Competing Risks Methodology in the Evaluation of Cardiovascular and Cancer Mortality as a Consequence of Albuminuria in Type 2 Diabetes

Benjamin Graham Feakins

St Edmund Hall

University of Oxford

A Thesis submitted for the degree of

Doctor of Philosophy in Primary Health Care

Trinity Term, 2016

Acknowledgements

First and foremost, I must thank to my supervisors, Prof Richard Stevens, Dr Emily McFadden & Prof Andrew Farmer, for their guidance, support, and encouragement throughout my research. I would also like to thank the NIHR SPCR, without whose funding this research would not have taken place.

The various reviews and systematic reviews contained within this Thesis would not have been possible without the help of Nia Roberts and Jennifer Hirst, both of whom provided me with valuable information and advice. Special thanks are also owed to Dr Helen Ashdown, Dr David McCartney & Prof Andrew Farmer for their clinical input into much of the database work performed.

Undertaking a DPhil degree can be a stressful pursuit. Therefore, special thanks are due to all those who supported me throughout the last three and a half years. It would be impossible to list everyone here, but I would like to extend particular thanks to all my colleagues in the Nuffield Department of Primary Care Health Sciences. Particular thanks go to Dr Emily McFadden, for her weekly tolerance of my problems, to Prof Richard Stevens for entertaining my 5pm statistical quandaries, and to Dr Helen Ashdown for all the sweets, cups of tea and coffee.

Finally, I would like to thank my partner, Karolina, for her unwavering support, and encouragement.

Statement of Originality

All writing and statistical analyses was carried out by myself between October 2012 and May 2016. To my knowledge all of the work in this Thesis is original, except where referenced and has not been submitted for a degree at any other university.

Competing Risks Methodology in the Evaluation of Cardiovascular and Cancer Mortality as a Consequence of Albuminuria in Type 2 Diabetes

Benjamin Graham Feakins
St Edmund Hall

Doctor of Philosophy in Primary Health Care
Trinity Term, 2016

Abstract

Background: 'Competing risks' are events that either preclude or alter the probability of experiencing the primary study outcome(s). Many standard survival models fail to account for competing risks, introducing an unknown level of bias in their measures of absolute and relative risk. Individuals with type 2 diabetes mellitus (T2DM) and albuminuria are at increased risk of multiple competing causes of mortality, including cardiovascular disease (CVD), cancer and renal disease, yet studies to date have not implemented competing risks methodology.

Aim: Using albuminuria in T2DM as a case study, this Thesis set out to quantify differences between standard- and competing-risks-adjusted survival analysis estimates of absolute and relative risk for the outcomes of cardiovascular and cancer mortality.

Methods: 86,962 patients aged ≥35 years with T2DM present on or before 2005 were identified in the Clinical Practice Research Datalink. To quantify differences in measures of absolute risk, cumulative risk estimates for cardiovascular and cancer mortality from standard survival analysis methods (Kaplan-Meier estimator) were compared to those from competing-risks-adjusted methods (cumulative incidence competing risk estimator). Cumulative risk estimates were stratified by patient albuminuria level (normoalbuminuria vs albuminuria). To quantify differences in measures of relative risk, estimates for the effect of albuminuria on the relative hazards of cardiovascular and cancer mortality were compared between standard cause-specific hazard (CSH) models (Cox-proportional-hazards regression), competing risk CSH models (unstratified Lunn-McNeil model), and competing risk subdistribution hazard (SDH) models (Fine-Gray model).

Results: Patients with albuminuria, compared to those with normoalbuminuria, were older ($p<0.001$), had higher systolic blood pressure ($p<0.001$), had worse glycaemic control ($p<0.001$), and were more likely to be current or ex-smokers ($p<0.001$). Over the course of nine years of follow-up 22,512 patients died; 8,800 from CVD, 5,239 from cancer, and 8,473 from other causes. Median follow-up was 7.7 years. In patients with normoalbuminuria, nine-year standard and competing-risks-adjusted cumulative risk estimates for cardiovascular mortality were 11.1% (95% confidence interval (CI): 10.8-11.5%) and 10.2% (95% CI: 9.9-10.5%), respectively. For cancer mortality, these figures were 8.0% (95% CI: 7.7-8.3%) and 7.2% (95% CI: 6.9-7.5%). In patients with albuminuria, standard and competing-risks-adjusted estimates for cardiovascular mortality were 21.8% (95% CI: 20.9-22.7%) and 18.5% (95% CI: 17.8-19.3%), respectively. For cancer mortality, these figures were 10.7% (95% CI: 10.0-11.5%) and 8.6% (8.1-9.2%). For the effect of albuminuria on cardiovascular mortality, hazard ratios from multivariable standard CSH, competing risks CSH, and subdistribution hazard ratios from competing risks SDH models were 1.75 (95% CI: 1.63-1.87), 1.75 (95% CI: 1.64-1.87), and 1.58 (95% CI: 1.48-1.69), respectively. For the effect of albuminuria on cancer mortality, these values were 1.27 (95% CI: 1.16-1.39), 1.28 (95% CI: 1.17-1.40), and 1.11 (95% CI: 1.01-1.21).

Conclusions: When evaluating measures of absolute risk, differences between standard and competing-risks-adjusted methods were small in absolute terms, but large in relative terms. For the investigation of epidemiological relationships using relative hazards models, standard survival analysis methods produced near-identical risk estimates to the CSH competing risks methods for the clinical associations evaluated in this Thesis. For the evaluation of risk prediction using relative hazards models, CSH models produced consistently higher risk estimates than SDH models, and their use may lead to over-estimation of the predictive effect of albuminuria on either outcome. Where outcomes are less common (like cancer) CSH models provide poor estimates of risk prediction, and SDH models should be used. This research demonstrates that differences can be present between risk estimates derived using CSH and SDH methods, and that the two are not necessarily interchangeable. Moreover, such differences may be present in other clinical areas.

Acronyms and Abbreviations

ACR - Albumin-Creatinine Ratio

AER - Albumin Excretion Rate

BMI - Body-Mass Index

BNF - British National Formulary

CI - Confidence Interval

CICR - Cumulative Incidence Competing Risk

CKD - Chronic Kidney Disease

Cox-PH - Cox Proportional Hazards

CPRD - Clinical Practice Research Datalink

CSH - Cause Specific Hazard

CVD - Cardiovascular Disease

DAFNE - Dose Adjustment for Normal Eating

DCCT - Diabetes Control and Complications Trial

DESMOND - Diabetes Education and Self-Management for Ongoing and Newly Diagnosed

DKA - Diabetic Ketoacidosis

GP - General Practitioner

HbA1c - Glycated Haemoglobin

HDL - High Density Lipoprotein

HONK - Hyperosmolar Nonketotic Syndrome

HR - Hazard Ratio

ICD - International Classification of Disease

IDDM - Insulin-Dependent Diabetes Mellitus

IFCC - International Federation of Clinical Chemistry

ISAC - Independent Scientific Advisory Committee

NIDDM - Non-Insulin-Dependent Diabetes Mellitus

ONS - Office for National Statistics

PCOS - Polycystic Ovary Syndrome

PCR - Protein-Creatinine Ratio

PER - Protein Excretion Rate

PPV - Positive Predictive Value

PRISMA - Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PROSPERO - Prospective Register of Systematic Reviews and Meta-Analyses

QOF - Quality Outcomes Framework

RCT - Randomised Controlled Trial

SBP - Systolic Blood Pressure

SDH - Subdistribution Hazard

T1DM - Type 1 Diabetes Mellitus

T2DM - Type 2 Diabetes Mellitus

Total : HDL - Total-to-High-Density-Lipoprotein

UKPDS - United Kingdom Prospective Diabetes Study

UTS - Up to Standard

Contents

1. Introduction

1.1. Background

1.1.1. Survival Analysis and Competing Risks

Survival analyses are a set of methods for the analysis of data in which time until the occurrence of one or more outcomes is of interest; otherwise known as 'time-to-event' data. Through such analyses, researchers may draw inferences on the association of risk factors with the probability of experiencing an outcome over time (termed 'cumulative risk') and how much more likely an outcome is to occur in subjects who are exposed to the risk factor than in those who are unexposed (termed 'relative hazard'). Standard survival analysis methods are only able to evaluate the occurrence of single outcome. When the outcome of interest is all-cause mortality, quantifying the effect of a risk factor is relatively straightforward. When the outcome of interest represents a particular type of mortality or a non-fatal event, inferences can be more problematic, as subjects are rarely (if ever) at risk of experiencing only a single type of outcome.

Competing risks are defined as events during follow-up that either preclude the observation of the primary outcome, or alter the probability of its occurrence[1]. The effect of competing risks was first acknowledged by d'Alembert and Bernoulli 1760s in relation to the effects of inoculation on short- and long-term mortality from smallpox[2]. To illustrate, one may consider a simplified multi-state model consisting of three states: a baseline state, *i*; a primary outcome state, *j*; and a competing outcome state, *k* (Figure 1.1). For competing risks that preclude the observation of the primary outcome, subjects move from *i* to *j* at the rate λ_j and from *i* to *k* at the rate λ_k. For example, subjects may move

from a state of healthiness to either the study primary outcome of cardiovascular mortality, or the competing outcome of non-cardiovascular mortality.

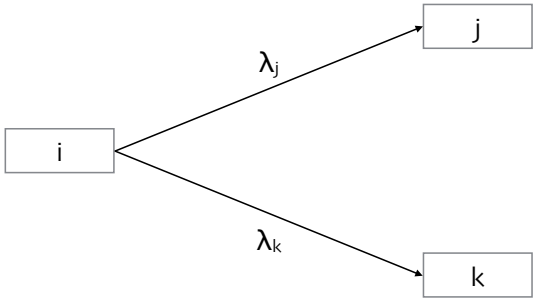


Figure 1.1 - A multi-state model depicting the competing risk process for competing outcomes that preclude the observation of the primary outcome.

In prognostic studies, time-to-event data may be incomplete for many subjects as a result of loss to follow-up, reaching the end of the study without experiencing the event of interest, or withdrawal from the study. The truncation of subject follow-up resulting from incomplete data is typically referred to as 'censoring'. Most standard survival models, like the Kaplan-Meier estimator[3] and Cox proportional hazards (Cox-PH) regression[4], assume that censoring is 'non-informative', meaning that a subject censored at a certain time point should be representative of those still under observation at the same time point[5], i.e. subject censoring times and event times should be independent. This assumption may be suitable for the majority of censoring events when evaluating all-cause mortality. Thus, in such studies, the use of standard survival analysis methods may be appropriate. However, in the evaluation of any outcome that is not all-cause mortality, the censoring of subjects who have died from competing causes represents a clear violation of the non-informative censoring assumption. In such studies, the use of standard survival models may lead to bias in risk estimates.

1.1.2. Survival Analysis Concepts and Estimators

Most non-Bayesian survival models and estimators may be classified as either 'cause-specific hazards' (CSHs) or 'subdistribution hazards' (SDHs) (Figure 1.2). These terms describe the segments of the population considered to be in the 'risk set' (i.e. 'at risk') at time *t*[6].

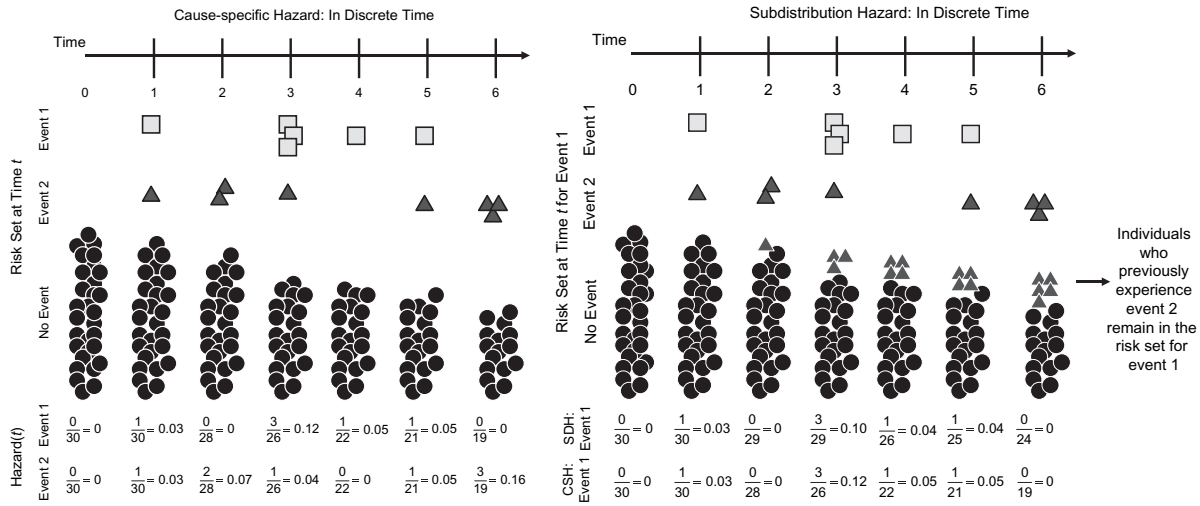


Figure 1.2 - (Reproduced from a paper by Lau et al.[7]). Diagrams representing the CSH (left) and SDH (right) populations for a primary outcome. "Event 1" is the primary outcome, while "Event 2" represents a competing outcome. Key: □ = primary outcome events; ▲ = Competing outcome events; ● = Subjects yet to experience either the primary or competing outcome events.

In the case of models assuming CSH risk sets (termed 'CSH models') subjects are considered to be at risk who have not experienced the study primary outcome, and have not experienced competing outcomes (Figure 1.2; left-hand plot). Comparatively, in the case of models assuming SDH risk sets (termed 'SDH models') subjects are considered to be at risk if they have not experienced the study primary outcome (Figure 1.2; right-hand plot). Thus, subjects who have experienced competing outcomes are considered to be at risk at time *t* by SDH models, but are not considered to be at risk by CSH models. Although the approach of the SDH models may appear counter-intuitive, maintaining subjects in the risk set who have experienced competing outcomes allows these subjects to act as place-holders that represent the proportion of the risk set that can never experience the primary outcome[7,8]. Although not represented in Figure 1.2, both CSH and SDH models can allow for censoring from 'true loss to follow-up' (i.e. subjects lost to follow-up as a result of moving away). In CSH and SDH models, subjects censored as a result of 'true loss to follow-up' are removed from the risk set.

1.1.3. Absolute Risk Measures

In standard survival analysis, one popular method for determining the cumulative hazard (a CSH measure of cumulative risk) is the Kaplan-Meier cumulative hazard estimator[3]. Where j_i represents the number of occurrences of the study primary outcome at time t_i, and n_i represents the number of subjects under observation at time t_i, the Kaplan-Meier cumulative hazard (H) may be expressed by Equation 1.1. In essence, the Kaplan-Meier estimator represents the product of a number of ratios representing the risks at each individual time point, up to and including time t_i.

Equation 1.1 - Specification of the Kaplan-Meier cumulative hazard estimator, for a primary outcome j.

$$\widetilde{H}_j(t) = 1 - \prod_{\substack{i \\ t_i \leq t}} \left(1 - \frac{j_i}{n_i}\right)$$

A second standard survival analysis method for determining the cumulative hazard is the Nelson-Aalen estimator[9]. The Nelson-Aalen cumulative hazard estimator differs from the Kaplan-Meier cumulative hazard estimator through being the sum of the risks at each individual time point, as opposed to 1- the product of the survival probabilities at each individual time point. Where j_i represents the number of subjects who have experienced the primary outcome j at time t_i, and n_i represents the number of subjects under observation at time t_i, the Nelson-Aalen cumulative hazard (H) may be expressed by Equation 1.2.

Equation 1.2 - Specification of the Nelson-Aalen cumulative hazard estimator, for a primary outcome j.

$$\widetilde{H}_j(t) = \sum_{\substack{i \\ t_i \leq t}} \left(\frac{j_i}{n_i}\right)$$

Both the Nelson-Aalen and Kaplan-Meier cumulative hazard estimators are only able to account for the occurrence of a single outcome j. All subjects experiencing competing outcomes would be censored by both estimators, impacting the number of subjects able to experience the primary outcome, and hence, their cumulative risk estimates[1,5,10]. This has led some researchers to refer to the

risk estimates from CSH estimators as estimates of "marginal" or "net" survival, as they relate to the probability of experiencing the primary outcome in a virtual world where all other outcomes are absent[11]. The effect of competing outcomes on the cumulative hazard may be observed through the summation of the individual cumulative hazards present in a population. When competing risks are present, the sum of the cumulative hazards can exceed 100%, implying that more than the entirety of the population can experience mutually exclusive outcomes[5].

In competing risks survival analysis, the cumulative incidence competing risk (CICR) estimator[5] represents a method of calculating the cumulative incidence (a SDH measure of cumulative risk) of a study outcome. Where *k* represents the number of subjects experiencing a second competing outcome *k* at time t_i, the cumulative incidence (C*I*) of an outcome *j* at time t_i may be given by Equation 1.3. Essentially, for two competing outcomes *j* and *k*, the cumulative incidence of *j* at time *ti* represents a function of the Nelson-Aalen cumulative hazard estimate of outcome *j* at time t_i, and the Kaplan-Meier product limit survival estimates of outcome *j* and *k* at time t_i.

Equation 1.3 - Specification of the CICR estimator, for a primary outcome *j* with one competing outcome *k*.

$$CI_j(t) = \sum_{\substack{i \\ t_i \leq t}} \frac{j_i}{n_i} \left(\prod_{\substack{l \\ t_l \leq t_i}} \left(1 - \frac{j_l}{n_l}\right) \right) \left(\prod_{\substack{l \\ t_l \leq t_i}} \left(1 - \frac{k_l}{n_l}\right) \right)$$

Unlike the cumulative hazard, the cumulative incidence is dependent upon the number of subjects experiencing the primary outcome and the number of subjects experiencing competing outcomes. Consequently, while competing events will result in higher estimates of cumulative hazard, they will result in lower estimates of cumulative incidence[1]. Moreover, whether competing risks are present or absent, the summation of the individual cumulative incidences present in a population will never exceed 100%.

1.1.4. Relative Risk Measures

A core concept to the interpretation of relative measures of risk in competing risks survival analysis is that of the type of risk being described by the model. In this field, models are broadly classified as "aetiological" or "prognostic" dependent upon whether they are describing "causal" or "predictive" relationships, respectively[7,8]. Although it is a well-established notion that neither of these attributes may be established by statistical modelling[12], these terms are deeply entrenched in the descriptions of competing risks models present in the literature[7,8,13,14]. Hence, competing risks models do not serve to establish the presence of causal or predictive relationships, and the use of the terms "aetiological association" and "prognostic association" may simply be translated to "the effects of a variable on the hazard of an outcome", and "the effects of a variable on the cumulative incidence of an outcome", respectively. The use of these terms relates to the fact that the hazard is thought to be a better means to quantify any causal relationships present, and the cumulative incidence is thought to be a better means to quantify any predictive relationships present. An example where the aetiological association between a risk factor and outcome is desired may be found in the research question: "to what degree is the presence of albuminuria directly associated with cardiovascular mortality?". Whereas, an example where the prognostic association between a risk factor and outcome is desired may be found in the research question: "Regardless of the direct association, are subjects with albuminuria more likely to experience cardiovascular mortality?".

A key component to many standard survival models is that of the CSH function (λ), representing the instantaneous risk of experiencing the primary outcome *j*, given the subject is still alive at time *t*[7]. In 'relative hazards' standard survival models, λ is assumed to be proportional to a baseline hazard function (λ_0). Where no distribution is assumed for λ_0 the resulting model is termed the 'Cox-PH model'[4] (Equation 1.4). Relative hazards survival models like the Cox-PH model may be used to describe relationships between risk factors and the CSH of an outcome. In the field of competing risks survival analysis this relationship is typically referred to as 'aetiological' or 'causal' association[7,8,13,14]. Measures of aetiological association, like the cause-specific hazard ratio (HR_{CS} =

e^{β}), are well suited to the investigation of aetiological relationships, as they quantify the risk of the primary outcome in subjects who are actually able to develop the primary outcome[1,7,8].

Equation 1.4 - Specification of the Cox proportional hazards CSH model, for a primary outcome *j*.

$$\lambda_j(t) = \lambda_{0,j}(t)e^{(\beta)}$$

Despite its conceptual use in describing aetiological relationships, it has been suggested that HR_{CS} derived from standard survival analysis methods, such as Cox-PH regression, could still be biased by the presence of competing risks[15], and that this bias may be mitigated through the use of a joint λ_0 for all present competing outcomes (thus, assuming proportionality between the CSHs of each)[16]. This adaptation of the Cox-PH model is termed the 'unstratified Lunn-McNeil model'. For a primary outcome *j* in the presence of a competing outcome *k* the specification of the unstratified Lunn-McNeil model for the CSH function λ_j is shown in Equation 1.5.

Equation 1.5 - Specification of the unstratified Lunn-McNeil CSH model, for a primary outcome *j* with one competing outcome *k*.

$$\lambda_j(t) = \lambda_{0,j,k}(t)e^{(\beta)}$$

In scenarios where subjects are more likely to experience competing outcomes than the study primary outcome, HR_{CS} can give misleading representations of prognostic risk, as risk factors that are aetiologically associated with the primary outcome may be unable to predict its occurrence (i.e. parameter estimates from CSH models cannot be directly interpreted in terms of the SDH)[17]. Further complications to the use of HR_{CS} may be found in the presence of violations of the proportional CSH assumption, where large amounts of censoring (i.e. strong competing risks) may result in biased estimates of HR_{CS}[18]. This has brought into question the value of HR_{CS} as a measure of prognostic association in the presence of competing risks[19], and has led to development of models based on SDH risk sets[20]. The SDH function l_j may be defined as the instantaneous risk of experiencing the

primary outcome *j*, given the subject has not already experienced the primary outcome *j*[7]. For a primary outcome *j*, I_j is dependent upon the CSH of primary outcome λ_j and the CSHs of all present competing outcomes λ_k[21] (Equation 1.6).

Equation 1.6 - Specification of the relationship between the SDH and CSH, for a primary outcome *j* with one competing outcome *k*.

$$I_j(t) = \left(\frac{\lambda_j(t)}{\lambda_j(t) + \lambda_k(t)}\right)(1 - e^{-(\lambda_j + \lambda_k)t})$$

In a relative hazards model framework, *I* may be assumed to be proportional to an unspecified arbitrary baseline SDH function I_0 to yield a proportional hazards model analogous to that of the Cox-PH model for λ[20], for which the proportionality of *I* to I_0 is often satisfied when strong competing risks are present[18], and thus, when the proportional CSH assumption is violated[22]. This adaptation of the Cox-PH model methodology is termed the 'Fine-Gray model'[20]. For a primary outcome *j* the specification of the Fine-Gray model for the SDH function I_j is shown in (Equation 1.7).

Equation 1.7 - Specification of the Fine-Gray SDH model, for a primary outcome *j*.

$$I_j(t) = I_{0,j}(t)e^{(\varphi)}$$

In the field of competing risks survival analysis, the relationship between risk factors and the SDH of an outcome is typically referred to as 'prognostic' or 'predictive' association[7,8,13]. Measures of prognostic association, like the subdistribution hazard ratio ($HR_{SD} = e^{\varphi}$), are well suited to the investigation of risk prediction as they are able to quantify the effect of risk factors on the primary outcome, adjusted for all present competing outcomes (i.e. they can be interpreted directly in terms of the SDH).

1.2. Type 2 Diabetes Mellitus and Diabetic Nephropathy

According to Diabetes UK in 2014 3.5 million people in the UK had been diagnosed with diabetes mellitus (using Quality Outcome Framework (QOF) diagnostic codes)[23]. Of these approximately 90% will have type 2 diabetes mellitus (T2DM)[24]. T2DM is a metabolic disorder that is characterised by defects in insulin sensitivity or secretion[25]. The disease typically manifests in later life and is progressive in nature[26], resulting in systemic macro- and microvascular complications[27]. One such complication is diabetic nephropathy whereby damage to the microvasculature of the glomerular capillaries of the kidneys results in a leakage of proteins into the urine (albuminuria/proteinuria), hypertension and reduced glomerular filtration rate[28]. As it is usually impractical to confirm by biopsy, the disease is typically diagnosed and monitored by means of subject albuminuria levels[28]. Although a continuous variable by nature, albuminuria may be categorised into three stages based on the ratio of albumin to creatinine in the urine, with the lowest stage, normoalbuminuria (<30mg/g), implying an absence of clinically significant nephropathy, and microalbuminuria (30-300mg/g) and macroalbuminuria (>300mg/g) implying the presence of clinically significant nephropathy[28]. Within 15 years of diagnosis of T2DM, approximately 35% of UK patients will have developed clinically significant albuminuria (microalbuminuria or more sever)[29].

Subjects with T2DM and albuminuria are at increased risk of mortality from a range of cardiovascular, renal and cancer causes[29,30]. Therefore, in the evaluation of any outcome that is not all-cause mortality, the degree of censoring representing failure from competing causes is likely to be high. Shared dependence among respective outcome distributions and both T2DM and albuminuria renders the non-informative censoring assumption unlikely to be valid, as the distributions of censoring events (primarily driven by competing mortality) and primary outcome events will both be associated with the severity of albuminuria and T2DM. Thus, censoring competing outcomes may impart significant bias on measures of risk estimated from standard survival methods in populations with T2DM and albuminuria.

Competing risks methodology is largely unused in prognostic modelling[13], and its application to the area of T2DM and albuminuria warrants further investigation. A review of the literature was

conducted to identify the current use of competing risks methodology in the evaluation of cardiovascular and cancer outcomes in patients with T2DM and albuminuria.

1.3. Review of Current Practice

1.3.1. Aims and Objectives

Primary Objective

To determine the current usage of competing risks methods in the evaluation of cardiovascular or cancer outcomes in patients with T2DM and albuminuria.

Specific Aims

i. To identify studies in which time-to-event analyses had been used to evaluate the risk of cardiovascular or cancer outcomes in patients with T2DM and albuminuria.

ii. To determine whether identified studies were potentially susceptible to bias from competing risks as a result of the implementation of inappropriate time-to-event analyses.

1.3.2. Methods

Data Sources and Searches

The CINAHL[31] database and Ovid versions of the MEDLINE[32] and EMBASE[33] databases were systematically searched from their inception until 25th May 2013. Relevant published studies were identified through a search strategy incorporating keywords and search terms, i.e. medical subject heading terms for CINAHL and MEDLINE, and EMTREE terms for EMBASE (see Appendix A.1). Outcome-specific keywords were derived by exploding the medical subject headings for "Cardiovascular Diseases" and "Neoplasms" from the PubMed MEDLINE interface[34]. Additional searches of the Prospective Register of Systematic Reviews (PROSPERO)[35] database were performed to identify any in-progress or published systematic reviews seeking to determine the usage of competing risks methodology in the evaluation of cardiovascular and cancer outcomes in

patients with T2DM and albuminuria. The reference lists of relevant systematic reviews were sought and screened for further studies.

Study Selection

Studies were eligible for inclusion if they were of randomised controlled trial (RCT) or prospective cohort design, had a minimum average follow-up of five years, and the full-text was available in English.

Population

Studies were required to have a minimum of 1,000 subjects aged 18 years or over with both T2DM and albuminuria present at baseline.

Exposure

The use of prognostic models in the evaluation of patient risk.

Outcome

Studies were eligible for inclusion if the outcomes under investigation were cardiovascular disease, cancer, a sub-type of either of these outcome classifications (e.g. “stroke” or “prostate cancer”), or all-cause mortality.

Data Extraction

A sample of study full-texts was used to pilot data extraction forms. Data were extracted on:

- Length of follow up.
- Number of subjects.
- Study design.
- Study primary outcomes.
- Study competing outcomes.
- The prognostic analysis models/methods used.

- The citation of competing risk model/method seminal papers.
- The discussion of competing risks in the study free text.

Data Synthesis and Analysis

Competing risk events were defined as:

- All-cause mortality, where study outcomes were non-fatal.
- Competing mortality, where study outcomes were cause-specific mortality, or fatal/non-fatal hybrid outcomes.

Susceptibility to competing risks was defined as:

- The presence of risk estimates produced from time-to-event analyses unadjusted for competing risks, (e.g. the Kaplan-Meier estimator and Cox-PH regression).
- The analysis of outcome structures that left scope for competing risk events.

1.3.3. Results

Database searches identified 5,913 studies. Of these, 2,329 were duplicates, leaving 3,886 studies for review. After the initial phase of title and abstract screening, 3,707 papers were excluded. Full texts were sourced for the remaining 179 papers. Upon further examination, an additional 150 papers were excluded, leaving 29 for review. The Preferred Reporting In Systematic Reviews and Meta-Analyses (PRISMA) flow diagram depicting the study selection process may be found in Figure 1.3.

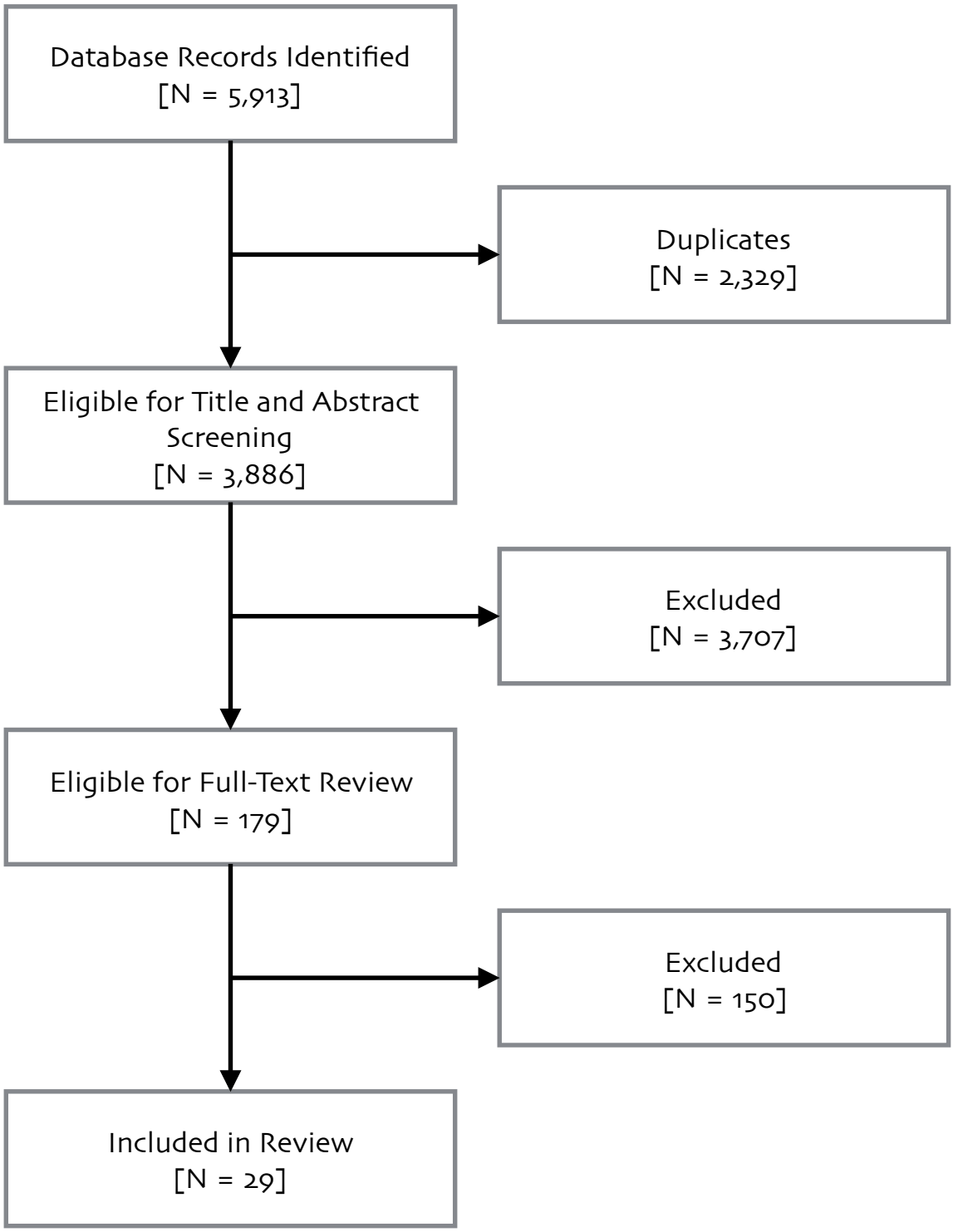


Figure 1.3 - PRISMA flow diagram depicting the screening process.

The median follow-up and median study population were 5.7 years and 6,027 subjects, respectively. Further details on the extracted studies may be found in Table 1.1. In 83% (24/29) studies, the structure of outcomes analysed provided scope for the presence of unanalysed competing risks.

Time-to-event models adjusted for the presence of competing risks were not used by any studies, and no studies discussed the topic of competing risks in free text.

Table 1.1 - Characteristics of the studies included in the review. "." Denotes missing data. *The measure of central tendency used by the study author to represent the 'average' was extracted. Where the mean and median were stated, the median was extracted.

