis of CKD i.e. the first CKD 3a or 3b staging.

##### 4.3.2.1.2.Outcome Measures

The primary outcome in the all cause mortality model was time to death from the diagnosis of CKD. This was defined by patient registration data.

##### 4.3.2.1.3.Co-Variables

The distribution of the variables is described below. For any independent variable, the closest measurement between three months before the diagnosis of CKD and one

month after were in the multivariable analysis. The table also demonstrates the files and methodology used to define the variables. Note when preparing the database, the same cleaning and preparation techniques were used as in chapter 2 and 3. In the absence of relevant Read Codes indicating smoking status or comorbidity it was assumed that that patient did not smoke or have a diagnosis for the condition of interest.

**Table 4-1. List of co-variables in the analysis for all cause mortality**

| Variable | Description | Justification | Variable type, categories and distribution |
|---|---|---|---|
| **Age** | Age at diagnosis of CKD | Age – stepwise increase in mortality[265] | Continuous variable |
| **Gender** | | Male gender is associated with worse mortality[239] | Categorical Variable<br>M= Male<br>F=Female |
| **Cardiovascular Disease**<br>**Peripheral Vascular Disease (PVD)**<br>**Cerebrovascular Disease (CVD)**<br>**Coronary Heart Disease (CHD)** | Defined by QOF business Read codes and definitions from other cohorts | Cardiovascular Disease is a risk factor for mortality | Categorical<br>No PVD vs. PVD<br>No CVD vs. CVD<br>No CHD vs. CHD |
| **Diabetes Mellitus type 1 and type 2** | Defined by QOF business Read codes and definitions from other cohorts | Diabetes is a risk factor for CVD in CKD[196] | Categorical<br>No Diabetes<br>Diabetes |

| Variable | Description | Justification | Variable type, categories and distribution |
|---|---|---|---|
| **Atrial Fibrillation** | Defined by predetermined Read codes including QOF business codes. Assume those with missing codes do not have AF | Predictive of mortality[102;197] | |
| **Ever Smoked** | Read codes<br>Assume missing Read codes are non smokers | Smoking is associated with CKD associated CVD[112] | Categorical<br>Never Smoked<br>Smoked |
| **Antihypertensive drugs defined in classes *<br>1. Angiotensin blockers (angiotensin receptor antagonists and ACE inhibitors<br>2. Beta Blockers<br>3. Calcium Channel<br>4. Blockers<br>5. Diuretics<br>6. Other** | Defined by Drug codes and must be present for at least 3 months. Each Drug category will be analysed as separate co-variable | BP lowering has impact on CVD mortality in CKD and non CKD patients and there are class specific effects of Angiotensin blockade[120;204;229;231;266;267] | Categorical<br>not on drug<br>on drug |
| **Anti platelet agents*** | Defined by Drug codes and must be present for at least 3 months | Protective in CKD patients[120] | As above |

| Variable | Description | Justification | Variable type, categories and distribution |
|---|---|---|---|
| **Statins*** | Defined by Drug codes and must be present for at least 3 months | Reduces CVD in Sharp study[121]* | As above |
| **Renal drugs such as**<br>**1. Sodium bicarbonate**<br>**2. Phosphate binders(inc calcium based and non calcium based agents)**<br>**3. Erythropoietin**<br>***** | Defined by Drug codes and must be present for at least 3 months. Each drug will be analysed as a separate co-variable | Bicarbonate is protective, phosphate binders reduce phosphate and hence calcification and EPO may be harmful and have excess CVD events[268-270]* | As above |
| **QOF CKD code** | QOF business rules code | Shown to be detrimental in other diseases[182] | Categorical<br>no code<br>code |
| **Ethnic Group** | Identified by Read codes and adapted from the census guidelines | Significant differences in ethnic groups in incidence of cardiovascular disease in general population. Blacks/Indians do better in ESRD in terms of mortality[129] | Categorical<br>Caucasian<br>South East Asian<br>South Asian<br>African/Caribbean |
| **Deprivation index – Townsend Quintiles** | Separate file for this | Associated with worse outcomes [242] | Categorical Variable |

<table>
<tr><th>Variable</th><th>Description</th><th>Justification</th><th>Variable type</th></tr>
<tr><td>Blood pressure systolic and diastolic ~</td><td>Defined by AHD code</td><td>See Antihypertensive section</td><td>Continuous Normally distributed variable</td></tr>
<tr><td>BMI~</td><td>Defined by AHD code</td><td>Higher BMI associated with CKD and CVD. No evidence in CKD patients however J shaped relationship with ESRD patients. [127;271]</td><td>Positively skewed</td></tr>
<tr><td>Cholesterol~</td><td>From AHD file</td><td>Associated with progression of CKD[105]</td><td>Continuous Variable<br>Positively Skewed</td></tr>
<tr><td>Proteinuria~</td><td>If patients have no urine dip or ACR they will be defined as having normal protein excretion.<br>The categories will be created.[35]
<table>
<tr><th>ACR</th><th>Urine Dip</th><th>Category</th></tr>
<tr><td><3.0</td><td>Nil</td><td>Normal</td></tr>
<tr><td>3-30</td><td>Trace/1+</td><td>High</td></tr>
<tr><td>>30</td><td>>1+</td><td>Very High</td></tr>
</table></td><td>Categories adapted from Lancet collaborative meta-analysis shows proteinuria is associated with CVD mortality[93]</td><td>Categorical<br>Normal<br>High<br>Very High</td></tr>
<tr><td>Haemoglobin in mmol/l</td><td>From AHD files</td><td>Low Hb is associated with mortality[271]</td><td>Negatively skewed.</td></tr>
<tr><td>Calculated eGFR~ in (eGFR < 60 ml/min/1.73m2</td><td>First calculated eGFR (that was used to identify cohort)</td><td></td><td>Continuous Variable<br>Negatively skewed</td></tr>
</table>

*Note for all the drugs they have to be prescribed for a minimum of 3 months

#### *4.3.2.2.**Step by Step Analysis for Normal Cox Regression Model (Figure 4-1)***

##### 4.3.2.2.1.Preparing the dataset and multiple Imputation (performed by SAS)

Step 1. Firstly the dataset was labelled for the co-variables as above. The database was examined to look at the proportion with complete cases. If the proportion of population with missing data was 15% then multiple imputation was carried out.

Step 2. As IVEWARE uses methods that require normal distribution of continuous variables, the following transformations were performed to approximate to the normal distribution[263]

a. Haemoglobin (Hb) was reflected and then log transformed
b. Cholesterol was log transformed
c. Body Mass Index (BMI)was log transformed

Step 3. Survival times were calculated from the CKD diagnosis date to the outcome date and patients were censored if they were alive at the end of the study period and deregistered during the study period

##### 4.3.2.2.2.Analyzing each imputed data set and determining the final model

Step 4. For each imputed dataset, linearity of continuous covariables was assumed and variables were entered into a Cox proportionally hazards model using the backwards elimination model. The median regression co-efficients and parameters for each co-variable from each imputed dataset analysis were combined.

Step 5. For each imputed dataset, all the co-variables specified were entered in to a Cox – proportional hazards model using the MFP algorithm using MFP package in R with age, BMI, eGFR, cholesterol, systolic blood pressure, diastolic blood pressure, haemoglobin modelled by $2^{nd}$ degree fractional polynomials (mfp). The MFP algorithm in R entered all the co-variables regardless of whether the variable was to be transformed into a mfp in a backward elimination variable selection procedure. The procedure then cycled through all the possible 36 transformations (up to $2^{nd}$ degree fractional polynomial

transformations) for each specified co variable and selected the optimal fractional polynomial using a closed test procedure.[255] The commonest mfp forms for each imputed dataset were noted. It was anticipated that fractional polynomials would be different between the analyses of imputed datasets 1-5. The commonest multiple fractional forms were recorded.

Step 6. The commonest mfp for the continuous variables were entered with the other variables into a Cox proportional hazards model. The regression co-efficients were combined. The median regression co-efficients and parameters for each co-variable from each imputed dataset analysis were combined.

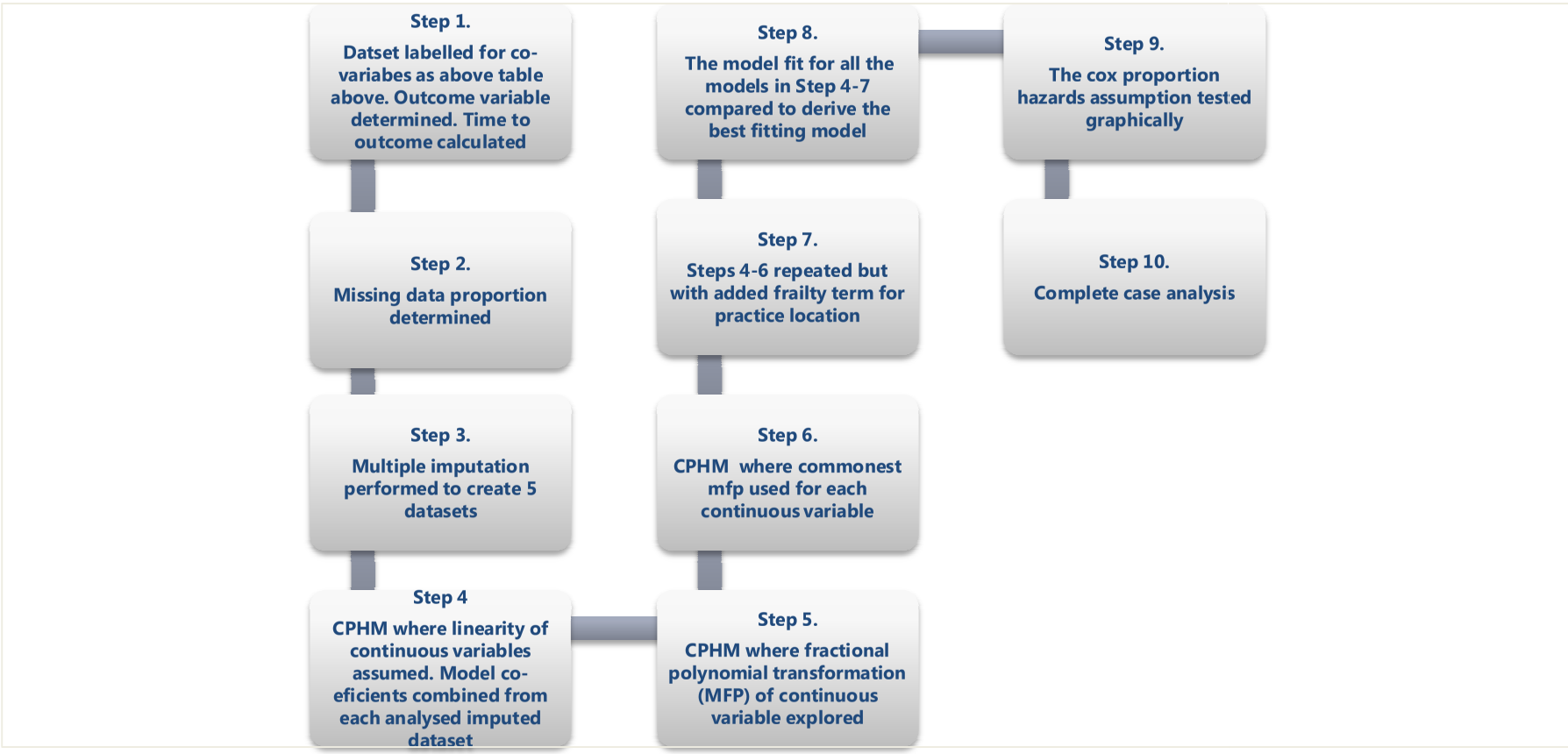
Step 7. Steps 4-6 were repeated with but with a frailty term for practice location added to each model.

Step 8. The model fit statistic the AIC of all the models in steps 4-7 were compared to demonstrate the best fitting model. i.e. the final model

Step 9. In the final model the proportional hazards assumption was examined for each co-variable using scaled Schoenfeld residuals against survival. The assumption was assumed to hold if there was a non zero slope.

Step 10. A complete case analysis i.e. a Cox proportional hazards model performed on dataset without missing data performed with mfp and frailty term for practice location was also estimated.

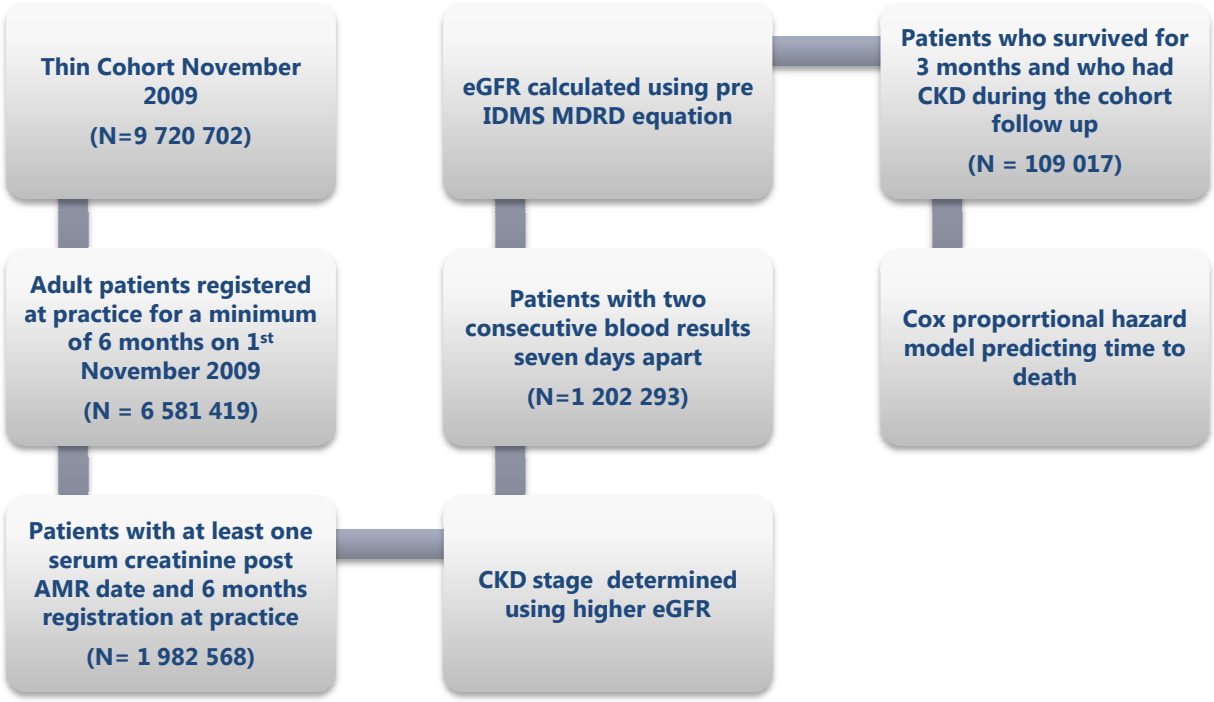
**Figure 4-1. Summary of all cause mortality model.**



## 4.4.Results

This section of the chapter describes the results of a model to predict all-cause mortality using routinely collected general practice data in CKD 3 patients. In total 109 017 patients were included in the model from the THIN cohort (Figure 4-2).

**Figure 4-2. Derivation of THIN CKD cohort for all cause mortality model**




### 4.4.1. Step 1: All cause mortality outcomes and description

The median follow up for the cohort was 1064 days (range 90 to 1079 days). There were 13 083 deaths. The cohort demographics and co-morbidities are shown in Table 4-2. The median age was 72.9 years and the majority of patients were female (62.2%). The majority of patients had either no recorded ethnicity or Caucasian race (98.3%), followed by African Caribbean (0.8%), Indian Sub-continental (0.9%) and South East Asian (0.1%) (Table 4-2). The majority of the cohort had the most affluent Townsend quintile.

#### *4.4.1.1.Risk factors/Medications*

Patients had a median eGFR of 54.32 mmol/litre/1.73m$^2$, median cholesterol of 4.9mmol/l, median BMI of 27.9 kg/m$^2$ and had mean blood pressure of 140.9/78.5 mmHg. The majority of patients with a urine dip or ACR had a high level of proteinuria (75.4%) followed by those with no proteinuria (17.6%) (Table 4-2). Patients had the following co-morbidities: 22.3% had diabetes mellitus, 11.5% had atrial fibrillation, 8.7% had heart failure and 19.8% were either currently smoking or had previously smoked.(Table 4-2) Many patients had some form of cardiovascular disease; 13.1% had Read codes for cerebro vascular accident (CVA), 25% had coronary heart disease (CHD) and 6.6% had peripheral vascular disease (PVD). Patients were on the following medications: anti-platelet agents (37.6%), angiotensin blockers (41.7%), beta-blockers (28.8%), diuretics (44.5%) and lipid lowering medication (41.6%) (Table 4-2) Nearly one in five (19.2%) patients had been prescribed non steroidal anti inflammatory drugs (NSAIDS) recently. (Table 4-2)

**Table 4-2. Demographics and clinical features of the whole for all cause mortality cohort (n= 109 017)**

| Demographic feature | Figure or proportion (either range or percentage in brackets) |
|---|---|
| **Median Age in years at time of diagnosis (years)** | 72.9(18-107) |
| **Gender – Female** | 68725(62.2) |
| **Not recorded/Caucasian Race** | 107178(98.3) |
| **African –Caribbean** | 813(0.8) |
| **Indian Subcontinent** | 975(0.9) |
| **South East Asian** | 81(0.1) |
| **Townsend quintiles (95.5) *** | |
| **1** | 26174(25.1) |
| **2** | 24810(23.8) |
| **3** | 21654(20.8) |
| **4** | 19066(18.3) |
| **5** | 12407(11.0) |
| **Diabetes Mellitus** | 24259 (22.3) |
| **Atrial Fibrillation** | 12555(11.5) |
| **Heart Failure** | 9504(8.7) |
| **Ever Smoked** | 21571(19.8) |
| **CVA** | 14231(13.1) |
| **CHD** | 27284(25.0) |
| **PVD** | 7178(6.6) |
| **Median eGFR in mmol/min/1.73m2** | 54.2 (3.6-59.9) |
| **Median Cholesterol (mmol/l) (66.1)*** | 4.9(1.7-13.9) |
| **Median Body Mass Index (kg/m2) (41.2)*** | 27.9(11.2-59.5) |
| **Median Haemoglobin (g/dl) (62.7)*** | 13.4(3.3-25.8) |
| **Mean Systolic blood pressure (mmHg) (84.8)*** | 140.9(20.6) |
| **Mean Diastolic blood pressure (mmHg) (84.8)*** | 78.5(11.4) |
| **Proteinuria levels (14.9)*** | |
| **None** | 2852(17.6) |
| **High** | 12225(75.4) |
| **Very High** | 1134(7.0) |
| **Anti platelets #** | 41004(37.6) |
| **Anticoagulation #** | 7056(6.5) |
| **Angiotensin Blockade #** | 51289(47.1) |
| **Beta blockers #** | 31378(28.8) |
| **Calcium Channel Blockers #** | 22510(20.7) |
| **Diuretics #** | 48525(44.5) |
| **Other Anti-hypertensives #** | 6959(6.4) |
| **Lipid lowering medication #** | 45320(41.6) |
| **Iron Supplementation #** | 6614(6.1) |
| **Vitamin D #** | 9357(8.6) |
| **Non Steroid Anti Inflammatory Drugs #** | 20968(19.2) |

* Proportion with available data # All medications prescribed between 3 months prior to and one month after CKD diagnosis

### 4.4.2.Step 2: Missing data

Only 5777 patients had a complete set of variables and the relative proportions are shown in Table 4-2**.** Townsend Quintiles (derived from postcode) and blood pressure had the highest proportion of complete data with proteinuria having the least complete data.

### 4.4.3. Step 3: Multiple imputation

Most records had missing data and so the pattern by which the data were missing was analysed and found to have an arbitrary missing pattern. In preparation for imputation, the incomplete data were graphed to assess normality. Serum cholesterol and BMI were positively skewed, hence log transformed and appeared normally distributed afterwards (Appendix A. Figures 1 to 2). Systolic and diastolic BP were normally distributed and Hb was negatively skewed and hence reflected and then log transformed (Appendix A. Figures 3 to 5). Subsequently multiple imputation of data in 5 datasets (Table 4-3) was undertaken using multiple chained equations. The distribution of imputed variables was examined and are shown in Table 4-3.

**Table 4-3. Variables with imputed data through imputation for all cause mortality model**

| Variable with imputed data | Original Data | Imp 1* | Imp 2 | Imp 3 | Imp 4 | Imp 5 |
|---|---|---|---|---|---|---|
| **Median Cholesterol (mmol/l)** | 4.9 (1.7-13.9) | 5.09 | 5.08 | 5.08 | 5.07 | 5.08 |
| **Median Body Mass Index (kg/m2)** | 27.9 (11.2-59.5) | 27.40 | 27.4 | 27.39 | 27.42 | 27.39 |
| **Median Haemoglobin (g/dl)** | 13.4 (3.3-25.8) | 13.50 | 13.50 | 13.50 | 13.50 | 13.50 |
| **Mean Systolic blood pressure (mmHg)** | 140.9 (20.6) | 140.48 | 140.54 | 140.52 | 140.51 | 140.50 |
| **Mean Diastolic blood pressure (mmHg)** | 78.5 (11.4) | 78.54 | 79.59 | 78.56 | 78.56 | 79.56 |
| **Proteinuria levels** | | | | | | |
| **None** | 2852 (17.6)* | 10740 (9.9) | 10516 (9.7) | 10095 (9.3) | 10342 (9.5) | 10366 (9.51) |
| **High** | 12225 (75.4) | 90092 (82.6) | 90236 (82.7) | 91118 (83.6) | 90473 (83.0) | 90368 (82.9) |
| **Very High** | 1134 (7.0) | 8185 (7.5) | 8265 (7.6) | 7804 (7.2) | 8202 (7.5) | 8283 (7.6) |
| **Townsend quintiles** | | | | | | |
| **1** | 26174 (25.1)* | 27347 (25.1) | 27390 (25.1) | 27339 (25.1) | 27367 (25.1) | 27372 (25.1) |
| **2** | 24810 (23.8) | 25982 (23.8) | 25931 (23.8) | 25994 (23.8) | 25934 (23.8) | 25897 (23.8) |
| **3** | 21654 (20.8) | 22680 (20.8) | 22727 (20.9) | 22683 (20.8) | 22694 (20.8) | 22699 (20.8) |
| **4** | 19066 (18.3) | 19998 (18.3) | 19956 (18.3) | 19974 (18.3) | 19997 (18.3) | 19981 (18.3) |
| **5** | 12407 (11.0) | 13010 (11.9) | 13013 (11.9) | 13027 (12.0) | 13025 (12.0) | 13068 (12.0) |

*abbreviation for imputation

(Note range will be the non imputed as minimum and maximum and proportions are in brackets (%))

### 4.4.4.Step 4. Results of analyses where the relationship between continuous co-variable and time to outcome was assumed linear.

The following sections detail the Cox proportional hazards model where the continuous co-variables were assumed to be linearly associated with time to outcome. The Cox regression analysis for each imputation is shown in Appendix A: Tables 1 to 5 and the combined estimates are shown in Table 4-4.

#### *4.4.4.1. Predictors associated with worse outcomes*

Increasing age (for every 100 years HR 513.7, 95%CI 414.2 to 637.1), male gender (compared to female HR 1.62, 95% CI 1.56 to 1.68), increasing Townsend quintile (Table 4-4), diabetes (HR 1.18, 95% CI 1.10 to 1.25), heart failure (HR 1.62, 95% CI 1.55 to 1.70), atrial fibrillation (HR 1.07, 95% CI 1.02 to 1.13) and patients who had previously smoked (HR 1.16, 95% CI 1.11 to 1.20) were independently associated with increased risk of death. In patients with CKD all forms of cardiovascular disease were associated (CHD HR 1.07, 95% CI 1.02 to 1.11, CVA 1.25, 95% CI 1.19 to 1.30, PVD 1.18, 95% CI 1.12 to 1.24) with poorer survival. Very but not high albuminuria was associated with worse survival. Anti-platelet medication (HR 1.08, 95% CI 1.04 to 1.27), anticoagulation (HR 1.29, 95% CI 1.21 to 1.38), diuretic use (HR 1.22 95% CI 1.18 to 1.27), iron supplements (HR 1.14, 95% CI 1.08 to 1.21) and vitamin D supplementation (HR 1.25, 95% CI 1.19 to 1.32) were associated with poorer survival (Table 4-4). The variables removed in backward selection were diastolic BP and NSAIDS. In the analysis of imputed dataset 3, calcium channel blockers were not significant but were significant in the overall combined model.

#### *4.4.4.2.Predictors associated with better outcomes*

African Caribbean ethnicity (HR 0.61, 95% CI 0.42 to 0.89) and Indian Sub-continental ethnicity (HR 0.47, 95% CI 0.35 to 0.63) were associated with better survival.(Table 4-4) Increasing haemoglobin (HR for every increase in 10g/dl 0.20, 95% CI 0.18 to 0.23), GFR (For an increase in 100ml/min/1.73m$^2$ HR 0.12, 95% CI 0.10 to 0.15), cholesterol (for an increase in 10mmol/l HR 0.50, 95% CI 0.43 to 0.59), systolic blood pressure (for an increase in 100mmHg, HR 0.51, 95% CI 0.46 to 0.55) and BMI (for an, increase in 10 kg/m$^2$, HR 0.83, 95% CI 0.80 to 0.86) were associated with improved survival. Angiotensin blockade (HR 0.85, 95% CI 0.81 to 0.88), beta blockade (HR 0.86, 95% CI 0.82 to 0.88), calcium channel blockers (HR 0.96, 95% CI 0.92 to 0.99) and other antihypertensive medication (HR 0.87, 95% CI 0.82 to 0.94) were associated with better survival. Lipid lowering medication was significantly associated with improved outcomes (HR 0.73, 95% CI 0.70 to 0.76). The Akaike Information Criteria (AIC), an indicator of model fit, improved with each model from a baseline of 293297 (Table 4-5).

**Table 4-4. All cause mortality: combined coefficients from Cox proportional hazards model and linear risk is assumed of numerical terms**

| Risk factor /Medication | Beta | SE | Hazard Ratio | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|---|
| **Age** | 6.242 | 0.110 | 513.685 | 414.181 | 637.093 |
| **Male Gender** | 0.481 | 0.020 | 1.618 | 1.557 | 1.682 |
| **Race**[a] | | | | | |
| **African Caribbean** | -0.494 | 0.190 | 0.610 | 0.421 | 0.885 |
| **Indian Subcontinent** | -0.763 | 0.151 | 0.466 | 0.347 | 0.628 |
| **South East Asian** | -1.183 | 1.000 | 0.306 | 0.043 | 2.176 |
| **Townsend Quintile**[b] | | | | | |
| **2nd** | 0.116 | 0.026 | 1.123 | 1.067 | 1.183 |
| **3rd** | 0.157 | 0.027 | 1.170 | 1.110 | 1.233 |
| **4th** | 0.176 | 0.027 | 1.192 | 1.129 | 1.258 |
| **5th** | 0.158 | 0.031 | 1.172 | 1.103 | 1.245 |
| **Diabetes Mellitus** | 0.168 | 0.023 | 1.183 | 1.130 | 1.239 |
| **Heart Failure** | 0.483 | 0.024 | 1.622 | 1.546 | 1.701 |
| **Atrial Fibrillation** | 0.070 | 0.025 | 1.073 | 1.021 | 1.126 |
| **Coronary Heart Disease** | 0.063 | 0.021 | 1.065 | 1.021 | 1.110 |
| **Cerebrovascular Accident** | 0.221 | 0.022 | 1.248 | 1.194 | 1.303 |
| **Peripheral Vascular Disease** | 0.161 | 0.029 | 1.175 | 1.109 | 1.244 |
| **Ever Smoked** | 0.146 | 0.023 | 1.157 | 1.106 | 1.210 |
| **Systolic BP** | -0.680 | 0.045 | 0.507 | 0.464 | 0.553 |
| **BMI** | -0.186 | 0.019 | 0.830 | 0.800 | 0.862 |
| **Haemoglobin** | -1.602 | 0.056 | 0.202 | 0.181 | 0.225 |
| **Glomerular Filtration Rate** | -2.131 | 0.101 | 0.119 | 0.097 | 0.145 |
| **Cholesterol** | -0.687 | 0.084 | 0.503 | 0.427 | 0.594 |
| **Albuminuria**[c] | | | | | |
| **High** | 0.029 | 0.035 | 1.030 | 0.961 | 1.103 |
| **Very High** | 0.217 | 0.044 | 1.243 | 1.141 | 1.354 |
| **Anti-platelets** | 0.078 | 0.021 | 1.081 | 1.037 | 1.126 |
| **Anticoagulation** | 0.255 | 0.033 | 1.291 | 1.210 | 1.377 |
| **Angiotensin Blockade** | -0.169 | 0.020 | 0.845 | 0.813 | 0.878 |
| **Beta blockade** | -0.151 | 0.021 | 0.860 | 0.825 | 0.896 |
| **Calcium Channel Blocker** | -0.045 | 0.022 | 0.956 | 0.916 | 0.998 |
| **Diuretic use** | 0.200 | 0.019 | 1.221 | 1.176 | 1.269 |
| **Other Antihypertensive** | -0.140 | 0.038 | 0.870 | 0.808 | 0.937 |
| **Lipid lowering agent** | -0.312 | 0.023 | 0.732 | 0.700 | 0.765 |
| **Iron supplements** | 0.135 | 0.029 | 1.144 | 1.081 | 1.211 |
| **Vitamin D supplementation** | 0.223 | 0.026 | 1.250 | 1.187 | 1.316 |

**Table 4-5. AIC for each analysed imputed dataset for analysis in Table 4-4**

| Imp Dataset | Akaike Information Criteria |
|---|---|
| **1** | 278848.2 |
| **2** | 278842.3 |
| **3** | 278887.3 |
| **4** | 278909 |
| **5** | 278946.2 |

### 4.4.5.Step 5: Analysis of transformed variables.

This section details the Cox proportional hazard models where continuous variables were transformed up to 2 fractional polynomial terms ($X^p + X^p$ where p is in the family 2, -1, -0.5, 0, 0.5, 1, 2 and 3, where 0 is the terminology for log). Each imputed model was analysed using the MFP algorithm in R. The details of each individual coefficient derived from each analysed imputed dataset are in Appendix A: Tables 11-15. The algorithm suggested a different transformation for the variables of, haemoglobin, BMI, cholesterol and GFR (Table 4-6).

**Table 4-6. Suggested transformation of variables from mfp algorithm in R for each imputation for mortality model**

| | **Imputed Dataset** | | | | |
|---|---|---|---|---|---|
| **Variable** | 1 | 2 | 3 | 4 | 5 |
| **Age/100** | $X + X*\text{Log } X$ | $X + X*\text{Log } X$ | $X + X*\text{Log } X$ | $X + X*\text{Log } X$ | $X + X*\text{Log } X$ |
| **Systolic BP /100** | $X + X^2$ | $X + X^2$ | $X + X^2$ | $X + X^2$ | $X + X^2$ |
| **Hb/10** | $X^{-1}+X^{-1}$ x LogX | $X^{-1}+X^{-1}$ x LogX | $X^{-1}+X^{-1}$ x LogX | $X^{-1}+X^{-1}$ x LogX | $X^{-1}+X^{-1}$ x LogX |
| **BMI/10** | $X+X^3$ | $\text{LogX} + X^2$ | $X^2 + X^2$ x LogX | $X^2 + X^2$ x LogX | $X^2 + X^2$ x LogX |
| **GFR/100** | $X^2$ | X | $X^2$ | $X^2$ | $X^2$ |
| **Cholesterol /10** | $X^2 + X^3$ | $X^2 + X^3$ | $X^2 + X^3$ | $X^2 + X^3$ | $X^2 + X^2$* Log X |

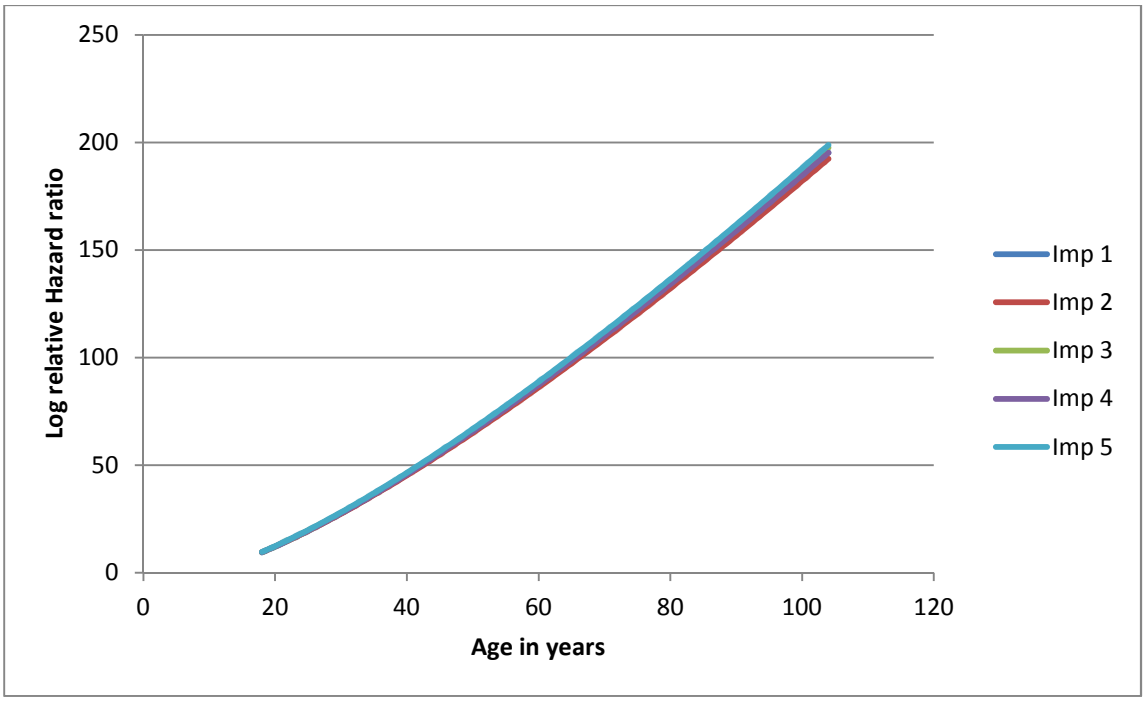
#### *4.4.5.1.Results: log relative hazard ratio for each transformation*

Using the coefficients for each variable, the log relative hazard ratio was plotted for all transformed co-variables against the original scale for each analysis of

- Age (Figure 4-3)
- Systolic blood pressure (Figure 4-4)
- Hb (Figure 4-5)
- BMI (Figure 4-6)
- Cholesterol (Figure 4-7)

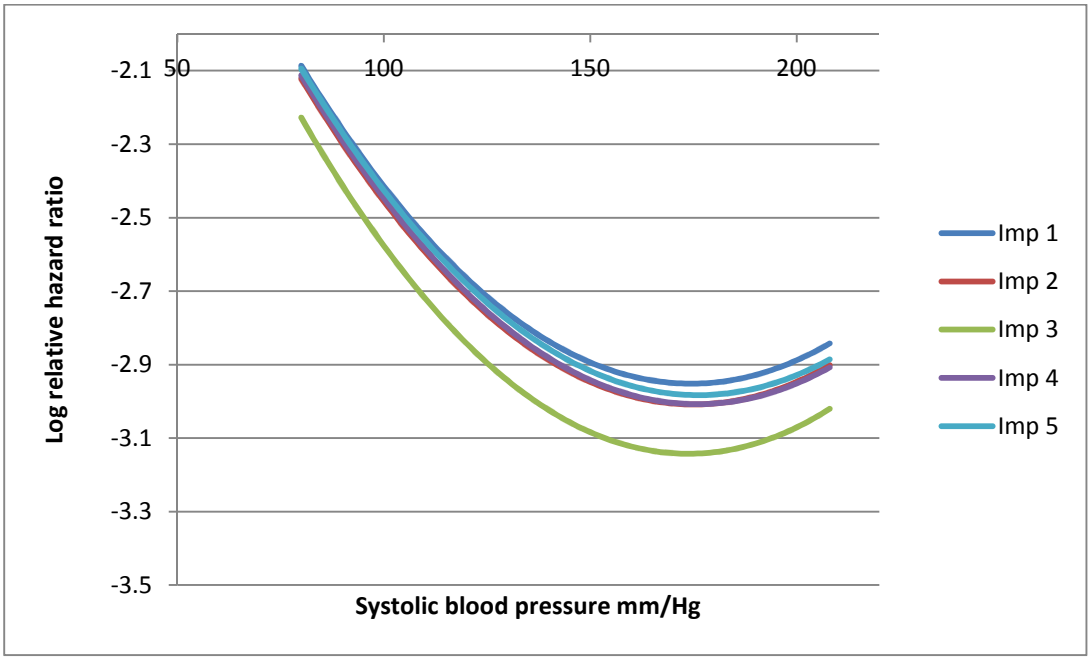
For age the relationship was identical for each imputation (Figure 4-3) and the log relative hazard ratio increased as age increased.

**Figure 4-3. Age versus Log relative Hazard Ratio for each imputation for mortality model (without frailty term)**



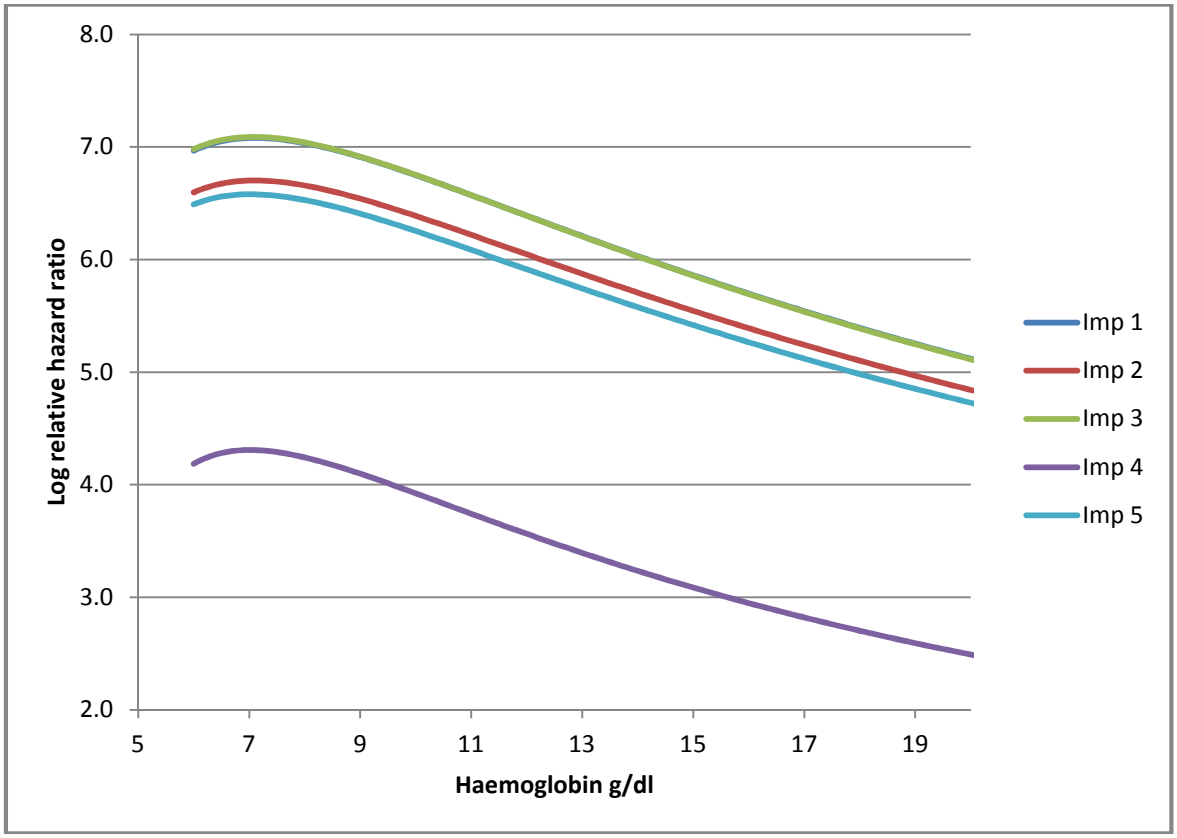
For systolic blood pressure (Figure 4-4), though the values were different, the log relative hazard ratio decreased until a systolic of 166mmHg and then increased slightly above this threshold.

**Figure 4-4. Log relative Hazard Ratio versus Systolic BP for each imputation for mortality model (without frailty term)**



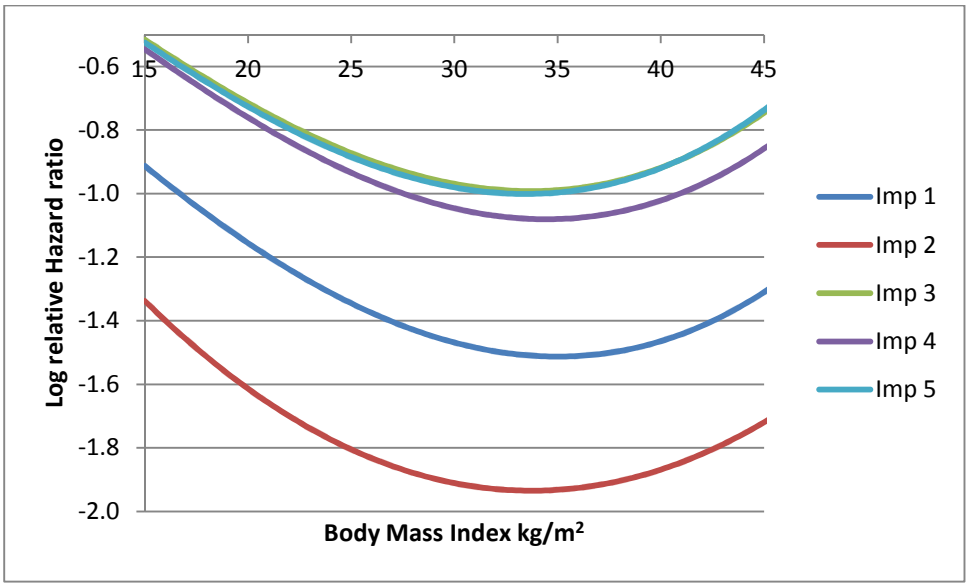

For haemoglobin (Figure 4-5), the values across imputations (imps) 1 to 3 and 5 are similar where the log relative hazard ratio decreased as haemoglobin increased. For the analysis of imputed dataset 4 the relationship is the same but the log relative hazard ratio was less.

**Figure 4-5. Haemoglobin versus Log relative Hazard Ratio for each imputation (without frailty term)**



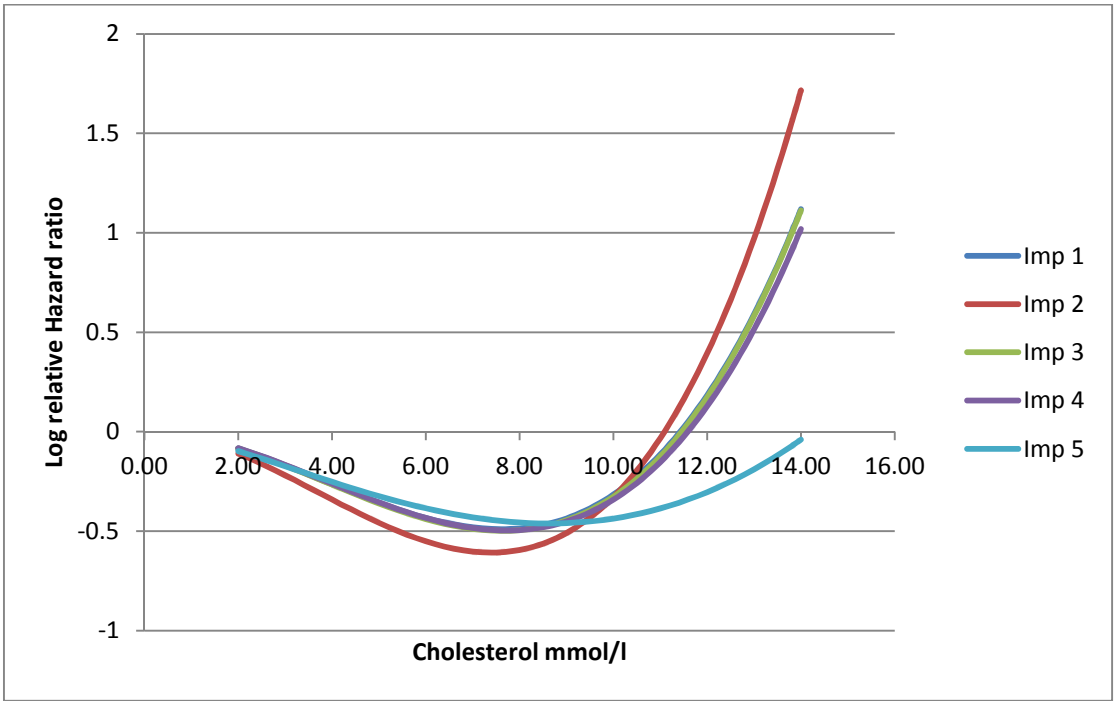

For the variable BMI (Figure 4-6), across Imps 1 to 5, the log relative hazard ratio decreased until BMI reaches 36 and then the log relative hazard ratio started to rise. However though the direction of the log relative hazard ratio was the same, the value for the beta coefficient for the log relative hazard ratio was different.

**Figure 4-6. Body Mass Index versus log relative hazards ratio for Imps 1-5 (no frailty term) for mortality model**




Similar to the above plot, as cholesterol increased from 2 to 8 mmol/l the log relative hazard ratio fell and then started to increase again through all the imputed datasets. (Figure 4-7)

**Figure 4-7. Cholesterol versus log relative hazards ratio for Imps 1-5 (no frailty term) for mortality model**



#### *4.4.5.2.Other non transformed continuous and categorical co variables*

Like the combined model results from step 4 in the results, diastolic blood pressure and NSAIDS were not significant and removed from the models for imputed datasets 1 to 5. Additionally calcium channel blockers were no longer significant and removed from the model. For each model for each imputation, with the exception of transformed variables, coefficients for each co-variable showed the same direction as the model in Step 4. Male gender, increasing Townsend quintile, diabetes mellitus, heart failure, atrial fibrillation, CHD, CVA, PVD, patients who smoke or had previously smoked, very high albuminuria, anti-platelets, anticoagulation, diuretic use, iron and vitamin D supplementation were associated with worse outcomes.

Patients with African-Caribbean race, Indian-Sub continental race, increasing GFR (at a squared rate were associated with improved survival), treatment with angiotensin blockers, beta-blockers, other anti-hypertensives and other lipid lowering agents were associated with improved survival. The AIC for the analysis for each imputation is shown in Table 4-7 and is significantly better than previous models in Step 4.

**Table 4-7. AIC for each analysed imputed dataset: comparing non linear and linear terms for mortality model**

| Imp Dataset | Akaike Information Criteria for models analysed in Step 4: No transformation of continuous co-variable | Akaike Information Criteria for models analysed in Step 5: Where transformations of co-variable in for each continuous co-variable |
|---|---|---|
| **1** | 278848.2 | 278528.2 |
| **2** | 278842.3 | 278507.4 |
| **3** | 278887.3 | 278556.3 |
| **4** | 278909 | 278604.1 |
| **5** | 278946.2 | 278596.5 |

### *4.4.6.* Step 6. Selection of the model with the most consistent transformations i.e. the alternative model

The analysis imputed dataset using the MFP algorithm suggests a different transformation for the BMI, Hb, cholesterol and GFR. As specified in the methods each imputed model was analysed using the commonest suggested transformation. The model fit statistic AIC of this model, the 'alternative' model was compared to the AIC of original model with mfp. The following were the commonest mfp for the transformed variable.

- Age - X + X*Log X (all Imps)
- Systolic BP - X + $X^2$ (all Imps)
- Hb - $X^{-1}$+$X^{1-}$ x Log X (1 to 3, 4)
- BMI - $X^2$+$X^2$*Log X (Imps 3 to 5)
- GFR - $X^2$ (Imps 1, 3 to 5)
- Cholesterol - $X^2$+$X^3$ (all Imps)

#### *4.4.6.1.Model fit statistic comparing the original mfp modes and the 'alternative' model*

To see if the 'alternative' model was significantly worse than the original mfp model then AIC had to increase by 4 ($p < 0.05$). However the model with the commonest transformation (alternative) model and the original mfp models were no different and the model fit improved for the best model in comparison with the original model.

**Table 4-8. A comparison between the original model with mfp and alternative model (no frailty term)**

| Imp Dataset | Akaike Information Criteria for original mfp models | Akaike Information for 'alternative' mfp model |
|---|---|---|
| **1** | 278528.2 | 278528.0 |
| **2** | 278507.4 | 278501.3 |
| **3** | 278556.3 | 278557.2 |
| **4** | 278604.1 | 278604.8 |
| **5** | 278596.5 | 278598.2 |

#### *4.4.6.2.The final combined (without frailty) (Table 4-9)*

The following co-variables were associated with poorer survival; male gender (HR 1.64, 95% CI 1.58 to 1.71), Increasing Townsend quintile (for e.g. 5$^{th}$ quintile versus the first quintile, HR 1.16, 1.10 to 1.24, diabetes mellitus (HR .1.19, 1.13 to 1.25), heart failure (HR 1.69, 1.59 to 1.79), atrial fibrillation (HR 1.18, 1.13 to 1.24).

Again as in the previous steps, all cardiovascular disease was associated with increased death (CHD HR 1.07, 1.02 to 1.13, CVD HR 1.24, 1.18 to 1,29, PVD HR 1.17, 1.11 to 1.24) Patients who were current or previous smokers (HR 1.26, 1.21 to 1.32), very high albuminuria but not high levels of proteinuria (very high albuminuria versus no proteinuria HR 1.24, 1.14 to 1.35), anti-platelets prescription (HR 1.08, 1.03 to 1.12), anticoagulation (HR 1.30, 1.22 to 1.39), diuretic use (HR 1.22, 1.17 to 1.26), iron supplementation (HR 1.13, 1.06 to 1.19) and vitamin D supplementation (HR 1.24, 1.18 to 1.31) were associated with lesser survival also.

African-Caribbean ethnicity (HR 0.58, 0.40 to 0.84), Indian Sub-continental ethnicity (HR 0.46, 0.34 to 0.62), angiotensin blockade (HR 0.85, 0.82 to 0.89), beta-blockade (HR 0.86, 0.83 to 0.90), other antihypertensives (HR 0.89, 0.82 to 0.96) and lipid lowering medication (HR 0.75, 0.71 to 0.78) were associated with better survival. Increasing GFR (for an increase in (100 mls/min/1.73m$^2$)$^2$, HR 0.10, 0.08 to 0.13) was associated with increased survival.

**Table 4-9. The combined coefficients using the 'alternative' models (No frailty term)**

| Risk factor/Prescription | Beta | SE | HR | 95% CI | |
|---|---|---|---|---|---|
| **Age/100** | 7.75 | 0.76 | 2328.55 | 529.29 | 10244.19 |
| **Age/100 x log(Age/100)** | 0.69 | 0.54 | 2.00 | 0.70 | 5.71 |
| **Male Gender** | 0.50 | 0.02 | 1.64 | 1.58 | 1.71 |
| **African –Caribbean**[a] | -0.54 | 0.19 | 0.58 | 0.40 | 0.84 |
| **Indian Subcontinent** | -0.77 | 0.15 | 0.46 | 0.34 | 0.62 |
| **South East Asian** | -1.25 | 1.00 | 0.29 | 0.04 | 2.03 |
| **2nd Townsend Quintile**[b] | 0.12 | 0.03 | 0.12 | 1.07 | 1.18 |
| **3rd Townsend Quintile**[b] | 0.15 | 0.03 | 1.16 | 1.10 | 1.23 |
| **4th Townsend Quintile**[b] | 0.17 | 0.03 | 1.18 | 1.12 | 1.25 |
| **5th Townsend Quintile**[b] | 0.15 | 0.03 | 1.16 | 1.10 | 1.24 |
| **Diabetes Mellitus** | 0.17 | 0.02 | 1.18 | 1.13 | 1.24 |
| **Heart Failure** | 0.46 | 0.02 | 1.59 | 1.52 | 1.67 |
| **Atrial Fibrillation** | 0.07 | 0.02 | 1.07 | 1.02 | 1.13 |
| **Coronary Heart Disease** | 0.06 | 0.02 | 1.07 | 1.02 | 1.11 |
| **Cerebrovascular Accident** | 0.21 | 0.02 | 1.24 | 1.18 | 1.29 |
| **Peripheral Vascular Disease** | 0.16 | 0.03 | 1.17 | 1.11 | 1.24 |
| **Ever Smoked** | 0.15 | 0.02 | 1.16 | 1.11 | 1.22 |
| **Systolic BP/100** | -3.41 | 0.31 | 0.03 | 0.02 | 0.06 |
| **(Systolic BP/100)$^2$** | 0.97 | 0.11 | 2.64 | 2.13 | 3.26 |
| **(BMI/10)$^2$ x log (BMI/10)** | 0.18 | 0.02 | 1.20 | 1.16 | 1.24 |
| **(BMI/10)$^2$** | -0.32 | 0.03 | 0.73 | 0.69 | 0.77 |
| **(Hb/10)$^{-1}$ x log(Hb/10)** | 4.74 | 0.36 | 114.66 | 56.40 | 233.12 |
| **(Haemoglobin/10)$^{-1}$** | 6.38 | 0.35 | 588.06 | 293.33 | 1178.92 |
| **(Glomerular Filtration Rate/100)2** | -2.29 | 0.12 | 0.10 | 0.08 | 0.13 |
| **(Cholesterol/10)$^2$** | -2.56 | 0.34 | 0.08 | 0.04 | 0.15 |
| **(Cholesterol/10)$^3$** | 2.24 | 0.36 | 9.36 | 4.59 | 19.09 |
| **High Albuminuria** [c] | 0.03 | 0.04 | 1.03 | 0.96 | 1.10 |
| **Very High Albuminuria** | 0.22 | 0.04 | 1.24 | 1.14 | 1.35 |
| **Anti-platelets** | 0.07 | 0.02 | 1.08 | 1.03 | 1.12 |
| **Anticoagulation** | 0.26 | 0.03 | 1.30 | 1.22 | 1.39 |
| **Angiotensin Blockade** | -0.17 | 0.02 | 0.85 | 0.82 | 0.88 |
| **Beta blockade** | -0.15 | 0.02 | 0.86 | 0.83 | 0.90 |
| **Diuretic use** | 0.20 | 0.02 | 1.22 | 1.17 | 1.26 |
| **Other** | -0.14 | 0.04 | 0.87 | 0.80 | 0.93 |
| **Lipid lowering agent** | -0.29 | 0.02 | 0.75 | 0.71 | 0.78 |
| **Iron supplements** | 0.12 | 0.03 | 1.13 | 1.06 | 1.19 |
| **Vitamin D supplementation** | 0.22 | 0.03 | 1.24 | 1.18 | 1.31 |

#### *4.4.6.3.Fractional polynomials in Model*

The relationship between log hazard ratios in age (Figure 4-8) is the same as for the original transformed mfp (Figure 4-3), as age increases the log relative hazard ratio increases. The log relative hazard decreased as systolic blood pressure started to decrease but when blood pressure was above 172 mmHg the risk started to increase (Figure 4-8). As haemoglobin increased the log relative hazard ratio decreased. When BMI increased from 15 to 36 kg/m$^2$ the log relative hazard ratio fell sharply and increased after this (Figure 4-9). As the cholesterol increased the log relative hazard ratio decreased until after a cholesterol of 7.6 mmol/l the log relative hazard ratio started to increase again. (Figure 4-9)

**Figure 4-8. Age versus log relative hazard ratio (no frailty term) in model using most consistent mfp for all cause mortality**

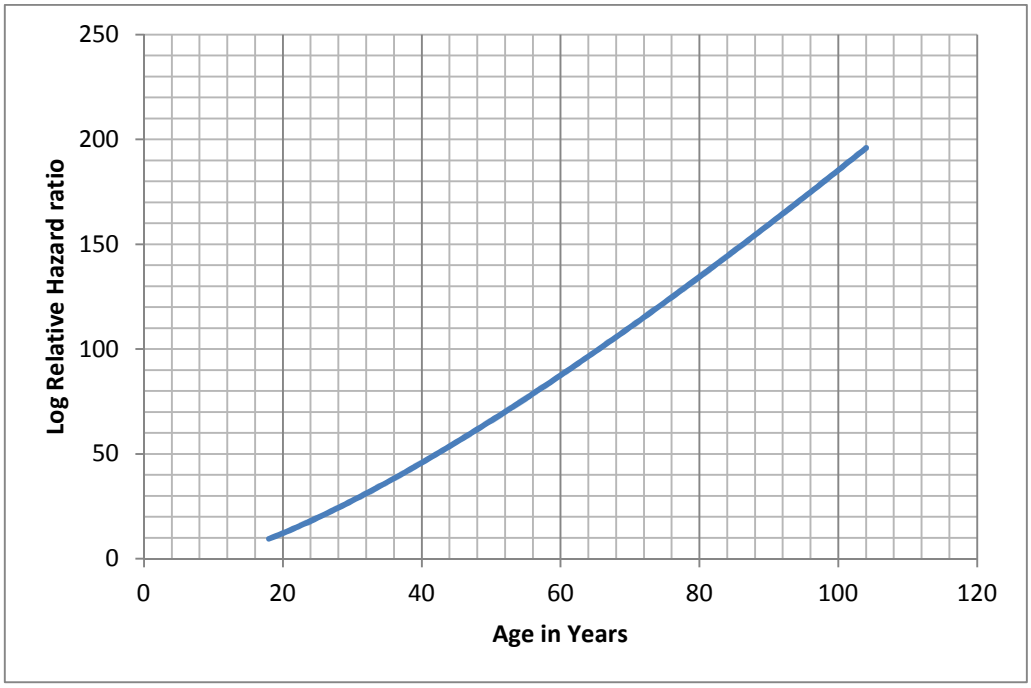
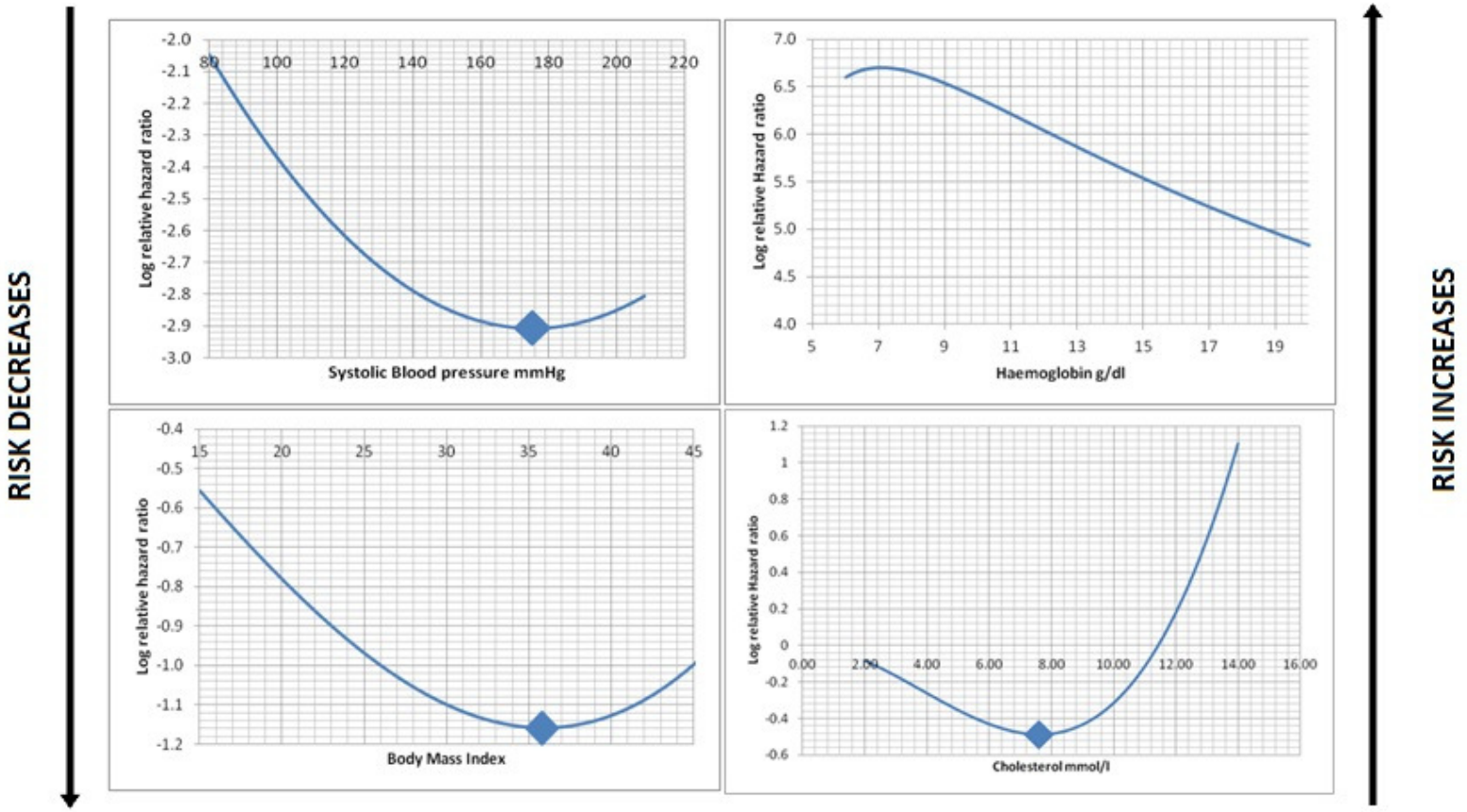


**Figure 4-9. Most consistent multiple fractional polynomials for death model (without frailty): Systolic blood pressure, Hb, BMI and cholesterol versus log relative hazard ratio**



### 4.4.7.Step 7. Frailty Models

Steps 4 to 6 were repeated with a frailty term (a random effects term) for practice location added to each model. These steps will be named Step 4F to 6F.

#### *4.4.7.1.Step 4F. Analysis of co-variables where relationship between continuous co-variable and time to outcome is assumed to be linear.*

When a frailty term (a random effects term) for practice was added to each imputation, similar to the model in step 4 without frailty, diastolic blood pressure and NSAIDs were not significant and removed from the final model. (Appendix A: Tables 11 to 15). Calcium channel blockers were included despite not being significant in 3$^{rd}$ Imp but were significant in the combined model. (Table 4-10).

**Table 4-10. Combined coefficients for all cause mortality model where linear risk is assumed of numerical terms with practice location as frailty term**

| Risk factor /Prescription | Beta | SE | HR | 95% CI | |
|---|---|---|---|---|---|
| AGE | 6.248 | 0.111 | 516.823 | 416.099 | 641.928 |
| MALE GENDER | 0.480 | 0.020 | 1.617 | 1.555 | 1.681 |
| African –Caribbean[a] | -0.498 | 0.191 | 0.608 | 0.418 | 0.884 |
| Indian Subcontinent | -0.750 | 0.153 | 0.473 | 0.350 | 0.638 |
| SOUTH EAST ASIAN | -1.197 | 1.001 | 0.302 | 0.043 | 2.148 |
| TOWNSEND QUINTILE[B] | | | | | |
| 2ND | 0.113 | 0.027 | 1.119 | 1.062 | 1.179 |
| 3RD | 0.152 | 0.027 | 1.164 | 1.103 | 1.228 |
| 4TH | 0.169 | 0.028 | 1.184 | 1.121 | 1.251 |
| 5TH | 0.154 | 0.032 | 1.166 | 1.096 | 1.242 |
| DIABETES MELLITUS | 0.165 | 0.024 | 1.179 | 1.126 | 1.234 |
| HEART FAILURE | 0.484 | 0.024 | 1.622 | 1.546 | 1.701 |
| ATRIAL FIBRILLATION | 0.065 | 0.025 | 1.067 | 1.016 | 1.121 |
| CORONARY HEART DISEASE | 0.063 | 0.022 | 1.065 | 1.021 | 1.111 |
| CEREBROVASCULAR ACCIDENT | 0.218 | 0.022 | 1.243 | 1.190 | 1.299 |
| PERIPHERAL VASCULAR DISEASE | 0.156 | 0.030 | 1.169 | 1.103 | 1.239 |
| EVER SMOKED | 0.144 | 0.024 | 1.154 | 1.101 | 1.210 |
| SYSTOLIC BP | -0.679 | 0.045 | 0.507 | 0.464 | 0.554 |
| BMI | -0.187 | 0.019 | 0.830 | 0.800 | 0.861 |
| HAEMOGLOBIN | -1.603 | 0.056 | 0.201 | 0.181 | 0.225 |
| GLOMERULAR FILTRATION RATE | -2.152 | 0.102 | 0.116 | 0.095 | 0.142 |
| CHOLESTEROL | -0.682 | 0.085 | 0.506 | 0.428 | 0.597 |
| ALBUMINURIA[C] | | | | | |
| HIGH | 0.028 | 0.035 | 1.029 | 0.960 | 1.103 |
| VERY HIGH | 0.228 | 0.044 | 1.257 | 1.153 | 1.369 |
| ANTI-PLATELETS | 0.077 | 0.021 | 1.080 | 1.036 | 1.126 |
| ANTICOAGULATION | 0.258 | 0.033 | 1.295 | 1.213 | 1.381 |
| ANGIOTENSIN BLOCKADE | -0.168 | 0.020 | 0.845 | 0.813 | 0.879 |
| BETA BLOCKADE | -0.150 | 0.021 | 0.861 | 0.826 | 0.897 |
| CALCIUM CHANNEL BLOCKER | -0.047 | 0.022 | 0.954 | 0.914 | 0.997 |
| DIURETIC USE | 0.198 | 0.020 | 1.219 | 1.173 | 1.266 |
| OTHER | -0.139 | 0.038 | 0.870 | 0.808 | 0.937 |
| LIPID LOWERING AGENT | -0.309 | 0.023 | 0.734 | 0.702 | 0.768 |
| IRON SUPPLEMENTS | 0.133 | 0.029 | 1.142 | 1.079 | 1.208 |
| VITAMIN D SUPPLEMENTATION | 0.224 | 0.027 | 1.250 | 1.187 | 1.318 |

The combined model with frailty was similar to the model above (Step 4. Non frailty models) To recap; increasing age, male gender, increasing Townsend quintile, diabetes mellitus, heart failure, atrial fibrillation, patients who had previously smoked, all cardiovascular disease, very high albuminuria, anti-platelets, anticoagulation, diuretic use, iron and vitamin D supplementation were associated with poorer survival.(Table 4-10)

African-Caribbean ethnicity, Indian sub-continental ethnicity, increasing systolic blood pressure, body mass index, Haemoglobin, GFR and cholesterol were associated with better survival. Treatment with Angiotensin blockade, beta blockers, calcium channel blockers, other anti-hypertensives and lipid lowering medication were also associated with better survival. The AIC for each model for each imputed dataset was better than the previous model without a frailty term for practice location (Table 4-11).

**Table 4-11. AIC for each model for all cause mortality model with and without frailty term**

| Imp Dataset | Akaike Information Criteria for models analysed in Step 4: No transformation of continuous co-variable without frailty | Akaike Information Criteria for models analysed in Step 4f: No transformation of continuous co-variable with frailty |
|---|---|---|
| **1** | 278848.2 | 278760.6 |
| **2** | 278842.3 | 278745.9 |
| **3** | 278887.3 | 278801.7 |
| **4** | 278909 | 278814.9 |
| **5** | 278946.2 | 278857.7 |

#### *4.4.7.2.Step 5f. Models with transformation*

When practice location was added to the algorithm to determine the best fitting fractional polynomial in Cox proportional hazards model for each imputed dataset; with the exception of calcium channel blockers, the same co-variables were significant and same transformation of continuous co variables as in Step 5 were suggested (Cox proportional hazards model are in Appendix A Tables 16 to 20) The plots of the transformed co-variables were almost identical to figure 4-8 and figure 4-9 and are shown in Appendix A Figures 6 to 10. The AIC for each imputation substantially improved in comparison with all previous models in step 2.(Table 4-12)

**Table 4-12. AIC for each model for all cause mortality with original suggested fractional polynomial with frailty term**

| Imp Dataset | Akaike Information Criteria for models analysed in Step 5: : Variables transformed according to Table 4-6 without frailty | Akaike Information Criteria for models analysed in Step 5f: Variables transformed according to Table 4-6 with frailty |
|---|---|---|
| **1** | 278528.2 | 278447.3 |
| **2** | 278507.4 | 278418 |
| **3** | 278556.3 | 278477.1 |
| **4** | 278604.1 | 278516 |
| **5** | 278596.5 | 278514.9 |

#### *4.4.7.3.Step 6f. Finding the most consistent transformations of continuous co-variables across the Imps*

Like the non frailty models (Step 6 non frailty model), the 'mfp' transformations suggested for Hb, BMI, Cholesterol and GFR differed across each Imp and therefore the

commonest transformations were compared against the original suggested transformations.