First Author	Publication Year	Study Design	Population (N)	Average* Follow-Up (Years)	Cumulative Risk Estimator	Relative Hazards Model	Unevaluated Competing Risk Events	Discussion of Competing Risks
Sasaki A[36]	1983	Cohort	1,221	7.5	None	None	No	No
Fuller J[37]	1996	Cohort	4,714	.	None	Cox-PH	Yes	No
Sievers M[38]	1996	Cohort	6,027	.	Unknown	Cox-PH	No	No
Wang S-L[39]	1996	Cohort	4,714	.	None	None	Yes	No
Davis T[40]	1999	Cohort	3,776	7.9	None	Cox-PH	Yes	No
Calhoun H[41]	2001	Cohort	3,558	8.4	None	Cox-PH	Yes	No
Fuller J[37]	2001	Cohort	4,743	.	None	Cox-PH	Yes	No
Keen H[42]	2001	Cohort	4,662	8.4	None	None	No	No
Miki E[43]	2001	Cohort	3,868	.	None	None	Yes	No
Morrish N[44]	2001	Cohort	4,713	.	None	None	No	No
Parkov M[45]	2004	Cohort	2,095	11.1	None	None	No	No
Bruno G[46]	2005	Cohort	1,565	7.0	None	Cox-PH	Yes	No
The FIELD Study Investigators[47]	2005	RCT	9,795	5.0	Kaplan-Meier	Cox-PH	Yes	No
Hippisley-Cox J[48]	2008	Cohort	2,300,000	7.0	None	Cox-PH	Yes	No
The ADVANCE Collaborative Group[49]	2008	RCT	11,140	5.0	Unknown	Cox-PH	Yes	No
Yang X[50]	2008	Cohort	6,445	5.4	None	Cox-PH	Yes	No
Yang X[51]	2008	Cohort	6,969	5.4	None	Cox-PH	Yes	No
Eeg-Olofsson K[52]	2009	Cohort	13,087	5.6	None	Cox-PH	Yes	No
Ma R[53]	2009	Cohort	3,861	5.6	None	Cox-PH	Yes	No
Nilsson P[54]	2009	Cohort	13,087	5.7	None	Cox-PH	Yes	No
Burgess D[55]	2010	RCT	9,795	5.0	Kaplan-Meier	Cox-PH	Yes	No
Lin E[56]	2010	Cohort	3,723	22.0	None	Unknown	Yes	No
Drury P[57]	2011	RCT	9,795	5.0	Kaplan-Meier	Cox-PH	Yes	No
Vupputuri S[58]	2011	Cohort	10,290	5.4	Unknown	Cox-PH	Yes	No
Girman C[59]	2012	Cohort	1,914,482	.	None	Cox-PH	Yes	No
Morton J[60]	2012	RCT	11,140	5.0	None	Cox-PH	Yes	No
Yang X[61]	2012	Cohort	3,476	5.1	Kaplan-Meier	Cox-PH	Yes	No
Hippisley-Cox J[62]	2013	Cohort	3,549,748	7.0	None	Cox-PH	Yes	No
Ting R[63]	2013	Cohort	7,835	7.4	Kaplan-Meier	Cox-PH	Yes	No

Use of time-to-event models susceptible to bias from competing risks was present in 76% (22/29) of studies. Of these, all (22/22) produced hazard ratios from Cox-PH regression, and 23% (5/22) produced cumulative hazard curves from the Kaplan-Meier estimator.

The use of standard or competing risks methods could not be determined for one or more analyses performed in 14% (4/29) studies, as the models used to generate the risk estimates were not specified. In 21% (6/29) of studies no time-to-event analyses were performed.

1.3.4. Discussion

Key Results

Competing risks methods were not used for the evaluation of cardiovascular or cancer outcomes in patients with T2DM and albuminuria by any studies identified in this review. Of the studies that performed time-to-event analyses, over a fifth (5/22) used CSH cumulative risk estimators susceptible to bias from competing risks, while all (22/22) used standard survival analysis CSH relative hazards models which can produce misleading estimates of prognostic risk in the presence of competing risks. In 83% (24/29) studies, the structure of study outcomes provided scope for the presence of unanalysed competing risk events.

Strengths and Limitations

A subset of the inclusion criteria (that were designed to ensure a minimum event-rate and act as surrogate markers of study quality) may have been overly exclusive. For example, no studies from the RENAAL cohort[64] (a large multicentre RCT evaluating the effect of angiotensin II receptor blockers on mortality and progression of diabetic nephropathy in patients with non-insulin-dependent diabetes) were included in this review due to insufficient follow-up (median 3.4 years). Similarly, many studies were excluded because they had an insufficient number of participants at baseline. Thus, it cannot be categorically stated that no studies investigating cardiovascular or cancer outcomes in populations of patients with T2DM and albuminuria accounted for competing risks; merely that for the predefined inclusion criteria, none were found. However, the fact that none of the reviewed studies used competing risk methodology may imply that the vast majority of researchers in this field are not aware of these methods.

It is possible that studies performing multiple Cox-PH regression analyses (one for each outcome under investigation) in the same population may have done so in a bid to follow the competing risks methods described by Kalbfleisch and Prentice, which advocates the use of multiple Cox-PH models to evaluate all CSHs present in a population[15]. However, this is unlikely, as the topic of competing risks was not discussed in the free text of any of the studies. Furthermore, few studies conducting multiple Cox-PH models structured their outcomes so as to account for all CSHs present in the population, and none cited the relevant methodological paper by Kalbfleisch and Prentice[15].

Some standard survival analysis methods, like the Kaplan-Meier estimator, are well established to be susceptible to bias from competing risks[1,7,8,10]. However, the susceptibility of some of the models to be biased by competing risks, like the Cox-PH model, has yet to be fully established[15,16]. As such, it is not clear to what degree and under which scenarios the estimates of these models may be biased. Where the estimates of competing risks models were not significantly different from those of standard methods, study authors may have opted not to use competing risks models in a bid to limit the complexity of their analyses.

The inconsistent use of terminology in the field of competing risks makes the determination of whether or not studies implemented competing risks methodology problematic. Within the realms of standard survival analysis, studies will often refer to distinct measures of risk interchangeably or vaguely. For example, some studies in the review referred to "HRs" as "relative risks" or "death rate ratios"[37,38,41,46], or used the terms "cumulative risk" or "cumulative progression" without reference to the estimator used[57,58]. These problems are compounded when evaluating the usage of competing risks methods, as in the field of competing risks survival analysis, studies may often use standard survival analysis terminology, while studies using standard survival analysis methods may inadvertently use competing risks terminology. For example, in this review, one study provided HR risk estimates without reference to the model used[56], three studies referred to cumulative hazard estimates from the Kaplan-Meier estimator as "cumulative incidences"[55,61,63], and two studies used the term "cumulative incidence" without reference to the estimator used[38,49]. In the latter of these examples, it is probable that the study by Sievers et al.[38] did not use competing risks methods to

produce cumulative incidence estimates as their study predates the seminal paper on the cumulative incidence competing risk (CICR) estimator[1].

Relationship to the Literature

This was the first study to evaluate the use of competing risks methodology in studies analysing populations with T2DM and albuminuria. As such it has no direct comparators with regards to its choice of study population and methods. A similar review was published by Koller et al. in 2012 on the use of competing risks methodology in high-impact medical journals[13]. However, the review by Koller et al.[13] was less exhaustive in its methods, as the authors chose not to analyse all studies eligible for inclusion.

The lack of studies reporting and accounting for competing risks in subjects with T2DM and albuminuria contrasts with other high risk patient populations, like cancer patients undergoing chemotherapy[20] or transplant recipients[16]. Where Koller et al.[13] identified the application of competing risks methodology in 20% of their study sample, and a further 18% as acknowledging competing risks as a potential issue; there was a complete absence in the use of this methodology in the papers identified by this study. The use of inappropriate statistical methods prone to bias from competing risks was a relatively common problem. It was considered that many of the study authors may not have been aware of appropriate competing risks methodologies to implement in their studies. However, all studies using the Kaplan-Meier estimator were published after the paper outlining methods to implement the CICR estimator[1], while all studies using Cox-PH models were published after the seminal paper for the Lunn-McNeil[16] competing risks proportional hazards model (with all but three of these also being published after the seminal paper for the Fine-Gray[20] competing risks proportional hazards model).

Implications

Failure to account for competing risks may have implications for the findings of prognostic studies evaluating cardiovascular or cancer outcomes in patients with T2DM and albuminuria. The absence of studies using competing-risks-adjusted methods identified by this review (despite the high

susceptibility of the population to competing outcomes) places a large degree of uncertainty upon the reliability of study risk estimates. Thus, there is a need to quantify the potential impact of competing outcomes on risk estimates derived from this population.

This Thesis will evaluate the effect of incorporating competing risks methodology on measures of absolute and relative risk in patients with T2DM. For the sake of simplicity, this Thesis will focus on a single statistical interpretation of probability (frequentist), two disease outcomes (cardiovascular and cancer mortality) and a single risk factor (albuminuria). All other scenarios are considered outside the scope of this Thesis.

1.4. Overview of Thesis

1.4.1. Thesis Objectives

1. To quantify the association between albuminuria and the risks of cardiovascular and cancer mortality in T2DM.
2. To derive and implement methods to identify diabetes mellitus and classify T2DM using Clinical Practice Research Datalink (CPRD) data.
3. To evaluate how competing risks methodology affects absolute risk estimates of cardiovascular and cancer mortality in patients with T2DM, with and without albuminuria.
4. To evaluate how competing risks methodology influences estimates for the effect of albuminuria on the relative hazards of cardiovascular and cancer mortality in patients with T2DM.

1.4.2. Thesis Chapters

Chapter 2: What is the Impact of Albuminuria on Cardiovascular and Cancer Mortality in Type 2 Diabetes?

Chapter 2 will address the first Thesis objective by systematically reviewing the literature to identify the risks posed by albuminuria towards each cardiovascular and cancer mortality. The risk estimates

of included studies will then be pooled by meta-analysis to serve as a point of comparison for the data-derived risk estimates of the later Thesis chapters.

Chapter 3: Identifying Diabetes Mellitus and Classifying Type 2 Diabetes Mellitus in Clinical Practice Research Datalink

Chapter 3 will address the second Thesis objective by establishing methods to identify T2DM in CPRD medical records. This will be performed in two parts: first, a review of the literature to determine the methods currently employed by researchers when identifying diabetes mellitus and classifying its sub-types in CPRD; second, the derivation and implementation of a T2DM classification strategy in CPRD data that will be guided by the findings of the review.

Chapter 4: Competing Risks and Measures of Absolute Risk

This chapter will use the methods outlined in the Chapter 3 to define a population with T2DM in CPRD. Using this population, the nine-year cumulative and absolute risks of cardiovascular and cancer mortality will be estimated through the use of cumulative risk estimators. The third Thesis objective will be addressed through comparing the risk estimates of standard and competing-risks-adjusted estimators, allowing inferences to be made on the effect of incorporating competing risks methodology on measures of absolute risk.

Chapter 5: Competing Risks and Measures of Relative Risk

This chapter will use the T2DM patient population identified in Chapter 4 to derive estimates for the effect of albuminuria on the nine-year relative hazards of cardiovascular and cancer mortality. The fourth Thesis objective will then be addressed through comparing risk estimates between standard and competing-risks-adjusted models for the effect of albuminuria on each outcome; allowing inferences to be made on the effect of incorporating competing risks methodology on measures of relative risk.

Chapter 6: Discussion

The main findings of the Thesis will be discussed in the context of the literature and their implications for future research considered.

2. What is the Impact of Albuminuria on Cardiovascular and Cancer Mortality in Type 2 Diabetes?

2.1. Background

2.1.1. Introduction

Albuminuria is an established risk factor for cardiovascular mortality[65–68]. However, consensus is lacking on the magnitude of its effect, with only a single systematic review estimate for the effect of albuminuria (hazard ratio (HR): 2.48)[67], and inconsistent review estimates for the effect of microalbuminuria (HR: 1.28 - 1.76) and macroalbuminuria (1.74 - 3.00)[65,67]. Studies have recently described links between albuminuria and cancer mortality[69–72], but the evidence is sparse, and the risks posed are less certain. Searches of the Prospective Register of Systematic Reviews (PROSPERO)[35] database have revealed three registered reviews determining the effect of albuminuria on cardiovascular mortality in patients with diabetes mellitus. However, the authors of these reviews failed to stratify their analyses by type of diabetes, thus, ignoring the aetiological mechanisms potentially underpinning the different albuminuria progression rates of type 1 diabetes mellitus (T1DM) and type 2 diabetes mellitus (T2DM)[73]. No reviews to determine the effect of albuminuria on cancer mortality currently exist in PROSPERO.

Through the process of systematic review, a solid evidence base may be consolidated for the relationship between albuminuria and both cardiovascular and cancer mortality in patients with T2DM. In turn, this will provide a point of comparison for the risk estimates derived from Clinical Practice Research Datalink (CPRD) data in Chapters 4 and 5.

2.1.2. Aims and Objectives

Primary Objective

To quantify the association between albuminuria and the risks of cardiovascular and cancer mortality in T2DM.

Specific Aims

i. To identify relative risk estimates, e.g. HRs, relative risks, odds ratios, for effect of albuminuria on cardiovascular and cancer mortality.
ii. To pool risk estimates via meta-analysis.
iii. To investigate potential sources of heterogeneity through meta-regression, and the use of stratification variables in meta-analyses.
iv. To examine risk estimates for the presence of publication bias.

2.2. Methods

2.2.1. Data Sources and Searches

The CINAHL[31] database, and Ovid versions of the MEDLINE[32] and EMBASE[33] databases were systematically searched from their inception until 13th January 2015. Relevant studies were identified through a search strategy incorporating keywords and search terms; medical subject heading terms for CINAHL and MEDLINE, and EMTREE terms for EMBASE (see Appendix B.1). Search terms and keywords were derived from those used in the methodological systematic review in Chapter 1. Additional searches were also performed in PROSPERO and PubMed[34] to identify studies from previously published systematic reviews assessing the effect of albuminuria on cardiovascular or cancer mortality.

The search strategy implemented consisted of a combination of keywords and subject headings for T2DM, cancer mortality, and cardiovascular mortality. These were combined to form a structure

whereby studies were required to have at least one keyword or search term for T2DM, in addition to at least one keyword or search term for cardiovascular mortality or cancer mortality. A search filter designed to identify only randomised controlled trials (RCTs) and observational studies conducted on human participants was constructed by combining the established filters of the Cochrane Collaboration[74], the British Medical Journal[75], and the Scottish Intercollegiate Guidelines Network[76] (Appendix B.1).

2.2.2. Study Selection

Studies were eligible for inclusion if they were of trial or cohort design with a minimum of 2 years of follow-up and 500 subjects, and if the study full-text was available in English.

Population

Studies were required to consist of adults (aged 18 years or over) with T2DM, but without pre-existing cardiovascular disease (CVD), cancer, end-stage renal disease, total pancreatectomy or renal replacement therapy. Studies without any subjects with both T2DM and albuminuria were excluded. Studies not including any subjects without CVD, cancer, end-stage renal disease, total pancreatectomy or renal replacement therapy were also excluded.

Exposure and Comparator

Studies were eligible for inclusion if albuminuria had been assessed as an:

- Albumin-creatinine ratio (ACR)
- Protein-creatinine ratio (PCR)
- 24-hour albumin excretion rate (AER)
- 24-hour protein excretion rate (PER)

Studies, in which albuminuria was assessed by means of “total urine albumin” or “total urine protein” output in the absence of a timescale qualifier, were ineligible for inclusion.

Outcomes

Studies were required to report either cardiovascular mortality or cancer mortality as outcomes.

2.2.3. Data Extraction and Quality Assessment

HRs or log-HRs from adjusted and unadjusted analyses were extracted for cardiovascular and cancer mortality. Where HRs or log-HRs were unavailable, relative risks then odds ratios were used. Where albuminuria was encoded as a binary variable, data were extracted providing the levels of the binary variable followed one of the structures described in Table 2.1. There were no restrictions for continuous variable structures. If not provided, confidence intervals (CIs) were derived using:

- Standard errors
- Standard deviations
- Precise p-values (e.g. "p=0.00543")
- Group and event numbers

Table 2.1 - Binary albuminuria variable structures eligible for inclusion.

Exposure Albuminuria Level	**Comparator Albuminuria Level**
Microalbuminuria	Normoalbuminuria
Macroalbuminuria	Normoalbuminuria
Albuminuria	Normoalbuminuria

A sample of eligible full-texts was used to pilot data extraction forms. In studies where multiple risk estimates from the same population were present, estimates were selected based upon the similarity of the underlying models used to those implemented in Chapter 5, i.e. those originating from models with the most similar adjustment variables to: age, gender, body-mass index (BMI), glycated haemoglobin (HbA1c), systolic blood pressure (SBP), total-to-high-density-lipoprotein (Total : HDL) cholesterol ratio, and smoking status; and the closest follow-up duration to nine years. Where RCTs presented risk estimates for both the intervention and placebo group, only risk estimates from the placebo group were extracted.

The results of the review were reported in accordance with the latest available version of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist[77].

Study Quality Assessment

The Quality in Prognostic Studies tool has six bias domains through which bias may be assessed in prognostic studies[78–80]:

- Study Participation
- Study Attrition
- Prognostic Factor Measurement
- Outcome Measure
- Study Confounding
- Statistical Analysis and Reporting

Study bias was assessed using the criteria listed under these bias domains in the most recent revision of the Quality in Prognostic Studies tool[78]. A detailed list of the criteria present under each domain may be found in this checklist.

2.2.4. Data Synthesis and Analysis

Where multiple studies using data from the same source cohort had been extracted (e.g. for RENAAL[64] and ZODIAC[81]) only data from the publication in which analyses were closest to those performed in Chapter 5 were entered into meta-analyses models.

Random-effects meta-analyses models were used to derive pooled estimates for the effect of albuminuria on the cause-specific hazards of cardiovascular and cancer mortality. Where albuminuria was stratified by other risk factors in studies, estimates from all strata were pooled by inverse-variance-weighted fixed-effect methods prior to use in random-effects models. The results from studies for which this was performed may be found in Appendix B.2. Prediction intervals for the 95% predictive distribution of a future study were calculated for all meta-analysis summary

estimates where the number of studies was greater than two[82]. Heterogeneity between studies was assessed using the I^2 statistic. Where substantial between-study heterogeneity existed ($I^2 \geq 75\%$) in analyses, pooled summary estimates were not displayed in forest plots. Potential sources of heterogeneity were explored through the use of meta-regression. For meta-analysis models, the between-study component of heterogeneity was estimated using the DerSimonian-Laird[83] method, while the analogous "method of moments"[84] method was used for meta-regression models.

Secondary analyses were conducted to determine the effect of the alternative albuminuria variable structuring from that detailed in the primary analysis. Random-effects meta-analysis models were performed for each binary and continuous variable structure of albuminuria encountered. Heterogeneity was evaluated as per the methods described in the primary analysis.

The presence of publication bias in prognostic studies is well acknowledged[85]. Egger's test[86] was used in conjunction with Funnel plots[87] to determine the presence of publication bias among studies.

All statistical analyses were performed using the metan[88] metabias[89] and metafunnel[90] packages in Stata (version 12.1)[91].

2.3. Results

Of the 2,758 studies returned by the initial database searches, 860 were duplicate entries. After the remaining 1,898 studies were screened, 32 studies were eligible for inclusion in the review. Four of these studies were later dropped as they constituted re-analyses of the same study populations using the same exposure and outcome structures as other studies eligible for inclusion. Of the remaining 28 studies: 13 were eligible for inclusion in the primary analysis, in which the effect of albuminuria (vs normoalbuminuria) was evaluated as a binary variable; 14 were eligible for inclusion in the secondary analyses, in which the effects of microalbuminuria or macroalbuminuria (vs normoalbuminuria) were evaluated as binary variables, or the effect of albuminuria was evaluated as a continuous variable; and one was eligible for inclusion in both. Database searches in PROSPERO

and PubMed identified four similar systematic reviews. No further eligible studies were identified within these reviews. A diagram of the study selection process is shown in the PRISMA flow diagram of Figure 2.1.

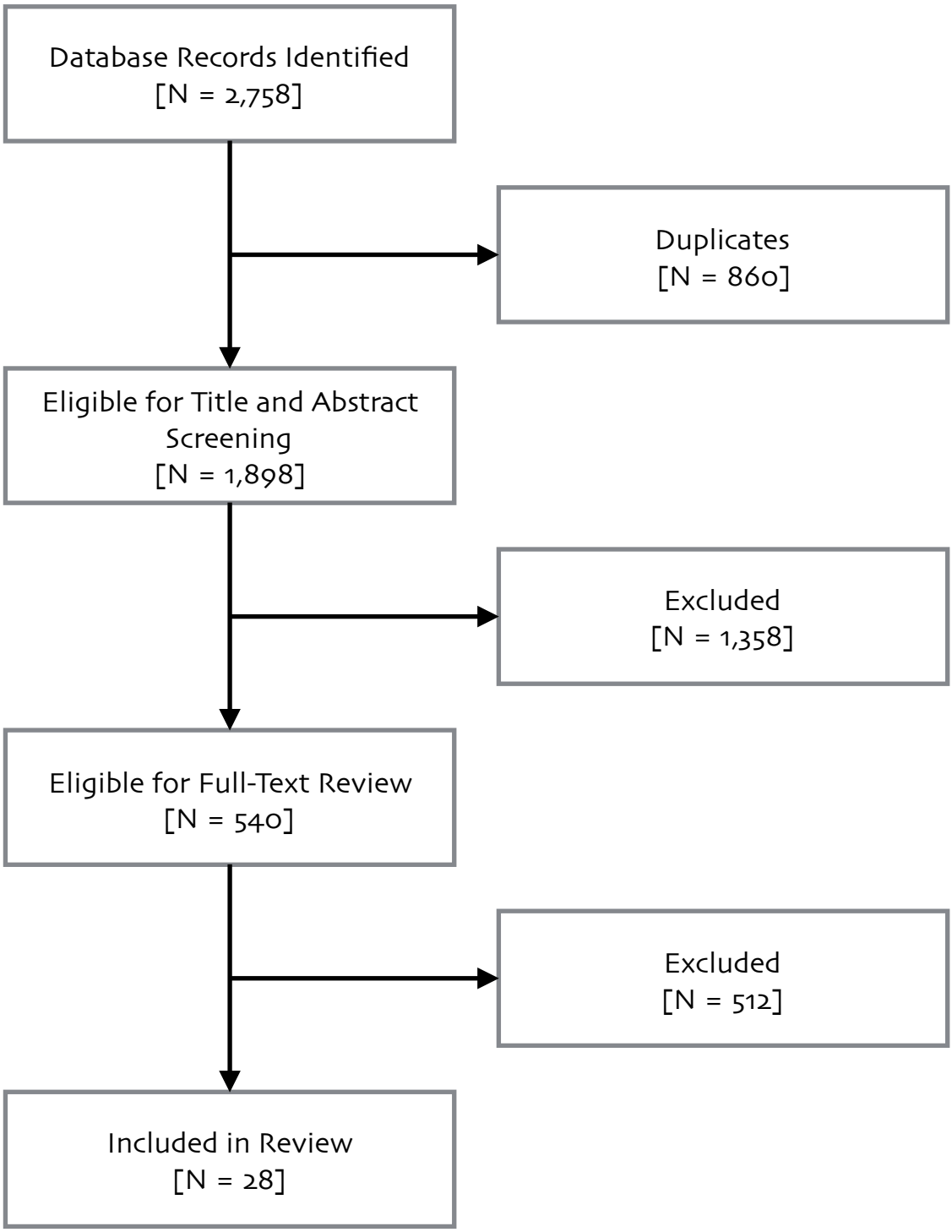


Figure 2.1 - PRISMA flow diagram representing the selection process.

2.3.1. Cardiovascular Mortality

Table 2.2 shows the baseline characteristics of 93,283 subjects from the 14 studies eligible for inclusion in the primary analysis. Average follow up ranged from 2.1 to 14.1 years.

Table 2.2 - Study characteristics of subjects in studies eligible for inclusion in the primary analysis.

First Author	Publication Year	Study Cohort	T2DM Population (N)	Cardiovascular Deaths (N)	Mean Age (Years)	Male (%)	Mean BMI (kg m-2)	Mean SBP (mm Hg)	Mean HbA1c (%)	Mean Diabetes Duration (Years)
Bo S[92]	2006	.	3,892	294	69.2	58.8	28.4	.	7.4	2.0
Bruno G[93]	2013	Casale Monferrato	1,825	175	67.6	49.0	.	146.0	7.0	10.7
Drion I[94]	2012	ZODIAC	810	344	74.3	38.9	28.2	157.7	7.5	6.8
Isomaa B[95]	2001	BOTNIA	1,697	168	58.6	52.6	29.5	137.2	7.5	.
Jager A[96]	1999	Hoorn Study	173	.	.	.	.	.	.	.
Jiminez-Corona A[97]	2006	.	1,605	190	.	42.2	.	.	.	.
Juutilainen A[98]	2006	.	1,059	271	57.5	54.9	29.4	.	.	7.5
Ko G[99]	2006	.	5,205	164	59.1	43.8	25.0	136.0	7.8	6.0
Lin C[100]	2012	Taichung Diabetes Study	6,523	105	54.2	51.2	.	.	8.6	5.2
Nelson R[101]	1988	.	1,426	.	.	.	.	.	.	.
Pavkov M[102]	2005	.	1,359	38	53.2	.	.	.	.	7.0
Sievers S[103]	1999	.	998	100	45.0	.	31.9	.	.	.
Svensson M[104]	2013	Swedish National Diabetes Register	66,065	1,590	63.9	58.9	29.4	140.0	7.1	7.4
Yu T[105]	2013	.	646	59	62.0	49.5	24.6	135.5	7.7	10.1

For adjusted risk estimates, the presence of albuminuria in subjects was associated with a twofold increase in the risk of cardiovascular mortality (HR: 2.07; 95% CI: 1.75 - 2.45) when compared with subjects with normoalbuminuria, (Figure 2.2). Prediction intervals suggest that future studies will identify a HR between 1.24 and 3.47 on 95% of occasions. Substantial heterogeneity was present between studies (I^2 = 63.0%, p=0.001).

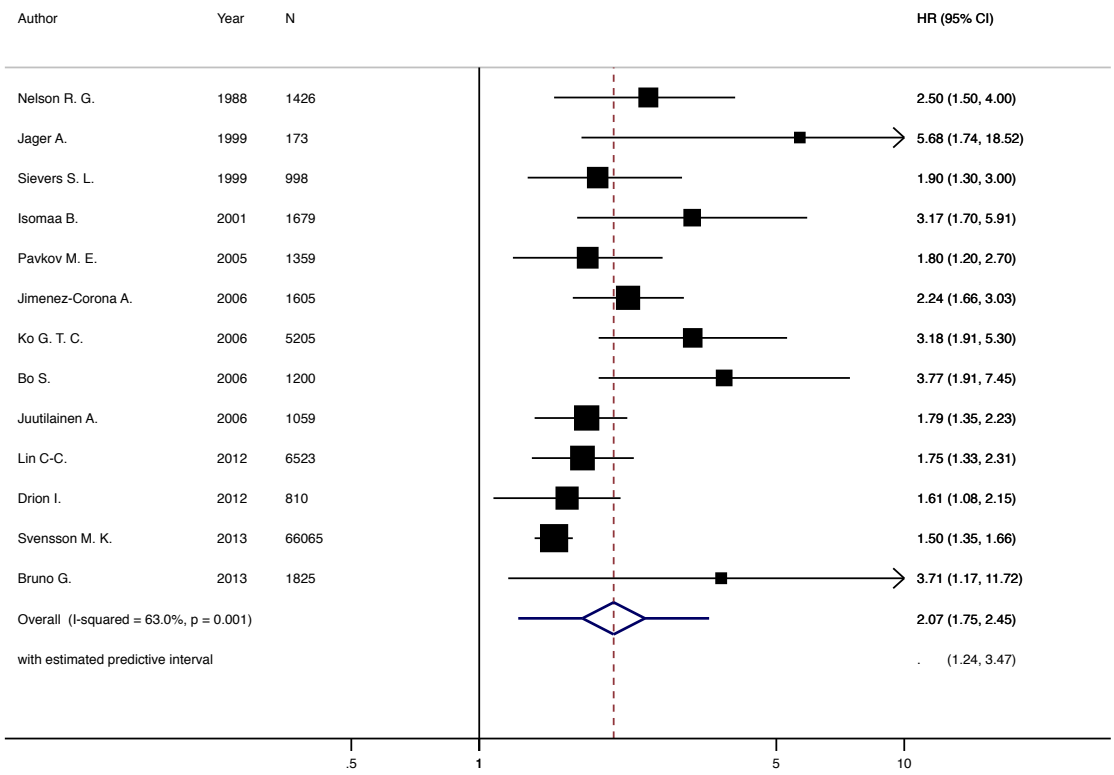


Figure 2.2 - Random-effects meta-analysis (including 95% prediction intervals[82]) for the effect of albuminuria on the relative hazard of cardiovascular mortality.

For unadjusted HRs, the pooling of estimates by meta-analysis was not performed as there were insufficient studies reporting this measure (N = 2).

Egger's test determined the presence of statistically significant asymmetry among studies evaluating the effect of albuminuria on cardiovascular mortality (P <0.001). The presence of publication bias was also suggested by the funnel plot in Figure 2.3.

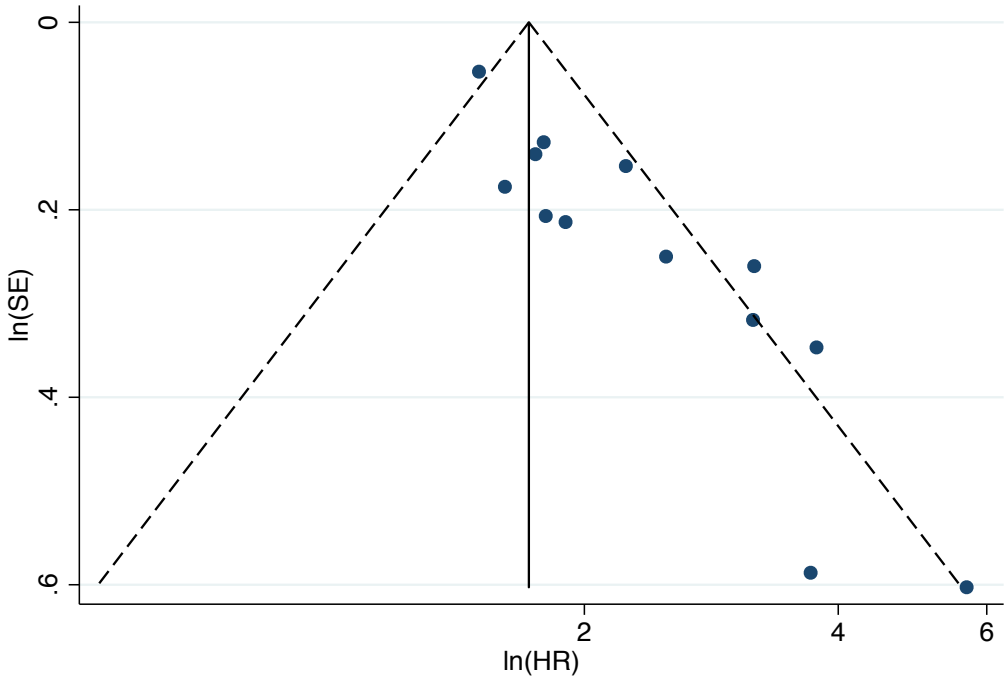


Figure 2.3 - Funnel plot with pseudo-95% CIs for the effect of albuminuria on cardiovascular mortality.

The addition of explanatory variables to the intercept-only meta-regression models did little to account for the high levels of between-study heterogeneity. Absolute change in the I^2 statistics ranged from a decrease of 11.29% to an increase of 6.42% (Figure 2.4). In all cases, the explanatory variables under assessment demonstrated HRs with no statistically significant effect.

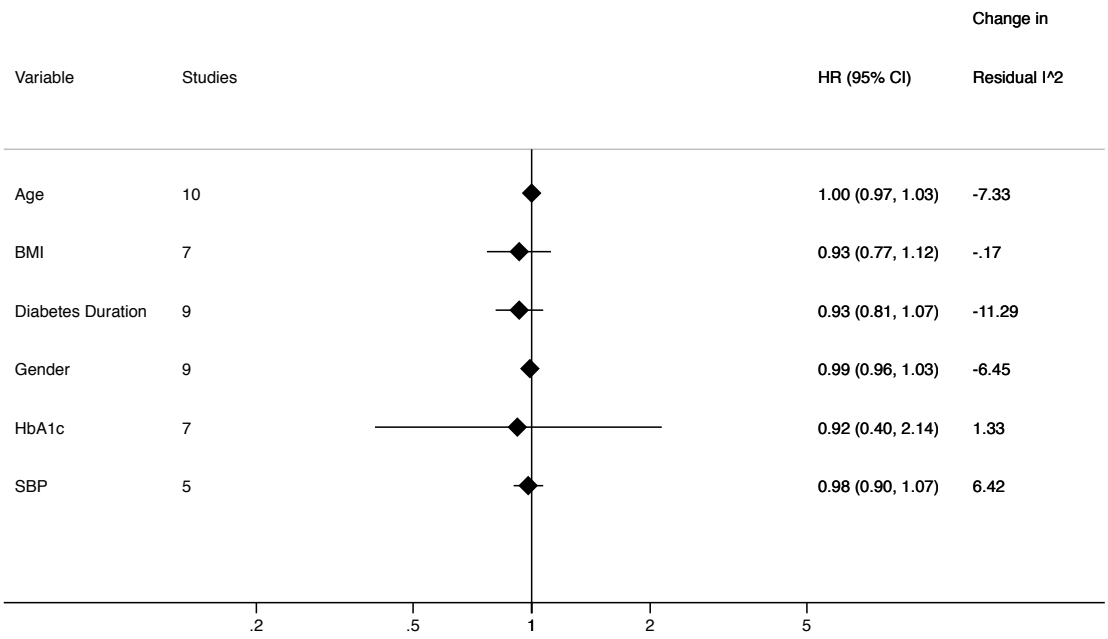


Figure 2.4 - Results of univariable meta-regression models. The change in residual I^2 represents the change from the intercept only model to the univariable model. Negative values indicate a decrease in residual heterogeneity.

Of the 15 studies eligible for inclusion in secondary analyses, nine evaluated albuminuria as a binary variable, four evaluated albuminuria as a continuous variable, and two evaluated albuminuria as both a binary and continuous variable. Table 2.3 shows the baseline characteristics of 111,788 subjects present in the studies eligible for the inclusion in secondary analyses. Average follow up ranged from 2.9 to 15.3 years.

Table 2.3 - Study characteristics of subjects in the studies present in the secondary analyses.