The commonest transformations to recap were:

Age/100 + Age/100 x log Age + Systolic blood pressure/100 + systolic blood pressure/$100^2$ + Haemoglobin $^{-1}$ + Haemoglobin/$10^{-1}$ x log Haemoglobin/10 + BMI $^2$ + BMI/$10^2$ x log BMI/10 + GFR/$100^2$ + Chol/$10^2$ + Chol/$10^3$

Like in step 6 in the non frailty model, the AIC statistic did not differ by 4 between models for each imputation except for the analysis of imputed dataset 3 where by the model improved with the most consistent transformations.(Table 4-13)

**Table 4-13. A comparison between the original model with mfp and the best model with a frailty term**

| Imp Dataset | Akaike Information Criteria for original Model | Akaike Information for 1st 'alternative' Model |
|---|---|---|
| **1** | 278447 | 278447 |
| **2** | 278418 | 278410 |
| **3** | 278477 | 278477 |
| **4** | 278516 | 278517 |
| **5** | 278515 | 278515 |

As the 'alternative' imputed model seemed the best fit across all analysed imputed datasets this was chosen as the final model and coefficients were combined for each imputation.

#### *4.4.7.4.Final combined model with mfp and frailty term for practice location*

The same co variables are significant in 'best model' without frailty (Step 3) and the 'best model' with frailty (Step 6f). The direction of the risk/protection did not change and the coefficients changed slightly (Table 4-14) especially for the transformed variables.

Table 4-14. Combined all cause model for each imputation using the alternative model incorporating a frailty term of practice location

| Risk factor /Prescription | Beta | SE | HR | 95% CI | |
|---|---|---|---|---|---|
| **Age/100** | 7.76 | 0.76 | 2339.28 | 529.49 | 10334.8 |
| **Age/100 x log(Age/100)** | 0.70 | 0.54 | 2.00 | 0.70 | 5.75 |
| **Male Gender** | 0.50 | 0.02 | 1.64 | 1.58 | 1.71 |
| **African –Caribbean**[a] | -0.55 | 0.19 | 0.58 | 0.40 | 0.84 |
| **Indian Subcontinent** | -0.76 | 0.15 | 0.47 | 0.35 | 0.63 |
| **2nd Quintile**[b] | 0.11 | 0.03 | 1.12 | 1.06 | 1.18 |
| **3rd Quintile**[b] | 0.15 | 0.03 | 1.16 | 1.10 | 1.22 |
| **4th Quintile**[b] | 0.16 | 0.03 | 1.17 | 1.11 | 1.24 |
| **5th Quintile**[b] | 0.15 | 0.03 | 1.16 | 1.09 | 1.23 |
| **Diabetes Mellitus** | 0.16 | 0.02 | 1.18 | 1.13 | 1.23 |
| **Heart Failure** | 0.46 | 0.02 | 1.59 | 1.52 | 1.67 |
| **Atrial Fibrillation** | 0.06 | 0.03 | 1.07 | 1.02 | 1.12 |
| **Coronary Heart Disease** | 0.06 | 0.02 | 1.07 | 1.02 | 1.11 |
| **Cerebrovascular Accident** | 0.21 | 0.02 | 1.23 | 1.18 | 1.29 |
| **Peripheral Vascular Disease** | 0.15 | 0.03 | 1.17 | 1.10 | 1.24 |
| **Ever Smoked** | 0.15 | 0.02 | 1.16 | 1.11 | 1.22 |
| **Systolic BP/100** | -3.35 | 0.31 | 0.04 | 0.02 | 0.06 |
| **(Systolic BP/100)$^2$** | 0.95 | 0.11 | 2.58 | 2.08 | 3.19 |
| **(BMI/10)$^2$ x log (BMI/10)** | 0.18 | 0.02 | 1.20 | 1.16 | 1.24 |
| **(BMI/10)$^2$** | -0.32 | 0.03 | 0.73 | 0.69 | 0.77 |
| **(Hb/10)$^{-1}$ x log(Hb/10)** | 4.74 | 0.36 | 114.53 | 56.27 | 233.11 |
| **(Hb/10)$^{-1}$** | 6.38 | 0.36 | 588.87 | 293.37 | 1182.02 |
| **(Glomerular Filtration Rate/100)2** | -2.32 | 0.12 | 0.10 | 0.08 | 0.12 |
| **(Cholesterol/10)$^2$** | -2.52 | 0.35 | 0.08 | 0.04 | 0.16 |
| **(Cholesterol/10)$^3$** | 2.20 | 0.37 | 9.04 | 4.42 | 18.50 |
| **High Albuminuria**[c] | 0.03 | 0.04 | 1.03 | 0.96 | 1.10 |
| **Very High Albuminuria** | 0.23 | 0.04 | 1.26 | 1.15 | 1.37 |
| **Anti-platelets** | 0.07 | 0.02 | 1.08 | 1.03 | 1.12 |
| **Anticoagulation** | 0.26 | 0.03 | 1.30 | 1.22 | 1.39 |
| **Angiotensin Blockade** | -0.17 | 0.02 | 0.85 | 0.82 | 0.88 |
| **Beta blockade** | -0.15 | 0.02 | 0.86 | 0.83 | 0.90 |
| **Diuretic use** | 0.19 | 0.02 | 1.21 | 1.17 | 1.26 |
| **Other** | -0.14 | 0.04 | 0.87 | 0.80 | 0.93 |
| **Lipid lowering agent** | -0.29 | 0.02 | 0.75 | 0.71 | 0.78 |
| **Iron supplements** | 0.12 | 0.03 | 1.12 | 1.06 | 1.19 |
| **Vitamin D supplementation** | 0.22 | 0.03 | 1.24 | 1.18 | 1.31 |

Reference for analysis a. Caucasian race b. 1st Townsend Quintile c. No Albuminuria

### 4.4.8. Step 8. Is model fit improved after incorporating mfp transformation and frailty terms?

The table below compares the model fit statistic AIC combined models from Step 4 (model without mfp and frailty), Step 6(model with 'best mfp') and Step 6f (model with 'best mfp' and frailty). A lower AIC implies a better model fit and a difference of 4 between the value of AIC for each model means there is a significantly better model ($p < 0.05$) and if this difference is 11 then the significance increases to $p < 0.001$. Incorporating fractional polynomials for age, systolic blood pressure, Hb and BMI, GFR and cholesterol led to a significantly better model. Incorporating the frailty term resulted in an even better fitting model. (Table 4-15)

**Table 4-15. Comparison of the AIC for the original all cause mortality model without frailty/mfp and then models with mfp and frailty**

| Imp Dataset | AIC for original Model i.e. no transformation or frailty term (Step 4) | AIC for alternative mfp model with but no frailty term (Step 6) | AIC for alternative mfp model with but with frailty term (Step 6f) |
|---|---|---|---|
| **1** | 278848 | 278528(-320) | 278447(81) |
| **2** | 278842 | 278501(341) | 278410(91) |
| **3** | 278887 | 278557(330) | 278477(80) |
| **4** | 278909 | 278604(304) | 278517(87) |
| **5** | 278946 | 278598(348) | 278515(83) |

*The value in brackets is the difference between the current column and the previous column

### 4.4.9.Step 9. Testing the proportional hazards assumption

The best most informative model is described in step 7 (Model in step 6f). The scaled Schoenfeld residuals against time are shown in Appendix B for Imputed dataset 3 analysis. All the graphs show a non zero slope hence maintain the proportional hazards assumption.

### 4.4.10.Step 10. Complete Case Analysis

As specified in the methods a complete case analysis was carried out for the dataset. This included 5777 patients who had a complete set of data for analysis. The median age was lower in the complete case analysis, but the gender proportion and ethnic mix was similar to the full dataset (Table 4-16).

**Table 4-16. Demographics of the whole all cause mortality cohort and cohort with complete data**

| Demographic feature | **Figure or proportion for entire Cohort (109 017)*** | **Figure or proportion Complete cohort analysis (n=5777)** |
|---|---|---|
| Median Age in years at time of diagnosis (years) | 72.9(18-107) | 71.8(23.4-99.8) |
| Gender – Female | 68725(62.2) | 3005(55.0) |
| Not recorded/Caucasian Race<br>African –Caribbean<br>Indian Subcontinent<br>South East Asian | 107178(98.3)<br>813(0.8)<br>975(0.9)<br>81(0.1) | 5617(97.2)<br>60(1.0)<br>97(1.7)<br>3(0.1) |
| Median eGFR mmol/min/1.73m$^2$ | 54.2 (3.6-59.9) | 55.1(4.4-59.9) |
| Diabetes Mellitus | 24259 (22.3) | 3307(57.2) |
| Atrial Fibrillation | 12555(11.5) | 703(12.2) |
| Heart Failure | 9504(8.7) | 561(9.7) |
| Ever Smoked | 21571(19.8) | 1478(25.6) |
| Antiplatelet agents | 41004(37.6) | 3054(52.9) |
| Anticoagulation | 7056(6.5) | 431(7.5) |
| Angiotensin Blockade | 51289(47.1) | 3752(65.0) |
| Beta blockers | 31378(28.8) | 1996(34.6) |
| Calcium Channel Blockers | 22510(20.7) | 1535(26.6) |
| Diuretics | 48525(44.5) | 2736(47.4) |
| Other Anti-hypertensives | 6959(6.4) | 266(9.8) |
| Lipid lowering medication | 45320(41.6) | 3870(67.0) |
| Iron Supplementation | 6614(6.1) | 469(8.1) |
| Vitamin D | 9357(8.6) | 395(6.8) |
| Non Steroid Anti Inflammatory Drugs | 20968(19.2) | 1018(17.6) |
| Median Cholesterol (mmol/l) (66.1) | 4.9(1.7-13.9) | 4.5(1.7-10.4) |
| Median Body Mass Index (kg/m2) (41.2) | 27.9(11.2-59.5) | 28.4(14-59.3) |
| Median Haemoglobin (g/dl) (62.7) | 13.4(3.3-25.8) | 13.5(3.9-20.8) |
| Mean Systolic blood pressure (mmHg) (84.8) | 140.9(20.6) | 139.0(13.8) |
| Mean Diastolic blood pressure (mmHg) (84.8) | 78.5(11.4) | 76.7(11.5) |
| Proteinuria levels (14.9)<br>None<br>High<br>Very High | <br>2852(17.6)*<br>12225(75.4)<br>1134(7.0) | <br>1088(18.8)<br>4389(80.0)<br>300(5.2) |
| Townsend quintiles (95.5)<br>1<br>2<br>3<br>4<br>5 | <br>26174(25.1)*<br>24810(23.8)<br>21654(20.8)<br>19066(18.3)<br>12407(11.0) | <br>1373(23.8)<br>1273(22.0)<br>1191(20.6)<br>1150(20.0)<br>790(13.7) |

* Proportion within available data

The median eGFR was slightly higher and more patients had diabetes and smoked but similar proportions had heart failure and atrial fibrillation (Table 4-16). In general patients were more likely to be on relevant prescription medication with the exception of Vitamin D supplementation and NSAID prescription (Table 4-16).

#### *4.4.10.1.Cox proportional hazards model for complete case analysis*

The median follow up was 967 days (90 to 1785) which is less than the previous analysis and 526 patients died during the study period. As models with frailty and transformation of variables had previously shown better fit, only this model is shown (Table 4-17)

Age and BMI will be discussed later but like the previous analysis the following were associated with worse survival; male gender, heart failure, 4$^{th}$ Townsend quintile, diabetes mellitus, heart failure, CVA, PVD, patients who had previously smoked, had very high albuminuria, diuretic use, anticoagulation, vitamin D supplementation.

Increasing haemoglobin, systolic blood pressure, and GFR and angiotensin blockade were associated with improved survival. Initially as age increased so did the log relative hazard ratio but when age reached 50 years then the rate at which the log relative hazard ratio increased fell and levelled out. As BMI increased the log relative hazard ratio decreased but not quite in a linear fashion.

**Table 4-17. Cox model with frailty term and fractional polynomials complete case analysis for all cause mortality model**

| Risk Factor/Prescription | Beta | SE | HR | 95% CI Lower | |
|---|---|---|---|---|---|
| Age/$100^{-1}$ | 9.85 | 1.45 | 18940.00 | 1113 | 322274. |
| Age/$100^{-0.5}$ | -0.38 | 3.88 | 0.00 | 0.00 | 0.00 |
| Gender | 0.58 | 0.10 | 1.78 | 1.48 | 2.15 |
| Townsend Quintile compared to 1st Quintile $2^{nd}$ | 0.06 | 0.14 | 1.06 | 0.81 | 1.40 |
| $3^{rd}$ | 0.09 | 0.14 | 1.10 | 0.84 | 1.44 |
| $4^{th}$ | 0.31 | 0.13 | 1.37 | 1.06 | 1.77 |
| $5^{th}$ | 0.17 | 0.15 | 1.19 | 0.88 | 1.60 |
| Diabetes Mellitus | 0.42 | 0.10 | 1.53 | 1.25 | 1.87 |
| Heart Failure | 0.76 | 0.11 | 2.13 | 1.72 | 2.63 |
| Cerebrovascular Accident | 0.25 | 0.10 | 1.28 | 1.04 | 1.57 |
| Peripheral Vascular Disease | 0.28 | 0.12 | 1.33 | 1.05 | 1.67 |
| Ever Smoked | 0.34 | 0.10 | 1.40 | 1.16 | 1.70 |
| Systolic BP/100 | -0.94 | 0.23 | 0.39 | 0.25 | 0.61 |
| Haemoglobin/10 | -1.70 | 0.27 | 0.18 | 0.11 | 0.31 |
| BMI/$10^{0.5}$ | -18.83 | 2.91 | 0.00 | 0.00 | 0.00 |
| BMI/10 | 5.30 | 0.84 | 200.60 | 38.64 | 1040.65 |
| GFR/100 | -1.87 | 0.54 | 0.15 | 0.05 | 0.45 |
| High Albuminuria | 0.12 | 0.13 | 1.12 | 0.88 | 1.44 |
| Very High Albuminuria | 0.85 | 0.18 | 2.34 | 1.63 | 3.35 |
| Anticoagulation | 0.31 | 0.13 | 1.37 | 1.06 | 1.77 |
| Angiotensin Blockade | -0.24 | 0.10 | 0.79 | 0.65 | 0.96 |
| Diuretic use | 0.30 | 0.10 | 1.35 | 1.12 | 1.64 |
| Vitamin D supplementation | 0.39 | 0.15 | 1.47 | 1.11 | 1.96 |

## 4.5. Discussion

### 4.5.1. Summary Of Results

This chapter has summarised the development of a model that examined which routinely collected primary care data are associated with survival in patients with CKD. The model development went through several iterations and included 109 017 patients with a median follow-up of 1064 days (i.e. about three years). Firstly the dataset had a high proportion of missing data and multiple imputations of missing data were undertaken with multiple chained equations to create five datasets with imputed data for blood pressure, cholesterol, haemoglobin, and body mass index. The first models included continuous co-variables linearly in the models. The 2$^{nd}$ iteration in the models were where the continuous variables were transformed up to 2$^{nd}$ degree fractional polynomials (fp) and the commonest and best fractional polynomials were determined. The 3$^{rd}$ iteration included models with fractional polynomials and a frailty term for practice location. The final model found to have the lowest AIC was the 3$^{rd}$ iteration. (Figure 4-10)

The forest plot in Figure 4-11 shows the co-variables that were not transformed in Model 3 and the model shows that African-Caribbean and Indian Sub-continental ethnicity, angiotensin blockers, beta blockers, other anti hypertensive drugs and anti lipid agents were associated with increased survival. Increasing Townsend quintile, AF, HF, diabetes, CHD, CVD, PVD, smoking, anti-platelets agents, anticoagulation, iron and

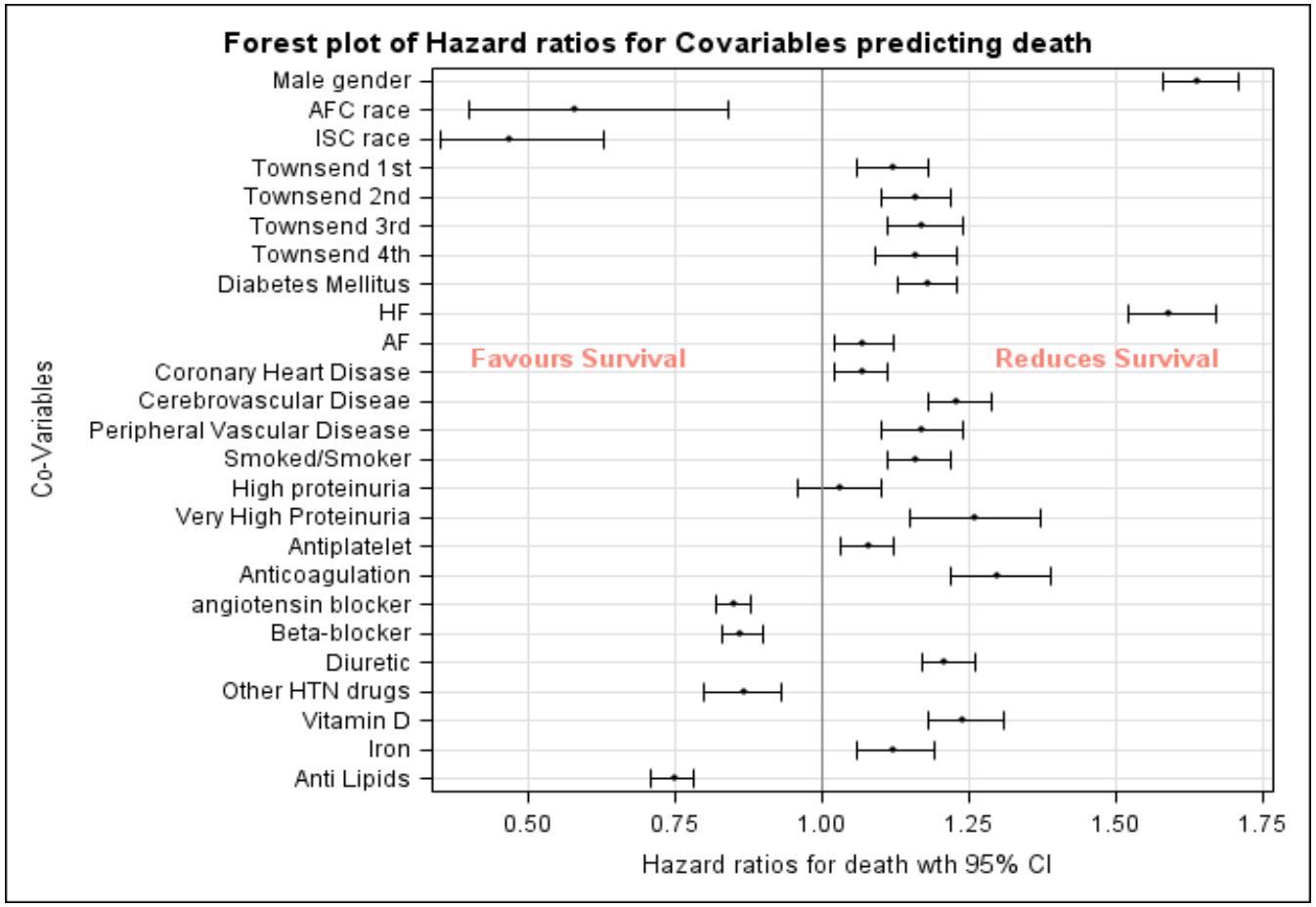
Vitamin D supplementation were all associated with worse survival. The non significant variables are described in (Figure 4-11).

With the exception of age where the risk of death approximately increases linearly as age increases, the relationship with systolic blood pressure, haemoglobin, cholesterol and BMI were more complex. The relationship with systolic blood pressure and hazard for death was inverse J shaped. As Hb increased generally the risk of death decreased. The relationship between BMI and risk was U shaped where this risk of death decreased until BMI of 36 $kg/m^2$ and then subsequently increased. This was similar to the relationship between the risk of death and cholesterol. When GFR increased the risk of death increased at squared rate. This model satisfied the non proportional hazards model as the co-variables had a non zero slope on Scaled Schoenfeld residual plots.

The complete case analysis consisted only of 5777 patients and the patients were younger than the whole cohort (Median age in years 71.8 vs. 72.9) and less likely to be female (55% vs. 62.2%). The median follow up was 967 days. The model was very similar to $1^{st}$ Model (Figure 4-10) however which contained a reduced number of protective factors.

**Figure 4-10. Iterations for Models predicting death**

**Figure 4-11. Forest plot of hazard rations for predictors of death**



### 4.5.2. Interpretation of Results

To ascertain the optimal model to determine co-variables associated with survival, the development underwent several iterations and ultimately the best model included the whole database with multiple imputations with fractional polynomials for age, systolic blood pressure, GFR, haemoglobin, cholesterol and BMI included in addition to a frailty (random effects) term for practice location. The suggested fractional polynomial for these co-variables differed across each model, however plots of these transformations with the log relative hazard ratio revealed very similar relationships. When the commonest transformations were used in the analysis for each imputed dataset, it was only for the analysis of $4^{th}$ imputed dataset that the model fit was worse than original suggested mfp. Additionally there was in improvement in model fit for dataset 2 analyses.

Before discussing each co-variable individually it is worth discussing the relationship between transformed co-variables and log hazard ratio. If a variable is linear or ordinal then the exponentiation of the log hazard ratio will be the hazard ratio, however when log relative hazard ratio is modelled on fractional polynomials especially $2^{nd}$ degree mfp then exponentiation will change this relationship and therefore further adjustment is required. However as log relative hazard ratio increases or decreases then there is a likewise change in hazard ratio.[251]

### 4.5.3.Comparison to existing literature

The relationship between age and log hazard ratio was almost linear and an increase in age was associated with increasing risk of death. This is similar to existing literature

where absolute risk for mortality and ESRD increased as age increased.[265] Male gender was associated with an increased risk of death and this is similar to existing literature in CKD patients and the general population.[239] Though ethnicity was missing in the majority of the population, non Caucasian patients were more likely to survive than those with unrecorded/Caucasian ethnicity. This may be due to several reasons. Ethnicity reporting in primary care electronic records software is either recorded by staff or volunteered by patient. Such patients are therefore more likely to consult their GP, have registered more recently and patients with under recorded ethnicity could have worse outcomes. Conversely these patients may be healthier and do not need to consult their GP. Favourable outcomes in Indian sub-continentals with CKD stage 3 have not been reported before but have been reported in patients on Renal replacement therapy.[129;241] The relationship with African – Caribbean patients with CKD is more complex. American studies suggest that Black patients with CKD are more likely to die in comparison with Caucasian patients.[240] However American Black patients have increased social deprivation and decreased access to health care. Though similar disparities exist in the UK, healthcare is free at the point of access and this group may relatively have better access to health care.[134] Additionally deprivation is accounted for in the model. Increased deprivation is associated with increased mortality and this may account for the differences in ethnic groups. The deprivation data is generally complete in the THIN dataset because it is derived from the post code from census data.[136]

As expected, all co-morbid conditions included in the analysis; diabetes, CVD, CHD, PVD, heart failure and atrial fibrillation were associated with increased mortality.[102;196;272;273] These conditions represent the most vulnerable of CKD patients as they predispose patients to CKD and also increased mortality.

#### *4.5.3.1.Transformed Variables and the relationship with medication*

##### 4.5.3.1.1.Body Mass index, Anaemia, GFR and Proteinuria

So far the results have not proven contradictory to current medical opinion. However the transformed continuous variables do not behave predictably. Patients who have a body mass index of 25 kg/$m^2$ or above in the general population are more likely to die or suffer from cardiovascular disease.[128] However, in patients with CKD the relationship is less clear as there is some evidence to suggest a U shaped relationship with mortality, i.e. being underweight or morbidly obese may be associated with worse outcomes but patients with overweight and obese body mass indices may have the best survival.[108;274;275] In my analysis body mass index convincingly showed a U shaped or inverse J shaped relationship with mortality. One possible hypothesis is that the MDRD equation may underestimate eGFR in obese individuals. This particularly happens in patients over a BMI over 29 kg/$m^2$.[276;277] and is particularly salient to the CKD cohort in my analysis as the median BMI was 29 kg/$m^2$. If the GFR is better at higher BMIs, then this may offset the increased risk of death usually associated with higher BMIs and explain the improved survival.

It is predictable that as haemoglobin rises the risk of death falls. A lower haemoglobin is associated with a higher risk of death in the general population.[132] In patients with CKD anaemia is primarily due to erythropoietin deficiency but may also be due to inflammation, iron deficiency and increases oxidative stress.[278] Patients treated with iron supplementation may be at greater risk for death as they are more likely to be anaemic. This is a form of confounding by indication,[279] where anaemia a risk factor of death and or CVD pre disposes to iron use and therefore death. This analysis also suggests that oral iron supplementation may not be effective at correcting iron deficiency anaemia and this may be due to impaired oral absorption of iron in CKD patients.[280]

The risk of death associated with decreasing GFR and increasing proteinuria is well documented by several other studies and this likely due to worsening metabolic complications as renal function worsens.[34;43;43;93-95;159;196;265;265]

#### *Lipid and Lipid Lowering*

The relationship in this analysis between cholesterol, cholesterol lowering agents and the risk of death was complex. This analysis showed on one hand that an increasing cholesterol until a relatively high total cholesterol is associated with reduced risk of death but on the other hand that cholesterol lowering agents such as statins are associated with reduced risk of death. Evidence from dialysis patients (i.e. those patients with the most severe form of kidney disease), suggests that they are more likely to survive if the cholesterol increases and this relationship was consistent with

my analysis.[124] However in patients with ESRD this is likely to be because a lower cholesterol is associated with systemic inflammation and malnutrition.[281] CKD patients, regardless of stage, are more likely to have systemic inflammation and other lipid abnormalities such as high triglycerides and lower HDL cholesterol. An increase in total cholesterol may represent patients with increased HDL cholesterol which is protective.

Patients treated with lipid lowering agents are more likely to survive than those not treated.[121] This contradicts the relationship between total cholesterol and mortality in my study but supports earlier trial evidence from the Study of Heart and Renal Protection where patients randomised to simvastatin and ezetimibe on average experienced LDL reduction and vascular risk reduction.[121] Interestingly, there was no reduction in mortality as perhaps this study was not powered to detect this. Treatment with statins may have additional effects other than lowering LDL cholesterol which may be protective.[105] This may partially explain the complex relationship between total cholesterol and risk of death. Patients with a total cholesterol above 5 mmol/l are more likely to be treated with lipid lowering agents and that is why the risk appears to reduce as cholesterol decreases at this threshold. However in treated patients the risk would increase if the cholesterol is very high and that is why there is U shaped relationship. This is another form of confounding by indication.

*Blood pressure and Anti-hypertensives*

The relationship between systolic blood pressure and mortality has always been described in the literature as U or J shaped.[119] In the general population the risk of cardiovascular disease and mortality increases above a systolic blood pressure of 135 mmHg in patients with diabetes and 145mmHg in those with hypertension.[111;225] However in my analysis, the turning point for where risk increases was considerably higher occurring at approximately a systolic blood pressure of 166mmHg. The recommendations for lowering systolic blood pressure beyond 140 mmHg are derived from patients with diabetic nephropathy.[116] In CKD patients, limited studies do not show a reduction below targets of 140 mmHg as either beneficial or harmful as shown in a meta-analysis.[282]

In a CKD Collaborative meta-analysis, patients with or without hypertension (defined as systolic blood pressure above 140mmHg) had a similar risk of mortality when their GFR fell below 95 ml/min/1.73m$^2$ and although not significant, the risk of death or ESRD was higher in those without hypertension.[283] A recent study examined the impact of blood pressure on mortality in 651 749 patients with CKD.[284] The group split blood pressure into 10mmHg groups and found that the risk of death was the same for patients with a blood pressure in the range of 130-160mmHg. The author of the study hypothesized that it was reduction in a normal diastolic blood pressure that was associated with greater mortality; however this was not significant in my analysis.[284]

This evidence from the literature conflicts with the protective use of some anti-hypertensives observed in my analysis (with the exception of diuretics and calcium channel blockers).[115] Angiotensin blockers have been shown in large meta-analysis to reduce cardiovascular outcomes in patients with CKD and my analysis shows that the risk of death is reduced in CKD patients on angiotensin blockers.[229] Angiotensin blockers reduce proteinuria in CKD patients.[117] As increasing proteinuria increases risk of death in my analysis, angiotensin blockers may reduce the risk of death independently of lowering blood pressure by reducing proteinuria.[117] Beta-blockers again may have additional protective effects, as CKD patients may have increased sympathetic activity and suppression of this may provide other benefits additional to blood pressure lowering agents.[120] Beta-blockers may additionally reduce the progression of renal disease in CKD patients and this may also protect from death, as worsening eGFR is associated with poorer survival and increased cardiovascular risk.[222;285]

Calcium Channel blockers were not significant in protecting from mortality in the final model and this is contradictory to previous evidence in the general population that showed a protective effect.[115] However in some studies with CKD patients this relationship has not been demonstrated.[222;231] The anti-hypertensive agents associated with better survival reduce blood pressure by blocking the renin angiotensin aldosterone system (RAAS) and calcium channel inhibitors act

independently of this. Blocking RAAS is reno-protective and therefore calcium channels may demonstrate no benefit.[286]

Another surprising finding is that diuretics (including thiazide and loop diuretics) were associated with increased death.[115] This may be because diuretics are more likely used in patients with fluid retention and or heart failure and therefore are a marker of increased co-morbidity.[7] More surprisingly, is that those on other hypertensives, i.e. mostly patients on Doxazosin, were more likely to survive than those not on these agents. This is again contrary to the ALLHAT study in the general population and there is no evidence that this agent is any less or more effective than other agents in reducing mortality. Perhaps it is confounding by indication as Doxazosin is not tolerated in patients with postural hypotension (a marker for mortality in its own right) and therefore healthier individuals may be on this drug and men due to use prostatic hypertrophy.[287] This may require further investigation. It is worth discussing confounding by indication in the context of these results. Treated patients may differ from those who not treated for example characteristics that to lead patients being selected for particular agents may increase or decrease their survival.

***Remaining risk factors***

Confounding by indication may also explain the fact that treatment with anti-platelet agents, anticoagulation (warfarin) and Vitamin D supplementation, was associated with poorer survival. Though anti-platelet agents and anti-coagulation are associated with better survival in some randomised trials and meta-analysis, these patients are

more likely to have co-morbidity that may not be accounted for in the analysis.[110;120] Additionally patients on vitamin D supplementation may be patients at risk for falls on Vitamin D3 replacement or patients with renal bone disease on alphacalcidol.[248] These factors are associated with increased mortality.[248] NSAID use was not significant in this study, even though widely used during the time of CKD diagnosis. NSAID use may only be significant in the very elderly and in higher cumulative doses.[288] Diastolic blood pressure was not a significant co-variable in the analysis and this supports evidence that systolic blood pressure is a better predictor of cardiovascular risk from trials in the general population.[289;290] This may also be the case in patients with CKD.[291]

In terms of prognostic models for mortality: this is one of the largest models conducted in CKD using routinely collected data.[292] All previous models have less candidate predictors then my model and may reflect the smaller sample sizes involved. In the complete case analysis the size of the study population was very small and there were far less significant predictors and they did not change direction or shape when analysed. This may be due to lack of power as optimism decreases when there are events per co-variable and this reduces overfitting. This emphasizes the use of the full cohort is preferential to discern the best predictors.[260]

**4.5.4.Limitations and Strengths**

This analysis had several limitations which will be discussed further in this section. The first is the generic bias associated with health care databases and UK primary care

databases.[146] There is a form of selection bias where patients with recorded information may be different from the general population.[144] Patients who attend for blood tests, blood pressure checks and who request repeat prescriptions are more likely to do better than non-compliant and non-engaging individuals.[293] However conversely healthy and well patients are less likely to have a check-up i.e. inverse care law. Additionally there was no linked data to secondary care and CKD 3 patients with complications may have been managed differently. Despite these limitations, patients selected from primary care are more likely to be representative of the general populations than those from hospital populations where selection bias mentioned above is likely to be more extreme.[125]

There were missing data especially for more 'specialised' measurements such as urine ACR but blood pressure data were quite complete. Multiple imputations were performed but assumptions were made that that the data were missing at random. This dataset relied heavily on Read code recording, and we have previously shown that for QOF CKD recording this may be inaccurate.(Chapter 3) However I used non QOF Read codes to identify comorbidities and it is known that for cardiovascular disease, Read code recording is reasonably accurate.[138;155]

There is poor recording of ethnicity, endemic to all GP computer systems and therefore THIN, GPRD and QRESEARCH and previous work has shown that this may result in non-coding of up to half of patients of non-white origin.[294] This may lead to the associations of better survival in ethnic minorities being questioned. There may be

patients with newer registrations included. The impact of ethnicity on CKD recognition will be discussed later. A further limitation is that although the THIN database provides an accurate record of prescribing there is no evidence that the drugs were dispensed and or taken.

In a complex multivariable analysis presented in this whether co-variables cause or prevent death cannot be determined. Firstly this is a retrospective observational study (although the data is prospectively collected). The information was not collected a priori for the specific analysis of outcomes in CKD patients and therefore only association of the co-variables with the outcome not causation.[109] This is because patient demographics collection, patient measurements and patient treatments are not determined by the investigators. This is exemplified by treatment assignment. Treatment is not randomly assigned and patient characteristics and clinician preference will govern this. This is confounding by indication and may lead to unusual relationships between co –variables and outcome for example blood pressure and death. [279]

Additionally in a large model like this interactions between co-variables were not assessed. However this would have been conceptually difficult as fractional polynomials were used and found to have a better fit in the model. Interaction terms between fractional polynomials and other polynomials and other co-variables are very complex to analyse and were outside the scope of this thesis. This could be considered

in the future using stata software which can analyse interactions between fractional polynomials and co-variables using the MFP gen algorithm.

Limitations specific to this analysis are firstly CKD recognition: this has been discussed earlier (Chapter 2) but this paragraph will reiterate the issues of method of creatinine analysis and the lack of black ethnicity data. To explain the first point, by using the lower creatinine of two blood tests at least seven days apart and using the non IDMS MDRD equation to calculate the eGFR led to under classification of CKD. However using two blood tests seven days apart will have increased the precision of creatinine measurement.[276] Regarding the question of black ethnicity: between 2001 and 2011 the UK national census reported that black ethnicity was between 2.2% and 3.3% and even if ten percent of these patients had CKD, the prevalence would only fall by 0.3%.

Additional limitations were that patients were censored after they had de registered leading to unequal lengths of follow up. Patient information were excluded before AMR date and after at least 6 registration at the practice. This was because it was unlikely that their records would be accurate prior to this, however this could have led to patients with CKD being excluded.[150] The cause of death was unknown as THIN was not linked to ONS. The model predominantly examined cardiovascular risk factors and an assumption was made that the majority of deaths were due to cardiovascular disease.

The major strengths of this study are that this was a large cohort generally representative of the UK population. Although there may be some selection bias i.e. those who got blood tests, measurements and prescriptions, the selection bias is likely to be less than in patients from hospitals and volunteers for screening studies.[136] The other advantage of this CKD population is that it is a UK primary care based cohort where health care is free at the point of access and therefore this makes this one of largest generalisable CKD cohort studies in the world. There is no recall bias as clinical data are collected prospectively. The prescribing information is accurate and has been verified for research.[146]

In many prognostic CKD studies when data is missing it is excluded or substituted with common values such as the mean or median.[284;292] Excluding patients with missing data can lead to considerably smaller sample sizes and their analyses may lose power and miss important predictors. Therefore multiple imputation was necessary in this analysis. Continuous data were not dichotomized and nor were they analysed linearly. By using fractional polynomials, important non-linear relationships were exposed and these transformed variables models had better fit than if analysed linearly. A random effects term for practice further improved model fit and again shows that this is important in such analyses where there may fundamental differences in clinical practice and case mix. The model included enough events to avoid substantial overfitting.[250]

## 4.6.Executive Summary

- Cardiovascular Disease and mortality in CKD patients are associated with traditional and CKD specific risk factors
- Cox proportional hazards model allow multivariable analysis of censored outcomes.
- When considering models: missing data and non linear functional forms have to be considered.
- In clustered data frailty models have to be considered.
- This dataset had greater than 15% missing data for several variables and therefore five imputed datasets were created.
- The optimal model was one that incorporated fractional polynomials and a frailty term.
- The model confirmed the association between traditional risk/prognostic factors for mortality such as increasing age, GFR, angiotensin blockage and statins.
- However other risk /prognostic factors such as blood pressure, cholesterol, blood thinning agents had an unconventional relationship with mortality.

# CHAPTER 5. CAN ROUTINELY COLLECTED PRIMARY CARE DATA PREDICT CARDIOVASCULAR DISEASE OR ALL CAUSE MORTALITY IN STAGE 3 CKD PATIENTS?

This Chapter describes the development of a new prognostic model which aims to identify potential predictors of the composite outcome of CVD or all cause mortality in CKD3-5 patients. No separate introduction and a brief overview of the methods are provided since the issues and approach used was similar to that in Chapter 4. The results are discussed in the context of the current literature.

## 5.1.Methods

### 5.1.1.Model Cohort

Patients with CKD stage 3a and 3b were characterized by two consecutive estimated Glomerular filtration rates (calculated from serum creatinine at least seven days apart using the non IDMS MDRD equation[13]). Patients had to have survived or suffered no cardiovascular events in subsequent three months to be included in the model. Patients were aged 18 or over at time of diagnosis of CKD and registered at the practice for greater than 6 months. All data post the AMR date was considered for the practice.[150] Patients entry in the cohort was the time of their initial diagnosis of CKD i.e. the first CKD 3a or 3b staging.

### 5.1.2.Outcome Measures:

The model outcome was a composite of time to cardiovascular event or all cause mortality. CVD or all-cause mortality was predefined by Read codes including the QOF business Read codes and additional codes identified by the doctoral researcher using her clinical expertise.[45] It was assumed that the absence of Read code or time of death meant that the patient did not incur an event. Patients were censored if they were alive at the end of the study period, had not suffered a CVD event or when they deregistered from the practice. Patients had to have survived a minimum of three months to allow appropriate modelling of risk factors.

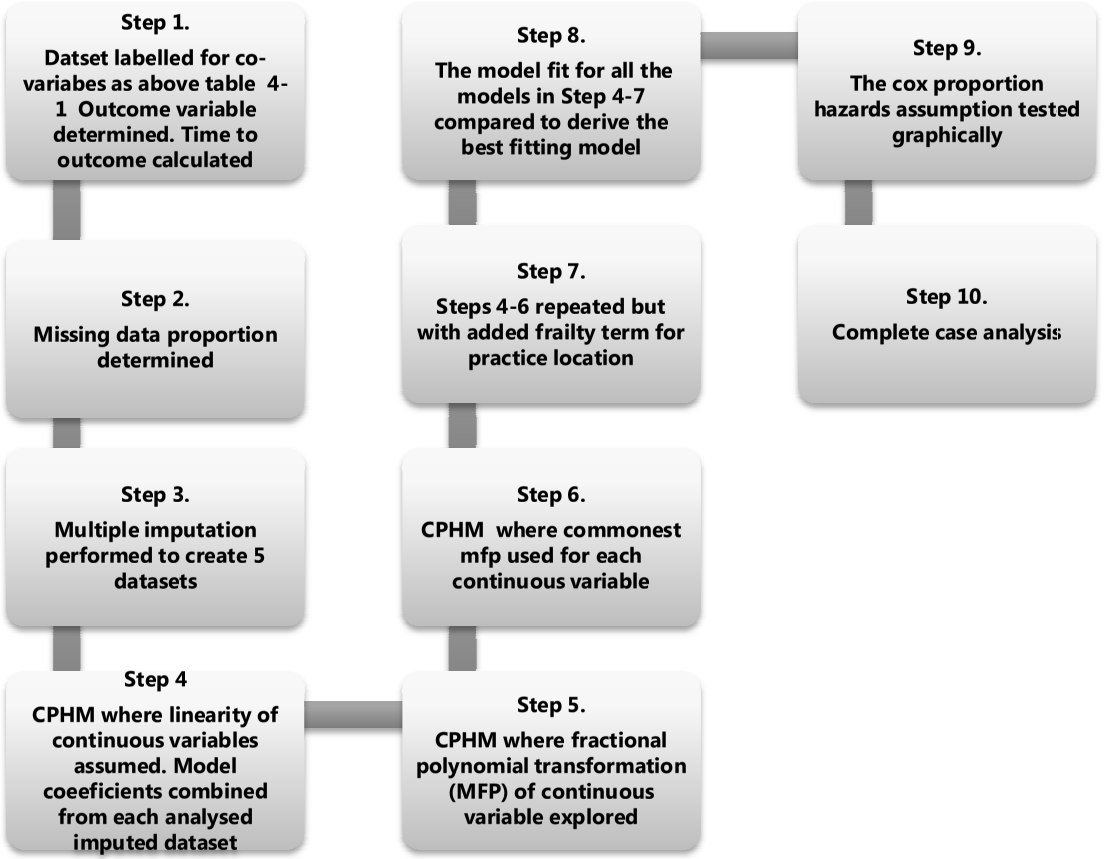
#### *5.1.2.1.Co-Variables*

The same covariables were included in the model as identified in the previous chapter (all-cause mortality model) (Table 4-1) with the exception of CVD as this was included in the composite outcome.

#### *5.1.2.2.Cox Proportional hazards-Model*

The same method of model development was adopted as described in Chapter 4 and is summarised in Figure 5-1.

Figure 5-1. Summary of Cardiovascular disease/all cause mortality model.

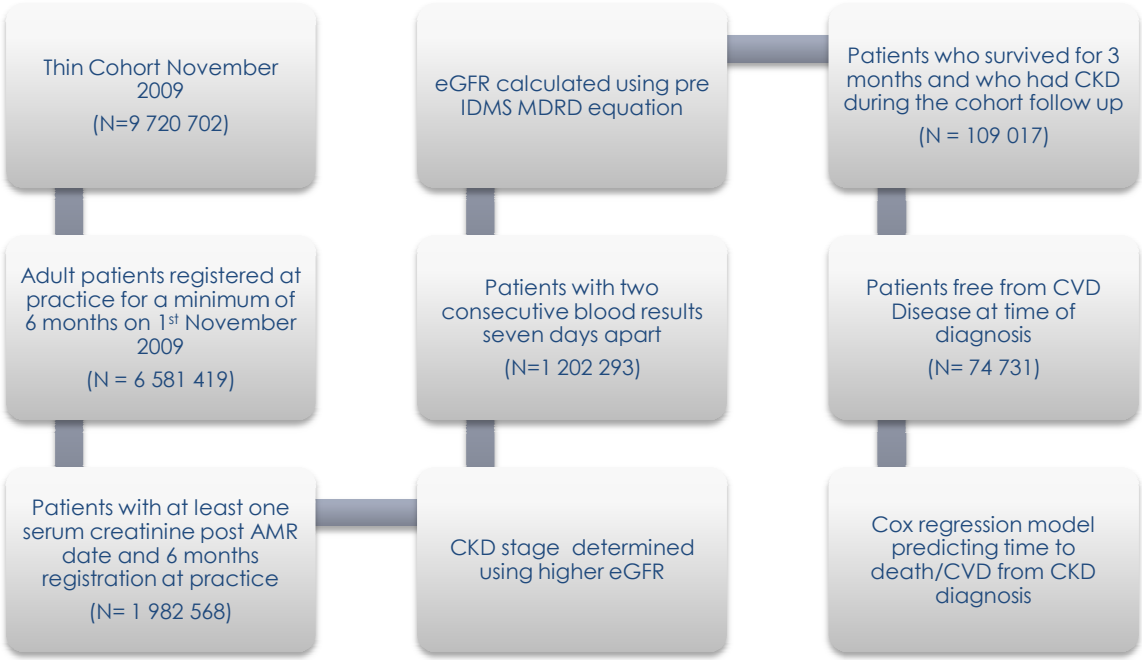


## 5.2. Results

### Cohort demographics

The cohort was derived from patients with two consecutive serum creatinines (N= 1 202 293) seven days apart where both calculated eGFRs were below 60 and the patient had survived a minimum of 3 months (N = 109 017). There were 74 731 patients with CKD 3 who were free of CVD at the time of diagnosis who were used for this analysis.(Figure 5-2)

**Figure 5-2. Derivation of THIN CKD cohort for composite outcome**




Patient demographics are shown in Table 5-1**.** The median age was 71.3 years and the majority of patients were female (67.6%). The majority of patients had either no recorded ethnicity or were of Caucasian ethnicity (98.1%) (Table 5-1). The median cholesterol was 5.3 mmol/l, the body mass index (BMI) was 28.1 kg/m$^2$, the haemoglobin (Hb) was 13.5 g/dl, blood pressure (BP) was 142.5/79.7 mmHg.

**Table 5-1. Demographics of the composite model cohort (if not stated in brackets then proportion of data complete)**

| Demographic feature | Figure or proportion |
|---|---|
| **Median Age in years at time of diagnosis (years)** | 71.3 (23-107) |
| **Gender – Female** | 50517 (67.6%) |
| **Not recorded/Caucasian Race**<br>**African –Caribbean**<br>**Indian Subcontinent**<br>**South East Asian** | 73317(98.1%)<br>683 (0.9%)<br>690(0.9%)<br>41(0.1%) |
| **Townsend quintiles (95.7%)***<br>**1**<br>**2**<br>**3**<br>**4**<br>**5** | <br>18855(26.4%)<br>17477(24.4%)<br>14762(20.6%)<br>12569(17.6%)<br>7857(11.0%) |
| **Median eGFR in mmol/min/1.73m$^2$** | 54.3 (3-64) |
| **Median Cholesterol (mmol/l) (60.8%)*** | 5.3(1.7-13.9) |
| **Median Body Mass Index (kg/m2) (38.2%)*** | 28.1(11.2-59.5) |
| **Median Haemoglobin (g/dl) (64.2%)*** | 13.5(3.3-25.8) |
| **Mean Systolic blood pressure (mmHg) (82.5%)*** | 142.5(20.5) |
| **Mean Diastolic blood pressure (mmHg) (82.5%)*** | 79.7(11.3) |
| **Proteinuria levels (13.6%)***<br>**None**<br>**High**<br>**Very High** | <br>1876(18.5%)<br>7586(74.8%)<br>675(6.7%) |
| **Angiotensin Blockade** | 30716(41.1%) |
| **Beta blockers** | 16522(22.1%) |
| **Calcium Channel Blockers** | 13143(17.6%) |
| **Diuretics** | 30826(41.3%) |
| **Other Anti-hypertensives** | 4421(5.9%) |
| **Lipid lowering medication** | 21455(28.7%) |
| **Iron Supplementation** | 3882(5.2%) |
| **Vitamin D** | 6153(8.2%) |
| **Non Steroid Anti Inflammatory Drugs** | 15301(20.2%) |

***PROPORTION OF DATA AVAILABLE**

Patients had a median eGFR of 54.3 mmol/litre/1.73m$^2$, 20.1% had diabetes mellitus, 8.5% had atrial fibrillation, 5.0% had heart failure and 17.4% were current or ex smokers (Table 5-1**).** A significant proportion of patients were on anti-platelets medication (22.6%), angiotensin blockade (41.1%), beta-blockers (22.1%), diuretics

(41.3%) and lipid lowering medication (28.7%) Interestingly 20.2% of patients had a recent prescription of non-steroidal anti-inflammatory drugs (NSAIDS) (Table 5-1).

### 5.2.1. Step 1. Outcomes experienced by the Cohort

The median follow up for the cohort was 1034 days (range 90 to 1079 days). There were 12 048 events, of which 6511 were deaths and 5537 cardiovascular events (i.e. first ever Read code for cardiovascular event). In patients who incurred cardiovascular events, there were 2859 patients who had a Read code for CHD, 2127 with a Read code for cerebrovascular accident and 947 patients who had a Read code for peripheral vascular disease. Please note that the total number of events were more than 5537 cardiovascular events as some patients had 1 or more new Read codes.

### 5.2.2.Step 2. Variables with missing data

Only 3247 patients had a complete set of variables and the relative proportions are shown in Table 5-1.

### 5.2.3. Step 3. Missing data analysis and subsequent imputation

High proportions of data were missing and found to have an arbitrary missing pattern (data not shown). The incomplete data were graphed to see if they were normally distributed. Serum cholesterol and BMI were positively skewed and hence were log transformed and following this they appeared normally distributed. (Appendix C. Figures 1 to 2). Systolic and diastolic BP were normally distributed but Hb was negatively skewed and hence reflected and then log transformed to generate a normal distribution (Appendix C. Figures 3 to 5). After multiple imputation was undertaken using multiple chained equations, the distribution of imputed variables were examined and are shown in Table 5-2.

**Table 5-2. Imputed datasets for composite model: The distribution of imputed variables**

| Variable with imputed data | Original Data | Imp 1 | Imp 2 | Imp 3 | Imp 4 | Imp 5 |
|---|---|---|---|---|---|---|
| **Median Cholesterol (mmol/l)** | 5.3(1.7-13.9) | 5.37 | 5.38 | 5.37 | 5.37 | 5.38 |
| **Median Body Mass Index (kg/m2)** | 28.1(11.2-59.5) | 27.57 | 27.50 | 27.59 | 27.60 | 27.53 |
| **Median Haemoglobin (g/dl)** | 13.5(3.3-25.8) | 13.51 | 13.50 | 13.51 | 13.50 | 13.50 |
| **Mean Diastolic blood pressure (mmHg)** | 142.5(20.5) | 79.92 | 79.89 | 79.90 | 79.91 | 79.88 |
| **Mean Systolic blood pressure (mmHg)** | 79.7(11.3) | 141.92 | 141.87 | 141.96 | 141.88 | 141.87 |
| **Proteinuria levels** | | | | | | |
| **None** | 1876 (18.5)* | 7827 (10.5) | 7738 (10.4) | 7751 (10.4) | 7711 (10.3) | 8102 (10.8) |
| **High** | 7586 (74.8) | 61085 (81.7) | 61695 (82.6) | 61408 (82.2) | 61345 (82.1) | 61321 (82.1) |
| **Very High** | 675 (6.7) | 5819 (7.8) | 5298 (7.1) | 5572 (7.46) | 5675 (7.6) | 5308 (7.1) |
| **Townsend quintiles** | | | | | | |
| **1** | 18855 (26.4)* | 19696 (26.4) | 19637 (26.3) | 19700 (26.4) | 19719 (26.4) | 19656 (26.3) |
| **2** | 17477 (24.4) | 18247 (24.4) | 18289 (24.4) | 18277 (24.5) | 18238 (24.4) | 18242 (24.4) |
| **3** | 14762 (20.6) | 15412 (20.6) | 15434 (20.6) | 15406 (20.6) | 15422 (20.6) | 15462 (20.7) |
| **4** | 12569 (17.6) | 13141 (17.5) | 13138 (17.6) | 13138 (17.6) | 13136 (176) | 13111 (17.5) |
| **5** | 7857 (11.0) | 8235 (11.0) | 8225 (11.0) | 8210 (11.0) | 8216 (11.0) | 8260 (11.1) |

(Note range will be the same as whole cohort as min and maximum specified in imputation)

### 5.2.4. Step 4. Analysis of co-variables where the relationship between continuous co-variables and time to outcome is assumed to be linear.

The following section details the results of a Cox proportional hazards model where the continuous co-variables were assumed to be linearly associated with time to outcome. The results of the Cox regression analyses for each imputation are shown in Appendix C: Tables 1-5 and the combined estimates are shown in Table 5-3. The variables removed in backward selection in each Imputation were diastolic BP, NSAIDS, beta-blockers and calcium channel blockers. As lipid lowering agents were significant only in the fifth Imputation, this co-variable was removed from the analysis.

**Table 5-3. Combined coefficients from Cox proportional hazards model for composite outcome where linear risk of numerical variables is assumed. (no frailty term)**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|---|
| **Age** | 0.05 | 0.00 | 1.05 | 1.05 | 1.05 |
| **Male Gender** | -0.49 | 0.02 | 1.64 | 1.56 | 1.69 |
| **Race with Caucasian race as reference**<br>**African –Caribbean** | -0.43 | 0.16 | 0.65 | 0.48 | 0.89 |
| **Indian Subcontinent** | -0.26 | 0.12 | 0.77 | 0.61 | 0.97 |
| **South East Asian** | -0.51 | 0.58 | 0.60 | 0.19 | 1.87 |
| **Townsend Quintile compared to 1st Quintile**<br>**2nd** | 0.09 | 0.03 | 1.09 | 1.04 | 1.15 |
| **3rd** | 0.12 | 0.03 | 1.13 | 1.07 | 1.19 |
| **4th** | 0.15 | 0.03 | 1.16 | 1.10 | 1.23 |
| **5th** | 0.20 | 0.03 | 1.22 | 1.14 | 1.30 |
| **Diabetes Mellitus** | 0.17 | 0.02 | 1.18 | 1.12 | 1.24 |
| **Heart Failure** | 0.54 | 0.03 | 1.72 | 1.62 | 1.82 |
| **Atrial Fibrillation** | 0.22 | 0.03 | 1.25 | 1.18 | 1.32 |
| **Ever Smoked** | 0.23 | 0.02 | 1.26 | 1.21 | 1.33 |
| **Systolic BP** | 0.00 | 0.00 | 0.9976 | 0.9967 | 0.9985 |
| **BMI** | -0.02 | 0.00 | 0.98 | 0.98 | 0.99 |
| **Haemoglobin** | -0.11 | 0.01 | 0.89 | 0.88 | 0.90 |
| **Glomerular Filtration Rate** | -0.02 | 0.00 | 0.98 | 0.98 | 0.99 |
| **Cholesterol** | -0.03 | 0.01 | 0.97 | 0.96 | 0.99 |
| **Proteinuria levels with none as reference**<br>**High** | 0.15 | 0.03 | 1.16 | 1.09 | 1.24 |
| **Very High** | 0.45 | 0.04 | 1.57 | 1.44 | 1.71 |
| **Aspirin** | 0.22 | 0.02 | 1.25 | 1.20 | 1.30 |
| **Anticoagulation** | 0.20 | 0.04 | 1.22 | 1.13 | 1.32 |
| **Angiotensin Blockade** | -0.17 | 0.02 | 0.85 | 0.81 | 0.88 |
| **Diuretic use** | 0.08 | 0.02 | 1.08 | 1.04 | 1.13 |
| **Other** | -0.12 | 0.04 | 0.88 | 0.82 | 0.95 |
| **Iron Medication** | 0.14 | 0.03 | 1.15 | 1.08 | 1.23 |
| **Vitamin D supplementation** | 0.17 | 0.03 | 1.19 | 1.12 | 1.26 |

#### *5.2.4.1.Covariables associated with worse outcomes (Table 5-3)*

Increasing age (for every year HR 1.05, 95% CI 1.05 to 1.05), male gender (compared to female HR 1.64, (1.56 to 1.69), increasing Townsend quintile diabetes (HR 1.18, 1.12 to 1.24), heart failure (HR 1.72, 1.62 to 1.82), atrial fibrillation (HR 1.25, 1.18 to 1.23) and patients who had previously smoked (HR 1.26, 1.21 to 1.33) were independently associated with increased rate of events (Table 5-3), Increasing proteinuria was associated with worse survival. Anti-platelet agents (HR 1.25,1.20 to 1.30), anticoagulation (HR 1.22,1.13 to 1.32), diuretic use (HR 1.08 1.04 to 1.13), iron medication (HR 1.15,1.08 to 1.23) and vitamin D supplementation (HR 1.19,1.12 to 1.26) were all associated with worse survival.

#### *5.2.4.2.Predictors associated with less CVD/Death (Table 5-3)*

African Caribbean ethnicity (HR 0.65, (95% CI 0.48 to 0.89) and Indian Sub-continental ethnicity (HR 0.77, 0.61 to 0.97) were associated with better survival but South East Asian ethnicity was not. Increasing haemoglobin (HR for every unit increase 0.89, 0.88 to 0.90), eGFR (HR 0.99, 0.98 to 0.99), cholesterol (HR 0.97, 0.96 to 0.99) systolic blood pressure (HR 0.9976, 0.9967 to 0.9985) and body mass index (HR 0.98, 0.98 to 0.99) were associated with better survival. Angiotensin blockade (HR 0.85, 0.81 to 0.88) and other anti-hypertensive medication (HR 0.88, 0.82 to 0.95) were also associated with better survival. The Akaike Information Criteria (AIC), in indicator of model fit, was improved with each model with a baseline of 260757. (Table 5-4).

**Table 5-4. AIC for each imputed dataset for Table 5-3**

| Imp Dataset | Akaike Information Criteria |
|---|---|
| **1** | 252133. |
| **2** | 252125. |
| **3** | 252220. |
| **4** | 252204. |
| **5** | 252139. |

### 5.2.5.Step 5: Analysis of transformed variables.

This section details the Cox proportional hazard models where continuous variables were transformed up to 2 fractional polynomial terms ($X^p$ + $X^p$ where p is in the family 2, -1, -0.5, 0, 0.5, 1, 2 and 3, where 0 is the terminology for log). Each imputed model was analysed using the MFP algorithm in R. The detail of each individual coefficient derived from each analysed imputed dataset are shown in Appendix C: Tables 6-10. The algorithm suggested a different mfp transformation for the variables of systolic blood pressure, haemoglobin, body mass index and GFR (Table 5-5).

**Table 5-5. Suggested transformation of variables from mfp algorithm in R for each imputation for composite model**

| Variable | Imputation 1 | Imputation 2 | Imputation 3 | Imputation 4 | Imputation 5 |
|---|---|---|---|---|---|
| **Age** | Age squared | Age squared | Age squared | Age squared | Age squared |
| **Systolic blood pressure** | $(x/100)^{0.5}$ + $(x/100)^{1}$ | $(x/100)^{2}$+ $(x/100)^{2}$* log(x/100) | $(x/100)^{1}$ + $(x/100)^{2}$ | $(x/100)^{1}$ + $(x/100)^{2}$ | $(x/100)^{1}$+ $(x/100)^{1}$* log(x/100) |
| **Haemoglobin** | $(x/10)^{-2}$+ $(x/10)^{-1}$ | $(x/10)^{3}$+ $(x/10)^{3}$* log(x/10) | $(x/10)^{3}$+ $(x/10)^{3}$* log(x/10) | $(x/10)^{3}$+ $(x/10)^{3}$* log(x/10) | $(x/10)^{3}$+ $(x/10)^{3}$* log(x/10) |
| **BMI** | $(x/10)^{-0.5}$+ $(x/10)^{3}$ | log(x/10)+ $(x/10)^{3}$ | $(x/10)^{-2}$ | $(x/10)^{-0.5}$+ $(x/10)^{3}$ | $(x/10)^{-2}$ |
| **GFR** | None | squared | None | None | None |
| **Cholesterol** | None | None | None | Not significant | None |

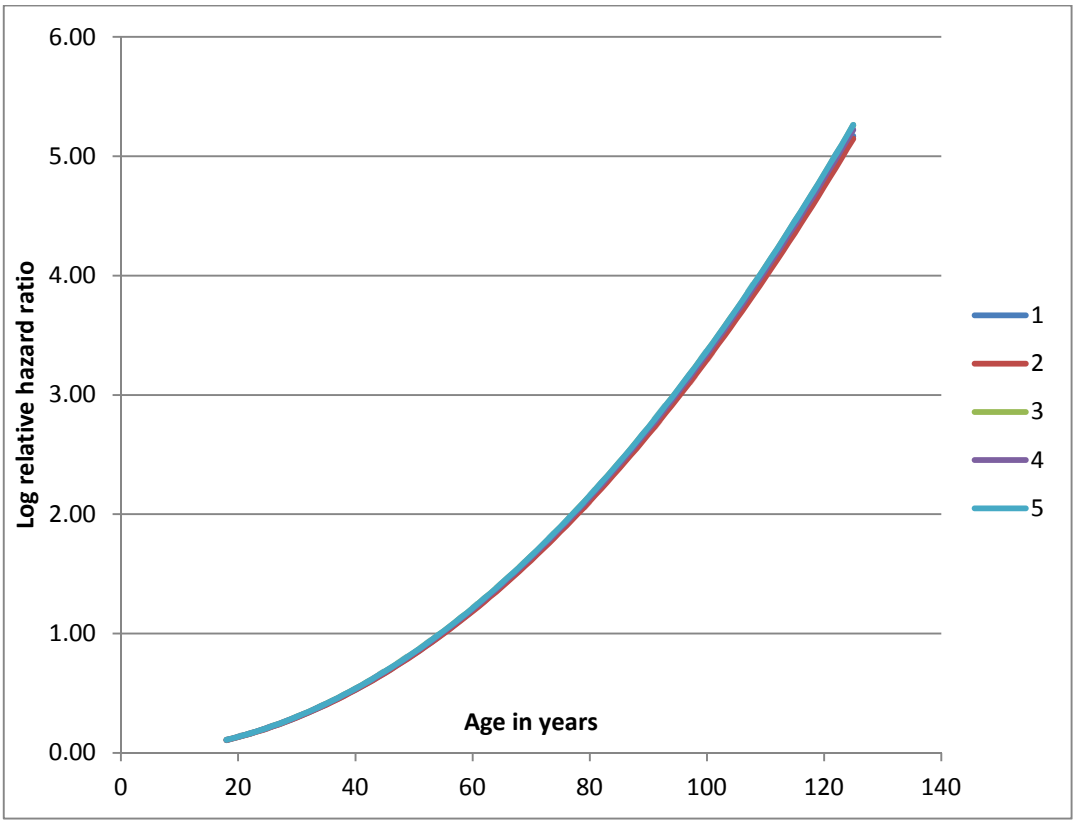
#### *5.2.5.1.Plotting the log relative hazard ratio for each transformation*

Using the coefficients for each variable, the log relative hazard ratio was plotted against the original untransformed variable for the following variables:

- Age (Figure 5-3)
- Systolic blood pressure (Figure 5-4)
- Hb (Figure 5-5)
- BMI(Figure 5-6)

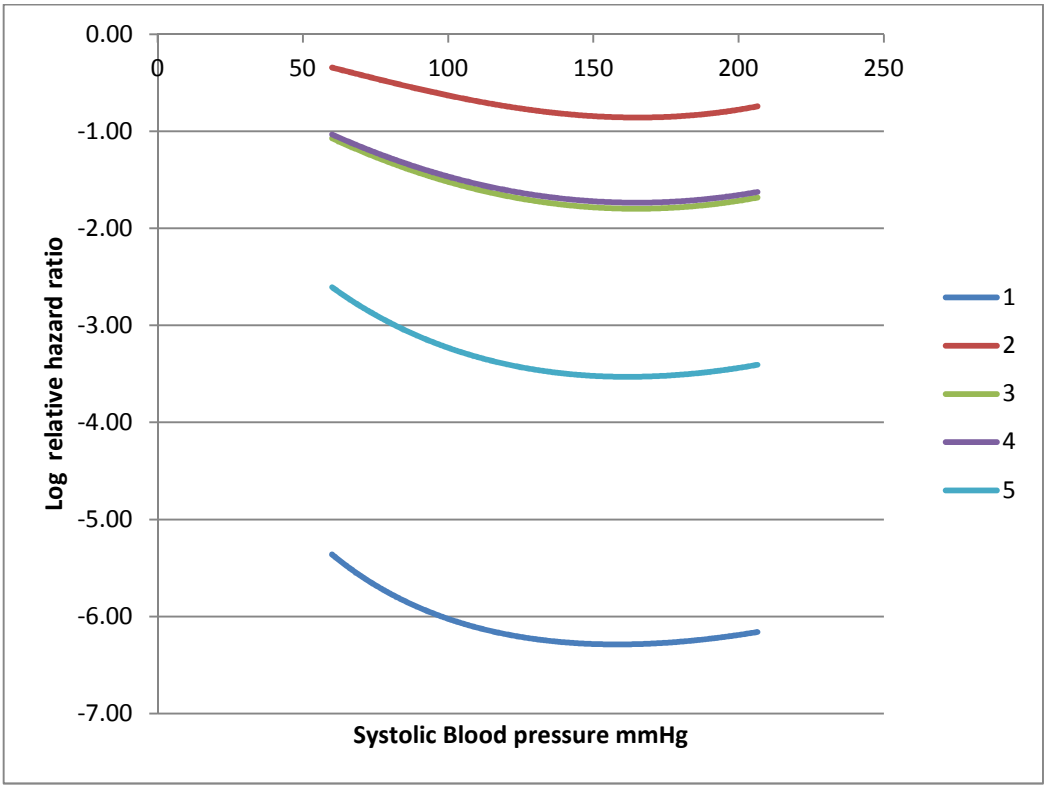
The relationship between age and the hazard function was identical for each imputation (Table 5-5) and the log relative hazard ratio increased at a squared rate as age increased.

**Figure 5-3. Age versus Log relative Hazard Ratio for each imputation (without frailty term)**



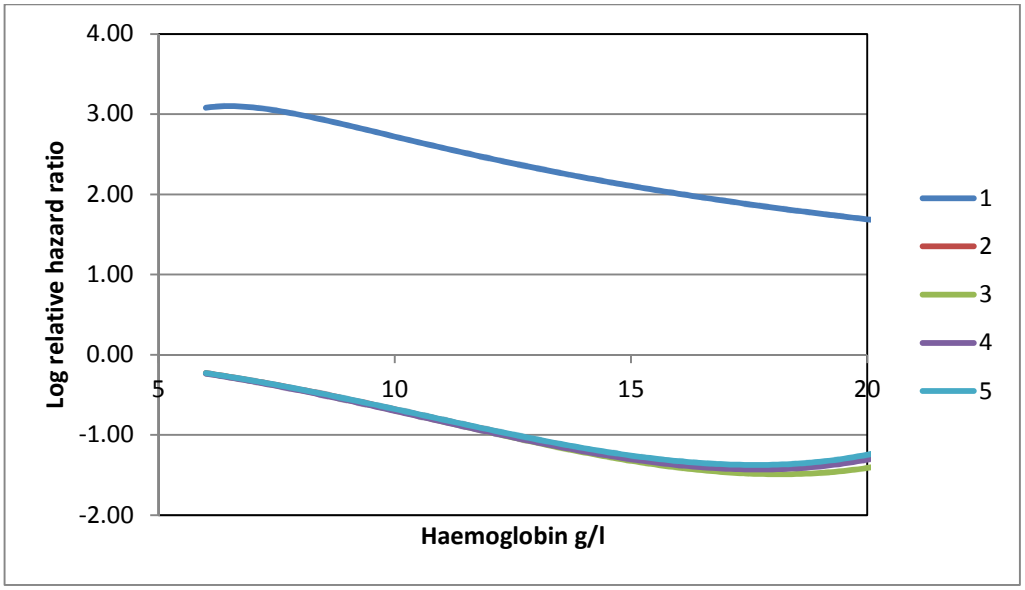
For systolic blood pressure (Figure 5-4), though the values were different for each imputation, the log relative hazard ratio decreased until a systolic of 166mmHg and then increased slightly above this threshold.

**Figure 5-4. Log relative Hazard Ratio versus Systolic BP for each imputation (without frailty term)**



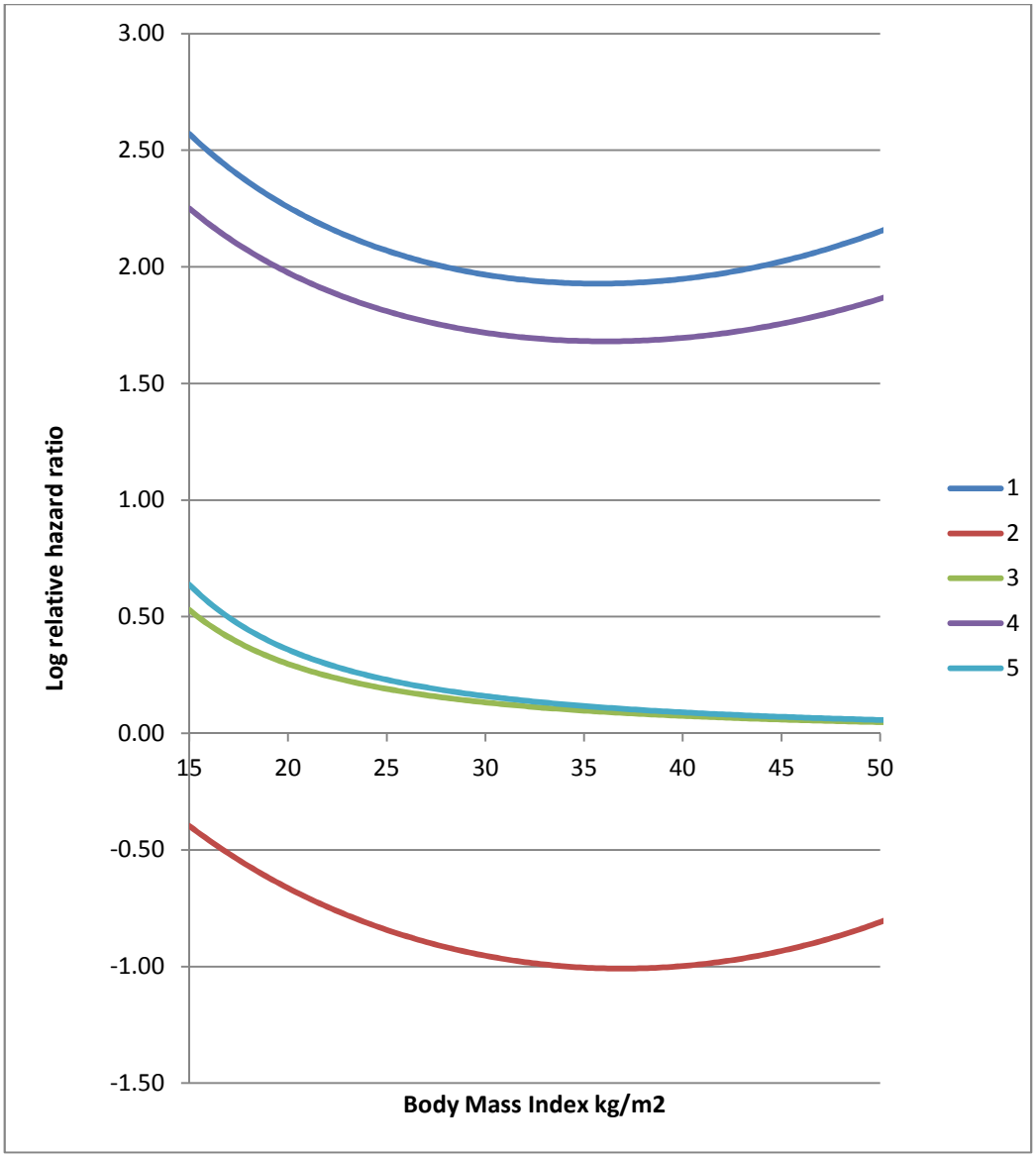

Haemoglobin (Figure 5-5) the relationships were identical across Imps 2-5 where the log relative hazard ratio decreased and then increased.

**Figure 5-5. Haemoglobin versus Log relative Hazard Ratio for each imputation (without frailty term)**




For the variable BMI (Figure 5-6), across Imps 1, 2 and 4, the log relative hazard ratio decreased until BMI reached 36 kg/m$^2$ and then the log relative hazard ratio started to rise. However though the direction of the log relative hazard ratio was the same the log relative hazard ratio is not the same. For Imps 3 and 5 the log relative hazard ratio decreased but plateaud around 36 kg/m$^2$ and the values were almost identical.

**Figure 5-6. Body Mass Index versus log relative hazards ratio for Imps 1 to 5 (no frailty term)**




***5.2.5.2.The relationship of other covariables and risk composite outcome***

Like the combined model in step 4, beta blockade, diastolic blood pressure, calcium channel blockers, non-steroidals, and lipid lowering agents were removed from the final model. In the imputed dataset 5, cholesterol was insignificant. For each model for each imputation, with the exception of transformed variables, coefficients for each co variable showed the same direction as the model in Step 4. Male gender, increasing

Townsend quintile, diabetes mellitus, heart failure, atrial fibrillation, patients who smoke or had previously smoked, increasing proteinuria, aspirin, anticoagulation, diuretic use, iron and vitamin D supplementation were associated with worse outcomes.

Patients with African-Caribbean ethnicity, Indian-Sub continental ethnicity, increasing eGFR, lipid lowering agents, angiotensin blockers and other antihypertensives had better outcomes. The AIC for the analysis for each imputation is shown in (Table 5-6) and was significantly better than previous model results from step 4.

**Table 5-6. AIC for each model with original suggested fractional polynomial for composite model (no frailty term)**

| Imp Dataset | Akaike Information Criteria for models analysed in Step 4: No transformation of continuous co-variable | Akaike Information Criteria for models analysed in Step 5: Where transformations of co-variable in Table 5-5 for each continuous co-variable |
|---|---|---|
| **1** | 252133.5 | 251977 |
| **2** | 252125.2 | 251989 |
| **3** | 252220.3 | 252108 |
| **4** | 252204.9 | 252071 |
| **5** | 252139.4 | 252010 |

### 5.2.6.Step 6. The ('alternative') model with the most consistent transformation

As the analysis of each imputed dataset showed a different a transformation for the systolic blood pressure, BMI, GFR, and Hb it was decided to explore if there was effect on model fit (AIC), if each imputed model was analysed using the commonest suggested transformation. The commonest transformations were:

- Systolic blood pressure - Systolic blood pressure + systolic blood pressure$^2$ (Imps 3 and 4),
- Haemoglobin - haemoglobin$^3$+haemoglobin/10$^3$*log haemoglobin (Imps 2-5)
- Body Mass Index - BMI$^{-2}$ (Imps 3 and 5) or BMI$^{-0.5}$+BMI$^3$ (Imps 2 and 4)
- Age - Age$^2$
- GFR and cholesterol no transformation.

Therefore two models for each imputation were carried out as BMI had two suggested mfp. The AIC was then compared between the original mfp model, and the two 'alternative' models.

#### *5.2.6.1.Model fit statistic for each model*

The 1$^{st}$ alternative model had a lower AIC for each analysed imputed dataset compared to the 2$^{nd}$ alternative model. When original mfp model was compared to the 1$^{st}$ alternative model, there was a significant difference between the model of imputed dataset 1 ($X^2$= 6.0, 1df p <0.02) and model of imputed dataset 2 ($X^2$= 11.0, 1df p <0.001), (Table 5-7). However for the analysis of imputed datasets 3 to 5 there was no significant difference between the AIC of the original mfp model and 1$^{st}$ alternative model.

**Table 5-7. A comparison between the original mode for composite outcome with mfp and the 1st and 2nd best model (no frailty term)**

| Imp Dataset | AIC for original Model | AIC for 1st 'alternative' Model | AIC for 2st 'alternative' Model |
|---|---|---|---|
| **1** | 251977 | 251983 | 251988 |
| **2** | 251989 | 252000 | 252021 |
| **3** | 252108 | 252108 | 252116 |
| **4** | 252071 | 252069 | 252071 |
| **5** | 252010 | 252011 | 252025 |

#### *5.2.6.2.The final combined model using mfp but without frailty*

The final combined model is represented in Table 5-8. The following co variables were associated with poorer survival; male gender(HR 1.65, 95% CI 1.58 to 1.71), increasing Townsend quintile (for e.g. 5th quintile versus the first quintile, HR 1.21, 1.14 to 1.29, diabetes mellitus (HR .1.19, 1.13 to 1.25), heart failure (HR 1.69, 1.59 to 1.79), atrial fibrillation (HR 1.24, 1.17 to 1.31), current/ex-smokers(HR 1.26, 1.21 to 1.32), increasing proteinuria (high versus no proteinuria HR 1.17, 1.09 to 1.25, very high versus no proteinuria HR 1.55, 1.43 to 1.69), anti-platelet prescription (HR 1.25, 1.20 to 1.30), anticoagulation (HR 1.23, 1.14 to 1.33), diuretic use (HR 1.09, 1.05 to 1.13), iron supplementation (HR 1.12, 1.05 to 1.20) and vitamin D supplementation (HR 1.17, 1.11 to 1.25).

**Table 5-8. Combined model for each Imputation for first alternative model for all Imputations. (No frailty term)**

| Risk factor/Prescription | Beta | SE | HR | 95% CI | |
|---|---|---|---|---|---|
| **$(Age/100)^2$** | 3.33 | 0.07 | 27.97 | 24.57 | 31.84 |
| **Male Gender** | 0.50 | 0.02 | 1.65 | 1.58 | 1.71 |
| **Race with African Caribbean**<br>**Race as reference**<br>**African –Caribbean** | -0.46 | 0.16 | 0.63 | 0.47 | 0.86 |
| **Indian Subcontinent** | -0.27 | 0.12 | 0.77 | 0.61 | 0.96 |
| **South East Asian** | -0.54 | 0.58 | 0.58 | 0.19 | 1.81 |
| **Townsend Quintile compared to 1st Quintile**<br>**$2^{nd}$** | 0.09 | 0.03 | 1.09 | 1.04 | 1.15 |
| **$3^{rd}$** | 0.12 | 0.03 | 1.12 | 1.06 | 1.19 |
| **$4^{th}$** | 0.15 | 0.03 | 1.16 | 1.10 | 1.23 |
| **$5^{th}$** | 0.19 | 0.03 | 1.21 | 1.14 | 1.29 |
| **Diabetes Mellitus** | 0.17 | 0.02 | 1.19 | 1.13 | 1.25 |
| **Heart Failure** | 0.52 | 0.03 | 1.69 | 1.59 | 1.79 |
| **Atrial Fibrillation** | 0.21 | 0.03 | 1.24 | 1.17 | 1.31 |
| **Ever Smoked** | 0.23 | 0.02 | 1.26 | 1.21 | 1.32 |
| **$(Systolic\ BP/100)^1$** | -2.19 | 0.34 | 0.11 | 0.06 | 0.22 |
| **$(Systolic\ BP/100)^2$** | 0.66 | 0.12 | 1.94 | 1.55 | 2.43 |
| **$(BMI/10)^{-2}$** | 1.66 | 0.15 | 5.29 | 3.96 | 7.06 |
| **$(Haemoglobin/10)^3$** | -0.69 | 0.05 | 0.50 | 0.46 | 0.55 |
| **$(Haemoglobin/10)^3$ x log (Haemoglobin)** | 0.76 | 0.07 | 2.13 | 1.87 | 2.43 |
| **Glomerular Filtration Rate/10** | -1.43 | 0.11 | 0.24 | 0.19 | 0.30 |
| **Cholesterol/10** | -0.28 | 0.08 | 0.76 | 0.65 | 0.89 |
| **Proteinuria compared to normal**<br>**High** | 0.15 | 0.03 | 1.17 | 1.09 | 1.25 |
| **Very High** | 0.44 | 0.04 | 1.55 | 1.43 | 1.69 |
| **Aspirin** | 0.22 | 0.02 | 1.25 | 1.20 | 1.30 |
| **Anticoagulation** | 0.21 | 0.04 | 1.23 | 1.14 | 1.33 |
| **Angiotensin Blockade** | -0.16 | 0.02 | 0.85 | 0.82 | 0.89 |
| **Diuretic use** | 0.08 | 0.02 | 1.09 | 1.05 | 1.13 |
| **Other** | -0.12 | 0.04 | 0.89 | 0.82 | 0.96 |
| **Iron Medication** | 0.11 | 0.03 | 1.12 | 1.05 | 1.20 |
| **Vitamin D supplementation** | 0.16 | 0.03 | 1.17 | 1.11 | 1.25 |

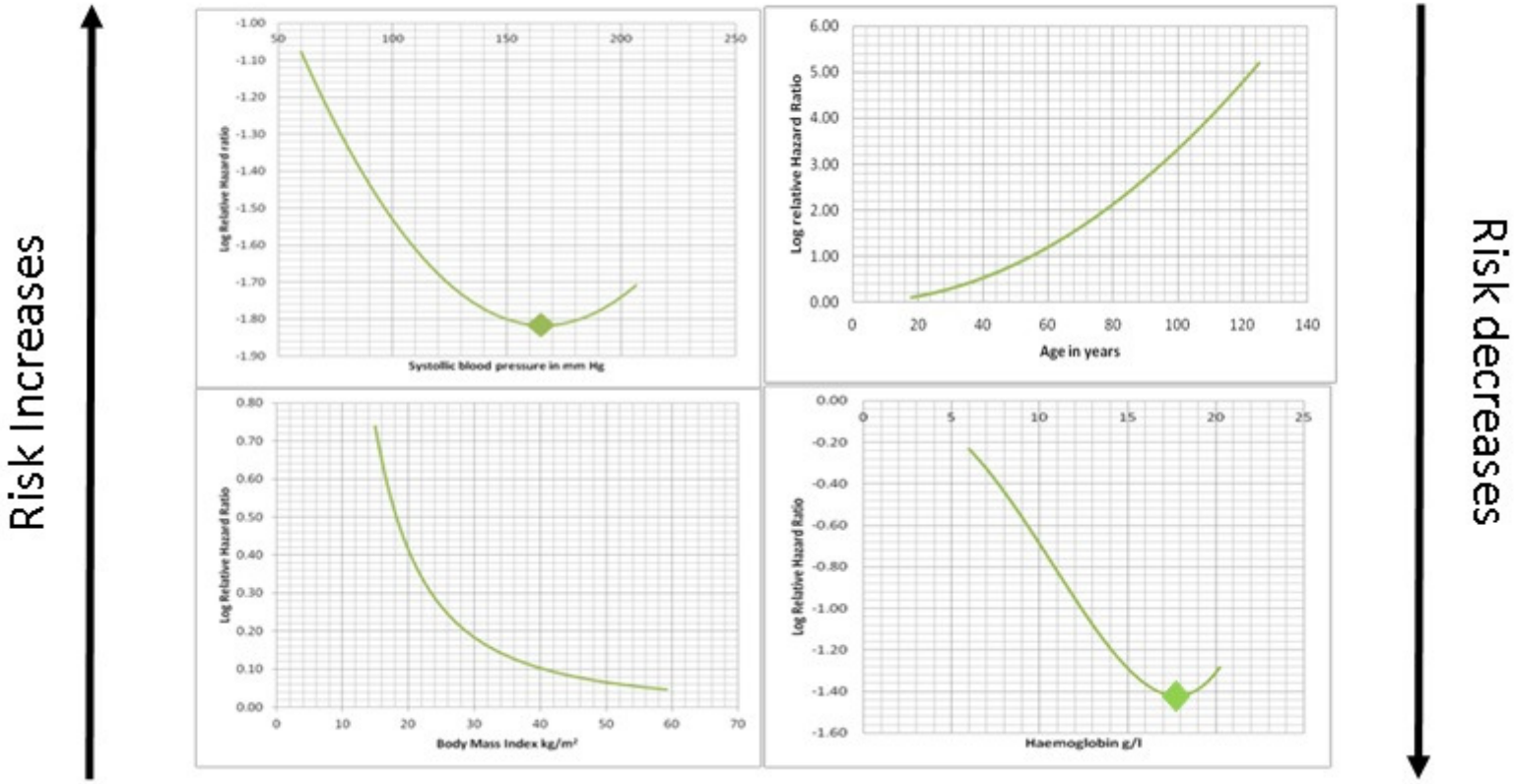
African-Caribbean ethnicity (HR 0.63, 0.47 to 0.86), Indian Sub-continental Ethnicity (HR 0.77, 0.61 to 0.96), angiotensin blockade (HR 0.85, 0.82 to 0.89) and use of other

antihypertensive (HR 0.89, 0.82 to 096) were associated with improved survival. Increasing GFR (for an increase in 100 mls/min/1.73m$^2$, HR 0.24, 0.19 to 0.30), and cholesterol (for an increase in 10mmol/l, HR 0.76, 0.65 to 0.89) were associated with increased survival. Beta blockade, diastolic blood pressure, calcium channel blockers and lipid lowering drugs were all non-significant and were removed from the final models.

#### *5.2.6.3.Fractional polynomials in final model*

The relationship between log hazard ratios and age is shown in Figure 5-7. When age increased, the log relative hazard ratio increased at a squared rate. The log relative hazard decreased as systolic blood pressure started to increase but when blood pressure is above 166 mmHg the risk started to increase (Figure 5-7). As Hb increased the log relative hazard ratio decreased but at a Hb of 17.5 g/dl the log relative hazard ratio started to rise (Figure 5-7). When BMI increased between 15 to 27 kg/m$^2$ the log relative hazard ratio fell sharply and then plateaued (Figure 5-7).

**Figure 5-7. Multiple fractional polynomials for cardiovascular disease/ death model without frailty. Systolic blood pressure, age BMI and Hb versus log relative hazard ratio**



### 5.2.7.Step 7: Frailty Models

Steps 4 to 6 were repeated with a frailty term (a random effects term) for practice location added to each model. These steps were named Step 4F – 6F.

#### *5.2.7.1.Step 4F. Analysis of co-variables where relationship between continuous co-variable and time to outcome is linear and a frailty term .*

When a frailty term (a random effects term) for practice was added to each imputation analysis diastolic blood pressure, NSAIDs, beta blockers and calcium channel blockers were removed from the final model. (Appendix B: Tables 11 to 15). Lipid lowering agents were significant only in the fifth Imputation and removed from combined model (Table 5-9). This is consistent with the findings observed in the model without the frailty term.

**Table 5-9. Combined coefficients from composite outcome model where linear risk of numerical variables is assumed with practice location as frailty term**

| Risk factor /Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Upper Limit |
|---|---|---|---|---|---|
| **Age** | 0.05 | 0.00 | 1.05 | 1.05 | 1.05 |
| **Male Gender** | 0.49 | 0.02 | 1.64 | 1.57 | 1.70 |
| **Race with African Caribbean Race as reference African –Caribbean** | -0.38 | 0.16 | 0.68 | 0.50 | 0.93 |
| **Indian Subcontinent** | -0.22 | 0.12 | 0.80 | 0.63 | 1.02 |
| **South East Asian** | -0.51 | 0.58 | 0.60 | 0.19 | 1.87 |
| **Townsend Quintile compared to 1st Quintile 2nd** | 0.08 | 0.03 | 1.09 | 1.03 | 1.14 |
| **3rd** | 0.11 | 0.03 | 1.12 | 1.06 | 1.18 |
| **4th** | 0.14 | 0.03 | 1.15 | 1.09 | 1.22 |
| **5th** | 0.19 | 0.03 | 1.21 | 1.13 | 1.30 |
| **Diabetes Mellitus** | 0.16 | 0.03 | 1.17 | 1.12 | 1.23 |
| **Heart Failure** | 0.55 | 0.03 | 1.73 | 1.63 | 1.84 |
| **Atrial Fibrillation** | 0.23 | 0.03 | 1.26 | 1.19 | 1.33 |
| **Ever Smoked** | 0.24 | 0.03 | 1.27 | 1.20 | 1.33 |
| **Systolic BP** | 0.00 | 0.00 | 0.997 | 0.996 | 0.998 |
| **BMI** | -0.02 | 0.00 | 0.98 | 0.98 | 0.99 |
| **Haemoglobin** | -0.11 | 0.01 | 0.89 | 0.88 | 0.90 |
| **GFR** | -0.01 | 0.00 | 0.99 | 0.98 | 0.99 |
| **Cholesterol** | -0.03 | 0.01 | 0.97 | 0.96 | 0.99 |
| **High Albuminuria** | 0.16 | 0.03 | 1.18 | 1.10 | 1.26 |
| **Very High albuminuria** | 0.46 | 0.04 | 1.59 | 1.46 | 1.73 |
| **Aspirin** | 0.22 | 0.02 | 1.24 | 1.19 | 1.29 |
| **Anticoagulation** | 0.18 | 0.04 | 1.19 | 1.10 | 1.29 |
| **Angiotensin Blockade** | -0.17 | 0.02 | 0.84 | 0.81 | 0.88 |
| **Diuretic use** | 0.07 | 0.02 | 1.07 | 1.03 | 1.12 |
| **Other** | -0.12 | 0.04 | 0.89 | 0.82 | 0.96 |
| **Iron Medication** | 0.15 | 0.03 | 1.16 | 1.08 | 1.24 |
| **Vitamin D supplementation** | 0.16 | 0.03 | 1.17 | 1.11 | 1.25 |

The combined model with frailty was similar to the model in step 4 above with the exception of Indian Sub-continental ethnicity which was no longer significant. To recap

increasing age, male gender, increasing Townsend quintile, diabetes mellitus, heart failure, atrial fibrillation, patients who had previously smoked, increasing proteinuria, aspirin, anticoagulation, diuretic use, iron and vitamin D supplementation were associated with poorer survival. African-Caribbean ethnicity, increasing systolic blood pressure, body mass index, haemoglobin, GFR and cholesterol were associated with better survival. Angiotensin blockade and other anti-hypertensives were associated with better survival also. The AIC for each model for each imputed dataset was better the previous model without a frailty term for practice location (Table 5-10).

**Table 5-10. AIC for each model for Table 5-9**

| Imp Dataset | Akaike Information Criteria for models analysed in Step 4: No transformation of continuous co-variable without frailty | Akaike Information Criteria for models analysed in Step 4f: No transformation of continuous co-variable with frailty |
|---|---|---|
| **1** | 252133.5 | 251900.1 |
| **2** | 252125.2 | 251895.2 |
| **3** | 252220.3 | 251993.8 |
| **4** | 252204.9 | 251979.6 |
| **5** | 252139.4 | 251924.6 |

#### *5.2.7.2.Results Step 5f. Models with transformation*

When a frailty term was added to the mfp algorithm to determine the best fitting fractional polynomial in Cox proportional hazards model for each imputed dataset, the with exception of Indian sub-continental ethnicity, the same co-variables were significant and same transformation of continuous co variables as in Table 5-5 were suggested (the results of the Cox proportional hazards model are presented in Appendix C Tables 16 to 20). The plots of the transformed co-variables were almost identical to those in Figure 36 to Figure 39 and are shown in Appendix C Figures 6 to 9.

The AIC for each imputation substantially improved in comparison with all previous models (Table 5-11).

**Table 5-11. AIC for each model with original suggested fractional polynomial with frailty term**

| Imp Dataset | Akaike Information Criteria for models analysed in Step 5: Where co variable is transformed as per Table 5-5 | Akaike Information Criteria for models analysed in Step 5f: Where co variable is transformed as per Table 5-5 with frailty |
|---|---|---|
| **1** | 251977 | 251745 |
| **2** | 251989 | 251760 |
| **3** | 252108 | 251884 |
| **4** | 252071 | 251846 |
| **5** | 252010 | 251797 |

***5.2.7.3.Results Step 6f. Finding the most consistent transformations of continuous co-variables across the Imputations***

Like the non frailty models, the mfp transformations suggested for age, systolic blood pressure, Hb and BMI differed across each Imputation and therefore the commonest transformations were compared against the original suggested transformations.

The commonest transformations to recap were (Table 5-5):

1st alternative Model: $Age^2$ + Systolic blood pressure + $systolic\ blood\ pressure^2$ + $haemoglobin^3$+haemoglobin/$10^3$*log haemoglobin + $BMI^{-2}$

2nd alternative Model: $Age^2$ + Systolic blood pressure + $systolic\ blood\ pressure^2$ + $haemoglobin^3$+haemoglobin/$10^3$*log haemoglobin + $BMI^{-0.5}$+$BMI^3$

For the frailty models again the 1st alternative Model had better fit compared to 2nd alternative Model. (Table 5-12)

**Table 5-12. A comparison between the ordinal model with mfp and the 1st and 2nd best model with frailty term for composite outcome**

| Imp Dataset | Akaike Information Criteria for original Model | Akaike Information for 1st 'alternative' Model | Akaike Information for 2st 'alternative Model |
|---|---|---|---|
| **1** | 251745 | 251751 | 251983 |
| **2** | 251760 | 251769 | 252000 |
| **3** | 251884 | 251884 | 252108 |
| **4** | 251846 | 251845 | 252069 |
| **5** | 251797 | 251798 | 252011 |

Similar to the analysis of the non frailty models in step 6, there was a significant difference in the fit in the analysis of imputation 1 ($X^2$= 6.0, 1df p <0.02) and analysis of imputation 2 ($X^2$= 9.0, 1df p <0.01) between the 1st best alternative model. There was no significant difference between models for imputed datasets 3-5. (Table 5-12) As the 1st alternative imputed model seemed the best fit across all imputed datasets using common transformations this was chosen as the final model and coefficients were combined for each imputation.

#### *5.2.7.4.Final model*

In the model with a frailty term for practice, Indian sub-continental ethnicity was no longer significantly protective (HR 0.80, 0.63 to 1.02), but as observed in the model produced in step 3 the remaining co-variables were still significant, the direction of the risk/protection did not change and the coefficients changed very slightly (Table 5-13). (The graphs for the data are not shown)

**Table 5-13. Combined model for composite outcome for each imputation using the 1st best model incorporating a frailty term of practice location**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Upper Limit |
|---|---|---|---|---|---|
| **$(Age/100)^2$** | 3.31 | 0.07 | 27.34 | 23.97 | 31.19 |
| **Male Gender** | 0.50 | 0.02 | 1.65 | 1.58 | 1.71 |
| **Race with Caucasian Race as reference** | | | | | |
| **African-Caribbean** | -0.40 | 0.16 | 0.67 | 0.49 | 0.92 |
| **Indian-Subcontinent** | -0.22 | 0.12 | 0.80 | 0.63 | 1.02 |
| **South East Asian** | -0.54 | 0.58 | 0.58 | 0.19 | 1.82 |
| **Townsend Quintile compared to 1st Quintile** | | | | | |
| **2** | 0.08 | 0.03 | 1.08 | 1.03 | 1.14 |
| **3** | 0.11 | 0.03 | 1.12 | 1.06 | 1.18 |
| **4** | 0.14 | 0.03 | 1.15 | 1.08 | 1.22 |
| **5** | 0.18 | 0.03 | 1.20 | 1.12 | 1.29 |
| **Diabetes Mellitus** | 0.16 | 0.02 | 1.18 | 1.12 | 1.24 |
| **Heart Failure** | 0.53 | 0.03 | 1.70 | 1.60 | 1.80 |
| **Atrial Fibrillation** | 0.23 | 0.03 | 1.25 | 1.18 | 1.32 |
| **$(Systolic\ BP/100)^1$** | -2.19 | 0.34 | 0.11 | 0.06 | 0.22 |
| **$(Systolic\ BP/100)^2$** | 0.64 | 0.12 | 1.90 | 1.52 | 2.39 |
| **$(BMI/10)^{-2}$** | 1.71 | 0.15 | 5.53 | 4.13 | 7.40 |
| **$(Haemoglobin/10)^3$** | -0.69 | 0.05 | 0.50 | 0.46 | 0.55 |
| **$(Haemoglobin/10)^3$ x log (Haemoglobin)** | 0.77 | 0.07 | 2.17 | 1.90 | 2.48 |
| **Glomerular Filtration Rate** | -1.34 | 0.11 | 0.26 | 0.21 | 0.33 |
| **(Cholesterol/10)** | -0.29 | 0.08 | 0.75 | 0.64 | 0.88 |
| **Albuminuria compared to normal** | | | | | |
| **High** | 0.16 | 0.03 | 1.18 | 1.10 | 1.26 |
| **Very High** | 0.45 | 0.04 | 1.57 | 1.44 | 1.71 |
| **Ever Smoked** | 0.24 | 0.03 | 1.27 | 1.20 | 1.33 |
| **Aspirin** | 0.22 | 0.02 | 1.24 | 1.19 | 1.30 |
| **Anticoagulation** | 0.19 | 0.04 | 1.20 | 1.11 | 1.30 |
| **Diuretic use** | 0.07 | 0.02 | 1.08 | 1.04 | 1.12 |
| **Other** | -0.12 | 0.04 | 0.89 | 0.82 | 0.96 |
| **Angiotensin Blockade** | -0.16 | 0.02 | 0.85 | 0.82 | 0.89 |
| **Iron Medication** | 0.12 | 0.03 | 1.12 | 1.05 | 1.20 |
| **Vitamin D supplementation** | 0.15 | 0.03 | 1.16 | 1.10 | 1.23 |

### 5.2.9. Step 8. Is model fit improved after incorporating mfp transformation and frailty terms?

The table below compares the model fit statistic AIC combined models from Step 4, Step 6 and Step 6f. A lower AIC implies a better model fit and a difference of 4 between the value of AIC for each model means there is a significantly better model ($p < 0.05$) and if this difference was 11 then the significance increased to $p < 0.001$. Therefore incorporating fractional polynomials for age, systolic blood pressure, Hb and BMI led to a significantly better model. However incorporating the frailty term results an even better fitting model. (Table 5-14)

**Table 5-14. Comparison of the AIC for the original model without frailty/mfp and models with mfp and frailty. The value in brackets is the difference between the current column and the previous column**

| Imp Dataset | AIC for original Model i.e. no transformation or frailty term (Step 4) | AIC for Model with mfp transformation (1st alternative Model) but no frailty term (Step 6) | AIC for Model with mfp transformation (1st Best Model) with frailty term (Step 6f) |
|---|---|---|---|
| **1** | 252133.5 | 251983(-150.5) | 251751(232) |
| **2** | 252125.2 | 252000(-125.2) | 251769(-231) |
| **3** | 252220.3 | 252108(112.3) | 251884(-224) |
| **4** | 252204.9 | 252069(-135.9) | 251845(-224) |
| **5** | 252139.4 | 252011(-128.4) | 251798(-213) |

### 5.2.10. Step 9: Testing the proportional hazards assumption

The best fitting, most informative model is described in step 6f and the scaled Schoenfeld residuals against time are shown in Appendix D. All the graphs showed a non zero slope hence maintaining the proportional hazards assumption.

### 5.2.11.Step 10. Complete Case Analysis

As specified in the methods a complete case analysis (CCA) was carried out for the dataset. This included 3243 patients who had a complete set of data for analysis. The Median age was lower in the complete case analysis, but the gender proportion and ethnic mix was similar (Table 5-15).

**Table 5-15. Demographics of the whole cohort and complete case cohort for composite outcome analysis**

| Demographic feature | Figure or proportion Whole Cohort (N=74 731) | Figure or proportion Complete cohort (n=3243) |
|---|---|---|
| **Median Age in years at time of diagnosis (years)** | 71.3 (23.-107) | 69.8 (23-95) |
| **Gender – Female** | 50517 (67.6%) | 1957 (67.4%) |
| **Not recorded/Caucasian Race**<br>**African –Caribbean**<br>**Indian Subcontinent**<br>**South East Asian** | <br>73317(98.1%)<br>683 (0.9%)<br>690(0.9%)<br>41(0.1%) | <br>3136(96.7%)<br>44 (1.4%)<br>62(1.9%)<br>1(0.03%) |
| **Townsend quintiles**<br>**1**<br>**2**<br>**3**<br>**4**<br>**5** | <br>18855(26.4%)*<br>17477(24.4%)<br>14762(20.6%)<br>12569(17.6%)<br>7857(11.0%) | <br>790(24.33)<br>1529(22.76)<br>2209(20.94)<br>2824(18.94)<br>423(13.03) |
| **Median eGFR in mmol/min/1.73m2** | 54.3 (3-64) | 55.0 (4-60) |
| **Diabetes Mellitus** | 15039(20.1%) | 1904(58.7%) |
| **Atrial Fibrillation** | 6341(8.5%) | 263(8.1%) |
| **Heart Failure** | 3705(5.0%) | 144(4.4%) |
| **Ever Smoked** | 12999(17.4%) | 729(22.5%) |
| **Aspirin** | 16875(22.6%) | 1168(36.0%) |
| **Anticoagulation** | 3281 (4.4%) | 149 (4.6%) |
| **Angiotensin Blockade** | 30716(41.1%) | 1961(60.5%) |
| **Beta blockers** | 16522(22.1%) | 757(23.3%) |
| **Calcium Channel Blockers** | 13143(17.6%) | 717(22.1%) |
| **Diuretics** | 30826(41.3%) | 1416(43.6%) |
| **Other Anti-hypertensives** | 4421(5.9%) | 301(9.3%) |
| **Lipid lowering medication** | 21455(28.7%) | 1804(55.7%) |
| **Iron Supplementation** | 3882(5.2%) | 234(7.2%) |
| **Vitamin D** | 6153(8.2%) | 214(6.6%) |

The median eGFR was slightly higher in the complete case cohort and more patients had diabetes and smoked but a similar proportion had heart failure and atrial

fibrillation (Table 5-15). In general patients were more likely to be on the prescription medication with the exception of Vitamin D supplementation and NSAID prescription (Table 5-15).

**Table 5-16. Variables with missing data and their mean, median value or proportion for the composite outcome analysis cohort**

| Variable with Missing Data (Proportion of population with complete data) | Figure or proportion Whole Cohort (N=74 731)* | Figure or proportion Complete cohort (n=3243) |
|---|---|---|
| **Median Cholesterol (mmol/l) (60.8%)** | 5.3(1.7-13.9) | 4.8(1.9-10.4) |
| **Median Body Mass Index (kg/m2) (38.2%)** | 28.1(11.2-59.5) | 28.8(14-59.3) |
| **Median Haemoglobin (g/dl) (64.2%)** | 13.5(3.3-25.8) | 13.5(4.7-20.8) |
| **Mean Systolic blood pressure (mmHg) (82.5%)** | 142.5(20.5) | 78.6(11.9) |
| **Mean Diastolic blood pressure (mmHg) (82.5%)** | 79.7(11.3) | 141.5(20.19) |
| **Proteinuria levels (13.6%)** | | |
| **None** | 1876(18.5%)* | 680(20.9) |
| **High** | 7586(74.8%) | 2402(74.0) |
| **Very High** | 675(6.7%) | 165(5.1) |

* Proportion within available data

#### *5.2.11.1.Cox proportional hazards model for complete case analysis*

The median follow-up was 914 days which was less than the previous analysis. As models with frailty and transformation of variables have previously shown better fit only this model is shown (Table 5-17). Like the previous analysis, male gender, heart failure, patients who had previously smoked, having very high albuminuria, diuretic use, anticoagulation and iron medication were all associated with worse survival.

**Table 5-17. Cox model for composite outcome of complete case analysis with frailty term and fractional polynomials**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% CI Lower | 95% CI Upper |
|---|---|---|---|---|---|
| **$(Age/100)^{-1}$** | -6.80 | 1.09 | 0.00 | 0.00 | 0.01 |
| **$(Age/100)^{-1}$ x log(Age/100)** | -3.55 | 0.63 | 0.03 | 0.01 | 0.10 |
| **Male Gender** | 0.51 | 0.10 | 1.66 | 1.36 | 2.02 |
| **Heart Failure** | 0.82 | 0.16 | 2.28 | 1.67 | 3.12 |
| **Ever Smoked** | 0.23 | 0.11 | 1.26 | 1.01 | 1.58 |
| **$(BMI/10)^{-1}$** | 36.01 | 7.95 | 4.35E+15 | 7.40E+08 | 2.56E+22 |
| **$(BMI/10)^{-0.5}$** | -41.66 | 9.75 | 0.00 | 0.00 | 0.00 |
| **Glomerular Filtration Rate ml/min/1.73m$^2$** | -0.01 | 0.01 | 0.99 | 0.97 | 1.00 |
| **Proteinuria compared to normal** | -0.09 | 0.13 | 0.91 | 0.71 | 1.16 |
| **High<br>Very High** | 0.72 | 0.20 | 2.05 | 1.38 | 3.05 |
| **Diuretics** | 0.26 | 0.10 | 1.30 | 1.06 | 1.59 |
| **Anticoagulation** | 0.45 | 0.17 | 1.57 | 1.11 | 2.20 |
| **Iron Medication** | 0.44 | 0.15 | 1.56 | 1.15 | 2.10 |

A higher eGFR and increasing BMI were protective. Initially as age increased so did the log relative hazard ratio but when the age reached 50 then the rate at which log relative hazard ratio increased fell and plateaued (Figure 5-8). As the BMI increased the log relative hazard ratio decreased sharply but then plateaued after the BMI reached over 35 (Figure 5-9).

**Figure 5-8. Age versus log relative hazard ratio in complete case analysis**

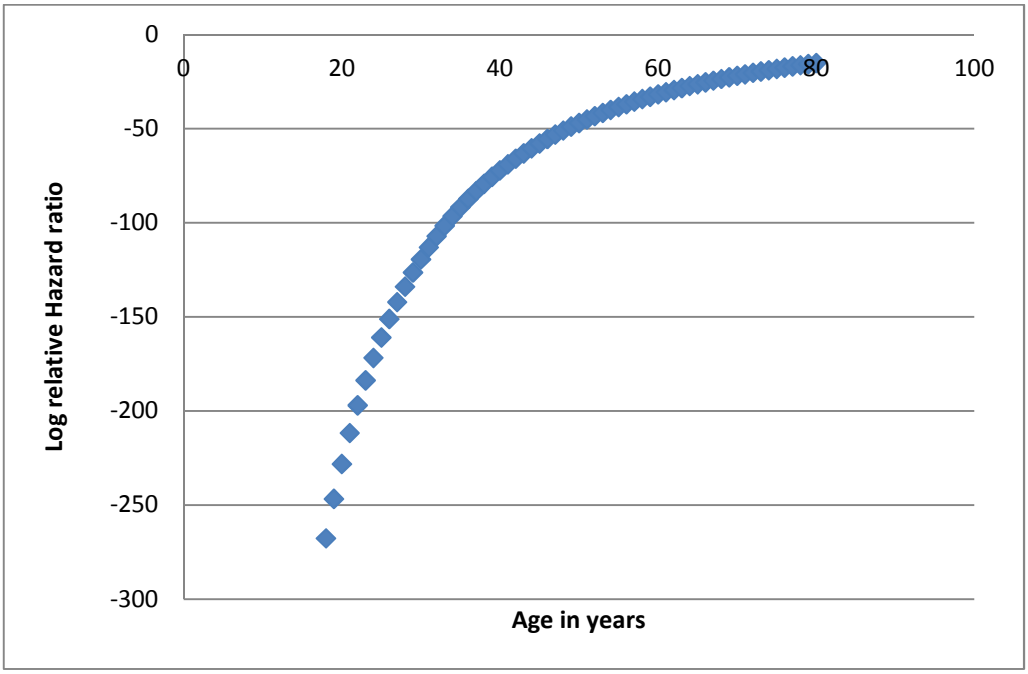
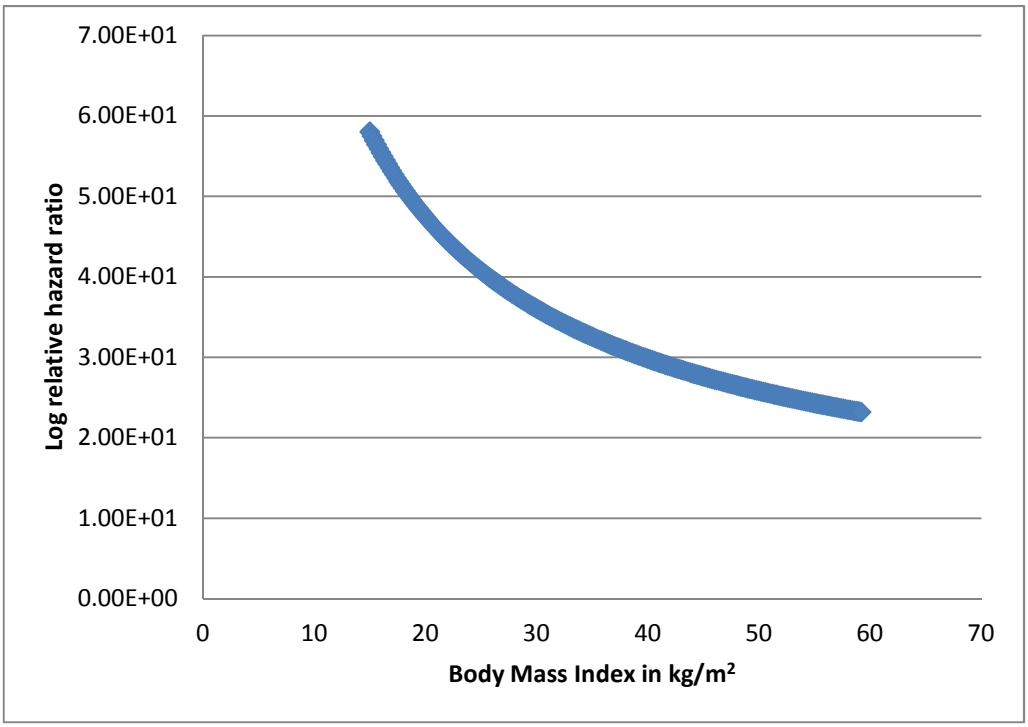



**Figure 5-9. Body mass Index versus log relative hazard ratio in complete case analysis**



## 5.3.Discussion

### 5.3.1.Summary of Results

The models have identified a number of potentially modifiable risk factors associated with CVD or all-cause mortality in CKD patients assessed routinely in primary care. There were 74 731 patients with CKD free from CVD at the time of diagnosis and the median follow up was 1034 days. Like the models in Chapter 4, there was a high proportion of missing data and 5 imputed datasets were created using chained equations. The model with transformed variables and frailty term for practice location had the best fit; i.e. the lowest AIC. In this model the relationship between age, systolic blood pressure, BMI and Hb were non-linear and are shown in Figure 5-7. As age increased the log relative hazard ratio increased at a squared rate. Blood pressure had an inverse J shaped relationship with log relative hazard ratio which was consistent with findings from Chapter 4. Particular to this model Hb had an inverse J shaped relationship with log relative hazard ratio. Initially, as BMI increased the log relative hazard ratio decreased but plateaued after a BMI of 30 $kg/m^2$.

The association of the rest of the risk factors with hazard for the composite outcome model are shown in the forest plot below (Figure 5-11). The predictors with the exception of cholesterol had similar risk profiles to the final model in Chapter 4. The complete case analysis consisted of 3243 patients who had a follow-up of 914 days. Very few variables had significance in this model, (Figure 5-10) but the direction or risk did not change for each of the co-variables.

**Figure 5-10. Iterations for Models predicting cardiovascular disease and or Death**

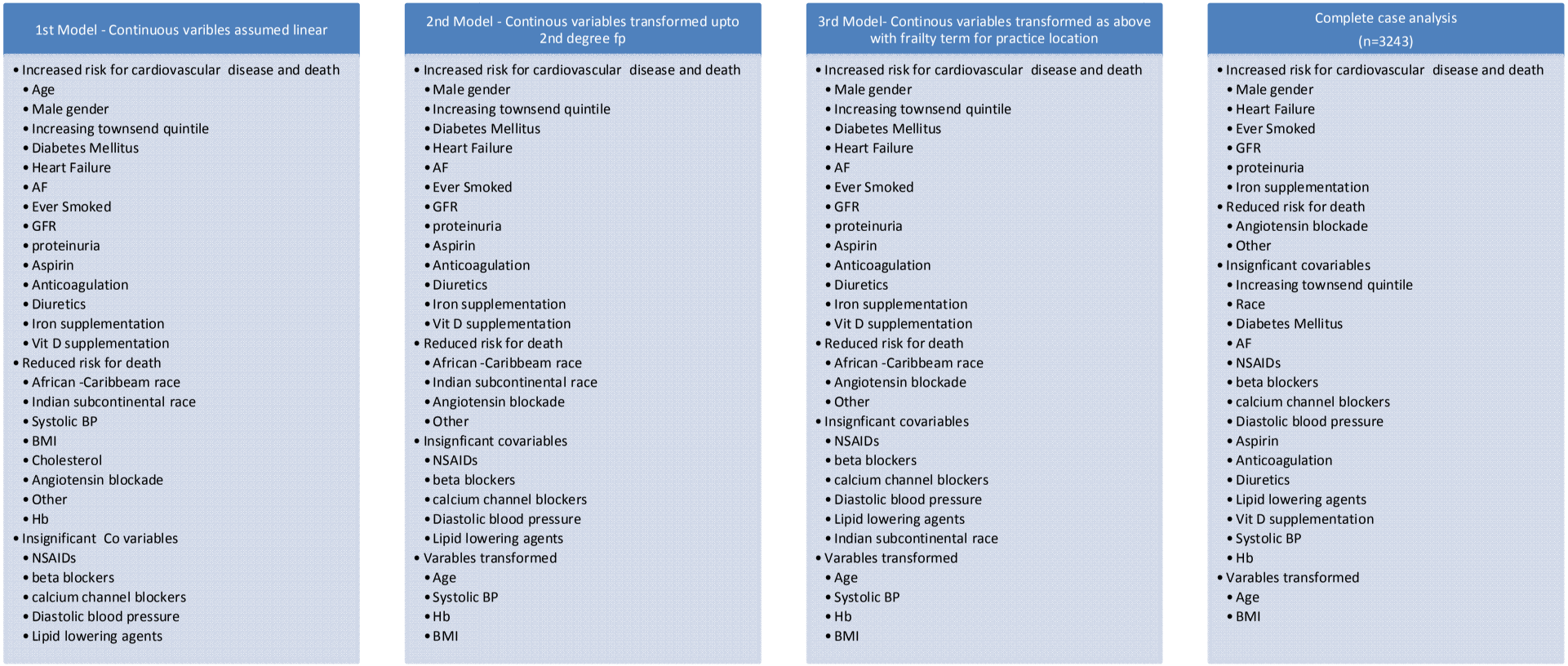
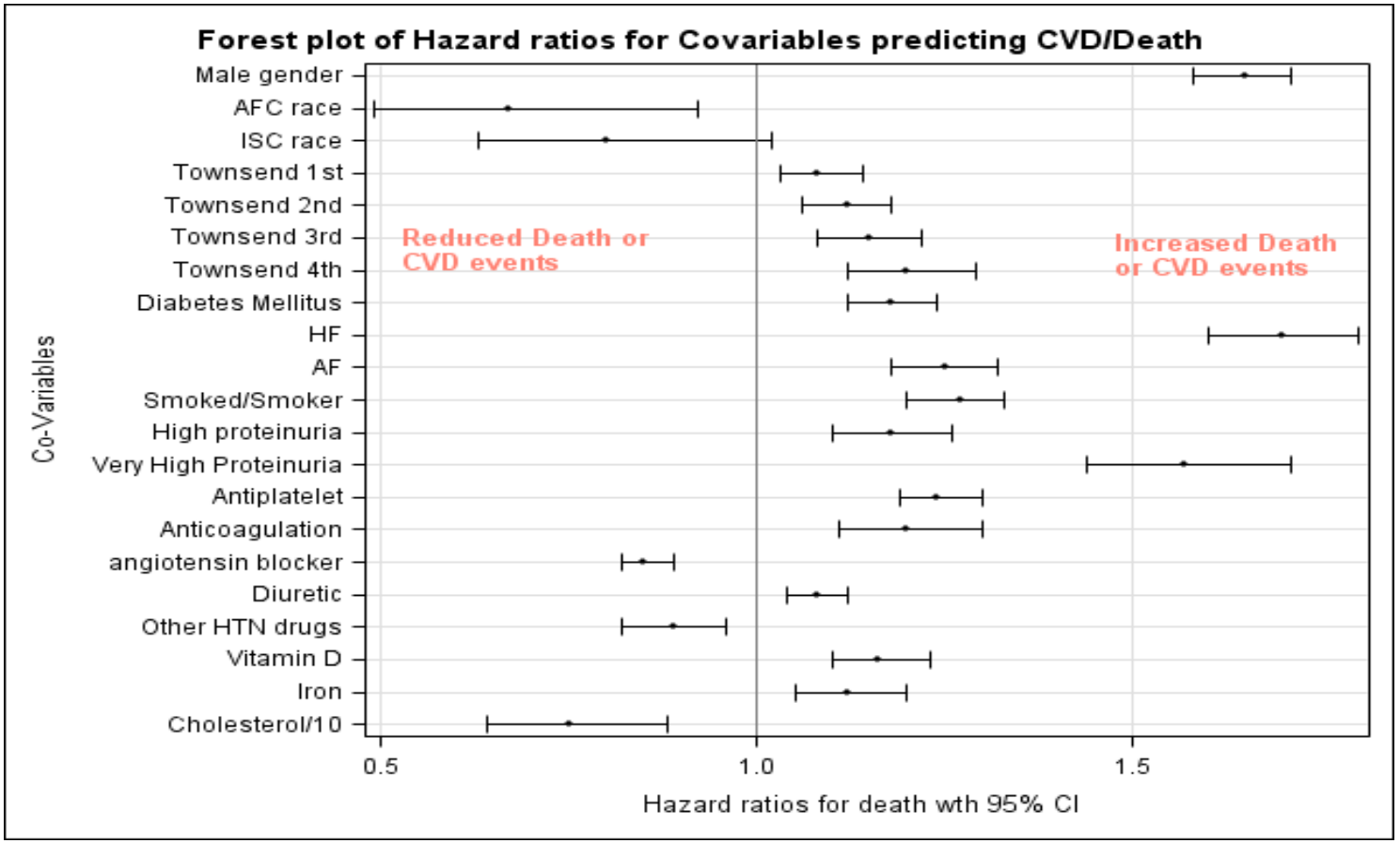


**Figure 5-11. Forest plot of hazard ratios for covariables predicting CVD/death**



In this chapter, time to the composite outcome of cardiovascular disease or death was modelled. The results for this model were consistent with the findings from Chapter 4 (outcome all cause mortality). These were gender, diabetes, AF, heart failure, treatment with angiotensin blockade, diuretics, other anti-hypertensives, anticoagulation, anti-platelet agents, Townsend quintiles, and vitamin D and iron supplementation. Additionally the relationship between systolic blood pressure and risk was the same between the two models. These factors and comparison to literature has been discussed and will not be revisited. Additionally NSAIDs, calcium channel blockers and diastolic blood pressure were also not significant and excluded from this analysis as well as the previous analysis and will not be discussed as they have been discussed before.(Chapter 4: Discussion)

Factors that are no longer significant in this analysis in comparison to the death model in chapter 4 were beta-blockers, lipid lowering agents, and Indian Sub-continental ethnicity. The relationships between haemoglobin, BMI and cholesterol and the risk of outcome were different in this analysis. There were less high risk individuals in this cohort and therefore there were less patients on beta blockade and lipid lowering agents. For example in the death model there were 42% of patients on lipid lowering agents versus 29% in the composite outcome model and 29% versus 22% on a beta-blocker respectively. Exclusion of patients in which these agents could provide secondary prevention of cardiovascular deaths, could lead to them appearing insignificant in this composite outcome analysis. Additionally, as less people are on these agents then the analysis will have less power to detect a difference. In the SHARP study the absolute risk difference though highly significant was only 2%.

Indian sub continental ethnicity may also be insignificant in this analysis for two reasons. Firstly perhaps higher risk patients with this ethnicity were removed as patients were free from cardiovascular disease. Secondly there may have been a loss of power in the second study as the sample size is smaller. Why cholesterol and BMI lost their J shaped relationship with risk of the composite outcome versus death alone is likely due to a loss of power or a more select group of patients in this analysis. Note when the MFP algorithm was run for this analysis, in the analysis of two of the imputed datasets a 'U' shaped relationship was suggested for BMI.

Like the complete case analysis for the death model there were fewer predictors in the complete case analysis of the composite outcome. The direction of the predictors did not change.

### 5.3.2. Limitations and Strengths

The limitations and strengths of this analysis have been discussed in the limitations and strengths of the death model in Chapter 4. The composite outcome is a smaller model with slightly fewer events and therefore may lack the power to detect the differences found in the death model.

## 5.4.Executive Summary

- Multiple imputations had to be implemented in this model consistent with the models in the previous chapter
- The optimal model was one that incorporated fractional polynomials and a frailty term.
- This model was similar to the all cause mortality model with the exception that beta blockers and statins lost their protective association. The relationship of BMI and log relative hazard ratio was no longer U shaped.

# CHAPTER 6. OVERALL DISCUSSION

The analyses presented in this thesis shows several important findings relating to the prevalence of CKD in the general population, the diagnosis and management of CKD in primary care and the use of a very large primary care dataset to identify potential risk factors associated with mortality or CVD and mortality in CKD patients. In this chapter, I summarise the key findings and discuss their implications.

## 6.1.Summary of Key Findings

**Chapter 2.** Main objective: to determine the prevalence of CKD in the UK from a very large primary care database.

Adapting the definition for stages 3-5 CKD from the NKDOQI and QOF guidelines, the prevalence of stages 3-5 CKD in the UK was 4.67% in 2009. The prevalence of stages 1-2 CKD was 0.34% using urine dip (with trace and above on urine dip as to define high albuminuria) and 0.33% using urine ACR. The overall prevalence of stages 1-5 CKD as identified was 5.01% (using urine dip data).

**Chapter 3.** Main objectives: 1) to determine if patients with stages 3-5 CKD were correctly diagnosed by primary care practitioners as required by the Quality Outcomes Framework and 2) to assess the impact of classification on indicators of quality of care.

A substantial proportion of patients with stages 3-5 CKD had not been placed on the practice CKD register; only 72% of patients with stages 3-5 CKD were appropriately coded for CKD. Furthermore, 44% of those on the practice CKD register did not have CKD. This indicates substantial over-classification and misclassification. Patients who were excluded from the practice CKD register were more likely be older and to have more co-morbidity

in a multivariable analysis. The likelihood of achieving QOF CKD indicators decreased in the following order; appropriately coded CKD patients, miscoded CKD patients; uncoded CKD patients. Although there were similar trends in the achievement of diabetic QOF indicators, diabetic patients were more likely to achieve QOF indicators regardless of the presence of CKD.

**Chapter 4.** Objective: to develop an optimal model to identify which routinely collected primary care data were associated with all cause mortality in stage 3 CKD patients. As there were a high proportion of missing data, which had the potential to exclude more than 15% of the cohort from the development of this model, 5 imputed datasets were generated using chained equations.

The best fitting model incorporated fractional polynomials for age, eGFR, systolic blood pressure, serum cholesterol, haemoglobin, body mass index and a frailty term for practice. In this model, male gender, increasing Townsend quintile, heart failure, AF, all CVD, smokers/ex-smokers, very high proteinuria, anti-platelet agents, anticoagulation, diuretics, vitamin D treatment and iron supplementation were associated with decreased survival. Non Caucasian race, angiotensin blockade, beta blockade, other hypertensive medication and anti-lipid drugs were associated with increased survival. NSAIDS, diastolic blood pressure and calcium channel blockers and high proteinuria had no relationship with survival.

Systolic blood pressure and body mass index had an inverse J shaped relationship with mortality where the risk of death decreases until a certain point (Systolic BP 166mmHg and BMI of 36 kg/$m^2$) and then started to rise. There was a similar relationship with log hazard ratio and cholesterol.

**Chapter 5.** Objective: to ascertain the optimal model for predicting the composite outcome of all cause mortality or cardiovascular disease using routinely collected primary care data in patients with stages 3-5 CKD (patients free from cardiovascular disease at the time of CKD diagnosis). Like the mortality model, there was a substantial proportion of missing data and five datasets with imputed data were generated.

The best model incorporated fractional polynomials for age, systolic blood pressure, haemoglobin and BMI. Consistent with the mortality model, there was an increased risk of the composite outcome with: male gender, increasing Townsend quintile, diabetes, heart failure, AF, previous/current smokers, increasing proteinuria (high or very high), anti-platelet agents, anticoagulation, diuretics, iron and vitamin D supplementation. The risk of the composite outcome was reduced with African–Caribbean ethnicity, angiotensin blockade and other anti-hypertensive usage. The relationship between blood pressure and the composite outcome was similar to the previous mortality model.

Some differences were observed when compared to the mortality model: Indian sub-continental race, beta blockers and lipid lowering agents were not significant in this model. This model also differs from the original model in that there was an inverse relationship between the cholesterol level and the risk of the composite outcome; interestingly, there was no inverse J shaped relationship with the outcome. The relationship between haemoglobin and the composite outcome was U shaped. The relationship between BMI and log relative hazard ratio was similar to the mortality model but when the BMI was greater than 36, the log relative hazard ratio for the composite outcome did not increase. NSAIDs, calcium channel blockers and diastolic blood pressure had no significant relationship with the composite outcome.

## 6.2.Clinical implications of the results reported in this thesis

### 6.2.1.Prevalence of CKD

The prevalence of stages 3-5 CKD in this large study population was lower than other UK estimates.[49;63;67] As discussed in the Chapter 2, this was because of two reasons. Firstly the number of patients with two blood tests was considerably less than those with single blood test used to identify CKD in screening or database studies.(Chapter 2.1) Secondly, as shown in the sensitivity analysis, using a single blood result to ascertain prevalence resulted in a higher prevalence of CKD and this was comparable to other studies.(Chapter 2.1) There is inherent variation in GFR; transient decreases in eGFR can occur due to medication, acute kidney injury, protein loading, hydration state or intra-assay variability in the creatinine measurement.[43] A further variation can be associated with different methods of measuring serum creatinine.[31] This is why all the international and national guidelines state that stages 3-5 CKD should be defined using two blood results.[1;43;44] In this study two blood results were used to stage CKD. Although the QICKD investigators did partially incorporate this definition when ascertaining stages 3-5 CKD prevalence, they also included patients where there was only a single blood test available.[49] Therefore this thesis is the only study to date that has ascertained the prevalence of CKD using two blood results to produce a reliable estimate of the burden of stages 3-5 CKD.

The estimates in this study for utilising two eGFRs may be conservative: the CKD stage was based on the blood test with the lower eGFR from the two blood tests and it was assumed that the creatinine analysis method was non IDMS and therefore could result in a higher eGFR if the creatinine analysis method was IDMS aligned.[31] Conversely the

whole population was assumed to be Caucasian and this may lead to overestimation of CKD, though the effects are likely to be negligible.

Despite these limitations he reduced prevalence of CKD found in this thesis could have a major impact on primary and secondary care. These reduced estimates could allow for more accurate planning of treatment and future financial planning of budgets for CKD care.[97]

The prevalence of stages 1-2 CKD was very low in comparison to previous estimates.[1;33;43;44] This was because only a small minority of the population with an eGFR above 60ml/min/1.73m$^2$ had two urine dips (8.34% by 2009) or two urine ACRs (2.16% by 2008). There is considerable variation in urine protein excretion and all the guidelines state that high or very high albuminuria should be confirmed on two separate occasions.[1;33;43;44] Many of the studies only used a single urine protein test and where studies report on a single test, larger proportions of patients had this measured.[62;71;295] The work from this thesis suggests that stages 1-2 CKD is under-recognised in primary care, even though the presence of albuminuria/proteinuria is common in patients with cardiovascular comorbidities including diabetes and associated with increased mortality and morbidity.[93] One important implication from my work is to focus on the need to test for urinary protein in patients: firstly in those with stages 3-5 CKD to stratify for risk and inform management (BP targets) and secondly screen people with an eGFR ≥ 60 ml/min for proteinuria where they are at risk i.e. those with diabetes, hypertension, cardiovascular disease and structural renal disease, as proteinuria is a marker for progression to stage 3 CKD and we can intervene to reduce this progression.

### 6.2.2.Recognition and management of patients with stages 3-5 CKD in primary care

The implication of this part of my research is that a significant proportion of people are being identified as having CKD without confirmation. Just through day to day biological variability and the measurement uncertainty associated with the assay a patient may receive a CKD diagnosis. However, a single eGFR as low as 54 ml/min cannot indicate stage 3 CKD without a repeat test, as this level is within the variability of an eGFR of 60 ml/min or more.[9] With the large numbers of patients with receiving a CKD diagnosis this has important consequences. For the individual patient this can mean an inappropriate disease label and on a population basis, the misuse of resources.

The current guidelines for primary care practitioners lack clarity in defining CKD. The summary NICE guidelines do not mention that two blood results are required to define stages 3-5 CKD and are mentioned in a footnote in the full guidelines.[161] This study may allow more accurate prediction of the costs of CKD, as previous estimates were based on QOF Read coded CKD.[97]

The Quality Outcomes Framework has incentivised stages 3-5 CKD recognition and management in primary care. However, my research shows that CKD recognition is poor as many patients are missed. Of note, patients who are less likely to be appropriately coded for CKD are younger patients or patients with less comorbidities and these patients may not be managed as well. Though it would seem intuitive that this subgroup would seem less likely to benefit from treatment, these patients have a far higher risk of death compared to the age-matched general population and therefore intervention may improve their outcomes as shown in the mortality models.

Patients with uncoded CKD were less likely to have attained CKD QOF indicators then those on the practice CKD register. Although this study, as a retrospective analysis, cannot confirm that under-recognition of CKD results in poorer management of uncoded patients, previous research in diabetes indicators suggests that this may be true.[141]

Although the differences in actual blood pressure between uncoded and those with coded CKD were only slightly higher (a difference in 1-2 mmHg), this resulted in significant differences in QOF achievement. One could speculate that patients at or just above the threshold BP of 140/85 may be recorded as lower. Primary care practices have to review the way CKD is recorded on a regular basis and actually target the high risk patients (young and older people with comorbidity) for inclusion into the QOF CKD register. Practically as primary care is all computerised, it would not be difficult to identify miscoded and uncoded individuals, as they can search for patients with 2 eGFRs under 60 ml/min/1.73m$^2$ and examine those who have the correct coding.This thesis is the first research to examine QOF recognition and achievement in patients with CKD and suggests that primary care practitioners may not utilise QOF appropriately.

Additionally there were no difference in cholesterol and only small differences in blood pressure between the uncoded and coded cohorts to suggest that primary care practitioners are managing hypertension and cholesterol to national guidelines. However, it is difficult to discern whether being labelled as a hypertensive or having hypertension may affect coding in the first place. For example patients with hypertension coding with CKD were more likely to be on the practice CKD register but have a lower blood pressure and therefore these subtle differences in blood pressure may reflect those preferentially managed.

### 6.2.3.Prognostic Models

The all cause mortality prognostic model has been developed from the largest cohort of stage 3 CKD patients to date. The analysis of this cohort and prognostic/risk factors associated with mortality reasserts the association of beta-blockers, angiotensin blockers, other antihypertensives, lipid lowering agents and possibly non Caucasian race with reduced mortality. However, paradoxically, the results suggest that only very high blood pressure and cholesterol is associated with mortality. The results presented in this thesis also suggest that calcium channel blockers, any blood thinning agents and diuretics may be associated with increased harm. Although these effects can partly be explained by confounding by indication, the association or risk with much higher blood pressure are supported by other large epidemiological studies.[283;284] Therefore this thesis suggests that beta blockers, other anti-hypertensives (mainly Doxazosin) and angiotensin blockade may reduce mortality independently of blood pressure. This may support the strong pathophysiological evidence that mortality and CVD in CKD patients are not all due to atherosclerotic disease; as kidney function becomes more advanced arterial stiffness and cardiac fibrosis becomes increasingly important.[112;206] Beta blockers and angiotensin blockers can reduce arterial stiffness independently of blood pressure.[296;297]

CKD patients are more likely to have gastrointestinal blood loss and this may explain why anti-platelet agents may cause more harm.[298] Additionally aspirin was found only to be efficacious in hypertensive patients and this cohort had relatively lower blood pressure than previously described.[112] Increased total cholesterol (7.6mmol/l) is associated with increased atherosclerosis but risk may be attenuated in patients with CKD.[105] It is also worth noting that this was an all-cause mortality model and therefore all deaths could not be attributed to atherosclerotic disease; CKD patients may suffer from increased

arrhythmogenic disease and sudden cardiac deaths, therefore reduction in cholesterol and blood pressure may not have the same effect.[112]

The composite outcome model had similar findings to the above model. This model only included CKD patients free from CVD at the time of CKD diagnosis. As discussed before there is a smaller number of significant risk/prognostic factors identified in this analysis compared to the all cause mortality model. This may be due to reduced power as this model had a smaller population.[250] This model also depended on coding from primary care practitioners and although cardiovascular disease recording is regarded as accurate, as described previously, that CKD coding is less accurate all CVD events may not be captured.[138;299] Nevertheless, the model reinforces previously described associations including the U shaped relationship of blood pressure and risk of morbidity and mortality.

Reassuringly primary care practitioners are using drugs such as angiotensin blockers, beta blockers and statins that appear to be associated with benefit in a high proportion of patients. (Though cause can not be inferred)

The models developed show that traditional risk/prognostic factors for mortality and cardiovascular morbidity in the general population may not be associated with the same risk/benefits in patients with stage 3 CKD suggesting alternative approaches against non-traditional risk factors as well as management with conventional therapies may be required to improve clinical outcomes. These include the reduction of arterial stiffness by agents such as Spironolactone, aggressive treatment of complications such as bone mineral disorders and acidosis and treatment of specific conditions such as adult polycystic kidney disease.[267;300-302]

## 6.3.Strengths and Limitations

During this thesis, I had the privilege to analyse an extremely large database and interrogate for CKD prevalence and CKD outcomes. The large majority of the remit of CKD patient care is with primary care in the UK and the care received is important as it concerns a clinically significant proportion of the population. This thesis provides reliable estimates of CKD prevalence and is the only study that critically examines the care received in primary care especially regarding QOF. The results reported should help primary care practitioners to redirect resources to appropriately diagnose and manage CKD patients. The model development is novel; this is the largest UK dataset and one of the largest internationally to examine mortality and morbidity in stage 3 CKD patients using prospectively collected data. Missing data were managed through multiple imputation and data transformation. This avoided the analysis of a much smaller dataset and the potential loss of association between key risk and prognostic factors. Also if data had not been transformed then spurious relationships between blood pressure, BMI and cholesterol may have been shown.

However the research reported in this thesis has limitations as it is a retrospective study (although it lacks recall bias as the data was prospectively collected). For example CKD prevalence was based on those with blood tests. Additioanlly there is likely confounding by indication.[279] Additionally as reverse causality can confound these results as the CKD disease state can modify the relationship of risk factors and outcomes such hypertension and cholesterol.[210]

## 6.4.Future Research

In preparation of this thesis it was found that there was a lack of prospective, appropriately powered, randomised control studies regarding key interventions in patients with CKD. There was inconclusive evidence regarding the blood pressure target and even adjunctive blood pressure agents.[118;282;303] Other agents that reduce cardiovascular risk such as aspirin may not be as efficacious and may cause harm by excess bleeding.[110]. This may be why risk predictive instruments such as Framingham and QRISK are ineffective in delineating risk to specific CKD stages.[51;125]

This thesis raises some thought provoking questions about how CKD is managed in primary care and the impact of demographic, clinical features and interventions in this group. There are numerous ideas and avenues for research and the following are a few limited examples.

- A prospective screening study to ascertain the prevalence of stages 1-5 of CKD in the UK, using two blood samples and two urine samples over 3 months to ascertain the exact prevalence of CKD. This could involve a follow up study of the Health Survey of England. This could alternatively be ascertained by using the THIN additional information services. Firstly those with a blood test in the last year could have a follow up blood test and urine test to assess what stage of CKD they have. Additionally high risk groups such as diabetics, those with CVD, AF and HTN could be screened using two consecutive blood and urine tests.
- Assessing strategies for enhancing the accuracy of disease registers for CKD and identify if further development of the CKD QOF has the potential to improve care for those surrogates that are associated with better long-term outcomes. Also to assess key interventions recommended by NICE should also be assessed.

- A randomised control study to determine which blood pressure target reduces morbidity and mortality in patients by CKD in primary care. As discussed in chapter 1 and 3, patients with stage 3 CKD are more likely to be managed in primary care and therefore this would be the ideal forum to conduct this research. This could be achieved by standard randomised control studies or novel trial designs such as prospective randomized open, blinded end-point cluster randomised control studies.
- Though this thesis concentrates primarily on stage 3 CKD, there is even less evidence in patients with stage 4 and 5 CKD and would need greater investigation in secondary and primary care.
- A randomised control study of the use of agents such as antiplatelet agents and calcium channel blockers to determine if they are safe and efficacious in patients with CKD. These agents have been shown to have limited efficacy CKD in previous studies and association with worse outcomes in this thesis and therefore need further investigation.
- A prospective cohort study in CKD patients incorporating routinely and non routinely collected primary care data such as bone mineral disorder management and inflammatory markers as examples to determine a complete set of predictors of mortality and morbidity. This could be a combined primary care and secondary care study as these complex non traditional risk factors are more likely to be managed in secondary care.
- The impact of different methods for estimating eGFR such CKD-EPI and other novel markers such as NGAL or Cystatin CKD on risk delineation in patients with kidney damage.

- Translation of basic scientific research in patients with CKD into treatments to reduce mortality and cardiovascular risk specific to patients with CKD. Cardiovascular disease and mortality may be substituted by surrogate markers for these outcomes such as arterial stiffness.

## 6.5. Clinical Implications and overall conclusions

- Stage 3-5 CKD prevalence (4.76%) is much lower than previous estimates. These 'more robust' evaluation of prevalence can allow clinicians to better predict the care these patients require.
- Stage 1-2 CKD prevalence is much less than previous estimates. The proportion of this high risk cohort need to be identified more accurately in the UK by primary care practitioners
- Primary care clinicians need to improve classification of CKD patients in their primary care register as this may impact upon management. Potential solutions and populations that need to be targeted are identified in this thesis
- Prognostic/Risk factor in patients with CKD such as blood pressure, cholesterol, blood thinning agents and diuretics have a different relationship to morbidity and mortality to the general population and this needs further investigation
- However the models in this thesis have limitations such as confounding by indication and reverse causation of CKD and risk factors.
- Despite these limitations clinicians are treating a high proportion of CKD patients with treatments such as statins, angiotensin blockers and beta blockers which appear to have a protective effect.

# BIBLIOGRAPHY


1. Levey AS, Coresh J, Balk E, Kausz AT, Levin A, Steffes MW, et al. National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification. Annals of Internal Medicine 2003 Jul 15;139(2):137-47.
2. Lindeman RD, Tobin JD, Shock NW. Association between blood pressure and the rate of decline in renal function with age. Kidney Int 1984 Dec;26(6):861-8.
3. Bank N. Mechanisms of diabetic hyperfiltration. Kidney Int 1991 Oct;40(4):792-807.
4. Rose GA. Measurement of Glomerular Filtration Rate by Inulin Clearance without Urine Collection. BMJ 1969 Apr 12;2.
5. Israelit AH, Long DL, White MG, Hull AR. Measurement of glomerular filtration rate utilizing a single subcutaneous injection of 125I-iothalamate. Kidney Int 1973 Nov;4(5):346-9.
6. Jelliffe RW. Letter: Creatinine clearance: bedside estimate. Nephron 1973 Oct;79(4):604-5.
7. Feehally J, Floege J, Jonhson RJ. Comprehensive Clinical Nephrology. 3rd Edition ed. Philedelphia: Mosby; 2007.
8. Stevens LA, Coresh J, Feldman HI, Greene T, Lash JP, Nelson RG, et al. Evaluation of the modification of diet in renal disease study equation in a large diverse population. J Am Soc Nephrol 2007;18:2749-57.
9. Myers GL, Miller WG, Coresh J, Fleming J, Greenberg N, Greene T, et al. Recommendations for improving serum creatinine measurement: a report from the Laboratory Working Group of the National Kidney Disease Education Program. Clin Chem 2006 Jan;52(1):5-18.
10. Chitalia VC, Kothari J, Wells EJ, Livesey JH, Robson RA, Searle M, et al. Cost-benefit analysis and prediction of 24-hour proteinuria from the spot urine protein-creatinine ratio. Clin Nephrol 2001 Jun;55(6):436-47.
11. Mitchell SC, Sheldon TA, Shaw AB. Quantification of proteinuria: a re-evaluation of the protein/creatinine ratio for elderly subjects. Age Ageing 1993 Nov;22(6):443-9.
12. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. Nephron 1976;16(1):31-41.
13. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D, et al. A More Accurate Method To Estimate Glomerular Filtration Rate from Serum Creatinine: A New Prediction Equation. Annals of Internal Medicine 1999 Mar 16;130(6):461-70.
14. Poggio ED, Wang X, Greene T, Van Lente F, Hall PM. Performance of the Modification of Diet in Renal Disease and Cockcroft-Gault Equations in the Estimation of GFR in Health and in Chronic Kidney Disease. J Am Soc Nephrol 2005 Feb 1;16(2):459-66.
15. Levey A, Greene T, Kusek JW, Beck GJ. A simplified equation to predict glomerular filtration rate from serum creatinine. J Am Soc Nephrol 11: A0828. 2000.