Lead Author	Publication Year	Study Cohort	T2DM Population (N)	Cardiovascular Deaths (N)	Mean Age (Years)	Male (%)	Mean BMI (kg m^{-2})	Mean SBP (mm Hg)	Mean HbA1c (%)	Mean Diabetes Duration (Years)
Beilin J[106]	1996	.	666	80	63.0	47.2	28.0	154.0	10.0	13.0
Bruce D[107]	2005	Freemantle	1,273	.	64.1	48.7	29.6	151.0	7.4	4.0
Bruno G[108]	2005	Casale Monferrato	1,565	324	68.9	.	.	.	.	10.8
Casiglia E[109]	2000	.	683	68	63.2	50.2	27.7	156.4	.	.
de Fine Olivarius N[110]	2010	.	1,323	164	.	.	.	150.0		0.0
Drury P[111]	2011	FIELD	9,795	266	60.2	66.0	30.6	139.0	7.0	4.8
Eckinci E[112]	2011	.	638	75	64.0	56.0	.	.	.	11.0
Fuller J[113]	2001	WHO MSVDD	3,179	.	46.9	47.2	.	.	.	7.7
McEwen L[114]	2012	TRIAD	8,334	.	.	.	.	.	.	.
Ninomiya T[115]	2009	ADVANCE	10,640	.	66.0	57.0	28.3	145.0	7.5	7.0
Nobakhthaghighi N[116]	2006	ABCD	880	35	58.6	.	31.5	145.3	11.6	8.9
Svensson M[104]	2013	Swedish National Diabetes Register	66,065	.	63.9	58.9	29.4	140.0	7.1	7.4
Targher G[117]	2011	Verona Diabetes Study	2,823	2,823	67.7	54.8	28.4	140.0	7.5	15.0
Valmadrid C[118]	2000	WESDR	840	431	67.9	45.0	29.4	143.3	9.3	15.1
Zoppini G[119]	2010	Verona Diabetes Study	3,084	.	67.5	45.1	28.2	.	7.5	15.5

For adjusted risk estimates, the presence of microalbuminuria and macroalbuminuria were associated with an increased risk of cardiovascular mortality 1.50 (95% CI: 1.27 - 1.76) and 2.22 (95% CI: 1.79 - 2.76) times higher than the presence of normoalbuminuria (Figure 2.5). There was no overlap between the pooled estimates of effect for the two albuminuria stages, indicating that the effects of microalbuminuria and macroalbuminuria statistically differed. The increase in the magnitude of effect for more severe stages of albuminuria indicates the presence of a 'dose-dependent' relationship between albuminuria and cardiovascular mortality. Prediction intervals suggest future studies may identify HRs between 0.93 - 2.41 for microalbuminuria and 1.15 - 4.30 for macroalbuminuria, on 95% of occasions. Substantial heterogeneity was present between studies for both albuminuria stages, with I^2 values of 60.9% ($p=0.004$) and 71.2% ($p<0.001$) for microalbuminuria and macroalbuminuria, respectively.

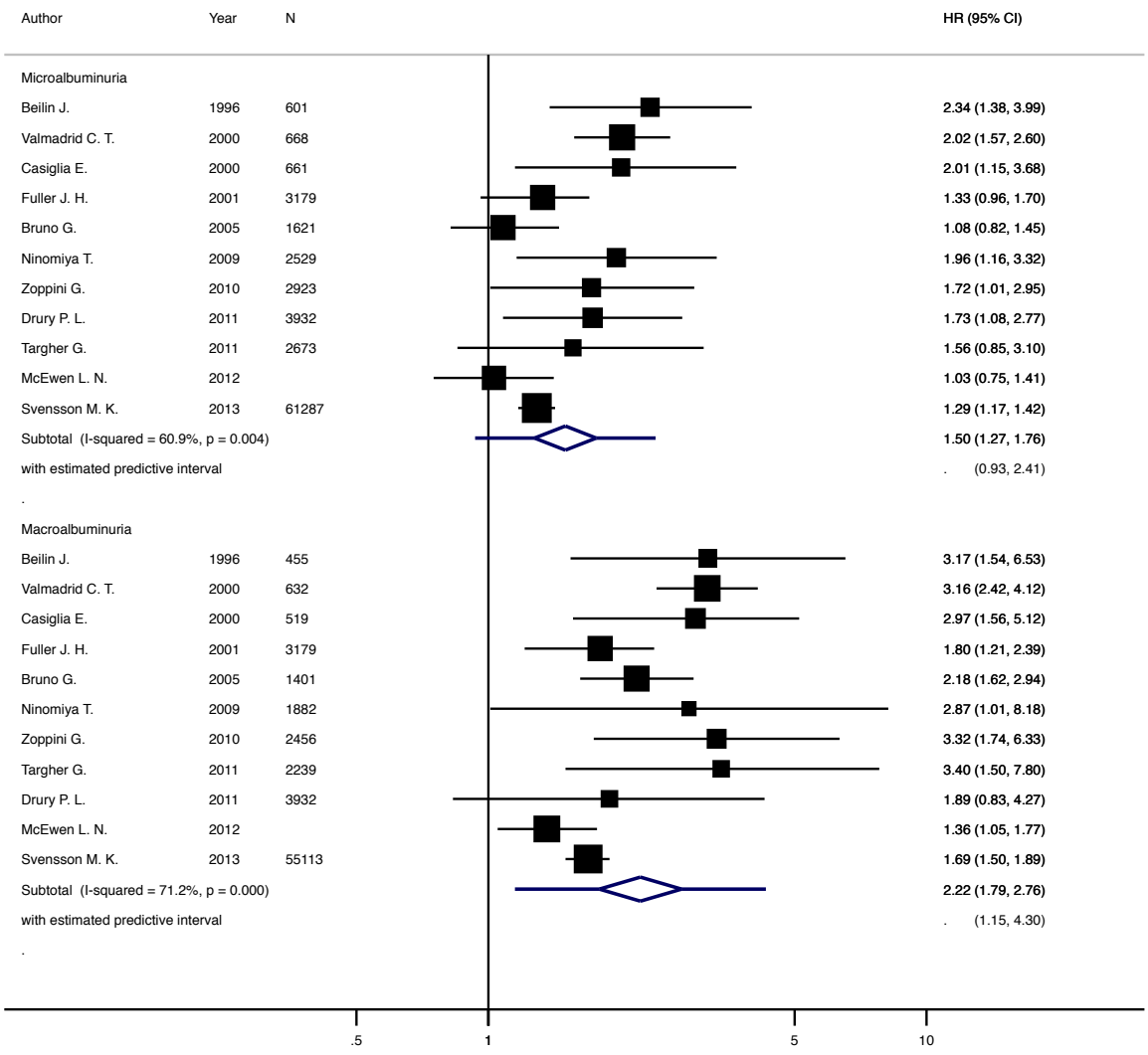


Figure 2.5 - Random-effects meta-analysis (including 95% prediction intervals[82]) for the effect of microalbuminuria and macroalbuminuria on the relative hazard of cardiovascular mortality.

For unadjusted HRs, the pooling of estimates by meta-analysis was not performed as there were insufficient studies reporting this measure (N = 2 for microalbuminuria and macroalbuminuria).

It was not possible to perform a meta-analysis of continuous measures of albuminuria, as too few studies reported consistent continuous measures of albuminuria, i.e. measures possible to be combined on a common scale.

A large degree of uniformity was demonstrated by studies across all six of the bias domains evaluated in the Quality in Prognostic Studies tool (Table 2.4). As such, no criteria or domains were

deemed suitable for use as stratification variables to investigate the large between-study heterogeneity demonstrated in the primary analysis.

Table 2.4 - Study quality assessed using the Quality in Prognostic Studies tool . Key: "T" indicates the total number of criteria within a bias domain where bias was not indicated; "Y" indicates that the requirement was fully satisfied; "P" indicates that the requirement was partially satisfied; "N" indicates that the requirement was not satisfied; "U" indicates that there was uncertainty regarding whether or not the requirement was satisfied.

Lead Author	Year	1 - Study Participation							2 - Study Attrition					
		a	b	c	d	e	f	T	a	b	c	d	e	T
Bo S.	2006	Y	Y	Y	Y	Y	N	5	Y	N	N	N	U	1
Bruno G.	2013	Y	Y	P	P	P	N	3.5	Y	N	N	N	U	1
Drion I.	2012	Y	Y	Y	Y	Y	Y	6	Y	N	N	N	U	1
Isomaa B.	2001	Y	Y	Y	Y	Y	Y	6	Y	N	N	N	U	1
Jager A.	1999	Y	Y	Y	Y	P	Y	5.5	Y	N	N	N	U	1
Jiminez-Corona A.	2006	Y	Y	Y	Y	Y	Y	6	Y	N	N	N	U	1
Juutilainen A.	2006	Y	Y	Y	Y	Y	Y	6	Y	N	N	N	U	1
Ko G. T. C.	2006	Y	Y	Y	Y	Y	Y	6	Y	N	N	N	U	1
Lin C.-C.	2012	Y	Y	Y	Y	Y	Y	6	Y	N	N	N	U	1
Nelson R. G.	1988	Y	Y	P	Y	Y	Y	5.5	Y	N	N	N	U	1
Pavkov M. E.	2005	Y	Y	Y	Y	Y	Y	6	Y	N	N	N	U	1
Sievers S. L.	1999	Y	Y	Y	Y	Y	Y	6	Y	N	N	N	U	1
Svensson M. K.	2013	Y	Y	Y	Y	Y	Y	6	Y	N	N	N	U	1
Yu T.-Y.	2013	Y	Y	Y	Y	Y	Y	6	Y	N	N	N	U	1

Lead Author	Year	3 - Prognostic Factor Measurement							4 - Outcome Measurement			
		a	b	c	d	e	f	T	a	b	c	T
Bo S.	2006	Y	Y	Y	Y	Y	U	5	Y	Y	Y	3
Bruno G.	2013	Y	U	P	U	Y	U	2.5	Y	Y	Y	3
Drion I.	2012	Y	Y	Y	Y	Y	U	5	P	Y	Y	2.5
Isomaa B.	2001	Y	U	P	Y	Y	U	3.5	N	Y	Y	2
Jager A.	1999	Y	Y	P	Y	Y	U	4.5	Y	Y	Y	3
Jiminez-Corona A.	2006	Y	U	Y	Y	Y	U	3	Y	Y	Y	3
Juutilainen A.	2006	Y	Y	Y	Y	Y	U	5	Y	Y	Y	3
Ko G. T. C.	2006	Y	Y	Y	Y	Y	U	5	Y	Y	Y	3
Lin C.-C.	2012	Y	P	P	Y	Y	U	3	Y	Y	Y	3
Nelson R. G.	1988	Y	Y	P	Y	Y	U	4	Y	Y	Y	3
Pavkov M. E.	2005	Y	Y	Y	Y	Y	U	5	Y	Y	Y	3
Sievers S. L.	1999	Y	Y	Y	Y	Y	U	5	Y	Y	Y	3
Svensson M. K.	2013	Y	Y	Y	P	Y	U	4.5	Y	Y	Y	3
Yu T.-Y.	2013	N	Y	P	Y	Y	U	3.5	Y	Y	Y	3

Lead Author	Year	5 - Study Confounding								6 - Statistical Analysis and Reporting				
		a	b	c	d	e	f	g	T	a	b	c	d	T
Bo S.	2006	Y	Y	Y	Y	U	Y	Y	6	Y	U	Y	Y	3
Bruno G.	2013	P	Y	Y	N	U	Y	P	4	Y	Y	Y	Y	4
Drion I.	2012	Y	P	U	Y	U	Y	Y	4.5	Y	Y	Y	Y	4
Isomaa B.	2001	P	Y	Y	U	U	Y	P	4	Y	N	P	Y	2.5
Jager A.	1999	P	Y	Y	Y	U	Y	P	5	Y	Y	Y	Y	4
Jiminez-Corona A.	2006	P	Y	Y	Y	U	Y	P	5	Y	Y	Y	Y	4
Juutilainen A.	2006	P	Y	Y	Y	U	Y	P	5	Y	Y	Y	Y	4
Ko G. T. C.	2006	Y	Y	Y	Y	U	Y	Y	6	Y	Y	Y	Y	4
Lin C.-C.	2012	P	P	Y	Y	U	Y	P	4.5	Y	Y	Y	Y	4
Nelson R. G.	1988	P	Y	Y	Y	U	P	P	4.5	Y	N	P	Y	2.5
Pavkov M. E.	2005	P	P	P	Y	U	Y	P	4	Y	N	P	Y	2.5
Sievers S. L.	1999	P	P	U	Y	U	Y	P	3.5	Y	P	Y	Y	3.5
Svensson M. K.	2013	Y	Y	Y	P	U	Y	Y	5.5	Y	Y	Y	Y	4
Yu T.-Y.	2013	Y	N	U	U	U	Y	Y	3	Y	Y	Y	Y	4

2.3.2. Cancer Mortality

For cancer mortality, one study was eligible for inclusion in the primary analysis (Yu et al., 2013[105]) and one study was eligible for inclusion in secondary analyses (Drury et al., 2011[111]). Demographic breakdowns of the study populations and the findings of these studies may be found in Table 2.5 and Figure 2.6, respectively.

Table 2.5 - Study characteristics of studies investigating the effect of albuminuria on cancer mortality.

Lead Author	Publication Year	Study Cohort	T2DM Population (N)	Cancer Deaths (N)	Mean Age (Years)	Male (%)	Mean BMI (kg m^{-2})	Mean SBP (mm Hg)	Mean HbA1c (%)	Mean Diabetes Duration (Years)
Drury P[111]	2011	FIELD	9,795	315	60.2	66.0	30.6	139.0	7.0	4.8
Yu T[105]	2013	.	646	59	62.0	49.5	.	.	7.7	10.1

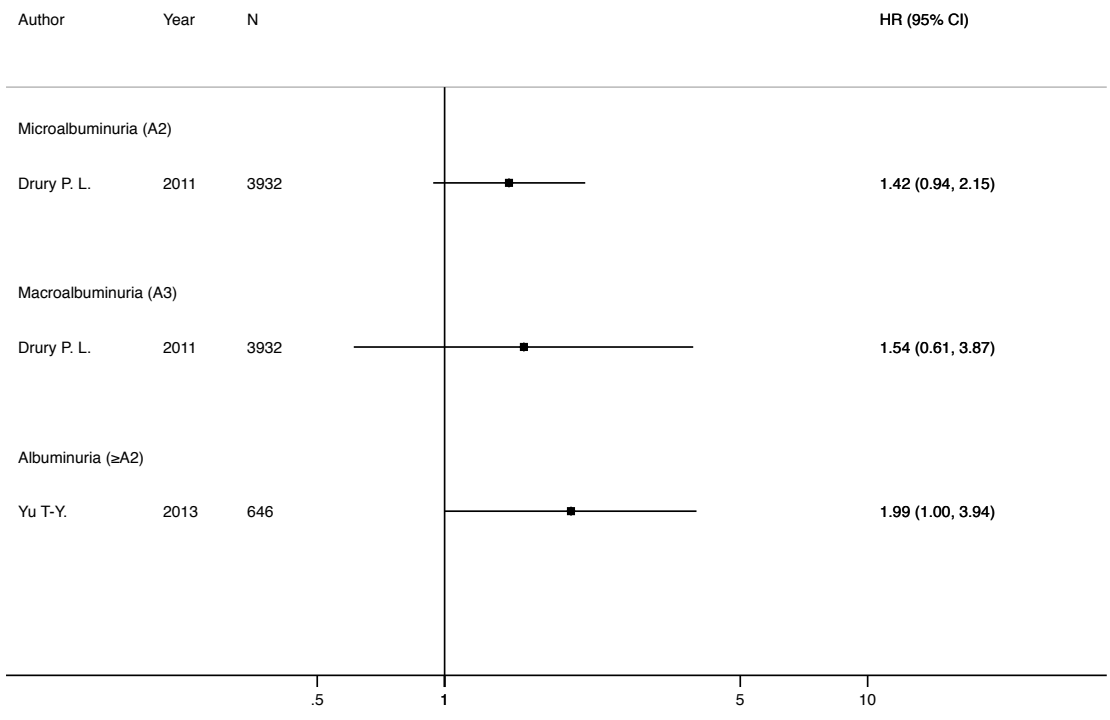


Figure 2.6 - Adjusted estimates for the effect of measures of albuminuria on cancer mortality.

Only Yu et al., 2013[105] reported unadjusted risk estimates for the effect of albuminuria on cancer mortality. This yielded a HR of 2.65 (95% CI: 1.52 - 4.62). No unadjusted risk estimates were reported for other measures of albuminuria.

2.4. Discussion

2.4.1. Key Results

This review confirms that an elevated level of albuminuria is a statistically significant risk factor for cardiovascular mortality in patients with T2DM. Secondary analyses evaluating the effect of microalbuminuria and macroalbuminuria reinforced these findings, demonstrating a dose-dependent response between albuminuria stage and cardiovascular mortality. Substantial between-study heterogeneity was present in the primary analysis (I^2 = 63.0%), and in both albuminuria stratifications in the secondary analyses (I^2 = 60.9% and 70.2%), indicating that albuminuria variable structure was not the source of this heterogeneity. The number of studies evaluating the effect of albuminuria on cancer mortality in patients with T2DM was not sufficient to perform a meta-analysis.

2.4.2. Strengths and Limitations

Search Criteria

The derivation and use of an un-validated hybrid search filter may have resulted in the omission of potentially relevant studies from the database searches. This filter was created and implemented as all screening stages had to be manageable for an individual researcher. Cross-referencing the studies identified by this review with those eligible for inclusion in three similar reviews[65–67] identified no additional studies eligible for inclusion in this review. Thus, the sensitivity of the hybrid search filters may also be viewed as a strength of this study.

The exclusion of studies consisting solely of high risk populations, like patients on renal dialysis, may have distorted the findings of the meta-analyses. The justification for this exclusion was that these patients did not constitute an accurate representation of most patients with T2DM, and may introduce

between-study heterogeneity. However, studies in which high-risk populations were excluded were eligible for inclusion in this review. The systematic exclusion of high-risk, but not low-risk populations may have altered pooled estimates of effect in the meta-analyses. However, as these high-risk groups generally only constitute a low proportion of patients with T2DM, any bias present is likely to be small.

The publication bias, apparent in the funnel plot in Figure 2.3, and the results of Egger's test, may have stemmed from the fact that in the majority of studies albuminuria constituted an adjustment variable rather than the primary study exposure. As such, its inclusion in analysis models was often conditional upon its significance in univariable or stepwise models. There is no easy way to rectify this form of publication bias, as unlike with the "file drawer problem"[120] (where the results of non-significant research are undesirable to publish) the analysed results for albuminuria would not reside in unpublished papers, but would be omitted from published ones.

Study Populations

Subjects in eligible studies represented older individuals from cohorts that were observed in the 1990s. Thus, inferences drawn from these studies may not be representative of patients with T2DM today, where the disease may have an earlier onset[121]. The evaluation of cardiovascular and cancer mortality in more recent population cohorts may help to address these issues for future systematic review updates in this area.

Analyses

Substantial heterogeneity was present in all the analyses of this review, and few studies were present with which to explore its possible sources. Thus, the univariable meta-regression models may have had insufficient power to determine statistical significance for any of the risk factors evaluated.

2.4.3. Relationship to the Literature

Two analyses by the United Kingdom Prospective Diabetes Study (UKPDS) potentially relevant to this review were deemed ineligible for inclusion after full-text review; UKPDS 64[29] and UKPDS 82[122]. In both cases, the studies contained no extractable data to pool. Three previous reviews have assessed the effect of albuminuria on cardiovascular mortality in patients with any form of diabetes mellitus[65–67], while a further review has assessed the effect of albuminuria on cardiovascular morbidity and mortality (as a composite outcome) in patients with non-insulin-dependent diabetes[68]. However, this is the first study to systematically review the literature to evaluate the effect of albuminuria on cardiovascular mortality in T2DM, and the first to attempt to review the literature to evaluate the effect of albuminuria on cancer mortality in patients with T2DM.

Comparable numbers of eligible studies were identified in this review (N = 28) and the three reviews by Fox et al., 2012[65] (N = 43), Savarese et al., 2014[66] (N = 30), and Toyama et al., 2013[67] (N = 13), while far less eligible studies were identified in the review by Dineen et al., 1997[68] (N = 12). The lower number of eligible studies identified in the review by Dineen et al.[68] may be explained by the fact that it was conducted before the publication of all but one of the studies included within the current review. Comparisons between the included studies of these reviews and the current one were complicated by the use of dissimilar eligibility criteria (Table 2.6). However, no additional studies were identified from those present within the previously conducted reviews.

Table 2.6 - Populations, Exposures and Outcomes of the four similar reviews by Fox et al.[65], Savarese et al.[66], Toyama et al.[67] and Dineen et al.[68].

Study	Studies (N)	Population	Exposures	Outcomes
Dineen et al., 1997[68]	12	Patients with non-insulin-dependent diabetes	Albuminuria	All-cause mortality, Cardiovascular morbidity and mortality
Fox et al., 2012[65]	43	Patients with any type of diabetes, Patients without diabetes	eGFR, Albuminuria	Cardiovascular mortality, All-cause mortality
Savarese et al., 2014[66]	30	Patients with any type of diabetes	Change in albuminuria	Cardiovascular events, Cardiovascular mortality (and sub-types of each)
Toyama et al., 2013[67]	13	Patients with any type of diabetes	eGFR, Albuminuria	Cardiovascular mortality, All-cause mortality, Renal events
This review	28	Patients with T2DM	Albuminuria	Cardiovascular mortality, Cancer mortality

Savarese et al.[66] investigated the effect of change in baseline albuminuria (as opposed to the effect of the magnitude of baseline albuminuria) on cardiovascular mortality. In their review, a 10% increase in baseline albuminuria was associated with a statistically significant increase in the risk of cardiovascular mortality (HR: 1.06; 95% CI: 1.03 - 1.09). Dineen et al.[68] investigated the effect of albuminuria on the composite outcome of cardiovascular morbidity and mortality (as opposed to just cardiovascular mortality). In their review, the presence of albuminuria was associated with a statistically significant increase in the risk of cardiovascular morbidity and mortality (OR: 2.0; 95% CI: 1.4 - 2.7).

In the remaining two reviews of patients with diabetes, strong evidence is presented for the association between albuminuria and cardiovascular mortality. Both Fox et al.[65] and Toyama et al.[67] described statistically significant HRs for the effect of macroalbuminuria on the risk cardiovascular mortality (Table 2.7). For the effect of albuminuria, only Toyama et al.[67] provided a risk estimate. In all prior reviews, risk estimates for cardiovascular mortality were not stratified by type of diabetes. Fox et al.[65] cited an inability to differentiate type of diabetes in their review. In the other

reviews, data regarding the type of diabetes were either not collected[66], or were collected but not used to provide stratified risk estimates in meta-analyses[67].

Table 2.7 - Pooled HRs for the effect of albuminuria on cardiovascular mortality described by systematic reviews in populations with diabetes. *Population represents all subjects with diabetes present across any analysis evaluating cardiovascular mortality.

Review	Analysis Population*	Albuminuria HR (95% CI)	Microalbuminuria HR (95% CI)	Macroalbuminuria HR (95% CI)
Fox et al., 2012[65]	128,505	.	1.81 (1.62 - 2.01)	2.44 (1.99 - 2.98)
Toyama et al., 2013[67]	19,745	2.48 (1.57 - 3.91)	1.76 (1.38 - 2.25)	3.00 (2.36 - 3.82)
This review	134,621	2.07 (1.75 - 2.45)	1.50 (1.27 - 1.76)	2.22 (1.79 - 2.76)

The effects of albuminuria on cardiovascular risk described in this review were lower than those described by Toyama et al.[67] and Fox et al.[65] (Table 2.7). This may be attributable to differences between the inclusion criteria of this review and the prior reviews; resulting in there being only five studies present in both this review and Toyama et al.[67], and two cohorts present in both this review and Fox et al.[65]. Neither Fox et al.[65] nor Toyama et al.[67] stated that they excluded studies conducted wholly in subjects constituting high risk groups, as was performed in this review through the exclusion of studies performed exclusively in subjects with either pre-existing CVD, cancer, end-stage renal disease, total pancreatectomy or renal-replacement therapy. Fox et al.[65] and Toyama et al.[67] also included studies in which patients had forms of diabetes other than T2DM, which were not included in this review. However, evidence is lacking as to whether or not a albuminuria may be differentially pathological in subjects with disparate forms of diabetes, and the vast majority of subjects in any given population with diabetes will be liable to have T2DM[123].

There were disparities in the evaluation of potential sources of heterogeneity in the three prior reviews. Savarese et al.[66] reported no significant heterogeneity in any pooled risk estimates. The inconsistency between the findings of Savarese et al.[66] and the findings of this review may be attributable to the fact that 'change in albuminuria' and 'albuminuria' are essentially different risk factors. Toyama et al.[67] stated the presence of moderate-to-high heterogeneity (I^2 = 40.0% -

66.0%), and stratified risk estimates by eGFR stage; however, the authors noted that neither this nor any of their other subgroup analyses could explain the source of the heterogeneity. Fox et al.[65] provided no estimates of between-study heterogeneity for albuminuria. In the current review, only the risk factors to be used in the survival models of Chapter 5 were evaluated as potential sources of heterogeneity. However, as previously stated, insufficient studies were present to determine the source of the heterogeneity. Thus, the heterogeneity may be attributable to any of the risk factors evaluated, or to other un-extracted components of the studies.

2.4.4. Implications

This review strengthens the conclusions of previous work describing associations between albuminuria and cardiovascular mortality in patients with T2DM. However, it also highlights the paucity of studies evaluating the effect of albuminuria on cancer mortality in patients with T2DM. Thus, more work is needed to assess the risks of cancer mortality in patients with T2DM and the role that albuminuria may play in predicting this outcome.

Chapter 5 will evaluate the effect of albuminuria on the relative hazards of cardiovascular and cancer mortality, in analyses both adjusted and unadjusted for the presence of competing risks. For cardiovascular mortality, derived risk estimates will be expected to fall between the prediction intervals for albuminuria, presented in Figure 2.2.

3. Identifying Diabetes Mellitus and Classifying Type 2 Diabetes Mellitus in Clinical Practice Research Datalink

Henceforth the identification of diabetes in medical records will be referred to as the "identification of diabetes", while the identification of a specific type of diabetes in medical records will be referred to as the "classification of diabetes".

3.1. Introduction

3.1.1. Diabetes Classification

In 1979 the National Diabetes Data Group published a paper seeking to standardise the definitions of diabetes mellitus in a system based on patients' treatment requirements[124], namely, insulin-dependent or non-insulin-dependent. This was later backed by the World Health Organisation as the gold standard for classification of diabetes[125,126]. However, as therapies evolved (e.g. the development of oral anti-diabetics) classification of the disease became progressively more difficult. The discovery of other forms of diabetes added further confusion and may have led to the misclassification of diabetes in patients.

In 1997, an international expert committee consisting of the American Diabetes Association and the World Health Organisation released a report with new recommendations for the classification of diabetes based on aetiology[127], (which herein will be referred to as the "aetiological" classification system). This remains the last major change in the classification of diabetes. Within this report, diabetes was classified under four main types:

- **Type 1 Diabetes (T1DM)** - characterised by destruction of pancreatic β-cells in an autoimmune process.
- **Type 2 Diabetes (T2DM)** - characterised by insulin secretory defect of pancreatic β-cells, and insulin resistance in peripheral tissue.
- **Gestational Diabetes** - characterised by the development of insulin insufficiency of insensitivity during pregnancy.
- **Other Specific Types** - various other forms of the disease, each characterised by its own specific aetiology

However, classification using these recommendations remains complex and difficult; an issue that has come to the fore with attempts to categorise patients where data is extracted from routinely collected clinical data.

3.1.2. Clinical Practice Research Datalink Data

Clinical Practice Research Datalink (CPRD)[128] is the world's largest database of anonymised primary health care records[129]. It consists of the records of over 14 million patients from 660 general practitioner (GP) practices across the UK. Its central repository may be linked to several external databases including Hospital Episode Statistics, Myocardial Ischaemia National Audit Project and the Office for National Statistics (ONS). Data collection from contributing GP practices began in 1980, potentially giving over 35 years of patient follow-up. The data is generated from the information computationally logged by GPs following their interactions with patients. For example, a patient visiting a GP for a prescription of an oral contraceptive will have records logged linking: the date and purpose of the visit, the medication prescribed, and information pertaining to blood pressure, body mass index, age and smoking status.

Clinical Practice Research Datalink Browser Dictionaries

CPRD issues two "browser dictionaries" to researchers, which act as graphical user interfaces to databases of "medical codes" and "product codes". The databases interfaced by the browser

dictionaries are compiled by the Health and Social Care Information Centre, and are updated with new codes every six months.

The Medical Browser Dictionary provides an interface to a database containing information on the medical codes usable by GPs. In this database, the "Read term" represents a plain text description of the purpose for a patient's visit, while the medical code represents a numerical identifier of this information. Keyword searches performed in the Read term field may be used to identify associated medical codes. Some sample entries that would be returned by the search term "**diab**" are illustrated in Table 3.1.

Table 3.1 - Sample medical codes with associated Read term field from the CPRD Medical Browser Dictionary.

Medical Code	Clinical Events	Read Code	Read Term
98723	19	C10FD11	Type II diabetes mellitus with hypoglycaemic coma
35288	300	C10E800	Type 1 diabetes mellitus - poor control
33254	1569	C105.00	Diabetes mellitus with ophthalmic manifestation

The "Product Browser Dictionary" provides an interface to a database of all the possible medications GPs may prescribe to patients. Each dosage of each medication is represented by a unique product code. In order to identify relevant anti-diabetic product codes, keyword searches may be performed in the following categories: the "Drug Substance Name" (the scientific name of the active drug compounds within the product), the "Product Name" (the name given to the drug product by the manufacturer), and the "BNF Header" (The British National Formulary (BNF) classification of the drug type). Some sample entries are illustrated in Table 3.2.

Table 3.2 - Sample product codes with entries from other associated fields present in the CPRD Product Browser Dictionary.

Product Code	Therapy Events	Drug Substance Name	Product Name	BNF Header
39560	1013	Metformin Hydrochloride	Bolamyn SR 500mg Tablets	Biguanides
12259	11	Glibornuride	Glibornuride 25mg Tablet	Sulphonylureas
9662	19593	Rosiglitazone Maleate	Avandia 4mg Tablets	Unknown

Clinical Practice Research Datalink Data Files

Four CPRD data files may be used to assist in the identification and classification of diabetes:

- ***Clinical File*** - contains all the medical history data (medical codes) entered in patient records, including symptoms, signs and diagnoses. This file can be used to identify any clinical diagnoses and the cause of mortality.
- ***Therapy File*** - contains data related to all prescriptions (product codes) for drugs and medical devices on in patient records.
- ***Patient File*** - contains basic patient information including date of birth, date of death and date of registration at the GP practice.
- ***Test File*** - contains information on any tests performed on patients, including the date, the specific test and the results from the test.

Linkage

The possible diagnostic data in the Read terms in the medical browser dictionary and the possible anti-diabetic products listed in the product browser dictionary may be linked to individual patient's records via the use of key linkage variables. These linkages are illustrated in Figure 3.1.

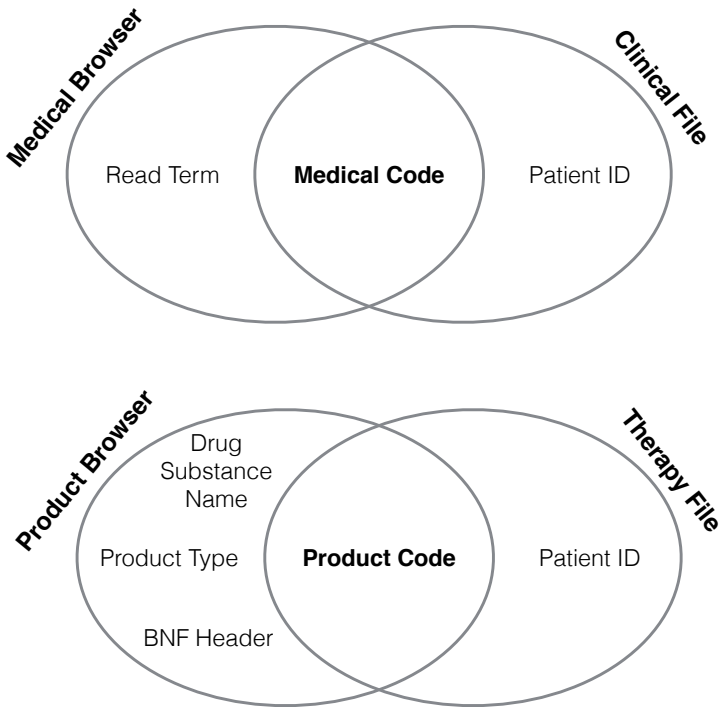


Figure 3.1 - Venn diagrams illustrating how data fields present in the CPRD Browser Dictionaries may be linked to both the Clinical and Therapy Files. The "Patient ID" is a unique identifier attributed to each patient in CPRD files.

3.1.3. Identifying and Classifying Diabetes in Clinical Practice Research Datalink

When using CPRD data there is no simple way to select a subset of patients with diabetes. All classification systems used by GPs since the inception of CPRD may be found in patients' medical records. Moreover, medical codes linked to these older classification systems are still present in the Medical Browser Dictionary, and may still be used by GPs when logging information regarding patient's consultations. Studies have identified large numbers of patients in primary care databases who were misclassified by GPs, i.e. contradictory codes given, or no evidence to support diagnosis[130]. Financial initiatives like Quality Outcomes Framework (QOF)[131], incentivise GPs to more consistently report diabetes using codes based on the aetiological classification system. However, some records in CPRD predate the arrival of the aetiological classification system, while some GPs may prefer to use the older classification systems (which map poorly to the aetiological one). Concomitantly, GPs may use type-agnostic medical codes (available in the QOF codes), e.g. "undefined DM", when a correct aetiological diagnosis is difficult to establish.

Potential Clinical Practice Research Datalink Components for Identifying and Classifying Diabetes

Several components of CPRD have been proposed to identify and classify diabetes:

- **Medical Codes for Diabetes** - diagnostic codes linked to the types of diabetes listed in section 3.1.1 or codes linked to undefined diabetes[130].
- **Product Codes** - codes related to pharmacotherapies used in the treatment of diabetes, e.g. insulin[130].
- **Diabetes Test Data** - QOF guidelines state that patients with diabetes should have the glycaemic control regularly tested[131]. Furthermore, certain laboratory or point of care tests are only ever performed for monitoring, rather than diagnostic purposes, e.g. HbA1c. Thus, both the presence and results from these tests may prove insightful in establishing the presence of diabetes[132].