Ref Type: Abstract

16. Ma YC, Zuo L, Chen JH, Luo Q, Yu XQ, Li Y, et al. Modified glomerular filtration rate estimating equation for Chinese patients with chronic kidney disease. J Am Soc Nephrol 2006 Oct;17(10):2937-44.
17. Lee CS, Cha RH, Lim YH, Kim H, Song KH, Gu N, et al. Ethnic coefficients for glomerular filtration rate estimation by the Modification of Diet in Renal Disease study equations in the Korean population. J Korean Med Sci 2010 Nov;25(11):1616-25.
18. Matsuo S, Imai E, Horio M, Yasuda Y, Tomita K, Nitta K, et al. Revised equations for estimated GFR from serum creatinine in Japan. American Journal of Kidney Diseases 2009 Jun;53(6):982-92.
19. Froissart M, Rossert J, Jacquot C, Paillard M, Houillier P. Predictive performance of the modification of diet in renal disease and Cockcroft-Gault equations for estimating renal function. J Am Soc Nephrol 2005 Mar;16(3):763-73.
20. Delanaye P, Mariat C, Maillard N, Krzesinski JM, Cavalier E. Are the creatinine-based equations accurate to estimate glomerular filtration rate in African American populations? Clin J Am Soc Nephrol 2011 Apr;6(4):906-12.
21. Poggio ED, Rule AD, Tanchanco R, Arrigain S, Butler RS, Srinivas T, et al. Demographic and clinical characteristics associated with glomerular filtration rates in living kidney donors. Kidney Int 2009 May;75(10):1079-87.
22. Rule AD, Larson TS, Bergstralh EJ, Slezak JM, Jacobsen SJ, Cosio FG. Using Serum Creatinine To Estimate Glomerular Filtration Rate: Accuracy in Good Health and in Chronic Kidney Disease. Annals of Internal Medicine 2004 Dec 21;141(12):929-37.
23. Holzmann MJ, Ivert T, Jungner I, Nordqvist T, Walldius G, Ostergren J, et al. Renal function assessed by two different formulas and incidence of myocardial infarction and death in middle-aged men and women. J Intern Med 2010 Apr;267(4):357-69.
24. Sebasky M, Kukla A, Leister E, Guo H, Akkina SK, El-Shahawy Y, et al. Appraisal of GFR-estimating equations following kidney donation. American Journal of Kidney Diseases 2009 Jun;53(6):1050-8.
25. Saleem M, Florkowski CM, George PM. Comparison of the Mayo Clinic Quadratic Equation with the Modification of Diet in Renal Disease equation and radionuclide glomerular filtration rate in a clinical setting. Nephrology (Carlton ) 2008 Dec;13(8):684-8.
26. Levey AS, Stevens LA, Schmid CH, Zhang Y, Castro AF, Feldman HI, et al. A New Equation to Estimate Glomerular Filtration Rate. Annals of Internal Medicine 2009;150(9):604-12.
27. Kilbride HS, Stevens PE, Eaglestone G, Knight S, Carter JL, Delaney MP, et al. Accuracy of the MDRD (Modification of Diet in Renal Disease) Study and CKD-EPI (CKD Epidemiology Collaboration) Equations for Estimation of GFR in the Elderly. American Journal of Kidney Diseases 2013 Jan;61(1):57-66.
28. Stevens LA, Schmid CH, Greene T, Zhang YL, Beck GJ, Froissart M, et al. Comparative performance of the CKD Epidemiology Collaboration (CKD-EPI) and the Modification of Diet in Renal Disease (MDRD) Study equations for estimating GFR levels above 60 mL/min/1.73 m2. American Journal of Kidney Diseases 2010 Sep;56(3):486-95.
29. Earley A, Miskulin D, Lamb EJ, Levey AS, Uhlig K. Estimating equations for glomerular filtration rate in the era of creatinine standardization: a systematic review. Annals of Internal Medicine 2012 Jun 5;156(11):785-270, W.

30. Murthy K, Stevens LA, Stark PC, Levey AS. Variation in the serum creatinine assay calibration: a practical application to glomerular filtration rate estimation. Kidney Int 2005 Oct;68(4):1884-7.
31. Mackenzie F. Notes from Birmingham Quality, UK NEQAS, February 2010. 27-5-2011.

Ref Type: Personal Communication

32. Levey AS, Coresh J, Greene T, Marsh J, Stevens LA, Kusek JW, et al. Expressing the Modification of Diet in Renal Disease Study equation for estimating glomerular filtration rate with standardized serum creatinine values. Clin Chem 2007 Apr;53(4):766-72.
33. Lamb EJ, MacKenzie F, Stevens PE. How should proteinuria be detected and measured? Ann Clin Biochem 2009 May;46(Pt 3):205-17.
34. van d, V, Matsushita K, Coresh J, Astor BC, Woodward M, Levey A, et al. Lower estimated glomerular filtration rate and higher albuminuria are associated with all-cause and cardiovascular mortality. A collaborative meta-analysis of high-risk population cohorts. Kidney Int 2011 Jun;79(12):1341-52.
35. Tonelli M, Muntner P, Lloyd A, Manns BJ, James MT, Klarenbach S, et al. Using Proteinuria and Estimated Glomerular Filtration Rate to Classify Risk in Patients With Chronic Kidney DiseaseA Cohort Study. Annals of Internal Medicine 2011 Jan 4;154(1):12-21.
36. Price CP, Newall RG, Boyd JC. Use of protein:creatinine ratio measurements on random urine samples for prediction of significant proteinuria: a systematic review. Clin Chem 2005 Sep;51(9):1577-86.
37. Methven S, MacGregor MS, Traynor JP, O'Reilly DS, Deighan CJ. Assessing proteinuria in chronic kidney disease: protein-creatinine ratio versus albumin-creatinine ratio. Nephrol Dial Transplant 2010 Sep;25(9):2991-6.
38. Ruggenenti P, Gaspari F, Perna A, Remuzzi G. Cross sectional longitudinal study of spot morning urine protein:creatinine ratio, 24 hour urine protein excretion rate, glomerular filtration rate, and end stage renal failure in chronic renal disease in patients without diabetes. BMJ 1998 Feb 14;316(7130):504-9.
39. Mogensen CE, Vestbo E, Poulsen PL, Christiansen C, Damsgaard EM, Eiskjaer H, et al. Microalbuminuria and potential confounders. A review and some observations on variability of urinary albumin excretion. Diabetes Care 1995 Apr;18(4):572-81.
40. Caring for Australasians with Renal Impairment (CARI). 2007.

Ref Type: Online Source

41. Clarke W, Frost SJ, Krause E, Ferris M, Jaar B, Wu J, et al. National Academy of Clinical Biochemistry. Renal function testing. Kidney Int Suppl . 2007.

Ref Type: Online Source

42. White SL, Yu R, Craig JC, Polkinghorne KR, Atkins RC, Chadban SJ. Diagnostic accuracy of urine dipsticks for detection of albuminuria in the general community. American Journal of Kidney Diseases 2011 Jul;58(1):19-28.
43. 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 Int 2011 Jul;80(1):17-28.
44. Crowe E, Halpin D, Stevens P. Early identification and management of chronic kidney disease: summary of NICE guidance. BMJ 2008 Sep 29;337.
45. NHS Primary Care Commissioning. QOF Implementation: Business Rules. 2009.

Ref Type: Online Source
46. Scottish Intercollegiate Network. Diagnosis and management of chronic kidney disease: A national clinical guideline. BMJ . 2008. NHS Scotland. 1-10-2012.
Ref Type: Online Source
47. Chapter 1: Definition and classification of CKD. Kidney Int Suppl 2013 Jan 1;3(1):19-62.
48. Coresh J, Eknoyan G, Levey AS, Clase CM, Garg AX, Kiberd BA. Estimating the Prevalence of Low Glomerular Filtration Rate Requires Attention to the Creatinine Assay Calibration. J Am Soc Nephrol 2002 Nov 1;13(11):2811-6.
49. de Lusignan S., Tomson C, Harris K, Van VJ, 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-c224.
50. Kramer H, Toto R, Peshock R, Cooper R, Victor R. Association between chronic kidney disease and coronary artery calcification: the Dallas Heart Study. J Am Soc Nephrol 2005;16:507-13.
51. Fox CS, Larson MG, Vasan RS, Guo CY, Parise H, Levy D, et al. Cross-sectional association of kidney function with valvular and annular calcification: the Framingham heart study. J Am Soc Nephrol 2006;17:521-7.
52. Coresh J, Selvin E, Stevens LA, Manzi J, Kusek JW, Eggers P, et al. Prevalence of chronic kidney disease in the United States. JAMA 2007 Nov 7;298(17):2038-47.
53. Shankar A, Klein R, Klein BE. The association among smoking, heavy drinking, and chronic kidney disease. Am J Epidemiol 2006;164:263-71.
54. McClellan W, Warnock DG, McClure L, Campbell RC, Newsome BB, Howard V, et al. Racial differences in the prevalence of chronic kidney disease among participants in the Reasons for Geographic and Racial Differences in Stroke (REGARDS) Cohort Study. J Am Soc Nephrol 2006 Jun;17(6):1710-5.
55. Manjunath G, Tighiouart H, Coresh J, Macleod B, Salem DN, Griffith JL, et al. Level of kidney function as a risk factor for cardiovascular outcomes in the elderly. Kidney Int 2003;63:1121-9.
56. Garg AX, Papaioannou A, Ferko N, Campbell G, Clarke JA, Ray JG. Estimating the prevalence of renal insufficiency in seniors requiring long-term care. Kidney Int 2004;65:649-53.
57. Hemmelgarn BR, Zhang J, Manns BJ, Tonelli M, Larsen E, Ghali WA, et al. Progression of kidney dysfunction in the community-dwelling elderly. Kidney Int 2006;69:2155-61.
58. de Zeeuw D, Hillege HL, de Jong PE. The kidney, a cardiovascular risk marker, and a new target for therapy. Kidney Int Suppl 2005;98:S25-S29.
59. Viktorsdottir O, Palsson R, Andresdottir MB, Aspelund T, Gudnason V, Indridason OS. Prevalence of chronic kidney disease based on estimated glomerular filtration rate and proteinuria in Icelandic adults. Nephrol Dial Transplant 2005;20:1799-807.
60. Cirillo M, Laurenzi M, Mancini M, Zanchetti A, Lombardi C, De Santo NG. Low glomerular filtration in the population: prevalence, associated disorders, and awareness. Kidney Int 2006 Aug;70(4):800-6.
61. Nitsch D, Felber DD, von EA, Gaspoz JM, Downs SH, Leuenberger P, et al. Prevalence of renal impairment and its association with cardiovascular risk factors

in a general population: results of the Swiss SAPALDIA study. Nephrol Dial Transplant 2006;21:935-44.
62. Hallan SI, Coresh J, Astor BC, Asberg A, Powe NR, Romundstad S, et al. International comparison of the relationship of chronic kidney disease prevalence and ESRD risk. J Am Soc Nephrol 2006;17(8):2275-84.
63. Aresu M, Chaudhury M, Diment E, Fuller E, Gordon-Dseagu V, Gunning N, et al. Health Survey of England 2009. The NHS Information Centre; 2010 Dec 16.
64. Vinhas J, Gardete-Correia L, Boavida JM, Raposo JF, Mesquita A, Fona MC, et al. Prevalence of chronic kidney disease and associated risk factors, and risk of end-stage renal disease: data from the PREVADIAB study. Nephron 2011;119(1):c35-c40.
65. Suleymanlar G, Utas C, Arinsoy T, Ates K, Altun B, Altiparmak MR, et al. A population-based survey of Chronic Renal Disease In Turkey--the CREDIT study. Nephrol Dial Transplant 2011 Jun;26(6):1862-71.
66. Carter JL, Stevens PE, Irving JE, Lamb EJ. Estimating glomerular filtration rate: comparison of the CKD-EPI and MDRD equations in a large UK cohort with particular emphasis on the effect of age. QJM 2011 Oct;104(10):839-47.
67. Gifford FJ, Methven S, Boag DE, Spalding EM, MacGregor MS. Chronic kidney disease prevalence and secular trends in a UK population: the impact of MDRD and CKD-EPI formulae. QJM 2011 Dec;104(12):1045-53.
68. Wasen E, Isoaho R, Mattila K, Vahlberg T, Kivela SL, Irjala K. Estimation of glomerular filtration rate in the elderly: a comparison of creatinine-based formulae with serum cystatin C. J Intern Med 2004;256:70-8.
69. Glynn LG, Anderson J, Reddan D, Murphy AW. Chronic kidney disease in general practice: prevalence, diagnosis, and standards of care. Ir Med J 2009 Oct;102(9):285-8.
70. Gambaro G, Yabarek T, Graziani MS, Gemelli A, Abaterusso C, Frigo AC, et al. Prevalence of CKD in northeastern Italy: results of the INCIPE study and comparison with NHANES. Clin J Am Soc Nephrol 2010 Nov;5(11):1946-53.
71. Chadban SJ, Briganti EM, Kerr PG, Dunstan DW, Welborn TA, Zimmet PZ, et al. Prevalence of kidney damage in Australian adults: The AusDiab kidney study. J Am Soc Nephrol 2003;14:S131-S138.
72. Domrongkitchaiporn S, Sritara P, Kitiyakara C, Stitchantrakul W, Krittaphol V, Lolekha P, et al. Risk factors for development of decreased kidney function in a southeast Asian population: a 12-year cohort study. J Am Soc Nephrol 2005;16:791-9.
73. Chen J, Wildman RP, Gu D, Kusek JW, Spruill M, Reynolds K, et al. Prevalence of decreased kidney function in Chinese adults aged 35 to 74 years. Kidney Int 2005;68:2837-45.
74. Wen CP, Cheng TY, Tsai MK, Chang YC, Chan HT, Tsai SP, et al. All-cause mortality attributable to chronic kidney disease: a prospective cohort study based on 462 293 adults in Taiwan. Lancet 2008 Jun 28;371(9631):2173-82.
75. Hsu CC, Hwang SJ, Wen CP, Chang HY, Chen T, Shiu RS, et al. High prevalence and low awareness of CKD in Taiwan: a study on the relationship between serum creatinine and awareness from a nationally representative survey. American Journal of Kidney Diseases 2006 Nov;48(5):727-38.

76. Kuo HW, Tsai SS, Tiao MM, Yang CY. Epidemiological Features of CKD in Taiwan. American Journal of Kidney Diseases 2007 Jan;49(1):46-55.
77. Zhang QL, Rothenbacher D. Prevalence of chronic kidney disease in population-based studies: systematic review. BMC Public Health 2008;8:117.
78. Chen N, Wang W, Huang Y, Shen P, Pei D, Yu H, et al. Community-based study on CKD subjects and the associated risk factors. Nephrol Dial Transplant 2009 Jul;24(7):2117-23.
79. Ong-Ajyooth L, Vareesangthip K, Khonputsa P, Aekplakorn W. Prevalence of chronic kidney disease in Thai adults: a national health survey. BMC Nephrology 2009;10:35.
80. Varma PP, Raman DK, Ramakrishnan TS, Singh P, Varma A. Prevalence of early stages of chronic kidney disease in apparently healthy central government employees in India. Nephrol Dial Transplant 2010 Sep;25(9):3011-7.
81. Ingsathit A, Thakkinstian A, Chaiprasert A, Sangthawan P, Gojaseni P, Kiattisunthorn K, et al. Prevalence and risk factors of chronic kidney disease in the Thai adult population: Thai SEEK study. Nephrol Dial Transplant 2010 May;25(5):1567-75.
82. Sabanayagam C, Lim SC, Wong TY, Lee J, Shankar A, Tai ES. Ethnic disparities in prevalence and impact of risk factors of chronic kidney disease. Nephrol Dial Transplant 2010 Aug;25(8):2564-70.
83. Jiang L, Liang Y, Qiu B, Wang F, Duan X, Yang X, et al. Prevalence of chronic kidney disease in a rural Chinese adult population: the Handan Eye Study. Nephron 2010;114(4):c295-c302.
84. Jang SY, Kim IH, Ju EY, Ahn SJ, Kim DK, Lee SW. Chronic kidney disease and metabolic syndrome in a general Korean population: the Third Korea National Health and Nutrition Examination Survey (KNHANES III) Study. J Public Health (Oxf) 2010 Dec;32(4):538-46.
85. Lee SW, Kim YC, Oh SW, Koo HS, Na KY, Chae DW, et al. Trends in the prevalence of chronic kidney disease, other chronic diseases and health-related behaviors in an adult Korean population: data from the Korean National Health and Nutrition Examination Survey (KNHANES). Nephrol Dial Transplant 2011 Dec;26(12):3975-80.
86. Sharma SK, Zou H, Togtokh A, Ene-Iordache B, Carminati S, Remuzzi A, et al. Burden of CKD, proteinuria, and cardiovascular risk among Chinese, Mongolian, and Nepalese participants in the International Society of Nephrology screening programs. American Journal of Kidney Diseases 2010 Nov;56(5):915-27.
87. White SL, Polkinghorne KR, Atkins RC, Chadban SJ. Comparison of the prevalence and mortality risk of CKD in Australia using the CKD Epidemiology Collaboration (CKD-EPI) and Modification of Diet in Renal Disease (MDRD) Study GFR estimating equations: the AusDiab (Australian Diabetes, Obesity and Lifestyle) Study. American Journal of Kidney Diseases 2010 Apr;55(4):660-70.
88. Cepoi V, Onofriescu M, Segall L, Covic A. The prevalence of chronic kidney disease in the general population in Romania: a study on 60,000 persons. International Urology & Nephrology 2012 Feb;44(1):213-20.
89. Zhang L, Wang F, Wang L, Wang W, Liu B, Liu J, et al. Prevalence of chronic kidney disease in China: a cross-sectional survey. Lancet 2012 Mar 3;379(9818):815-22.

90. Li ZY, Xu GB, Xia TA, Wang HY. Prevalence of chronic kidney disease in a middle and old-aged population of Beijing. Clin Chim Acta 2006;366:209-15.
91. Ninomiya T, Kiyohara Y, Kubo M, Tanizaki Y, Doi Y, Okubo K, et al. Chronic kidney disease and cardiovascular disease in a general Japanese population: the Hisayama Study. Kidney Int 2005;68:228-36.
92. Keith DS, Nichols GA, Gullion CM, Brown JB, Smith DH. Longitudinal follow-up and outcomes among a population with chronic kidney disease in a large managed care organization. Archives of Internal Medicine 2004;164:659-63.
93. Matsushita K, van d, V, Astor BC, Woodward M, Levey AS, de Jong PE, et al. Association of estimated glomerular filtration rate and albuminuria with all-cause and cardiovascular mortality in general population cohorts: a collaborative meta-analysis. Lancet 2010 Jun 12;375(9731):2073-81.
94. Astor BC, Matsushita K, Gansevoort RT, van d, V, Woodward M, Levey AS, et al. Lower estimated glomerular filtration rate and higher albuminuria are associated with mortality and end-stage renal disease. A collaborative meta-analysis of kidney disease population cohorts. Kidney Int 2011 Jun;79(12):1331-40.
95. Gansevoort RT, Matsushita K, van d, V, Astor BC, Woodward M, Levey AS, et al. Lower estimated GFR and higher albuminuria are associated with adverse kidney outcomes. A collaborative meta-analysis of general and high-risk population cohorts. Kidney Int 2011 Jul;80(1):93-104.
96. Perico N, Remuzzi G. Chronic kidney disease: a research and public health priority. Nephrol Dial Transplant 2012 Oct;27 Suppl 3:iii19-iii26.
97. Kerr M, Bray B, Medcalf J, O'Donoghue DJ, Matthews B. Estimating the financial cost of chronic kidney disease to the NHS in England. Nephrol Dial Transplant 2012 Oct;27 Suppl 3:iii73-iii80.
98. Australian Institute of Health and Welfare. Dialysis and kidney transplanation in Australia: 1991-2010. Canberra: Canberra:AIHW; 2011.
99. Foley RN, Collins AJ. End-stage renal disease in the United States: an update from the United States Renal Data System. J Am Soc Nephrol 2007 Oct;18(10):2644-8.
100. Byrne C, Steenkamp R, Castledine C, Ansell D, Feehally J. Chapter 4: UK ESRD Prevalent Rates in 2008: national and centre-specific analyses. Nephron Clinical Practice 2010;115(Suppl. 1):c41-c68.
101. Collins AJ, Chen SC, Gilbertson DT, Foley RN. CKD surveillance using administrative data: impact on the health care system. American Journal of Kidney Diseases 2009 Mar;53(3 Suppl 3):S27-S36.
102. Watanabe H, Watanabe T, Sasaki S, Nagai K, Roden DM, Aizawa Y. Close bidirectional relationship between chronic kidney disease and atrial fibrillation: the Niigata preventive medicine study. Am Heart J 2009;158(4):629-36.
103. Yacoub R, Habib H, Lahdo A, Al AR, Varjabedian L, Atalla G, et al. Association between smoking and chronic kidney disease: a case control study. BMC Public Health 2010;10:731.
104. Bakris GL, Williams M, Dworkin L, Elliott WJ, Epstein M, Toto R, et al. Preserving renal function in adults with hypertension and diabetes: a consensus approach. National Kidney Foundation Hypertension and Diabetes Executive Committees Working Group. American Journal of Kidney Diseases 2000 Sep;36(3):646-61.
105. Walker DB, Walker TJ, Jacobson TA. Chronic kidney disease and statins: improving cardiovascular outcomes. Curr Atheroscler Rep 2009;11(4):301-8.

106. Orchard TJ, Dorman JS, Maser RE, Becker DJ, Drash AL, Ellis D, et al. Prevalence of complications in IDDM by sex and duration. Pittsburgh Epidemiology of Diabetes Complications Study II. Diabetes 1990 Sep;39(9):1116-24.
107. Nathan DM, Zinman B, Cleary PA, Backlund JY, Genuth S, Miller R, et al. Modern-day clinical course of type 1 diabetes mellitus after 30 years' duration: the diabetes control and complications trial/epidemiology of diabetes interventions and complications and Pittsburgh epidemiology of diabetes complications experience (1983-2005). Archives of Internal Medicine 2009 Jul 27;169(14):1307-16.
108. Kwan BC, Murtaugh MA, Beddhu S. Associations of body size with metabolic syndrome and mortality in moderate chronic kidney disease. Clin J Am Soc Nephrol 2007 Sep;2(5):992-8.
109. Altman DG. Practical Statistics for Medical Research. London: Chapman and Hall/CRC; 2012.
110. Jardine MJ, Ninomiya T, Perkovic V, Cass A, Turnbull F, Gallagher MP, et al. Aspirin is beneficial in hypertensive patients with chronic kidney disease: a post-hoc subgroup analysis of a randomized controlled trial. J Am Coll Cardiol 2010 Sep 14;56(12):956-65.
111. Redon J, Mancia G, Sleight P, Schumacher H, Gao P, Pogue J, et al. Safety and efficacy of low blood pressures among patients with diabetes: subgroup analyses from the ONTARGET (ONgoing Telmisartan Alone and in combination with Ramipril Global Endpoint Trial). J Am Coll Cardiol 2012 Jan 3;59(1):74-83.
112. Muntner P, He J, Astor BC, Folsom AR, Coresh J. Traditional and Nontraditional Risk Factors Predict Coronary Heart Disease in Chronic Kidney Disease: Results from the Atherosclerosis Risk in Communities Study. J Am Soc Nephrol 2005 Feb 1;16(2):529-38.
113. Shlipak MG, Fried LF, Cushman M, Manolio TA, Peterson D, Stehman-Breen C, et al. Cardiovascular mortality risk in chronic kidney disease: comparison of traditional and novel risk factors. JAMA 2005 Apr 13;293(14):1737-45.
114. SHEP Cooperative Research Group. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. Final results of the Systolic Hypertension in the Elderly Program (SHEP). JAMA 1991 Jun 26;265(24):3255-64.
115. The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: The antihypertensive and lipid-lowering treatment to prevent heart attack trial (ALLHAT). JAMA 2002 Dec 18;288(23):2981-97.
116. Berl T, Hunsicker LG, Lewis JB, Pfeffer MA, Porush JG, Rouleau JL, et al. Impact of achieved blood pressure on cardiovascular outcomes in the Irbesartan Diabetic Nephropathy Trial. J Am Soc Nephrol 2005 Jul;16(7):2170-9.
117. Kunz R, Friedrich C, Wolbers M, Mann JF. Meta-analysis: effect of monotherapy and combination therapy with inhibitors of the renin angiotensin system on proteinuria in renal disease. Annals of Internal Medicine 2008 Jan 1;148(1):30-48.
118. Ruggenenti P, Perna A, Loriga G, Ganeva M, Ene-Iordache B, Turturro M, et al. Blood-pressure control for renoprotection in patients with non-diabetic chronic

renal disease (REIN-2): multicentre, randomised controlled trial. Lancet 2005;365(9463):939-46.
119. Hansson L, Zanchetti A, Carruthers SG, Dahlof B, Elmfeldt D, Julius S, et al. Effects of intensive blood-pressure lowering and low-dose aspirin in patients with hypertension: principal results of the Hypertension Optimal Treatment (HOT) randomised trial. HOT Study Group. Lancet 1998 Jun 13;351(9118):1755-62.
120. McCullough PA, Sandberg KR, Borzak S, Hudson MP, Garg M, Manley HJ. Benefits of aspirin and beta-blockade after myocardial infarction in patients with chronic kidney disease. American Heart Journal 2002 Aug;144(2):226-32.
121. Baigent C, Landray MJ, Reith C, Emberson J, Wheeler DC, Tomson C, et al. The effects of lowering LDL cholesterol with simvastatin plus ezetimibe in patients with chronic kidney disease (Study of Heart and Renal Protection): a randomised placebo-controlled trial. Lancet 2011 Jun 25;377(9784):2181-92.
122. Stamler J, Wentworth D, Neaton JD. Is relationship between serum cholesterol and risk of premature death from coronary heart disease continuous and graded? Findings in 356,222 primary screenees of the Multiple Risk Factor Intervention Trial (MRFIT). JAMA 1986 Nov 28;256(20):2823-8.
123. Weiner DE, Sarnak MJ. Managing dyslipidemia in chronic kidney disease. J Gen Intern Med 2004 Oct;19(10):1045-52.
124. Liu Y, Coresh J, Eustace JA, Longenecker JC, Jaar B, Fink NE, et al. Association between cholesterol level and mortality in dialysis patients: role of inflammation and malnutrition. JAMA 2004 Jan 28;291(4):451-9.
125. 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 Jun 28;336(7659):1475-82.
126. Soliman EZ, Prineas RJ, Go AS, Xie D, Lash JP, Rahman M, et al. Chronic kidney disease and prevalent atrial fibrillation: the Chronic Renal Insufficiency Cohort (CRIC). Am Heart J 2010 Jun;159(6):1102-7.
127. Leavey SF, McCullough K, Hecking E, Goodkin D, Port FK, Young EW. Body mass index and mortality in healthier as compared with sicker haemodialysis patients: results from the Dialysis Outcomes and Practice Patterns Study (DOPPS). Nephrol Dial Transplant 2001 Dec 1;16(12):2386-94.
128. Elsayed EF, Sarnak MJ, Tighiouart H, Griffith JL, Kurth T, Salem DN, et al. Waist-to-hip ratio, body mass index, and subsequent kidney disease and death. American Journal of Kidney Diseases 2008 Jul;52(1):29-38.
129. Jain P, Cockwell P, Little J, Ferring M, Nicholas J, Richards N, et al. Survival and transplantation in end-stage renal disease: a prospective study of a multiethnic population. Nephrol Dial Transplant 2009 Dec 1;24(12):3840-6.
130. Gregory I. Comparisons between geographies of mortality and deprivation from the 1900s and 2001: spatial analysis of census and mortality statistics. BMJ 2009 Sep 10;339.
131. Phrommintikul A, Haas SJ, Elsik M, Krum H. Mortality and target haemoglobin concentrations in anaemic patients with chronic kidney disease treated with erythropoietin: a meta-analysis. Lancet 2007 Feb 3;369(9559):381-8.
132. Kengne AP, Czernichow S+, Hamer M, Batty GD, Stamatakis E. Anaemia, Haemoglobin Level and Cause-Specific Mortality in People with and without Diabetes. PLoS ONE 2012 Aug 2;7(8):e41875.

133. Stevens PE, O'Donoghue DJ, de Lusignan S., Van VJ, Klebe B, Middleton R, et al. Chronic kidney disease management in the United Kingdom: NEOERICA project results. Kidney Int 2007 Jul;72(1):92-9.
134. National Health Service. Connecting for Health. 2013.
Ref Type: Online Source
135. Office for National Statistics. All releases of Population Estimates for UK, England and Wales, Scotland and Northern Ireland. 2012.
Ref Type: Online Source
136. Epidemiology and Pharmacology Information Core. THIN Data. 2012.
Ref Type: Online Source
137. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol 2010 Jan;69(1):4-14.
138. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract 2010 Mar;60(572):e128-e136.
139. Hippisley-Cox J, Coupland C. Predicting the risk of Chronic Kidney Disease in Men and Women in England and Wales: prospective derivation and external validation of the QKidney(R) Scores. BMC Family Practice 2010;11(1):49.
140. NHS Information Centre. Quality and Outcomes Framework Achievement Data 2009/10. 2010.
Ref Type: Pamphlet
141. Calvert M, Shankar A, McManus RJ, Lester H, Freemantle N. Effect of the quality and outcomes framework on diabetes care in the United Kingdom: retrospective cohort study. BMJ 2009 Jan 1;338.
142. Jolly S, Navaneethan SD, Schold JD, Arrigain S, Sharp JW, Jain AK, et al. **Does Recognition of Chronic Kidney Disease in the Electronic Health Record Problem List Improve Care?** 2011 p. 192.
143. Medicines & Healthcare products Regulatory Agency. Clinical Practice Research Datalink. 2012.
Ref Type: Online Source
144. Szatkowski L, Lewis S, McNeill A, Huang Y, Coleman T. Can data from primary care medical records be used to monitor national smoking prevalence? Journal of Epidemiology and Community Health 2011 May 13.
145. Bourke A, Dattani H, Robinson M. Feasibility study and methodology to create a quality-evaluated database of primary care data. Inform Prim Care 2004;12(3):171-7.
146. Lewis JD, Schinnar R, Bilker WB, Wang X, Strom BL. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. Pharmacoepidemiol Drug Saf 2007 Apr;16(4):393-401.
147. Smeeth L, Douglas I, Hall AJ, Hubbard R, Evans S. Effect of statins on a wide range of health outcomes: a cohort study validated by comparison with randomized trials. British Journal of Clinical Pharmacology 2009;67(1):99-109.
148. Langley T, Szatkowski L, Wythe S, Lewis S. Can primary care data be used to monitor regional smoking prevalence? An analysis of The Health Improvement Network primary care data. BMC Public Health 2011;11(1):773.

149. Davies AR, Smeeth L, Grundy EMD. Contribution of changes in incidence and mortality to trends in the prevalence of coronary heart disease in the UK: 1996GÇô2005. European Heart Journal 2007 Sep 1;28(17):2142-7.
150. Maguire A, Blak BT, Thompson M. The importance of defining periods of complete mortality reporting for research using automated data from primary care. Pharmacoepidemiol Drug Saf 2009 Jan;18(1):76-83.
151. 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 Jul 21;335(7611):136.
152. Meal A, Leonardi-Bee J, Smith C, Hubbard R, Bath-Hextall F. Validation of THIN data for non-melanoma skin cancer. Quality in Primary Care 16[1], 49-52. 2008.
Ref Type: Abstract
153. Seminara NM, Abuabara K, Shin DB, Langan SM, Kimmel SE, Margolis D, et al. Validity of The Health Improvement Network (THIN) for the study of psoriasis. British Journal of Dermatology 2011;164(3):602-9.
154. Townsend P. Deprivation. Journal of Social Policy 1987;16(02):125-46.
155. Smith N, Wilson A, Weekes T. Use of Read codes in development of a standard data set. BMJ 1995 Jul 29;311.
156. Perkovic V, Cass A, Patel AA, Suriyawongpaisal P, Barzi F, Chadban S, et al. High prevalence of chronic kidney disease in Thailand. Kidney Int 2008 Feb;73(4):473-9.
157. Coresh J, Astor BC, McQuillan G, Kusek J, Greene T, Van LF, et al. Calibration and random variation of the serum creatinine assay as critical elements of using equations to estimate glomerular filtration rate. American Journal of Kidney Diseases 2002 May;39(5):920-9.
158. Armitage P, Berry G, Matthews JNS. Statistical Methods in Medical Research. 4th ed. Blackwell Publishing; 2012.
159. Matsushita K, Mahmoodi BK, Woodward M, Emberson JR, Jafar TH, Jee SH, et al. Comparison of risk prediction using the CKD-EPI equation and the MDRD study equation for estimated glomerular filtration rate. JAMA 2012 May 9;307(18):1941-51.
160. Department of Health. The National Service Framework for renal services. Pt. 2, Chronic Kidney Disease, Acute Renal Failure and End of Life Care. London: Department of Health; 2005.
161. National Collaborating Centre for Chronic Conditions. Chronic kidney disease: national clinical guideline for early indentification and management in adults in primaryand secondary care. London: Royal College of Physicians; 2008.
162. Phillips LA, Donovan KL, Phillips AO. Renal quality outcomes framework and eGFR: impact on secondary care. QJM 2009 Jun 1;102(6):415-23.
163. Meran S, Don K, Shah N, Donovan K, Riley S, Phillips AO. Impact of chronic kidney disease management in primary care. QJM 2011 Jan 1;104(1):27-34.
164. Lester H, Campbell S. Developing Quality and Outcomes Framework (QOF) indicators and the concept of 'QOFability'. Quality in Primary Care 18[2], 103-109. 2010.
Ref Type: Abstract
165. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al. Pay-for-Performance Programs in Family Practices in the United Kingdom. N Engl J Med 2006 Jul 27;355(4):375-84.

166. Larsson S, Lawyer P, Garellick G+, Lindahl B, Lundstr+Âm M. Use Of 13 Disease Registries In 5 Countries Demonstrates The Potential To Use Outcome Data To Improve Health CareΓÇÖs Value. Health Affairs 2011 Dec 7.
167. Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. Pay for Performance in Commercial HMOs. N Engl J Med 2006 Nov 2;355(18):1895-902.
168. The Centers for Medicare & Medicaid Services. Physician Quality Reporting System. 2013. 1-8-2013.
Ref Type: Online Source
169. Canadian Chronic Disease Surveillance System. Canadian Chronic Disease Surveillance System (Diabetes). 2013. 1-8-2013.
Ref Type: Online Source
170. Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ 2010;340:c2442.
171. Public Health Agency of Cananda. Chronic Diseases. 2013.
Ref Type: Online Source
172. Salomaa V, Havulinna AS, Koukkunen H, Karja-Koskenkari P, Pietila A, Mustonen J, et al. Aging of the population may not lead to an increase in the numbers of acute coronary events: a community surveillance study and modelled forecast of the future. Heart 2013 Jul;99(13):954-9.
173. Swedish National Health Registries. 2013. 1-8-2013.
Ref Type: Online Source
174. Alshamsan R, Majeed A, Ashworth M, Car J, Millett C. Impact of pay for performance on inequalities in health care: systematic review. J Health Serv Res Policy 2010 Jul;15(3):178-84.
175. Alshamsan R, Millett C, Majeed A, Khunti K. Has pay for performance improved the management of diabetes in the United Kingdom? Prim Care Diabetes 2010 Jul;4(2):73-8.
176. Ashworth M, Medina J, Morgan M. Effect of social deprivation on blood pressure monitoring and control in England: a survey of data from the quality and outcomes framework. BMJ 2008;337:a2030.
177. McGovern MP, Boroujerdi MA, Taylor MW, Williams DJ, Hannaford PC, Lefevre KE, et al. The effect of the UK incentive-based contract on the management of patients with coronary heart disease in primary care. Family Practice 2008 Feb;25(1):33-9.
178. Simpson CR, Hannaford PC, Lefevre K, Williams D. Effect of the UK incentive-based contract on the management of patients with stroke in primary care. Stroke 2006 Sep;37(9):2354-60.
179. Martin S, Smith PC, Dushelko M, Gravelle H, Rice N. Do quality improvments in primary care reduce secondary care costs? London: The Health Foundation; 2010 Jul 1.
180. Dusheiko M, Doran T, Gravelle H, Fullwood C, Roland M. Does Higher Quality of Diabetes Management in Family Practice Reduce Unplanned Hospital Admissions? Health Services Research 2011;46(1p1):27-46.
181. Starfield B, Mangin D. An international perspective on the basis for payment for performance. Quality in Primary Care 2010;18(6):399-404.

182. Calvert MJ, Shankar A, McManus RJ, Ryan R, Freemantle N. Evaluation of the management of heart failure in primary care. Family Practice 2009;26(2):145.
183. Moulton BR. Random group effects and the precision of regression estimates. Journal of Econometrics 1986 Aug;32(3):385-97.
184. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. International Journal of Epidemiology 1999 Oct 1;28(5):964-74.
185. Stone.C.J., Koo CJ. Additive Splines in statistics. Washington DC 1985 p. 45-8.
186. Denburg MR, Haynes K, Shults J, Lewis JD, Leonard MB. Validation of The Health Improvement Network (THIN) database for epidemiologic studies of chronic kidney disease. Pharmacoepidemiol Drug Saf 2011 Nov;20(11):1138-49.
187. Kern EFO, Maney M, Miller DR, Tseng CL, Tiwari A, Rajan M, et al. Failure of ICD-9-CM Codes to Identify Patients with Comorbid Chronic Kidney Disease in Diabetes. Health Services Research 2006 Apr 1;41(2):564-80.
188. Jain P, Calvert M, Cockwell P, McManus RJ. The Need for Improved Identification and Accurate Classification of Stages 3ΓÇô5 Chronic Kidney Disease in Primary Care: Retrospective Cohort Study. PLoS ONE 2014 Aug 12;9(8):e100831.
189. Johnston ME, Gibson ES, Terry CW, Haynes RB, Taylor DW, Gafni A, et al. Effects of labelling on income, work and social function among hypertensive employees. J Chronic Dis 1984;37(6):417-23.
190. Kuznik A, Mardekian J, Tarasenko L. Evaluation of cardiovascular disease burden and therapeutic goal attainment in US adults with chronic kidney disease: an analysis of national health and nutritional examination survey data, 2001-2010. BMC Nephrology 2013;14:132.
191. Yamamoto S, Kon V. Mechanisms for increased cardiovascular disease in chronic kidney dysfunction. Curr Opin Nephrol Hypertens 2009 May;18(3):181-8.
192. Drueke TB, Massy ZA. Atherosclerosis in CKD: differences from the general population. Nat Rev Nephrol 2010 Dec;6(12):723-35.
193. Wild S, Roglic G, Green A, Sicree R, King H. Global Prevalence of Diabetes: Estimates for the year 2000 and projections for 2030. Diabetes Care 2004 May 1;27(5):1047-53.
194. Kannel WB, McGee DL. Diabetes and cardiovascular risk factors: the Framingham study. Circulation 1979 Jan;59(1):8-13.
195. Nathan DM. The role of glycemia management in the prevention of cardiovascular disease--starting over? Annals of Internal Medicine 2009 Dec 15;151(12):888-9.
196. Fox CS, Matsushita K, Woodward M, Bilo HJ, Chalmers J, Heerspink HJ, et al. Associations of kidney disease measures with mortality and end-stage renal disease in individuals with and without diabetes: a meta-analysis. Lancet 2012 Nov 10;380(9854):1662-73.
197. Baber U, Howard VJ, Halperin JL, Soliman EZ, Zhang X, McClellan W, et al. Association of chronic kidney disease with atrial fibrillation among adults in the United States: REasons for Geographic and Racial Differences in Stroke (REGARDS) Study. Circulation: Arrhythmia and Electrophysiology 2011 Feb;4(1):26-32.
198. McManus DD, Corteville DC, Shlipak MG, Whooley MA, Ix JH. Relation of kidney function and albuminuria with atrial fibrillation (from the Heart and Soul Study). Am J Cardiol 2009 Dec 1;104(11):1551-5.

199. Hart RG, Pearce LA, Asinger RW, Herzog CA. Warfarin in atrial fibrillation patients with moderate chronic kidney disease. Clin J Am Soc Nephrol 2011 Nov;6(11):2599-604.
200. The NHS Information Centre. Health Survey for England - 2010. The NHS Information Centre for health and social care; 2011 Nov 15.
201. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. Lancet 2005 Jan 15;365(9455):217-23.
202. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 2002 Dec 14;360(9349):1903-13.
203. Whaley-Connell AT, Sowers JR, Stevens LA, McFarlane SI, Shlipak MG, Norris KC, et al. CKD in the United States: Kidney Early Evaluation Program (KEEP) and National Health and Nutrition Examination Survey (NHANES) 1999-2004. American Journal of Kidney Diseases 2008 Apr;51(4 Suppl 2):S13-S20.
204. Sarnak MJ, Greene T, Wang X, Beck G, Kusek JW, Collins AJ, et al. The effect of a lower target blood pressure on the progression of kidney disease: long-term follow-up of the modification of diet in renal disease study. Annals of Internal Medicine 2005 Mar 1;142(5):342-51.
205. Udani S, Lazich I, Bakris GL. Epidemiology of hypertensive kidney disease. Nat Rev Nephrol 2011 Jan;7(1):11-21.
206. Briet M, Bozec E, Laurent S, Fassot C, London GM, Jacquot C, et al. Arterial stiffness and enlargement in mild-to-moderate chronic kidney disease. Kidney Int 2006 Jan 1;69(2):350-7.
207. Catapano F, Chiodini P, De NL, Minutolo R, Zamboli P, Gallo C, et al. Antiproteinuric response to dual blockade of the renin-angiotensin system in primary glomerulonephritis: meta-analysis and metaregression. American Journal of Kidney Diseases 2008 Sep;52(3):475-85.
208. Neumann J, Ligtenberg G, Klein II, Koomans HA, Blankestijn PJ. Sympathetic hyperactivity in chronic kidney disease: pathogenesis, clinical relevance, and treatment. Kidney Int 2004 May;65(5):1568-76.
209. Portaluppi F, Montanari L, Massari M, Di C, V, Capanna M. Loss of nocturnal decline of blood pressure in hypertension due to chronic renal failure. Am J Hypertens 1991 Jan;4(1 Pt 1):20-6.
210. 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. Nephrol Dial Transplant 2006 May;21(5):1257-62.
211. Lewis EJ, Hunsicker LG, Bain RP, Rohde RD. The effect of angiotensin-converting-enzyme inhibition on diabetic nephropathy. The Collaborative Study Group. N Engl J Med 1993 Nov 11;329(20):1456-62.
212. Brenner et al. Effects of Losartan on Renal and Cardiovascular Outcomes in Patients with Type 2 Diabetes and Nephropathy. N Engl J Med 2001 Sep 20;345(12):861-9.
213. UK Prospective Diabetes Study Group. Tight blood pressure control and risk of macrovascular and microvascular complications in type 2 diabetes: UKPDS 38. BMJ 1998 Sep 12;317(7160):703-13.

214. Yusuf S, Sleight P, Pogue J, Bosch J, Davies R, Dagenais G. Effects of an angiotensin-converting-enzyme inhibitor, ramipril, on cardiovascular events in high-risk patients. The Heart Outcomes Prevention Evaluation Study Investigators. N Engl J Med 2000 Jan 20;342(3):145-53.
215. Patel A, Macmahon S, Chalmers J, Neal B, Woodward M, Billot L, et al. Effects of a fixed combination of perindopril and indapamide on macrovascular and microvascular outcomes in patients with type 2 diabetes mellitus (the ADVANCE trial): a randomised controlled trial. Lancet 2007 Sep 8;370(9590):829-40.
216. Cushman WC, Evans GW, Byington RP, Goff DC, Jr., Grimm RH, Jr., Cutler JA, et al. Effects of intensive blood-pressure control in type 2 diabetes mellitus. N Engl J Med 2010 Apr 29;362(17):1575-85.
217. Schrier RW, Estacio RO, Esler A, Mehler P. Effects of aggressive blood pressure control in normotensive type 2 diabetic patients on albuminuria, retinopathy and strokes. Kidney Int 2002 Mar;61(3):1086-97.
218. Bakris GL, Weir MR, Shanifar S, Zhang Z, Douglas J, van Dijk DJ, et al. Effects of blood pressure level on progression of diabetic nephropathy: results from the RENAAL study. Archives of Internal Medicine 2003 Jul 14;163(13):1555-65.
219. de Zeeuw D, Remuzzi G, Parving HH, Keane WF, Zhang Z, Shahinfar S, et al. Albuminuria, a therapeutic target for cardiovascular protection in type 2 diabetic patients with nephropathy. Circulation 2004 Aug 24;110(8):921-7.
220. Appel LJ, Wright JT, Greene T, Agodoa LY, Astor BC, Bakris GL, et al. Intensive Blood-Pressure Control in Hypertensive Chronic Kidney Disease. N Engl J Med 2010 Sep 1;363(10):918-29.
221. Mann JF, Gerstein HC, Pogue J, Bosch J, Yusuf S. Renal insufficiency as a predictor of cardiovascular outcomes and the impact of ramipril: the HOPE randomized trial. Annals of Internal Medicine 2001 Apr 17;134(8):629-36.
222. Wright JT, Bakris G, Greene T, Agodoa LY, Appel LJ, Charleston J, et al. Effect of Blood Pressure Lowering and Antihypertensive Drug Class on Progression of Hypertensive Kidney Disease. JAMA: The Journal of the American Medical Association 2002 Nov 20;288(19):2421-31.
223. Jafar TH, Stark PC, Schmid CH, Landa M, Maschio G, de Jong PE, et al. Progression of chronic kidney disease: the role of blood pressure control, proteinuria, and angiotensin-converting enzyme inhibition: a patient-level meta-analysis. Annals of Internal Medicine 2003 Aug 19;139(4):244-52.
224. Klahr S, Levey AS, Beck GJ, Caggiula AW, Hunsicker L, Kusek JW, et al. The Effects of Dietary Protein Restriction and Blood-Pressure Control on the Progression of Chronic Renal Disease. N Engl J Med 1994 Mar 31;330(13):877-84.
225. Arguedas JA, Perez MI, Wright JM. Treatment blood pressure targets for hypertension. Cochrane Database Syst Rev 2009;(3):CD004349.
226. Bakris GL, Copley JB, Vicknair N, Sadler R, Leurgans S. Calcium channel blockers versus other antihypertensive therapies on progression of NIDDM associated nephropathy. Kidney Int 1996 Nov;50(5):1641-50.
227. Pohl MA, Blumenthal S, Cordonnier DJ, De AF, Deferrari G, Eisner G, et al. Independent and additive impact of blood pressure control and angiotensin II receptor blockade on renal outcomes in the irbesartan diabetic nephropathy trial: clinical implications and limitations. J Am Soc Nephrol 2005 Oct;16(10):3027-37.

228. Ruggenenti P, Perna A, Gherardi G, Garini G, Zoccali C, Salvadori M, et al. Renoprotective properties of ACE-inhibition in non-diabetic nephropathies with non-nephrotic proteinuria. Lancet 1999 Jul 31;354(9176):359-64.
229. Balamuthusamy S, Srinivasan L, Verma M, Adigopula S, Jalandhara N, Hathiwala S, et al. Renin angiotensin system blockade and cardiovascular outcomes in patients with chronic kidney disease and proteinuria: a meta-analysis. Am Heart J 2008 May;155(5):791-805.
230. Yusuf S, Teo KK, Pogue J, Dyal L, Copland I, Schumacher H, et al. Telmisartan, ramipril, or both in patients at high risk for vascular events. N Engl J Med 2008 Apr 10;358(15):1547-59.
231. Ruggenenti P, Perna A, Loriga G, Ganeva M, Ene-Iordache B, Turturro M, et al. Blood-pressure control for renoprotection in patients with non-diabetic chronic renal disease (REIN-2): multicentre, randomised controlled trial. Lancet 2005 Mar 12;365(9463):939-46.
232. Heerspink HL, de Zeeuw D. Composite renal endpoints: was ACCOMPLISH accomplished? Lancet 2010 Apr 3;375(9721):1140-2.
233. Parving HH, Persson F, Lewis JB, Lewis EJ, Hollenberg NK. Aliskiren combined with losartan in type 2 diabetes and nephropathy. N Engl J Med 2008 Jun 5;358(23):2433-46.
234. Baigent C, Blackwell L, Collins R, Emberson J, Godwin J, Peto R, et al. Aspirin in the primary and secondary prevention of vascular disease: collaborative meta-analysis of individual participant data from randomised trials. Lancet 2009 May 30;373(9678):1849-60.
235. Eknoyan G, Wacksman SJ, Glueck HI, Will JJ. Platelet function in renal failure. N Engl J Med 1969 Mar 27;280(13):677-81.
236. Palmer SC, Craig JC, Navaneethan SD, Tonelli M, Pellegrini F, Strippoli GF. Benefits and harms of statin therapy for persons with chronic kidney disease: a systematic review and meta-analysis. Annals of Internal Medicine 2012 Aug 21;157(4):263-75.
237. Agarwal R, Curley TM. The role of statins in chronic kidney disease. Am J Med Sci 2005 Aug;330(2):69-81.
238. Navaneethan SD, Pansini F, Perkovic V, Manno C, Pellegrini F, Johnson DW, et al. HMG CoA reductase inhibitors (statins) for people with chronic kidney disease not requiring dialysis. Cochrane Database Syst Rev 2009;(2):CD007784.
239. Nitsch D, Grams M, Sang Y, Black C, Cirillo M, Djurdjev O, et al. Associations of estimated glomerular filtration rate and albuminuria with mortality and renal failure by sex: a meta-analysis. BMJ 2013 Jan 29;346.
240. Mehrotra R, Kermah D, Fried L, Adler S, Norris K. Racial Differences in Mortality Among Those with CKD. J Am Soc Nephrol 2008 Jul;19(7):1403-10.
241. Pei YPC, Greenwood CMT, Chery AL, Wu GG. Racial differences in survival of patients on dialysis. Kidney Int 2000 Sep;58(3):1293-9.
242. Hossain MP, Palmer D, Goyder E, El Nahas AM. Social deprivation and prevalence of chronic kidney disease in the UK: workload implications for primary care. QJM 2012 Feb 1;105(2):167-75.
243. Bruce MA, Beech BM, Crook ED, Sims M, Wyatt SB, Flessner MF, et al. Association of Socioeconomic Status and CKD Among African Americans: The Jackson Heart Study. American Journal of Kidney Diseases 55[6], 1001-1008. 1-6-2010.
Ref Type: Abstract

244. Kramer H, Shoham D, McClure LA, Durazo-Arvizu R, Howard G, Judd S, et al. Association of waist circumference and body mass index with all-cause mortality in CKD: The REGARDS (Reasons for Geographic and Racial Differences in Stroke) Study. American Journal of Kidney Diseases 2011 Aug;58(2):177-85.
245. Astor BC, Muntner P, Levin A, Eustace JA, Coresh J. Association of kidney function with anemia: the Third National Health and Nutrition Examination Survey (1988-1994). Archives of Internal Medicine 2002 Jun 24;162(12):1401-8.
246. Zakai NA, Katz R, Hirsch C, Shlipak MG, Chaves PH, Newman AB, et al. A prospective study of anemia status, hemoglobin concentration, and mortality in an elderly cohort: the Cardiovascular Health Study. Archives of Internal Medicine 2005 Oct 24;165(19):2214-20.
247. Pfeffer MA, Burdmann EA, Chen CY, Cooper ME, de ZD, Eckardt KU, et al. A trial of darbepoetin alfa in type 2 diabetes and chronic kidney disease. N Engl J Med 2009 Nov 19;361(21):2019-32.
248. Palmer SC, Hayen A, Macaskill P, Pellegrini F, Craig JC, Elder GJ, et al. Serum levels of phosphorus, parathyroid hormone, and calcium and risks of death and cardiovascular disease in individuals with chronic kidney disease: a systematic review and meta-analysis. JAMA 2011 Mar 16;305(11):1119-27.
249. Block GA, Wheeler DC, Persky MS, Kestenbaum B, Ketteler M, Spiegel DM, et al. Effects of phosphate binders in moderate CKD. J Am Soc Nephrol 2012 Aug;23(8):1407-15.
250. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996 Feb 28;15(4):361-87.
251. Collet D. Modelling Survival Data in Medical Research. 2nd ed. Chapman and Hall/CRC; 2003.
252. Moons KGM, Biesheuvel CJ, Grobbee DE. Test Research versus Diagnostic Research. Clinical Chemistry 2004 Mar 1;50(3):473-6.
253. Mantel N. Why Stepdown Procedures in Variable Selection. Technometrics 1970 Aug 1;12(3):621-5.
254. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 2006;25(1):127-41.
255. Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. Computational Statistics & Data Analysis 2006 Aug;50(12):3464-85.
256. Akaike H. Likelihood of a model and information criteria. Journal of Econometrics 1981;16:3-14.
257. Stocken D. Statistical modelling for the prognostic classification of patients with pancreatic cancer for optimisation of treatment allocation University of Birmingham; 2012.
258. Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. BMJ 2009;339:b2584.
259. Rubin D. Multiple Imputation for Nonresponse in Surveys. New York: Wiley; 1987.
260. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. Stat Med 1991 Apr;10(4):585-98.

261. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 2006 Oct;59(10):1087-91.
262. Harrell FE, Jr. Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis. Springer; 2001.
263. Raghunathan TE, Lepkowski JM, Van Hoewyk J, Solenberger P. A Multivariate Technique for Multiply Imputing Missing Values
Using a Sequence of Regression Models. Survey Methodology 2001 Jun;27(1):85-95.
264. Duchateau L, Janssen, Paul. The Frailty Model. 2010.
265. Hallan SI, Matsushita K, Sang Y, Mahmoodi BK, Black C, Ishani A, et al. Age and association of kidney measures with mortality and end-stage renal disease. JAMA 2012 Dec 12;308(22):2349-60.
266. Muntner P, Anderson A, Charleston J, Chen Z, Ford V, Makos G, et al. Hypertension awareness, treatment, and control in adults with CKD: results from the Chronic Renal Insufficiency Cohort (CRIC) Study. American Journal of Kidney Diseases 2010 Mar;55(3):441-51.
267. Edwards NC, Steeds RP, Stewart PM, Ferro CJ, Townend JN. Effect of Spironolactone on Left Ventricular Mass and Aortic Stiffness in Early-Stage Chronic Kidney Disease: A Randomized Controlled Trial. J Am Coll Cardiol 2009 Aug 4;54(6):505-12.
268. de Brito-Ashurst I, Varagunam M, Raftery MJ, Yaqoob MM. Bicarbonate Supplementation Slows Progression of CKD and Improves Nutritional Status. J Am Soc Nephrol 2009 Sep 1;20(9):2075-84.
269. Singh AK, Szczech L, Tang KL, Barnhart H, Sapp S, Wolfson M, et al. Correction of Anemia with Epoetin Alfa in Chronic Kidney Disease. N Engl J Med 2006 Nov 16;355(20):2085-98.
270. Chertow GM, Burke SK, Raggi P, Treat to Goal Working Group. Sevelamer attenuates the progression of coronary and aortic calcification in hemodialysis patients. Kidney Int 2002 Jul;62(1):245-52.
271. Miller A, Smaglo B, McFarlane S. Obesity, anemia, and cardiovascular risk in patients with chronic kidney disease: Overview and pathophysiologic insights. Current Cardiovascular Risk Reports 2009 Mar 1;3(2):102-8.
272. Smith GL, Lichtman JH, Bracken MB, Shlipak MG, Phillips CO, DiCapua P, et al. Renal impairment and outcomes in heart failure: systematic review and meta-analysis. J Am Coll Cardiol 2006 May 16;47(10):1987-96.
273. Anavekar NS, McMurray JJV, Velazquez EJ, Solomon SD, Kober L, Rouleau JL, et al. Relation between Renal Dysfunction and Cardiovascular Outcomes after Myocardial Infarction. N Engl J Med 2004 Sep 23;351(13):1285-95.
274. Elsayed EF, Tighiouart H, Weiner DE, Griffith J, Salem D, Levey AS, et al. Waist-to-hip ratio and body mass index as risk factors for cardiovascular events in CKD. American Journal of Kidney Diseases 2008 Jul;52(1):49-57.
275. Obermayr RP, Temml C, Gutjahr G, Kainz A, Klauser-Braun R, Fugger R, et al. Body mass index modifies the risk of cardiovascular death in proteinuric chronic kidney disease. Nephrol Dial Transplant 2009 Aug;24(8):2421-8.
276. Bird NJ, Peters C, Michell AR, Peters AM. Reliability of the MDRD method for estimating glomerular filtration rate in relation to gender, body mass index and extracellular fluid volume. Eur J Clin Invest 2008 Jul;38(7):486-93.

277. Chudleigh RA, Dunseath G, Peter R, Harvey JN, Ollerton RL, Luzio S, et al. Influence of Body Weight on the Performance of Glomerular Filtration Rate Estimators in Subjects With Type 2 Diabetes. Diabetes Care 2008 Jan 1;31(1):47-9.
278. Horl WH. Anaemia management and mortality risk in chronic kidney disease. Nat Rev Nephrol 2013 May;9(5):291-301.
279. Salas M, Hotman A, Stricker BH. Confounding by Indication: An Example of Variation in the Use of Epidemiologic Terminology. American Journal of Epidemiology 1999 Jun 1;149(11):981-3.
280. Besarab A, Coyne DW. Iron supplementation to treat anemia in patients with chronic kidney disease. Nat Rev Nephrol 2010 Dec;6(12):699-710.
281. Lowrie EG, Lew NL. Death risk in hemodialysis patients: the predictive value of commonly measured variables and an evaluation of death rate differences between facilities. American Journal of Kidney Diseases 1990 May;15(5):458-82.
282. Upadhyay A, Earley A, Haynes SM, Uhlig K. Systematic review: blood pressure target in chronic kidney disease and proteinuria as an effect modifier. Annals of Internal Medicine 2011 Apr 19;154(8):541-8.
283. 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. Lancet 2012 Nov 10;380(9854):1649-61.
284. Kovesdy CP, Bleyer AJ, Molnar MZ, Ma JZ, Sim JJ, Cushman WC, et al. Blood Pressure and Mortality in U.S. Veterans With Chronic Kidney DiseaseA Cohort Study. Annals of Internal Medicine 2013 Aug 20;159(4):233-42.
285. Bakris GL, Hart P, Ritz E. Beta blockers in the management of chronic kidney disease. Kidney Int 2006 Oct 4;70(11):1905-13.
286. Ruster C, Wolf G. Renin-Angiotensin-Aldosterone System and Progression of Renal Disease. Journal of the American Society of Nephrology 2006 Nov 1;17(11):2985-91.
287. Chapple CR. A Comparison of Varying alpha-Blockers and Other Pharmacotherapy Options for Lower Urinary Tract Symptoms. Rev Urol 2005;7 Suppl 4:S22-S30.
288. Gooch K, Culleton BF, Manns BJ, Zhang J, Alfonso H, Tonelli M, et al. NSAID use and progression of chronic kidney disease. Am J Med 2007 Mar;120(3):280-7.
289. Leonetti G, Cuspidi C, Facchini M, Stramba-Badiale M. Is systolic pressure a better target for antihypertensive treatment than diastolic pressure? J Hypertens Suppl 2000 Jul;18(3):S13-S20.
290. Nielsen WB, Lindenstrom E, Vestbo J, Jensen GB. Is diastolic hypertension an independent risk factor for stroke in the presence of normal systolic blood pressure in the middle-aged and elderly? Am J Hypertens 1997 Jun;10(6):634-9.
291. Mentari E, Rahman M. Blood pressure and progression of chronic kidney disease: importance of systolic, diastolic, or diurnal variation. Curr Hypertens Rep 2004 Oct;6(5):400-4.
292. Tangri N, Kitsios GD, Inker LA, Griffith J, Naimark DM, Walker S, et al. Risk Prediction Models for Patients With Chronic Kidney DiseaseA Systematic Review. Annals of Internal Medicine 2013 Apr 16;158(8):596-603.
293. Arana A, Wentworth CE, Ayuso-Mateos JL, Arellano FM. Suicide-related events in patients treated with antiepileptic drugs. N Engl J Med 2010 Aug 5;363(6):542-51.

294. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Brindle P. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. Heart 2008 Jan 1;94(1):34-9.
295. Coresh J, Byrd-Holt D, Astor BC, Briggs JP, Eggers PW, Lacher DA, et al. Chronic kidney disease awareness, prevalence, and trends among U.S. adults, 1999 to 2000. J Am Soc Nephrol 2005;16:180-8.
296. Mahmud A, Feely J. Beta-blockers reduce aortic stiffness in hypertension but nebivolol, not atenolol, reduces wave reflection. Am J Hypertens 2008 Jun;21(6):663-7.
297. Shahin Y, Khan JA, Chetter I. Angiotensin converting enzyme inhibitors effect on arterial stiffness and wave reflections: A meta-analysis and meta-regression of randomised controlled trials. Atherosclerosis 2012 Mar;221(1):18-33.
298. Liang CC, Muo CH, Wang IK, Chang CT, Chou CY, Liu JH, et al. Peptic Ulcer Disease Risk in Chronic Kidney Disease: Ten-Year Incidence, Ulcer Location, and Ulcerogenic Effect of Medications. PLoS ONE 2014 Feb 3;9(2):e87952.
299. Violan C, Foguet-Boreu Q, Hermosilla-Perez E, Valderas JM, Bolibar B, Fabregas-Escurriola M, et al. Comparison of the information provided by electronic health records data and a population health survey to estimate prevalence of selected health conditions and multimorbidity. BMC Public Health 2013;13:251.
300. Walz G, Budde K, Mannaa M, N++rnberger J, Wanner C, Sommerer C, et al. Everolimus in Patients with Autosomal Dominant Polycystic Kidney Disease. New England Journal of Medicine 2010 Jun 26;363(9):830-40.
301. Jamal SA, Vandermeer B, Raggi P, Mendelssohn DC, Chatterley T, Dorgan M, et al. Effect of calcium-based versus non-calcium-based phosphate binders on mortality in patients with chronic kidney disease: an updated systematic review and meta-analysis. The Lancet 2012;382(9900):1268-77.
302. de Brito-Ashurst I, Varagunam M, Raftery MJ, Yaqoob MM. Bicarbonate supplementation slows progression of CKD and improves nutritional status. J Am Soc Nephrol 2009 Sep;20(9):2075-84.
303. Wright, Jr JT, Bakris G, Greene T. Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: Results from the aask trial. JAMA 2002 Nov 20;288(19):2421-31.