- **Medical Codes for Diabetic Complications** - some medical conditions like diabetic ketoacidosis (DKA) and hyperosmolar nonketotic syndrome (HONK) may demonstrate strong association with a particular type of diabetes[133].
- **Medical Codes for Concurrent Conditions** - medical conditions like acanthosis nigricans, polycystic ovary syndrome (PCOS), obesity and hypertension may demonstrate associations with particular forms of diabetes[133].
- **Ethnicity** - ethnicity may be associated with the presence or type of diabetes[133].

3.1.4. Aims and Objectives

Primary Objectives

To derive and implement methods for querying CPRD data to identify diabetes and classify T2DM.

Specific Aims

i. To identify the current methods and strategies employed in the identification of diabetes, and in the classification of T1DM and T2DM, in studies using CPRD data.

ii. To identify medical codes for diabetes and product codes for anti-diabetic therapies in the CPRD Browser Dictionaries and to map:

   a. Medical codes to the aetiological classification system of diabetes.

   b. Product codes to classes of anti-diabetic therapy.

iii. To develop and implement a strategy to identify diabetes in CPRD patient records.

iv. To develop and implement a strategy to classify T2DM in the CPRD patient records.

v. To assess the validity of mapping further medical codes for diabetes to the aetiological classification system, and incorporate any new code mappings into the diabetes identification and T2DM classification strategies.

3.2. Review of the Literature

3.2.1. Methods

Search Methods and Eligibility Criteria

The CPRD publications database was searched for studies of any design published from 1st June 1987 to 29th June 2015. As the CPRD database is updated on a quarterly basis, the initially intended search was constrained to being from 1st June 1987 to 31st January 2015. All studies containing the keyword stem "*diabet*" in the title were considered eligible for review. Studies were included in the review if they defined diabetes, T1DM or T2DM as either an outcome or a risk factor in study analyses, or if these populations were defined for the purposes of providing summary tables or statistics.

Data Extraction

Two reviewers extracted data on the type(s) of diabetes identified or classified, in addition to whether any of the following CPRD components were used in the identification or classification:

- Medical codes
- Product codes
- Tests performed
- Test results
- Age at diagnosis

Additional means of identifying and classifying diabetes were also recorded, if present. Where diabetes was identified as part of an intermediate step in classifying a type of diabetes, or where a type of diabetes was classified as an exclusion group, no data were extracted, unless these were analysed or presented as summary data.

3.2.2. Results

The CPRD publications database listed 1,345 research papers; 103 of which were eligible for full-text screening. Of these, nine publications were excluded because they failed to specify how diabetes was defined, and one study was excluded as it defined a form of diabetes outside the scope of this review (i.e. gestational diabetes). Within the remaining 93 publications: 15 identified diabetes, 12 classified T1DM and 75 classified T2DM, as part of their analyses. The sum total of these studies exceeds 93, as some studies defined more than one diabetic population for analysis. The "Preferred Reporting Items for Systematic Reviews" (PRISMA) flow diagram representing the identification process is shown in Figure 3.2.

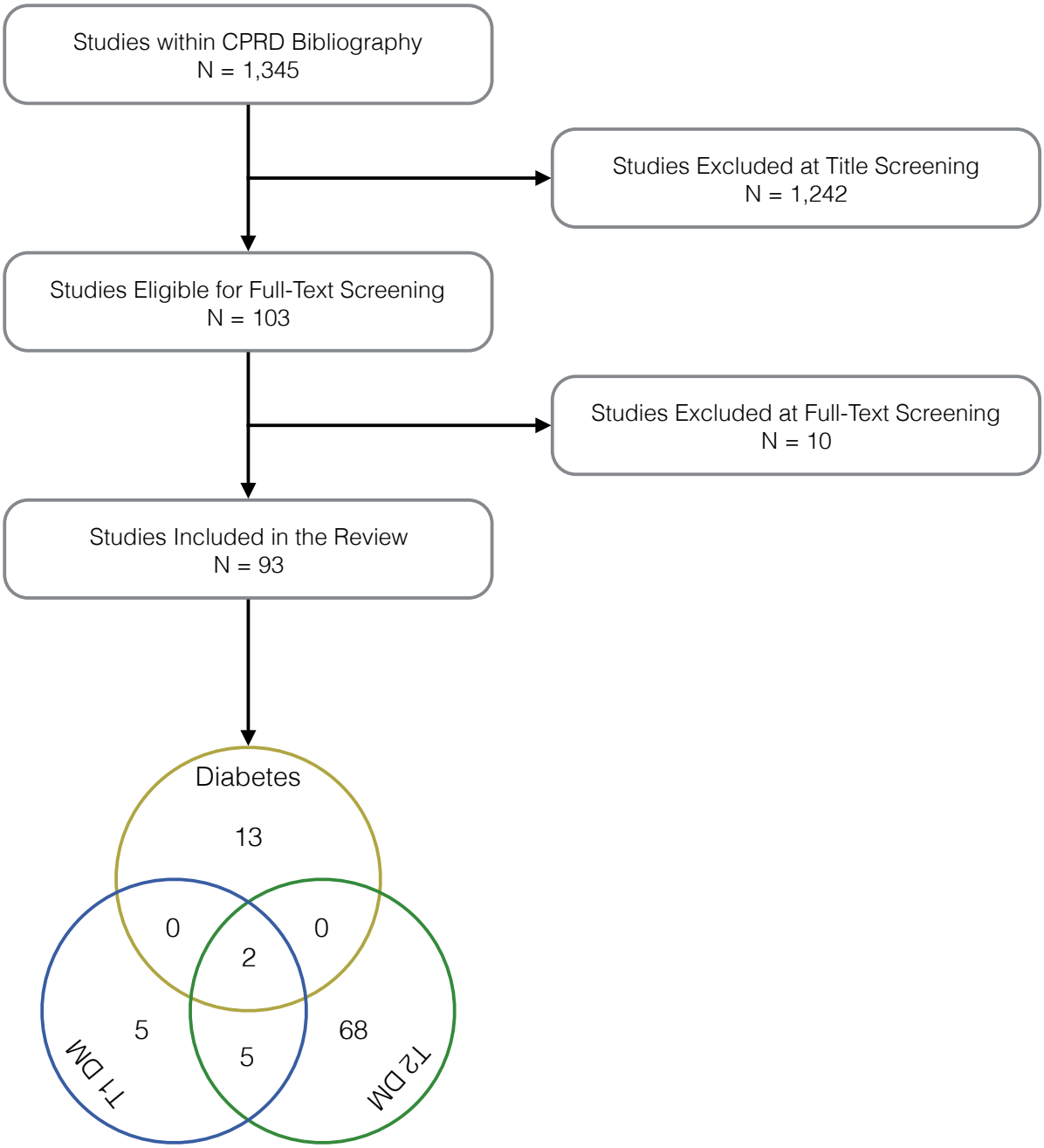


Figure 3.2 - PRISMA flow diagram to illustrate the review process.

Studies Identifying Diabetes

Table 3.3 depicts the CPRD strategies and components used in the identification of diabetes. The most common strategy incorporated both medical codes and product codes. These two components were present together in nine out of 15 identified studies. The second most common strategy involved the use of medical codes in isolation. This was performed by six out of 15 studies.

Table 3.3 - The strategies employed by studies to identify diabetes (grouped by strategy and ordered by the frequency of occurrence of each strategy).

	CPRD Components (n)						Studies (N)
	Medical Codes	Product Codes	Age	Test Presence	Test Value	Other	
Strategy	Y	Y	-	-	-	-	7
	Y	-	-	-	-	-	6
	Y	Y	-	-	-	Y	1
	Y	Y	-	Y	-	-	1
Total (n)	15	9	0	1	0	1	15

The most commonly used CPRD components, independent of strategy, were: medical codes (in all 15 studies), and product codes (in eight studies). The only other identification medium used was diabetes clinic notes.

Studies Classifying Type 1 Diabetes

Table 3.4 depicts the CPRD strategies and components used in the classification of T1DM. The most common strategy incorporated medical codes, product codes, and age. This was performed by six out of 12 identified studies. Medical codes and product codes were used without age by three out of 12 studies. No strategies used age without product codes, although four out of 12 studies used product codes without age in their classification strategies.

Table 3.4 - The strategies employed by studies to classify T1DM (grouped by strategy and ordered by the frequency of occurrence of each strategy).

	CPRD Components (n)						Studies (N)
	Medical Codes	Product Codes	Age	Test Presence	Test Value	Other	
Strategy	Y	Y	Y	-	-	-	6
	Y	Y	-	-	-	-	2
	-	Y	Y	-	-	-	2
	Y	Y	-	-	-	Y	1
	-	Y	-	-	-	-	1
Total (n)	9	12	8	0	0	1	12

The most commonly used CPRD components, independent of strategy, were: product codes (in all 12 studies), medical codes (in nine studies), and age (in eight studies). Where present, the most commonly

used age boundary was 35 years (range 15 - 35). The only other classification medium used was the number of GP visits.

Studies Classifying Type 2 Diabetes

Table 3.5 depicts the CPRD strategies and components used in the classification of T2DM. The most common strategy incorporated medical codes, product codes and age. This was performed by 24 out of 75 identified studies; while these three components featured together as part of the strategies of 28 out of 75 studies. 26 out of 75 studies reported strategies that used medical codes and product codes without age, although many of these also used further CPRD components, like the presence or results of tests. One out of 75 studies used age without product codes, and 29 out of 75 studies used product codes without age in their classification strategies.

Table 3.5 - The strategies employed by studies to classify T2DM (grouped by strategy and ordered by the frequency of occurrence of each strategy).

	CPRD Components (n)						Studies (N)
	Medical Codes	Product Codes	Age	Test Presence	Test Value	Other	
Strategy	Y	Y	Y	-	-	-	24
	Y	Y	-	-	-	-	18
	Y	-	-	-	-	-	8
	-	Y	Y	-	-	-	6
	-	Y	-	-	-	-	5
	Y	Y	-	-	Y	-	3
	Y	Y	Y	Y	-	-	3
	Y	Y	-	-	-	Y	2
	Y	-	-	-	-	-	2
	Y	Y	Y	Y	Y	-	1
	Y	Y	Y	-	Y	-	1
	Y	-	Y	-	-	-	1
	Y	-	-	Y	-	-	1
Total (n)	64	64	37	4	5	4	75

The most commonly used CPRD components, independent of strategy, were: medical codes (in 64 studies), product codes (in 64 studies), and age (in 37 studies). Where present, the most commonly used age boundary was 35 years (range 18 - 40). Other classification mediums used by studies included the number of GP visits, the presence of ketoacidosis, and diabetes clinic notes.

In the next section, the findings of the review will be used to develop a new strategy to classify T2DM in the CPRD medical records of a general patient population.

3.3. Classification of Type 2 Diabetes in Clinical Practice Research Datalink

3.3.1. Methods

Code-List Derivation

Browser Dictionary Search Strategy

The Medical and Product Browser Dictionaries were searched using keyword stems (fragments of keywords) to identify text associated with relevant medical and product codes, respectively. In the Medical Browser Dictionary the "Read term" text field was searched, whereas, in the Product Browser Dictionary the "drug substance name", "product name", and "BNF header" text fields were searched. An adaptive strategy was employed whereby keyword stems were used to identify further potential keyword stems, in addition to relevant codes. For the Medical Browser Dictionary, the base keyword stems used were: "*diab*" and "*insulin*", while for the Product Browser Dictionary the keyword stems were: "*antidiabetic*", "*insulin*" and "*metform*". Where further keyword stems were identified, the search was re-run using the existing and new keyword stems as part of an iterative process. The process was terminated once no further keyword stems were identified. The final lists of keyword stems are in Appendices C.1 and C.2. Code-lists generated by the searches were independently reviewed by BF and three GPs (HA, AF, and DM). Disagreements were resolved through consensus. The reviewed lists were subsequently compared to other code-lists for diabetes and anti-diabetic therapies present in the Nuffield Department of Primary Care Health Sciences, at the University of Oxford. The final versions of these lists may be found in Appendices C.3 and C.4.

Classification of Medical Codes

Where possible, medical codes returned from keyword searches in the Medical Browser Dictionary were classified as the following types of diabetes:

- T1DM
- T2DM
- Drug-Induced Diabetes
- Bronzed Diabetes (Haemochromatosis)
- Gestational Diabetes
- Monogenic Diabetes
- Malnutrition-Related Diabetes

Where medical codes conveyed the presence of diabetes but not the type, codes were indexed as "Undefined Diabetes".

Classification of Product Codes

All product codes returned from keyword searches in the Product Browser Dictionary were classified depending upon evidence presented in the associated "Product Name", "BNF Header" and "Drug Substance Name" fields. The following classifications were used:

- Insulin-Based Therapy - anti-diabetic medication containing insulin.
- Metformin Monotherapy - anti-diabetic medications, in which metformin constituted the sole drug class.
- Other Anti-Diabetic Therapy - other anti-diabetic medications, not fulfilling the above two criteria.

Study Population

There was no distinct patient population (N) for this aspect of the study. Instead, the population constituted the codes (n) present in the Medical and Product Browser Dictionaries.

Identifying Patients with Diabetes

Identifying Diabetes in Patient Data

To identify diabetes in patient records, the Clinical and Therapy Files were searched using the medical codes derived in the previous section of the Methods. Where the presence of the disease was marked by either medical codes in the Clinical File or product codes in the Therapy File, patients were determined to have diabetes. All codes found without an associated event date were discarded.

Metformin, a drug that is commonly prescribed to patients with diabetes, is also prescribed for other conditions such as PCOS. To avoid potential misclassification, patients whose records contained medical codes for PCOS were excluded from the study. These codes were identified by searching the Medical Browser Dictionary for the keyword stems "*pcos*" and "*polycystic*" in the Read term text field. Code-lists generated by the search were independently reviewed by BF, HA, AF, and DM. The final code-lists for PCOS may be found in Appendix C.5.

Study Population

The dataset comprised of a random sample of 300,000 general patients from the CPRD database whose records were linked to Office for National Statistics (ONS) mortality data (protocol 12_091R: "Optimal strategies for monitoring lipid levels in patients at risk or with cardiovascular disease: best marker for monitoring and cost effectiveness of different monitoring frequencies."). Patients were aged 40 years or more on 1st January 2006, and had at least one year of follow-up data, either before or after this date, up to the 31st December 2011 (or the latest date available).

The earliest date of eligible patient data was the latest from:

- The date the patient registered at their current GP practice.
- The date the GP practice became up-to-standard (UTS).

The latest date of eligible patient data was the earliest from:

- The date the GP practice last committed data to CPRD.

- The date the patient transferred out of the GP practice.
- The patient's date of death.

After applying the eligible data criteria, 299,994 patients were eligible for inclusion in the study.

Classifying Patients with Type 2 Diabetes

Removal of Gestational and Other Specific Types of Diabetes

Patient records in the Clinical File were searched for medical codes for: bronzed diabetes (haemochromatosis), gestational diabetes, drug-induced diabetes, monogenetic diabetes, and malnutrition-related diabetes. The medical codes used to classify these types of diabetes may be found in Appendix C.3. Patients whose records contained medical codes for any of these conditions were excluded.

Differentiating Between Type 1 and Type 2 Diabetes

To differentiate between T1DM and T2DM a stratified classification strategy was employed consisting of three tiers:

1. For patients whose records contained one or more medical codes associated with either T1DM or T2DM, they were classified accordingly. For patients who records contained medical codes associated with both T1DM and T2DM, they were classified according to the last occurrence of these codes (chronologically) in their records.
2. For patients whose diabetes was not classifiable via medical codes, classification was performed using of a combination of product codes and age. Where insulin-based therapies were prescribed before the age of 35, patients were classified as having T1DM. No other classifications were made using product codes and age.
3. For patients whose diabetes was not classifiable via either medical codes, or a combination of product codes and age, classification was performed on the basis of the high background prevalence of T2DM relative to other types of the disease. Within this, all remaining patients were classified as having T2DM.

Study Population

A dataset of 33,051 patients identified as having diabetes in the previous analysis section was used. The eligibility criteria used were the same as in the previous section.

These patients constituted a subset of the initial 299,994 sample whose records contained:

- Medical codes conveying the presence of diabetes, or product codes for anti-diabetic therapies.
- No medical codes for PCOS.

Sensitivity Analyses

Further Medical Code Classification

It was theorised that further medical codes obtained from searches of the Medical Browser Dictionary could be associated with the presence of diabetes, or mapped to the aetiological classification system. Terminology traditionally associated with any diabetes, T1DM or T2DM was used to further classify medical codes in the sensitivity analyses. The terminology used was broadly classifiable in terms of: diabetes education programmes, age of onset, insulin-dependence, and diabetes-related complications. Table 3.6 depicts the terminology identified and the form of diabetes associated with it.

Table 3.6 - Terminology traditionally associated with the presence or type of diabetes. Key: IDDM = "Insulin-Dependent Diabetes Mellitus", NIDDM = "Non-Insulin-Dependent Diabetes Mellitus", DAFNE = "Dose Adjustment for Normal Eating", DESMOND = "Diabetes Education and Self-Management for Ongoing and Newly Diagnosed", DKA = "Diabetic Ketoacidosis", and HONK = "Hyperosmolar Nonketotic Syndrome".

Terminology	Type of Diabetes Associated
Juvenile-/Infant-/Child-Onset DM	T1DM
IDDM	T1DM
NIDDM	T2DM
Any Diabetes Education Programme	Any Diabetes
DAFNE	T1DM
DESMOND	T2DM
DKA	T1DM
HONK	T2DM

Code-List Generation

Code-lists pertaining to each terminology association under investigation were drawn up as subsets of those used to identify diabetes in the main study analysis.

Medical codes were eligible for inclusion if their associated Read terms explicitly stated any of the terminology present in Table 3.6 (or their synonyms). Medical codes were not eligible for inclusion if it had been previously possible to map them to a type of diabetes in this study (i.e. if they explicitly stated a type of diabetes). For example, the Read term "Diabetes mellitus with ketoacidosis" would be eligible for inclusion in the sensitivity analyses, as it does not directly refer to a type of diabetes, whereas the Read term "type 1 diabetes mellitus with ketoacidosis" would not be eligible for inclusion as it directly refers to a type of diabetes. A complete list of the code-lists used may be found in Appendix C.6.

For the purposes of comparison, separate type-explicit code-lists were also created as subsets of the medical codes used in the primary analysis. These consisted of medical codes that unequivocally stated either T1DM or T2DM. Medical codes were eligible for inclusion if their associated Read terms explicitly stated the terms "type 1", "type I", "type 2" or "type II" in reference to diabetes (e.g. the Read term "Type II diabetes mellitus with nephropathy" would be eligible for inclusion, whereas the Read term "non-insulin dependent diabetes mellitus with nephropathy" would not). A complete list of the code-lists used may be found in Appendix C.7.

Can Education Programmes be used to Identify Prevalent Diabetes?

Where patients had medical codes for diabetes education programmes, the dates these programmes were prescribed were cross-referenced against the date of their earliest diabetes record. This was defined as the earliest date from:

- The first anti-diabetic prescription.
- The first medical code indicating the presence of diabetes that was not also being evaluated in sensitivity analyses.

Analyses were stratified by education programme to evaluate whether each were prescribed to patients with or without diabetes. The code-lists used for the identification of diabetes education programmes may be found in Appendix C.8.

Which Terminology May be used to Classify Diabetes?

Initial assessment of the terminology traditionally associated with diabetes or a type of diabetes was performed using medical codes previously mapped to the aetiological classification system. Patients were identified whose records included one or more medical codes containing terminology traditionally associated with T1DM or T2DM. These codes were then removed from patient records, and patients who had previously had them were assigned to a type of diabetes based on the last type-classifiable medical code remaining in their records. Patients were not classified where no type-classifiable medical codes remained in their records.

Secondary assessment of the terminology traditionally associated with diabetes or type of diabetes was performed using the prescription of anti-diabetic therapies. Medical codes containing terminology traditionally associated with T1DM or T2DM were identified in patient records. These were then cross-referenced with the associated patients' prescription records to determine the anti-diabetic drug classes prescribed six months either side of the code's occurrence. For the purpose of comparison, this process was repeated for the medical codes explicitly stating a type of diabetes present in the type-explicit code-list (Appendix C.7).

How does the Classification of Further Medical Codes Affect the Classification of Type 2 Diabetes?

The mappings of the code lists in Appendix C.3 were amended to reflect the findings of the aforementioned sensitivity analyses. All other methods outlined in section 3.3.1 remained unchanged, and were implemented on the same original study population. The effects of these diabetes code-list re-mappings were then evaluated on the populations classified with each form of diabetes.

How Accurate Was the Classification of All Unclassified Patients as Type 2 Diabetes?

Patients were identified whose diabetes was unclassifiable via the use of medical codes or product codes and age. The records of these patients were searched for prescriptions of anti-diabetic therapies occurring at any point following their first diabetes-related record, i.e. either the first medical code linked to the presence of diabetes (see Appendix C.3) or product code for an anti-diabetic therapy (see Appendix C.4). Summary analysis was subsequently performed, in which the presence and class of anti-diabetic therapies prescribed, and the time from the first diabetes-related record to the first prescription of insulin, were evaluated.

Study Population

Three data cuts were used in the procedures outlined in the preceding paragraphs:

1. A data set of 33,051 patients was used to assess whether terminology traditionally associated with diabetes or a type of diabetes could be used to classify further medical codes. This study population was used previously in this chapter for the classification of diabetes.
2. A data set of 299,994 patients was used to assess the effect of classifying further medical codes on the classification of T2DM. This study population was used previously in this chapter for the identification of diabetes.
3. A data set of 8,650 patients was used to assess the validity of the classification of patients' diabetes as T2DM, based on the high background prevalence of the disease. This study population was a subset of the one used in the re-run of the identification and classification strategies; constituting patients whose diabetes was unclassifiable through the use of either medical codes or product codes and age.

3.3.2. Results

Code-List Derivation

Medical Codes

The database interfaced by the Medical Browser Dictionary contained 109,240 unique medical codes. Keyword searches for Read terms related to diabetes in the medical browser returned 744 unique medical codes, 657 of which were determined to convey the presence of one or more forms of diabetes. Of these codes, 210 (32.0%) could be mapped to the aetiological classification system. The remaining 447 (68.0%) codes were not mapped to any individual type of diabetes in the aetiological classification system and were classified as "Undefined Diabetes". The number of codes assigned to each type of diabetes is listed below in Table 3.7. The codes used to define each of these categories may be found in Appendix C.3.

Table 3.7 - Results from the mapping of medical codes to the aetiological classification system.

Type of Diabetes	Codes (n)
T1DM	75
T2DM	106
Gestational Diabetes	13
Drug-Induced Diabetes	4
Bronzed Diabetes	3
Monogenic Diabetes	4
Malnutrition-Related Diabetes	5
Undefined Diabetes	447

Patients whose diabetes was identified via medical codes had a median of 2 medical codes that could be mapped to any specific type of diabetes. This comprised 9.1% of the median total number of all diabetes codes in their records.

Product Codes

The database interfaced by the Product Browser Dictionary contained 62,297 unique product codes. Keyword searches of the drug substance name, BNF header and product name of the Product Browser Dictionary returned 973 unique product codes, 898 of which were anti-diabetic pharmacotherapies. Of these codes, 595 were insulin-based interventions, 76 were metformin monotherapies, and 227 were other anti-diabetic therapies. The number of codes assigned to each

class of anti-diabetic therapy is listed in Table 3.8. The codes used to define each of these categories may be found in Appendix C.4.

Table 3.8 - Results from the mapping of product codes to classes of anti-diabetic therapy. *Combination therapies were product codes combining one or more non-insulin therapies into a single product code.

Anti-Diabetic Class	Codes (n)
Insulin	595
Biguanides	76
Thiazolidinediones	17
Sulphonylureas	104
Meglitinides	18
Alpha-Glucosidase Inhibitors	4
GLP-1 Agonists	14
DPP-4 Inhibitors	19
SGLT2 Inhibitors	12
Guar Gum	5
Combination*	34

Identifying Patients with Diabetes

The overall process of identifying diabetes in patient records is outlined in Figure 3.3. Searching the Clinical File for diabetes medical codes returned 920,642 records belonging to 32,638 patients. Searching the Therapy File for anti-diabetic product codes returned 2,326,105 prescriptions issued to 23,288 patients.

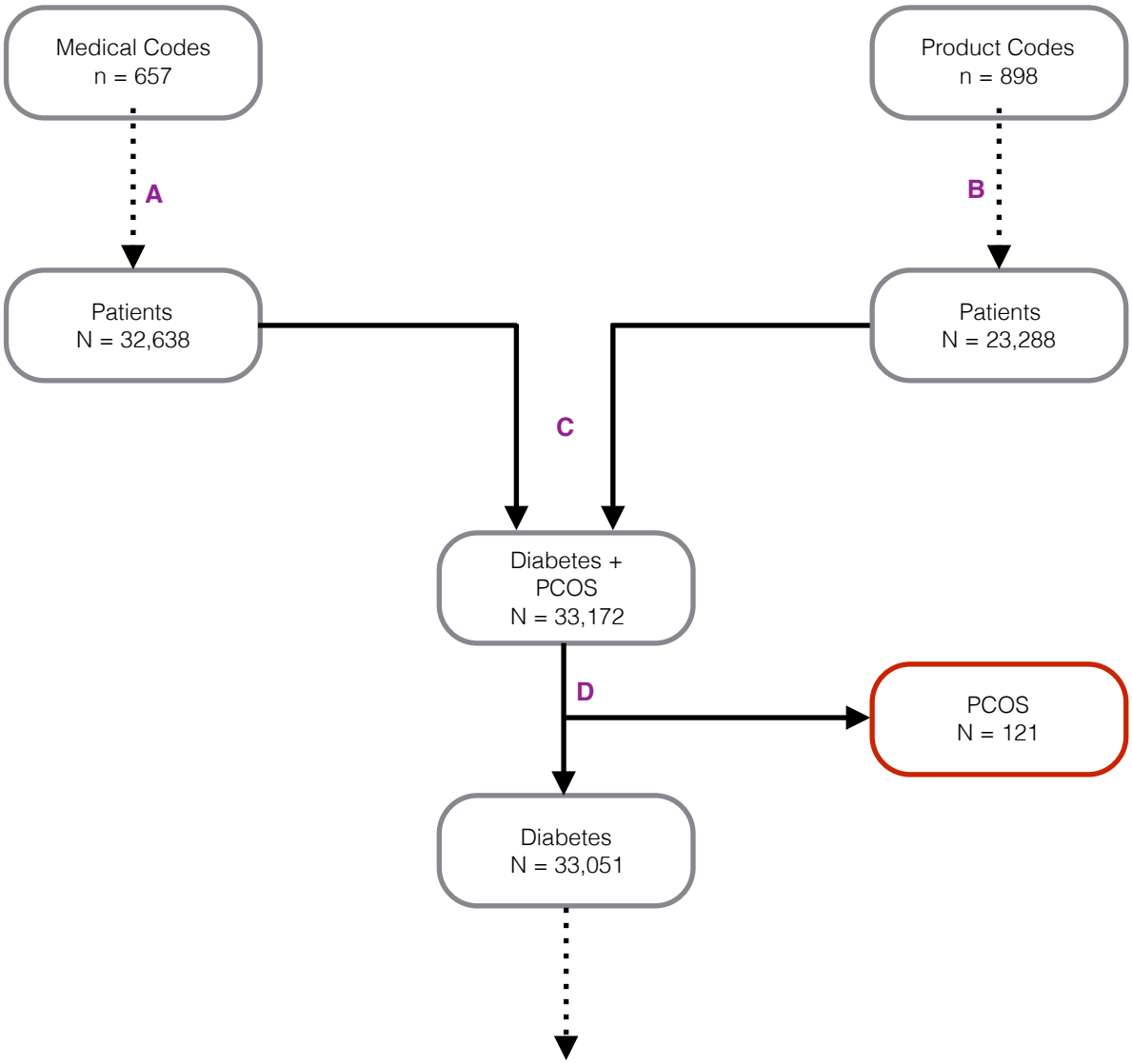


Figure 3.3 - Stages in identifying a population with diabetes, in addition to number of patients present at each stage. Key: A = Identification of patients with diabetes via medical codes; B = Identification of patients with diabetes via product codes; C = Pooling of patients identified via medical codes and product codes; D = Classification of patients with PCOS via medical codes.

Combining the search results from the medical and product code strategies yielded a total of 33,172 patients: 22,754 of these were identified by both strategies, while 534 were identified only by product codes, and 9,884 only by medical codes.

Searching the Clinical File records for medical codes pertaining to PCOS identified 121 patients as having the condition. These were removed from the dataset, yielding a final population of 33,051 patients with diabetes (Table 3.9).

Table 3.9 - The number of patients found to have PCOS and diabetes from the initial starting population.

Population	Patients (N)
PCOS	121
Undefined Diabetes	33,051

Classifying Patients with Type 2 Diabetes

Medical codes for other specific types of diabetes (drug-induced diabetes, bronzed diabetes, monogenic diabetes, malnutrition diabetes) and gestational diabetes were found in the records of 243 patients with diabetes. These patients were classified as having other specific types of the disease and were excluded from further classification.

Of the 32,808 patients determined to have either T1DM or T2DM, 22,894 had medical codes that could be used to classify their type of diabetes. Within these patients, 889 were classified as having T1DM and 22,005 were classified as having T2DM.

In the 9,914 patients without any diabetes-classifiable medical codes, 108 had product codes for insulin prescriptions before the age of 35; these patients were classified as having T1DM.

The remaining 9,806 patients had no medical codes that could be mapped to any specific type of diabetes, and were not prescribed insulin before the age of 35; these patients were classified as having T2DM. The final numbers of each type of diabetes classified are shown in Table 3.10, while the number of patients classified by each stage of the diabetes classification strategy is presented in Figure 3.4.

Table 3.10 - The number of patients with each type of diabetes identified from the initial starting population with diabetes.

Type of Diabetes	Patients (N)
T1DM	997
T2DM	31,811
Gestational & Other Diabetes	243

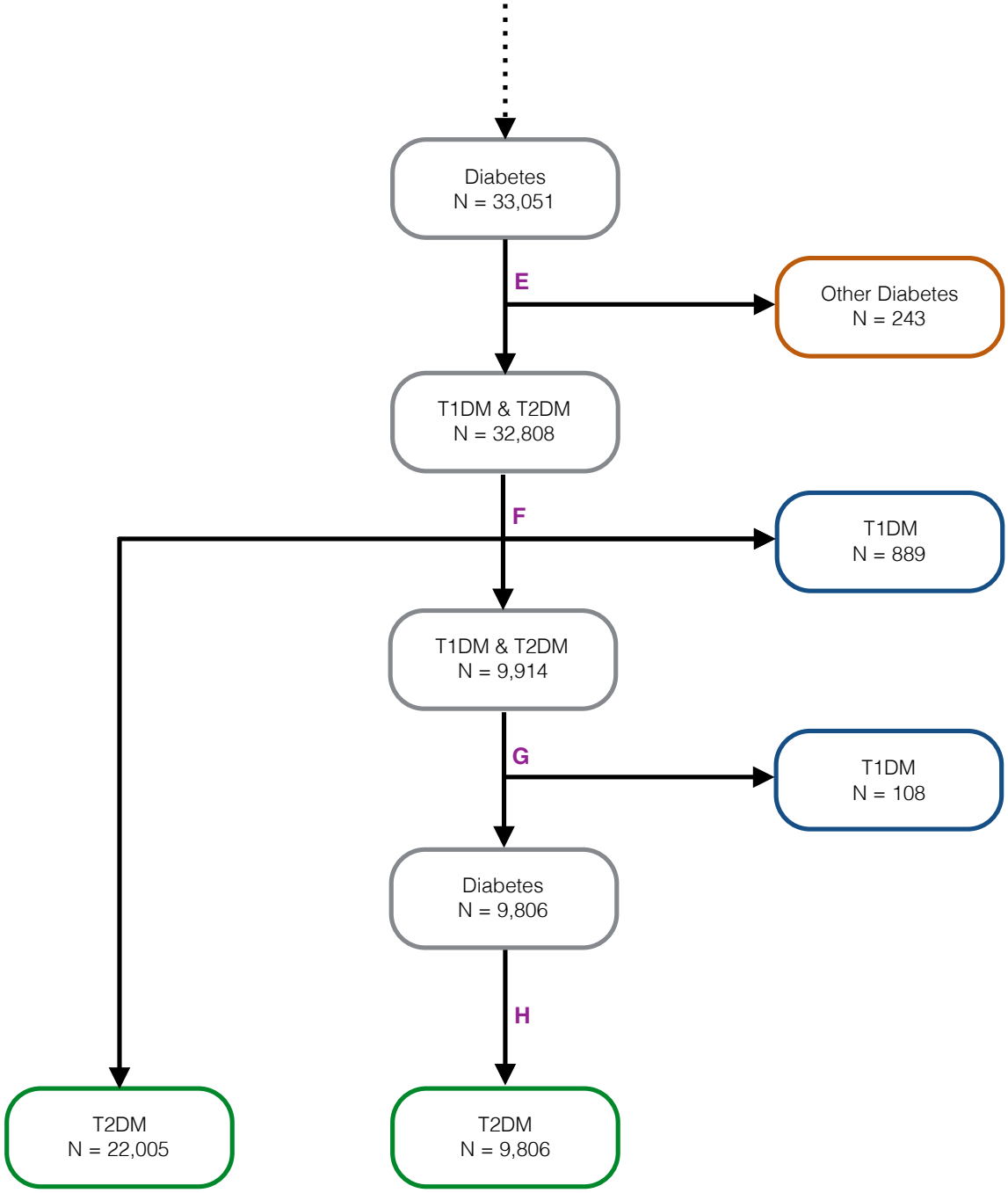


Figure 3.4 - Stages in classifying T2DM in a population with diabetes. The numbers of patients present and excluded at each stage are also depicted. "Other Diabetes" denotes other specific types of diabetes and gestational diabetes. Key: E = Classification of patients with others specific types of diabetes and gestational diabetes via medical codes; F = Classification of patients with T1DM and T2DM via medical codes; G = Classification of patients with T1DM via product codes and age; H = Classification of T2DM on the basis of the background prevalence.