# Appendix A

Figure 1. Histogram of original Cholesterol data pre imputation

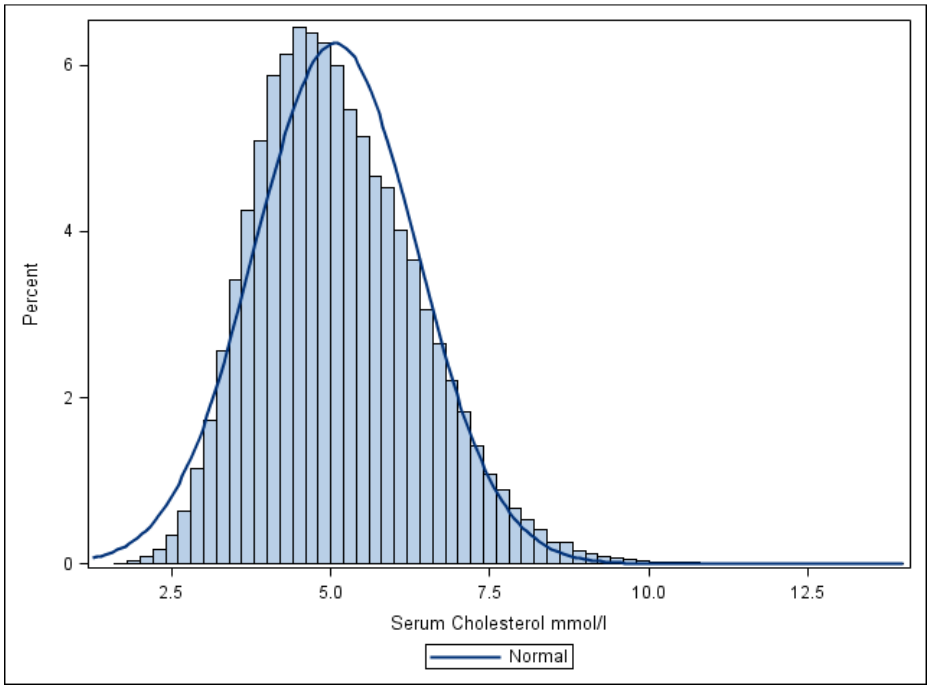



Figure 2. Histogram of original Body Mass index pre imputation

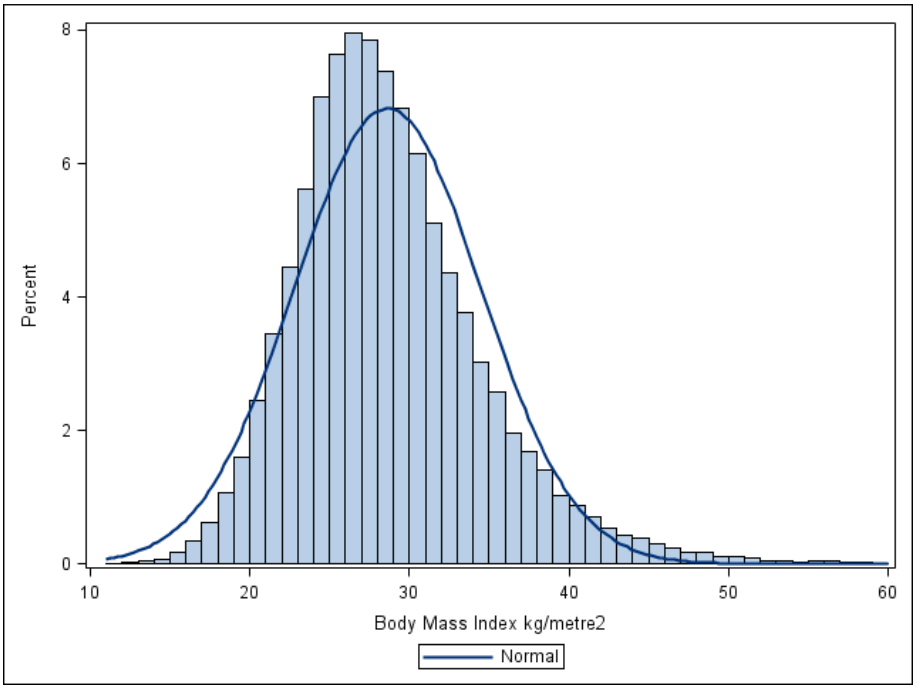


Figure 3. Histogram of original Systolic blood pressure data pre imputation

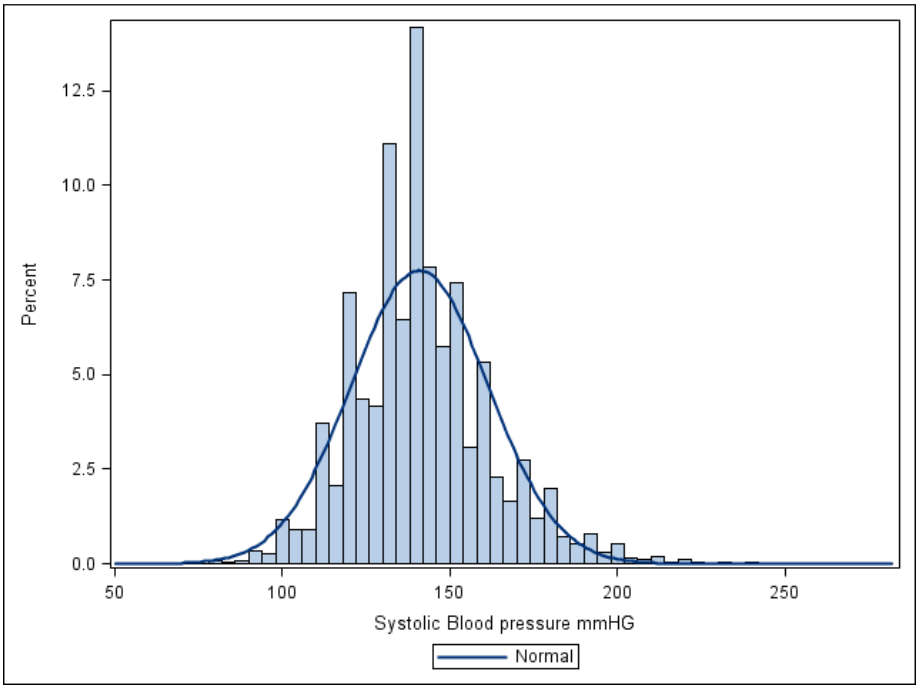



Figure 4. Histogram of original Diastolic blood pressure data pre imputation

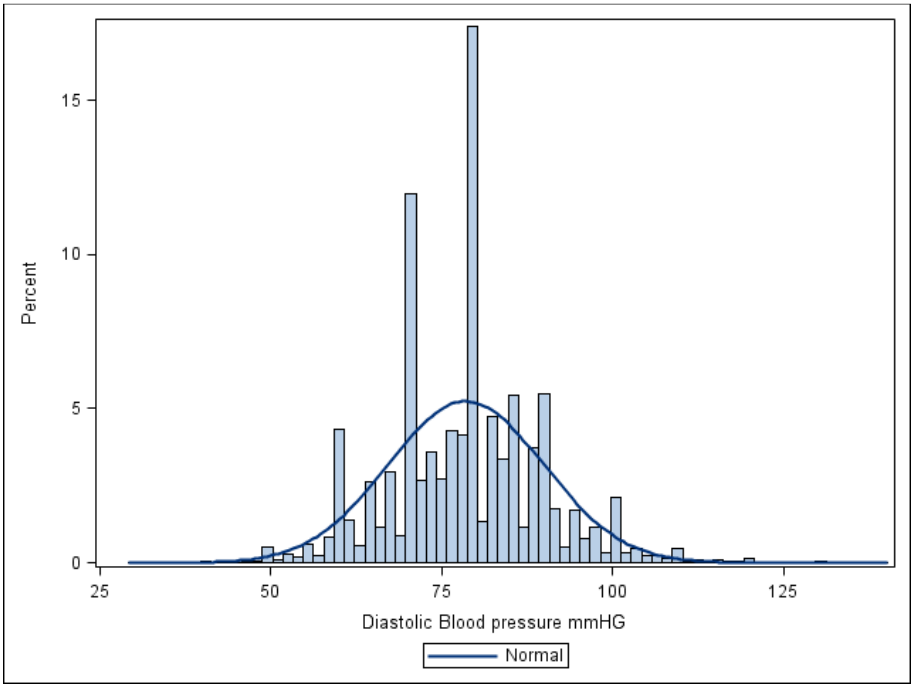


Figure 5. Histogram of original Haemoglobin data pre imputation

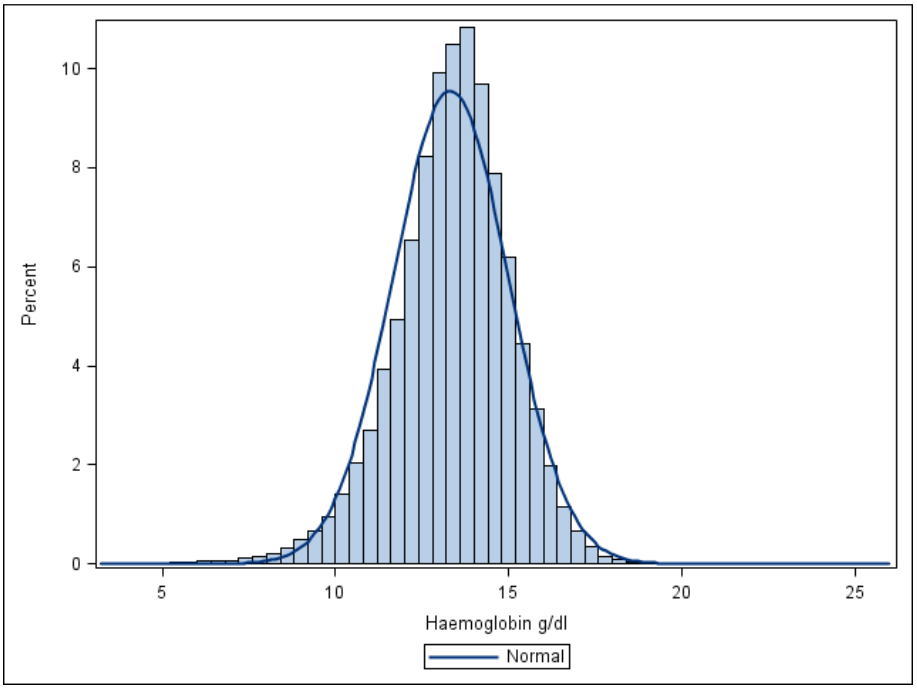


Table 1. Cox regression analysis for Imputation 1 without transformation and frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age | 6.154 | 0.110 | 470.761 | 379.572 | 583.857 |
| Male Gender | 0.485 | 0.020 | 1.625 | 1.563 | 1.689 |
| Race with Caucasian race as reference African – Caribbean | -0.517 | 0.190 | 0.596 | 0.411 | 0.865 |
| Indian Subcontinent | -0.769 | 0.151 | 0.464 | 0.344 | 0.624 |
| South East Asian | -1.183 | 1.000 | 0.306 | 0.043 | 2.176 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.112 | 0.026 | 1.118 | 1.062 | 1.177 |
| $3^{rd}$ | 0.157 | 0.027 | 1.170 | 1.110 | 1.233 |
| $4^{th}$ | 0.179 | 0.027 | 1.196 | 1.133 | 1.262 |
| $5^{th}$ | 0.158 | 0.031 | 1.172 | 1.103 | 1.245 |
| Diabetes Mellitus | 0.160 | 0.023 | 1.173 | 1.121 | 1.228 |
| Heart Failure | 0.488 | 0.024 | 1.630 | 1.554 | 1.709 |
| Atrial Fibrillation | 0.072 | 0.025 | 1.074 | 1.023 | 1.128 |
| Coronary Heart Disease | 0.063 | 0.021 | 1.065 | 1.021 | 1.110 |
| Cerebrovascular Accident | 0.221 | 0.022 | 1.248 | 1.194 | 1.303 |
| Peripheral Vascular Disease | 0.161 | 0.029 | 1.175 | 1.109 | 1.244 |
| Ever Smoked | 0.145 | 0.023 | 1.157 | 1.106 | 1.210 |
| Systolic BP | -0.655 | 0.045 | 0.519 | 0.476 | 0.567 |
| BMI | -0.213 | 0.019 | 0.808 | 0.779 | 0.839 |
| Haemoglobin | -1.673 | 0.055 | 0.188 | 0.168 | 0.209 |
| Glomerular Filtration Rate | -2.131 | 0.101 | 0.119 | 0.097 | 0.145 |
| Cholesterol | -0.676 | 0.084 | 0.509 | 0.431 | 0.600 |
| Proteinuria levels with none as reference High | -0.050 | 0.033 | 0.951 | 0.891 | 1.015 |
| Very High | 0.090 | 0.043 | 1.094 | 1.006 | 1.189 |
| Anti-platelets | 0.080 | 0.021 | 1.083 | 1.040 | 1.129 |
| Anticoagulation | 0.256 | 0.033 | 1.291 | 1.211 | 1.377 |
| Angiotensin Blockade | -0.164 | 0.020 | 0.849 | 0.816 | 0.882 |
| Beta blockade | -0.150 | 0.021 | 0.861 | 0.826 | 0.897 |
| Calcium Channel Blocker | -0.048 | 0.022 | 0.953 | 0.913 | 0.995 |
| Diuretic use | 0.208 | 0.019 | 1.231 | 1.185 | 1.279 |
| Other | -0.140 | 0.038 | 0.870 | 0.808 | 0.937 |
| Lipid | -0.312 | 0.023 | 0.732 | 0.700 | 0.765 |
| Iron Medication | 0.129 | 0.029 | 1.137 | 1.075 | 1.204 |
| Vitamin D supplementation | 0.218 | 0.026 | 1.243 | 1.180 | 1.309 |

Table 2. Cox regression analysis for Imputation 2 without transformation and frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age | 6.224 | 0.110 | 504.617 | 406.974 | 625.688 |
| Male Gender | 0.475 | 0.020 | 1.608 | 1.547 | 1.672 |
| Race with Caucasian race as reference African – Caribbean | -0.494 | 0.190 | 0.610 | 0.421 | 0.885 |
| Indian Subcontinent | -0.777 | 0.151 | 0.460 | 0.342 | 0.619 |
| South East Asian | -1.174 | 1.000 | 0.309 | 0.044 | 2.195 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.123 | 0.026 | 1.131 | 1.074 | 1.191 |
| $3^{rd}$ | 0.163 | 0.027 | 1.177 | 1.117 | 1.240 |
| $4^{th}$ | 0.178 | 0.027 | 1.195 | 1.133 | 1.261 |
| $5^{th}$ | 0.155 | 0.031 | 1.168 | 1.099 | 1.241 |
| Diabetes Mellitus | 0.183 | 0.023 | 1.200 | 1.147 | 1.256 |
| Heart Failure | 0.480 | 0.024 | 1.616 | 1.541 | 1.695 |
| Atrial Fibrillation | 0.070 | 0.025 | 1.073 | 1.021 | 1.126 |
| Coronary Heart Disease | 0.059 | 0.021 | 1.061 | 1.017 | 1.106 |
| Cerebrovascular Accident | 0.217 | 0.022 | 1.242 | 1.189 | 1.297 |
| Peripheral Vascular Disease | 0.155 | 0.029 | 1.167 | 1.102 | 1.236 |
| Ever Smoked | 0.148 | 0.023 | 1.159 | 1.109 | 1.212 |
| Systolic BP | -0.675 | 0.045 | 0.509 | 0.466 | 0.556 |
| BMI | -0.194 | 0.019 | 0.823 | 0.793 | 0.855 |
| Haemoglobin | -1.602 | 0.056 | 0.202 | 0.181 | 0.225 |
| Glomerular Filtration Rate | -2.078 | 0.102 | 0.125 | 0.103 | 0.153 |
| Cholesterol | -0.788 | 0.084 | 0.455 | 0.386 | 0.537 |
| Proteinuria levels with none as reference High | 0.070 | 0.035 | 1.072 | 1.001 | 1.149 |
| Very High | 0.320 | 0.044 | 1.378 | 1.265 | 1.501 |
| Anti-platelets | 0.070 | 0.021 | 1.073 | 1.030 | 1.118 |
| Anticoagulation | 0.251 | 0.033 | 1.286 | 1.205 | 1.371 |
| Angiotensin Blockade | -0.172 | 0.020 | 0.842 | 0.810 | 0.875 |
| Beta blockade | -0.151 | 0.021 | 0.860 | 0.825 | 0.896 |
| Calcium Channel Blocker | -0.046 | 0.022 | 0.955 | 0.914 | 0.997 |
| Diuretic use | 0.205 | 0.019 | 1.227 | 1.181 | 1.275 |
| Other | -0.142 | 0.038 | 0.868 | 0.806 | 0.935 |
| Lipid | -0.312 | 0.023 | 0.732 | 0.700 | 0.765 |
| Iron Medication | 0.135 | 0.029 | 1.144 | 1.082 | 1.211 |
| Vitamin D supplementation | 0.225 | 0.026 | 1.252 | 1.189 | 1.318 |

Table 3. Cox regression analysis for Imputation 3 without transformation and frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age | 6.248 | 0.110 | 517.200 | 416.918 | 641.604 |
| Male Gender | 0.491 | 0.020 | 1.634 | 1.572 | 1.698 |
| Race with Caucasian race as reference African – Caribbean | -0.508 | 0.190 | 0.601 | 0.415 | 0.872 |
| Indian Subcontinent | -0.763 | 0.151 | 0.466 | 0.347 | 0.628 |
| South East Asian | -1.174 | 1.000 | 0.309 | 0.044 | 2.196 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.116 | 0.026 | 1.123 | 1.067 | 1.182 |
| $3^{rd}$ | 0.157 | 0.027 | 1.171 | 1.111 | 1.234 |
| $4^{th}$ | 0.176 | 0.027 | 1.192 | 1.129 | 1.258 |
| $5^{th}$ | 0.166 | 0.031 | 1.180 | 1.111 | 1.254 |
| Diabetes Mellitus | 0.168 | 0.023 | 1.183 | 1.130 | 1.239 |
| Heart Failure | 0.483 | 0.024 | 1.622 | 1.546 | 1.701 |
| Atrial Fibrillation | 0.070 | 0.025 | 1.073 | 1.022 | 1.127 |
| Coronary Heart Disease | 0.058 | 0.021 | 1.060 | 1.017 | 1.105 |
| Cerebrovascular Accident | 0.226 | 0.022 | 1.254 | 1.201 | 1.310 |
| Peripheral Vascular Disease | 0.161 | 0.029 | 1.175 | 1.109 | 1.244 |
| Ever Smoked | 0.151 | 0.023 | 1.163 | 1.112 | 1.216 |
| Systolic BP | -0.685 | 0.045 | 0.504 | 0.462 | 0.550 |
| BMI | -0.157 | 0.019 | 0.854 | 0.823 | 0.886 |
| Haemoglobin | -1.692 | 0.055 | 0.184 | 0.165 | 0.205 |
| Glomerular Filtration Rate | -2.141 | 0.101 | 0.117 | 0.096 | 0.143 |
| Cholesterol | -0.669 | 0.084 | 0.512 | 0.435 | 0.604 |
| Proteinuria levels with none as reference High | 0.029 | 0.036 | 1.030 | 0.960 | 1.104 |
| Very High | 0.130 | 0.045 | 1.138 | 1.042 | 1.244 |
| Anti-platelets | 0.078 | 0.021 | 1.081 | 1.037 | 1.126 |
| Anticoagulation | 0.261 | 0.033 | 1.298 | 1.216 | 1.384 |
| Angiotensin Blockade | -0.173 | 0.020 | 0.841 | 0.810 | 0.875 |
| Beta blockade | -0.152 | 0.021 | 0.859 | 0.824 | 0.895 |
| Calcium Channel Blocker | -0.038 | 0.022 | 0.962 | 0.922 | 1.005 |
| Diuretic use | 0.199 | 0.019 | 1.220 | 1.175 | 1.268 |
| Other | -0.137 | 0.038 | 0.872 | 0.810 | 0.939 |
| Lipid | -0.310 | 0.023 | 0.733 | 0.702 | 0.767 |
| Iron Medication | 0.129 | 0.029 | 1.138 | 1.075 | 1.204 |
| Vitamin D supplementation | 0.224 | 0.027 | 1.251 | 1.188 | 1.318 |

Table 4, Cox regression analysis for Imputation 4 without transformation and frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age | 6.242 | 0.110 | 513.685 | 414.271 | 636.956 |
| Male Gender | 0.478 | 0.020 | 1.613 | 1.552 | 1.677 |
| Race with Caucasian race as reference African – Caribbean | -0.486 | 0.190 | 0.615 | 0.424 | 0.892 |
| Indian Subcontinent | -0.752 | 0.151 | 0.471 | 0.350 | 0.634 |
| South East Asian | -1.225 | 1.000 | 0.294 | 0.041 | 2.086 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.124 | 0.026 | 1.132 | 1.075 | 1.192 |
| $3^{rd}$ | 0.151 | 0.027 | 1.163 | 1.104 | 1.226 |
| $4^{th}$ | 0.171 | 0.027 | 1.186 | 1.124 | 1.252 |
| $5^{th}$ | 0.158 | 0.031 | 1.171 | 1.103 | 1.244 |
| Diabetes Mellitus | 0.178 | 0.023 | 1.195 | 1.142 | 1.251 |
| Heart Failure | 0.487 | 0.024 | 1.627 | 1.552 | 1.706 |
| Atrial Fibrillation | 0.069 | 0.025 | 1.071 | 1.020 | 1.125 |
| Coronary Heart Disease | 0.064 | 0.021 | 1.066 | 1.022 | 1.111 |
| Cerebrovascular Accident | 0.219 | 0.022 | 1.245 | 1.192 | 1.301 |
| Peripheral Vascular Disease | 0.155 | 0.029 | 1.167 | 1.102 | 1.237 |
| Ever Smoked | 0.146 | 0.023 | 1.157 | 1.106 | 1.210 |
| Systolic BP | -0.684 | 0.045 | 0.505 | 0.462 | 0.551 |
| BMI | -0.186 | 0.019 | 0.830 | 0.800 | 0.862 |
| Haemoglobin | -1.597 | 0.056 | 0.203 | 0.182 | 0.226 |
| Glomerular Filtration Rate | -2.165 | 0.101 | 0.115 | 0.094 | 0.140 |
| Cholesterol | -0.687 | 0.084 | 0.503 | 0.426 | 0.594 |
| Proteinuria levels with none as reference High | 0.033 | 0.035 | 1.034 | 0.964 | 1.108 |
| Very High | 0.249 | 0.044 | 1.283 | 1.176 | 1.398 |
| Anti-platelets | 0.078 | 0.021 | 1.081 | 1.038 | 1.127 |
| Anticoagulation | 0.255 | 0.033 | 1.291 | 1.210 | 1.377 |
| Angiotensin Blockade | -0.168 | 0.020 | 0.846 | 0.814 | 0.879 |
| Beta blockade | -0.156 | 0.021 | 0.856 | 0.821 | 0.892 |
| Calcium Channel Blocker | -0.045 | 0.022 | 0.956 | 0.916 | 0.998 |
| Diuretic use | 0.200 | 0.019 | 1.221 | 1.176 | 1.269 |
| Other | -0.134 | 0.038 | 0.874 | 0.812 | 0.942 |
| Lipid | -0.314 | 0.023 | 0.730 | 0.699 | 0.763 |
| Iron Medication | 0.137 | 0.029 | 1.147 | 1.084 | 1.214 |
| Vitamin D supplementation | 0.223 | 0.026 | 1.250 | 1.187 | 1.316 |

Table 5. Cox regression analysis for Imputation 5 without transformation and frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age | 6.292 | 0.110 | 540.497 | 435.740 | 670.439 |
| Male Gender | 0.481 | 0.020 | 1.618 | 1.557 | 1.682 |
| Race with Caucasian race as reference African – Caribbean | -0.487 | 0.190 | 0.614 | 0.424 | 0.891 |
| Indian Subcontinent | -0.753 | 0.151 | 0.471 | 0.350 | 0.634 |
| South East Asian | -1.210 | 1.000 | 0.298 | 0.042 | 2.118 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.116 | 0.026 | 1.123 | 1.067 | 1.183 |
| $3^{rd}$ | 0.153 | 0.027 | 1.165 | 1.106 | 1.228 |
| $4^{th}$ | 0.173 | 0.027 | 1.188 | 1.126 | 1.254 |
| $5^{th}$ | 0.159 | 0.031 | 1.172 | 1.103 | 1.245 |
| Diabetes Mellitus | 0.159 | 0.023 | 1.173 | 1.120 | 1.228 |
| Heart Failure | 0.481 | 0.024 | 1.617 | 1.542 | 1.696 |
| Atrial Fibrillation | 0.065 | 0.025 | 1.068 | 1.017 | 1.121 |
| Coronary Heart Disease | 0.065 | 0.021 | 1.067 | 1.023 | 1.112 |
| Cerebrovascular Accident | 0.222 | 0.022 | 1.248 | 1.195 | 1.304 |
| Peripheral Vascular Disease | 0.164 | 0.029 | 1.178 | 1.112 | 1.248 |
| Ever Smoked | 0.143 | 0.023 | 1.154 | 1.103 | 1.207 |
| Systolic BP | -0.680 | 0.045 | 0.507 | 0.464 | 0.553 |
| BMI | -0.149 | 0.019 | 0.861 | 0.830 | 0.894 |
| Haemoglobin | -1.586 | 0.056 | 0.205 | 0.184 | 0.228 |
| Glomerular Filtration Rate | -2.096 | 0.102 | 0.123 | 0.101 | 0.150 |
| Cholesterol | -0.725 | 0.085 | 0.484 | 0.410 | 0.572 |
| Proteinuria levels with none as reference High | -0.018 | 0.035 | 0.982 | 0.917 | 1.052 |
| Very High | 0.217 | 0.043 | 1.243 | 1.143 | 1.352 |
| Anti-platelets | 0.073 | 0.021 | 1.076 | 1.032 | 1.121 |
| Anticoagulation | 0.254 | 0.033 | 1.290 | 1.209 | 1.376 |
| Angiotensin Blockade | -0.169 | 0.020 | 0.845 | 0.813 | 0.878 |
| Beta blockade | -0.146 | 0.021 | 0.864 | 0.830 | 0.901 |
| Calcium Channel Blocker | -0.044 | 0.022 | 0.957 | 0.916 | 0.999 |
| Diuretic use | 0.199 | 0.019 | 1.220 | 1.174 | 1.267 |
| Other | -0.140 | 0.038 | 0.869 | 0.807 | 0.936 |
| Lipid | -0.319 | 0.023 | 0.727 | 0.695 | 0.760 |
| Iron Medication | 0.142 | 0.029 | 1.152 | 1.089 | 1.219 |
| Vitamin D supplementation | 0.218 | 0.027 | 1.243 | 1.180 | 1.310 |

Table 6. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 1

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age/100 | 7.751 | 0.758 | 2323.895 | 526.526 | 10256.839 |
| Age/100 x log(Age/100) | 0.617 | 0.537 | 1.853 | 0.647 | 5.308 |
| Male Gender | 0.504 | 0.020 | 1.655 | 1.592 | 1.721 |
| Race with Caucasian race as reference African – Caribbean | -0.570 | 0.190 | 0.566 | 0.390 | 0.821 |
| Indian Subcontinent | -0.779 | 0.152 | 0.459 | 0.341 | 0.617 |
| South East Asian | -1.239 | 1.000 | 0.290 | 0.041 | 2.056 |
| Townsend Quintile compared to 1th Quintile | 0.112 | 0.026 | 1.118 | 1.062 | 1.177 |
| $2^{nd}$ | 0.152 | 0.027 | 1.164 | 1.105 | 1.227 |
| $3^{rd}$ | 0.174 | 0.027 | 1.190 | 1.128 | 1.256 |
| $4^{th}$ | 0.152 | 0.031 | 1.164 | 1.096 | 1.237 |
| $5^{th}$ | 0.159 | 0.023 | 1.172 | 1.119 | 1.227 |
| Diabetes Mellitus | 0.472 | 0.024 | 1.603 | 1.529 | 1.681 |
| Heart Failure | 0.071 | 0.025 | 1.074 | 1.023 | 1.128 |
| Atrial Fibrillation | 0.064 | 0.021 | 1.066 | 1.023 | 1.111 |
| Coronary Heart Disease | 0.212 | 0.022 | 1.236 | 1.183 | 1.291 |
| Cerebrovascular Accident | 0.160 | 0.029 | 1.173 | 1.107 | 1.242 |
| Peripheral Vascular Disease | 0.150 | 0.023 | 1.162 | 1.111 | 1.216 |
| Ever Smoked | -3.384 | 0.311 | 0.034 | 0.018 | 0.062 |
| Systolic BP/100 | 0.970 | 0.108 | 2.638 | 2.135 | 3.260 |
| (Systolic BP/100)2 | -0.648 | 0.050 | 0.523 | 0.474 | 0.577 |
| BMI/10 | 0.018 | 0.002 | 1.018 | 1.014 | 1.021 |
| (BMI/10)3 | 5.030 | 0.363 | 152.933 | 75.092 | 311.464 |
| (Haemoglobin/10)-1 x log(Heaemoglobin/10) | 6.749 | 0.356 | 853.205 | 424.306 | 1715.644 |
| (Haemoglobin/10)-1 | -2.303 | 0.117 | 0.100 | 0.079 | 0.126 |
| (Glomerular Filtration Rate/100)2 | -2.534 | 0.353 | 0.079 | 0.040 | 0.159 |
| (Cholesterol/10)2 | 2.218 | 0.376 | 9.189 | 4.394 | 19.216 |
| (Cholesterol/10)3 | -0.046 | 0.033 | 0.955 | 0.895 | 1.020 |
| Proteinuria levels with none as reference High | 0.088 | 0.042 | 1.092 | 1.004 | 1.186 |
| Very High | 0.075 | 0.021 | 1.078 | 1.035 | 1.124 |
| Anti-platelets | 0.261 | 0.033 | 1.298 | 1.217 | 1.385 |
| Anticoagulation | -0.161 | 0.020 | 0.851 | 0.819 | 0.885 |
| Angiotensin Blockade | -0.146 | 0.021 | 0.864 | 0.829 | 0.901 |
| Beta blockade | 0.205 | 0.019 | 1.227 | 1.181 | 1.275 |
| Diuretic use | -0.148 | 0.038 | 0.863 | 0.801 | 0.929 |
| Other | -0.292 | 0.023 | 0.747 | 0.714 | 0.781 |
| Lipid | 0.115 | 0.029 | 1.122 | 1.060 | 1.188 |
| Iron Medication | 0.209 | 0.027 | 1.232 | 1.169 | 1.298 |
| Vitamin D supplementation | | | | | |

Table 7. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 2

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age/100 | 7.518 | 0.759 | 1840.514 | 415.871 | 8145.528 |
| Age/100 x log(Age/100) | 0.893 | 0.538 | 2.443 | 0.851 | 7.018 |
| Male Gender | 0.491 | 0.020 | 1.633 | 1.571 | 1.698 |
| Race with Caucasian race as reference African – Caribbean | -0.541 | 0.190 | 0.582 | 0.401 | 0.845 |
| Indian Subcontinent | -0.788 | 0.151 | 0.455 | 0.338 | 0.612 |
| South East Asian | -1.256 | 1.000 | 0.285 | 0.040 | 2.023 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.123 | 0.026 | 1.131 | 1.074 | 1.191 |
| $3^{rd}$ | 0.159 | 0.027 | 1.173 | 1.113 | 1.236 |
| $4^{th}$ | 0.175 | 0.027 | 1.191 | 1.129 | 1.257 |
| $5^{th}$ | 0.147 | 0.031 | 1.158 | 1.090 | 1.231 |
| Diabetes Mellitus | 0.178 | 0.023 | 1.195 | 1.141 | 1.251 |
| Heart Failure | 0.458 | 0.024 | 1.580 | 1.507 | 1.657 |
| Atrial Fibrillation | 0.068 | 0.025 | 1.071 | 1.020 | 1.124 |
| Coronary Heart Disease | 0.059 | 0.021 | 1.061 | 1.018 | 1.106 |
| Cerebrovascular Accident | 0.207 | 0.022 | 1.230 | 1.178 | 1.285 |
| Peripheral Vascular Disease | 0.156 | 0.029 | 1.168 | 1.103 | 1.238 |
| Ever Smoked | 0.152 | 0.023 | 1.165 | 1.114 | 1.218 |
| Systolic BP/100 | -3.438 | 0.312 | 0.032 | 0.017 | 0.059 |
| (Systolic BP/100)2 | 0.982 | 0.108 | 2.671 | 2.161 | 3.301 |
| BMI/10 | -1.146 | 0.100 | 0.318 | 0.261 | 0.386 |
| (BMI/10)2 | 0.170 | 0.017 | 1.185 | 1.146 | 1.226 |
| (Haemoglobin/10)-1 x log(Heaemoglobin/10) | 4.757 | 0.363 | 116.365 | 57.176 | 236.827 |
| (Haemoglobin/10)-1 | 6.387 | 0.355 | 593.941 | 296.027 | 1191.668 |
| (Glomerular Filtration Rate/100) | -1.946 | 0.102 | 0.143 | 0.117 | 0.174 |
| (Cholesterol/10)2 | -3.336 | 0.327 | 0.036 | 0.019 | 0.068 |
| (Cholesterol/10)3 | 3.008 | 0.341 | 20.252 | 10.383 | 39.499 |
| Proteinuria levels with none as reference High | 0.072 | 0.035 | 1.075 | 1.003 | 1.152 |
| Very High | 0.310 | 0.044 | 1.364 | 1.252 | 1.485 |
| Anti-platelets | 0.062 | 0.021 | 1.064 | 1.021 | 1.109 |
| Anticoagulation | 0.264 | 0.033 | 1.302 | 1.220 | 1.389 |
| Angiotensin Blockade | -0.169 | 0.020 | 0.845 | 0.812 | 0.878 |
| Beta blockade | -0.146 | 0.021 | 0.864 | 0.829 | 0.900 |
| Diuretic use | 0.202 | 0.019 | 1.224 | 1.178 | 1.271 |
| Other | -0.148 | 0.038 | 0.862 | 0.801 | 0.928 |
| Lipid | -0.298 | 0.023 | 0.743 | 0.710 | 0.777 |
| Iron Medication | 0.123 | 0.029 | 1.131 | 1.069 | 1.197 |

| | | | | | |
|---|---|---|---|---|---|
| Vitamin D supplementation | 0.216 | 0.026 | 1.241 | 1.178 | 1.307 |

Table 8. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 3

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age/100 | 7.830 | 0.756 | 2514.225 | 571.494 | 11061.059 |
| Age/100 x log(Age/100) | 0.665 | 0.536 | 1.945 | 0.681 | 5.558 |
| Male Gender | 0.509 | 0.020 | 1.664 | 1.600 | 1.730 |
| Race with Caucasian race as reference African – Caribbean | -0.561 | 0.190 | 0.571 | 0.393 | 0.828 |
| Indian Subcontinent | -0.775 | 0.151 | 0.461 | 0.342 | 0.620 |
| South East Asian | -1.229 | 1.000 | 0.293 | 0.041 | 2.078 |
| Townsend Quintile compared to 1th Quintile | 0.112 | 0.026 | 1.119 | 1.063 | 1.178 |
| $2^{nd}$ | | | | | |
| $3^{rd}$ | 0.151 | 0.027 | 1.163 | 1.104 | 1.226 |
| $4^{th}$ | 0.168 | 0.027 | 1.183 | 1.121 | 1.248 |
| $5^{th}$ | 0.154 | 0.031 | 1.167 | 1.098 | 1.240 |
| Diabetes Mellitus | 0.168 | 0.024 | 1.183 | 1.129 | 1.239 |
| Heart Failure | 0.464 | 0.024 | 1.590 | 1.516 | 1.668 |
| Atrial Fibrillation | 0.070 | 0.025 | 1.072 | 1.021 | 1.126 |
| Coronary Heart Disease | 0.059 | 0.021 | 1.061 | 1.018 | 1.106 |
| Cerebrovascular Accident | 0.217 | 0.022 | 1.242 | 1.189 | 1.297 |
| Peripheral Vascular Disease | 0.160 | 0.029 | 1.174 | 1.108 | 1.243 |
| Ever Smoked | 0.157 | 0.023 | 1.170 | 1.118 | 1.223 |
| Systolic BP/100 | -3.618 | 0.308 | 0.027 | 0.015 | 0.049 |
| (Systolic BP/100)2 | 1.041 | 0.107 | 2.833 | 2.297 | 3.494 |
| (BMI/10)2 x log (BMI/10) | 0.175 | 0.017 | 1.191 | 1.153 | 1.230 |
| (BMI/10)2 | -0.300 | 0.027 | 0.741 | 0.703 | 0.781 |
| (Haemoglobin/10)-1 x log(Heaemoglobin/10) | 5.017 | 0.362 | 151.032 | 74.285 | 307.068 |
| (Haemoglobin/10)-1 | 6.751 | 0.355 | 854.529 | 425.890 | 1714.571 |
| (Glomerular Filtration Rate/100)2 | -2.290 | 0.117 | 0.101 | 0.081 | 0.127 |
| (Cholesterol/10)2 | -2.563 | 0.345 | 0.077 | 0.039 | 0.151 |
| (Cholesterol/10)3 | 2.236 | 0.364 | 9.358 | 4.587 | 19.091 |
| Proteinuria levels with none as reference High | 0.030 | 0.036 | 1.031 | 0.961 | 1.105 |
| Very High | 0.130 | 0.045 | 1.139 | 1.042 | 1.244 |
| Anti-platelets | 0.073 | 0.021 | 1.075 | 1.032 | 1.120 |
| Anticoagulation | 0.268 | 0.033 | 1.307 | 1.225 | 1.395 |
| Angiotensin Blockade | -0.171 | 0.020 | 0.843 | 0.810 | 0.876 |
| Beta blockade | -0.148 | 0.021 | 0.863 | 0.828 | 0.899 |
| Diuretic use | 0.196 | 0.019 | 1.216 | 1.171 | 1.264 |
| Other | -0.140 | 0.038 | 0.869 | 0.807 | 0.936 |
| Lipid | -0.291 | 0.023 | 0.747 | 0.715 | 0.782 |

| | | | | | |
|---|---|---|---|---|---|
| Iron Medication | 0.114 | 0.029 | 1.121 | 1.059 | 1.186 |
| Vitamin D supplementation | 0.218 | 0.027 | 1.244 | 1.180 | 1.310 |

Table 9. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 4

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age/100 | 7.684 | 0.756 | 2174.100 | 494.017 | 9567.902 |
| Age/100 x log(Age/100) | 0.771 | 0.536 | 2.162 | 0.756 | 6.186 |
| Male Gender | 0.493 | 0.020 | 1.637 | 1.574 | 1.702 |
| Race with Caucasian race as reference African – Caribbean | -0.537 | 0.190 | 0.585 | 0.403 | 0.848 |
| Indian Subcontinent | -0.762 | 0.151 | 0.467 | 0.347 | 0.628 |
| South East Asian | -1.253 | 1.000 | 0.286 | 0.040 | 2.030 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.121 | 0.026 | 1.129 | 1.072 | 1.189 |
| $3^{rd}$ | 0.148 | 0.027 | 1.159 | 1.100 | 1.222 |
| $4^{th}$ | 0.167 | 0.027 | 1.181 | 1.119 | 1.247 |
| $5^{th}$ | 0.150 | 0.031 | 1.162 | 1.094 | 1.234 |
| Diabetes Mellitus | 0.178 | 0.023 | 1.195 | 1.142 | 1.251 |
| Heart Failure | 0.468 | 0.024 | 1.596 | 1.522 | 1.674 |
| Atrial Fibrillation | 0.068 | 0.025 | 1.071 | 1.020 | 1.125 |
| Coronary Heart Disease | 0.065 | 0.021 | 1.067 | 1.023 | 1.112 |
| Cerebrovascular Accident | 0.210 | 0.022 | 1.233 | 1.181 | 1.288 |
| Peripheral Vascular Disease | 0.152 | 0.029 | 1.164 | 1.099 | 1.233 |
| Ever Smoked | 0.151 | 0.023 | 1.163 | 1.112 | 1.216 |
| Systolic BP/100 | -3.418 | 0.314 | 0.033 | 0.018 | 0.061 |
| (Systolic BP/100)2 | 0.971 | 0.109 | 2.641 | 2.134 | 3.269 |
| (BMI/10)2 x log (BMI/10) | 0.182 | 0.017 | 1.200 | 1.161 | 1.239 |
| (BMI/10)2 | -0.316 | 0.027 | 0.729 | 0.692 | 0.768 |
| (Haemoglobin/10)-1 | 6.042 | 0.354 | 420.582 | 209.993 | 842.358 |
| (Haemoglobin/10)-2 | -2.118 | 0.176 | 0.120 | 0.085 | 0.170 |
| (Glomerular Filtration Rate/100)2 | -2.331 | 0.117 | 0.097 | 0.077 | 0.122 |
| (Cholesterol/10)2 | -2.495 | 0.346 | 0.083 | 0.042 | 0.162 |
| (Cholesterol/10)3 | 2.154 | 0.365 | 8.616 | 4.215 | 17.613 |
| Proteinuria levels with none as reference High | 0.034 | 0.035 | 1.034 | 0.965 | 1.109 |
| Very High | 0.243 | 0.044 | 1.275 | 1.170 | 1.390 |
| Anti-platelets | 0.074 | 0.021 | 1.077 | 1.033 | 1.122 |
| Anticoagulation | 0.264 | 0.033 | 1.302 | 1.221 | 1.389 |
| Angiotensin Blockade | -0.165 | 0.020 | 0.848 | 0.816 | 0.882 |
| Beta blockade | -0.152 | 0.021 | 0.859 | 0.824 | 0.895 |
| Diuretic use | 0.197 | 0.019 | 1.218 | 1.172 | 1.265 |
| Other | -0.138 | 0.038 | 0.871 | 0.809 | 0.938 |
| Lipid | -0.294 | 0.023 | 0.745 | 0.713 | 0.780 |

| | | | | | |
|---|---|---|---|---|---|
| Iron Medication | 0.113 | 0.029 | 1.120 | 1.058 | 1.185 |
| Vitamin D supplementation | 0.217 | 0.027 | 1.243 | 1.180 | 1.309 |

Table 10. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 5

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age/100 | 7.852 | 0.755 | 2571.771 | 585.354 | 11299.155 |
| Age/100 x log(Age/100) | 0.691 | 0.536 | 1.997 | 0.699 | 5.703 |
| Male Gender | 0.497 | 0.020 | 1.644 | 1.581 | 1.709 |
| Race with Caucasian race as reference African – Caribbean | -0.539 | 0.190 | 0.583 | 0.402 | 0.846 |
| Indian Subcontinent | -0.768 | 0.151 | 0.464 | 0.345 | 0.625 |
| South East Asian | -1.263 | 1.000 | 0.283 | 0.040 | 2.008 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.115 | 0.026 | 1.122 | 1.066 | 1.182 |
| $3^{rd}$ | 0.148 | 0.027 | 1.160 | 1.101 | 1.222 |
| $4^{th}$ | 0.167 | 0.027 | 1.181 | 1.119 | 1.246 |
| $5^{th}$ | 0.153 | 0.031 | 1.166 | 1.097 | 1.238 |
| Diabetes Mellitus | 0.154 | 0.024 | 1.166 | 1.114 | 1.221 |
| Heart Failure | 0.458 | 0.024 | 1.581 | 1.508 | 1.659 |
| Atrial Fibrillation | 0.065 | 0.025 | 1.067 | 1.016 | 1.120 |
| Coronary Heart Disease | 0.064 | 0.021 | 1.066 | 1.023 | 1.111 |
| Cerebrovascular Accident | 0.213 | 0.022 | 1.237 | 1.185 | 1.292 |
| Peripheral Vascular Disease | 0.162 | 0.029 | 1.176 | 1.110 | 1.245 |
| Ever Smoked | 0.150 | 0.023 | 1.162 | 1.111 | 1.215 |
| Systolic BP/100 | -3.387 | 0.314 | 0.034 | 0.018 | 0.063 |
| (Systolic BP/100)2 | 0.961 | 0.109 | 2.615 | 2.113 | 3.235 |
| (BMI/10)2 x log (BMI/10) | 0.179 | 0.016 | 1.196 | 1.158 | 1.235 |
| (BMI/10)2 | -0.306 | 0.027 | 0.737 | 0.699 | 0.776 |
| (Haemoglobin/10)-1 | 4.620 | 0.352 | 101.490 | 50.888 | 202.409 |
| (Haemoglobin/10)-2 | 6.253 | 0.348 | 519.538 | 262.917 | 1026.633 |
| (Glomerular Filtration Rate/100)2 | -2.245 | 0.117 | 0.106 | 0.084 | 0.133 |
| (Cholesterol/10)2 | 2.855 | 0.334 | 17.377 | 9.038 | 33.409 |
| (Cholesterol/10)2*log(Cholesterol/10) | -0.437 | 0.072 | 0.646 | 0.561 | 0.743 |
| Proteinuria levels with none as reference High | -0.017 | 0.035 | 0.983 | 0.918 | 1.053 |
| Very High | 0.218 | 0.043 | 1.243 | 1.143 | 1.352 |
| Anti-platelets | 0.068 | 0.021 | 1.071 | 1.028 | 1.116 |
| Anticoagulation | 0.262 | 0.033 | 1.299 | 1.218 | 1.386 |
| Angiotensin Blockade | -0.166 | 0.020 | 0.847 | 0.815 | 0.881 |
| Beta blockade | -0.142 | 0.021 | 0.868 | 0.832 | 0.904 |
| Diuretic use | 0.194 | 0.019 | 1.214 | 1.169 | 1.261 |
| Other | -0.144 | 0.038 | 0.866 | 0.804 | 0.932 |

| | | | | | |
|---|---|---|---|---|---|
| Lipid | -0.306 | 0.023 | 0.737 | 0.704 | 0.771 |
| Iron Medication | 0.126 | 0.029 | 1.135 | 1.072 | 1.201 |
| Vitamin D supplementation | 0.211 | 0.027 | 1.235 | 1.172 | 1.301 |

Table 11.Cox regression analysis for Imputation 1 without transformation but with frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age | 6.162 | 0.111 | 474.518 | 382.039 | 589.383 |
| Male Gender | 0.485 | 0.020 | 1.624 | 1.562 | 1.688 |
| Race with Caucasian race as reference African –Caribbean | -0.516 | 0.191 | 0.597 | 0.411 | 0.868 |
| Indian Subcontinent | -0.756 | 0.153 | 0.470 | 0.348 | 0.634 |
| South East Asian | -1.197 | 1.000 | 0.302 | 0.043 | 2.147 |
| Townsend Quintile compared to 1th Quintile 2nd | 0.107 | 0.027 | 1.113 | 1.056 | 1.173 |
| 3rd | 0.152 | 0.027 | 1.164 | 1.103 | 1.228 |
| 4th | 0.174 | 0.028 | 1.190 | 1.126 | 1.257 |
| 5th | 0.154 | 0.032 | 1.166 | 1.095 | 1.242 |
| Diabetes Mellitus | 0.157 | 0.024 | 1.170 | 1.118 | 1.226 |
| Heart Failure | 0.489 | 0.024 | 1.630 | 1.554 | 1.710 |
| Atrial Fibrillation | 0.067 | 0.025 | 1.069 | 1.018 | 1.123 |
| Coronary Heart Disease | 0.063 | 0.022 | 1.065 | 1.021 | 1.111 |
| Cerebrovascular Accident | 0.218 | 0.022 | 1.243 | 1.190 | 1.299 |
| Peripheral Vascular Disease | 0.156 | 0.030 | 1.169 | 1.103 | 1.239 |
| Ever Smoked | 0.144 | 0.024 | 1.154 | 1.101 | 1.210 |
| Systolic BP | -0.655 | 0.045 | 0.519 | 0.476 | 0.567 |
| BMI | -0.213 | 0.019 | 0.808 | 0.778 | 0.838 |
| Haemoglobin | -1.671 | 0.056 | 0.188 | 0.169 | 0.210 |
| Glomerular Filtration Rate | -2.152 | 0.102 | 0.116 | 0.095 | 0.142 |
| Cholesterol | -0.674 | 0.085 | 0.510 | 0.432 | 0.602 |
| Proteinuria levels with none as reference High | -0.048 | 0.034 | 0.953 | 0.892 | 1.018 |
| Very High | 0.103 | 0.043 | 1.109 | 1.020 | 1.206 |
| Anti-platelets | 0.079 | 0.021 | 1.082 | 1.038 | 1.128 |
| Anticoagulation | 0.258 | 0.033 | 1.295 | 1.213 | 1.381 |
| Angiotensin Blockade | -0.164 | 0.020 | 0.849 | 0.817 | 0.883 |
| Beta blockade | -0.149 | 0.021 | 0.861 | 0.826 | 0.898 |
| Calcium Channel Blocker | -0.050 | 0.022 | 0.952 | 0.911 | 0.994 |
| Diuretic use | 0.206 | 0.020 | 1.229 | 1.182 | 1.276 |
| Other | -0.139 | 0.038 | 0.870 | 0.808 | 0.937 |
| Lipid | -0.309 | 0.023 | 0.734 | 0.702 | 0.768 |

| | | | | | |
|---|---|---|---|---|---|
| Iron Medication | 0.128 | 0.029 | 1.137 | 1.074 | 1.203 |
| Vitamin D supplementation | 0.219 | 0.027 | 1.244 | 1.181 | 1.311 |

Table 12. Cox regression analysis for Imputation 2 without transformation but with frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age | 6.236 | 0.111 | 510.811 | 411.340 | 634.337 |
| Male Gender | 0.475 | 0.020 | 1.608 | 1.546 | 1.671 |
| Race with Caucasian race as reference African – Caribbean | -0.498 | 0.191 | 0.608 | 0.418 | 0.884 |
| Indian Subcontinent | -0.764 | 0.153 | 0.466 | 0.345 | 0.629 |
| South East Asian | -1.186 | 1.000 | 0.306 | 0.043 | 2.171 |
| Townsend Quintile compared to 1th Quintile 2$^{nd}$ | 0.119 | 0.027 | 1.126 | 1.069 | 1.187 |
| 3$^{rd}$ | 0.158 | 0.027 | 1.171 | 1.110 | 1.235 |
| 4$^{th}$ | 0.174 | 0.028 | 1.189 | 1.126 | 1.257 |
| 5$^{th}$ | 0.152 | 0.032 | 1.165 | 1.094 | 1.240 |
| Diabetes Mellitus | 0.180 | 0.023 | 1.197 | 1.143 | 1.253 |
| Heart Failure | 0.480 | 0.024 | 1.616 | 1.541 | 1.696 |
| Atrial Fibrillation | 0.065 | 0.025 | 1.067 | 1.016 | 1.121 |
| Coronary Heart Disease | 0.060 | 0.022 | 1.061 | 1.018 | 1.107 |
| Cerebrovascular Accident | 0.214 | 0.022 | 1.238 | 1.185 | 1.293 |
| Peripheral Vascular Disease | 0.149 | 0.030 | 1.161 | 1.096 | 1.230 |
| Ever Smoked | 0.146 | 0.024 | 1.157 | 1.104 | 1.213 |
| Systolic BP | -0.677 | 0.045 | 0.508 | 0.465 | 0.555 |
| BMI | -0.194 | 0.019 | 0.824 | 0.793 | 0.855 |
| Haemoglobin | -1.603 | 0.056 | 0.201 | 0.181 | 0.225 |
| Glomerular Filtration Rate | -2.100 | 0.102 | 0.123 | 0.100 | 0.150 |
| Cholesterol | -0.785 | 0.085 | 0.456 | 0.387 | 0.539 |
| Proteinuria levels with none as reference High | 0.072 | 0.035 | 1.074 | 1.002 | 1.152 |
| Very High | 0.337 | 0.044 | 1.400 | 1.285 | 1.526 |
| Anti-platelets | 0.070 | 0.021 | 1.073 | 1.029 | 1.118 |
| Anticoagulation | 0.255 | 0.033 | 1.290 | 1.209 | 1.376 |
| Angiotensin Blockade | -0.171 | 0.020 | 0.843 | 0.811 | 0.877 |
| Beta blockade | -0.150 | 0.021 | 0.861 | 0.826 | 0.897 |
| Calcium Channel Blocker | -0.049 | 0.022 | 0.952 | 0.912 | 0.994 |
| Diuretic use | 0.202 | 0.020 | 1.224 | 1.178 | 1.272 |
| Other | -0.141 | 0.038 | 0.869 | 0.806 | 0.936 |
| Lipid | -0.308 | 0.023 | 0.735 | 0.703 | 0.768 |
| Iron Medication | 0.133 | 0.029 | 1.142 | 1.079 | 1.208 |
| Vitamin D supplementation | 0.226 | 0.027 | 1.254 | 1.190 | 1.321 |

Table 13. Cox regression analysis for Imputation 3 without transformation but with frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age | 6.257 | 0.111 | 521.495 | 419.779 | 647.858 |
| Male Gender | 0.490 | 0.020 | 1.633 | 1.570 | 1.697 |
| Race with Caucasian race as reference African – Caribbean | -0.508 | 0.191 | 0.602 | 0.414 | 0.875 |
| Indian Subcontinent | -0.750 | 0.153 | 0.473 | 0.350 | 0.638 |
| South East Asian | -1.188 | 1.001 | 0.305 | 0.043 | 2.166 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.113 | 0.027 | 1.119 | 1.062 | 1.179 |
| $3^{rd}$ | 0.152 | 0.027 | 1.164 | 1.104 | 1.228 |
| $4^{th}$ | 0.169 | 0.028 | 1.184 | 1.121 | 1.251 |
| $5^{th}$ | 0.163 | 0.032 | 1.177 | 1.105 | 1.253 |
| Diabetes Mellitus | 0.165 | 0.024 | 1.179 | 1.126 | 1.235 |
| Heart Failure | 0.484 | 0.024 | 1.622 | 1.546 | 1.701 |
| Atrial Fibrillation | 0.066 | 0.025 | 1.068 | 1.017 | 1.122 |
| Coronary Heart Disease | 0.058 | 0.022 | 1.060 | 1.016 | 1.106 |
| Cerebrovascular Accident | 0.224 | 0.022 | 1.251 | 1.197 | 1.306 |
| Peripheral Vascular Disease | 0.156 | 0.030 | 1.169 | 1.103 | 1.239 |
| Ever Smoked | 0.149 | 0.024 | 1.160 | 1.107 | 1.216 |
| Systolic BP | -0.687 | 0.045 | 0.503 | 0.461 | 0.549 |
| BMI | -0.158 | 0.019 | 0.854 | 0.823 | 0.886 |
| Haemoglobin | -1.689 | 0.056 | 0.185 | 0.166 | 0.206 |
| Glomerular Filtration Rate | -2.165 | 0.102 | 0.115 | 0.094 | 0.140 |
| Cholesterol | -0.668 | 0.084 | 0.513 | 0.435 | 0.605 |
| Proteinuria levels with none as reference High | 0.028 | 0.036 | 1.029 | 0.959 | 1.103 |
| Very High | 0.138 | 0.045 | 1.148 | 1.050 | 1.255 |
| Anti-platelets | 0.077 | 0.021 | 1.080 | 1.036 | 1.126 |
| Anticoagulation | 0.263 | 0.033 | 1.301 | 1.219 | 1.388 |
| Angiotensin Blockade | -0.172 | 0.020 | 0.842 | 0.810 | 0.875 |
| Beta blockade | -0.151 | 0.021 | 0.860 | 0.825 | 0.896 |
| Calcium Channel Blocker | -0.040 | 0.022 | 0.961 | 0.920 | 1.003 |
| Diuretic use | 0.196 | 0.020 | 1.217 | 1.171 | 1.264 |
| Other | -0.136 | 0.038 | 0.873 | 0.810 | 0.940 |
| Lipid | -0.307 | 0.023 | 0.736 | 0.703 | 0.769 |
| Iron Medication | 0.129 | 0.029 | 1.138 | 1.075 | 1.204 |
| Vitamin D supplementation | 0.225 | 0.027 | 1.253 | 1.189 | 1.320 |

Table 14. Cox regression analysis for Imputation 4 without transformation but with frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age | 6.248 | 0.111 | 516.823 | 416.181 | 641.802 |
| Male Gender | 0.478 | 0.020 | 1.612 | 1.551 | 1.676 |
| Race with Caucasian race as reference African – Caribbean | -0.488 | 0.191 | 0.614 | 0.422 | 0.893 |
| Indian Subcontinent | -0.739 | 0.153 | 0.478 | 0.354 | 0.645 |
| South East Asian | -1.246 | 1.001 | 0.288 | 0.040 | 2.044 |
| Townsend Quintile compared to 1th Quintile 2nd | 0.121 | 0.027 | 1.128 | 1.071 | 1.189 |
| 3rd | 0.145 | 0.027 | 1.156 | 1.096 | 1.220 |
| 4th | 0.164 | 0.028 | 1.178 | 1.115 | 1.245 |
| 5th | 0.154 | 0.032 | 1.166 | 1.096 | 1.242 |
| Diabetes Mellitus | 0.176 | 0.024 | 1.192 | 1.138 | 1.248 |
| Heart Failure | 0.487 | 0.024 | 1.628 | 1.552 | 1.707 |
| Atrial Fibrillation | 0.064 | 0.025 | 1.066 | 1.015 | 1.120 |
| Coronary Heart Disease | 0.064 | 0.022 | 1.066 | 1.022 | 1.112 |
| Cerebrovascular Accident | 0.216 | 0.022 | 1.241 | 1.188 | 1.297 |
| Peripheral Vascular Disease | 0.150 | 0.030 | 1.161 | 1.096 | 1.231 |
| Ever Smoked | 0.143 | 0.024 | 1.154 | 1.101 | 1.209 |
| Systolic BP | -0.686 | 0.045 | 0.504 | 0.461 | 0.550 |
| BMI | -0.187 | 0.019 | 0.830 | 0.800 | 0.861 |
| Haemoglobin | -1.597 | 0.056 | 0.202 | 0.181 | 0.226 |
| Glomerular Filtration Rate | -2.188 | 0.101 | 0.112 | 0.092 | 0.137 |
| Cholesterol | -0.682 | 0.085 | 0.506 | 0.428 | 0.597 |
| Proteinuria levels with none as reference High | 0.035 | 0.036 | 1.035 | 0.966 | 1.110 |
| Very High | 0.259 | 0.044 | 1.296 | 1.188 | 1.413 |
| Anti-platelets | 0.079 | 0.021 | 1.082 | 1.038 | 1.127 |
| Anticoagulation | 0.258 | 0.033 | 1.295 | 1.214 | 1.382 |
| Angiotensin Blockade | -0.167 | 0.020 | 0.846 | 0.814 | 0.879 |
| Beta blockade | -0.155 | 0.021 | 0.856 | 0.821 | 0.892 |
| Calcium Channel Blocker | -0.047 | 0.022 | 0.954 | 0.914 | 0.997 |
| Diuretic use | 0.198 | 0.020 | 1.219 | 1.173 | 1.266 |
| Other | -0.133 | 0.038 | 0.876 | 0.813 | 0.943 |
| Lipid | -0.311 | 0.023 | 0.733 | 0.701 | 0.766 |
| Iron Medication | 0.136 | 0.029 | 1.145 | 1.082 | 1.212 |
| Vitamin D supplementation | 0.224 | 0.027 | 1.250 | 1.187 | 1.318 |

Table 15. Cox regression analysis for Imputation 5 without transformation but with frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age | 6.301 | 0.111 | 545.008 | 438.791 | 676.936 |
| Male Gender | 0.480 | 0.020 | 1.617 | 1.555 | 1.681 |
| Race with Caucasian race as reference African – Caribbean | -0.486 | 0.191 | 0.615 | 0.423 | 0.894 |
| Indian Subcontinent | -0.741 | 0.153 | 0.476 | 0.353 | 0.643 |
| South East Asian | -1.221 | 1.001 | 0.295 | 0.042 | 2.097 |
| Townsend Quintile compared to 1th Quintile 2nd | 0.112 | 0.027 | 1.119 | 1.062 | 1.179 |
| 3rd | 0.147 | 0.027 | 1.158 | 1.098 | 1.222 |
| 4th | 0.166 | 0.028 | 1.180 | 1.117 | 1.247 |
| 5th | 0.154 | 0.032 | 1.167 | 1.096 | 1.242 |
| Diabetes Mellitus | 0.157 | 0.024 | 1.170 | 1.117 | 1.225 |
| Heart Failure | 0.481 | 0.024 | 1.617 | 1.542 | 1.696 |
| Atrial Fibrillation | 0.060 | 0.025 | 1.062 | 1.011 | 1.116 |
| Coronary Heart Disease | 0.065 | 0.022 | 1.067 | 1.023 | 1.113 |
| Cerebrovascular Accident | 0.219 | 0.022 | 1.244 | 1.191 | 1.300 |
| Peripheral Vascular Disease | 0.159 | 0.030 | 1.172 | 1.106 | 1.242 |
| Ever Smoked | 0.142 | 0.024 | 1.152 | 1.099 | 1.208 |
| Systolic BP | -0.679 | 0.045 | 0.507 | 0.464 | 0.554 |
| BMI | -0.149 | 0.019 | 0.861 | 0.830 | 0.894 |
| Haemoglobin | -1.582 | 0.056 | 0.206 | 0.184 | 0.229 |
| Glomerular Filtration Rate | -2.120 | 0.102 | 0.120 | 0.098 | 0.147 |
| Cholesterol | -0.728 | 0.085 | 0.483 | 0.409 | 0.570 |
| Proteinuria levels with none as reference High | -0.015 | 0.035 | 0.985 | 0.919 | 1.055 |
| Very High | 0.228 | 0.043 | 1.257 | 1.155 | 1.367 |
| Anti-platelets | 0.072 | 0.021 | 1.075 | 1.031 | 1.120 |
| Anticoagulation | 0.257 | 0.033 | 1.293 | 1.212 | 1.380 |
| Angiotensin Blockade | -0.168 | 0.020 | 0.845 | 0.813 | 0.879 |
| Beta blockade | -0.145 | 0.021 | 0.865 | 0.830 | 0.902 |
| Calcium Channel Blocker | -0.046 | 0.022 | 0.955 | 0.914 | 0.997 |
| Diuretic use | 0.196 | 0.020 | 1.217 | 1.171 | 1.264 |
| Other | -0.140 | 0.038 | 0.870 | 0.807 | 0.937 |
| Lipid | -0.316 | 0.023 | 0.729 | 0.697 | 0.762 |
| Iron Medication | 0.141 | 0.029 | 1.151 | 1.088 | 1.219 |
| Vitamin D supplementation | 0.219 | 0.027 | 1.245 | 1.181 | 1.311 |

Table 16. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 1 with frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% CI Lower Limit | 95% CI Lower Limit |
|---|---|---|---|---|---|
| Age/100 | 7.755 | 0.759 | 2334.143 | 526.851 | 10341.113 |
| Age/100 x log(Age/100) | 0.625 | 0.538 | 1.867 | 0.651 | 5.359 |
| Male Gender | 0.503 | 0.020 | 1.654 | 1.591 | 1.720 |
| Race with Caucasian race as reference African – Caribbean | -0.567 | 0.191 | 0.567 | 0.390 | 0.825 |
| Indian Subcontinent | -0.765 | 0.153 | 0.466 | 0.345 | 0.628 |
| South East Asian | -1.248 | 1.000 | 0.287 | 0.040 | 2.039 |
| Townsend Quintile compared to 1th Quintile 2nd | 0.107 | 0.027 | 1.113 | 1.056 | 1.172 |
| 3rd | 0.147 | 0.027 | 1.158 | 1.098 | 1.221 |
| 4th | 0.169 | 0.028 | 1.184 | 1.120 | 1.251 |
| 5th | 0.147 | 0.032 | 1.159 | 1.088 | 1.234 |
| Diabetes Mellitus | 0.157 | 0.024 | 1.169 | 1.117 | 1.225 |
| Heart Failure | 0.472 | 0.024 | 1.604 | 1.529 | 1.682 |
| Atrial Fibrillation | 0.066 | 0.025 | 1.069 | 1.017 | 1.122 |
| Coronary Heart Disease | 0.064 | 0.021 | 1.066 | 1.023 | 1.112 |
| Cerebrovascular Accident | 0.209 | 0.022 | 1.232 | 1.180 | 1.287 |
| Peripheral Vascular Disease | 0.155 | 0.030 | 1.167 | 1.102 | 1.237 |
| Ever Smoked | 0.149 | 0.024 | 1.161 | 1.107 | 1.217 |
| Systolic BP/100 | -3.318 | 0.313 | 0.036 | 0.020 | 0.067 |
| (Systolic BP/100)2 | 0.947 | 0.108 | 2.577 | 2.084 | 3.187 |
| BMI/10 | -0.649 | 0.050 | 0.522 | 0.473 | 0.577 |
| (BMI/10)3 | 0.018 | 0.002 | 1.018 | 1.014 | 1.022 |
| (Haemoglobin/10)-1 x log(Heaemoglobin/10) | 5.006 | 0.363 | 149.247 | 73.236 | 304.147 |
| (Haemoglobin/10)-1 | 6.725 | 0.357 | 832.889 | 413.870 | 1676.138 |
| (Glomerular Filtration Rate/100)2 | -2.329 | 0.118 | 0.097 | 0.077 | 0.123 |
| (Cholesterol/10)2 | -2.520 | 0.354 | 0.080 | 0.040 | 0.161 |
| (Cholesterol/10)3 | 2.203 | 0.377 | 9.049 | 4.320 | 18.958 |
| Proteinuria levels with none as reference High | -0.043 | 0.034 | 0.958 | 0.897 | 1.023 |
| Very High | 0.101 | 0.043 | 1.106 | 1.017 | 1.203 |
| Anti-platelets | 0.075 | 0.021 | 1.077 | 1.034 | 1.123 |
| Anticoagulation | 0.264 | 0.033 | 1.302 | 1.220 | 1.389 |
| Angiotensin Blockade | -0.161 | 0.020 | 0.852 | 0.819 | 0.886 |
| Beta blockade | -0.146 | 0.021 | 0.865 | 0.829 | 0.901 |
| Diuretic use | 0.203 | 0.020 | 1.224 | 1.178 | 1.272 |
| Other | -0.147 | 0.038 | 0.863 | 0.801 | 0.930 |
| Lipid | -0.289 | 0.023 | 0.749 | 0.716 | 0.784 |
| Iron Medication | 0.114 | 0.029 | 1.121 | 1.059 | 1.187 |
| Vitamin D supplementation | 0.209 | 0.027 | 1.233 | 1.170 | 1.299 |

Table 17. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 2 with frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age/100 | 7.517 | 0.761 | 1839.226 | 413.784 | 8175.164 |
| Age/100 x log(Age/100) | 0.910 | 0.540 | 2.484 | 0.863 | 7.151 |
| Male Gender | 0.490 | 0.020 | 1.633 | 1.570 | 1.698 |
| Race with Caucasian race as reference African – Caribbean | -0.544 | 0.191 | 0.580 | 0.399 | 0.844 |
| Indian Subcontinent | -0.774 | 0.153 | 0.461 | 0.342 | 0.623 |
| South East Asian | -1.263 | 1.000 | 0.283 | 0.040 | 2.010 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.118 | 0.027 | 1.126 | 1.068 | 1.186 |
| $3^{rd}$ | 0.154 | 0.027 | 1.167 | 1.106 | 1.231 |
| $4^{th}$ | 0.170 | 0.028 | 1.185 | 1.121 | 1.252 |
| $5^{th}$ | 0.144 | 0.032 | 1.155 | 1.084 | 1.230 |
| Diabetes Mellitus | 0.175 | 0.024 | 1.191 | 1.138 | 1.248 |
| Heart Failure | 0.459 | 0.024 | 1.582 | 1.508 | 1.660 |
| Atrial Fibrillation | 0.063 | 0.025 | 1.065 | 1.014 | 1.119 |
| Coronary Heart Disease | 0.060 | 0.022 | 1.061 | 1.018 | 1.107 |
| Cerebrovascular Accident | 0.205 | 0.022 | 1.227 | 1.175 | 1.282 |
| Peripheral Vascular Disease | 0.151 | 0.030 | 1.163 | 1.097 | 1.232 |
| Ever Smoked | 0.151 | 0.024 | 1.163 | 1.110 | 1.219 |
| Systolic BP/100 | -3.368 | 0.313 | 0.034 | 0.019 | 0.064 |
| (Systolic BP/100)2 | 0.957 | 0.109 | 2.603 | 2.104 | 3.220 |
| BMI/10 | -1.154 | 0.100 | 0.315 | 0.259 | 0.384 |
| (BMI/10)2 | 0.171 | 0.017 | 1.187 | 1.147 | 1.227 |
| (Haemoglobin/10)-1 x log(Heaemoglobin/10) | 4.756 | 0.363 | 116.222 | 57.044 | 236.791 |
| (Haemoglobin/10)-1 | 6.389 | 0.356 | 594.963 | 296.171 | 1195.194 |
| (Glomerular Filtration Rate/100) | -1.965 | 0.103 | 0.140 | 0.115 | 0.171 |
| (Cholesterol/10)2 | -3.326 | 0.328 | 0.036 | 0.019 | 0.068 |
| (Cholesterol/10)3 | 3.000 | 0.341 | 20.075 | 10.282 | 39.199 |
| Proteinuria levels with none as reference High | 0.074 | 0.035 | 1.077 | 1.004 | 1.154 |
| Very High | 0.325 | 0.044 | 1.384 | 1.271 | 1.509 |
| Anti-platelets | 0.063 | 0.021 | 1.064 | 1.021 | 1.109 |
| Anticoagulation | 0.267 | 0.033 | 1.306 | 1.224 | 1.394 |
| Angiotensin Blockade | -0.168 | 0.020 | 0.846 | 0.813 | 0.879 |
| Beta blockade | -0.146 | 0.021 | 0.865 | 0.829 | 0.901 |
| Diuretic use | 0.199 | 0.020 | 1.221 | 1.175 | 1.268 |
| Other | -0.148 | 0.038 | 0.863 | 0.801 | 0.929 |
| Lipid | -0.295 | 0.023 | 0.745 | 0.712 | 0.779 |
| Iron Medication | 0.121 | 0.029 | 1.129 | 1.066 | 1.195 |
| Vitamin D supplementation | 0.218 | 0.027 | 1.243 | 1.180 | 1.310 |

Table 18. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 3 with frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age/100 | 7.831 | 0.758 | 2517.446 | 569.820 | 11121.995 |
| Age/100 x log(Age/100) | 0.676 | 0.537 | 1.966 | 0.686 | 5.630 |
| Male Gender | 0.509 | 0.020 | 1.663 | 1.599 | 1.729 |
| Race with Caucasian race as reference African – Caribbean | -0.561 | 0.191 | 0.571 | 0.393 | 0.830 |
| Indian Subcontinent | -0.760 | 0.153 | 0.468 | 0.346 | 0.631 |
| South East Asian | -1.239 | 1.001 | 0.290 | 0.041 | 2.059 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.109 | 0.027 | 1.115 | 1.058 | 1.175 |
| $3^{rd}$ | 0.146 | 0.027 | 1.157 | 1.097 | 1.220 |
| $4^{th}$ | 0.161 | 0.028 | 1.175 | 1.112 | 1.241 |
| $5^{th}$ | 0.152 | 0.032 | 1.164 | 1.093 | 1.239 |
| Diabetes Mellitus | 0.164 | 0.024 | 1.179 | 1.125 | 1.235 |
| Heart Failure | 0.465 | 0.024 | 1.591 | 1.517 | 1.669 |
| Atrial Fibrillation | 0.065 | 0.025 | 1.067 | 1.016 | 1.121 |
| Coronary Heart Disease | 0.059 | 0.022 | 1.061 | 1.017 | 1.107 |
| Cerebrovascular Accident | 0.214 | 0.022 | 1.239 | 1.186 | 1.294 |
| Peripheral Vascular Disease | 0.156 | 0.030 | 1.168 | 1.103 | 1.238 |
| Ever Smoked | 0.155 | 0.024 | 1.167 | 1.114 | 1.223 |
| Systolic BP/100 | -3.568 | 0.310 | 0.028 | 0.015 | 0.052 |
| (Systolic BP/100)2 | 1.023 | 0.108 | 2.780 | 2.252 | 3.432 |
| (BMI/10)2 x log (BMI/10) | 0.176 | 0.017 | 1.192 | 1.154 | 1.232 |
| (BMI/10)2 | -0.301 | 0.027 | 0.740 | 0.702 | 0.780 |
| (Haemoglobin/10)-1 x log(Heaemoglobin/10) | 5.002 | 0.363 | 148.725 | 73.026 | 302.894 |
| (Haemoglobin/10)-1 | 6.733 | 0.356 | 839.914 | 417.861 | 1688.257 |
| (Glomerular Filtration Rate/100)2 | -2.320 | 0.118 | 0.098 | 0.078 | 0.124 |
| (Cholesterol/10)2 | -2.534 | 0.346 | 0.079 | 0.040 | 0.156 |
| (Cholesterol/10)3 | 2.203 | 0.365 | 9.054 | 4.426 | 18.523 |
| Proteinuria levels with none as reference High | 0.029 | 0.036 | 1.030 | 0.960 | 1.104 |
| Very High | 0.137 | 0.045 | 1.147 | 1.050 | 1.254 |
| Anti-platelets | 0.072 | 0.021 | 1.075 | 1.032 | 1.120 |
| Anticoagulation | 0.271 | 0.033 | 1.312 | 1.229 | 1.400 |
| Angiotensin Blockade | -0.171 | 0.020 | 0.843 | 0.811 | 0.877 |
| Beta blockade | -0.147 | 0.021 | 0.863 | 0.828 | 0.900 |
| Diuretic use | 0.193 | 0.020 | 1.213 | 1.167 | 1.260 |
| Other | -0.139 | 0.038 | 0.870 | 0.808 | 0.937 |
| Lipid | -0.288 | 0.023 | 0.749 | 0.716 | 0.784 |
| Iron Medication | 0.113 | 0.029 | 1.120 | 1.058 | 1.186 |
| Vitamin D supplementation | 0.219 | 0.027 | 1.245 | 1.181 | 1.312 |