A cross-reference of 32,808 patients whose medical records contained medical codes or product codes to indicate the presence of diabetes is presented in Table 3.11. The table contains only patients deemed to have either T1DM or T2DM, as other forms of diabetes and PCOS had previously been removed from the population.

Table 3.11 - A cross-tabulation of how patients would be classified using only medical codes, or only product codes in conjunction with age. Green shading indicates agreement between the classification mediums, while orange shading indicates disagreement between the classification mediums.

		Product Code & Age Classification			
		T1DM	**Undefined Diabetes**	**.**	**Total**
Medical Code Classification	**T1DM**	78	802	9	**889**
	T2DM	15	17,341	4,649	**22,005**
	Undefined Diabetes	94	4,244	4,972	**9,310**
	.	14	590	N/A	**604**
	Total	**201**	**22,977**	**9,630**	**32,808**

In its current form, the overall strategy used medical codes in preference to product codes for the classification of T1DM and T2DM. However, it would have been possible to classify 201 (0.6%) patients with T1DM through the use of product code alone, while 22,894 (69.8%) patients could be classified through the use of medical codes alone.

An overlap in the ability to classify diabetes through the use of either medical codes or product codes and age was present in 93 (0.3%) patients (highlighted in Table 3.11). Agreement was present between the two means of classification in 78 (83.8%) patients, while disagreement was present in 15 (16.1%) patients.

Sensitivity Analyses

Can Education Programmes be used to Identify Prevalent Diabetes?

Twenty-four medical codes were identified that could be mapped to diabetes education programmes. These identified a total of 1,382 (4.2%) patients. The number of patients prescribed each medical programme is depicted in Table 3.12. Prior diabetes-related codes were found in the records of more than 97.8% of patients whose records contained medical codes for diabetes education programmes (Table 3.12).

Table 3.12 - The number of patients whose records contained medical codes for each diabetes education programme, and whether or not they had prior medical of product codes indicative of diabetes. * Un-named diabetes education programmes.

Education Programme	Browser Medical Codes	Total Patients Identified	Patients with Previous Diabetes Codes		Patients with No Previous Diabetes Codes	
	n	N	N	%	N	%
DAFNE	6	9	9	100.0	0	0.0
DESMOND	4	334	327	97.9	7	2.1
X-PERT	5	277	275	99.3	2	0.7
Other*	9	812	801	98.6	11	1.4

Which Terminology May be used to Classify Diabetes?

A total of 90 medical codes were identified that contained terminology traditionally associated with T1DM or T2DM. Of these 11 were related to the age-of-onset classification system, 59 were related to the insulin-dependence classification system, 10 were related to diabetes education programmes, and 13 were related to complications of diabetes. Table 3.13 shows the total number of patients whose medical records contained these codes, in addition to the type of diabetes they would be assigned based on their last-classifiable medical code. Of the 33,051 patients with diabetes present

in the initial dataset, codes for terminology traditionally associated with T1DM or T2DM identified 6,623 (20.0%) patients.

Table 3.13 - The number of patients with medical codes containing terminology traditionally associated with T1DM or T2DM. These are presented alongside the number of last classifiable medical codes for each type of diabetes. * refers to both other specific types of DM and gestational DM.

Code Mapping	Browser Medical Codes	Associated Type of Diabetes	Patients Identified in CPRD	T1DM		T2DM		Other Diabetes*		Undefined Diabetes	
	n		N	N	%	N	%	N	%	N	%
IDDM	36	T1DM	1,124	235	20.9	444	39.5	5	0.4	440	39.1
DKA	9		123	33	26.8	40	32.5	0	0.0	50	40.7
DAFNE	6		9	1	11.1	3	33.3	0	0.0	5	55.6
Juvenile	11		8	3	37.5	2	25.0	0	0.0	3	37.5
NIDDM	23	T2DM	5,228	54	1.0	4,093	78.3	12	0.2	1,069	20.4
DESMOND	4		334	2	0.6	323	96.7	0	0.0	9	2.7
HONK	4		5	0	0.0	0	0.0	0	0.0	5	100.0

Less than 26.8% of patients with medical codes for either IDDM or DKA had last-classifiable medical codes for T1DM, while greater than 78.3% of patients with medical codes for NIDDM or DESMOND had last classifiable medical codes for T2DM. Fewer than 10 patients had medical codes for DAFNE, juvenile diabetes or HONK.

The number of medical codes found in patient medical records that contained terminology traditionally associated with either T1DM or T2DM is presented in Table 3.14. This is referenced against the number (and percentage) of occasions on which there were prescriptions for anti-diabetic therapies six months before or after the date of the medical code being issued. Of the 920,642 medical codes relating to diabetes in the Clinical File, 12,178 were linked to terminology traditionally associated with T1DM or T2DM. A further 37,554 medical codes were identified using the type-explicit code-list, i.e. codes that explicitly stated a type of diabetes.

Table 3.14 - The number of codes containing terminology traditionally associated with T1DM or T2DM, and the number occasions these were associated with prescriptions of anti-diabetic therapies, six months prior or post the issuing of the code. Type-explicit codes are shown in bold.

Code Mapping	Browser Medical Codes	Associated Type of Diabetes	Medical Codes Identified in CPRD	Insulin Therapies Prescribed		Non-Insulin Therapies Prescribed		Insulin & Non-Insulin Therapies Prescribed	
	n		n	n	%	n	%	n	%
T1DM	***73***		***1,442***	***1,330***	***92.2***	***339***	***23.5***	***272***	***18.9***
IDDM	36		2,021	1,907	94.4	697	34.5	648	32.1
DKA	9	T1DM	166	145	87.3	35	21.1	20	12.0
DAFNE	6		9	9	100.0	3	33.3	3	33.3
Juvenile	11		8	6	75.0	1	12.5	0	0.0
T2DM	***89***		***35,308***	***1,330***	***12.5***	***21,938***	***62.1***	***3,191***	***9.0***
NIDDM	23		9,495	568	6.0	6,439	67.8	399	4.2
DESMOND	4	T2DM	409	16	3.9	221	54.0	14	3.4
HONK	4		5	2	40.0	3	60.0	1	20.0

For the T1DM-explicit medical codes, on 92.2% of occasions there was a prescription for an insulin therapy six months before or after it, while 23.5% had a prescription for a non-insulin therapy in this time window. For the T2DM-explicit medical codes, a prescription for an insulin therapy was present six months before or after the date the code was issued on 12.5% of occasions, while this figure was 62.1% for non-insulin therapies.

For medical codes mapped to IDDM, DKA and DAFNE, prescriptions for insulin-based therapies were present within six months (before or after) on greater than 87.3% of occasions. This figure was 92.2% for type-explicit medical codes explicitly stating synonyms of T1DM. Prescriptions for non-insulin-based therapies were present within six months of the medical codes for IDDM, DKA and DAFNE on 21.1% - 34.5% of occasions. This figure was 23.5% for T1DM-explicit medical codes.

For medical codes mapped to NIDDM and DESMOND, prescriptions for insulin-based therapies were present within six months on less than 6.0% of occasions. This figure was 12.5% for aetiological comparator medical codes explicitly stating synonyms of T2DM. Prescriptions for non-insulin-based therapies were present within six months of medical codes for NIDDM and DESMOND on 54.0% - 67.8% of occasions. This figure was 62.1% for T2DM-explicit medical codes.

Fewer than 10 medical codes were present in the database for DAFNE, Juvenile diabetes and HONK.

How does the Classification of Further Medical Codes Affect the Classification of Type 2 Diabetes?

Following the sensitivity analysis, 27 NIDDM and DESMOND medical codes, originally mapped to "Undefined Diabetes", were remapped to T2DM. A breakdown of the final medical code mappings is shown in Table 3.15. Of these codes, 236 (35.9%) were now mappable to the aetiological classification system with a reasonable degree of certainty. The remaining 421 (64.1%) codes were left as indicating the presence of undefined diabetes.

Table 3.15 - Results from the mapping of medical codes to the aetiological classification system, (revised to reflect the findings of the sensitivity analyses.

Type of Diabetes	Codes (n)
T1DM	75
T2DM	132
Gestational	13
Drug-Induced	4
Bronzed	3
Monogenic	4
Malnutrition-Related	5
Undefined	421

The additional medical codes for T2DM identified by the sensitivity analyses allowed for the classification of diabetes via medical codes in a further 1,147 patients. Figure 3.5 depicts the final populations identified by each tier of the diabetes identification and classification system. The number of patients classified as having each type of diabetes may be found in Table 3.16.

Table 3.16 - The number of patients with each type of diabetes identified from the initial starting population with diabetes.

Type of Diabetes	Patients (N)
T1DM	996
T2DM	31,812
Gestational & Other Diabetes	243

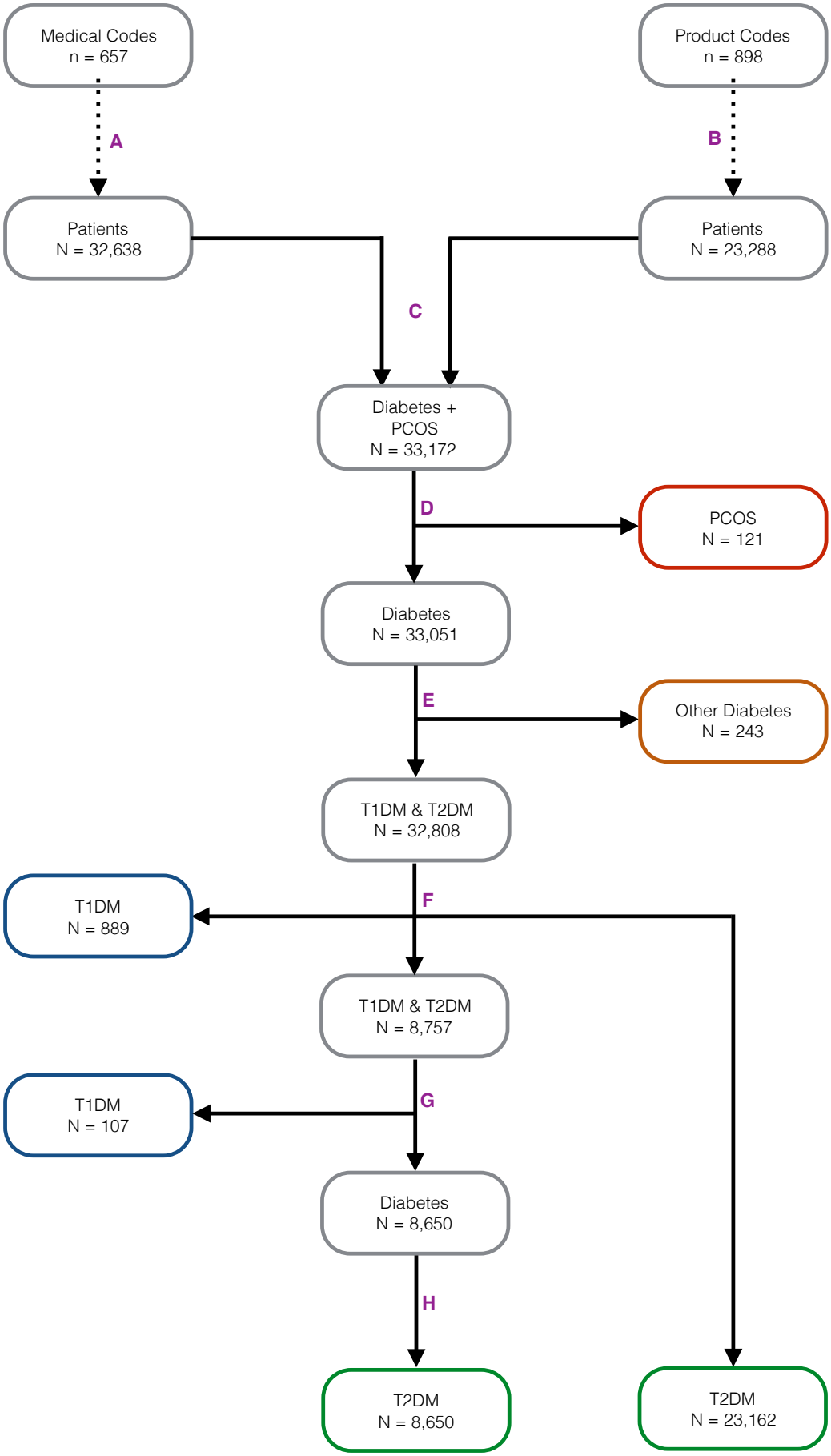

Medical Codes
n = 657
Product Codes
n = 898
A
B
Patients
N = 32,638
Patients
N = 23,288
C
Diabetes +
PCOS
N = 33,172
D
PCOS
N = 121
Diabetes
N = 33,051
E
Other Diabetes
N = 243
T1DM & T2DM
N = 32,808
F
T1DM
N = 889
T1DM & T2DM
N = 8,757
G
T1DM
N = 107
Diabetes
N = 8,650
H
T2DM
N = 8,650
T2DM
N = 23,162

Figure 3.5 - The various stages in identifying a population with T2DM. The numbers of patients present and excluded at each stage are also depicted. "Other Diabetes" denotes other specific types of diabetes and gestational diabetes. Key: A = Identification of patients with diabetes via medical codes; B = Identification of patients with diabetes via product codes; C = Pooling of patients identified via medical codes and product codes; D = Classification of patients with PCOS via medical codes; E = Classification of patients with others specific types of diabetes and gestational diabetes via medical codes; F = Classification of patients with T1DM and T2DM via medical codes; G = Classification of patients with T1DM via product codes and age; H = Classification of T2DM on the basis of the background prevalence.

How Accurate Was the Classification of All Unclassified Patients as Type 2 Diabetes?

Following the reclassification of medical codes for NIDDM and DESMOND as indicating the presence of T2DM, 8,650 patients remained unclassifiable via the use of either medical codes or product codes and age. The number of patients requiring anti-diabetics, and the class of anti-diabetic prescribed is presented in Table 3.17.

Table 3.17 - The anti-diabetic therapies prescribed to patients whose diabetes was not classifiable through the use of either medical codes, or product codes and age.

Therapies Prescribed	Patients	
	N	%
Prescribed No Anti-Diabetic Therapies	4,700	54.3
Prescribed Only Non-Insulin-Based Therapies	273	3.2
Prescribed Insulin Therapies within 6 Months	2,745	31.4
Prescribed Insulin Therapies within 24 Months	3,038	35.1
Prescribed Insulin Therapies at Any Point	3,677	42.5

3.4. Discussion

3.4.1. Review of the Literature

Key Results of the Review of the Literature

The results of the review indicate that there is no clear consensus on the best strategy to identify diabetes, T1DM or T2DM in CPRD. For each study, researchers devised strategies dependent upon the levels of sensitivity and specificity they required for their study populations, i.e. whether they

required a high degree of certainty that all patients had diabetes or a particular type of diabetes, or whether they wanted the largest possible population sample at the expense of potential misclassification.

When identifying diabetes, the most frequently used strategies combined both medical and product codes (nine out of 15 studies). A combination of medical codes, product codes and age was the most frequently used strategy by studies for the classification of both T1DM (six out of 12 studies) and T2DM (24 out of 75 studies). The combined use of medical and product codes featured in the strategies of 10 out of 15 studies classifying T1DM and 52 out of 75 studies classifying T2DM. Medical codes and product codes also constituted the most commonly used CPRD components (independent of strategy) for studies identifying diabetes, and classifying T1DM, and T2DM. For both T1DM and T2DM, age was the third most frequently used CPRD component independent of strategy. These results suggest that studies tend to favour the use of medical and product codes to identify diabetes, while they favour medical codes, product codes and age to classify T1DM and T2DM.

Few studies used the presence of test criteria (the presence of tests or test results) in the identification of diabetes, or classification of T1DM or T2DM. None of the 12 studies classifying T1DM used test criteria, while one out of 15 studies and eight out of 75 studies used these criteria to identify diabetes and classify T2DM, respectively. In the studies that used test criteria to classify T2DM, the tests were used to identify DM, prior to the implementation of subsequent strategies to distinguish T1DM from T2DM. A lack in the use of test criteria could be attributed to the potential for misclassification of patients by these tests. GPs may use tests (such as fasting plasma glucose and HbA1c) in order to diagnose and monitor patients' diabetes. Therefore, the use of 'test presence' in medical records may be liable to flag patients as having diabetes when the test may have been performed for diagnostic purposes and yielded a negative result. Some of these tests may also be used in patients without diabetes to evaluate hyperglycaemia or hypoglycaemia brought about by some medical conditions or drug therapies. Thus, positive test results in medical records may not confer the presence of diabetes, as in certain circumstances the results may not be representative of the patients' true

underlying glycaemic control. Despite the shortcomings of test criteria, many studies used these criteria as a means of validating the diabetes identified by other components of their strategy. However, as this did not constitute a means of "identifying" patients with diabetes and is was considered outside the scope of this review.

No studies used product codes as the only CPRD component in their diabetes identification strategies. Conversely, product codes constituted the only CPRD component in the strategies of one of the 12 studies classifying T1DM and five of the 75 studies classifying T2DM. Within these studies, prescriptions for insulin were used to classify T1DM, while an absence of insulin prescriptions, or prescriptions for non-insulin therapies, were used to classify T2DM. The use of prescriptions for insulin therapies to classify T1DM may be flawed as it could misclassify patients with insulin-dependent T2DM. Alternatively, the use of prescriptions for non-insulin therapies to classify T2DM may be flawed as it could misclassify patients prescribed metformin for the treatment of PCOS, and patients prescribed metformin for weight control in T1DM. The reliance on non-insulin therapies to classify patients with T2DM would also miss patients whose T2DM was controlled by diet.

When age was used as a diagnostic criterion to classify T1DM or T2DM, it was combined with product codes by all but one study, which combined it with medical codes. Within these studies, the age and product code strategy either took the form of identifying patients above a certain age and on non-insulin therapies, or below a certain age and on insulin therapies. The age boundaries used by studies to classify T1DM and T2DM by these strategies were largely similar, with 35 being the most frequently used age in the classification of both types of diabetes. This may appear odd in the context of the possible age and product code combinations stated above, as one would presume that studies would use a higher age in conjunction with non-insulin therapies to classify T2DM, and a lower age in conjunction with insulin therapies to classify T1DM. However, in the classification of both forms of diabetes, the strategy using a lower age in conjunction with prescriptions for insulin therapies was preferred. Therefore, studies that classified T1DM used insulin prescription to patients under the age of 35 as a diagnostic criterion. For studies that classified T2DM, the same process was performed, but for the purpose of removing these patients from the dataset; the assumption being

that all remaining patients had T2DM. This method has benefits and drawbacks in the classification of both T1DM and T2DM. On the one hand, this may allow for the classifying of patients with T1DM who do not have medical codes to indicate the presence of the disease. On the other hand, the use of this method makes the assumption that T1DM can only onset before a specified cut-off age (often 35). However, T1DM is more likely to onset after the age of 35 than before it[26]. Conversely, it is possible to develop T2DM before the age of 35. It is merely the case that any diabetes that onsets before the age of 35 is more likely to be T1DM, and any diabetes that onsets after the age of 35 is more likely to be T2DM. The combined use of age and prescription data was only used to positively classify T2DM (i.e. the use of being over a certain age and on non-insulin therapies) by four of the 75 studies that classified T2DM. In these studies, the age boundary was 40 years. "Use of insulin therapies combined with being below a certain age" was preferred by authors as a means of classifying both T1DM and T2DM.

Strengths and Limitations of the Review of the Literature

This is the first study to compare the methodologies used to define diabetes, T1DM and T2DM in CPRD. The use of CPRD's bibliography as a database to find eligible studies is a strength of the review. Performing searches in this database allowed for the use of a more simplistic search strategy based solely around keywords for diabetes and anti-diabetic medications. Had this review been conducted in a large database of medical studies (like MEDLINE) keywords would also have had to have been specified for the study population source (CPRD). This would have been problematic as database searches normally only scrutinise the titles and abstracts of publications for the specified keywords. As abstract word limits are tight for many journals, it is unlikely that the study population source used would have made it into the abstract. Thus, the use of CPRD's bibliography minimised the amount of screening required, simplified the search strategy and minimised the possibility of omitting relevant studies.

A current limitation of this review is that it has yet to be completely co-reviewed. It is common practice for a second reviewer to co-screen articles retrieved from database searches and co-extract data from eligible articles[134]. As of this moment, data have been co-extracted for 23 of the 82 retrieved

publications, and there are no plans to co-screen the titles returned by the CPRD bibliography search. Until the co-extraction phase is completed several of the results presented in this study may be liable to change.

The limited information presented by eligible studies on their diabetes identification and classification strategies limited the ability of this review to suggest how components of CPRD should be used. Furthermore, there was little information regarding the reasons studies chose to implement their strategies in the way they did. For instance, some studies may have desired a high specificity in their selection criteria, whereas others may have wished to optimise sensitivity to maximise the available analysis population. In the absence of this information, it was only possible to consistently extract data on the CPRD components used for each strategy.

3.4.2. Classification of Type 2 Diabetes in Clinical Practice Research Datalink

Key Results of the Classification of Type 2 Diabetes in Clinical Practice Research Datalink

Diabetes identification strategies using only medical codes or product codes would have been able to identify 98.7% or 70.4% of the total number of patients identified by the combined strategy of this study. The inclusion of product codes resulted in the identification of few additional patients with diabetes, at the expense of potentially incorporating patients with PCOS instead of diabetes. However, the relative ease of identifying patients by product codes, and removing patients with PCOS via medical codes, justified the inclusion of product codes in the strategy used. Furthermore, the removal of patients with PCOS had little impact on the final diabetes population identified, effecting only 0.4% of patients with codes associated with diabetes. Further analyses demonstrated that diabetes education programmes could effectively be used to identify diabetes, as only 0.0% - 2.1% of the patients prescribed these programmes had no pre-existing diabetes-related codes.

Few patients with diabetes had medical codes for other specific types of diabetes or gestational diabetes, and the removal of patients with these types of diabetes had virtually no impact on the classification of T2DM. However, little extra effort was required to classify medical codes relating to these forms of diabetes from those returned by searches of the Medical Browser Dictionary. Furthermore, many of these conditions may be liable to misclassification as T2DM while demonstrating clear differences in aetiology of onset. Thus, these patients should be classified and removed.

Patients had a median of two classifiable medical codes, despite there being a median of 22 medical codes per patient. This allowed for the classification of diabetes in 70% of patients. Hence, despite the relative effectiveness of medical codes in identifying diabetes, it is important to have additional mechanisms for the classification of diabetes. In transitioning from previous diabetes classifications systems to the aetiological system, NIDDM was presumed to be synonymous with T2DM, and IDDM was presumed to be synonymous with T1DM. Medical codes for NIDDM demonstrated a high positive predictive value (PPV) for the presence of T2DM, while medical codes for IDDM demonstrated a low PPV for the presence of T1DM. The education programmes of DESMOND and DAFNE state that they are only intended for prescription to patients with T2DM and T1DM, respectively. Further analysis found that medical codes for DESMOND demonstrated a high PPV for the presence of T2DM, while medical codes for DAFNE demonstrated a low PPV for the presence of T1DM. DKA is a diabetes complication that was presumed to be strongly associated with T1DM. However, medical codes for DKA demonstrated a low PPV for T1DM. The results of these analyses indicate that only medical codes stating synonyms of T1DM are appropriate to classify T1DM, while medical codes stating NIDDM, DESMOND or synonyms of T2DM may be used to classify T2DM. However, caution should be taken when interpreting these results, as PPVs are influenced by the prevalence of the outcomes they are used to evaluate.

The use of product codes in combination with age had minimal effect on the overall diabetes classification, resulting in the classification of 0.3% patients with diabetes. However, without this step,

these patients would have been misclassified as having T2DM in the later stages of the overall classification strategy.

Of the 8,650 patients that could not be classified through the use of medical codes or product codes in combination with age 4,700 (54.3%) had no prescriptions for anti-diabetic therapies. Within the remaining 3,950 patients, 3,677 had prescriptions for insulin-based therapies. For 2,745 of these patients, insulin was prescribed within six months of their first medical code or product code associated with diabetes. Hence, of the 33,051 patients identified as having diabetes there is uncertainty regarding the classification of only 2,745 (8.3%).

Strengths and Limitations of the Classification of Type 2 Diabetes in Clinical Practice Research Datalink

Strengths and Limitations of the Dataset

This was a scoping exercise conducted using a pre-existing data cut, rather than a data cut made specifically for this study. Therefore, the data cut incorporated design decisions that were sub-optimal for use in this study. For every patient in the dataset, follow-up data were only available for the period of time they were registered at their most recent CPRD-linked GP practice, and only from the age of 18. This was problematic, as data potentially usable in the identification or classification of diabetes was missing. A strength of the dataset was the date chosen to be the index date, 1st January 2006. This roughly coincides with the advent of QOF in April 2004[135]. QOF financially incentivised GPs to use medical codes for diabetes from the aetiological classification system, to differentiate between T1DM and T2DM, and to see their patients with diabetes at least once per year. Thus, much of the data present in the dataset will be from after this date, and liable to be of a high quality.

The use of medical codes placed a heavy reliance on GPs being able to establish correct diagnoses in patients, being adherent to good practice, and being accurate in their record keeping. Where this was not the case, bias may be present in medical codes as a result of misdiagnosis and miscoding.

Strengths and Limitations of Methods

The methods and CPRD components used in this study have been primarily influenced by the findings of the review conducted in section 3.2. This was a major strength of this study as it meant that the methods implemented were based on real-world pragmatic approaches as opposed to theoretical techniques. Thus, all methods would be relatively simple to implement in subsequent CPRD research.

The methods employed in the identification of medical codes and product codes in the Browser Dictionaries were novel and able to detect codes that might be missed by search strategies that used an entirely pre-specified list of search terms. This was corroborated through comparing the lists generated to those already present within the Nuffield Department of Primary Care Health Sciences, which yielded no additional codes.

With no access to individual patients to verify type of diabetes, the classification of patients' diabetes and the assessment of their diabetes classifications were limited to 'internal coding checks', i.e. in assessing whether patients with medical codes for NIDDM had T2DM, verification of their disease classification was sought from the presence of further medical codes and prescription data.

The methods implemented in the diabetes classification strategy devised by this study were restricted to those identified in the review of the literature. Therefore, this chapter does not constitute an exhaustive appraisal of all possible methods to identify diabetes and classify T2DM. Many more strategies are possible using CPRD components both identified and not identified in this chapter. However, describing all such strategies is outside of the scope of this study.

3.4.3 Relationship to the Literature

This is the first study to investigate the methods for identifying diabetes and classifying T2DM in the CPRD database. Studies have previously investigated how diabetes should be identified and classified using data obtained from a small number of GP practices in the "Cutting Out Needless Deaths Using Information Technology" and "Quality Improvement In Chronic Kidney Disease" trials[130,132,136,137]. A collection of studies using this data form part of a Royal College of General

Practitioners report that is often cited in reference to the identification and classification of diabetes[138]. Data from these studies were extracted directly from the computer systems of contributing GP practices, as opposed to from CPRD. Therefore, the inferences made from such data would only be useful in the identification and classification of diabetes in primary care computer systems, as opposed to in CPRD. Evidence for this may be seen in the classification algorithms present in the Royal College of General Practitioners report, which incorporate the date of diabetes diagnosis. This data is not readily available in a CPRD data cut. A further limitation in the use of this report is that its primary objective was to combine evidence to allow for the identification of patients whose diabetes classification was uncertain. There is a subtle but important difference between this and the primary objective of this study, which was to combine evidence to allow for the classification of patients' diabetes. Nevertheless, some of the relevant studies forming the Royal College of General Practitioners report, in addition to other identified studies that detail methods to classify diabetes in primary care data, are detailed in Table 3.18.

Table 3.18 - The possible methods of classifying T1DM and T2DM in primary health care data suggested by one systematic review (ψ), and three studies using the combined trial data of the "Cutting Out Needless Deaths Using Information Technology" and "Quality Improvement In Chronic Kidney Disease" studies (ε).

Paper	Type of Diabetes	Medical Codes	Product Codes	Age	BMI	Test Presence	Test Results	Diabetes Complications	Associated Diseases	Other*
ψBagheri et al., 2009[133]	T1DM and T2DM	Y	Y	Y	Y	Y	Y	Y	Y	Y
εde Lusignan et al., 2010[130]	T1DM and T2DM	Y	Y	-	Y	Y	-	-	-	-
εde Lusignan et al., 2012[132]	T1DM	Y	Y	-	Y	-	-	-	-	-
εde Lusignan et al., 2012[132]	T2DM	Y	Y	-	-	-	Y	-	-	-
εSadek et al., 2012[137]	T1DM and T2DM	Y	Y	Y	-	-	-	-	-	-

The review in this chapter found that a combination of medical codes and product codes was most frequently used strategy to identify diabetes in CPRD data. As such, these components were also used to identify diabetes in the strategy implemented in this chapter. A major difference between the studies identified by the review and the methods implemented in the strategy to identify diabetes came in the identification of PCOS. The strategy performed by this study opted to identify and remove patients with PCOS to minimise the risk of these patients being misclassified as having diabetes. Few studies in the review opted to perform this step. However, reporting of the minutiae of identification strategies was often poor. Therefore, it is possible that this step was performed and not reported.

Few studies suggested the presence of standard monitoring tests for diabetes (like HbA1c) as diagnostic criteria for identifying diabetes in primary care[130,133]. Similarly, few studies suggested the results of standard monitoring and diagnostic tests for diabetes (like HbA1c and fasting plasma glucose) as diagnostic criteria for identifying diabetes in primary care. In addition to standard tests, Bagheri et al.[133] recommend the use of tests for serum C-peptides, islet cell autoantibodies, antibodies to insulin and other biomarkers. However, these are not standard diagnostic tests for diabetes and many are infrequently performed in general practice. Consequently, their results are not likely to be consistently recorded in CPRD, limiting their use as diagnostic criteria. The lack of advocation for the use of test criteria was in line with the findings of the review in this chapter, where the presence or results of diagnostic and monitoring tests seldom formed a part of the identification strategies of studies using CPRD data. However the lack of use of test criteria by the studies in the review may have resulted from the fact that such data are often missing in primary care records[137].

In the classification of diabetes, the existence of gestational diabetes and other specific types of diabetes was often not acknowledged in either the methods identified by the review of CPRD-based studies, or the methods advocated by studies investigating means to classify diabetes in primary care data (Table 3.18). Failing to acknowledge these types of diabetes would lead to their misclassification via the use of product codes and age as diagnostic criteria; a step performed by

many of the CPRD-based studies in the review. Where these forms of diabetes were classified in the review, the classification was performed either directly (via the use of medical codes linked to the type of diabetes) or indirectly (via the identification of patients prescribed non-insulin therapies below a certain age). Within the strategy implemented in this study the direct method was chosen. This was because not all of these types of diabetes would be liable to require non-insulin therapies before a certain age, but all may appear in patient records as medical codes.

The use of age to classify diabetes was not commonly advocated by other studies investigating means of classifying diabetes in primary care data (Table 3.18). However, a combination of medical codes, product codes and age was found to be the most frequently used strategy to classify diabetes in CPRD. Within this, two sub-strategies were identified using age and product codes:

- Patients prescribed insulin therapies before a specified age had T1DM, and everyone else had T2DM; the most frequently used age cut-off being 35.
- Patients prescribed non-insulin therapies after a specified age had T2DM; the most frequently used age cut-off being 40.

The former of these two was the most frequently used strategy in studies classifying both T1DM and T2DM in CPRD. This may be attributable to information previously described in the literature, which states a high PPV in using age and product codes for the direct classification of T1DM, yet a comparatively low PPV in using these combined diagnostic criteria for the direct classification T2DM[133]. Thus, age and product codes were combined to indirectly classify T2DM in the strategy implemented in this study.

BMI was advocated as a diagnostic criterion to classify diabetes by three studies investigating methods to classify diabetes in primary care data (Table 3.18). However, BMI was not present in the classification strategies of any CPRD-based studies identified by the review. The lack of use of BMI as a diagnostic criterion for classifying diabetes was surprising as figures from studies identified in the literature have stated a high PPV in the classification of T1DM when using a combination of age, product codes and BMI[133]. The lack in use of BMI as a diagnostic criterion in CPRD may be

attributable to its inconsistent recording, in which data may be either missing, or have height and weight values recorded at different appointment dates; often with a high degree of temporal separation. As the classification strategies identified in the review did not use BMI, it was not implemented in the classification strategy of this study.

The tendency for certain diseases and complications to be more strongly associated with certain forms of diabetes has been well documented by the American Diabetes Association and the Expert Committee on the Diagnosis and Classification of Diabetes[25,124]. However, only one study investigating methods to classify diabetes in primary care data suggested the use of these criteria (Table 3.18). This was also reflected in the CPRD-based studies identified by the review, where there was only one study that used the presence of ketoacidosis to classify diabetes. One explanation for the lack of use of these criteria to classify diabetes may be their scarcity. Complications and associated diseases like ketoacidosis, PCOS or acanthosis nigricans are rare. Thus, their incorporation into any strategy is likely to have minimal effect, especially when they are used to classify less common forms of diabetes. Another explanation may be a potential lack of communication between clinicians and database researchers. Whereas the use of CPRD components such as medical codes, product codes and age requires little knowledge of diabetes, the use of associated complications or diseases requires more clinical input. Hence, the difficulty in establishing the degree of association between these conditions and type of diabetes may result in their lack of use.

It is possible that the degree of disconnect between the methods for identifying diabetes and classifying diabetes in primary care, and those identified in the review of CPRD-based studies, may be the result of a difference in approach. Some of the strategies identified by studies investigating methods to classify diabetes in primary care data could be considered to be experimental or theoretical, whereas the approaches identified in CPRD-based studies could be considered to be applied or pragmatic. Another explanation for the differences could lie in the range of primary care datasets in existence in the UK. Methods derived from the use of one of these datasets may be less viable for use in another.

3.4.4 Implications

This chapter set out to develop methods to identify and classify diabetes in CPRD. Initial review of the literature identified the strategies previously employed by researchers to identify diabetes and classify T1DM and T2DM in CPRD. The review found that the most pragmatic means to identify diabetes is via the use of medical codes and product codes, while the most pragmatic means of classifying T1DM and T2DM is via a combination of medical codes, product codes and age.

Sensitivity analyses determined that researchers may use medical codes for NIDDM and the DESMOND education programme to establish the presence of T2DM in CPRD. However, medical codes for IDDM and DKA should not be used to identify T1DM in CPRD. Sensitivity analyses also determined that researchers may use medical codes for any diabetes education programme to establish the presence of diabetes in CPRD.

In the next chapters these methods will be used to identify a cohort in which analyses will be conducted to assess the impact of competing outcomes on measures of absolute and relative risk.

4. Competing Risks and Measures of Absolute Risk

4.1. Introduction

4.1.1. Background

The review in Chapter 1 identified an absence of studies evaluating the impact of competing outcomes on estimates of absolute and relative risk in patients with type 2 diabetes mellitus (T2DM) and albuminuria, while the systematic review in Chapter 2 failed to identify any studies, in which the absolute risks of either cardiovascular or cancer mortality had been stratified by albuminuria level.

In Chapter 3, the methods used by authors to identify and classify diabetes mellitus in Clinical Practice Research Datalink (CPRD) were systematically reviewed to derive methods to identify T2DM in CPRD. In this chapter, the methods to identify T2DM will be implemented to define a population with which to compare estimates from standard and competing risk estimators for the absolute risks of cardiovascular and cancer mortality in patients with and without albuminuria.

4.1.2. Aims and Objectives

Primary Objective

To evaluate how competing risks methodology affects absolute risk estimates of cardiovascular and cancer mortality in patients with T2DM, with and without albuminuria.

Specific Aims

To compare estimates of cumulative incidence of cardiovascular and cancer mortality stratified by the presence of albuminuria, from competing risks methodology (cumulative incidence competing risk) vs. standard methodology (Kaplan-Meier).

4.2. Methods

4.2.1. Study Population

Population

CPRD source data was obtained subsequent to a successful application to the Independent Scientific Advisory Committee (ISAC); Protocol 15_011A (Appendix D.1). The dataset comprised of 280,728 patients with diabetes, registered at general practitioner (GP) practices deemed "up-to-standard" before the study start date.

Eligibility was conditional upon: the ability to successfully link patient CPRD and Office for National Statistics (ONS) records, the patient being aged 35 years or over with T2DM on or before the study start date, the patient having unambiguously recorded gender, and the patient being registered at their practice for a minimum of two years prior to the study start date. In this study 'baseline' was taken as reference to the two-year period preceding the study start date, during which time risk factors and patient characteristics were evaluated. The presence of T2DM was established using the methods outlined in Chapter 3.

Patients were excluded if their records contained medical codes for gestational or other specific types of diabetes mellitus, polycystic ovary syndrome, total pancreatectomy, or pancreatic or renal transplant, at any point prior to the study start date.

Patient eligibility was defined using only data uploaded by GP practices after the date they were classified as up-to-standard.

Follow-Up

Follow-up commenced on the study start date of 1st January 2005 and terminated on the study end date of 1st January 2014. Patient records were censored on the earliest event from: the study end date, the date of patient mortality, the date of last data collection from the patient's GP practice, the date the patient transferred out of their current practice, the date the patient underwent total pancreatectomy, or pancreatic or renal transplant.

4.2.2. Stratification Variables

Albuminuria

Patients were classified into those with normoalbuminuria and albuminuria, according to the 2013 Kidney Disease improving Global Outcomes classification structure[28]. Three CPRD data sources were used to establish albuminuria; classifications made using diagnostic medical codes were used in preference to classifications made using numerical test values, and classifications made using numerical test values were used in preference to classifications made using categorical test values.

Diagnostic Medical Codes

Relevant medical codes were identified through searching the Medical Browser Dictionary using the following keyword stems: "*albumin*", "*protein*", "*nephropath*", and "*kimmelstiel*wilson*". Retrieved codes were screened to determine the relevance of each to albuminuria, and subsequently indexed according to whether or not they conveyed the presence of albuminuria. Guidance from GPs (HA, AF, and DM) was sought at both the screening and indexing stages. The final code-list may be found in Appendix D.2.

To identify the presence or absence of albuminuria in patient records, the Clinical File was searched using the derived code-list for albuminuria (Appendix D.2)

Numerical Test Measures

The Test File was searched for potentially relevant entity types for albuminuria tests using the medical codes in Appendix D.2. The returned list of entity types was screened to determine the relevance of each to albuminuria testing, with relevant entity types subsequently categorised according to whether or not they were associated with numerical data. Values linked to entity types for: "Albumin Creatinine Ratio", "Urine Microalbumin", "Urine Biochemistry (Routine)", "Other Laboratory Tests", and "Other Biochemistry Test"; were used to define albuminuria.

For numerical values linked to the "Albumin Creatinine Ratio" entity type, the biomarker was assumed to be albumin, while the test performed was deduced from the associated units. For numerical values linked to other relevant entity types, tests and biomarkers were identified by means of associated medical codes (Appendices D.2 and D.3) and associated units.

Numerical values were only used to define albuminuria where they were in the form of ACRs, Protein-to-creatinine ratios (PCRs), 24-hour albumin excretion rates (AERs), or 24-hour protein-excretion rates (PERs). Where urine albumin/protein and creatinine tests were conducted on the same date the data were combined to form ACRs and PCRs. AERs and PERs were only used where the units were expressed per 24 hours or per day. Albumin and protein measures with units expressed per unit of time less than 24 hours were not scaled to a 24-hour period.

Extracted or calculated numerical test values were then mapped to the dichotomised variable structure described above using the boundaries depicted in Table 4.1.

Table 4.1 - Classification of albuminuria from Kidney Disease Improving Global Outcomes[28] and National Institute for Health and Care Excellence[139] guidelines.

			Categories	
			Normoalbuminuria	**Albuminuria**
Biomarkers & Tests	**Albuminuria**	**AER (mg/day)**	<30	≥30
		ACR (mg/mmol)	<3	≥3
		ACR (mg/g)	<30	≥30
	Proteinuria	**PER (mg/day)**	<150	≥150
		PCR (mg/mmol)	<15	≥15
		PCR (mg/g)	<150	≥150

Categorical Test Measures

The Test File was searched using the medical codes in Appendix D.2 for potentially relevant entity types for albuminuria tests. The returned list of entity types was screened to determine the relevance of each to albuminuria testing, with relevant entity types subsequently categorised according to whether or not they were associated with categorical data. Values linked to entity types for: “Urine Test”, “Urinalysis Protein”, and “Urine Dipstick for Protein”; were used to define albuminuria.

For categorical values linked to relevant entity types, tests and biomarkers were identified using associated medical codes for albuminuria (Appendix D.2). Where encountered, the categorical terms: "Absent", "Nil", "Very Low", "Significantly Low", "Low", "Trace", "Normal", "Negative", and "Below low reference limit"; were assumed to refer to normoalbuminuria; while the categorical terms: “Present”, "High", "Significantly High", "Very High", "Abnormal", “Very Abnormal", "Positive", "+",

"++", "+++", "Outside ref range", "Outside reference range", and "Above high reference limit"; were assumed to refer to albuminuria.

4.2.3. Outcomes

All outcomes were established using a combination of UK ONS death certificates and CPRD. Where patients in CPRD could be successfully matched to ONS data on National Health Service number, gender and date of birth, or where the dates of death in CPRD and ONS differed by ≤30 days, the ONS date of death was used to define the date of patient mortality. Where patients in CPRD could not be successfully matched to ONS data on National Health Service number, gender and date of birth, and where the dates of death in CPRD and ONS differed by >30 days, patients were censored at the earliest date of mortality in CPRD and ONS.

Cardiovascular Mortality

Death from cardiovascular causes was defined as death with International Classification of Diseases version 10 (ICD-10) codes I10-I79 listed as the "underlying cause" on the death certificate.

Cancer Mortality

Death from neoplastic causes was defined as death with ICD-10 codes C00-C97 listed as the "underlying cause" on the death certificate.

Other Mortality

Other mortality was defined as death from causes other than those listed under cardiovascular or cancer mortality above.

4.2.4 Sample Size Calculation

For the methodological comparisons in this Thesis, standard sample size methods are not applicable. Previous studies have suggested that the difference between standard- and competing risks estimates may not be large, but confidence intervals were wide and the standard error (0.11) was

larger than the effect size (0.07)[70]. Within these studies the standard error estimates around the regression coefficients (β's) demonstrated proportionality towards the total number of events as opposed to the event rate[70,140]. Thus, scaling up the expected number of events to represent a population of 150,000 people with type 2 diabetes and 10 years of follow up would reduce the standard errors of the regression coefficients of previous studies by a factor of 10[70]. Conservative hazard ratio estimates from the literature suggest the effect of albuminuria on major cardiovascular events and incident cancer to be 1.84 (1.46 - 2.31)[141] and 1.53 (0.76 - 3.08)[140] respectively. In 150,000 people with type 2 diabetes, the regression coefficient standard errors would be reduced to 0.017 and 0.022. This would yield projected 95% confidence intervals around the hazard ratio estimates of 1.81 - 1.87 and 1.47 - 1.58 respectively, giving the best estimates to date of the importance of competing risks methodology in diabetes outcomes research.

4.2.5. Statistical Analyses

Baseline Demographics

Comparisons between patient albuminuria level and baseline patient characteristics were performed using unpaired two-tailed t-tests[142] for continuous outcomes, and Fisher's exact test[143] for categorical outcomes. All p-values were adjusted for multiple comparisons via Bonferroni's method whereby each p-value was multiplied by the total number of baseline comparisons[144–146].

Primary Analysis

Equality of the survival functions of each albuminuria level was tested using the log-rank test[147–150], while equality of the cumulative incidence functions of each albuminuria level was tested using Gray's K-sample test[151]. These tests were performed for both outcomes.

Standard estimates for the nine-year absolute risks of cardiovascular and cancer mortality were derived using the Kaplan-Meier estimator[3]. Competing-risks-adjusted estimates of the nine-year absolute risks of cardiovascular and cancer mortality were derived using the cumulative incidence competing risk (CICR) estimator[5]. All cumulative risk plots were stratified by albuminuria level. Bias

from competing risks was evaluated as the nine-year absolute and relative difference between the Kaplan-Meier and CICR estimators for each albuminuria group in each outcome. The absolute and relative differences were defined in Equation 4.1 and Equation 4.2, where $p_{km}(t)$ and $p_{CICR}(t)$ represent Kaplan-Meier and CICR cumulative risk probabilities at time t:

Equation 4.1 - Specification of the absolute difference between Kaplan-Meier and CICR cumulative risk estimates.

$$Absolute\ Difference = [p_{KM}(t) - p_{CICR}(t)] \times 100\%$$

Equation 4.2 - Specification of the relative difference between Kaplan-Meier and CICR cumulative risk estimates.

$$Relative\ Difference = \left[\frac{p_{KM}(t) - p_{CICR}(t)}{p_{CICR}(t)}\right] \times 100\%$$

Statistical Software and Implementation

Kaplan-Meier Estimator

Kaplan-Meier cumulative risk estimates were produced using a combination of the "**stset**" and "**sts generate**" commands in Stata 12.1[91]. "**stset**" declared the data to be survival/time-to-event data and recorded the primary outcome of interest for subsequent commands. In the following code example, "**stime**" is a variable of survival times and "**event**" is a variable of outcome types (coded "**0**" for censoring, "**1**" for cardiovascular mortality, "**2**" for cancer mortality, and "**3**" for other mortality). "**sts generate**" was used subsequent to "**stset**" to generate a new variable ("**KM**") containing Kaplan-Meier product limit estimates (specified by the "**s**" object). The "**by**" option informed the function to do this for each albuminuria level. Cumulative risk estimates were generated by subtracted the Kaplan-Meier product limit estimates from one. These were plotted using the "**twoway line**" command.

```
stset stime, failure(event == 1)
sts generate KM = s, by(albuminuria)
replace KM = 1 - KM
```

```
generate KM_normo = KM if event == 1 & albuminuria == 0
generate KM_micromacro = KM if event == 1 & albuminuria == 1

twoway line KM_normo stime, sort lcolor(dkgreen) ///
|| line KM_micromacro stime, sort lcolor(navy) ///
xlabel(0(1)9, labsize(small)) xscale(range(0 9)) ///
xtitle("Time (Years)", size(small)) ///
ylabel(0(0.05)0.25, labsize(small)) yscale(range(0 0.25)) ///
ytitle("Probability", size(small)) ///
legend(off)
```

Cumulative Incidence Competing Risk Estimator

CICR cumulative risk estimates were produced using a combination of the "**stset**" and user-written "**stcompet**"[152] commands in Stata 12.1[91]. "**stset**" declared the data to be survival/time-to-event data and recorded the primary outcome of interest for subsequent commands. The variables used in the following code example are identical to those specified for the Kaplan-Meier estimator. "stcompet" was used subsequent to "**stset**" to generate a new variable ("**CI**") containing CICR cumulative risk estimates (specified by the "**ci**" object). The "**by**" option informed the function to do this for each albuminuria level, while the "**compet1(2)**" and "**compet2(3)**" options informed the "**stcompet**" command which events constituted competing outcomes. Cumulative risk estimates were plotted using the "**twoway line**" command.

```
stset stime, failure(event == 1)
stcompet CI = ci, by(albuminuria) compet1(2) compet2(3)

generate CI_normo = CI if event == 1 & albuminuria == 0
generate CI_micromacro = CI if event == 1 & albuminuria == 1

twoway line CI_normo stime, sort lcolor(dkgreen) ///
|| line CI_micromacro stime, sort lcolor(navy) ///
```

```
xlabel(0(1)9, labsize(small)) xscale(range(0 9)) ///
xtitle("Time (Years)", size(small)) ///
ylabel(0(0.05)0.25, labsize(small)) yscale(range(0 0.25)) ///
ytitle("Probability", size(small)) ///
legend(off)
```

4.2.6. Sensitivity Analyses

The Nelson-Aalen Estimator

The Nelson-Aalen estimator[9] is an infrequently used standard survival analysis method to estimate cumulative risk. The methods of the primary analysis were repeated with the Nelson-Aalen estimator used in place of the Kaplan-Meier estimator to determine whether inferences from the primary analysis would be different if a different standard survival analysis cumulative risk estimator had been used. For these comparisons, the absolute and relative differences were defined in Equation 4.3 and Equation 4.4, where $p_{NA}(t)$ *and* $p_{CICR}(t)$ represent Nelson-Aalen and CICR cumulative risk probabilities at time *t*:

Equation 4.3 - Specification of the absolute difference between Nelson-Aalen and CICR cumulative risk estimates.

$$Absolute\ Difference = [p_{NA}(t) - p_{CICR}(t)] \times 100\%$$

Equation 4.4 - Specification of the relative difference between Nelson-Aalen and CICR cumulative risk estimates.

$$Relative\ Difference = \left[\frac{p_{NA}(t) - p_{CICR}(t)}{p_{CICR}(t)}\right] \times 100\%$$

ICD-10 Recoding

A death certificate lists a single primary cause of death, alongside up to 15 contributing causes. In the primary analysis, only the primary cause was used to define the cause of death. However, ambiguity may arise in cases where cancer contributed to a patient's death from cardiovascular

causes, and vice versa. To assess the potential for this to bias risk estimates, the models of the primary analysis were repeated under the following scenarios:

1. Patient mortality was defined as cardiovascular disease where the disease was listed as either the underlying cause or a contributory cause on the death certificate (even where cancer was stated as the underlying cause). In this scenario patient mortality was only classified as cancer mortality or other mortality if one of these was stated as the underlying cause of mortality, and cardiovascular mortality was not listed as a contributory cause or mortality.
2. Patient mortality was defined as cancer where the disease was listed as either the underlying cause or a contributory cause on the death certificate (even where cardiovascular disease was stated as the underlying cause). In this scenario patient mortality was only classified as cardiovascular mortality or other mortality if one of these was stated as the underlying cause of mortality, and cancer mortality was not listed as a contributory cause or mortality.

Albuminuria Variable Coding

As previously stated in the Methods section, the coding of the albuminuria variable was performed using a hierarchical evidence structure. To assess the agreement between different CPRD sources, a combination of Fleiss'[153] and Cohen's[154] κ coefficients was used; Fleiss' κ to assess the overall agreement and Cohen's κ to assess the agreement between CPRD sources pairs.

To determine the effect of conducting a study based around only a single one of these sources, the methods of the primary analysis were repeated using albuminuria variables defined using each of the three sources in isolation:

1. **Medical codes** - linked to diagnostic codes for albuminuria present in the Clinical File.
2. **Numerical test results** - values for ACRs/PCRs, and 24-hr AERs and PERs, present in the Test File.
3. **Categorical test results** - categorical identifiers of albuminuria, present as "test qualifier units" associated with medical codes for albuminuria in the Test File.

Missing Data

Missing data can bias study results where the data are either 'missing at random' or 'missing not at random'. In this study, it was possible that albuminuria values could be 'missing not at random' as a result of GPs less frequently recording the results of tests that indicated an absence of albuminuria. To test this assumption, the study primary analysis was repeated with missing albuminuria imputed as normoalbuminuria.

4.3. Results

4.3.1. Study Population

Of the initial 280,728 patient sample, 86,962 were eligible for inclusion in this study. Figure 4.1 depicts the selection process outlined in the "Study Population" section of the Methods.

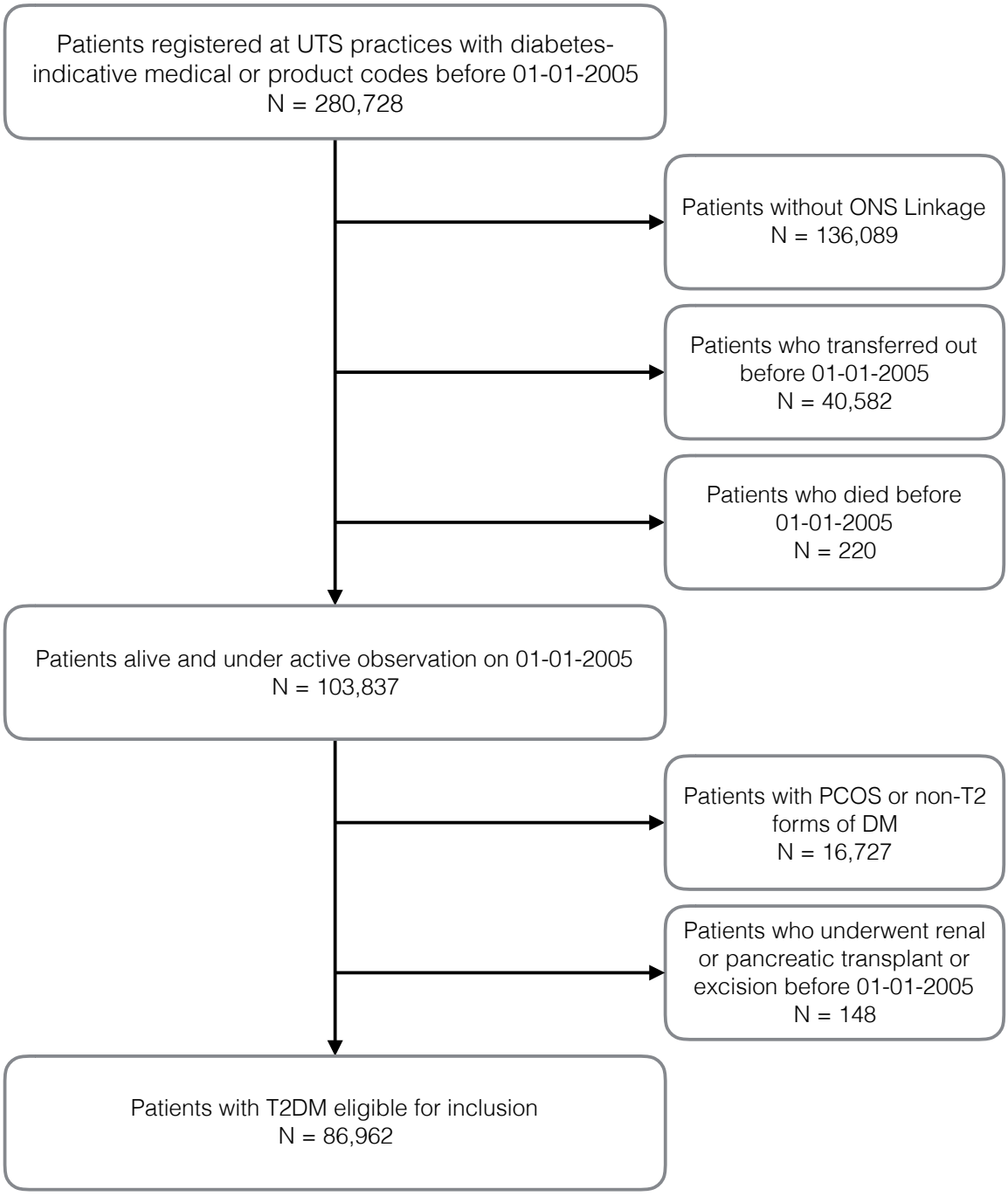


Figure 4.1 - Flow diagram depicting the selection process for the patient population.

4.3.2. Baseline Demographics

The baseline characteristics of the study population are presented in Table 4.2. Patients with albuminuria at baseline were older ($p<0.001$), had higher systolic blood pressure (SBP) ($p<0.001$), had worse glycaemic control (HbA1c) ($p<0.001$), and were more likely to be current or ex-smokers

($p<0.001$). No significant associations were found between baseline albuminuria and: gender, body-mass index (BMI) or total-to-high-density-lipoprotein (Total : HDL) cholesterol ratio.

Table 4.2 - Baseline characteristics of study cohort, stratified by baseline albuminuria level. Data represent means ± SD unless a percentage (%) is stated.

Baseline Variable	All (n = 86,962)	Normoalbuminuria (n = 42,662)	Albuminuria (n = 12,139)
Men	47,011 (54.1%)	23,611 (55.3%)	6,553 (54.0%)
Age (years)	66.6 ± 12.6	66.8 ± 11.8	69.4 ± 12.2
BMI (kg/m^2)	29.8 ± 6.1	29.7 ± 5.9	29.8 ± 6.2
SBP (mm Hg)	138.5 ± 17.4	137.7 ± 16.4	140.8 ± 18.4
Total : HDL Cholesterol Ratio	3.9 ± 1.2	3.8 ± 1.2	3.8 ± 1.2
HbA1c (mmol/mol)	57.5 ± 16.1	56.1 ± 14.7	59.6 ± 17.0
Never Smoked	25,164 (28.9%)	10,499 (24.6%)	2,809 (23.1%)
Ex-Smokers	41,645 (47.9%)	22,604 (52.9%)	6,151 (50.7%)
Current Smokers	20,153 (23.2%)	9,559 (22.4%)	3,179 (26.2%)

4.3.3. Primary Analysis

Over the course of nine years of follow-up 14,201 patients died: 9,558 (67.3%) with normoalbuminuria at baseline and 4,643 (32.7%) with albuminuria at baseline. Of these 14,201 patients: 5,574 (39.3%) died from cardiovascular causes, 3,455 (24.3%) from cancer, and 5,172 (36.4%) died from other causes.

Among patients who died from cardiovascular causes, 3,656 (65.6%) had normoalbuminuria at baseline and 1,918 (34.4%) had albuminuria at baseline. Among patients who died from cancer causes, 2,566 (74.3%) had normoalbuminuria at baseline and 889 (25.7%) had albuminuria at baseline. For both outcomes, log-rank tests determined statistically significant differences between the survival functions of patients with normoalbuminuria and albuminuria ($p<0.001$), while Gray's K-sample test determined statistically significant differences between the cumulative incidence functions of patients with normoalbuminuria and albuminuria ($p<0.001$).

Cumulative risk estimates for cardiovascular and cancer mortality, stratified by albuminuria level, are presented in Figure 4.2, while the associated nine-year absolute risk estimates are presented in Table 4.3.

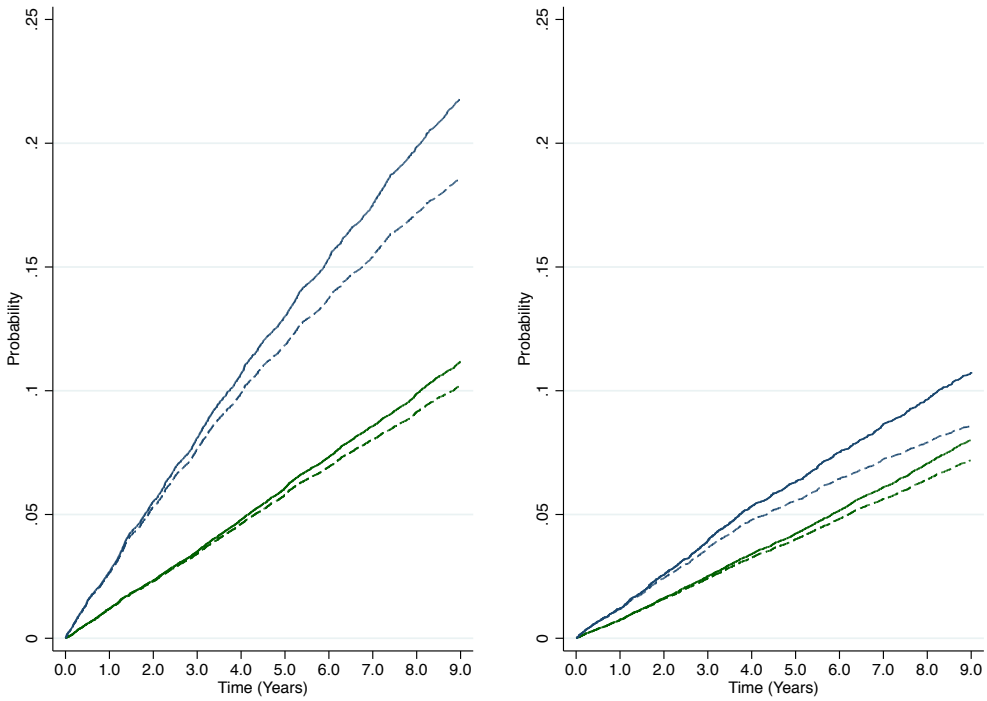


Figure 4.2 - Cumulative risk estimates for cardiovascular (left) and cancer (right) mortality. Solid lines = "Kaplan-Meier", dashed lines = "cumulative incidence competing risk", green lines = "normoalbuminuria", blue lines = "albuminuria".

Table 4.3 - Nine-year absolute risk estimates for cardiovascular and cancer mortality from Kaplan-Meier and cumulative incidence competing risk estimators.

Albuminuria Level	**Cardiovascular Mortality Risk %, (95% CI)**		**Cancer Mortality Risk %, (95% CI)**	
	Kaplan-Meier	**CICR**	**Kaplan-Meier**	**CICR**
Normoalbuminuria	11.1 (10.8 - 11.5)	10.2 (9.9 - 10.5)	8.0 (7.7 - 8.3)	7.2 (6.9 - 7.5)
Albuminuria	21.8 (20.9 - 22.7)	18.5 (17.8 - 19.3)	10.7 (10.0 - 11.5)	8.6 (8.1 - 9.2)

For both outcomes in patients with and without albuminuria, cumulative hazard estimates from Kaplan-Meier estimators produced higher risk estimates than cumulative incidence estimates from CICR estimators. Estimates from Kaplan-Meier and CICR estimators were similar for the first two to three years, but progressively diverged with the passage of time. Absolute and relative differences

between the Kaplan-Meier and CICR estimators for each outcome and albuminuria level may be found in Table 4.4.

Table 4.4 - Absolute and relative differences between Kaplan-Meier and CICR estimators.

Albuminuria Level	Cardiovascular Mortality		Cancer Mortality	
	Absolute Difference (%)	Relative Difference (%)	Absolute Difference (%)	Relative Difference (%)
Normoalbuminuria	0.9	8.8	0.8	11.1
Albuminuria	3.3	17.8	2.1	24.4

For a given albuminuria level, absolute differences between the Kaplan-Meier and CICR estimators were greater for cardiovascular mortality, while relative differences were greater for cancer mortality. For a given outcome absolute and relative differences between the Kaplan-Meier and CICR estimators were greater for patients with albuminuria than normoalbuminuria.

4.3.4. Sensitivity Analyses

The Nelson-Aalen Estimator

Cumulative risk estimates from the Nelson-Aalen estimator for both outcomes are presented alongside the those of the Kaplan-Meier and CICR estimators in Figure 4.3. Nine-year absolute risk estimates from the aforementioned risk estimators are presented in Table 4.5.

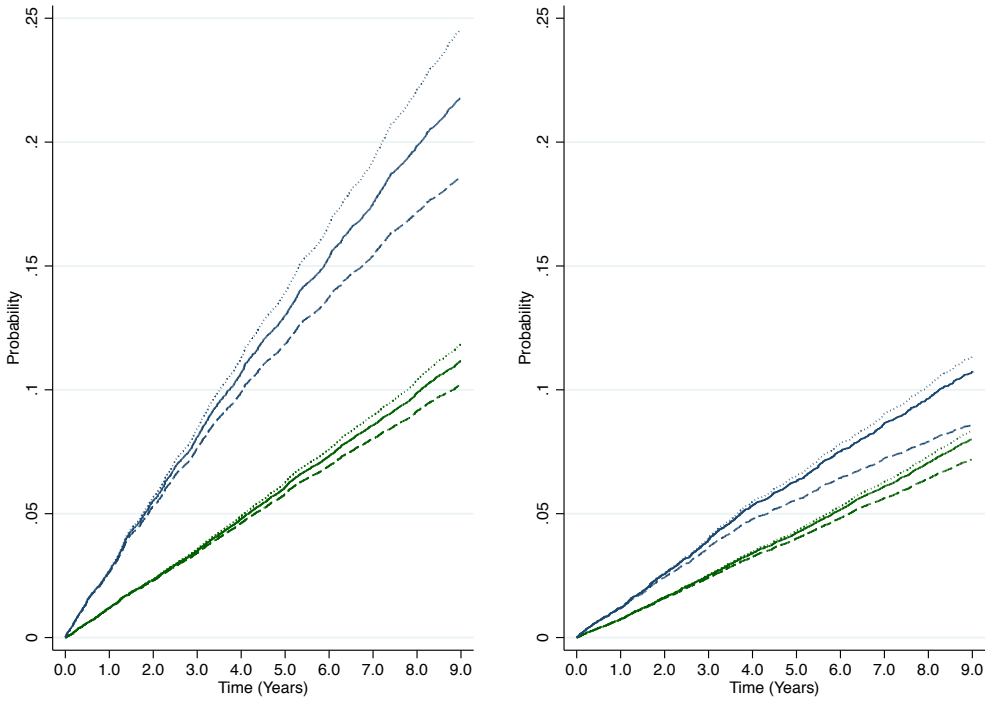


Figure 4.3 - Cumulative risk estimates for cardiovascular (left) and cancer (right) mortality. Dotted lines = "Nelson-Aalen", solid lines = "Kaplan-Meier", dashed lines = "cumulative incidence competing risk", green lines = "normoalbuminuria", blue lines = "albuminuria".

Table 4.5 - Nine-year absolute risk estimates for cardiovascular and cancer mortality from Nelson-Aalen, Kaplan-Meier and cumulative incidence competing risk estimators.

Albuminuria Level	**Cardiovascular Mortality Risk %, (95% CI)**			**Cancer Mortality Risk %, (95% CI)**		
	Nelson-Aalen	**Kaplan-Meier**	**CICR**	**Nelson-Aalen**	**Kaplan-Meier**	**CICR**
Normoalbuminuria	11.8 (11.5 - 12.3)	11.1 (10.8 - 11.5)	10.2 (9.9 - 10.5)	8.3 (8.0 - 8.7)	8.0 (7.7 - 8.3)	7.2 (6.9 - 7.5)
Albuminuria	24.6 (23.5 - 25.8)	21.8 (20.9 - 22.7)	18.5 (17.8 - 19.3)	11.3 (10.6 - 12.2)	10.7 (10.0 - 11.5)	8.6 (8.1 -9.2)

For both outcomes in patients with and without albuminuria, cumulative hazard estimates from Nelson-Aalen estimators were greater than those from Kaplan-Meier estimators, and greater still than cumulative incidence estimates from CICR estimators. Estimates from all estimators were similar for the first two to three years, but progressively diverged thereafter. Absolute and relative differences between the Nelson-Aalen and CICR estimators for each outcome and albuminuria level may be found in Table 4.6.

Table 4.6 - Absolute and relative differences between Nelson-Aalen and CICR estimators.

Albuminuria Level	Cardiovascular Mortality		Cancer Mortality	
	Absolute Difference (%)	Relative Difference (%)	Absolute Difference (%)	Relative Difference (%)
Normoalbuminuria	1.6	15.7	1.1	15.3
Albuminuria	6.1	33.0	2.7	31.4

For a given albuminuria level, absolute and relative differences between the Nelson-Aalen and CICR estimators were greater for cardiovascular mortality. Concomitantly, for a given outcome absolute and relative differences between the Nelson-Aalen and CICR estimators were greater for patients with albuminuria than with normoalbuminuria.

ICD-10 Recoding

Reclassification of mortality as cardiovascular (where cardiovascular mortality was listed as either the underlying or contributory cause of mortality) resulted in cause of mortality being altered in 2,998 patients. Reclassification of mortality as cancer (where cancer mortality was listed as either the underlying or contributory cause of mortality) resulted in the cause of mortality being altered in 275 patients. More detail on the reclassifications is presented in Table 4.7.

Table 4.7 - The number of patients with each outcome in the primary analysis, and after the reclassification of mortality according to the presence of causes (cardiovascular or cancer) listed as being contributory.

Outcome	Primary Analysis	Reclassification as Cardiovascular Mortality	Reclassification as Cancer Mortality
Cardiovascular Mortality	5,574 (39.3%)	8,572 (60.4%)	5,299 (37.3%)
Cancer Mortality	3,455 (24.3%)	2,787 (19.6%)	3,998 (28.2%)
Other Mortality	5,172 (36.4%)	2,842 (20.0%)	4,904 (34.5%)
Total Mortality	14,201 (100%)	14,201 (100%)	14,201 (100%)

Cumulative risk estimates following the reclassification of mortality (as cardiovascular and cancer) are presented in Figure 4.4 and Figure 4.5, while the associated nine-year absolute risk estimates

are presented in Table 4.8 and Table 4.9. Reclassification of mortality as cardiovascular disease produced higher cumulative and absolute risk estimates for cardiovascular mortality, and lower cumulative risk estimates for cancer mortality. Comparatively, reclassification of mortality as cancer produced lower cumulative and absolute risk estimates for cardiovascular mortality, and higher cumulative risk estimates for cancer mortality.