Table 19. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 4 with frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age/100 | 7.681 | 0.758 | 2166.569 | 490.207 | 9575.586 |
| Age/100 x log(Age/100) | 0.783 | 0.538 | 2.188 | 0.763 | 6.275 |
| Male Gender | 0.493 | 0.020 | 1.637 | 1.574 | 1.702 |
| Race with Caucasian race as reference African – Caribbean | -0.537 | 0.191 | 0.584 | 0.402 | 0.850 |
| Indian Subcontinent | -0.748 | 0.153 | 0.473 | 0.351 | 0.639 |
| South East Asian | -1.269 | 1.001 | 0.281 | 0.040 | 1.998 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.117 | 0.027 | 1.124 | 1.067 | 1.185 |
| $3^{rd}$ | 0.141 | 0.027 | 1.152 | 1.092 | 1.215 |
| $4^{th}$ | 0.159 | 0.028 | 1.173 | 1.110 | 1.239 |
| $5^{th}$ | 0.145 | 0.032 | 1.156 | 1.086 | 1.231 |
| Diabetes Mellitus | 0.176 | 0.024 | 1.192 | 1.139 | 1.249 |
| Heart Failure | 0.468 | 0.024 | 1.597 | 1.523 | 1.676 |
| Atrial Fibrillation | 0.064 | 0.025 | 1.066 | 1.015 | 1.119 |
| Coronary Heart Disease | 0.065 | 0.022 | 1.067 | 1.023 | 1.113 |
| Cerebrovascular Accident | 0.207 | 0.022 | 1.230 | 1.178 | 1.285 |
| Peripheral Vascular Disease | 0.147 | 0.030 | 1.158 | 1.093 | 1.227 |
| Ever Smoked | 0.148 | 0.024 | 1.160 | 1.106 | 1.215 |
| Systolic BP/100 | -3.354 | 0.315 | 0.035 | 0.019 | 0.065 |
| (Systolic BP/100)2 | 0.948 | 0.109 | 2.581 | 2.083 | 3.197 |
| (BMI/10)2 x log (BMI/10) | 0.183 | 0.017 | 1.201 | 1.162 | 1.241 |
| (BMI/10)2 | -0.318 | 0.027 | 0.728 | 0.690 | 0.767 |
| (Haemoglobin/10)-1 | 6.023 | 0.355 | 412.733 | 206.021 | 826.847 |
| (Haemoglobin/10)-2 | -2.106 | 0.176 | 0.122 | 0.086 | 0.172 |
| (Glomerular Filtration Rate/100)2 | -2.359 | 0.117 | 0.095 | 0.075 | 0.119 |
| (Cholesterol/10)2 | -2.460 | 0.347 | 0.085 | 0.043 | 0.168 |
| (Cholesterol/10)3 | 2.120 | 0.366 | 8.327 | 4.064 | 17.062 |
| Proteinuria levels with none as reference High | 0.036 | 0.036 | 1.036 | 0.966 | 1.111 |
| Very High | 0.253 | 0.044 | 1.288 | 1.181 | 1.404 |
| Anti-platelets | 0.074 | 0.021 | 1.077 | 1.033 | 1.123 |
| Anticoagulation | 0.268 | 0.033 | 1.307 | 1.224 | 1.395 |
| Angiotensin Blockade | -0.164 | 0.020 | 0.848 | 0.816 | 0.882 |
| Beta blockade | -0.151 | 0.021 | 0.860 | 0.825 | 0.896 |
| Diuretic use | 0.195 | 0.020 | 1.215 | 1.169 | 1.262 |
| Other | -0.137 | 0.038 | 0.872 | 0.810 | 0.939 |
| Lipid | -0.291 | 0.023 | 0.748 | 0.715 | 0.783 |
| Iron Medication | 0.111 | 0.029 | 1.118 | 1.056 | 1.183 |
| Vitamin D supplementation | 0.218 | 0.027 | 1.243 | 1.180 | 1.310 |

Table 20. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 5 with frailty term

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals: Lower Limit | 95% Confidence intervals: Lower Limit |
|---|---|---|---|---|---|
| Age/100 | 7.865 | 0.757 | 2605.552 | 590.688 | 11493.213 |
| Age/100 x log(Age/100) | 0.696 | 0.537 | 2.005 | 0.700 | 5.739 |
| Male Gender | 0.496 | 0.020 | 1.643 | 1.580 | 1.708 |
| Race with Caucasian race as reference African – Caribbean | -0.538 | 0.191 | 0.584 | 0.402 | 0.849 |
| Indian Subcontinent | -0.755 | 0.153 | 0.470 | 0.348 | 0.635 |
| South East Asian | -1.271 | 1.001 | 0.281 | 0.039 | 1.994 |
| Townsend Quintile compared to 1th Quintile $2^{nd}$ | 0.111 | 0.027 | 1.118 | 1.061 | 1.178 |
| $3^{rd}$ | 0.142 | 0.027 | 1.153 | 1.093 | 1.216 |
| $4^{th}$ | 0.159 | 0.028 | 1.173 | 1.110 | 1.239 |
| $5^{th}$ | 0.148 | 0.032 | 1.160 | 1.090 | 1.235 |
| Diabetes Mellitus | 0.152 | 0.024 | 1.164 | 1.111 | 1.219 |
| Heart Failure | 0.459 | 0.024 | 1.582 | 1.508 | 1.660 |
| Atrial Fibrillation | 0.059 | 0.025 | 1.061 | 1.010 | 1.115 |
| Coronary Heart Disease | 0.064 | 0.022 | 1.067 | 1.023 | 1.112 |
| Cerebrovascular Accident | 0.210 | 0.022 | 1.234 | 1.181 | 1.289 |
| Peripheral Vascular Disease | 0.157 | 0.030 | 1.170 | 1.104 | 1.240 |
| Ever Smoked | 0.149 | 0.024 | 1.161 | 1.107 | 1.217 |
| Systolic BP/100 | -3.316 | 0.315 | 0.036 | 0.020 | 0.067 |
| (Systolic BP/100)2 | 0.936 | 0.109 | 2.550 | 2.059 | 3.158 |
| (BMI/10)2 x log (BMI/10) | 0.180 | 0.016 | 1.197 | 1.159 | 1.236 |
| (BMI/10)2 | -0.307 | 0.027 | 0.736 | 0.698 | 0.775 |
| (Haemoglobin/10)-1 | 4.617 | 0.353 | 101.180 | 50.654 | 202.104 |
| (Haemoglobin/10)-2 | 6.246 | 0.348 | 515.997 | 260.716 | 1021.235 |
| (Glomerular Filtration Rate/100)2 | -2.273 | 0.118 | 0.103 | 0.082 | 0.130 |
| (Cholesterol/10)2 | -0.442 | 0.072 | 0.643 | 0.558 | 0.740 |
| (Cholesterol/10)2*log(Cholesterol/10) | 2.839 | 0.334 | 17.102 | 8.882 | 32.931 |
| Proteinuria levels with none as reference High | -0.014 | 0.035 | 0.986 | 0.921 | 1.057 |
| Very High | 0.229 | 0.043 | 1.257 | 1.155 | 1.367 |
| Anti-platelets | 0.068 | 0.021 | 1.070 | 1.027 | 1.115 |
| Anticoagulation | 0.265 | 0.033 | 1.303 | 1.221 | 1.390 |
| Angiotensin Blockade | -0.165 | 0.020 | 0.848 | 0.815 | 0.881 |
| Beta blockade | -0.141 | 0.021 | 0.868 | 0.833 | 0.905 |
| Diuretic use | 0.192 | 0.020 | 1.211 | 1.166 | 1.258 |
| Other | -0.144 | 0.038 | 0.866 | 0.804 | 0.933 |
| Lipid | -0.303 | 0.023 | 0.739 | 0.706 | 0.773 |
| Iron Medication | 0.125 | 0.029 | 1.133 | 1.071 | 1.200 |
| Vitamin D supplementation | 0.212 | 0.027 | 1.236 | 1.173 | 1.303 |

Figure 6. Age versus log relative hazards ratio for imputations 1-5 with frailty term

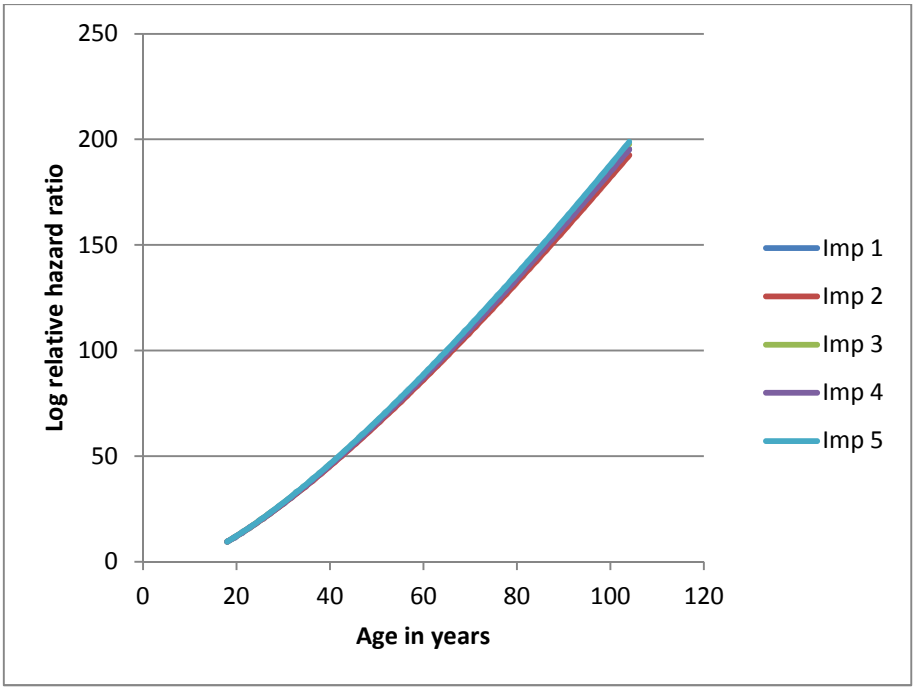



Figure 7. Systolic Blood pressure versus log relative hazards ratio for imputations 1-5 with frailty term

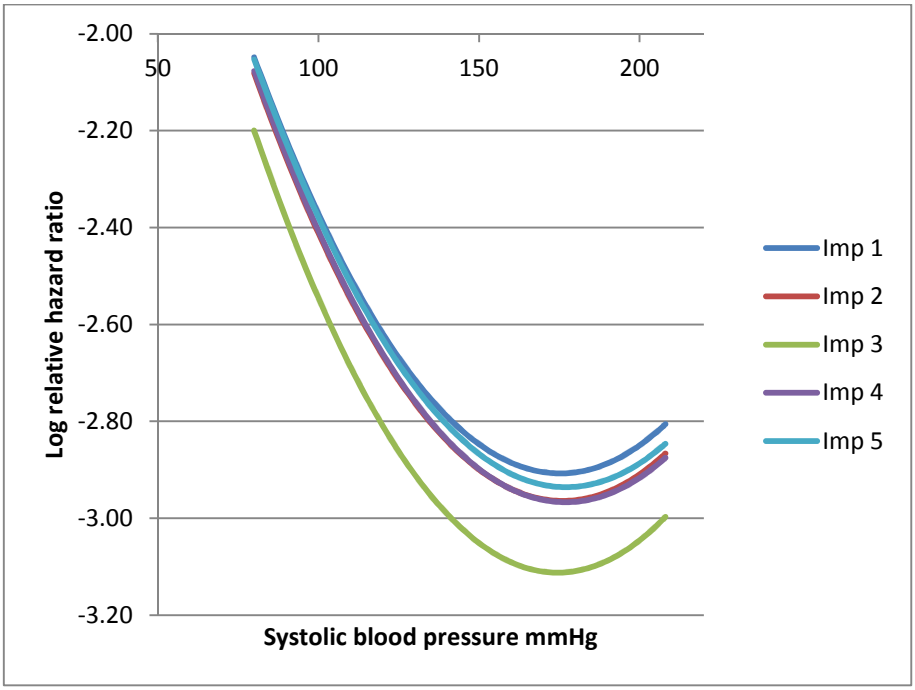


Figure 8. Haemoglobin versus log relative hazards ratio for imputations 1-5 with frailty term

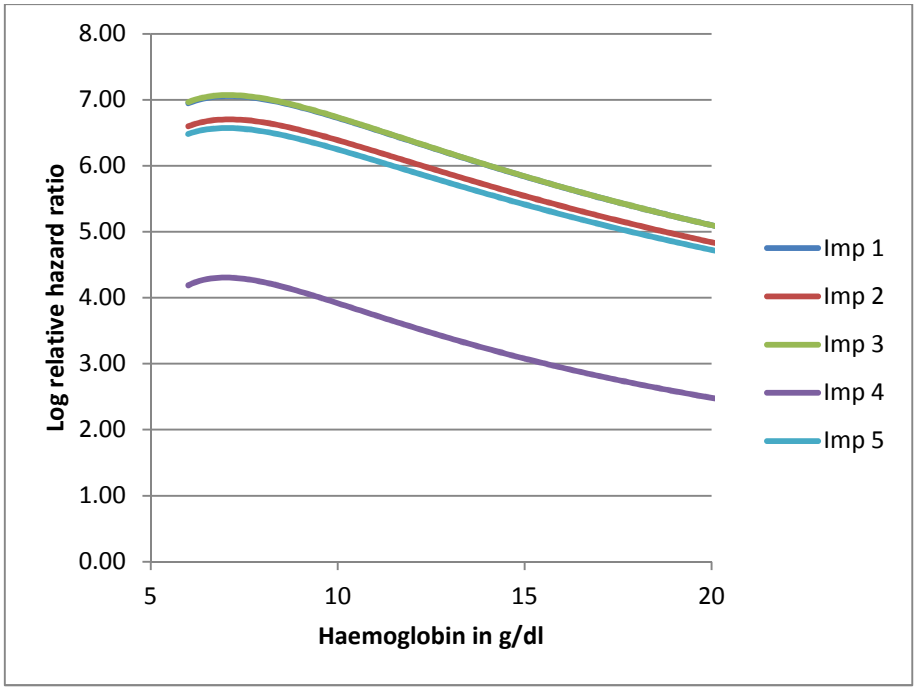



Figure 9. Body Mass Index versus log relative hazards ratio for imputations 1-5 with frailty term

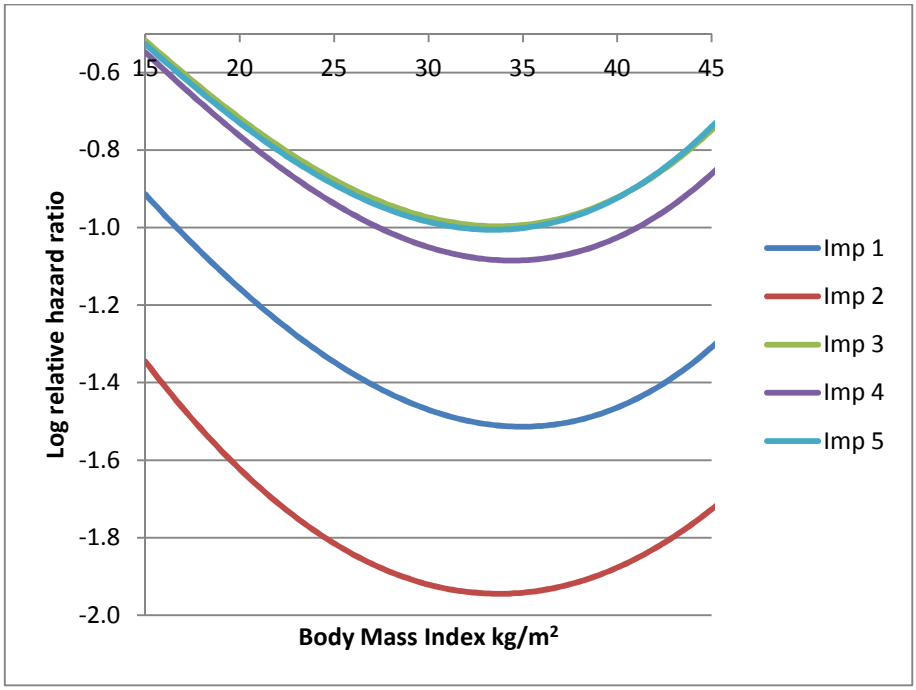
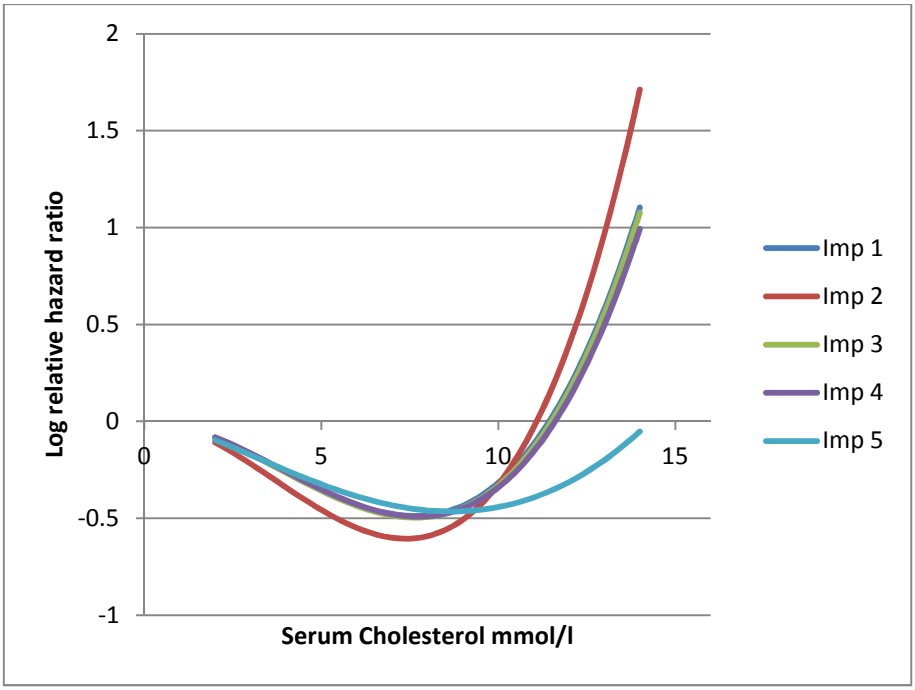


Figure 10. Cholesterol versus log relative hazards ratio for imputations 1-5 with frailty term



# Appendix B. Schoenfeld residuals for Mortality Model

**Figure 1. Scaled Schoenfeld Residuals for Antiplatelet versus Time**

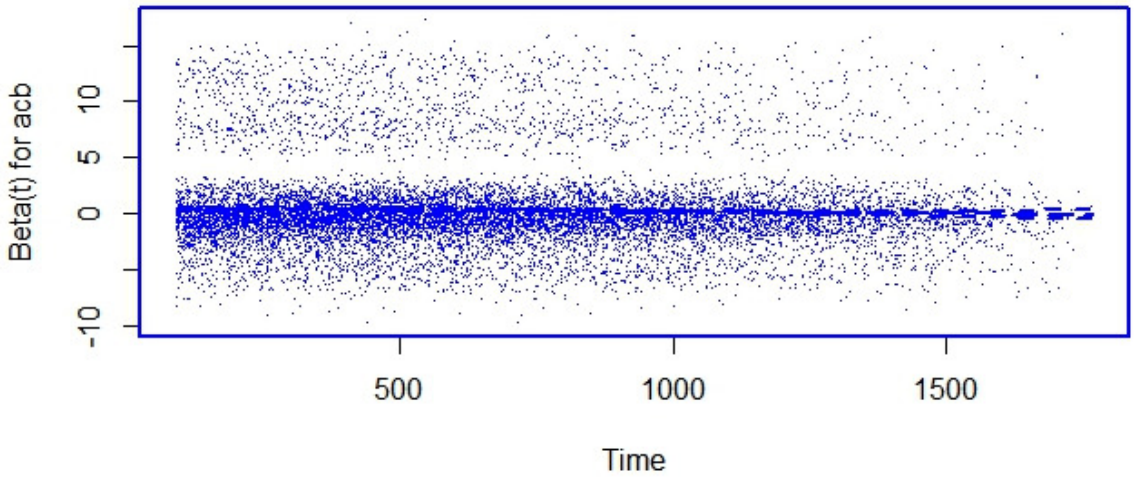



**Figure 2. Scaled Schoenfeld Residuals for Atrial Fibrillation versus Time**

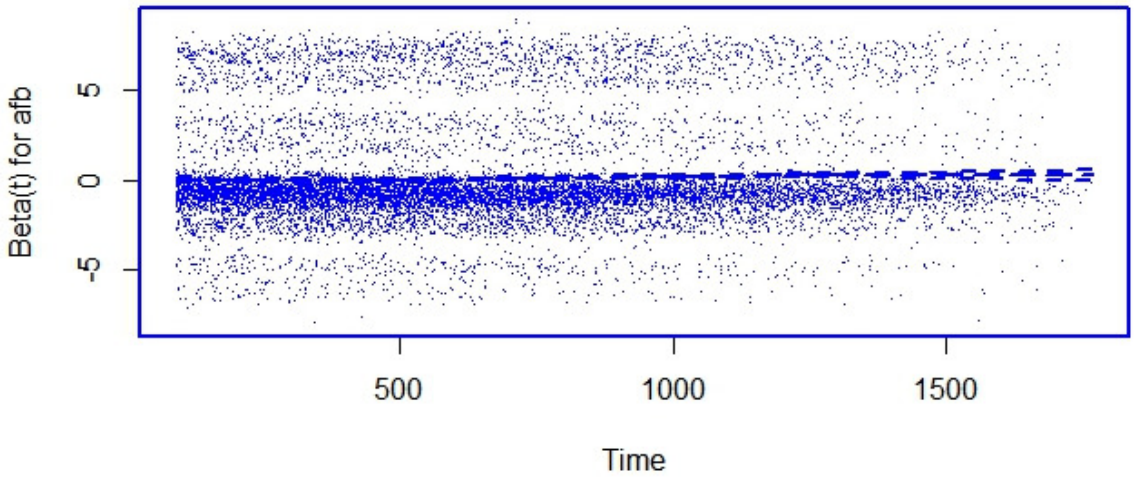


**Figure 3. Scaled Schoenfeld Residuals for AFC race versus Time**

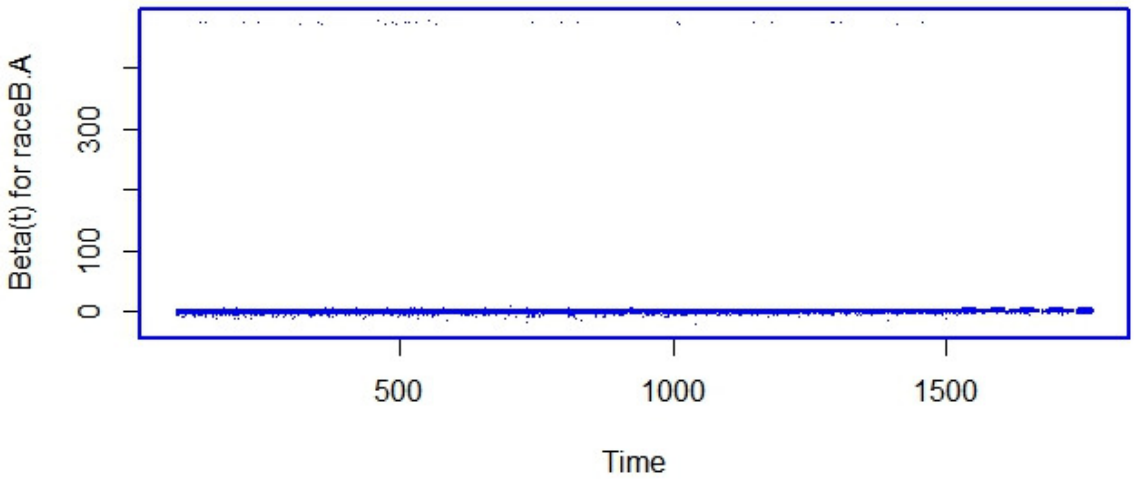



**Figure 4. Scaled Schoenfeld Residuals for Age versus Time**

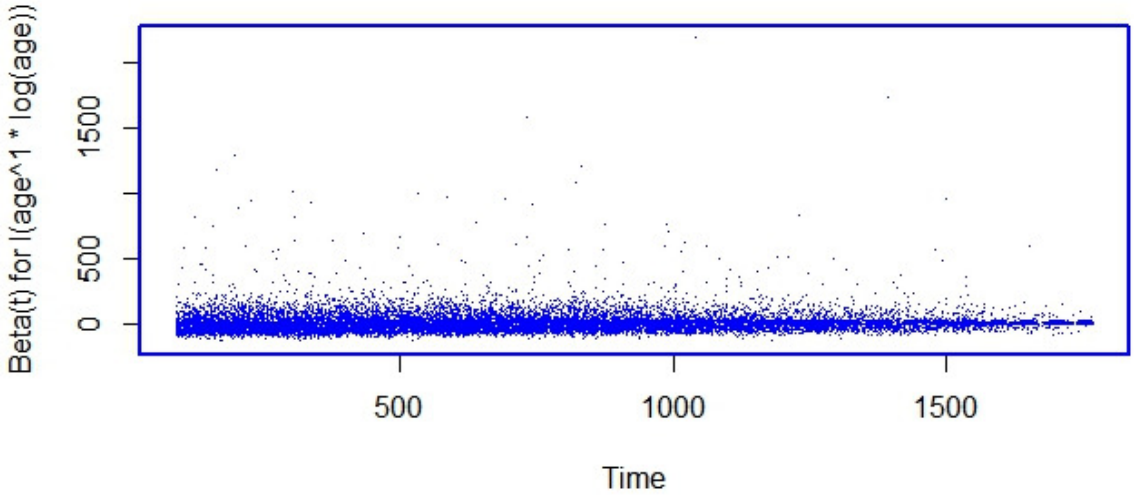


**Figure 5. Scaled Schoenfeld Residuals Age for versus Time**

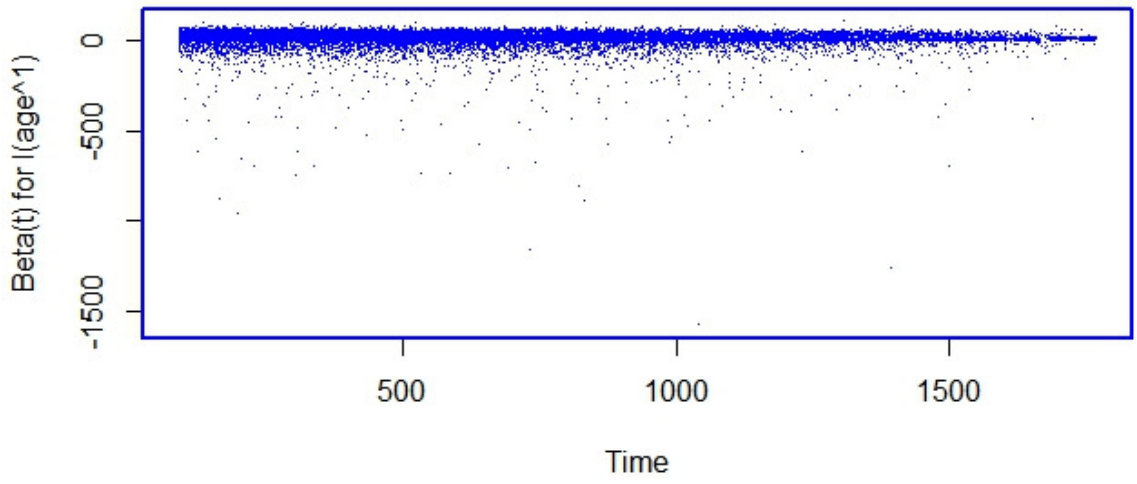



**Figure 6. Scaled Schoenfeld Residuals for betablockade versus Time**

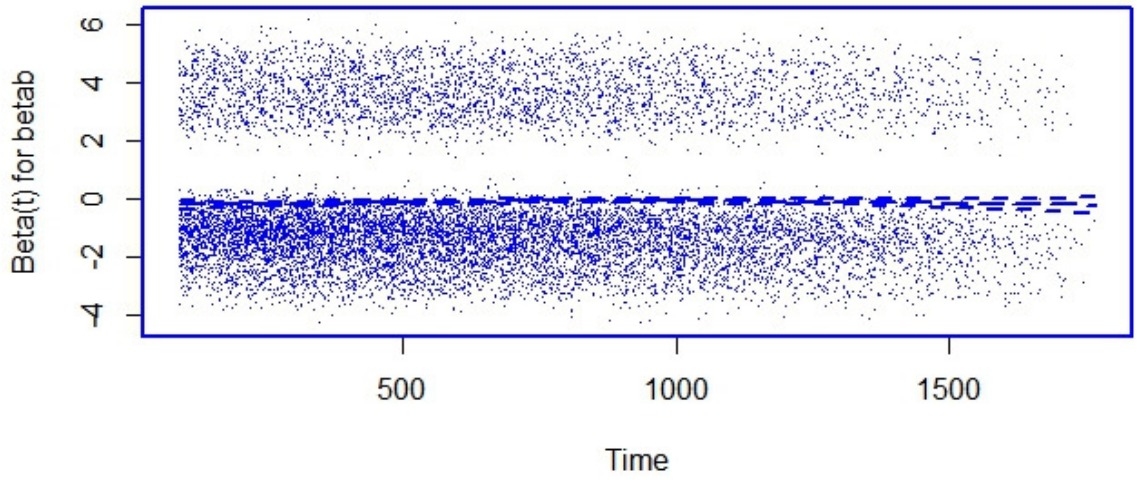


**Figure 7. Scaled Schoenfeld Residuals BMI for versus Time**

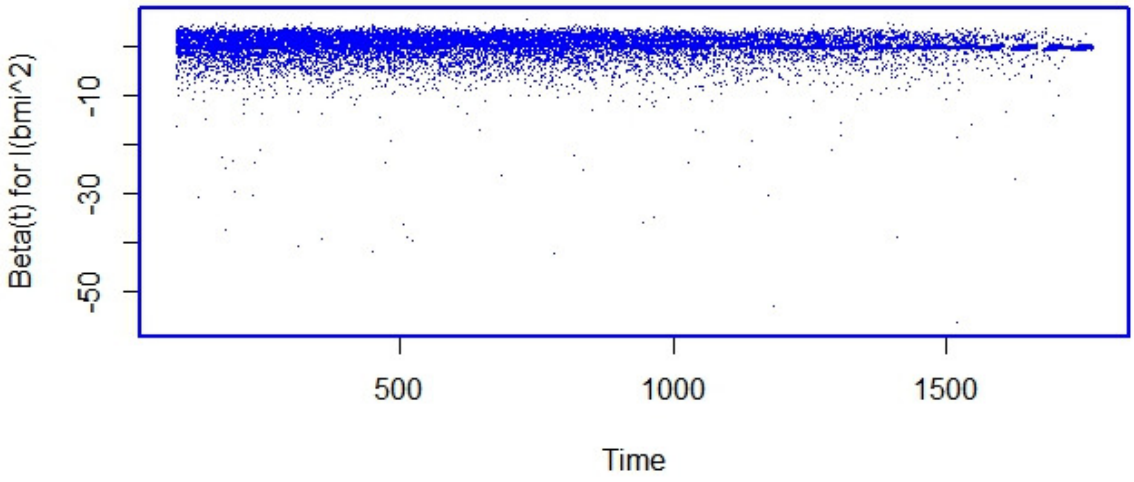



**Figure 8. Scaled Schoenfeld Residuals for BMI versus Time**

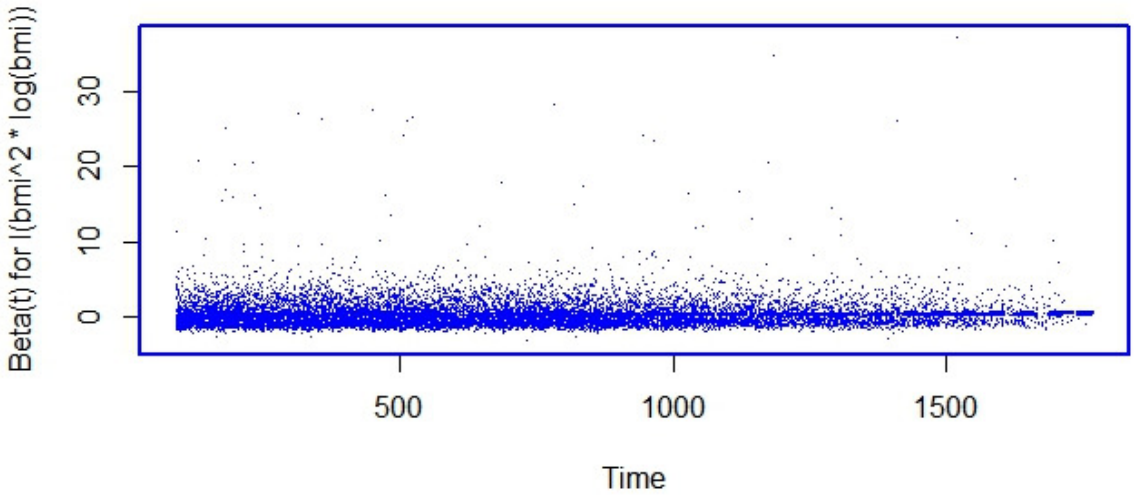


**Figure 9. Scaled Schoenfeld Residuals CHD for versus Time**

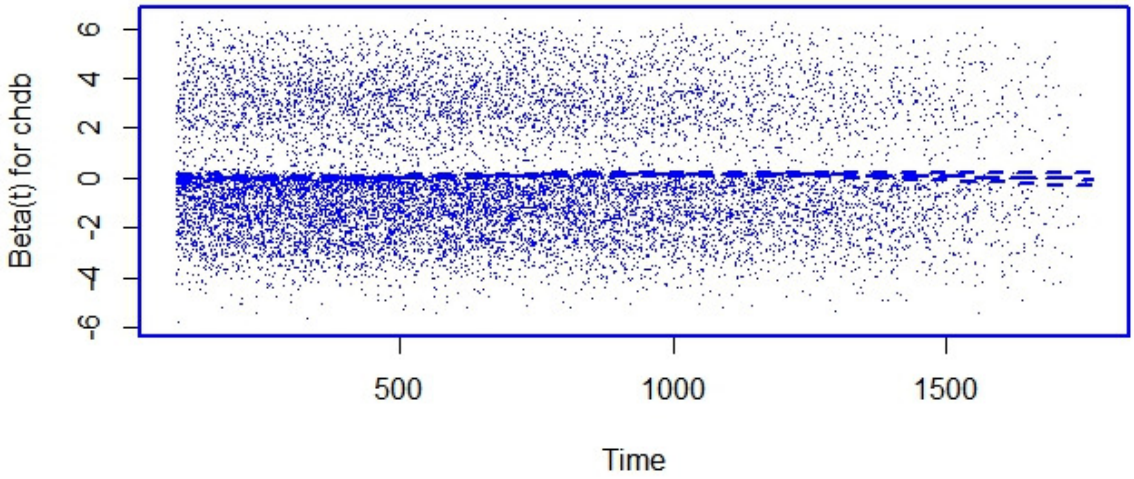



Figure 10. Scaled Schoenfeld Residuals cholesterol for versus Time

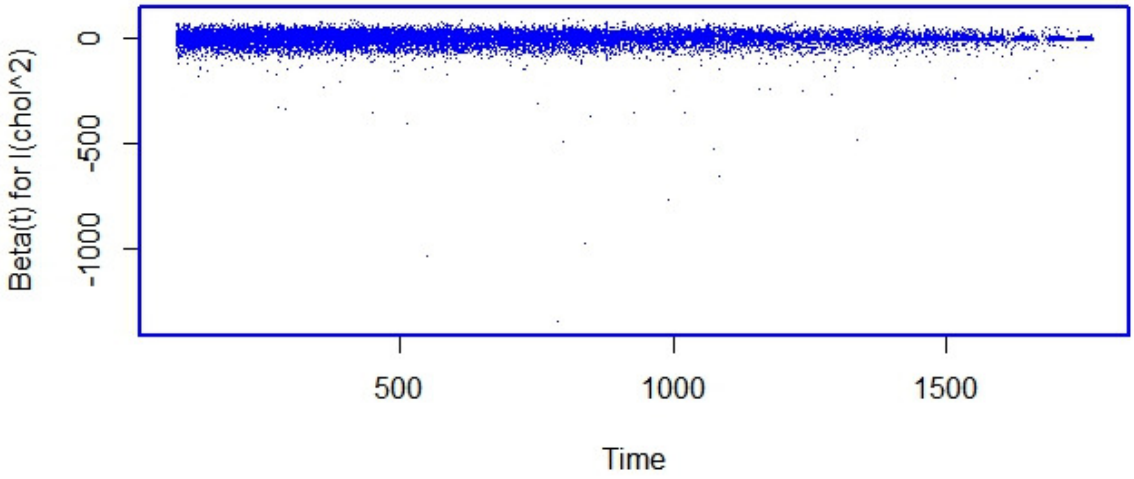


**Figure 11. Scaled Schoenfeld Residuals Cholesterol versus Time**

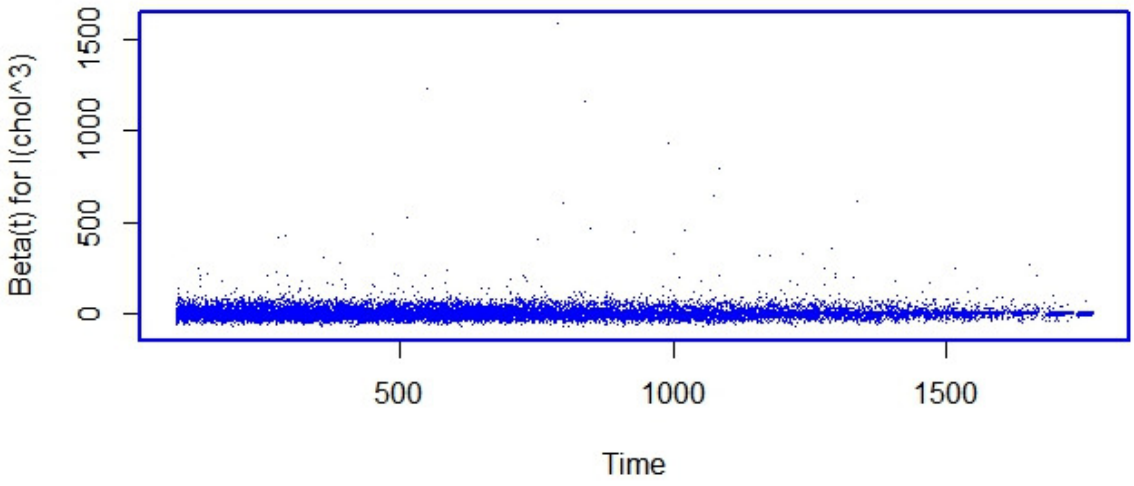


**Figure 12. Scaled Schoenfeld Residuals CVA versus Time**

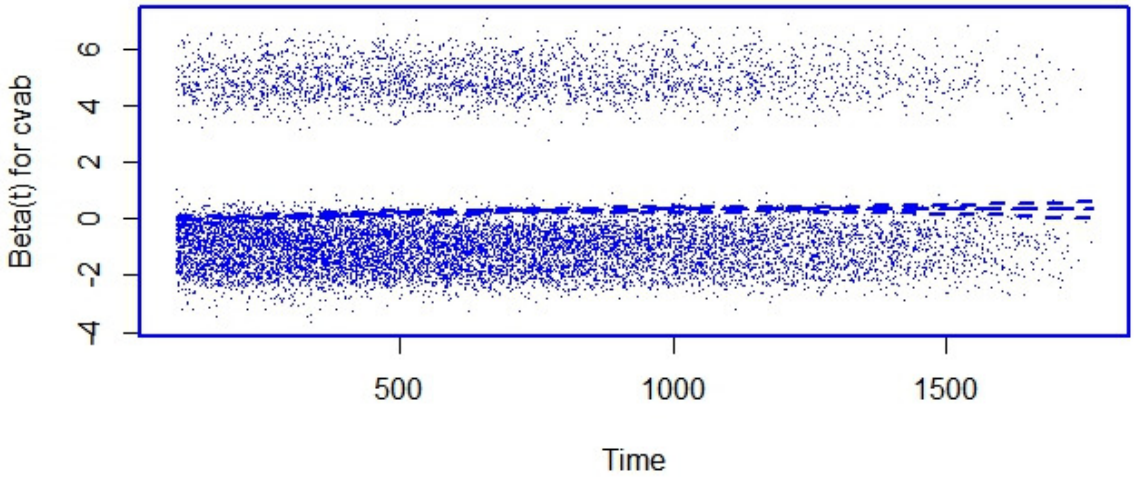



**Figure 13. Scaled Schoenfeld Residuals for Diuretics versus Time**

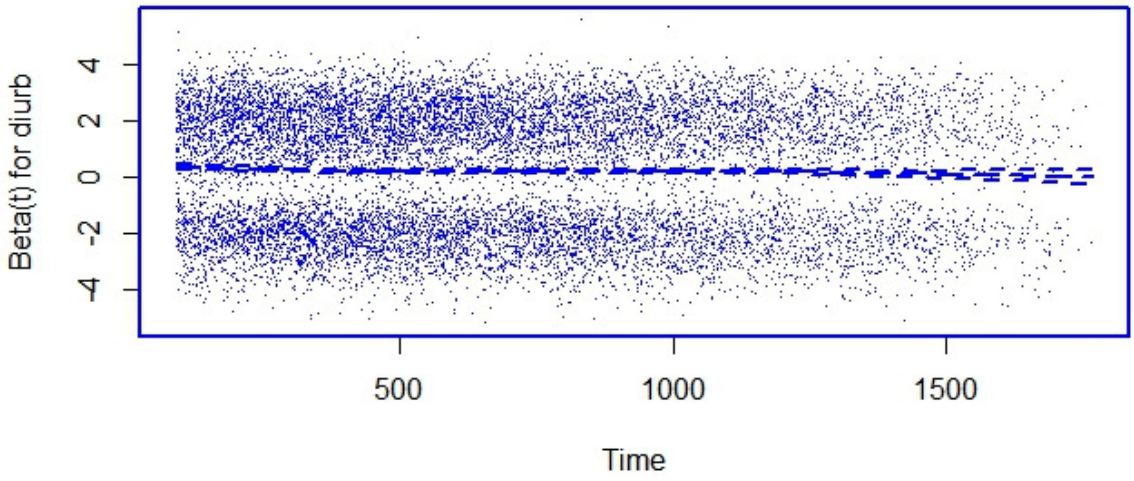


**Figure 14. Scaled Schoenfeld Residuals Iron supplements for versus Time**

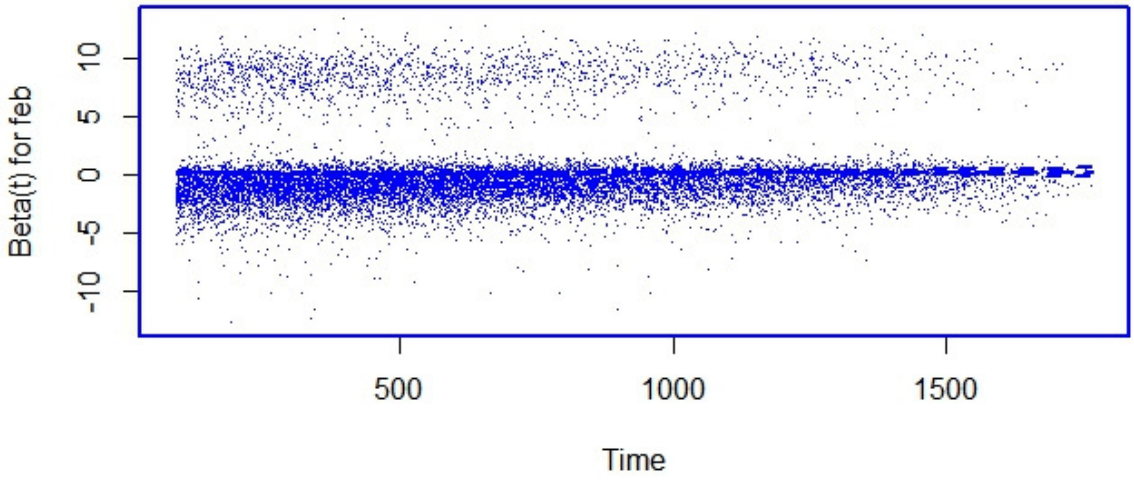



**Figure 15. Scaled Schoenfeld Residuals GFR for versus Time**

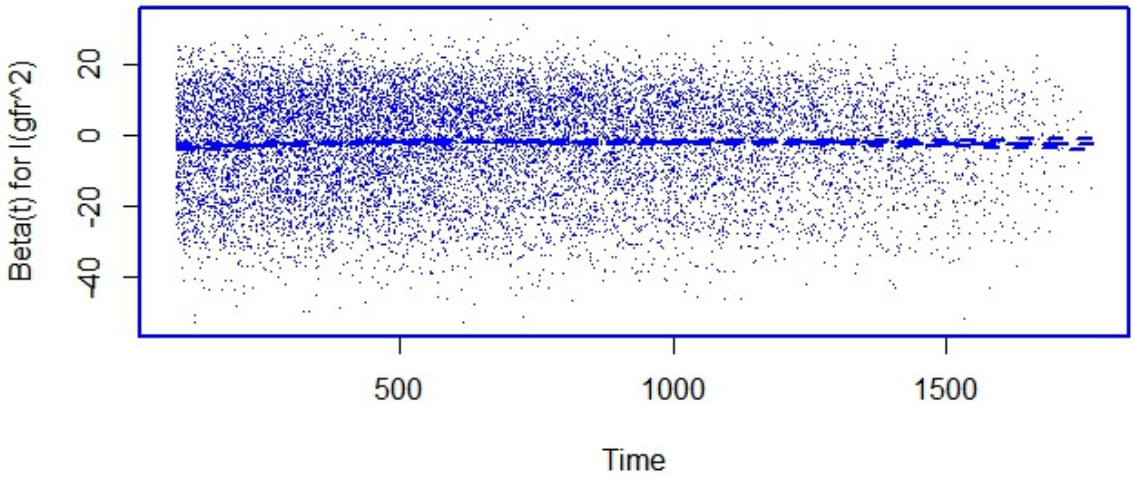


**Figure 16. Scaled Schoenfeld Residuals for HB versus Time**

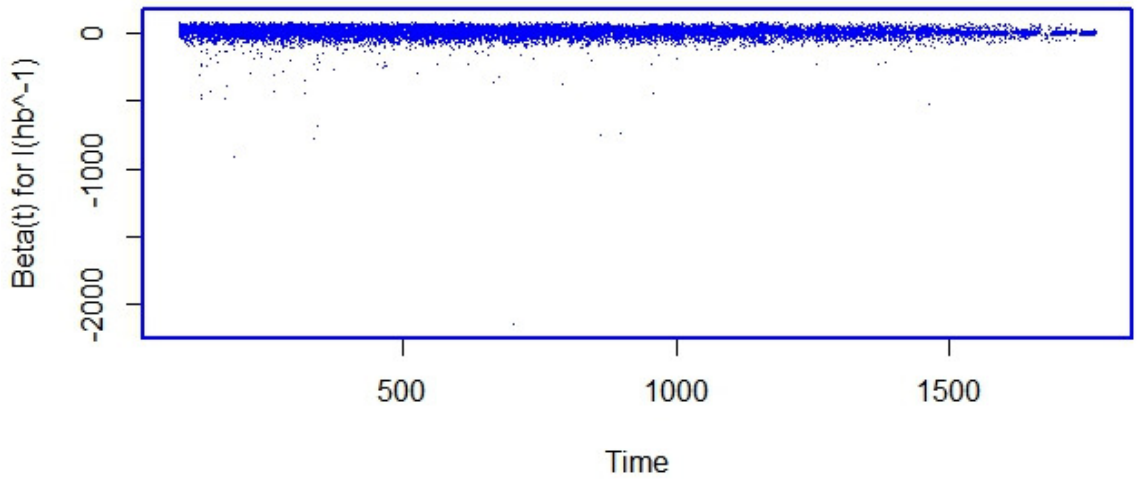



**Figure 17. Scaled Schoenfeld Residuals for versus Time**

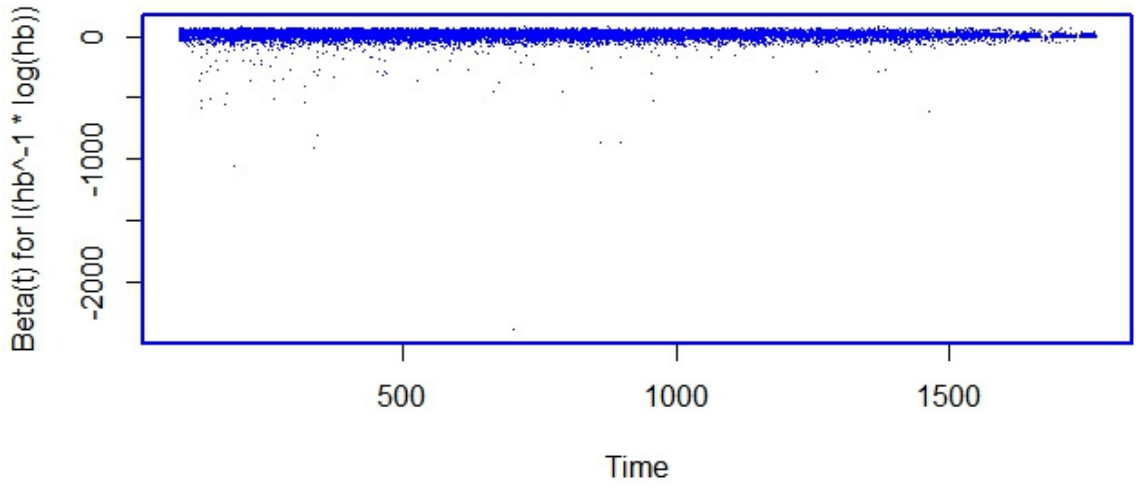


**Figure 18. Scaled Schoenfeld Residuals for Heart failure versus Time**

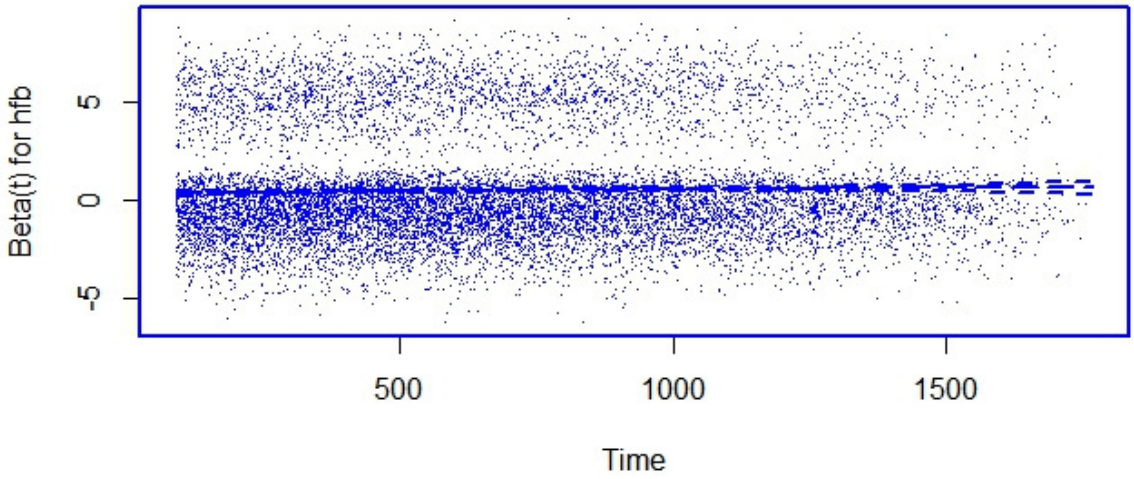



**Figure 19. Scaled Schoenfeld Residuals for ISC race versus Time**

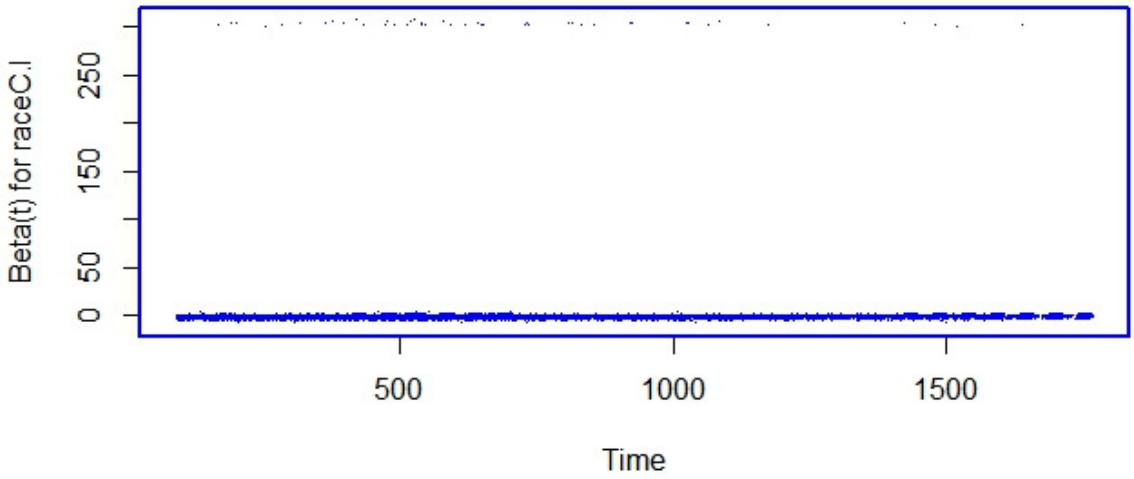


**Figure 20. Scaled Schoenfeld Residuals for lipid lowering agents versus Time**

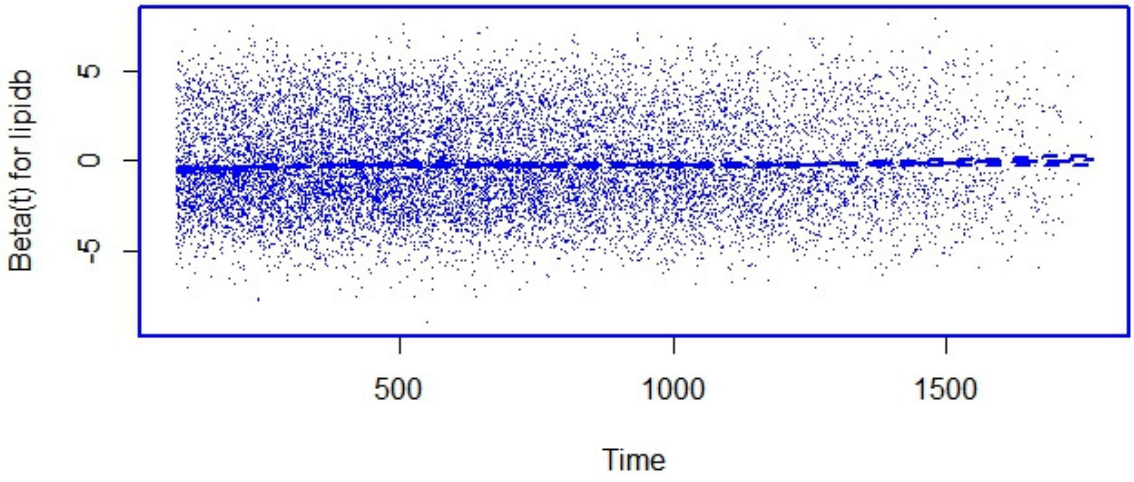



**Figure 21. Scaled Schoenfeld Residuals for other anti-hypertensive versus Time**

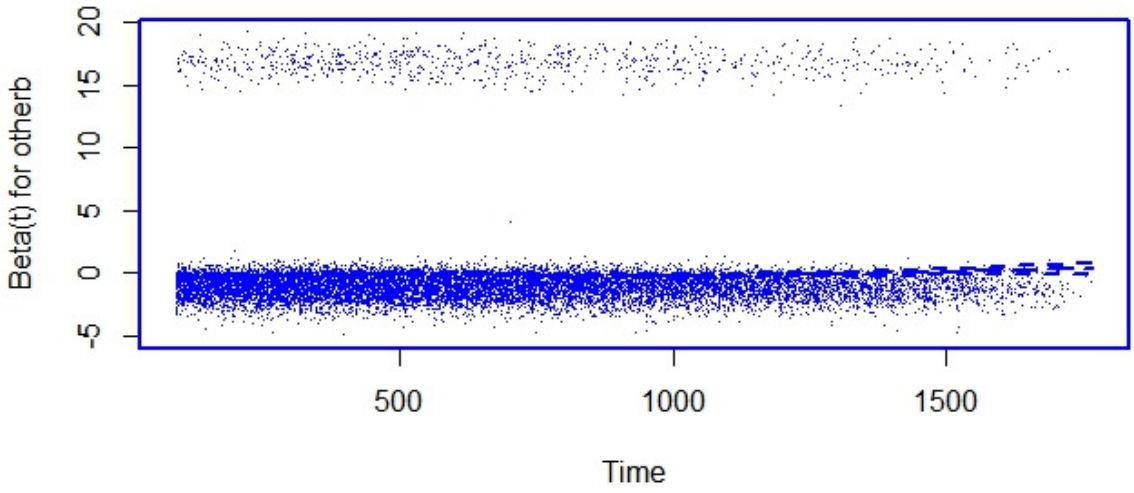


**Figure 22. Scaled Schoenfeld Residuals for Very High proteinuria versus Time**

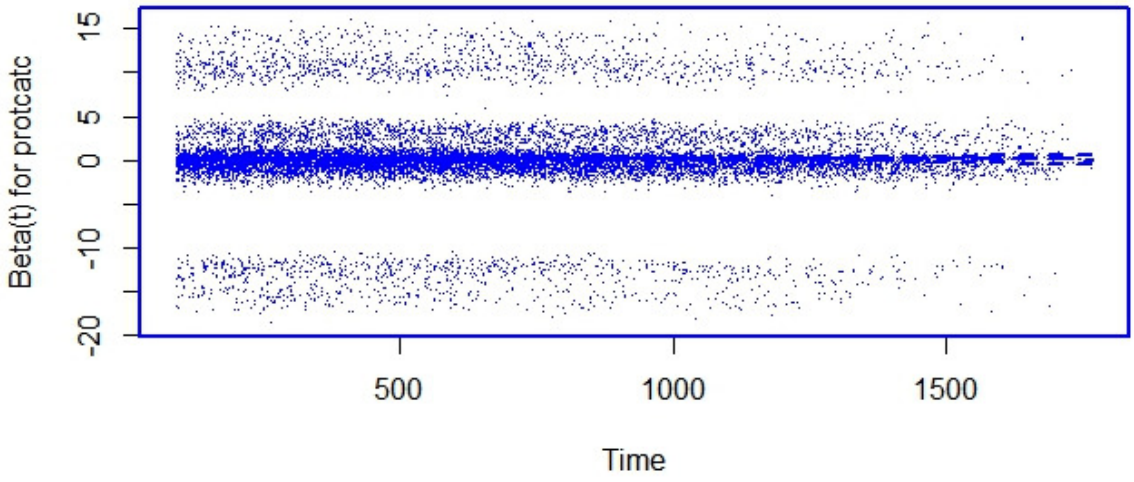



**Figure 23. Scaled Schoenfeld Residuals for versus Time**

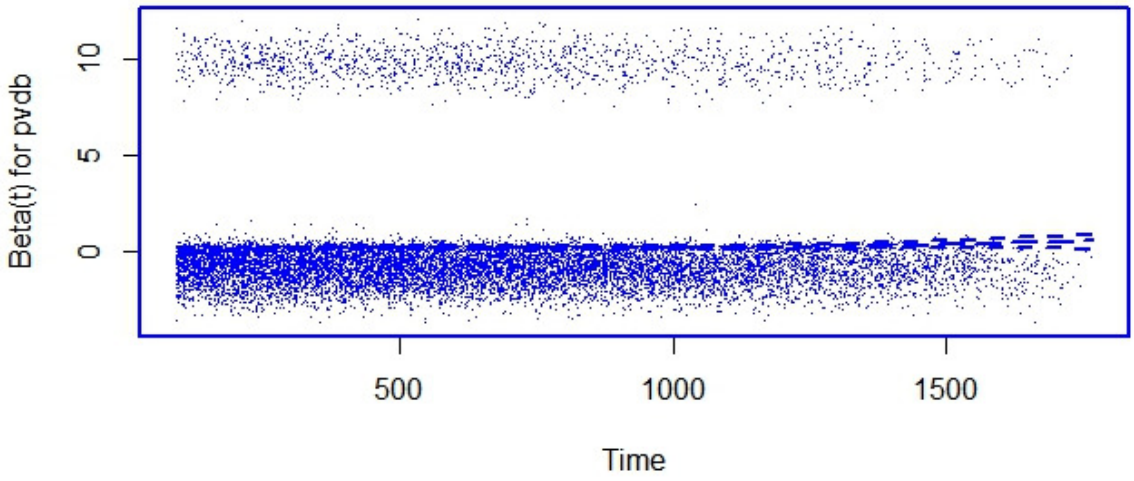


**Figure 24. Scaled Schoenfeld Residuals for Angiotensin blockade versus Time**

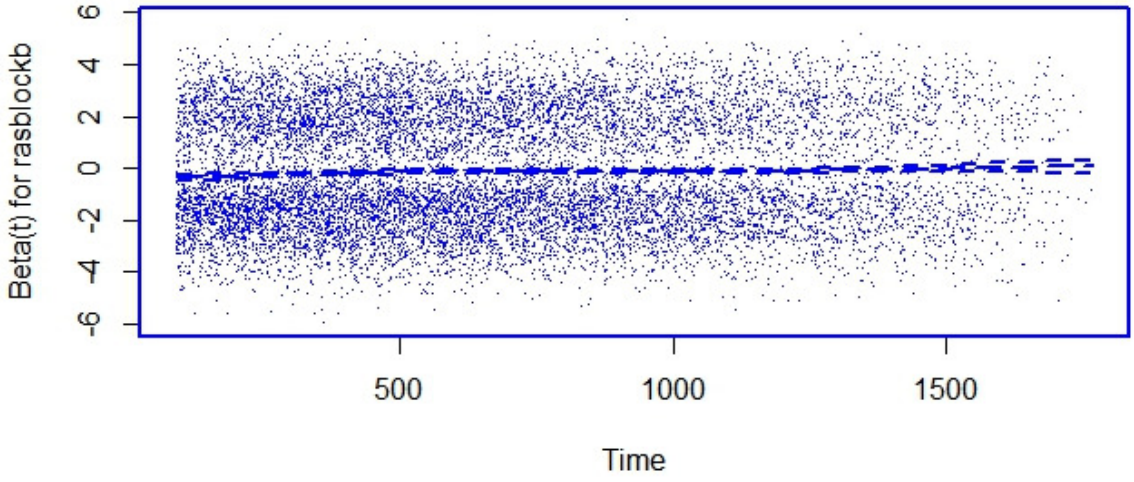



**Figure 25. Scaled Schoenfeld Residuals for gender versus Time**

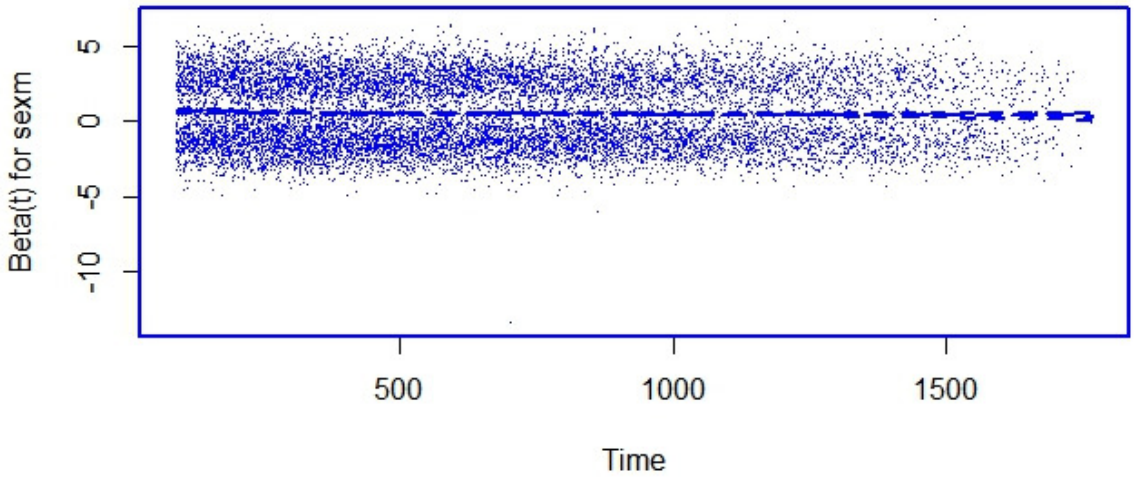


**Figure 26. Scaled Schoenfeld Residuals for Smokers/Ex Smokers versus Time**

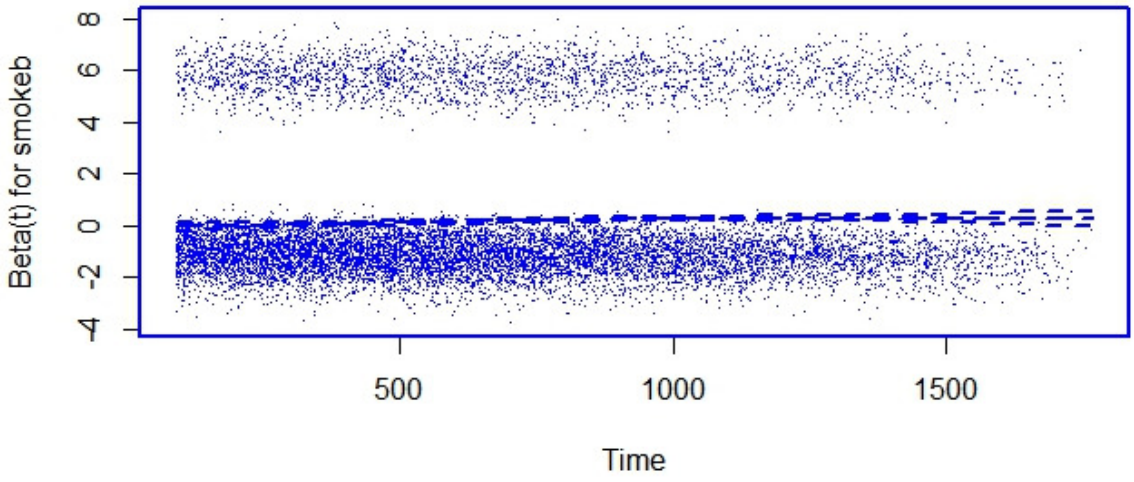



**Figure 27. Scaled Schoenfeld Residuals for systolic BP versus Time**

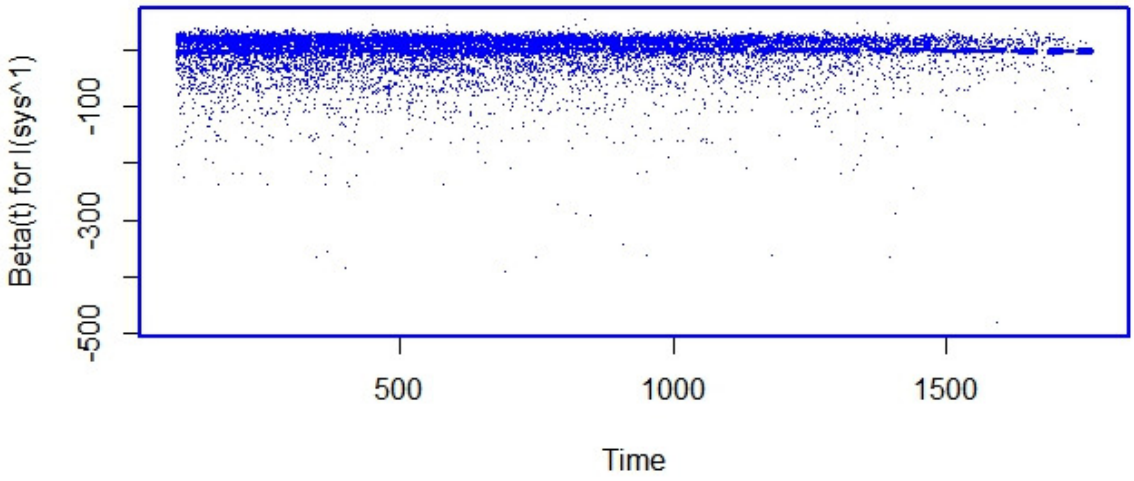


**Figure 28. Scaled Schoenfeld Residuals for Systolic BP versus Time**

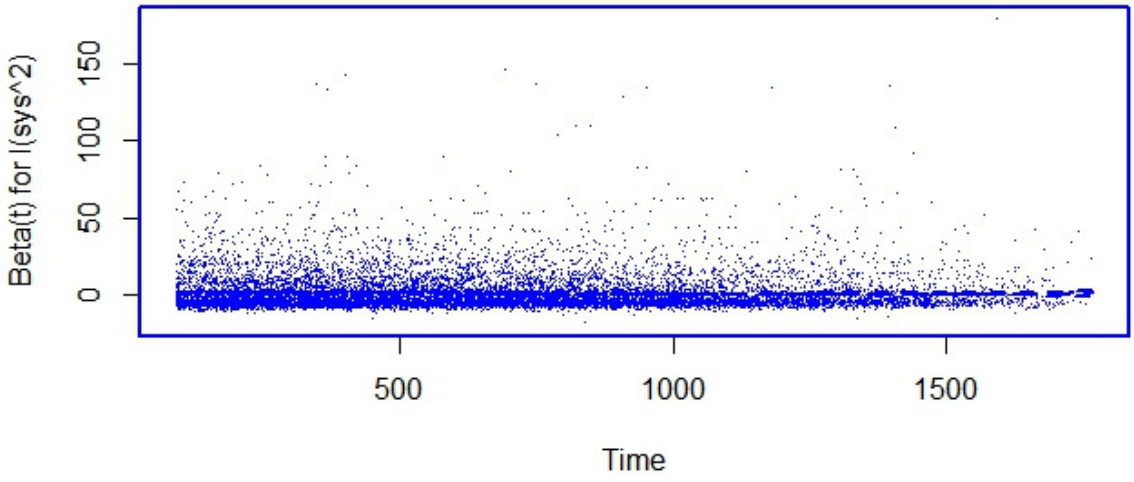



**Figure 29. Scaled Schoenfeld Residuals for Townsend Quintile 2 versus Time**

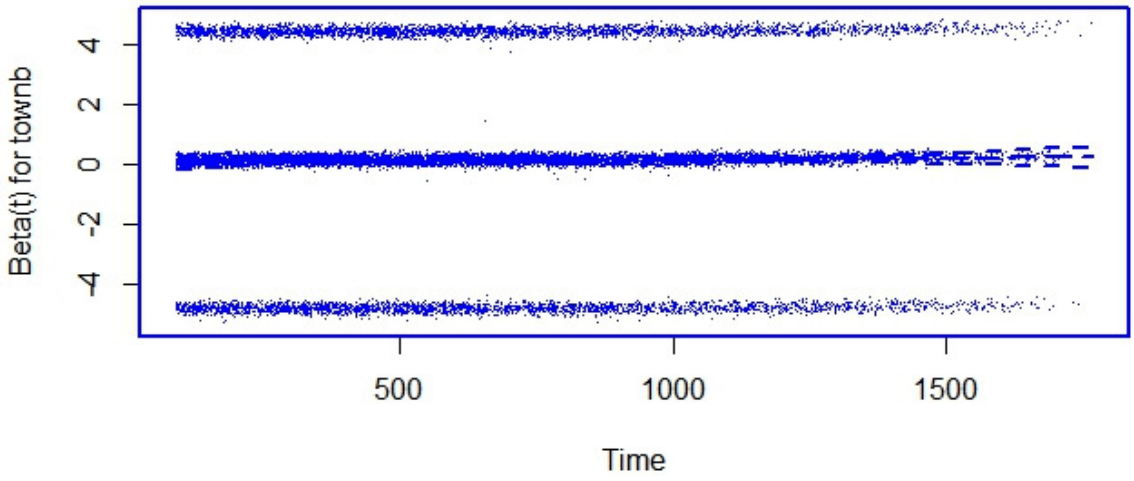


**Figure 30. Scaled Schoenfeld Residuals for Townsend Quintile 3 versus Time**

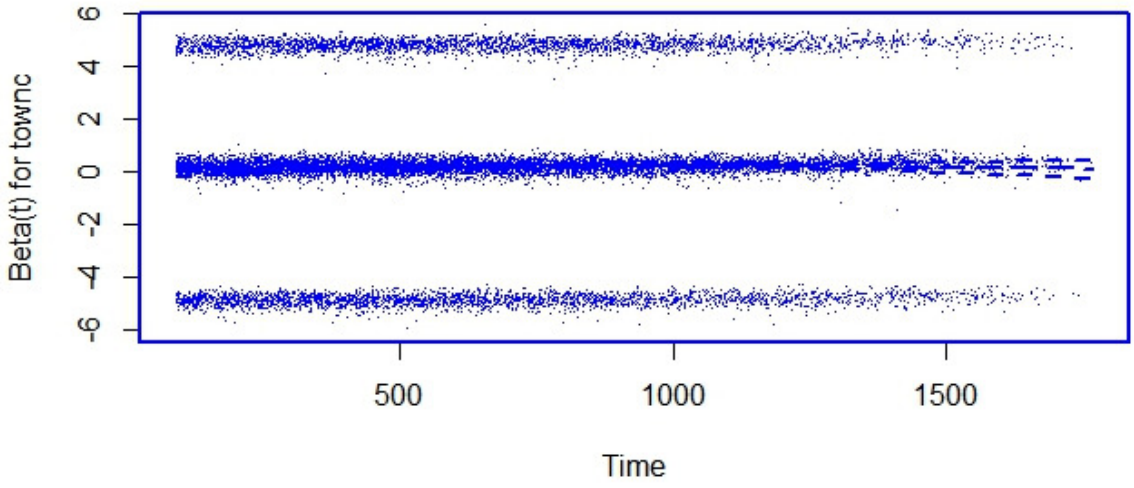



**Figure 31. Scaled Schoenfeld Residuals for Townsend Quintile 4 versus Time**

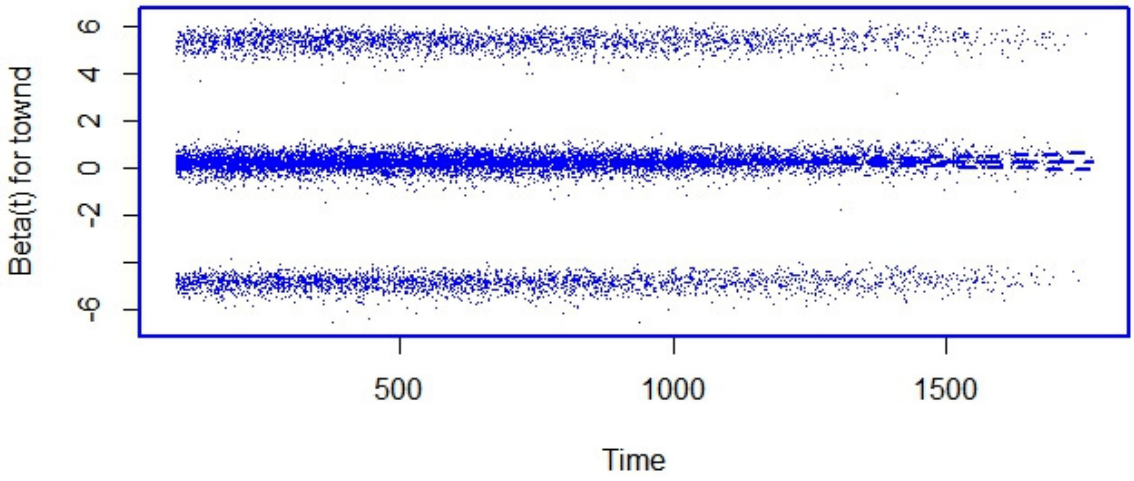


**Figure 32. Scaled Schoenfeld Residuals for Townsend Quintile 5 versus Time**

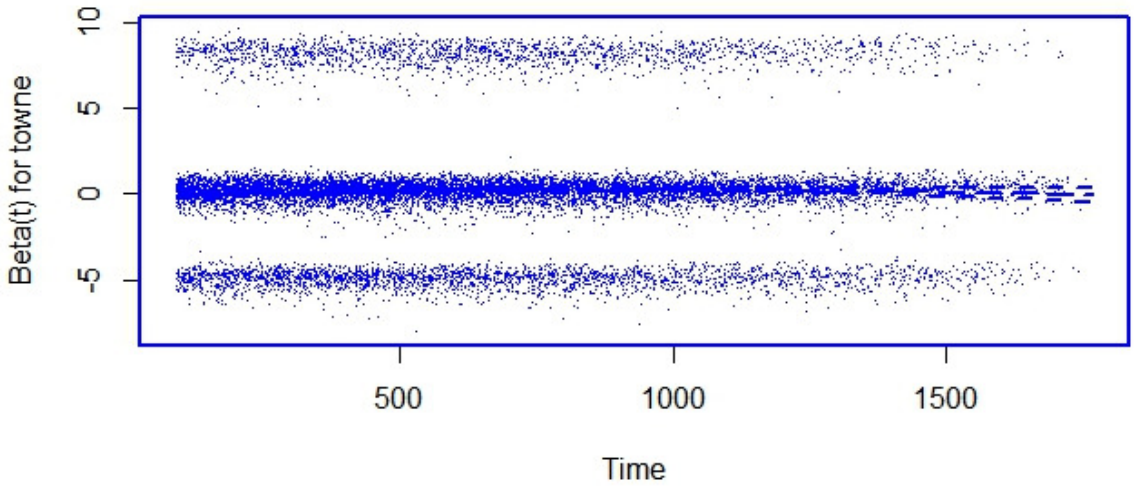
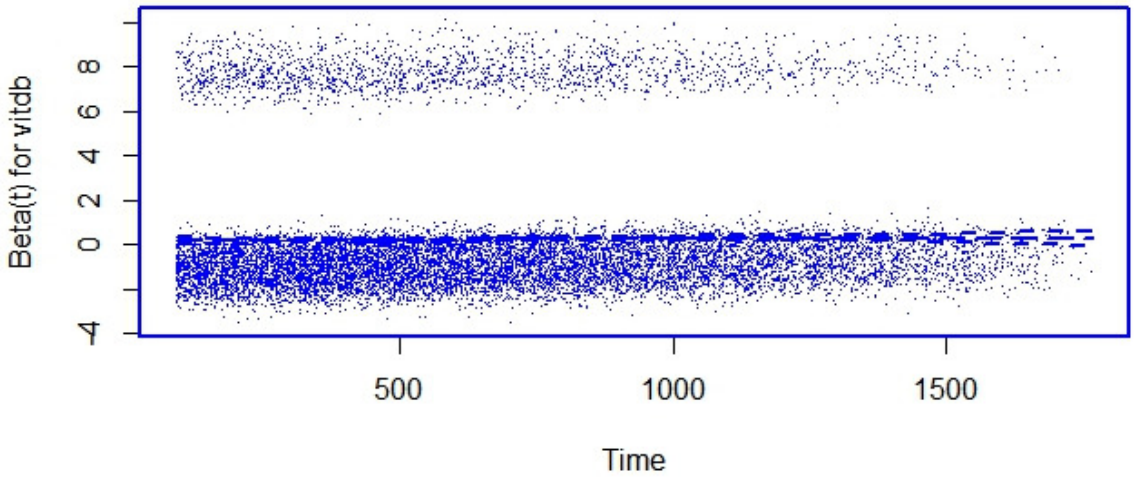