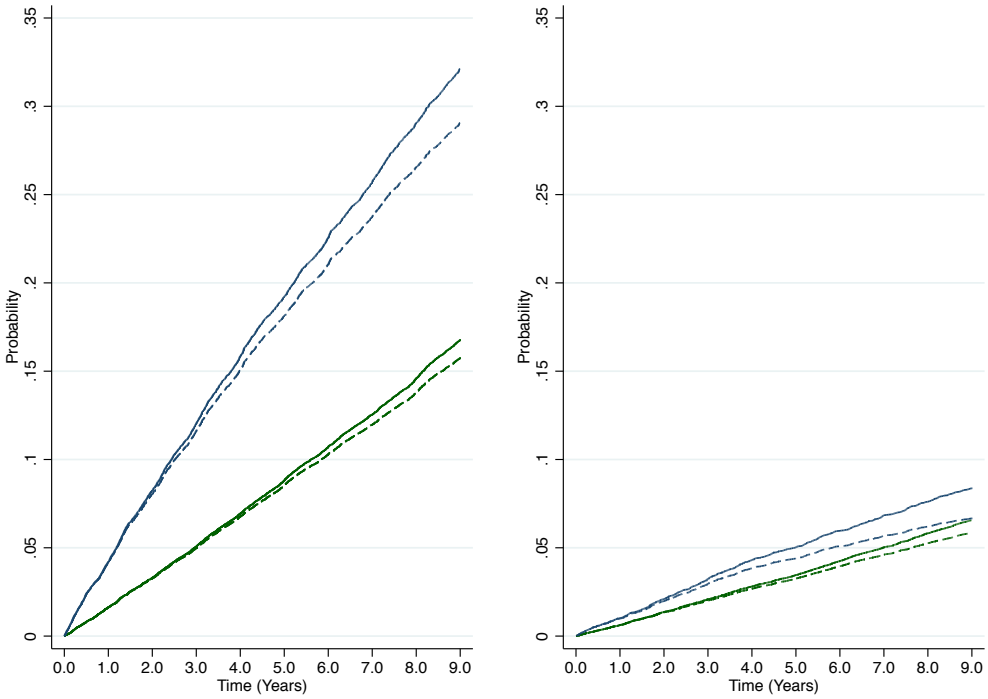


Figure 4.4 - Cumulative risk estimates for cardiovascular (left) and cancer (right) mortality, with mortality reclassified as cardiovascular where it was listed as a contributory cause on patient death certificates. Solid lines = "Kaplan-Meier", dashed lines = "cumulative incidence competing risk", green lines = "normoalbuminuria", blue lines = "albuminuria". The y-axis is rescaled from 0-0.35 in the above plots.

Table 4.8 - Nine-year absolute risk estimates for cardiovascular and cancer mortality from Kaplan-Meier and cumulative incidence competing risk estimators, with mortality reclassified as cardiovascular where it was listed as a contributory cause on patient death certificates.

Albuminuria Level	**Cardiovascular Mortality Risk %, (95% CI)**		**Cancer Mortality Risk %, (95% CI)**	
	Kaplan-Meier	**CICR**	**Kaplan-Meier**	**CICR**
Normoalbuminuria	16.8 (16.4 - 17.2)	15.7 (15.4 - 16.1)	6.6 (6.3 - 6.8)	5.8 (5.6 - 6.1)
Albuminuria	32.1 (31.1 - 33.1)	29.1 (28.2 - 29.9)	8.4 (7.7 - 9.0)	6.7 (6.2 - 7.1)

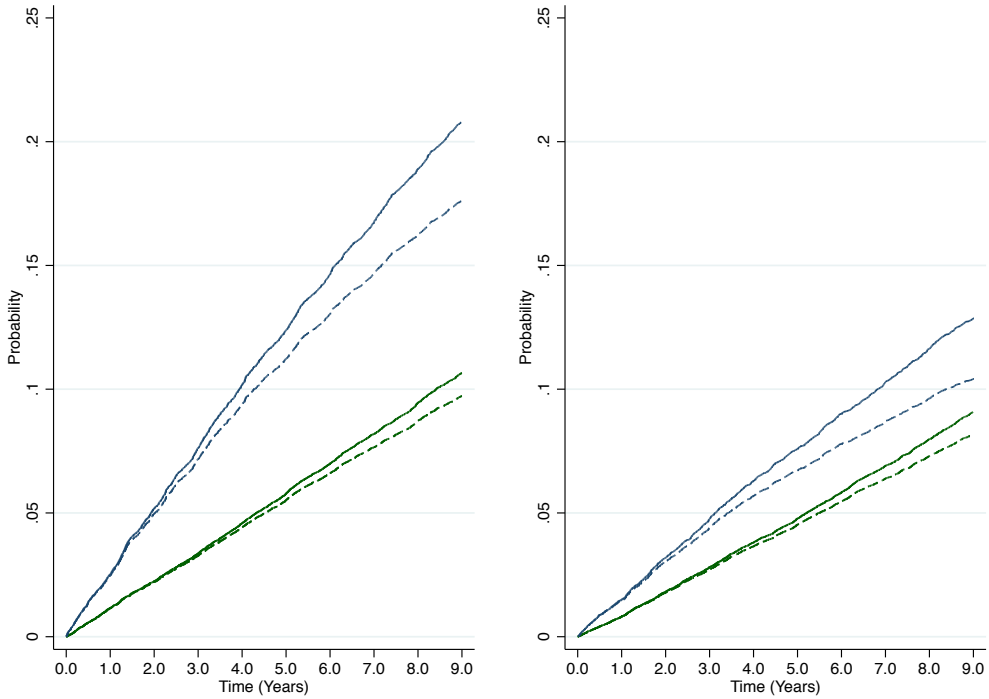


Figure 4.5 - Cumulative risk estimates for cardiovascular (left) and cancer (right) mortality, with mortality reclassified as cardiovascular where it was listed as a contributory cause on patient death certificates. Solid lines = "Kaplan-Meier", dashed lines = "cumulative incidence competing risk", green lines = "normoalbuminuria", blue lines = "albuminuria".

Table 4.9 - Nine-year absolute risk estimates for cardiovascular and cancer mortality from Kaplan-Meier and cumulative incidence competing risk estimators, with mortality reclassified as cancer where it was listed as a contributory cause on patient death certificates.

Albuminuria Level	**Cardiovascular Mortality Risk %, (95% CI)**		**Cancer Mortality Risk %, (95% CI)**	
	Kaplan-Meier	**CICR**	**Kaplan-Meier**	**CICR**
Normoalbuminuria	10.7 (10.3 - 11.0)	9.7 (9.4 - 10.0)	9.1 (8.8 - 9.4)	8.2 (7.9 - 8.5)
Albuminuria	20.8 (19.9 - 21.7)	17.6 (16.8 - 18.3)	12.9 (12.1 - 13.6)	10.4 (9.8 - 11.0)

Absolute and relative differences between the Kaplan-Meier and CICR estimators at nine years after each reclassification are presented in Table 4.10. Estimates for the absolute difference varied from those obtained in the primary analysis by 0.1 - 0.4%. Estimates for the relative difference varied from those obtained in the primary analysis by 0.4 - 2.7%.

Table 4.10 - Absolute and relative differences between Kaplan-Meier and CICR estimators, where mortality was reclassified dependent upon the presence of cardiovascular disease or cancer as contributory causes.

Mortality Reclassified as	Albuminuria Level	Cardiovascular Mortality		Cancer Mortality	
		Absolute Difference (%)	Relative Difference (%)	Absolute Difference (%)	Relative Difference (%)
Cardiovascular Mortality	Normoalbuminuria	1.1	7.0	0.8	13.8
Cancer Mortality		1.0	10.3	0.9	11.0
Cardiovascular Mortality	Albuminuria	3.0	10.3	1.7	25.4
Cancer Mortality		3.2	18.2	2.5	24.0

Albuminuria Variable Coding

Recoding the albuminuria variable resulted in large variability in the numbers and proportions of patients with and without albuminuria at baseline (Table 4.11). The use of numerical test values alone resulted in an albuminuria population split most closely resembling that of the primary analysis. Low overall agreement was demonstrated between the three CPRD sources used to define the albuminuria variable (Fleiss' $\kappa = 0.170$). Looking at complete cases only, the highest agreement was found between medical codes and numerical test values (Cohen's $\kappa = 0.234$), followed by numerical and categorical test values (Cohen's $\kappa = 0.229$), with medical codes and categorical test values having the lowest agreement (Cohen's $\kappa = 0.057$).

Table 4.11 - Patients classifiable using different CPRD data sources for albuminuria.

CPRD Source	Albuminuria Level, n (%)		
	Normoalbuminuria	Albuminuria	Total
Medical Codes Only	1,102 (18.4%)	4,873 (81.6%)	5,975 (100%)
Numerical Test Values Only	21,330 (74.0%)	7,498 (26.0%)	28,828 (100%)
Categorical Test Values Only	36,227 (91.9%)	3,183 (8.1%)	39,410 (100%)
All (Hierarchical Structure)	42,666 (77.9%)	12,135 (22.1%)	54,801 (100%)

Cumulative risk estimates for cardiovascular and cancer mortality with albuminuria defined using different CPRD sources, stratified by albuminuria level, are presented in Figure 4.6, Figure 4.7 and

Figure 4.8, while the associated nine-year absolute risk estimates are presented in Table 4.12, Table 4.13 and Table 4.14. Cumulative and absolute risk estimates were similar between CPRD sources, with nine-year absolute risks differing from those of the primary analysis by 0.2 - 1.4%.

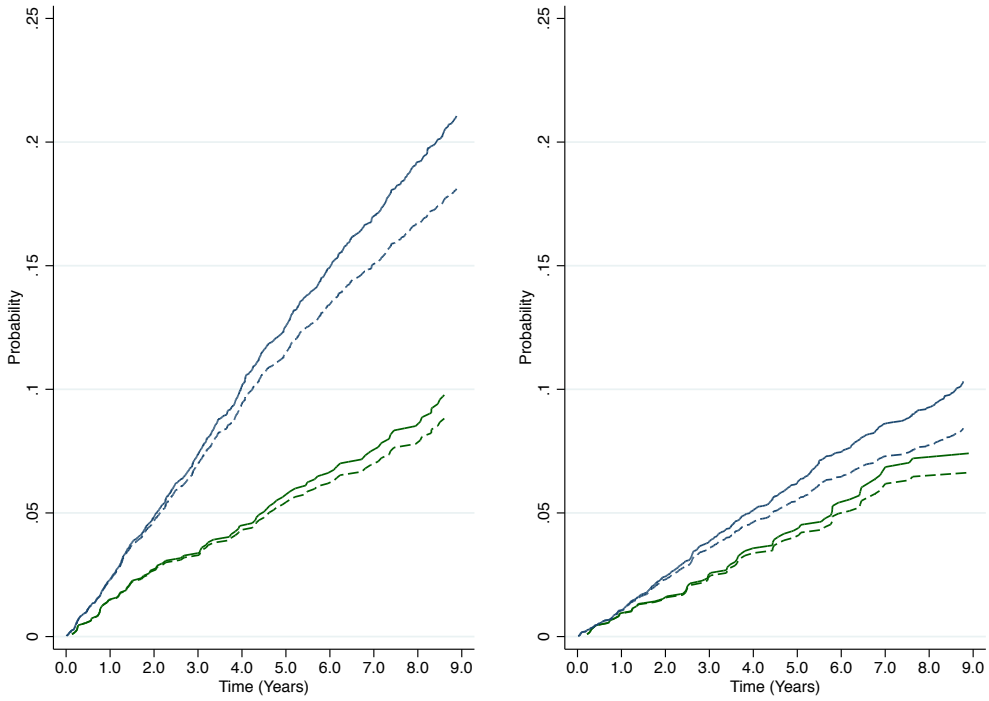


Figure 4.6 - Cumulative risk estimates for cardiovascular (left) and cancer (right) mortality, with albuminuria defined using medical codes only. Solid lines = "Kaplan-Meier", dashed lines = "cumulative incidence competing risk", green lines = "normoalbuminuria", blue lines = "albuminuria".

Table 4.12 - Nine-year absolute risk estimates for cardiovascular and cancer mortality from Kaplan-Meier and cumulative incidence competing risk estimators, with albuminuria defined using medical codes only.

Albuminuria Level	**Cardiovascular Mortality Risk %, (95% CI)**		**Cancer Mortality Risk %, (95% CI)**	
	Kaplan-Meier	**CICR**	**Kaplan-Meier**	**CICR**
Normoalbuminuria	9.8 (7.9 - 12.1)	8.8 (7.1 - 10.8)	7.4 (5.8 - 9.5)	6.6 (5.1 - 8.4)
Albuminuria	21.1 (19.7 - 22.5)	18.1 (16.9 - 19.3)	10.3 (9.3 - 11.4)	8.4 (7.6 - 9.3)

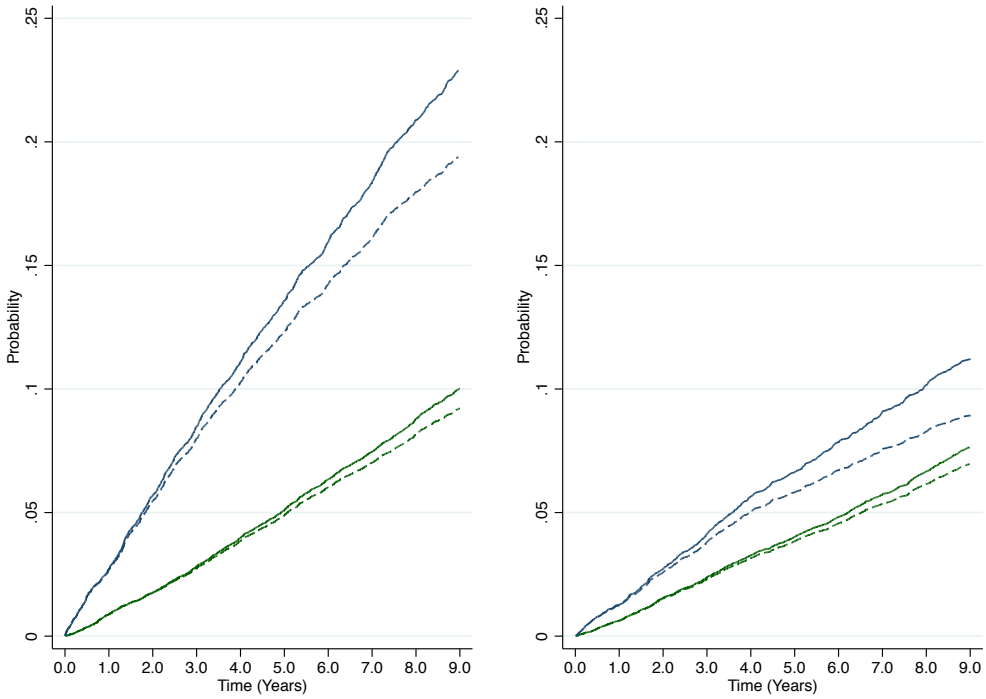


Figure 4.7 - Cumulative risk estimates for cardiovascular (left) and cancer (right) mortality, with albuminuria defined using numerical test values only. Solid lines = "Kaplan-Meier", dashed lines = "cumulative incidence competing risk", green lines = "normoalbuminuria", blue lines = "albuminuria".

Table 4.13 - Nine-year absolute risk estimates for cardiovascular and cancer mortality from Kaplan-Meier and cumulative incidence competing risk estimators, with albuminuria defined using numerical test values.

Albuminuria Level	**Cardiovascular Mortality Risk %, (95% CI)**		**Cancer Mortality Risk % (95% CI)**	
	Kaplan-Meier	**CICR**	**Kaplan-Meier**	**CICR**
Normoalbuminuria	10.0 (9.6 - 10.5)	9.2 (8.8 - 9.7)	7.6 (7.2 - 8.1)	7.0 (6.6 - 7.3)
Albuminuria	22.9 (21.7 - 24.1)	19.4 (18.4 - 20.4)	11.2 (10.4 - 12.2)	8.9 (8.3 - 9.7)

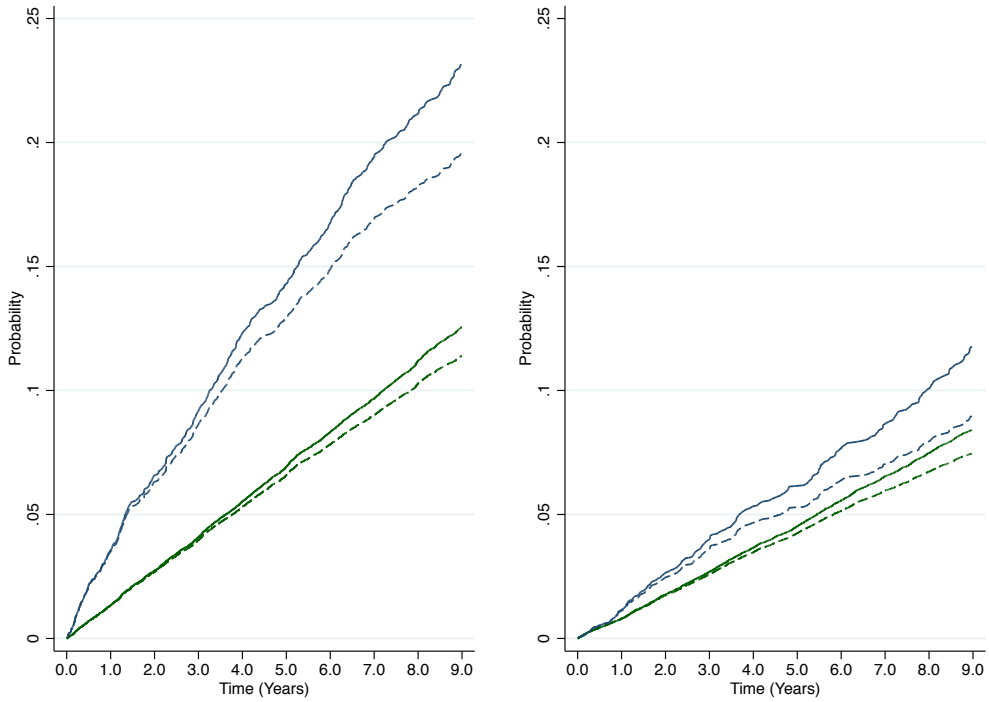


Figure 4.8 - Cumulative risk estimates for cardiovascular (left) and cancer (right) mortality, with albuminuria defined using categorical test values only. Solid lines = "Kaplan-Meier", dashed lines = "cumulative incidence competing risk", green lines = "normoalbuminuria", blue lines = "albuminuria".

Table 4.14 - Nine-year absolute risk estimates for cardiovascular and cancer mortality from Kaplan-Meier and cumulative incidence competing risk estimators, with albuminuria defined using categorical test values only.

Albuminuria Level	**Cardiovascular Mortality Risk %, (95% CI)**		**Cancer Mortality Risk %, (95% CI)**	
	Kaplan-Meier	**CICR**	**Kaplan-Meier**	**CICR**
Normoalbuminuria	12.6 (12.2 - 13.0)	11.4 (11.0 - 11.7)	8.4 (8.1 - 8.7)	7.5 (7.1 - 7.8)
Albuminuria	23.2 (21.4 - 25.1)	19.6 (18.1 - 21.1)	11.8 (10.4 - 13.4)	9.0 (7.9 - 10.1)

Absolute and relative differences between the Kaplan-Meier and CICR estimators at nine years after each reclassification are presented in Table 4.15. Estimates for the absolute difference varied from those obtained in the primary analysis by 0.0 - 0.7%. Estimates for the relative difference varied from those obtained in the primary analysis by 0.1 - 6.7%.

Table 4.15 - Absolute and relative differences between Kaplan-Meier and CICR estimators for albuminuria levels defined using different CPRD sources.

CPRD Source	Albuminuria Level	Cardiovascular Mortality		Cancer Mortality	
		Absolute Difference (%)	Relative Difference (%)	Absolute Difference (%)	Relative Difference (%)
Medical Codes	Normoalbuminuria	1.0	11.3	0.8	12.1
Numerical Test Values		0.8	8.7	0.6	8.6
Categorical Test Values		1.2	10.5	0.9	12.0
Medical Codes	Albuminuria	3.0	16.6	1.9	22.6
Numerical Test Values		3.5	18.0	2.3	25.8
Categorical Test Values		3.6	18.4	2.8	31.1

Missing Data

The reclassification of missing albuminuria values as normoalbuminuria increased the number of patients that could be included in analysis by 32,161. It also increased the number of observed events, as depicted in Table 4.16.

Table 4.16 - The number of patients experiencing each outcome in the primary analysis, and after the recoding of missing albuminuria data as normoalbuminuria.

Outcome	Primary Analysis (N = 54,801)	Missing Data Analysis (N = 86,962)
Cardiovascular Mortality	5,574	8,800
Cancer Mortality	3,455	5,239
Other Mortality	5,172	8,473

Cumulative risk estimates for cardiovascular and cancer mortality, with missing albuminuria values imputed as normoalbuminuria, are presented in Figure 4.9, while the associated nine-year absolute risk estimates are presented in Table 4.17. Cumulative and absolute risk estimates for patients with

albuminuria at baseline were identical to those of the primary analysis. For patients with normoalbuminuria at baseline, cumulative and absolute risks estimates were greater than those of the primary analysis, with nine-year absolute risks differing from those of the primary analysis by 0.1 - 0.9%.

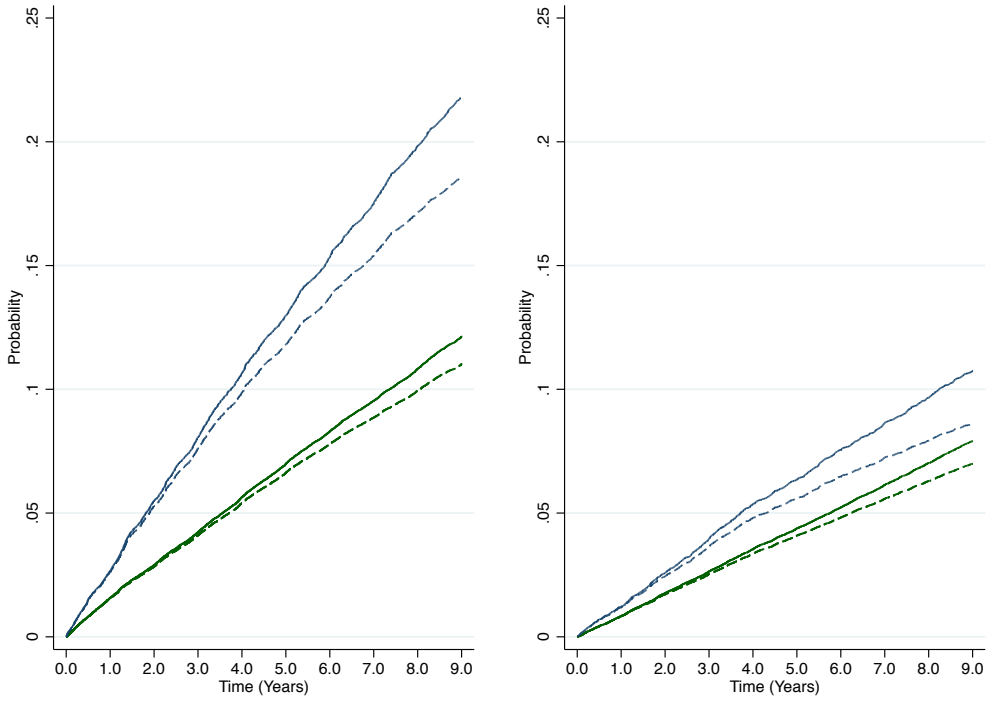


Figure 4.9 - Cumulative risk estimates for cardiovascular (left) and cancer (right) mortality, with missing albuminuria data recoded as normoalbuminuria. Solid lines = "Kaplan-Meier", dashed lines = "cumulative incidence competing risk", green lines = "normoalbuminuria", blue lines = "albuminuria".

Table 4.17 - Nine-year absolute risk estimates for cardiovascular and cancer mortality from Kaplan-Meier and cumulative incidence competing risk estimators, with missing albuminuria data recoded as normoalbuminuria.

Albuminuria Level	**Cardiovascular Mortality Risk % (95% CI)**		**Cancer Mortality Risk %, (95% CI)**	
	Kaplan-Meier	**CICR**	**Kaplan-Meier**	**CICR**
Normoalbuminuria	12.1 (11.8 - 12.4)	11.0 (10.7 - 11.3)	7.9 (7.7 - 8.1)	7.0 (6.8 - 7.2)
Albuminuria	21.8 (20.9 - 22.7)	18.5 (17.8 - 19.3)	10.7 (10.1 - 11.5)	8.6 (8.1 - 9.2)

Absolute and relative differences between the Kaplan-Meier and CICR estimators at nine years after each reclassification are presented in Table 4.18 Estimates for the absolute and relative

difference for patients with albuminuria were identical to those obtained in the primary analysis. Estimates for the absolute and relative difference for patients with normoalbuminuria differed from those obtained in the primary analysis by 0.0 - 0.3% and 1.2 - 1.8%, respectively.

Table 4.18 - Absolute and relative differences between Kaplan-Meier and CICR estimators after the imputation of missing albuminuria data as normoalbuminuria.

Albuminuria Level	Cardiovascular Mortality		Cancer Mortality	
	Absolute Difference (%)	Relative Difference (%)	Absolute Difference (%)	Relative Difference (%)
Normoalbuminuria	1.1	10.0	0.9	12.9
Albuminuria	3.3	17.8	2.1	24.4

4.4. Discussion

4.4.1. Key Results

Cumulative hazards from the Kaplan-Meier estimator produced consistently higher risk estimates than cumulative incidence estimates from the CICR estimators for both albuminuria strata, and both outcomes. Absolute and relative differences between the Kaplan-Meier and CICR estimators were greater in patients with albuminuria than normoalbuminuria. With respect to the outcomes of cardiovascular and cancer mortality, greater absolute differences were present for cardiovascular mortality, while greater relative differences were present for cancer mortality.

Cumulative hazard estimates from the Nelson-Aalen estimator were even greater than those of the Kaplan-Meier estimator, resulting in absolute and relative differences between the CICR and Nelson-Aalen estimators more substantial than those described between the CICR and Kaplan-Meier estimators in the primary analysis.

Estimates for absolute and relative differences between the Kaplan-Meier and CICR estimators proved robust to most sensitivity analyses performed. However, some relative difference estimates

for patients with albuminuria were found to be sensitive to the reclassification of mortality as cardiovascular and to the use of categorical test values to define the albuminuria variable.

Log-rank tests determined significant differences in survivor functions between albuminuria strata for both cardiovascular and cancer mortality, implying a statistically significant effect of albuminuria on both outcomes.

4.4.2. Strengths and Limitations

Strengths and Limitations of the Dataset

This was a large study, comprising almost 87,000 patients with T2DM sourced from the world renowned CPRD database[129]. Furthermore, patient mortality was established using ONS data, a source for which there is no equal in the field of database research. Therefore, this study provides the best estimates to date for the effects of incorporating competing risks methodology on absolute risk estimates for cardiovascular and cancer mortality in this population.

Despite its advantages in providing reliable mortality data, ONS data is not currently available for linkage to CPRD for the whole of the UK, restricting the source population sample to patients registered at practices in England. Hence, in opting to establish mortality using ONS data, an element of geographical generalisability was lost. Although ONS mortality data captures the majority of deaths occurring in CPRD, it may have been possible to extract the date and cause of mortality from CPRD alone. This would have enabled the use of a more geographically representative population sample. However, such data may be biased, as there would be no way of knowing whether the cause of death listed was the primary cause of death, or a contributory cause.

The large number of patients with missing albuminuria data (N = 32,161) was a limitation to the accuracy of all risk estimates generated. One possible explanation for the missing data is that the study start date was less than a year after the advent of the Quality Outcomes Framework (QOF), in April 2004[135]. Under QOF, GP practices are financially incentivised to monitor the kidney function of patients with T2DM at least once per year, and to record the data in a uniform manner. Sensitivity

analyses tested the possible scenario of GPs being less likely to record patient albuminuria for negative test results; something which may have been more likely before QOF. However, imputing missing values as normoalbuminuria had limited impact on results.

Urine albumin levels typically exhibit a high degree of variability within subjects[155,156]. This meant there was potential for the misclassification of patients' true underlying albuminuria level from test results that were abnormally high or low at any particular visit. In turn, this may have explained some of the low agreement between the different albuminuria data sources in the sensitivity analyses. The National Institute for Health and Care Excellence's 2014 CKD assessment guidelines require that GPs perform a further two tests following an initial positive ACR test, if the test result is <70mg/mmol[157]. However, while this may aid in the prevention of incorrectly diagnosing patients with albuminuria, it increases the potential for patients to be incorrectly diagnosed with normoalbuminuria. Hence, an unknown proportion of patients classified as having normoalbuminuria in this study might have had albuminuria. There are no simple means to estimate a patient's true underlying level of albuminuria from 'messy' biomarkers, and the methods to do so are outside the scope of this Thesis[158,159].

Death certificates are known to be an unreliable source from which to establish the cause of patient mortality that was never intended for use in research[160]. One study of autopsy cases in America demonstrated an under-reporting of malignancies and over-reporting of vascular deaths[161]. This may explain the large shifts in absolute risk estimates demonstrated in sensitivity analyses in this study (where the cause of mortality was altered to reflect the presence of cardiovascular disease or cancer being listed as contributory causes of mortality on death certificates). Other studies have also identified major disagreements between death certificates and post mortem findings that resulted in the reclassification of the cause of mortality listed on the death certificate to one of a completely different disease classification in 29% of cases[162]. Unfortunately, the cause of mortality is rarely certain in the absence of post mortem[160]; a procedure which is implausible for historic cohorts, while the cost and practicality render it prohibitive for many projects.

Strengths and Limitations of the Analyses

The results of this study were robust to a range of sensitivity analysis, allowing great confidence to be placed on their accuracy. Moreover, the use of cumulative risk estimators that are frequently used in the literature and readily implementable in statistical software alongside simple means to quantify the bias present from competing risks, added to the general interpretability of the findings.

Non-parametric estimators can only provide unadjusted estimates of risk, which can prove problematic from the perspective of producing unbiased risk estimates in the presence of confounding. In this study, several patient baseline characteristics demonstrated statistically significant associations with albuminuria (age, SBP, HbA1c, and smoking status). Thus, it is not clear from the analyses performed whether the risk difference between albuminuria strata for each outcome was truly a result of patient albuminuria, or a result of other risk factors associated with it. Further stratification by the potentially confounding risk factors may have helped to clarify this. However, the determination of associations between these risk factors and cardiovascular and cancer mortality was not the aim of this study.

The 'absolute difference' and 'relative difference' were both susceptible to various forms of bias, for which there was no simple means to rectify. Studies comparing the cumulative risk estimates of competing risk and standard survival analysis estimators typically either do not quantify the difference between estimators[1] or use the 'absolute difference' as a means to quantify the bias present in standard methods[8]. However, the absolute difference is a quantity that is not invariant to time, rendering this metric difficult to compare between studies. Furthermore, the information provided by this metric may be considered largely meaningless without qualifying information in the form of the cumulative risk of the primary outcome, or all competing outcomes. For example, a 6% absolute difference between Kaplan-Meier and CICR estimators for an outcome with a Kaplan-Meier cumulative risk of 12% could be considered very different from the same absolute difference between estimators for a Kaplan-Meier cumulative risk of 80%. In order to qualify this difference, the relative difference was devised. However, both the absolute difference and relative difference

were susceptible to floor and ceiling effects that limited the scope for the presentation of differences where cumulative risks were near 0% or 100%, respectively.

It is possible to estimate cumulative risk curves from the hazard or subdistribution hazard functions of proportional hazards models[163]; however, this was not performed in this chapter. Thus, the differences described between the cumulative risk estimates of Kaplan-Meier and CICR estimators may also be present between cumulative risk estimates derived from Cox-PH and Fine-Gray models. Furthermore, such differences have been described in the literature[17]. However, the use of such models for this purpose is uncommon, and is considered outside the scope of this Thesis.

Analyses in this chapter were limited to the evaluation of a single censoring mechanism for the Kaplan-Meier and Nelson-Aalen estimators. However, three distinct censoring mechanisms may be used in the aforementioned estimators[10]:

- **The "Censor All" Mechanism** - censoring of patients experiencing fatal or non-fatal competing events or who are lost to follow up.
- **The "Censor Death Only" Mechanism** - censoring of patients experiencing fatal competing events or who are lost to follow-up only. Patients experiencing non-fatal competing events are ignored by this method.
- **The "Ignore All" Mechanism** - censoring of patients who are lost to follow-up only. Patients experiencing fatal and non-fatal competing events are ignored by this method.

The "censor all" mechanism was used for the Kaplan-Meier and Nelson-Aalen estimators in this study, as it was the mechanism liable to demonstrate the greatest risk difference between standard- and competing-risks-adjusted methods. However, the majority of studies will fall between the "censor all" and "censor death only", either by design or through an inability to correctly identify all non-fatal competing events.

This study did not evaluate the effect of non-fatal competing outcomes on the absolute and relative difference between the Kaplan-Meier and CICR estimators. The definition of competing risks allows

for the possibility of non-fatal events to compete with the occurrence of the study primary outcome if they either fundamentally alter the probability of the main event occurring or preclude it entirely[1]. However, establishing which non-fatal events compete with the primary outcome is not a simple task and requires extensive knowledge of the diseases and study outcomes under evaluation. Thus, while it may seem obvious that death from cancer precludes the occurrence of cardiovascular mortality, the degree to which, for example, renal replacement therapy alters the probability of occurrence of either cardiovascular or cancer mortality is less clear. Moreover, the degree to which an event must alter the probability of occurrence of the primary study outcome in order to be classified as a non-fatal competing event is not addressed in the literature. For these reasons, non-fatal competing events were not evaluated in this study.