**Figure 33. Scaled Schoenfeld Residuals for Vit D versus Time**



# Appendix C

**Figure 1. Histogram of original Cholesterol data pre imputation**

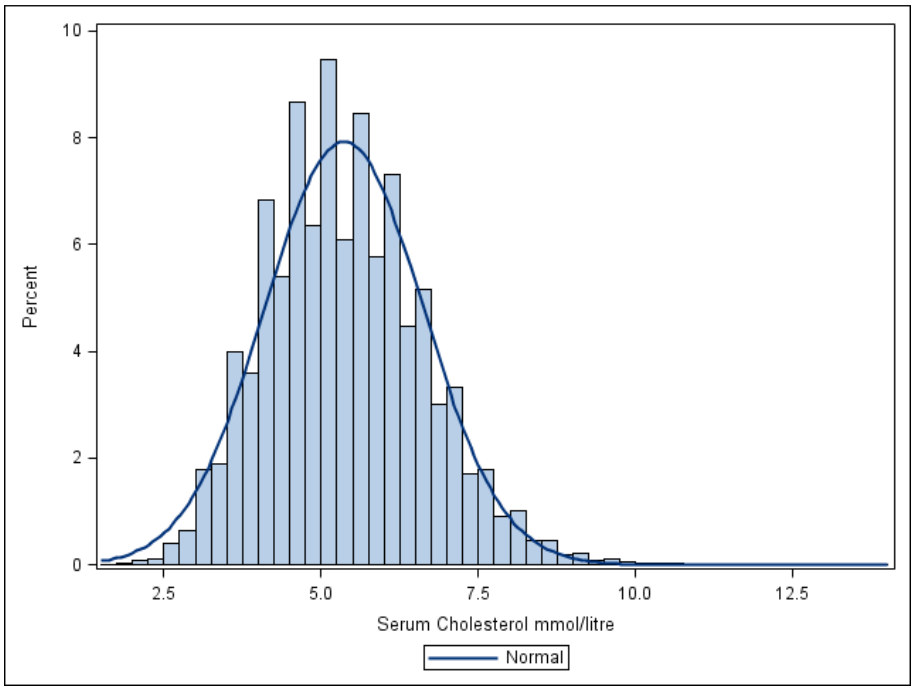



**Figure 2. Histogram of original Body Mass index pre imputation**

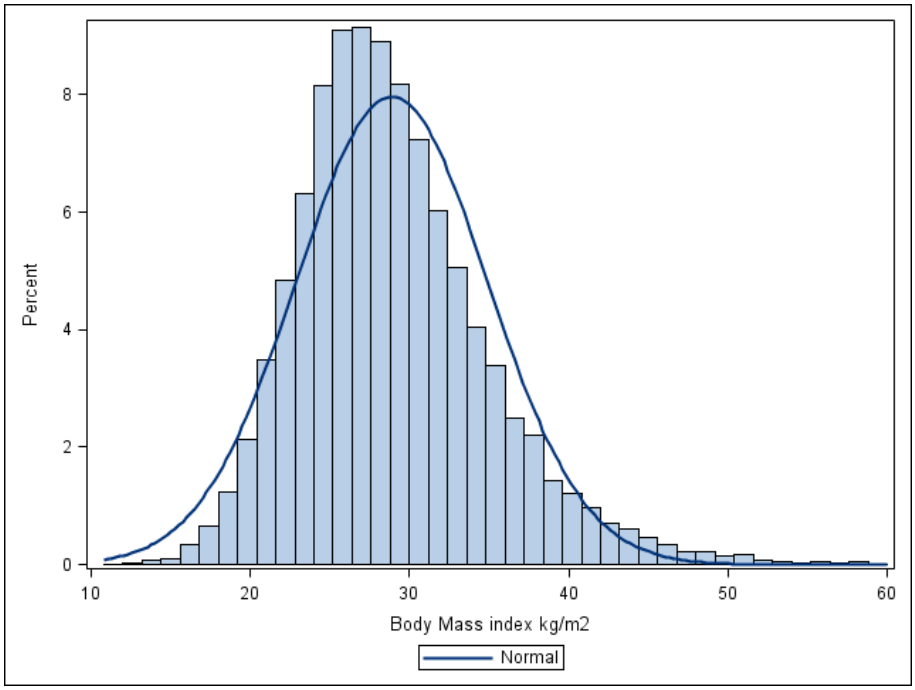


**Figure 3. Histogram of original Systolic blood pressure data pre imputation**

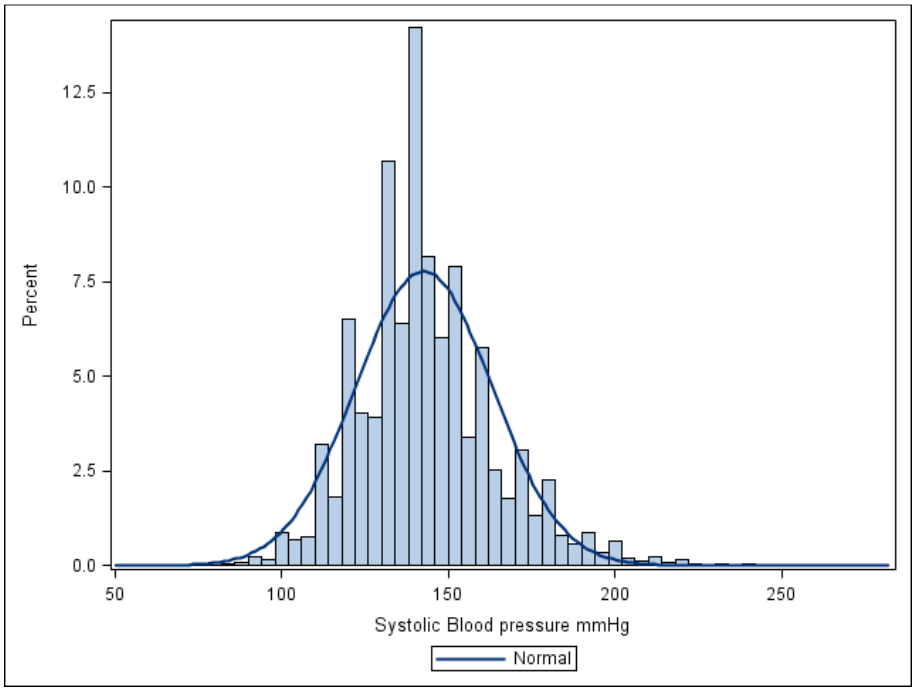



**Figure 4. Histogram of original Diastolic blood pressure data pre imputation**

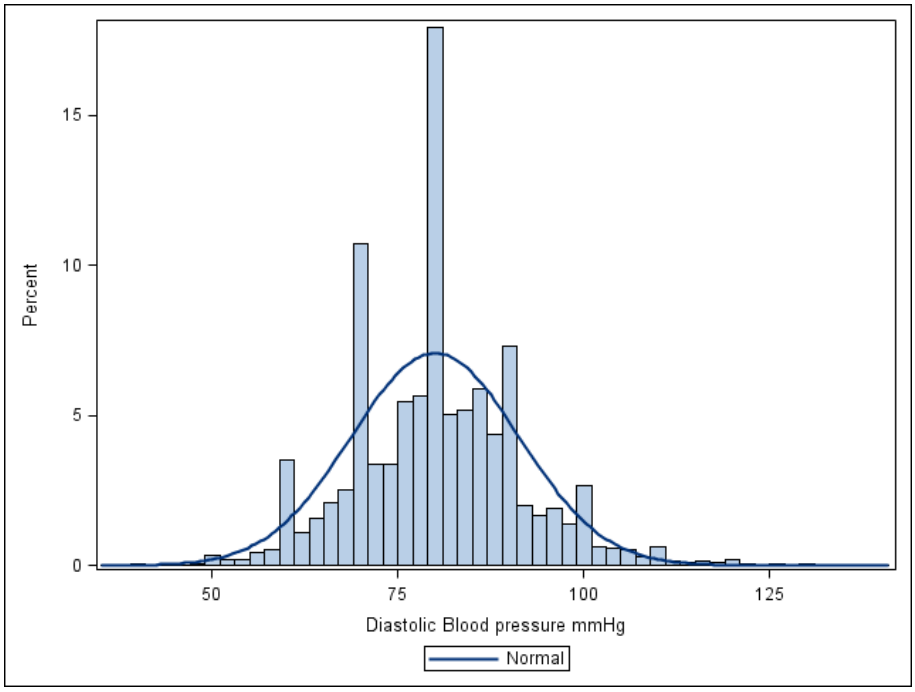


**Figure 5. Histogram of original Haemoglobin data pre imputation**

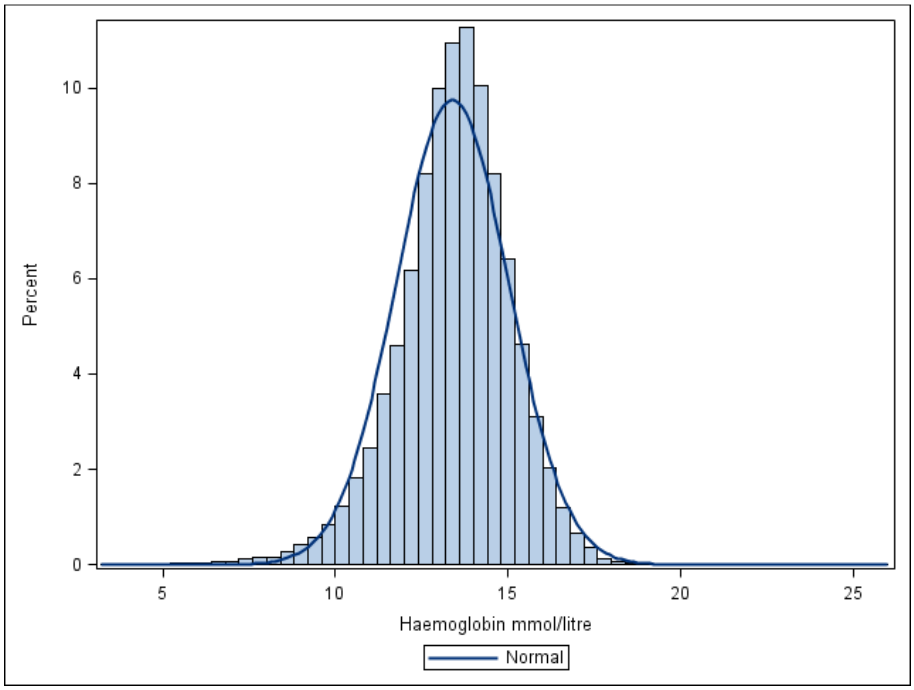


**Table 1. Cox regression analysis for Imputation 1 without transformation and frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| Age | 0.04842 | 0.00101 | 1.05 | 1.05 | 1.05 |
| Male Gender | -0.49218 | 0.02041 | 0.61 | 0.59 | 0.64 |
| Race with African Caribbean<br>Race as reference<br>African –Caribbean | -0.41615 | 0.15713 | 0.66 | 0.48 | 0.90 |
| Indian Subcontinent | -0.25705 | 0.11647 | 0.77 | 0.62 | 0.97 |
| South East Asian | -0.50717 | 0.57761 | 0.60 | 0.19 | 1.87 |
| Townsend Quintile compared to $1^{th}$ Quintile<br>$2^{nd}$ | 0.08576 | 0.02671 | 1.09 | 1.03 | 1.15 |
| $3^{rd}$ | 0.11812 | 0.02753 | 1.13 | 1.07 | 1.19 |
| $4^{th}$ | 0.15209 | 0.02844 | 1.16 | 1.10 | 1.23 |
| $5^{th}$ | 0.20154 | 0.03229 | 1.22 | 1.15 | 1.30 |
| Diabetes Mellitus | 0.17665 | 0.02483 | 1.19 | 1.14 | 1.25 |
| Heart Failure | 0.54364 | 0.03024 | 1.72 | 1.62 | 1.83 |
| Atrial Fibrillation | 0.22331 | 0.02836 | 1.25 | 1.18 | 1.32 |
| Ever Smoked | 0.23456 | 0.02396 | 1.26 | 1.21 | 1.33 |
| Systolic BP | -0.00219 | 0.000462 | 1.00 | 1.00 | 1.00 |
| BMI | -0.01941 | 0.00188 | 0.98 | 0.98 | 0.98 |
| Haemoglobin | -0.11735 | 0.00592 | 0.89 | 0.88 | 0.90 |
| Glomerular Filtration Rate | -0.01552 | 0.00113 | 0.98 | 0.98 | 0.99 |
| Cholesterol | -0.02845 | 0.00818 | 0.97 | 0.96 | 0.99 |
| Proteinuria levels with none as reference<br>High | 0.15116 | 0.03363 | 1.16 | 1.09 | 1.24 |
| Very High | 0.49561 | 0.04197 | 1.64 | 1.51 | 1.78 |
| Aspirin | 0.22012 | 0.02101 | 1.25 | 1.20 | 1.30 |
| Anticoagulation | 0.21485 | 0.03899 | 1.24 | 1.15 | 1.34 |
| Diuretic use | 0.08411 | 0.01985 | 1.09 | 1.05 | 1.13 |
| Other | -0.1224 | 0.03904 | 0.88 | 0.82 | 0.96 |
| Angiotensin Blockade | -0.17006 | 0.02036 | 0.84 | 0.81 | 0.88 |
| Iron Medication | 0.13689 | 0.03419 | 1.15 | 1.07 | 1.23 |
| Vitamin D supplementation | 0.17716 | 0.02953 | 1.19 | 1.13 | 1.26 |

**Table 2. Cox regression analysis for Imputation 2 without transformation and frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| Age | 0.04841 | 0.00101 | 1.05 | 1.05 | 1.05 |
| Male Gender | -0.48303 | 0.02048 | 0.62 | 0.59 | 0.64 |
| Race with African Caribbean Race as reference African –Caribbean | -0.42724 | 0.15714 | 0.65 | 0.48 | 0.89 |
| Indian Subcontinent | -0.27402 | 0.11654 | 0.76 | 0.61 | 0.96 |
| South East Asian | -0.50589 | 0.57767 | 0.60 | 0.19 | 1.87 |
| Townsend Quintile compared to 1$^{th}$ Quintile 2$^{nd}$ | 0.08885 | 0.02667 | 1.09 | 1.04 | 1.15 |
| 3$^{rd}$ | 0.11346 | 0.02756 | 1.12 | 1.06 | 1.18 |
| 4$^{th}$ | 0.14299 | 0.02853 | 1.15 | 1.09 | 1.22 |
| 5$^{th}$ | 0.1781 | 0.03254 | 1.19 | 1.12 | 1.27 |
| Diabetes Mellitus | 0.16943 | 0.02471 | 1.18 | 1.13 | 1.24 |
| Heart Failure | 0.53526 | 0.03032 | 1.71 | 1.61 | 1.81 |
| Atrial Fibrillation | 0.21496 | 0.02834 | 1.24 | 1.17 | 1.31 |
| Ever Smoked | 0.22948 | 0.02398 | 1.26 | 1.20 | 1.32 |
| Systolic BP | -0.00224 | 0.000459 | 1.00 | 1.00 | 1.00 |
| BMI | -0.02041 | 0.00189 | 0.98 | 0.98 | 0.98 |
| Haemoglobin | -0.11185 | 0.00594 | 0.89 | 0.88 | 0.90 |
| Glomerular Filtration Rate | -0.01557 | 0.00113 | 0.98 | 0.98 | 0.99 |
| Cholesterol | -0.04343 | 0.00823 | 0.96 | 0.94 | 0.97 |
| Proteinuria levels with none as reference High | 0.17037 | 0.03429 | 1.19 | 1.11 | 1.27 |
| Very High | 0.48867 | 0.04323 | 1.63 | 1.50 | 1.77 |
| Aspirin | 0.20915 | 0.02104 | 1.23 | 1.18 | 1.28 |
| Anticoagulation | 0.19884 | 0.0389 | 1.22 | 1.13 | 1.32 |
| Diuretic use | 0.08948 | 0.0199 | 1.09 | 1.05 | 1.14 |
| Other | -0.12648 | 0.03903 | 0.88 | 0.82 | 0.95 |
| Angiotensin Blockade | -0.16935 | 0.02039 | 0.84 | 0.81 | 0.88 |
| Iron Medication | 0.14908 | 0.03413 | 1.16 | 1.09 | 1.24 |
| Vitamin D supplementation | 0.1675 | 0.02952 | 1.18 | 1.12 | 1.25 |

**Table 3. Cox regression analysis for Imputation 3 without transformation and frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| Age | 0.0493 | 0.00101 | 1.05 | 1.05 | 1.05 |
| Male Gender | -0.49306 | 0.02051 | 0.61 | 0.59 | 0.64 |
| Race with African Caribbean Race as reference African –Caribbean | -0.43145 | 0.15713 | 0.65 | 0.48 | 0.88 |
| Indian Subcontinent | -0.26388 | 0.11645 | 0.77 | 0.61 | 0.96 |
| South East Asian | -0.42055 | 0.57767 | 0.66 | 0.21 | 2.04 |
| Townsend Quintile compared to 1th Quintile 2nd | 0.0793 | 0.02668 | 1.08 | 1.03 | 1.14 |
| 3rd | 0.11506 | 0.02752 | 1.12 | 1.06 | 1.18 |
| 4th | 0.1507 | 0.02842 | 1.16 | 1.10 | 1.23 |
| 5th | 0.17923 | 0.03244 | 1.20 | 1.12 | 1.27 |
| Diabetes Mellitus | 0.16297 | 0.02472 | 1.18 | 1.12 | 1.24 |
| Heart Failure | 0.54316 | 0.03027 | 1.72 | 1.62 | 1.83 |
| Atrial Fibrillation | 0.22588 | 0.02829 | 1.25 | 1.19 | 1.32 |
| Ever Smoked | 0.24077 | 0.02399 | 1.27 | 1.21 | 1.33 |
| Systolic BP | -0.00265 | 0.000463 | 1.00 | 1.00 | 1.00 |
| BMI | -0.01266 | 0.00184 | 0.99 | 0.98 | 0.99 |
| Haemoglobin | -0.12049 | 0.00593 | 0.89 | 0.88 | 0.90 |
| Glomerular Filtration Rate | -0.01588 | 0.00113 | 0.98 | 0.98 | 0.99 |
| Cholesterol | -0.02782 | 0.00823 | 0.97 | 0.96 | 0.99 |
| Proteinuria levels with none as reference High | 0.20771 | 0.03443 | 1.23 | 1.15 | 1.32 |
| Very High | 0.4515 | 0.04326 | 1.57 | 1.44 | 1.71 |
| Aspirin | 0.2233 | 0.02103 | 1.25 | 1.20 | 1.30 |
| Anticoagulation | 0.19659 | 0.03896 | 1.22 | 1.13 | 1.31 |
| Diuretic use | 0.0767 | 0.0199 | 1.08 | 1.04 | 1.12 |
| Other | -0.12347 | 0.03904 | 0.88 | 0.82 | 0.95 |
| Angiotensin Blockade | -0.16834 | 0.02039 | 0.85 | 0.81 | 0.88 |
| Iron Medication | 0.14261 | 0.03419 | 1.15 | 1.08 | 1.23 |
| Vitamin D supplementation | 0.1718 | 0.02955 | 1.19 | 1.12 | 1.26 |

**Table 4, Cox regression analysis for Imputation 4 without transformation and frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| Age | 0.04885 | 0.00101 | 1.05 | 1.05 | 1.05 |
| Male Gender | -0.4939 | 0.02047 | 0.61 | 0.59 | 0.64 |
| Race with African Caribbean Race as reference African –Caribbean | -0.42634 | 0.15714 | 0.65 | 0.48 | 0.89 |
| Indian Subcontinent | -0.26383 | 0.11647 | 0.77 | 0.61 | 0.97 |
| South East Asian | -0.6401 | 0.57811 | 0.53 | 0.17 | 1.64 |
| Townsend Quintile compared to 1th Quintile 2nd | 0.09896 | 0.02669 | 1.10 | 1.05 | 1.16 |
| 3rd | 0.13878 | 0.02754 | 1.15 | 1.09 | 1.21 |
| 4th | 0.15799 | 0.0285 | 1.17 | 1.11 | 1.24 |
| 5th | 0.1973 | 0.03245 | 1.22 | 1.14 | 1.30 |
| Diabetes Mellitus | 0.16631 | 0.02482 | 1.18 | 1.12 | 1.24 |
| Heart Failure | 0.53911 | 0.0303 | 1.71 | 1.62 | 1.82 |
| Atrial Fibrillation | 0.21938 | 0.02832 | 1.25 | 1.18 | 1.32 |
| Ever Smoked | 0.23315 | 0.02396 | 1.26 | 1.20 | 1.32 |
| Systolic BP | -0.00244 | 0.000464 | 1.00 | 1.00 | 1.00 |
| BMI | -0.0175 | 0.00187 | 0.98 | 0.98 | 0.99 |
| Haemoglobin | -0.11354 | 0.00596 | 0.89 | 0.88 | 0.90 |
| Glomerular Filtration Rate | -0.0156 | 0.00113 | 0.98 | 0.98 | 0.99 |
| Cholesterol | -0.02268 | 0.00813 | 0.98 | 0.96 | 0.99 |
| Proteinuria levels with none as reference High | 0.13478 | 0.03457 | 1.14 | 1.07 | 1.22 |
| Very High | 0.42549 | 0.0431 | 1.53 | 1.41 | 1.67 |
| Aspirin | 0.22096 | 0.02101 | 1.25 | 1.20 | 1.30 |
| Anticoagulation | 0.1976 | 0.03892 | 1.22 | 1.13 | 1.32 |
| Diuretic use | 0.08123 | 0.01986 | 1.08 | 1.04 | 1.13 |
| Other | -0.12372 | 0.03908 | 0.88 | 0.82 | 0.95 |
| Angiotensin Blockade | -0.16016 | 0.02038 | 0.85 | 0.82 | 0.89 |
| Iron Medication | 0.16259 | 0.03412 | 1.18 | 1.10 | 1.26 |
| Vitamin D supplementation | 0.17201 | 0.02953 | 1.19 | 1.12 | 1.26 |

**Table 5. Cox regression analysis for Imputation 5 without transformation and frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| Age | 0.04898 | 0.00102 | 1.05 | 1.05 | 1.05 |
| Male Gender | -0.48506 | 0.02046 | 0.62 | 0.59 | 0.64 |
| Race with African Caribbean Race as reference African –Caribbean | -0.43105 | 0.15717 | 0.65 | 0.48 | 0.88 |
| Indian Subcontinent | -0.26216 | 0.11654 | 0.77 | 0.61 | 0.97 |
| South East Asian | -0.66425 | 0.57783 | 0.51 | 0.17 | 1.60 |
| Townsend Quintile compared to 1$^{th}$ Quintile 2$^{nd}$ | 0.0967 | 0.02671 | 1.10 | 1.05 | 1.16 |
| 3$^{rd}$ | 0.12154 | 0.02758 | 1.13 | 1.07 | 1.19 |
| 4$^{th}$ | 0.14219 | 0.02854 | 1.15 | 1.09 | 1.22 |
| 5$^{th}$ | 0.19873 | 0.03233 | 1.22 | 1.14 | 1.30 |
| Diabetes Mellitus | 0.14489 | 0.02514 | 1.16 | 1.10 | 1.21 |
| Heart Failure | 0.54025 | 0.03031 | 1.72 | 1.62 | 1.82 |
| Atrial Fibrillation | 0.21821 | 0.02844 | 1.24 | 1.18 | 1.32 |
| Ever Smoked | 0.2385 | 0.02398 | 1.27 | 1.21 | 1.33 |
| Systolic BP | -0.00273 | 0.0004626 | 1.00 | 1.00 | 1.00 |
| BMI | -0.01567 | 0.00188 | 0.98 | 0.98 | 0.99 |
| Haemoglobin | -0.11058 | 0.00597 | 0.90 | 0.88 | 0.91 |
| Glomerular Filtration Rate | -0.01481 | 0.00113 | 0.99 | 0.98 | 0.99 |
| Cholesterol | -0.03418 | 0.00831 | 0.97 | 0.95 | 0.98 |
| Proteinuria levels with none as reference High | 0.00606 | 0.03223 | 1.01 | 0.94 | 1.07 |
| Very High | 0.41909 | 0.04114 | 1.52 | 1.40 | 1.65 |
| Aspirin | 0.22485 | 0.02146 | 1.25 | 1.20 | 1.31 |
| Anticoagulation | 0.20493 | 0.03907 | 1.23 | 1.14 | 1.33 |
| Diuretic use | 0.07983 | 0.01991 | 1.08 | 1.04 | 1.13 |
| Other | -0.12723 | 0.03908 | 0.88 | 0.82 | 0.95 |
| Angiotensin Blockade | -0.16572 | 0.0206 | 0.85 | 0.81 | 0.88 |
| Iron Medication | 0.13598 | 0.03426 | 1.15 | 1.07 | 1.23 |
| Vitamin D supplementation | 0.17614 | 0.02953 | 1.19 | 1.13 | 1.26 |

**Table 6. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 1**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| $Age^2$ | 3.31 | 0.07 | 27.32 | 23.99 | 31.14 |
| Male Gender | 0.50 | 0.02 | 1.65 | 1.59 | 1.72 |
| Race with African Caribbean Race as reference African –Caribbean | -0.45 | 0.16 | 0.64 | 0.47 | 0.86 |
| Indian Subcontinent | -0.26 | 0.12 | 0.77 | 0.61 | 0.97 |
| South East Asian | -0.54 | 0.58 | 0.58 | 0.19 | 1.80 |
| Townsend Quintile compared to $1^{th}$ Quintile $2^{nd}$ | 0.08 | 0.03 | 1.09 | 1.03 | 1.15 |
| $3^{rd}$ | 0.12 | 0.03 | 1.12 | 1.06 | 1.18 |
| $4^{th}$ | 0.15 | 0.03 | 1.16 | 1.10 | 1.23 |
| $5^{th}$ | 0.20 | 0.03 | 1.22 | 1.14 | 1.30 |
| Diabetes Mellitus | 0.19 | 0.02 | 1.20 | 1.15 | 1.26 |
| Heart Failure | 0.52 | 0.03 | 1.68 | 1.59 | 1.79 |
| Atrial Fibrillation | 0.22 | 0.03 | 1.24 | 1.18 | 1.32 |
| Ever Smoked | 0.24 | 0.02 | 1.27 | 1.21 | 1.33 |
| $(Systolic\ BP/100)^{0.5}$ | -10.00 | 1.59 | 0.00 | 0.00 | 0.00 |
| $(Systolic\ BP/100)$ | 3.98 | 0.66 | 53.32 | 14.53 | 195.56 |
| $(BMI/10)^{-0.5}$ | 3.12 | 0.28 | 22.70 | 13.04 | 39.48 |
| $(BMI/10)^3$ | 0.01 | 0.00 | 1.01 | 1.00 | 1.01 |
| $(Hb/10)^{-1}$ | 4.03 | 0.34 | 56.17 | 28.84 | 109.33 |
| $(Hb/10)^{-2}$ | -1.31 | 0.17 | 0.27 | 0.19 | 0.38 |
| $Gfr/100$ | -1.42 | 0.11 | 0.24 | 0.19 | 0.30 |
| $Cholesterol/10$ | -0.28 | 0.08 | 0.76 | 0.64 | 0.89 |
| Proteinuria levels with none as reference High | 0.15 | 0.03 | 1.17 | 1.09 | 1.25 |
| Very High | 0.49 | 0.04 | 1.64 | 1.51 | 1.78 |
| Aspirin | 0.23 | 0.02 | 1.26 | 1.20 | 1.31 |
| Anticoagulation | 0.23 | 0.04 | 1.26 | 1.16 | 1.36 |
| Diuretic use | -0.16 | 0.02 | 0.85 | 0.82 | 0.88 |
| Other | 0.09 | 0.02 | 1.09 | 1.05 | 1.13 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.88 | 0.82 | 0.95 |
| Iron Medication | 0.10 | 0.03 | 1.11 | 1.04 | 1.19 |
| Vitamin D supplementation | 0.16 | 0.03 | 1.18 | 1.11 | 1.25 |

**Table 7. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 2**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| $Age^2$ | 3.30 | 0.07 | 26.98 | 23.66 | 30.76 |
| Male Gender | 0.49 | 0.02 | 1.63 | 1.57 | 1.70 |
| Race with African Caribbean Race as reference | | | | | |
| African –Caribbean | -0.46 | 0.16 | 0.63 | 0.46 | 0.86 |
| Indian Subcontinent | -0.28 | 0.12 | 0.75 | 0.60 | 0.95 |
| South East Asian | -0.53 | 0.58 | 0.59 | 0.19 | 1.82 |
| Townsend Quintile compared to $1^{th}$ Quintile | | | | | |
| $2^{nd}$ | 0.09 | 0.03 | 1.09 | 1.04 | 1.15 |
| $3^{rd}$ | 0.11 | 0.03 | 1.12 | 1.06 | 1.18 |
| $4^{th}$ | 0.14 | 0.03 | 1.15 | 1.09 | 1.21 |
| $5^{th}$ | 0.17 | 0.03 | 1.19 | 1.11 | 1.26 |
| Diabetes Mellitus | 0.17 | 0.02 | 1.19 | 1.13 | 1.25 |
| Heart Failure | 0.52 | 0.03 | 1.67 | 1.58 | 1.78 |
| Atrial Fibrillation | 0.21 | 0.03 | 1.24 | 1.17 | 1.31 |
| Ever Smoked | 0.23 | 0.02 | 1.26 | 1.20 | 1.32 |
| $(Systolic\ BP/100)^2$ | -0.63 | 0.11 | 0.53 | 0.43 | 0.65 |
| $(Systolic\ BP/100)^2 \times log(Systolic\ BP/100)$ | 0.63 | 0.11 | 1.88 | 1.50 | 2.35 |
| $(BMI/10)^3$ | 0.01 | 0.00 | 1.01 | 1.00 | 1.01 |
| Log (BMI/10) | -1.04 | 0.10 | 0.35 | 0.29 | 0.43 |
| $(Hb/10)^3$ | -0.68 | 0.05 | 0.51 | 0.46 | 0.55 |
| $(Hb/10)^3 X\ log(Hb/10)$ | 0.76 | 0.07 | 2.13 | 1.87 | 2.43 |
| $(Gfr/100)^2$ | -1.63 | 0.13 | 0.19 | 0.15 | 0.25 |
| $Cholesterol/10$ | -0.44 | 0.08 | 0.64 | 0.55 | 0.75 |
| Proteinuria levels with none as reference | | | | | |
| High | 0.17 | 0.03 | 1.19 | 1.11 | 1.27 |
| Very High | 0.48 | 0.04 | 1.61 | 1.48 | 1.75 |
| Aspirin | 0.21 | 0.02 | 1.24 | 1.19 | 1.29 |
| Anticoagulation | 0.21 | 0.04 | 1.23 | 1.14 | 1.33 |
| Diuretic use | -0.16 | 0.02 | 0.85 | 0.82 | 0.89 |
| Other | 0.09 | 0.02 | 1.10 | 1.05 | 1.14 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.89 | 0.82 | 0.96 |
| Iron Medication | 0.12 | 0.03 | 1.13 | 1.05 | 1.20 |
| Vitamin D supplementation | 0.16 | 0.03 | 1.17 | 1.11 | 1.24 |

**Table 8. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 3**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| $Age^2$ | 3.37 | 0.07 | 29.00 | 25.48 | 33.01 |
| Male Gender | 0.50 | 0.02 | 1.65 | 1.58 | 1.72 |
| Race with African Caribbean Race as reference | | | | | |
| African –Caribbean | -0.46 | 0.16 | 0.63 | 0.46 | 0.86 |
| Indian Subcontinent | -0.26 | 0.12 | 0.77 | 0.61 | 0.97 |
| South East Asian | -0.47 | 0.58 | 0.63 | 0.20 | 1.95 |
| Townsend Quintile compared to $1^{th}$ Quintile | | | | | |
| $2^{nd}$ | 0.08 | 0.03 | 1.08 | 1.03 | 1.14 |
| $3^{rd}$ | 0.11 | 0.03 | 1.12 | 1.06 | 1.18 |
| $4^{th}$ | 0.15 | 0.03 | 1.16 | 1.10 | 1.23 |
| $5^{th}$ | 0.17 | 0.03 | 1.19 | 1.12 | 1.27 |
| Diabetes Mellitus | 0.08 | 0.02 | 1.08 | 1.04 | 1.13 |
| Heart Failure | 0.52 | 0.03 | 1.69 | 1.59 | 1.79 |
| Atrial Fibrillation | 0.22 | 0.03 | 1.25 | 1.18 | 1.32 |
| Ever Smoked | 0.24 | 0.02 | 1.27 | 1.22 | 1.34 |
| (Systolic BP/100) | -2.19 | 0.34 | 0.11 | 0.06 | 0.22 |
| $(Systolic\ BP/100)^2$ | 0.66 | 0.12 | 1.94 | 1.55 | 2.44 |
| $(BMI/10)^{-2}$ | 1.19 | 0.15 | 3.28 | 2.45 | 4.40 |
| $(Hb/10)^3$ | -0.70 | 0.05 | 0.50 | 0.46 | 0.55 |
| $(Hb/10)^3$ X log(Hb/10) | 0.75 | 0.07 | 2.12 | 1.86 | 2.41 |
| Gfr/100 | -1.46 | 0.11 | 0.23 | 0.19 | 0.29 |
| Cholesterol/10 | -0.27 | 0.08 | 0.76 | 0.65 | 0.89 |
| Proteinuria levels with none as reference | | | | | |
| High | 0.20 | 0.03 | 1.23 | 1.15 | 1.31 |
| Very High | 0.44 | 0.04 | 1.55 | 1.43 | 1.69 |
| Aspirin | 0.23 | 0.02 | 1.25 | 1.20 | 1.31 |
| Anticoagulation | 0.21 | 0.04 | 1.23 | 1.14 | 1.33 |
| Diuretic use | -0.16 | 0.02 | 0.85 | 0.82 | 0.89 |
| Other | 0.17 | 0.02 | 1.18 | 1.13 | 1.24 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.89 | 0.82 | 0.96 |
| Iron Medication | 0.11 | 0.03 | 1.12 | 1.05 | 1.20 |
| Vitamin D supplementation | 0.16 | 0.03 | 1.17 | 1.11 | 1.25 |

**Table 9. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 4**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| $Age^2$ | 3.34 | 0.07 | 28.31 | 24.84 | 32.26 |
| Male Gender | 0.52 | 0.02 | 1.67 | 1.61 | 1.74 |
| Race with African Caribbean Race as reference | | | | | |
| African –Caribbean | -0.45 | 0.16 | 0.64 | 0.47 | 0.87 |
| Indian Subcontinent | -0.26 | 0.12 | 0.77 | 0.61 | 0.97 |
| South East Asian | -0.67 | 0.58 | 0.51 | 0.16 | 1.58 |
| Townsend Quintile compared to $1^{th}$ Quintile | | | | | |
| $2^{nd}$ | 0.10 | 0.03 | 1.10 | 1.05 | 1.16 |
| $3^{rd}$ | 0.14 | 0.03 | 1.15 | 1.09 | 1.21 |
| $4^{th}$ | 0.15 | 0.03 | 1.17 | 1.10 | 1.23 |
| $5^{th}$ | 0.19 | 0.03 | 1.21 | 1.14 | 1.29 |
| Diabetes Mellitus | 0.18 | 0.02 | 1.20 | 1.15 | 1.26 |
| Heart Failure | 0.52 | 0.03 | 1.69 | 1.59 | 1.79 |
| Atrial Fibrillation | 0.22 | 0.03 | 1.24 | 1.17 | 1.31 |
| Ever Smoked | 0.23 | 0.02 | 1.26 | 1.21 | 1.32 |
| (Systolic BP/100) | -2.10 | 0.35 | 0.12 | 0.06 | 0.24 |
| $(Systolic\ BP/100)^2$ | 0.64 | 0.12 | 1.89 | 1.51 | 2.38 |
| $(BMI/10)^{-0.5}$ | 2.73 | 0.29 | 15.41 | 8.80 | 26.97 |
| $(BMI/10)^3$ | 0.01 | 0.00 | 1.01 | 1.00 | 1.01 |
| $(Hb/10)^3$ | -0.70 | 0.04 | 0.50 | 0.46 | 0.54 |
| $(Hb/10)^3$ X log(Hb/10) | 0.77 | 0.06 | 2.17 | 1.92 | 2.46 |
| $Gfr/100$ | -1.42 | 0.11 | 0.24 | 0.19 | 0.30 |
| Proteinuria levels with none as reference | | | | | |
| High | 0.13 | 0.03 | 1.14 | 1.07 | 1.22 |
| Very High | 0.41 | 0.04 | 1.51 | 1.39 | 1.65 |
| Aspirin | 0.23 | 0.02 | 1.26 | 1.21 | 1.31 |
| Anticoagulation | 0.21 | 0.04 | 1.24 | 1.15 | 1.34 |
| Diuretic use | -0.15 | 0.02 | 0.86 | 0.83 | 0.90 |
| Other | 0.08 | 0.02 | 1.09 | 1.05 | 1.13 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.89 | 0.82 | 0.96 |
| Iron Medication | 0.14 | 0.03 | 1.15 | 1.07 | 1.23 |
| Vitamin D supplementation | 0.16 | 0.03 | 1.18 | 1.11 | 1.25 |

**Table 10. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 5**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| Age$^{2}$ | 3.37 | 0.07 | 29.01 | 25.47 | 33.03 |
| Male Gender | 0.49 | 0.02 | 1.63 | 1.57 | 1.70 |
| Race with African Caribbean Race as reference<br>African –Caribbean | -0.46 | 0.16 | 0.63 | 0.47 | 0.86 |
| Indian Subcontinent | -0.27 | 0.12 | 0.76 | 0.61 | 0.96 |
| South East Asian | -0.73 | 0.58 | 0.48 | 0.16 | 1.50 |
| Townsend Quintile compared to 1$^{th}$ Quintile<br>2$^{nd}$ | 0.10 | 0.03 | 1.10 | 1.04 | 1.16 |
| 3$^{rd}$ | 0.12 | 0.03 | 1.13 | 1.07 | 1.19 |
| 4$^{th}$ | 0.14 | 0.03 | 1.15 | 1.09 | 1.21 |
| 5$^{th}$ | 0.19 | 0.03 | 1.21 | 1.14 | 1.29 |
| Diabetes Mellitus | 0.08 | 0.02 | 1.09 | 1.04 | 1.13 |
| Heart Failure | 0.52 | 0.03 | 1.69 | 1.59 | 1.79 |
| Atrial Fibrillation | 0.21 | 0.03 | 1.24 | 1.17 | 1.31 |
| Ever Smoked | 0.24 | 0.02 | 1.27 | 1.21 | 1.33 |
| (Systolic BP/100)$^{1}$ | -3.23 | 0.46 | 0.04 | 0.02 | 0.10 |
| (Systolic BP/100)$^{1}$ x log(Systolic BP/100) | 2.18 | 0.34 | 8.87 | 4.58 | 17.19 |
| (BMI/10)-$^{2}$ | 1.44 | 0.15 | 4.20 | 3.14 | 5.62 |
| (Hb/10)$^{3}$ | -0.68 | 0.05 | 0.51 | 0.46 | 0.55 |
| (Hb/10)$^{3}$ X log(Hb/10) | 0.75 | 0.07 | 2.13 | 1.87 | 2.42 |
| Gfr/100 | -1.35 | 0.11 | 0.26 | 0.21 | 0.32 |
| Cholesterol/10 | -0.34 | 0.08 | 0.72 | 0.61 | 0.84 |
| Proteinuria levels with none as reference<br>High | 0.01 | 0.03 | 1.01 | 0.95 | 1.07 |
| Very High | 0.41 | 0.04 | 1.51 | 1.39 | 1.63 |
| Aspirin | 0.22 | 0.02 | 1.24 | 1.19 | 1.30 |
| Anticoagulation | 0.21 | 0.04 | 1.23 | 1.14 | 1.33 |
| Diuretic use | -0.16 | 0.02 | 0.85 | 0.82 | 0.88 |
| Other | 0.14 | 0.02 | 1.15 | 1.09 | 1.20 |
| Angiotensin Blockade | -0.13 | 0.04 | 0.88 | 0.82 | 0.95 |
| Iron Medication | 0.10 | 0.03 | 1.11 | 1.04 | 1.19 |
| Vitamin D supplementation | 0.16 | 0.03 | 1.18 | 1.11 | 1.25 |

**Table 11. Cox regression analysis for Imputation 1 without transformation but with frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| Age | 0.05 | 0.00 | 1.05 | 1.05 | 1.05 |
| Male Gender | 0.49 | 0.02 | 1.64 | 1.57 | 1.70 |
| Race with African Caribbean Race as reference African – Caribbean | -0.37 | 0.16 | 0.69 | 0.50 | 0.95 |
| Indian Subcontinent | -0.21 | 0.12 | 0.81 | 0.64 | 1.03 |
| South East Asian | -0.51 | 0.58 | 0.60 | 0.19 | 1.87 |
| Townsend Quintile compared to 1$^{th}$ Quintile 2$^{nd}$ | 0.08 | 0.03 | 1.08 | 1.03 | 1.14 |
| 3$^{rd}$ | 0.11 | 0.03 | 1.12 | 1.06 | 1.18 |
| 4$^{th}$ | 0.14 | 0.03 | 1.15 | 1.09 | 1.22 |
| 5$^{th}$ | 0.20 | 0.03 | 1.22 | 1.14 | 1.30 |
| Diabetes Mellitus | 0.17 | 0.03 | 1.19 | 1.13 | 1.25 |
| Heart Failure | 0.55 | 0.03 | 1.73 | 1.63 | 1.84 |
| Atrial Fibrillation | 0.23 | 0.03 | 1.26 | 1.19 | 1.34 |
| Ever Smoked | 0.24 | 0.03 | 1.27 | 1.20 | 1.33 |
| Systolic BP | 0.00 | 0.00 | 0.998 | 0.997 | 0.999 |
| BMI | -0.02 | 0.00 | 0.98 | 0.98 | 0.98 |
| Haemoglobin | -0.12 | 0.01 | 0.89 | 0.88 | 0.90 |
| Glomerular Filtration Rate | -0.01 | 0.00 | 0.99 | 0.98 | 0.99 |
| Cholesterol | -0.03 | 0.01 | 0.97 | 0.96 | 0.99 |
| Proteinuria levels with none as reference High | 0.16 | 0.03 | 1.18 | 1.10 | 1.26 |
| Very High | 0.52 | 0.04 | 1.68 | 1.54 | 1.82 |
| Aspirin | 0.21 | 0.02 | 1.24 | 1.19 | 1.29 |
| Anticoagulation | 0.19 | 0.04 | 1.21 | 1.12 | 1.31 |
| Diuretic use | -0.17 | 0.02 | 0.84 | 0.81 | 0.88 |
| Other | 0.07 | 0.02 | 1.08 | 1.04 | 1.12 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.89 | 0.82 | 0.96 |
| Iron Medication | 0.14 | 0.03 | 1.15 | 1.07 | 1.23 |
| Vitamin D supplementation | 0.17 | 0.03 | 1.18 | 1.11 | 1.25 |

**Table 12. Cox regression analysis for Imputation 2 without transformation but with frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| Age | 0.05 | 0.00 | 1.05 | 1.05 | 1.05 |
| Male Gender | 0.49 | 0.02 | 1.62 | 1.56 | 1.69 |
| Race with African Caribbean<br>Race as reference<br>African –Caribbean | -0.39 | 0.16 | 0.68 | 0.50 | 0.93 |
| Indian Subcontinent | -0.23 | 0.12 | 0.80 | 0.63 | 1.01 |
| South East Asian | -0.50 | 0.58 | 0.61 | 0.19 | 1.88 |
| Townsend Quintile compared to $1^{th}$ Quintile<br>$2^{nd}$ | 0.08 | 0.03 | 1.09 | 1.03 | 1.14 |
| $3^{rd}$ | 0.11 | 0.03 | 1.11 | 1.05 | 1.18 |
| $4^{th}$ | 0.13 | 0.03 | 1.14 | 1.08 | 1.21 |
| $5^{th}$ | 0.17 | 0.03 | 1.19 | 1.11 | 1.27 |
| Diabetes Mellitus | 0.17 | 0.02 | 1.18 | 1.12 | 1.24 |
| Heart Failure | 0.54 | 0.03 | 1.72 | 1.62 | 1.83 |
| Atrial Fibrillation | 0.23 | 0.03 | 1.26 | 1.19 | 1.33 |
| Ever Smoked | 0.23 | 0.03 | 1.26 | 1.20 | 1.32 |
| Systolic BP | 0.00 | 0.00 | 0.998 | 0.997 | 0.999 |
| BMI | -0.02 | 0.00 | 0.98 | 0.98 | 0.98 |
| Haemoglobin | -0.11 | 0.01 | 0.89 | 0.88 | 0.90 |
| Glomerular Filtration Rate | -0.01 | 0.00 | 0.99 | 0.98 | 0.99 |
| Cholesterol | -0.04 | 0.01 | 0.96 | 0.94 | 0.97 |
| Proteinuria levels with none as reference<br>High | 0.18 | 0.03 | 1.19 | 1.11 | 1.27 |
| Very High | 0.51 | 0.04 | 1.66 | 1.52 | 1.81 |
| Aspirin | 0.20 | 0.02 | 1.22 | 1.17 | 1.28 |
| Anticoagulation | 0.17 | 0.04 | 1.19 | 1.10 | 1.28 |
| Diuretic use | -0.17 | 0.02 | 0.84 | 0.81 | 0.88 |
| Other | 0.08 | 0.02 | 1.08 | 1.04 | 1.13 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.88 | 0.82 | 0.95 |
| Iron Medication | 0.15 | 0.03 | 1.16 | 1.09 | 1.25 |
| Vitamin D supplementation | 0.16 | 0.03 | 1.17 | 1.10 | 1.24 |

**Table 13. Cox regression analysis for Imputation 3 without transformation but with frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| Age | 0.05 | 0.00 | 1.05 | 1.05 | 1.05 |
| Male Gender | 0.49 | 0.02 | 1.64 | 1.58 | 1.71 |
| Race with African Caribbean<br>Race as reference<br>African –Caribbean | -0.38 | 0.16 | 0.68 | 0.50 | 0.93 |
| Indian Subcontinent | -0.22 | 0.12 | 0.81 | 0.63 | 1.02 |
| South East Asian | -0.43 | 0.58 | 0.65 | 0.21 | 2.01 |
| Townsend Quintile compared to $1^{th}$ Quintile<br>$2^{nd}$ | 0.07 | 0.03 | 1.07 | 1.02 | 1.13 |
| $3^{rd}$ | 0.11 | 0.03 | 1.12 | 1.05 | 1.18 |
| $4^{th}$ | 0.14 | 0.03 | 1.15 | 1.09 | 1.22 |
| $5^{th}$ | 0.17 | 0.03 | 1.19 | 1.11 | 1.27 |
| Diabetes Mellitus | 0.16 | 0.02 | 1.17 | 1.11 | 1.23 |
| Heart Failure | 0.55 | 0.03 | 1.74 | 1.63 | 1.84 |
| Atrial Fibrillation | 0.24 | 0.03 | 1.27 | 1.20 | 1.34 |
| Ever Smoked | 0.24 | 0.03 | 1.28 | 1.21 | 1.34 |
| Systolic BP | 0.00 | 0.00 | 0.997 | 0.996 | 0.998 |
| BMI | -0.01 | 0.00 | 0.99 | 0.98 | 0.99 |
| Haemoglobin | -0.12 | 0.01 | 0.89 | 0.88 | 0.90 |
| Glomerular Filtration Rate | -0.02 | 0.00 | 0.99 | 0.98 | 0.99 |
| Cholesterol | -0.03 | 0.01 | 0.97 | 0.96 | 0.99 |
| Proteinuria levels with none as reference<br>High | 0.22 | 0.03 | 1.24 | 1.16 | 1.33 |
| Very High | 0.46 | 0.04 | 1.59 | 1.46 | 1.73 |
| Aspirin | 0.22 | 0.02 | 1.24 | 1.19 | 1.29 |
| Anticoagulation | 0.17 | 0.04 | 1.19 | 1.10 | 1.28 |
| Diuretic use | -0.17 | 0.02 | 0.84 | 0.81 | 0.88 |
| Other | 0.07 | 0.02 | 1.07 | 1.03 | 1.11 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.89 | 0.82 | 0.96 |
| Iron Medication | 0.15 | 0.03 | 1.16 | 1.08 | 1.24 |
| Vitamin D supplementation | 0.16 | 0.03 | 1.17 | 1.11 | 1.25 |

**Table 14. Cox regression analysis for Imputation 4 without transformation but with frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| Age | 0.05 | 0.00 | 1.05 | 1.05 | 1.05 |
| Male Gender | 0.49 | 0.02 | 1.64 | 1.58 | 1.71 |
| Race with African Caribbean<br>Race as reference<br>African –Caribbean | -0.38 | 0.16 | 0.69 | 0.50 | 0.94 |
| Indian Subcontinent | -0.22 | 0.12 | 0.80 | 0.63 | 1.02 |
| South East Asian | -0.64 | 0.58 | 0.53 | 0.17 | 1.64 |
| Townsend Quintile compared to $1^{th}$ Quintile<br>$2^{nd}$ | 0.09 | 0.03 | 1.10 | 1.04 | 1.16 |
| $3^{rd}$ | 0.13 | 0.03 | 1.14 | 1.08 | 1.21 |
| $4^{th}$ | 0.15 | 0.03 | 1.16 | 1.09 | 1.23 |
| $5^{th}$ | 0.19 | 0.03 | 1.21 | 1.13 | 1.30 |
| Diabetes Mellitus | 0.16 | 0.03 | 1.17 | 1.12 | 1.23 |
| Heart Failure | 0.55 | 0.03 | 1.73 | 1.63 | 1.83 |
| Atrial Fibrillation | 0.23 | 0.03 | 1.26 | 1.19 | 1.33 |
| Ever Smoked | 0.23 | 0.03 | 1.26 | 1.20 | 1.33 |
| Systolic BP | 0.00 | 0.00 | 0.998 | 0.997 | 0.998 |
| BMI | -0.02 | 0.00 | 0.98 | 0.98 | 0.99 |
| Haemoglobin | -0.11 | 0.01 | 0.89 | 0.88 | 0.90 |
| Glomerular Filtration Rate | -0.01 | 0.00 | 0.99 | 0.98 | 0.99 |
| Cholesterol | -0.02 | 0.01 | 0.98 | 0.96 | 0.99 |
| Proteinuria levels with none as reference<br>High | 0.14 | 0.03 | 1.15 | 1.07 | 1.23 |
| Very High | 0.44 | 0.04 | 1.55 | 1.43 | 1.69 |
| Aspirin | 0.22 | 0.02 | 1.24 | 1.19 | 1.29 |
| Anticoagulation | 0.18 | 0.04 | 1.19 | 1.10 | 1.29 |
| Diuretic use | -0.16 | 0.02 | 0.85 | 0.82 | 0.89 |
| Other | 0.07 | 0.02 | 1.07 | 1.03 | 1.12 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.89 | 0.82 | 0.96 |
| Iron Medication | 0.17 | 0.03 | 1.18 | 1.11 | 1.26 |
| Vitamin D supplementation | 0.16 | 0.03 | 1.17 | 1.11 | 1.25 |

**Table 15. Cox regression analysis for Imputation 5 without transformation but with frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| Age | 0.05 | 0.00 | 1.05 | 1.05 | 1.05 |
| Male Gender | 0.49 | 0.02 | 1.63 | 1.56 | 1.69 |
| Race with African Caribbean<br>Race as reference<br>African –Caribbean | -0.39 | 0.16 | 0.68 | 0.49 | 0.93 |
| Indian Subcontinent | -0.22 | 0.12 | 0.80 | 0.63 | 1.02 |
| South East Asian | -0.68 | 0.58 | 0.51 | 0.16 | 1.58 |
| Townsend Quintile compared to $1^{th}$ Quintile<br>$2^{nd}$ | 0.09 | 0.03 | 1.09 | 1.04 | 1.16 |
| $3^{rd}$ | 0.12 | 0.03 | 1.12 | 1.06 | 1.19 |
| $4^{th}$ | 0.13 | 0.03 | 1.14 | 1.08 | 1.21 |
| $5^{th}$ | 0.19 | 0.03 | 1.21 | 1.14 | 1.30 |
| Diabetes Mellitus | 0.14 | 0.03 | 1.15 | 1.10 | 1.21 |
| Heart Failure | 0.55 | 0.03 | 1.73 | 1.63 | 1.84 |
| Atrial Fibrillation | 0.23 | 0.03 | 1.26 | 1.19 | 1.33 |
| Ever Smoked | 0.24 | 0.03 | 1.27 | 1.21 | 1.34 |
| Systolic BP | 0.00 | 0.00 | 0.997 | 0.996 | 0.998 |
| BMI | -0.02 | 0.00 | 0.98 | 0.98 | 0.99 |
| Haemoglobin | -0.11 | 0.01 | 0.90 | 0.89 | 0.91 |
| Glomerular Filtration Rate | -0.01 | 0.00 | 0.99 | 0.98 | 0.99 |
| Cholesterol | -0.03 | 0.01 | 0.97 | 0.95 | 0.98 |
| Proteinuria levels with none as reference<br>High | 0.01 | 0.03 | 1.01 | 0.95 | 1.08 |
| Very High | 0.43 | 0.04 | 1.54 | 1.42 | 1.67 |
| Aspirin | 0.22 | 0.02 | 1.24 | 1.19 | 1.30 |
| Anticoagulation | 0.18 | 0.04 | 1.20 | 1.11 | 1.30 |
| Diuretic use | -0.17 | 0.02 | 0.84 | 0.81 | 0.88 |
| Other | 0.07 | 0.02 | 1.07 | 1.03 | 1.12 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.88 | 0.82 | 0.95 |
| Iron Medication | 0.14 | 0.03 | 1.15 | 1.08 | 1.23 |
| Vitamin D supplementation | 0.17 | 0.03 | 1.18 | 1.11 | 1.25 |

**Table 16. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 1 with frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| $Age^2$ | 3.28 | 0.07 | 26.67 | 23.36 | 30.45 |
| Male Gender | 0.50 | 0.02 | 1.65 | 1.59 | 1.72 |
| Race with African Caribbean Race as reference<br>African –Caribbean | -0.40 | 0.16 | 0.67 | 0.49 | 0.92 |
| Indian Subcontinent | -0.21 | 0.12 | 0.81 | 0.64 | 1.03 |
| South East Asian | -0.54 | 0.58 | 0.58 | 0.19 | 1.82 |
| Townsend Quintile compared to $1^{th}$ Quintile<br>$2^{nd}$ | 0.08 | 0.03 | 1.08 | 1.02 | 1.14 |
| $3^{rd}$ | 0.11 | 0.03 | 1.11 | 1.05 | 1.18 |
| $4^{th}$ | 0.14 | 0.03 | 1.15 | 1.08 | 1.21 |
| $5^{th}$ | 0.19 | 0.03 | 1.21 | 1.13 | 1.30 |
| Diabetes Mellitus | 0.18 | 0.03 | 1.20 | 1.14 | 1.26 |
| Heart Failure | 0.53 | 0.03 | 1.69 | 1.59 | 1.80 |
| Atrial Fibrillation | 0.23 | 0.03 | 1.26 | 1.19 | 1.33 |
| Ever Smoked | 0.24 | 0.03 | 1.27 | 1.21 | 1.33 |
| $(Systolic\ BP/100)^{0.5}$ | -9.77 | 1.60 | 0.00 | 0.00 | 0.00 |
| $(Systolic\ BP/100)$ | 3.88 | 0.67 | 48.21 | 12.99 | 178.93 |
| $(BMI/10)^{-0.5}$ | 3.18 | 0.28 | 24.00 | 13.77 | 41.83 |
| $(BMI/10)^{3}$ | 0.01 | 0.00 | 1.01 | 1.00 | 1.01 |
| $(Hb/10)^{-1}$ | 3.96 | 0.34 | 52.69 | 27.10 | 102.45 |
| $(Hb/10)^{-2}$ | -1.27 | 0.17 | 0.28 | 0.20 | 0.39 |
| $Gfr/100$ | -1.33 | 0.11 | 0.26 | 0.21 | 0.33 |
| $Cholesterol/10$ | -0.29 | 0.08 | 0.75 | 0.64 | 0.88 |
| Proteinuria levels with none as reference<br>High | 0.16 | 0.03 | 1.18 | 1.10 | 1.26 |
| Very High | 0.51 | 0.04 | 1.67 | 1.54 | 1.82 |
| Aspirin | 0.22 | 0.02 | 1.25 | 1.20 | 1.30 |
| Anticoagulation | 0.21 | 0.04 | 1.23 | 1.14 | 1.33 |
| Diuretic use | -0.17 | 0.02 | 0.85 | 0.81 | 0.88 |
| Other | 0.08 | 0.02 | 1.08 | 1.04 | 1.12 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.89 | 0.82 | 0.96 |
| Iron Medication | 0.11 | 0.03 | 1.11 | 1.04 | 1.19 |
| Vitamin D supplementation | 0.15 | 0.03 | 1.16 | 1.10 | 1.23 |

**Table 17. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 2 with frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| $Age^2$ | 3.27 | 0.07 | 26.35 | 23.07 | 30.10 |
| Male Gender | 0.49 | 0.02 | 1.64 | 1.57 | 1.70 |
| Race with African Caribbean Race as reference | | | | | |
| African –Caribbean | -0.41 | 0.16 | 0.66 | 0.48 | 0.91 |
| Indian Subcontinent | -0.23 | 0.12 | 0.79 | 0.62 | 1.01 |
| South East Asian | -0.53 | 0.58 | 0.59 | 0.19 | 1.83 |
| Townsend Quintile compared to $1^{th}$ Quintile | | | | | |
| $2^{nd}$ | 0.08 | 0.03 | 1.08 | 1.03 | 1.14 |
| $3^{rd}$ | 0.10 | 0.03 | 1.11 | 1.05 | 1.17 |
| $4^{th}$ | 0.13 | 0.03 | 1.13 | 1.07 | 1.20 |
| $5^{th}$ | 0.16 | 0.03 | 1.18 | 1.10 | 1.26 |
| Diabetes Mellitus | 0.17 | 0.02 | 1.18 | 1.13 | 1.24 |
| Heart Failure | 0.52 | 0.03 | 1.69 | 1.59 | 1.79 |
| Atrial Fibrillation | 0.22 | 0.03 | 1.25 | 1.18 | 1.32 |
| Ever Smoked | 0.23 | 0.03 | 1.26 | 1.20 | 1.33 |
| $(Systolic\ BP/100)^2$ | -0.62 | 0.11 | 0.54 | 0.44 | 0.67 |
| $(Systolic\ BP/100)^2$ x log(Systolic BP/100) | 0.61 | 0.12 | 1.84 | 1.47 | 2.31 |
| $(BMI/10)^3$ | 0.01 | 0.00 | 1.01 | 1.00 | 1.01 |
| Log (BMI/10) | -1.06 | 0.10 | 0.35 | 0.28 | 0.42 |
| $(Hb/10)^3$ | -0.69 | 0.05 | 0.50 | 0.46 | 0.55 |
| $(Hb/10)^3$ X log(Hb/10) | 0.77 | 0.07 | 2.16 | 1.89 | 2.47 |
| $(Gfr/100)^2$ | -1.53 | 0.13 | 0.22 | 0.17 | 0.28 |
| Cholesterol/10 | -0.44 | 0.08 | 0.64 | 0.55 | 0.76 |
| Proteinuria levels with none as reference | | | | | |
| High | 0.17 | 0.03 | 1.19 | 1.11 | 1.27 |
| Very High | 0.49 | 0.04 | 1.63 | 1.50 | 1.78 |
| Aspirin | 0.21 | 0.02 | 1.23 | 1.18 | 1.28 |
| Anticoagulation | 0.18 | 0.04 | 1.20 | 1.11 | 1.29 |
| Diuretic use | -0.16 | 0.02 | 0.85 | 0.82 | 0.89 |
| Other | 0.08 | 0.02 | 1.09 | 1.05 | 1.13 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.89 | 0.82 | 0.96 |
| Iron Medication | 0.12 | 0.03 | 1.13 | 1.06 | 1.21 |
| Vitamin D supplementation | 0.15 | 0.03 | 1.16 | 1.10 | 1.23 |

**Table 18. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 3 with frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| $Age^2$ | 3.34 | 0.07 | 28.25 | 24.77 | 32.22 |
| Male Gender | 0.50 | 0.02 | 1.65 | 1.59 | 1.72 |
| Race with African Caribbean<br>Race as reference<br>African –Caribbean | -0.41 | 0.16 | 0.66 | 0.48 | 0.91 |
| Indian Subcontinent | -0.21 | 0.12 | 0.81 | 0.64 | 1.03 |
| South East Asian | -0.48 | 0.58 | 0.62 | 0.20 | 1.93 |
| Townsend Quintile compared to $1^{th}$ Quintile<br>$2^{nd}$ | 0.07 | 0.03 | 1.07 | 1.02 | 1.13 |
| $3^{rd}$ | 0.11 | 0.03 | 1.11 | 1.05 | 1.18 |
| $4^{th}$ | 0.14 | 0.03 | 1.15 | 1.08 | 1.22 |
| $5^{th}$ | 0.17 | 0.03 | 1.18 | 1.11 | 1.27 |
| Diabetes Mellitus | 0.16 | 0.02 | 1.18 | 1.12 | 1.24 |
| Heart Failure | 0.53 | 0.03 | 1.70 | 1.61 | 1.81 |
| Atrial Fibrillation | 0.23 | 0.03 | 1.26 | 1.19 | 1.33 |
| Ever Smoked | 0.24 | 0.03 | 1.28 | 1.21 | 1.34 |
| (Systolic BP/100) | -2.13 | 0.35 | 0.12 | 0.06 | 0.23 |
| $(Systolic\ BP/100)^2$ | 0.64 | 0.12 | 1.90 | 1.51 | 2.39 |
| $(BMI/10)^{-2}$ | 1.21 | 0.15 | 3.34 | 2.49 | 4.48 |
| $(Hb/10)^3$ | -0.71 | 0.05 | 0.49 | 0.45 | 0.54 |
| $(Hb/10)^3$ X log(Hb/10) | 0.77 | 0.07 | 2.17 | 1.90 | 2.48 |
| Gfr/100 | -1.37 | 0.11 | 0.25 | 0.20 | 0.32 |
| Cholesterol/10 | -0.28 | 0.08 | 0.76 | 0.64 | 0.89 |
| Proteinuria levels with none as reference<br>High | 0.21 | 0.03 | 1.24 | 1.16 | 1.32 |
| Very High | 0.45 | 0.04 | 1.57 | 1.44 | 1.71 |
| Aspirin | 0.22 | 0.02 | 1.24 | 1.19 | 1.30 |
| Anticoagulation | 0.18 | 0.04 | 1.20 | 1.11 | 1.29 |
| Diuretic use | -0.16 | 0.02 | 0.85 | 0.82 | 0.89 |
| Other | 0.07 | 0.02 | 1.07 | 1.03 | 1.12 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.89 | 0.82 | 0.96 |
| Iron Medication | 0.12 | 0.03 | 1.12 | 1.05 | 1.20 |
| Vitamin D supplementation | 0.15 | 0.03 | 1.16 | 1.09 | 1.23 |

**Table 19. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 4 with frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| $Age^2$ | 3.32 | 0.07 | 27.67 | 24.23 | 31.59 |
| Male Gender | 0.52 | 0.02 | 1.68 | 1.61 | 1.74 |
| Race with African Caribbean Race as reference<br>African –Caribbean | -0.40 | 0.16 | 0.67 | 0.49 | 0.92 |
| Indian Subcontinent | -0.21 | 0.12 | 0.81 | 0.64 | 1.03 |
| South East Asian | -0.67 | 0.58 | 0.51 | 0.16 | 1.59 |
| Townsend Quintile compared to $1^{th}$ Quintile<br>$2^{nd}$ | 0.09 | 0.03 | 1.10 | 1.04 | 1.16 |
| $3^{rd}$ | 0.13 | 0.03 | 1.14 | 1.08 | 1.21 |
| $4^{th}$ | 0.14 | 0.03 | 1.15 | 1.09 | 1.22 |
| $5^{th}$ | 0.18 | 0.03 | 1.20 | 1.12 | 1.29 |
| Diabetes Mellitus | 0.18 | 0.02 | 1.19 | 1.14 | 1.25 |
| Heart Failure | 0.53 | 0.03 | 1.70 | 1.60 | 1.81 |
| Atrial Fibrillation | 0.23 | 0.03 | 1.25 | 1.19 | 1.33 |
| Ever Smoked | 0.24 | 0.03 | 1.27 | 1.20 | 1.33 |
| (Systolic BP/100) | -2.05 | 0.35 | 0.13 | 0.07 | 0.25 |
| $(Systolic\ BP/100)^2$ | 0.62 | 0.12 | 1.86 | 1.48 | 2.33 |
| $(BMI/10)^{-0.5}$ | 2.79 | 0.29 | 16.36 | 9.33 | 28.69 |
| $(BMI/10)^3$ | 0.01 | 0.00 | 1.01 | 1.00 | 1.01 |
| $(Hb/10)^3$ | -0.71 | 0.05 | 0.49 | 0.45 | 0.54 |
| $(Hb/10)^3$ X log(Hb/10) | 0.79 | 0.07 | 2.20 | 1.93 | 2.50 |
| $Gfr/100$ | -1.33 | 0.11 | 0.26 | 0.21 | 0.33 |
| Proteinuria levels with none as reference<br>High | 0.14 | 0.03 | 1.15 | 1.07 | 1.23 |
| Very High | 0.43 | 0.04 | 1.54 | 1.41 | 1.67 |
| Aspirin | 0.23 | 0.02 | 1.25 | 1.20 | 1.31 |
| Anticoagulation | 0.19 | 0.04 | 1.21 | 1.12 | 1.31 |
| Diuretic use | -0.15 | 0.02 | 0.86 | 0.83 | 0.90 |
| Other | 0.07 | 0.02 | 1.08 | 1.04 | 1.12 |
| Angiotensin Blockade | -0.11 | 0.04 | 0.89 | 0.83 | 0.96 |
| Iron Medication | 0.14 | 0.03 | 1.15 | 1.08 | 1.23 |
| Vitamin D supplementation | 0.15 | 0.03 | 1.16 | 1.10 | 1.23 |

**Table 20. Cox proportional hazards model with fractional polynomials suggested by mfp algorithm for Imputed Dataset 5 with frailty term**

| Risk factor/Prescription | Beta | Standard Error | Hazard Ratio | 95% Confidence intervals Lower Limit | 95% Confidence intervals Lower Limit |
|---|---|---|---|---|---|
| $Age^2$ | 3.34 | 0.07 | 28.26 | 24.77 | 32.25 |
| Male Gender | 0.49 | 0.02 | 1.63 | 1.57 | 1.70 |
| Race with African Caribbean Race as reference<br>African –Caribbean | -0.41 | 0.16 | 0.66 | 0.48 | 0.91 |
| Indian Subcontinent | -0.22 | 0.12 | 0.80 | 0.63 | 1.02 |
| South East Asian | -0.74 | 0.58 | 0.48 | 0.15 | 1.48 |
| Townsend Quintile compared to $1^{th}$ Quintile<br>$2^{nd}$ | 0.09 | 0.03 | 1.09 | 1.04 | 1.15 |
| $3^{rd}$ | 0.11 | 0.03 | 1.12 | 1.06 | 1.19 |
| $4^{th}$ | 0.13 | 0.03 | 1.14 | 1.07 | 1.20 |
| $5^{th}$ | 0.19 | 0.03 | 1.21 | 1.13 | 1.29 |
| Diabetes Mellitus | 0.07 | 0.02 | 1.08 | 1.03 | 1.12 |
| Heart Failure | 0.53 | 0.03 | 1.70 | 1.60 | 1.81 |
| Atrial Fibrillation | 0.23 | 0.03 | 1.25 | 1.18 | 1.32 |
| Ever Smoked | 0.24 | 0.03 | 1.27 | 1.21 | 1.34 |
| $(Systolic\ BP/100)^1$ | -3.16 | 0.47 | 0.04 | 0.02 | 0.11 |
| $(Systolic\ BP/100)^1$ x log(Systolic BP/100) | 2.12 | 0.34 | 8.36 | 4.29 | 16.27 |
| $(BMI/10)^{-2}$ | 1.46 | 0.15 | 4.30 | 3.21 | 5.75 |
| $(Hb/10)^3$ | -0.69 | 0.05 | 0.50 | 0.46 | 0.55 |
| $(Hb/10)^3$ X log(Hb/10) | 0.77 | 0.07 | 2.16 | 1.89 | 2.46 |
| Gfr/100 | -1.27 | 0.11 | 0.28 | 0.22 | 0.35 |
| Cholesterol/10 | -0.33 | 0.08 | 0.72 | 0.61 | 0.84 |
| Proteinuria levels with none as reference<br>High | 0.01 | 0.03 | 1.01 | 0.95 | 1.08 |
| Very High | 0.42 | 0.04 | 1.52 | 1.40 | 1.65 |
| Aspirin | 0.21 | 0.02 | 1.24 | 1.18 | 1.29 |
| Anticoagulation | 0.19 | 0.04 | 1.21 | 1.12 | 1.30 |
| Diuretic use | -0.17 | 0.02 | 0.85 | 0.81 | 0.88 |
| Other | 0.13 | 0.02 | 1.14 | 1.09 | 1.20 |
| Angiotensin Blockade | -0.12 | 0.04 | 0.88 | 0.82 | 0.96 |
| Iron Medication | 0.11 | 0.03 | 1.12 | 1.04 | 1.19 |
| Vitamin D supplementation | 0.16 | 0.03 | 1.17 | 1.10 | 1.24 |

**Figure 6. Systolic Blood pressure versus log relative hazards ratio for imputations 1-5 with frailty term**

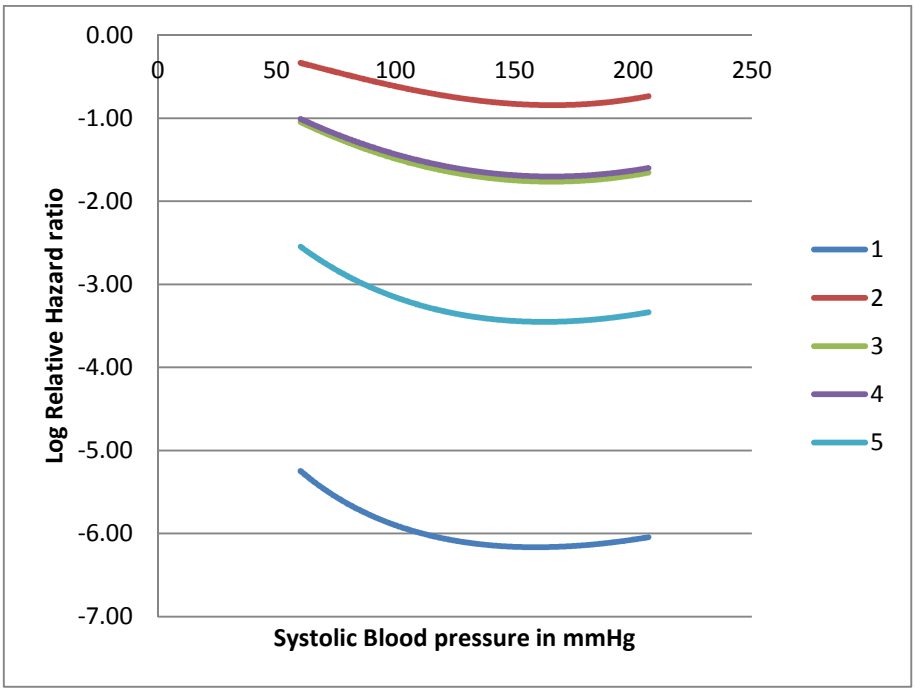



**Figure 7. Age versus log relative hazards ratio for imputations 1-5 with frailty term**

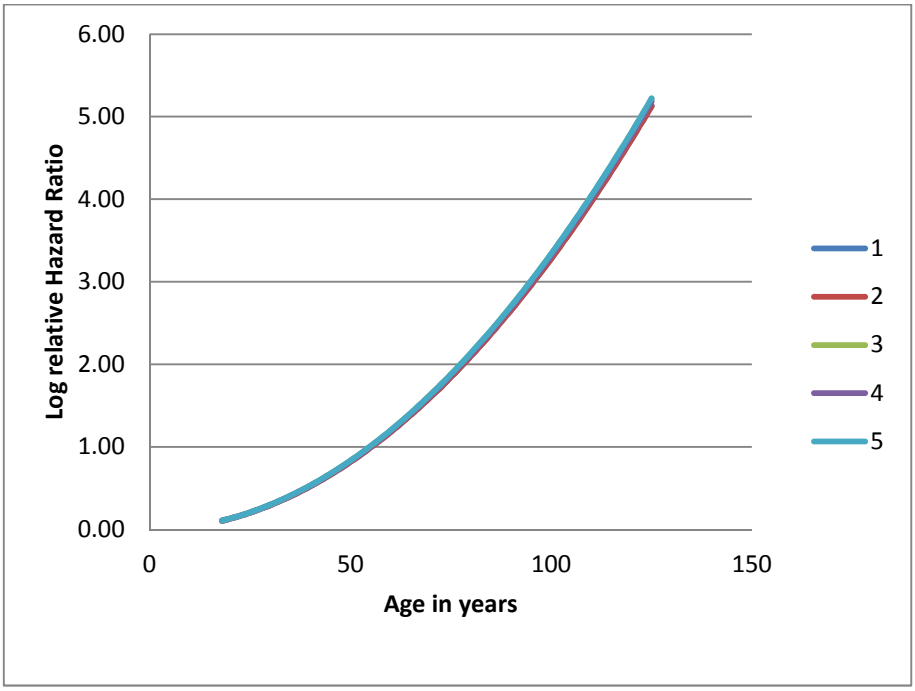


**Figure 8. Haemoglobin versus log relative hazards ratio for imputations 1-5 with frailty term**

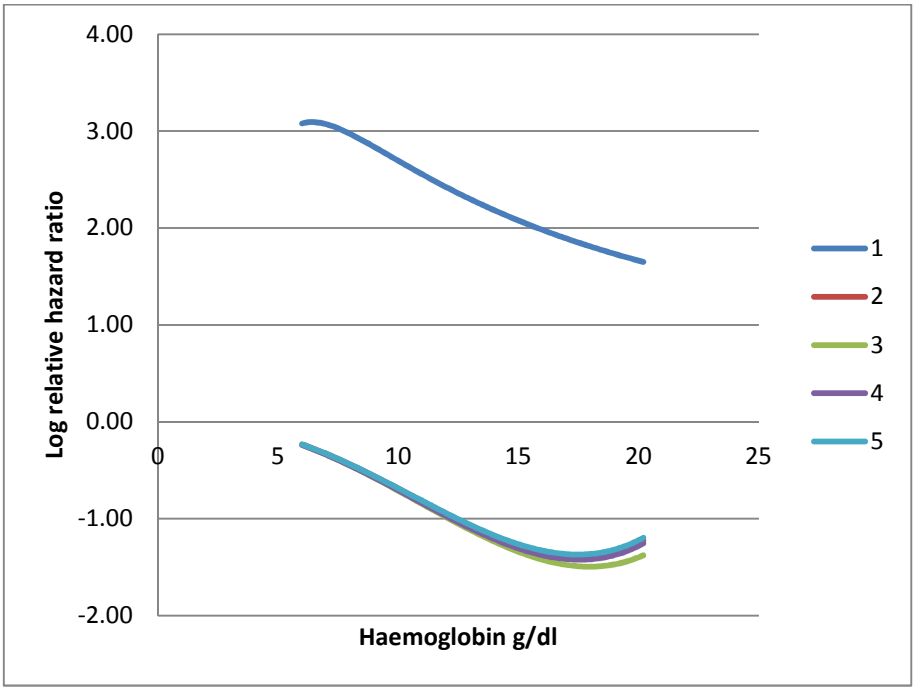
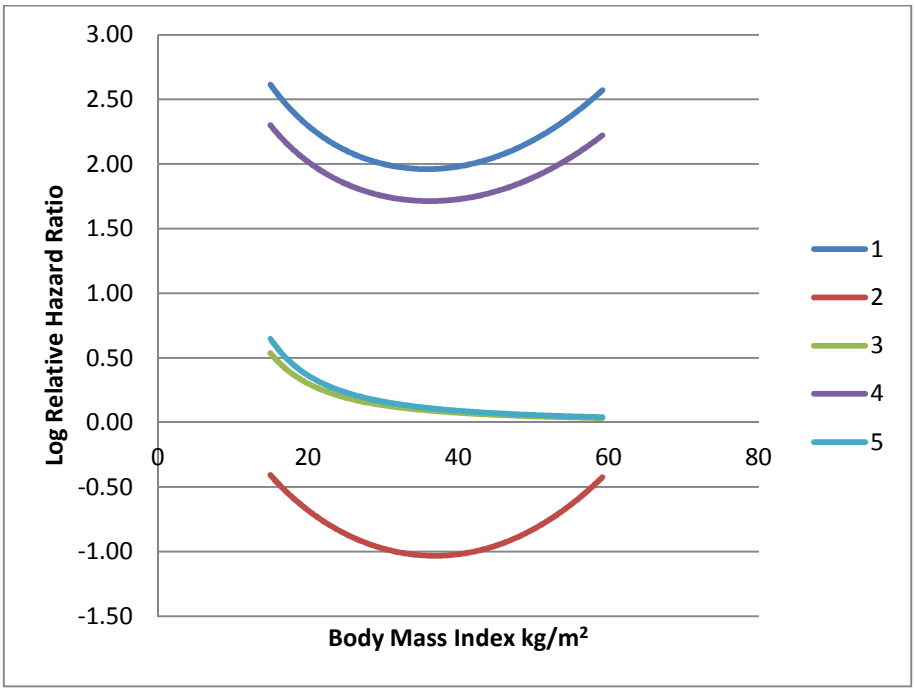



**Figure 9. Body Mass Index versus log relative hazards ratio for imputations 1-5 with frailty term**



# Appendix D. Schoenfeld residuals for Composite Model

**1. Scaled Schoenfeld residuals for age versus Time in days**

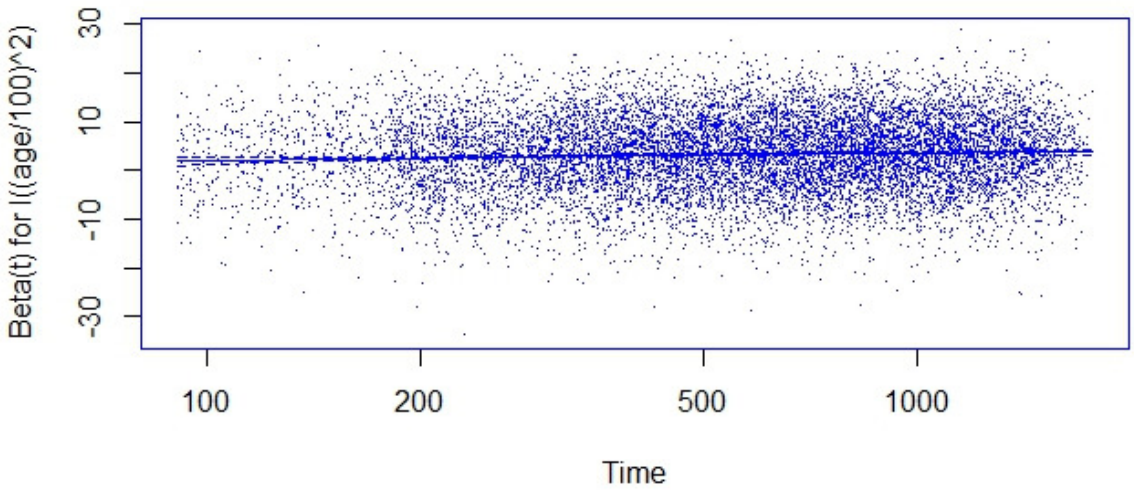



**Figure 2. Scaled Schoenfeld Residuals for$(Hb/10)^3$ versus Time**

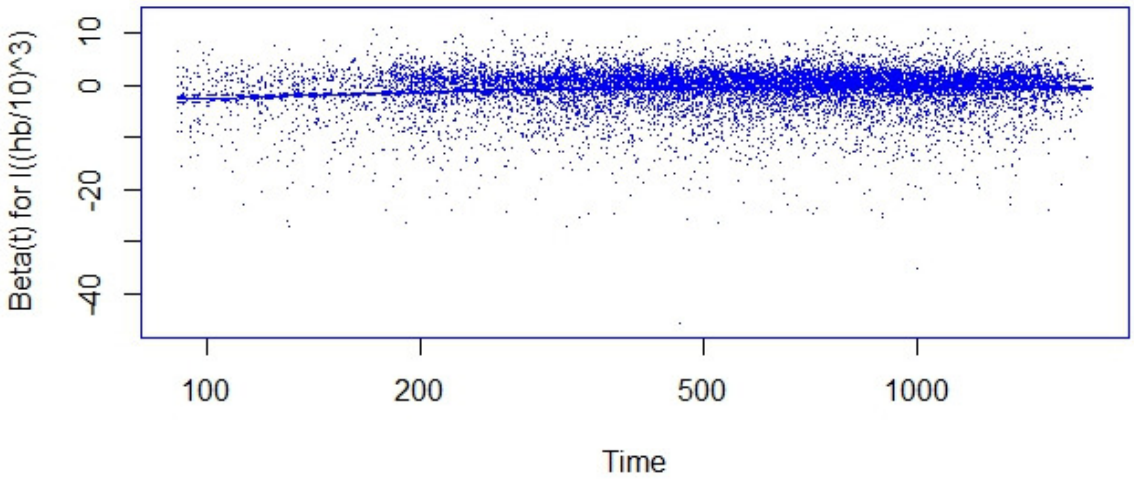


**Figure 3. Scaled Schoenfeld Residuals for Hb/10*3 x log Hb/10 versus Time**

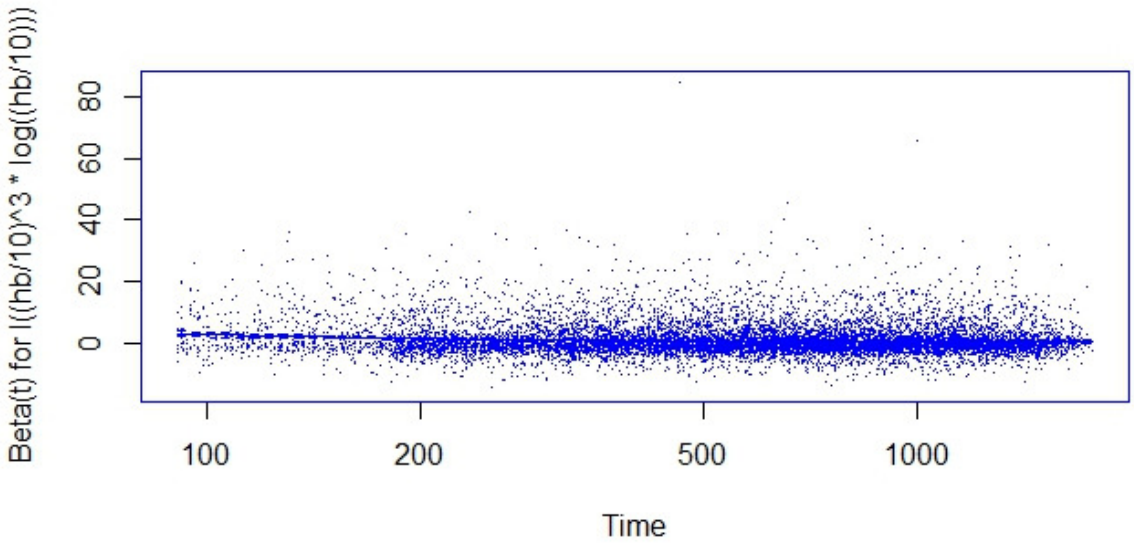



**Figure 4. Scaled Schoenfeld Residuals for Heart Failure versus Time**

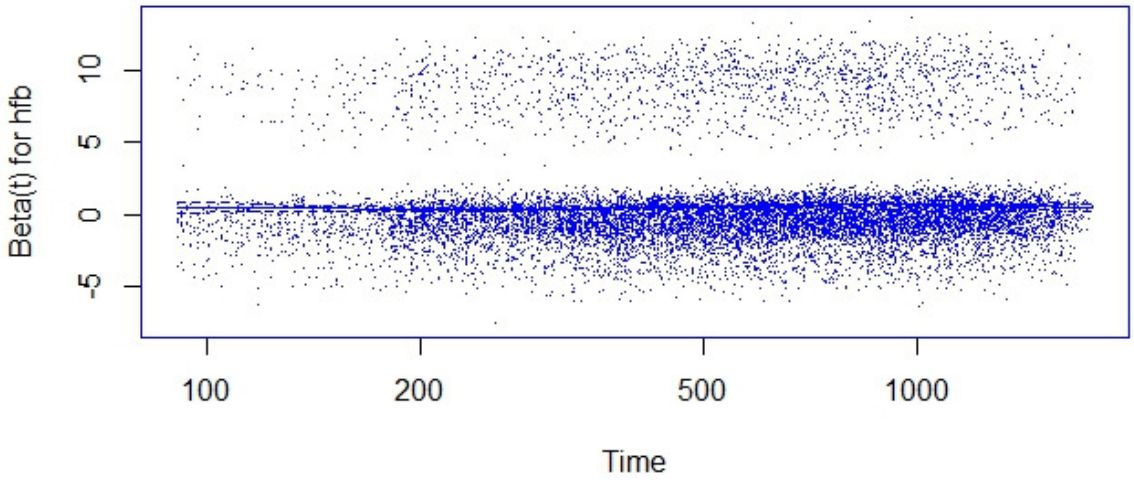


**Figure 5. Scaled Schoenfeld Residuals for Gender versus Time**

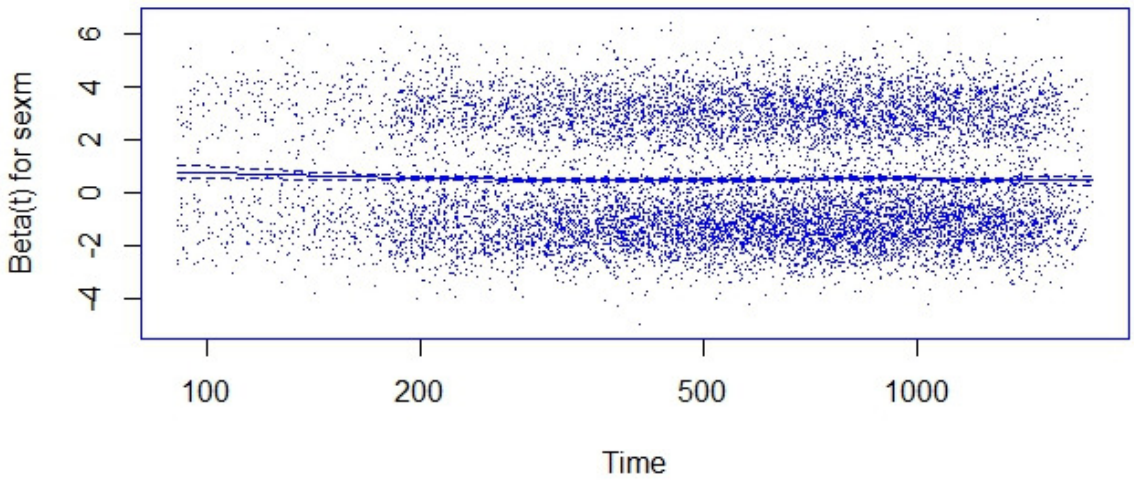



**Figure 6. Scaled Schoenfeld Residuals for Anticoagulation versus Time**

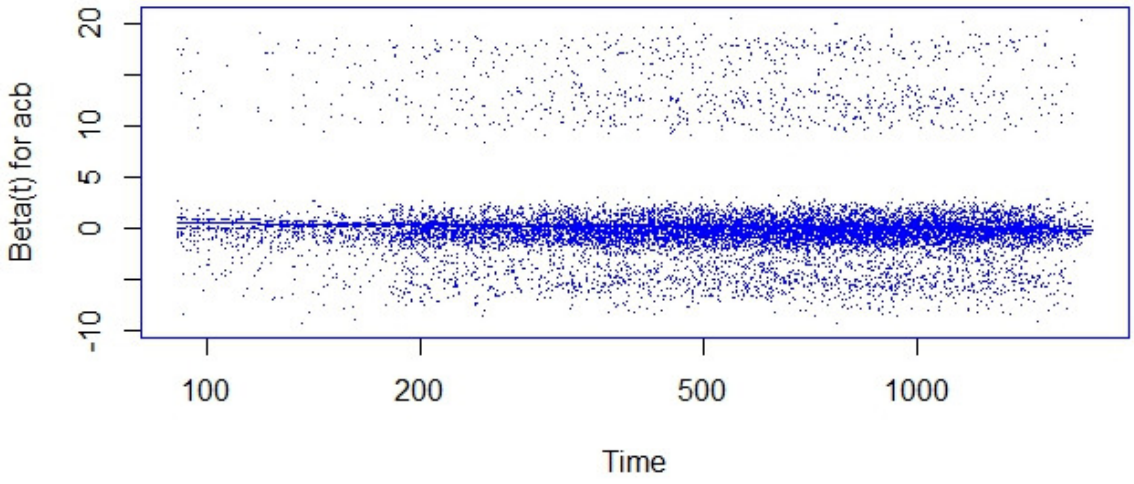


**Figure 7. Scaled Schoenfeld Residuals for Angiotensin blockade versus Time**

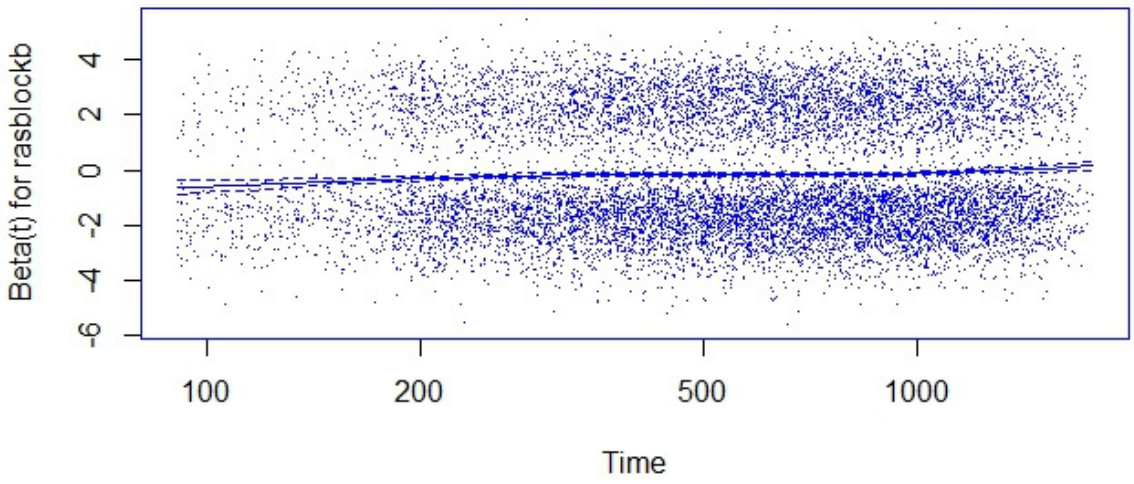



**Figure 8. Scaled Schoenfeld Residuals for Atrial Fibrillation versus Time**

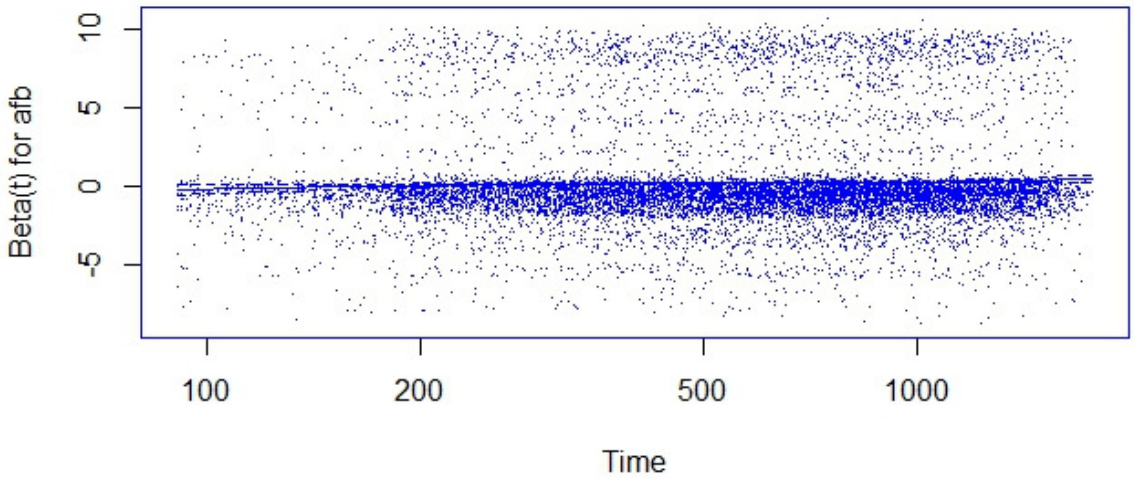


**Figure 9. Scaled Schoenfeld Residuals for AFC race versus Time**

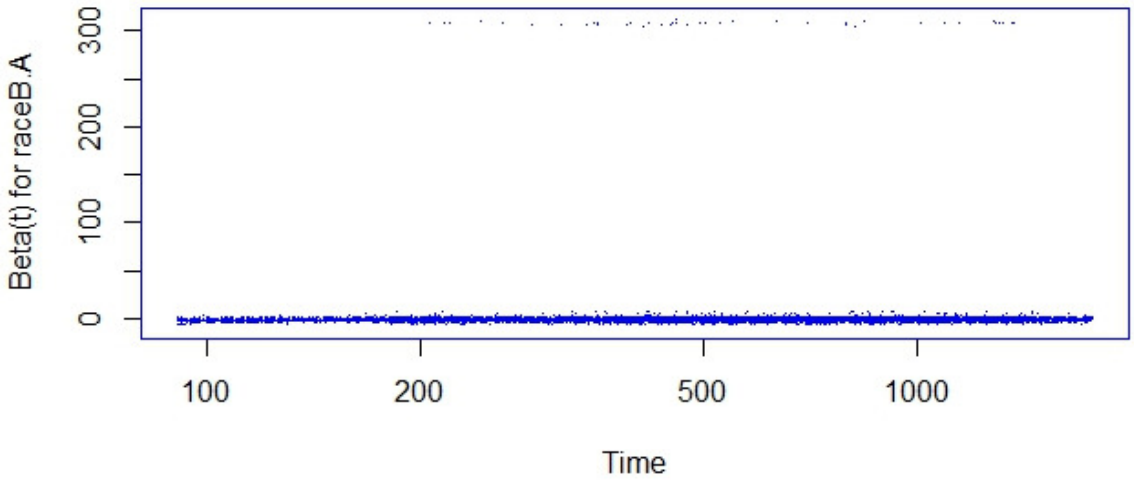



**Figure 10. Scaled Schoenfeld Residuals for Aspirin versus Time**

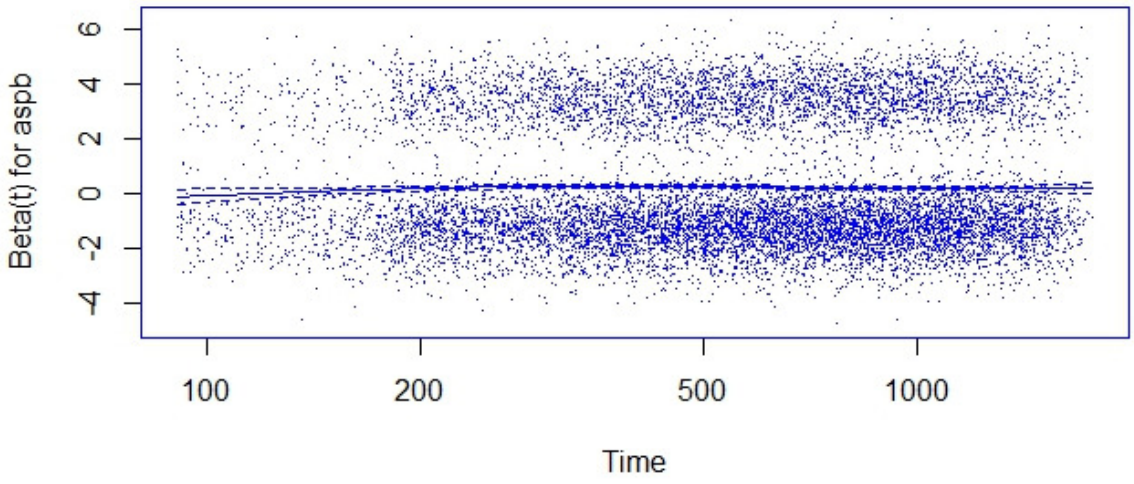


**Figure 11. Scaled Schoenfeld Residuals for BMI/10*-2 versus Time**

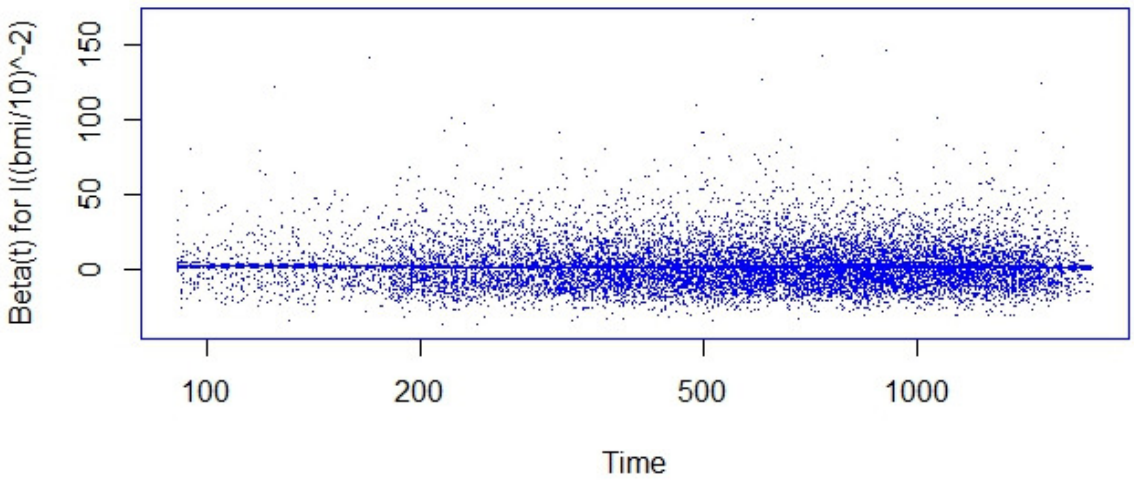



**Figure 12. Scaled Schoenfeld Residuals for Cholesterol/10 versus Time**

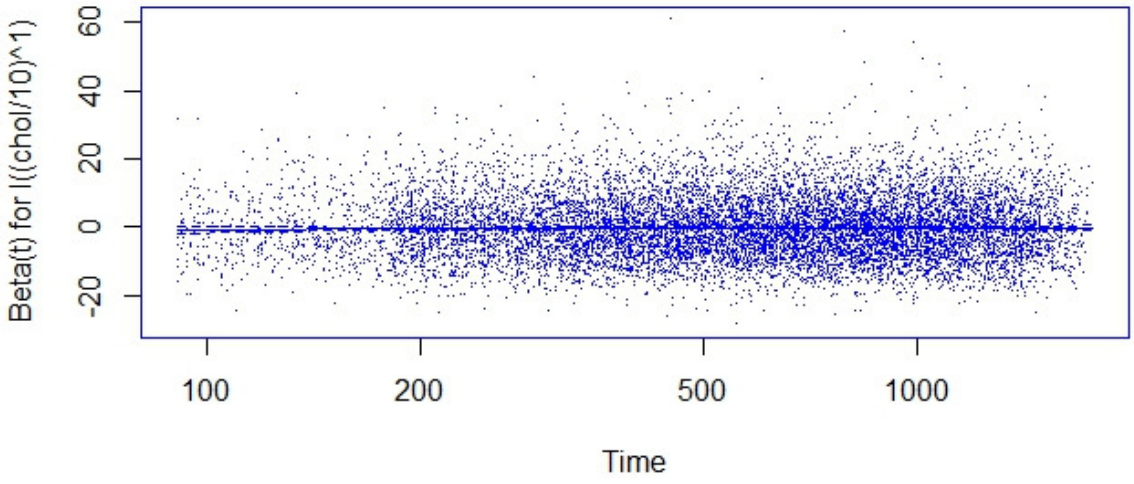


**Figure 13. Scaled Schoenfeld Residuals for Diuretics versus Time**

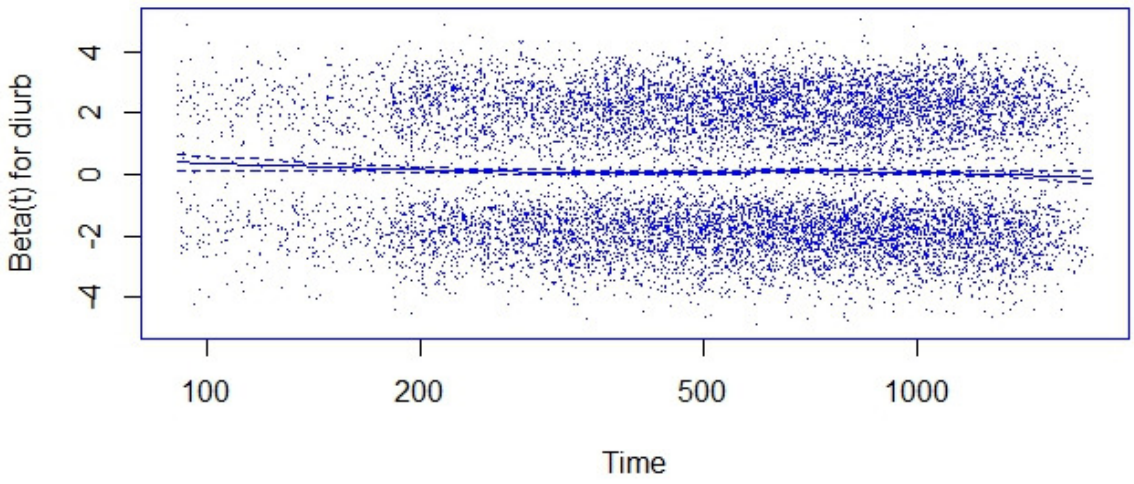



**Figure 14. Scaled Schoenfeld Residuals for Diabetes versus Time**

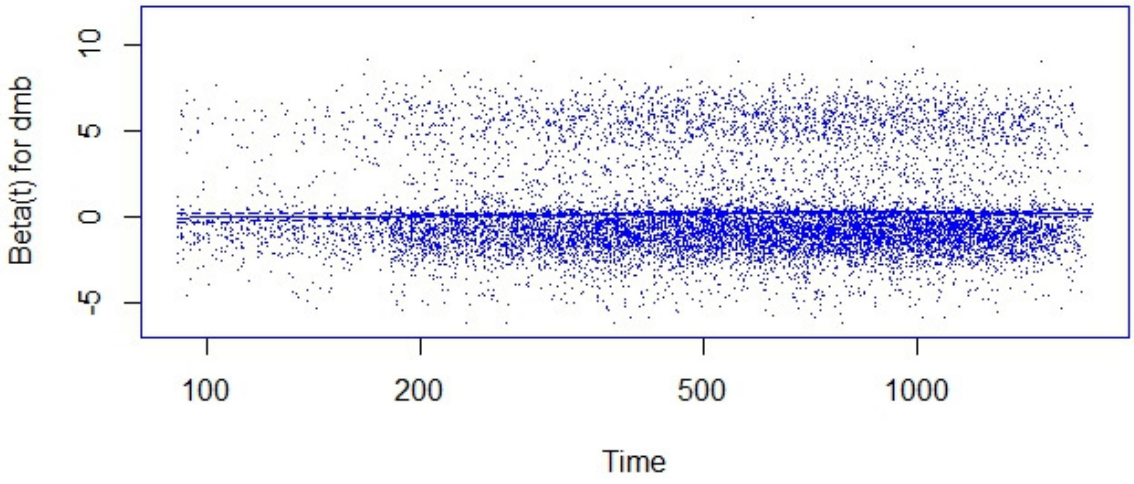


**Figure 15. Scaled Schoenfeld Residuals for iron supplements versus Time**

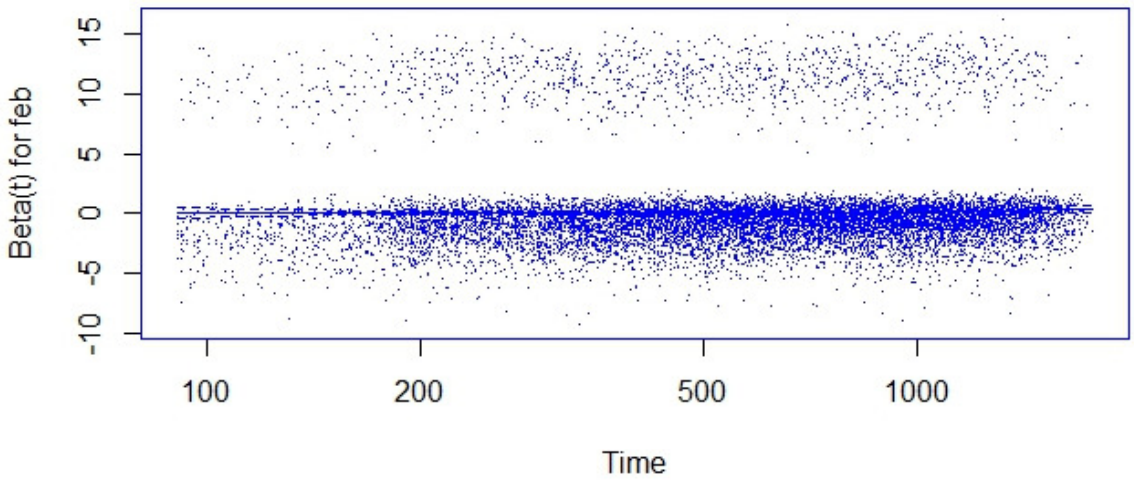



**Figure 16. Scaled Schoenfeld Residuals for other hypertensive versus Time**

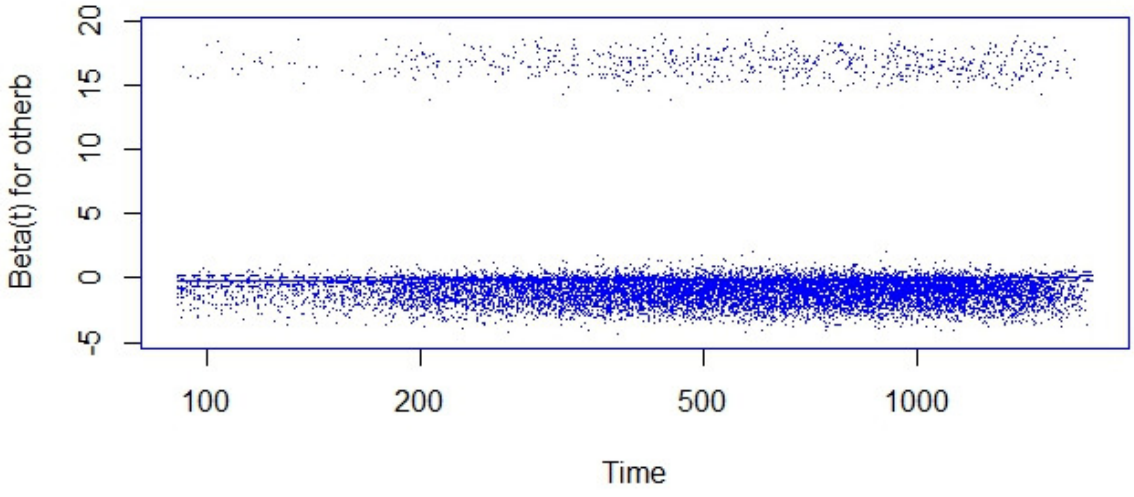


**Figure 17. Scaled Schoenfeld Residuals for High proteinuria versus Time**

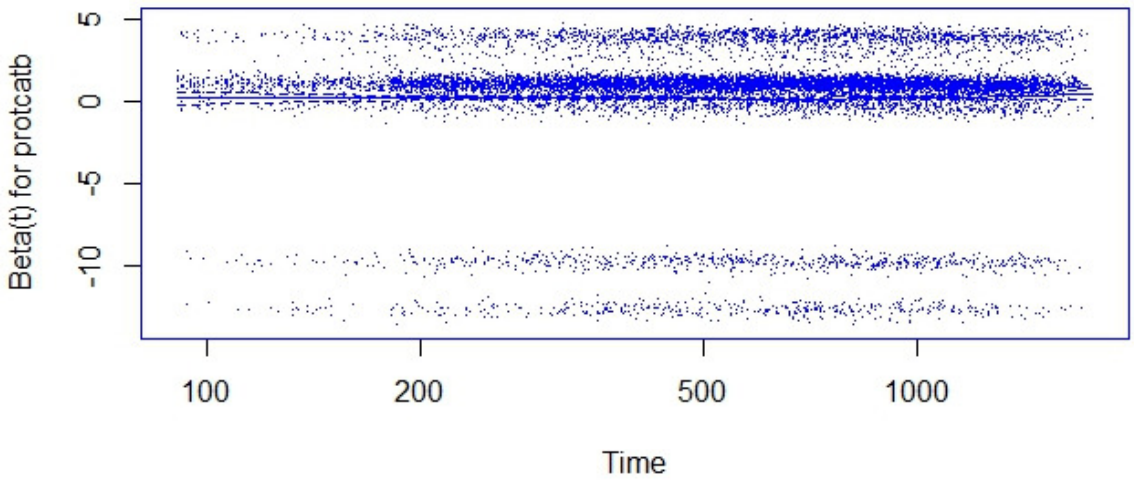



**Figure 18. Scaled Schoenfeld Residuals for Very high proteinuria versus Time.**

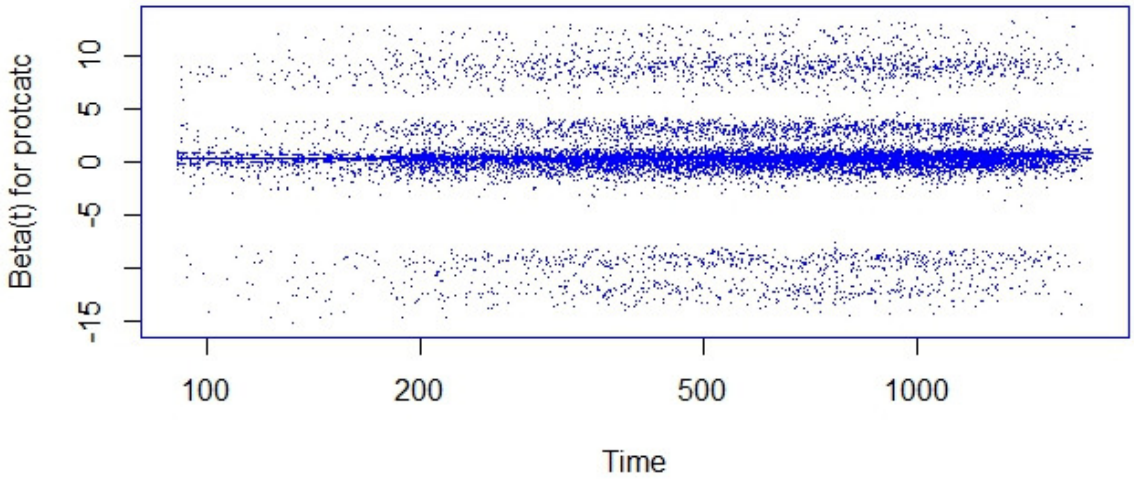


**Figure 19. Scaled Schoenfeld Residuals for Ex/Current Smokers versus Time**

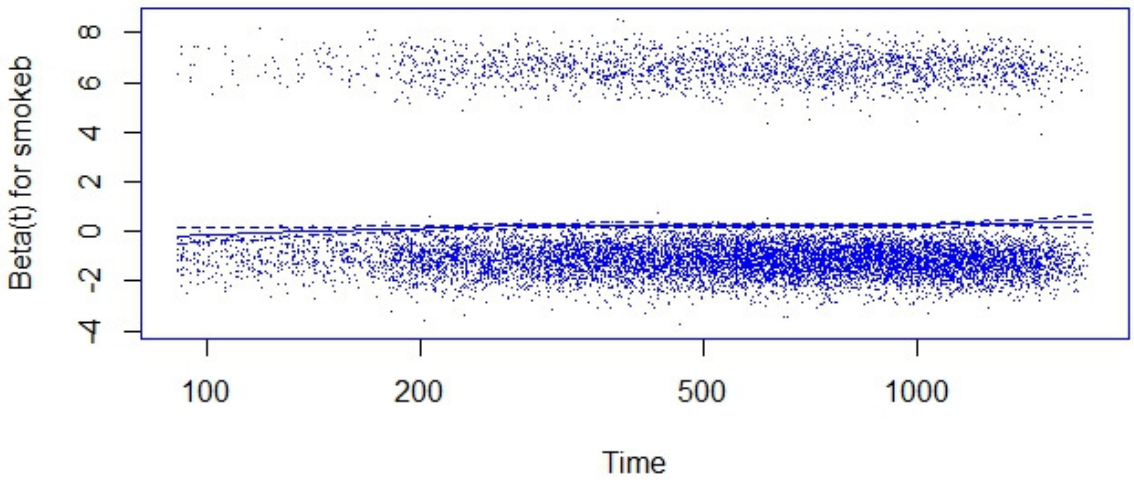



**Figure 20. Scaled Schoenfeld Residuals for sys bp/100 versus Time**

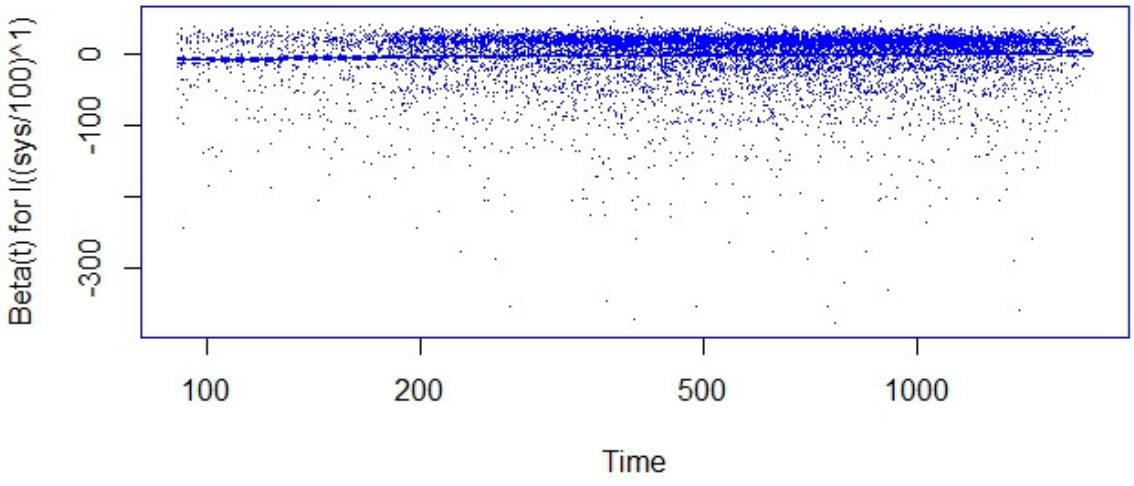


**Figure 21. Scaled Schoenfeld Residuals for Systolic  bp/100 squared versus Time**

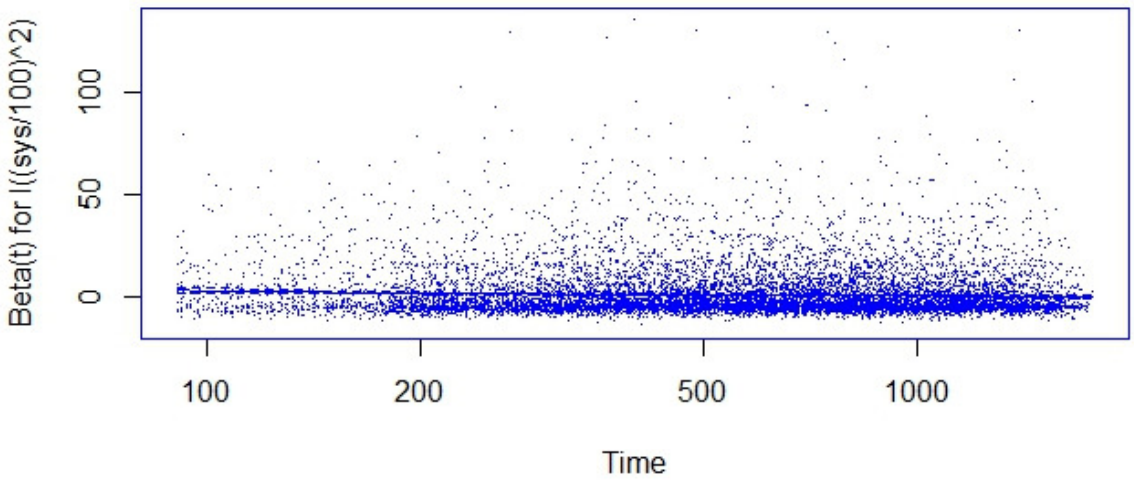



**Figure 22. . Scaled Schoenfeld Residuals for Townsend Quintile 2 versus Time**

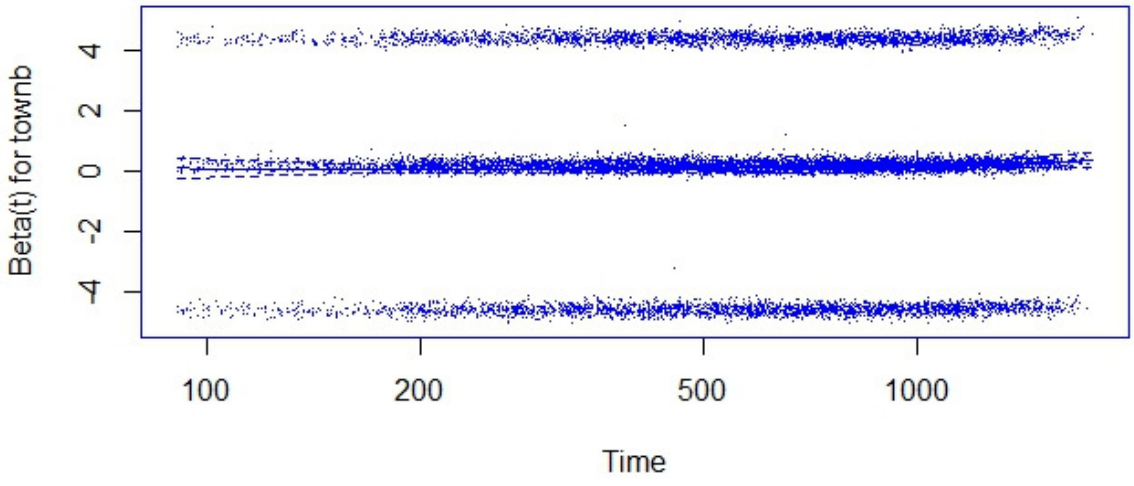


**Figure 23. Scaled Schoenfeld Residuals for Townsend Quintile 3 versus Time**

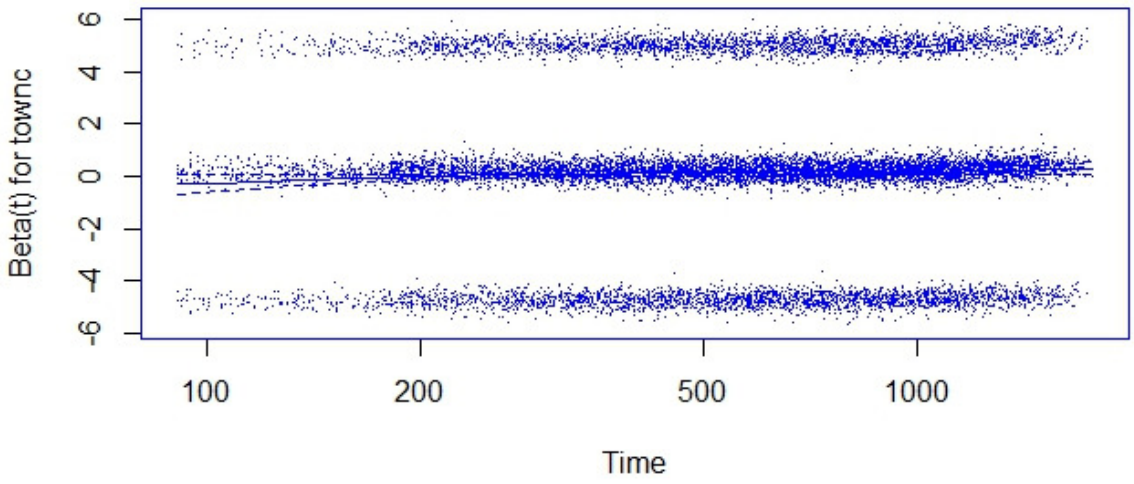



**Figure 24. Scaled Schoenfeld residuals for Townsend Quintile 4 versus time**

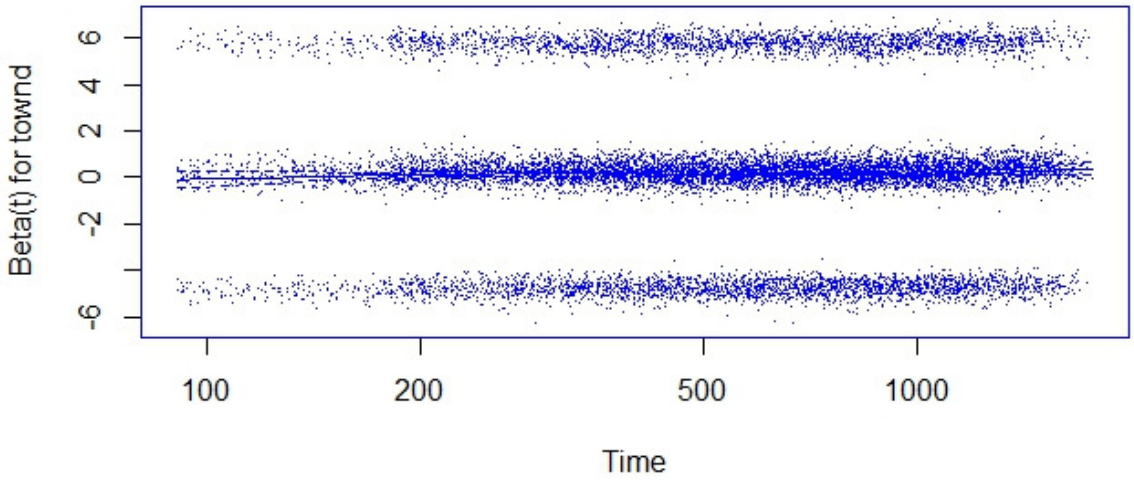


**Figure 25. Scaled Schoenfeld residuals for Townsend Quintile 4 versus time**

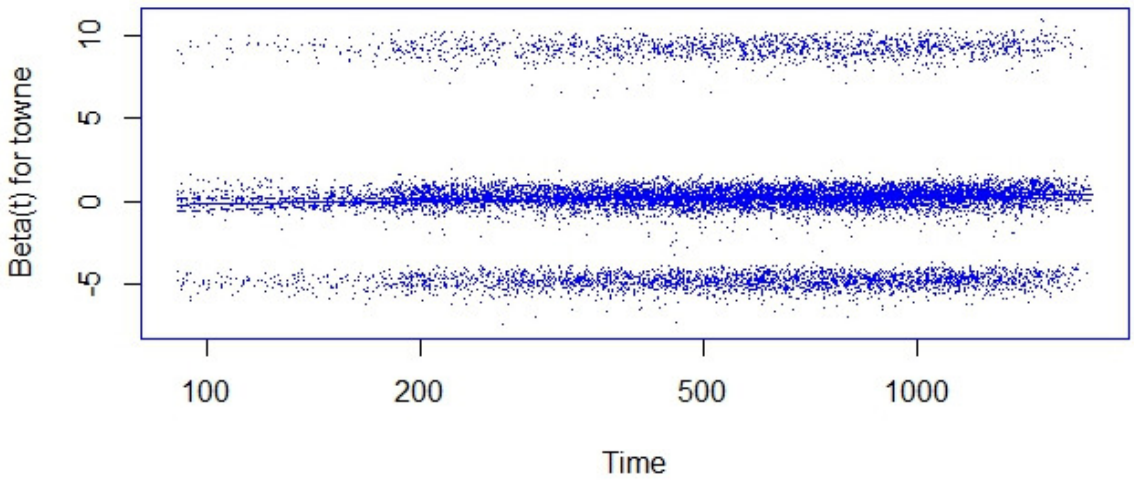



**Figure 26. Scaled Schoenfeld residuals for Vit D supplementation versus time**

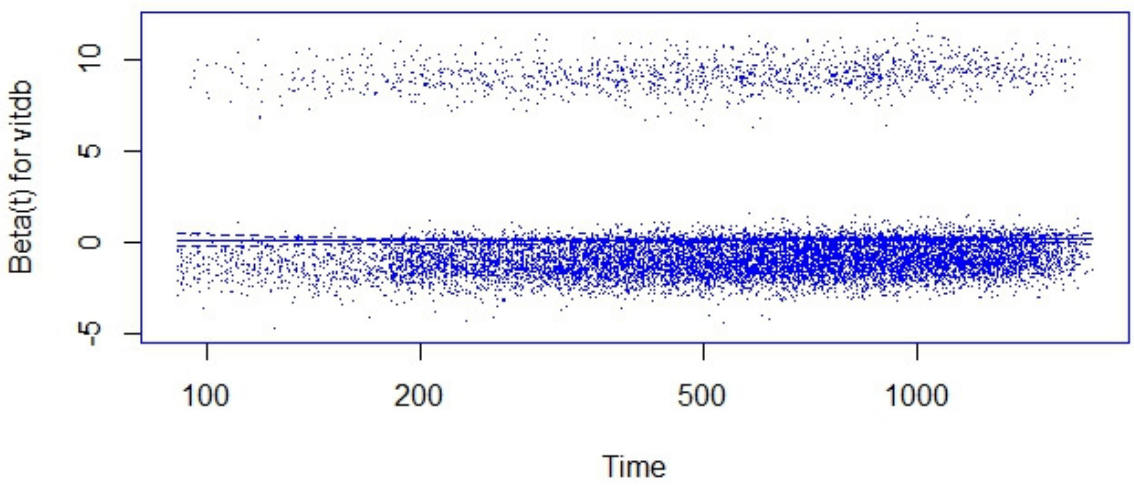


**Figure 27. Scaled Schoenfeld residuals for ISC race versus time**

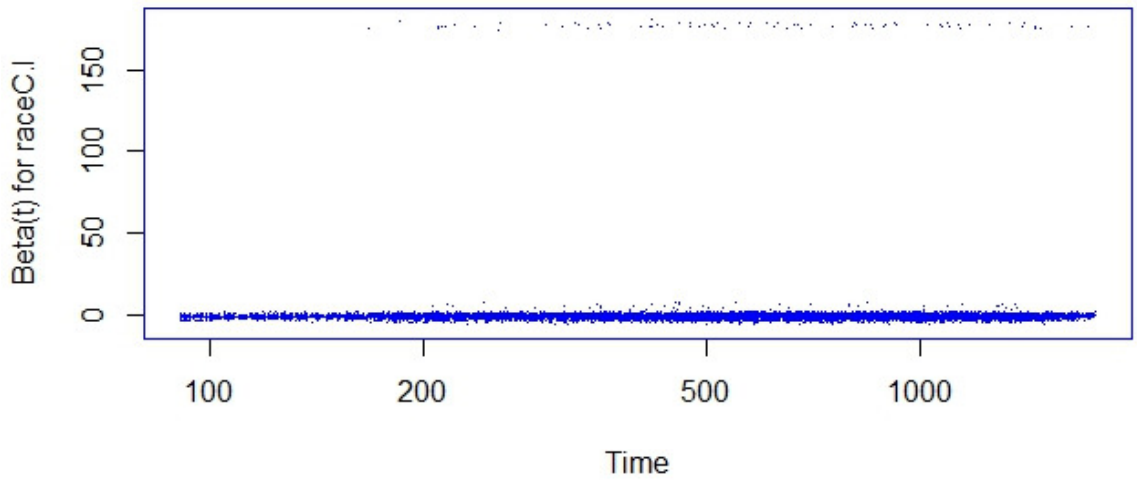



**Figure 28. Scaled Schoenfeld residuals for GFR versus time**