4.4.3. Implications

When competing risks are present, large absolute and relative differences may be present between CSH and SDH methods of calculating absolute risk (like the Kaplan-Meier and CICR estimators, respectively). Relative differences in absolute risk estimates may be particularly apparent for less common outcomes or exposures (like cancer mortality or the presence of albuminuria, respectively). No clear trends in absolute difference were evident in absolute risk estimates.

Differences between The Kaplan-Meier and CICR estimator reflect the strength of the competing risks present, with the two estimators only providing identical estimates of cumulative risk in the evaluation of outcomes for which there are no competing outcomes, like all-cause mortality[1]. Thus, differences between these estimators will also be present in other clinical areas in which outcomes are present that compete with the primary outcome.

In the next chapter univariable and multivariable relative hazards models will be used to assess the effect of albuminuria (as a covariate) on cardiovascular and cancer mortality.

5. Competing Risks and Measures of Relative Risk

5.1. Introduction

5.1.1. Background

The previous chapter set out to identify the effect of competing risks on the absolute risks of cardiovascular and cancer mortality in patients with type 2 diabetes mellitus (T2DM), stratified by albuminuria level. This chapter will determine the effect of albuminuria on the relative hazards of cardiovascular and cancer mortality in Clinical Practice Research Datalink (CPRD), and the effect of incorporating competing risks methodology on risk estimates for these outcomes. As an extension to this, this chapter will compare the effects of albuminuria on the cause-specific hazards (CSHs) and subdistribution hazards (SDHs) of cardiovascular and cancer mortality. From this, inferences may be made on the suitability of CSH survival models for the evaluation of the effect of albuminuria on each outcome, and on the ability of albuminuria to describe aetiological and predictive relationships between itself and each outcome.

5.1.2. Aims and Objectives

Primary Objective

To evaluate how competing risks methodology influences estimates for the effect of albuminuria on the relative hazards of cardiovascular and cancer mortality in patients with T2DM.

Specific Aims

i. To compare estimates derived from standard (Cox proportional hazards (Cox-PH)) and competing risks (Lunn-McNeil) methodology for the effect of albuminuria on the CSHs of cardiovascular and cancer mortality.

ii. To compare the effect of albuminuria on the CSHs (Lunn-McNeil) and SDHs (Fine-Gray) of cardiovascular and cancer mortality, in a competing risks framework.

5.2. Methods

5.2.1. Study Population

Details on the data source and the selection process for the study population are described in Chapters 3 and 4. Briefly, the analysis population consisted of 86,962 patients (47,011 men and 39,951 women) from the CPRD database whose medical records indicated the presence of T2DM.

5.2.2. Main Exposure Variable

Albuminuria [No Units]

Albuminuria was defined as a dichotomous variable, i.e. as normoalbuminuria and albuminuria. The methods used to classify patients into these two categories are outlined in Chapter 4.

5.2.3. Study Adjustment Variables

Adjustment Variable Selection

After consulting with my clinical supervisor, a standard set of cardiovascular risk factors were selected for people with T2DM. These included: age, gender, body mass index (BMI), smoking, systolic blood pressure (SBP), total-to-high-density-lipoprotein (Total : HDL) cholesterol ratio[164] and glycated haemoglobin (HbA1c)[165].

Age [Years]

Age was defined using information contained in the "yob" (year of birth) variable of the Patient File. No information regarding the day or month of birth was provided on the grounds of patient anonymity. For each patient, the date of birth was assumed to be the midpoint of the year, i.e. 30th June. This date was subsequently subtracted from the study start date to provide an estimate of each patient's age.

Gender [No Units]

Gender was defined using the "gender" variable in the Patient File.

Body Mass Index [kg/m^2]

BMI was defined using values linked to patient BMI, height and weight in the Additional File. Where the BMI was not explicitly stated, it was calculated as the patient's weight in kg divided by their height in metres squared.

Human height has been shown to be relatively stable past the age of 20[166]. A period of three years was added to this age to allow for patients who may have been late developers. The median recorded height between the date the patient became 23 and the study start date was used to represent patient height. The median was used in preference to the mean as it was more robust to data entry errors. For patient weight, the most recent value from up to two years prior to the study start date was used. BMI was coded as missing if no data on BMI or height and weight were recorded.

Smoking Status [No Units]

Smoking status was defined through the use of medical codes related to smoking in the Clinical File (see Appendix E.1) and numerical values linked to smoking in the Additional File. Patients were classified into 'never smoked', 'ex-smoker' and 'current smoker' categories, as follows:

- **Never Smoked** - Patient records contained no codes indicating that they smoked.

- **Ex-Smoker** - Patient records contained codes more than two years before the study start date to indicate that they smoked, with either no codes recorded in the two years preceding the study start date, or a final code recorded in these two years to indicate smoking cessation. Where cessation therapy was prescribed or monitoring for cessation was performed, the patient was assumed to have ceased smoking. However, where informed dissent to cessation or to cessation monitoring was mentioned, the patient was assumed to have resumed smoking.
- **Current Smoker** - Patient records in the two years preceding the study start date contained codes to indicate that they smoked, with the final code listed being non-indicative of smoking cessation. Without further context, tests quantifying the level of addiction of patients to either nicotine or smoking were assumed to indicate the continued presence of smoking in the patient.

Systolic Blood Pressure [mmHg]

SBP was defined from numerical SBP test results present in the Additional File. The most recent value, from up to two years prior to the study start date, was used to represent patient SBP. SBP was coded as missing if no data was recorded.

Glycated Haemoglobin [mmol/mol]

HbA1c was defined from numerical HbA1c test results present in the Test File. The International Federation of Clinical Chemistry (IFCC) units (mmol/mol) were used for all analyses. Where values were stated in the Diabetes Control and Complications Trial (DCCT) units (%), they were converted using Equation 5.1.

Equation 5.1 - Formula to convert HbA1c values from DCCT to IFCC units.

$$IFCC_{mmol/mol} = (DCCT_{\%} - 2.15) \times 10.929$$

The most recent value, from up to two years prior to the study start date, was used as a representation of HbA1c for each patient. HbA1c was coded as missing if no data was recorded.

Total-to-High-Density-Lipoprotein Cholesterol Ratio [No Units]

Total : HDL cholesterol ratio was defined from numerical test values for HDL cholesterol, total cholesterol or total : HDL cholesterol ratio in the Test File.

Separate total cholesterol and HDL cholesterol values recorded on the same date were combined to form total : HDL cholesterol ratios under the assumption that both measures were taken at the same general practitioner (GP) visit. The most recent value, from up to two years prior to the study start date, was used as a representation of total : HDL cholesterol ratio. Total : HDL cholesterol ratio was coded as missing if no data was recorded.

5.2.4. Outcomes

All outcomes were established using a combination of UK ONS death certificates and CPRD. Where patients in CPRD could be successfully matched to ONS data on National Health Service number, gender and date of birth, or where the dates of death in CPRD and ONS differed by ≤30 days, the ONS date of death was used to define the date of patient mortality. Where patients in CPRD could not be successfully matched to ONS data on National Health Service number, gender and date of birth, and where the dates of death in CPRD and ONS differed by >30 days, patients were censored at the earliest date of mortality in CPRD and ONS.

Cardiovascular Mortality

Death from cardiovascular causes was defined as death with International Classification of Diseases version 10 (ICD-10) codes I10-I79 listed as the “underlying cause” on the death certificate.

Cancer Mortality

Death from neoplastic causes was defined as death with ICD-10 codes C00-C97 listed as the “underlying cause” on the death certificate.

Other Mortality

Other mortality was defined as death from causes other than those listed under cardiovascular or cancer mortality above.

5.2.5. Statistical Analyses

All analyses were carried out using Stata 12.1[91]. Time since cohort entry was used as the underlying time function. Table 5.1 gives an overview the analysis models and outcomes evaluated in this chapter. All multivariable models were adjusted for age, gender, BMI, smoking status, SBP, HbA1c, and Total : HDL cholesterol ratio.

Table 5.1 - Overview of analysis model structures. "Other mortality" refers to all mortality not attributable to either cardiovascular or cancer.

Model ID	Models	Model Type	Analysis Type	Output Metric	Univariable / Multivariable	Encapsulation of Albuminuria	Outcomes
1	Cox Proportional Hazards[4]	CSH	Standard	Hazard Ratio	Univariable and Multivariable	Covariate	CVD Mortality
							Cancer Mortality
2	Unstratified Lunn-McNeil[16]	CSH	Competing Risks	Hazard Ratio	Univariable and Multivariable	Covariate	[1] CVD Mortality [2] Cancer Mortality [3] Other Mortality
3	Fine-Gray[20]	SDH	Competing Risks	Subhazard Ratio	Univariable and Multivariable	Covariate	[1] CVD Mortality [2] Non-CVD Mortality
							[1] Cancer Mortality [2] Non-Cancer Mortality

Primary Analysis

Covariate-adjusted and unadjusted risk estimates for the effect of albuminuria on the cause-specific hazards (CSHs) of cardiovascular and cancer mortality were produced from Cox-PH (Table 5.1, Model ID 2) and unstratified Lunn-McNeil (Table 5.1, Model ID 2) survival models. The correct parameterisation of the (unstratified) Lunn-McNeil model (herein, merely referred to as the "Lunn-McNeil model") necessitates the evaluation of all competing CSHs present in a population. As other mortality was unlikely to share a common hazard distribution with either cardiovascular mortality or

cancer mortality, the decision was made to model it alongside the CSHs of cardiovascular mortality and cancer mortality in the Lunn-McNeil model. Other mortality was not subdivided into further CSHs for the sake of simplicity.

Covariate-adjusted and unadjusted risk estimates for the effect of albuminuria on the subdistribution hazards (SDHs) of cardiovascular and cancer mortality were produced from the Fine-Gray survival model (Table 5.1, Model ID 3). As the Fine-Gray model can only address two outcomes, the competing outcome structure differed from that used in the parameterisation of the Lunn-McNeil model (Table 5.1, Model ID 2), i.e. competing failure was modelled as "all mortality not related to the incident event under investigation" (Table 5.1, Model ID 3).

Statistical Software and Implementation

Cox-PH Model

Cox-PH model risk estimates were produced using a combination of the stset and stcox commands in Stata 12.1[91]. "`stset`" specified which outcome was the primary outcome, and to declared the data as survival/time-to-event for subsequent statistical packages. In the following code example, "`stime`" is a variable containing survival time data and "`event`" is a variable of outcome types (coded "`0`" for censoring, "`1`" for cardiovascular mortality, "`2`" for cancer mortality, and "`3`" for other mortality). "`stcox`" was used subsequent to "`stset`" to implement the Cox-PH model. Sandwich-type variance estimators were used for all relative hazards models (specified via the "`vce`" option) so as to maintain consistency with the Fine-Gray model, which (at the time of writing) was not able to use likelihood-based variance estimators in Stata.

```
stset stime, failure(event == 1)
stcox albuminuria, vce(robust)
```

Lunn-McNeil Model

To perform the Lunn-McNeil model the data was expanded so as to provide one record of each competing outcome for each patient. Three "`type`" variables were created; one for each competing

outcome, equal to "1" where the outcome occurred, and "0" where it did not. Where the patient was censored, all "type" variables were coded as "0". For the Unstratified variant of the Lunn-McNeil model (used in the analyses of this chapter), interaction terms were created for each variable and the competing outcome "type" variables. The data were then "stset" as was performed for the Cox-PH model. The "stcox" command was then used to implement a Cox-PH model, clustered by patient identifier ("patid"), with interaction terms present between each competing outcome type and each adjustment variable.

```
expand 3
bysort patid: generate type = _n
tabulate type, generate(type)
generate status = (type == event)

stset stime, failure(status)
stcox i.albuminuria#i.type, cluster(patid) vce(cluster patid)
```

Fine-Gray Model

The "stset" command was run, as per the previous two models. The Fine-Gray model was performed using the "stcrreg" command. This had a similar syntax to the "stcox" command used for the Cox-PH model. The additional specification of the "compete" option informed the command which outcomes were competing.

```
stset stime, failure(event == 1)
stcrreg albuminuria, compete(event == 2 3) vce(robust)
```

5.2.6. Sensitivity Analyses

Parameterisation of the Lunn-McNeil model

As previously stated, the correct parameterisation of the Lunn-McNeil model necessitates the evaluation of all competing CSHs present in a population. For the sake of simplicity, only three

outcomes were used in the Lunn-McNeil models of primary analysis, i.e. the CSHs of the main two outcomes under investigation and the CSH of all other mortality. However, this parameterisation reduced the comparability of risk estimates to those of the Fine-Gray model, which is limited to the evaluation of two SDHs per model, i.e. cardiovascular mortality and non-cardiovascular mortality in one model, with cancer mortality and non-cancer mortality in a second model.

To determine whether the structuring of the CSHs had any effect on risk estimates, the outcome and model structures of the Lunn-McNeil model were altered to reflect those used in the Fine-Gray models, i.e. two separate Lunn-McNeil models were performed: one with the CSHs of cardiovascular mortality and non-cardiovascular mortality as outcomes, and one with the CSHs of cancer mortality and non-cancer mortality as outcomes.

Time Interactions

A common method for the evaluation of the proportionality assumption of survival analysis involves the plotting of scaled Schoenfeld-type residuals[167,168] (analogous to ordinary least squares residuals in linear regression) for each variable against survival time. A locally weighted scatterplot smoother is then added to each plot. Violations of the proportionality assumption, defined as significant deviations of the smoother from the line $y = 0$, imply that the variable demonstrates an interaction with a function of time. The methods to evaluate the proportionality assumption using Schoenfeld residuals were initially developed to evaluate the proportional CSHs assumption[167]. However, these have since been adapted to evaluate the proportional SDHs assumption using Schoenfeld-like residuals[168]. From here on both Schoenfeld residuals and Schoenfeld-like residuals will be referred to non-specifically as "Schoenfeld-type residuals".

Where violations of the proportionality assumption are demonstrated in Schoenfeld-type residual plots, common practice is to include time-interactions with the affected variables in order to remove the relationship between the variable and time (and hence, the violation of the proportionality assumption). However, as proportional CSHs often precludes proportional SDHs (and vice versa)[22], it may be the case that time interactions are required for models based on one hazards distribution

type, but not the other. In these situations, it may be difficult to be certain that differences between model estimates are a result of differential effects of the variable on the CSH and SDH of the outcome, as opposed to differences artificially introduced by the inclusion of time interactions. A further complication to the inclusion of time interactions is that for the variables on which they are used, the magnitude of effect is no longer invariant on time. This makes comparisons between models problematic as differences in effect will only hold true at the time points selected for comparison. Thus, in order to maintain direct comparability between model estimates, time-interactions were not used in the study primary analyses.

Analyses were conducted in order to determine the effect of correcting any violations of the proportionality assumption. The proportionality assumption was evaluated for each variable present in the models with respect to each outcome. For proportional CSHs models, this was assessed using diagnostic plots based on scaled Schoenfeld residuals, developed by Therneau and Grambsch[167]. For proportional SDH models, this was assessed using diagnostic plots based on scaled Schoenfeld-like residuals, developed by Zhou, Fine and Laird[168]. Time interactions were used to remove any time trends demonstrated in model variables. Where time interactions were included in the model adjustment variables the effect of the interactions was evaluated through examining the change in the magnitude of the HR or SHR for albuminuria in multivariable models.

Where time interactions were necessitated for the albuminuria variable itself, the value of the HR or SHR for albuminuria and the interaction term between albuminuria and time were quantified to determine the degree to which the models of the primary analyses may misrepresent the relationships between albuminuria and each outcome.

ICD-10 Recoding

A death certificate lists a single primary cause of death, alongside up to 15 contributing causes. In the primary analysis, only the primary cause was used to define the cause of death. However, ambiguity may arise in cases where cancer contributed to a patient's death from cardiovascular

causes, and vice versa. To assess the potential for this to bias risk estimates, the models of the primary analysis were repeated under the following scenarios:

3. Patient mortality was defined as cardiovascular disease where the disease was listed as either the underlying cause or a contributory cause on the death certificate (even where cancer was stated as the underlying cause). In this scenario patient mortality was only classified as cancer mortality or other mortality if one of these was stated as the underlying cause of mortality, and cardiovascular mortality was not listed as a contributory cause or mortality.
4. Patient mortality was defined as cancer where the disease was listed as either the underlying cause or a contributory cause on the death certificate (even where cardiovascular disease was stated as the underlying cause). In this scenario patient mortality was only classified as cardiovascular mortality or other mortality if one of these was stated as the underlying cause of mortality, and cancer mortality was not listed as a contributory cause or mortality.

Albuminuria Variable Coding

As previously stated in Chapter 4, the coding of the albuminuria variable used a hierarchical evidence structure to define each patient's albuminuria level. However, Chapter 4 also described disagreement between the three CPRD evidence sources used to define this variable. To determine the effect of planning a study purely based around one of these sources, the models in the study primary analysis were repeated using albuminuria variable levels defined using each of the following sources:

1. **Medical codes** - linked to diagnostic codes for albuminuria, present in the Clinical File.
2. **Numerical test results** - values for ACRs/PCRs, and 24-hr AERs and PERs, present in the Test File.
3. **Categorical test results** - categorical identifiers of albuminuria, present as "test qualifier units" associated with medical codes for albuminuria in the Test File.

Missing Data

Missing data may be liable to produce biased risk estimates unless data are 'missing completely at random'. In the study primary analysis, a significant proportion of patients had missing albuminuria data. One potential cause of this was that GPs failed to record patient albuminuria values in the absence of clinically significant albuminuria. In this scenario, the likelihood of albuminuria data being missing would be linked to unobserved variables, i.e. albuminuria data would be 'missing not at random'. Therefore, to evaluate the potential scenario that albuminuria values tended not to be recorded for patients with normoalbuminuria, the study primary analysis was re-run will all missing albuminuria data imputed as normoalbuminuria.

5.3. Results

5.3.1. Baseline Variables

The baseline characteristics of the study participants have been described in the Results section of Chapter 4. Briefly, of the 86,962 patients with T2DM eligible to participate at baseline, 54,801 had albuminuria data recorded. After nine years of follow-up 14,201 of these patients died (5,574 from cardiovascular causes, 3,455 from cancer and 5,172 from other causes). Further detail pertaining to the entire study cohort is presented in Table 4.2. Distributions of the continuous variables presented in Table 4.2 may be found in Appendix E.2.

Table 5.2 - Baseline characteristics of study cohort, stratified by outcome. Data represent means ± SD unless a percentage (%) is stated.

Baseline Variable	Variable Data Completeness [n, (%)]	All (n = 86,962)	Death from CVD (n = 8,800)	Death from Cancer (n = 5,239)
Men	86,962 (100%)	47,011 (54.1%)	4,879 (55.4%)	3,176 (60.62%)
Age (years)	86,962 (100%)	66.6 ± 12.6	75.5 ± 34.5	72.6 ± 9.2
BMI (kg m^{-2})	78,253 (90.0%)	29.8 ± 6.1	30.0 ± 6.0	28.9 ± 5.5
SBP (mm Hg)	84,806 (97.5%)	138.5 ± 17.4	140.4 ± 19.6	138.9 ± 16.8
Total : HDL Cholesterol	62,459 (71.8%)	3.9 ± 1.2	3.8 ± 1.2	3.8 ± 1.2
HbA1c (mmol mol^{-1})	74,449 (85.6%)	57.5 ± 16.1	58.0 ± 16.8	55.9 ± 15.0
Never Smoked	86,962 (100%)	25,164 (28.9%)	2,293 (26.0%)	1,073 (20.5%)
Ex-Smokers	86,962 (100%)	41,645 (47.9%)	4,585 (52.1%)	2,672 (51.0%)
Current Smokers	86,962 (100%)	20,153 (23.2%)	1,922 (21.8%)	1,494 (28.5%)
Normoalbuminuria	54,801 (63.0%)	42,668 (77.9%)	3,660 (65.7%)	2,565 (74.2%)
Albuminuria	54,801 (63.0%)	12,133 (22.1%)	1,914 (34.3%)	890 (25.8%)

5.3.2. Primary Analysis

Cardiovascular Mortality

Albuminuria was significantly associated with an increase in the hazard of cardiovascular mortality in both univariable and multivariable Cox-PH models, and in univariable and multivariable Lunn-McNeil models (Columns 1 and 2, Table 5.3). Differences between Cox-PH and Lunn-McNeil model estimates were negligible, if present.

Albuminuria was significantly associated with the cumulative incidence of cardiovascular mortality in univariable and multivariable Fine-Gray models (Column 3, Table 5.3). The SHRs for albuminuria from Fine-Gray models were lower than the HRs of Lunn-McNeil models. Hence, albuminuria's aetiological associations with cardiovascular mortality were stronger than its prognostic associations. However, the 75% increase in risk described by the HR of the multivariable Lunn-McNeil model would have similar clinical implications to the 58% increase in risk described by the SHR of the multivariable Fine-Gray model, were both to be interpreted as interchangeable measures of risk.

Table 5.3 - Estimates for the effect of albuminuria on cardiovascular mortality from Cox-PH, Lunn-McNeil and Fine-Gray Models. *Adjusted for age, gender, BMI, smoking status, SBP, HbA1c and Total : HDL cholesterol ratio. Values for adjustment variables may be found in Appendix E.3.

Univariable / Multivariable	Cox-PH	Lunn-McNeil	Fine-Gray
	HR (95% CI)	**HR (95% CI)**	**SHR (95% CI)**
Univariable	2.13 (2.01 - 2.25)	2.14 (2.02 - 2.26)	1.95 (1.85 - 2.07)
Multivariable*	1.75 (1.63 - 1.87)	1.75 (1.64 - 1.87)	1.58 (1.48 - 1.69)

Cancer Mortality

Albuminuria was significantly associated with an increase in the hazard of cancer mortality in both univariable and multivariable Cox-PH models and univariable and multivariable Lunn-McNeil models (Columns 1 and 2, Table 5.4). Differences between Cox-PH and Lunn-McNeil model estimates were negligible, if present.

Albuminuria was significantly associated with the cumulative incidence of cancer mortality in univariable and multivariable Fine-Gray models (Column 3, Table 5.4). The SHRs for albuminuria from Fine-Gray models were lower than the HRs of Lunn-McNeil models. Hence, albuminuria's aetiological associations with cancer mortality were stronger than its prognostic associations. However, the 28% increase in risk described by the HR of the multivariable Lunn-McNeil model would have much larger clinical implications than the 11% increase in risk described by the SHR of the multivariable Fine-Gray model, were both to be interpreted as interchangeable measures of risk.

Table 5.4 - Estimates for the effect of albuminuria on cancer mortality from Cox-PH, Lunn-McNeil and Fine-Gray Models. *Adjusted for age, gender, BMI, smoking status, SBP, HbA1c and Total : HDL cholesterol ratio. Values for adjustment variables may be found in Appendix E.3.

Univariable / Multivariable	Cox-PH	Lunn-McNeil	Fine-Gray
	HR (95% CI)	HR (95% CI)	SHR (95% CI)
Univariable	1.42 (1.31 - 1.53)	1.42 (1.32 - 1.53)	1.25 (1.16 - 1.35)
Multivariable*	1.27 (1.16 - 1.39)	1.28 (1.17 - 1.40)	1.11 (1.01 - 1.21)

5.3.3. Sensitivity Analyses

Parameterisation of the Lunn-McNeil Model

Restructuring the Lunn-McNeil model to have only two outcomes had minimal effect on the estimates for the effect of albuminuria on cardiovascular (Table 5.5) and cancer (Table 5.6) mortality in univariable and multivariable models.

Table 5.5 - Estimates for the effect of albuminuria on cardiovascular mortality from two-outcome- and three-outcome-structure Lunn-McNeil models. *Adjusted for age, gender, BMI, smoking status, SBP, HbA1c and Total : HDL cholesterol ratio.

Model	Univariable / Multivariable	Primary Analysis (3 Outcomes)	Alternative Outcome Structure (2 Outcomes)
		HR (95% CI)	HR (95% CI)
Lunn-McNeil	Univariable	2.14 (2.02 - 2.26)	2.14 (2.02 - 2.26)
	Multivariable*	1.75 (1.64 - 1.87)	1.75 (1.64 - 1.87)

Table 5.6 - Estimates for the effect of albuminuria on cancer mortality from two-outcome- and three-outcome-structure Lunn-McNeil models. *Adjusted for age, gender, BMI, smoking status, SBP, HbA1c and Total : HDL cholesterol ratio.

Model	Univariable / Multivariable	Primary Analysis (3 Outcomes)	Alternative Outcome Structure (2 Outcomes)
		HR (95% CI)	HR (95% CI)
Lunn-McNeil	Univariable	1.42 (1.32 - 1.53)	1.42 (1.31 - 1.53)
	Multivariable*	1.28 (1.17 - 1.40)	1.27 (1.16 - 1.39)

Time Interactions

No violations of the proportional CSH assumption were detected in the Schoenfeld-residual plots for cardiovascular mortality, cancer mortality or other mortality (Appendix E.4). Violations in the proportional SDH assumption were detected for the variable of albuminuria for both cardiovascular and cancer mortality (Figure 5.1). In both instances, the addition of interaction terms between albuminuria and time was found to remove the association (Figure 5.1). For all other variables, the proportional SDH assumption was satisfied for cardiovascular and cancer mortality (Appendix E.5).

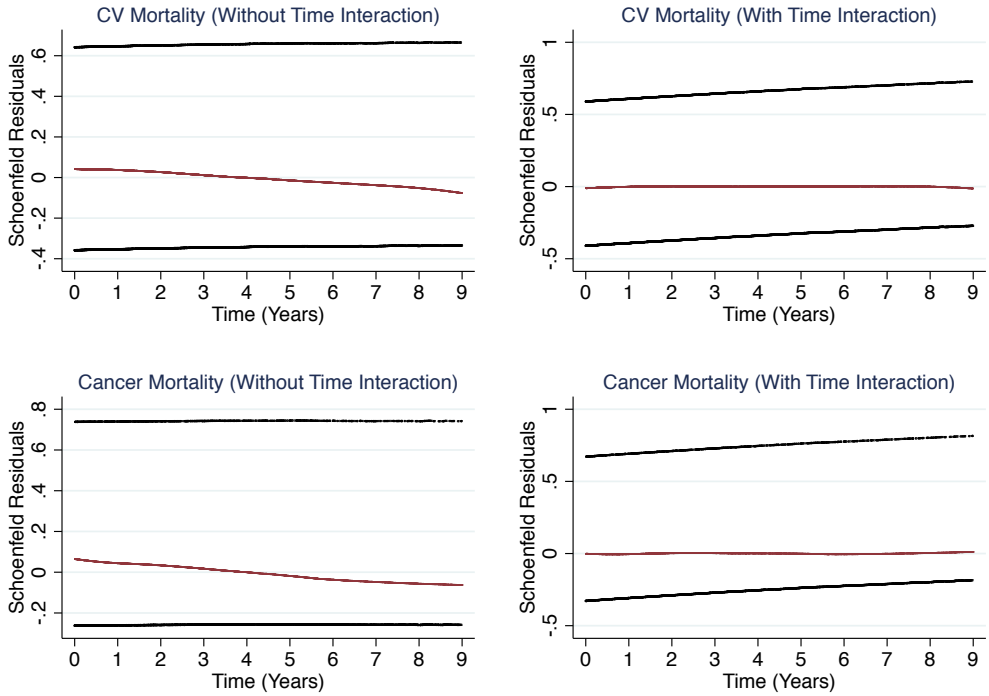


Figure 5.1 - Scaled Schoenfeld-like residual plots of the albuminuria variable for the SDHs of cardiovascular (CV) mortality and cancer mortality.

The inclusion of time interactions for the albuminuria variable had little to no effect on the magnitude of the SHRs of the study adjustment variables for either outcome, as shown by the values in Table 5.7 and Table 5.8.

Table 5.7 - Fine-Gray model estimates for the effect of albuminuria on the subdistribution hazard of cardiovascular mortality from the primary analysis and after the inclusion of time interactions.

Variable	Primary Analysis	Albuminuria Time Interaction Analysis
	SHR (95% CI)	**SHR (95% CI)**
Albuminuria (vs normoalbuminuria)	1.58 (1.48 - 1.69)	1.95 (1.73 - 2.20)
Albuminuria × Time (Years)	.	0.95 (0.92 - 0.97)
Male Gender (vs Female Gender)	1.30 (1.22 - 1.39)	1.30 (1.22 - 1.39)
Age (per Year)	1.09 (1.09 - 1.10)	1.09 (1.09 - 1.10)
BMI (per km/m²)	1.02 (1.01 - 1.02)	1.02 (1.01 - 1.02)
Total : HDL Cholesterol	1.05 (1.02 - 1.08)	1.05 (1.02 - 1.08)
HbA1c (per mmol/mol)	1.01 (1.01 - 1.01)	1.01 (1.01 - 1.01)
SBP (per mmHg)	1.00 (1.00 - 1.00)	1.00 (1.00 - 1.00)
Ex-Smoker (vs Never Smoked)	1.16 (1.07 - 1.26)	1.16 (1.07 - 1.26)
Current Smoker (vs Never Smoked)	1.46 (1.32 - 1.61)	1.46 (1.32 - 1.61)

Table 5.8 - Fine-Gray model estimates for the effect of albuminuria on the subdistribution hazard of cancer mortality from the primary analysis and after the inclusion of time interactions.

Variable	Primary Analysis	Albuminuria Time Interaction Analysis
	SHR (95% CI)	**SHR (95% CI)**
Albuminuria (vs normoalbuminuria)	1.11 (1.01 - 1.21)	1.50 (1.28 - 1.77)
Albuminuria × Time (Years)	.	0.93 (0.89 - 0.96)
Male Gender (vs Female Gender)	1.40 (1.29 - 1.52)	1.40 (1.29 - 1.52)
Age (per Year)	1.05 (1.05 - 1.06)	1.05 (1.05 - 1.06)
BMI (per kg/m²)	1.01 (1.00 - 1.02)	1.01 (1.00 - 1.02)
Total : HDL Cholesterol	1.01 (0.97 - 1.05)	1.01 (0.97 - 1.05)
HbA1c (per mmol/mol)	1.00 (1.00 - 1.00)	1.00 (1.00 - 1.00)
SBP (per mmHg)	1.00 (1.00 - 1.00)	1.00 (1.00 - 1.00)
Ex-Smoker (vs Never Smoked)	1.40 (1.25 - 1.56)	1.40 (1.25 - 1.56)
Current Smoker (vs Never Smoked)	2.08 (1.85 - 2.36)	2.08 (1.85 - 2.36)

The presence of time-interactions for albuminuria meant that the magnitude of effect of albuminuria was no longer invariant on time. Where time interactions were present in the models they demonstrated a decrease in the magnitude of the ln(SHR) for the effect of albuminuria on cardiovascular mortality of 0.058/year in univariable analysis and 0.052/year in multivariable analysis. For cancer mortality, there was a decrease in the magnitude of the ln(SHR) for the effect of albuminuria of 0.083/year and 0.075/year in univariable and multivariable analysis, respectively. The range of values the SHR for the effect of albuminuria could take for both outcomes within the study observation period is shown in Figure 5.2. At the beginning of follow-up (0 years), the presence of time interactions resulted in SHRs in excess of the HRs of the Cox-PH and Lunn-McNeil models of the primary analysis, while at the end of follow-up (9 years) the presence of time interactions resulted in SHRs that were substantially lower than those of the Fine-Gray models of the primary analysis.

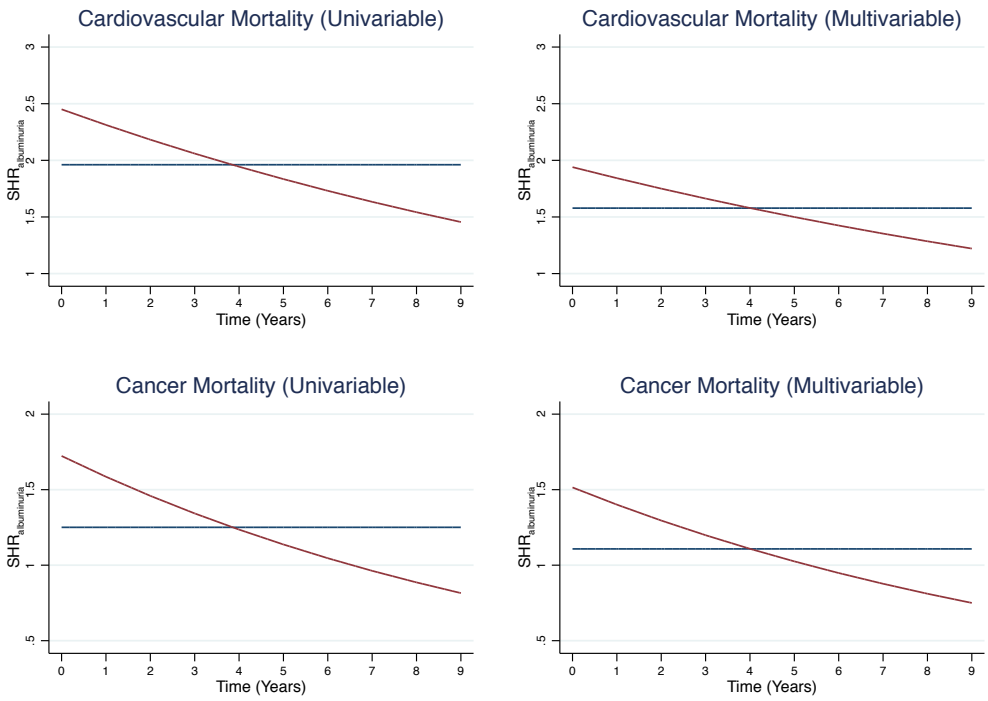


Figure 5.2 - The magnitude of effect of albuminuria during study follow from the Fine-Gray models of the primary analysis (blue lines) and from Fine-Gray models incorporating time-interactions with albuminuria (red lines).

ICD-10 Recoding

The effects of reclassifying mortality were more pronounced for estimates on cancer mortality than for cardiovascular mortality. The number of patients whose cause of mortality was reclassified is shown in Table 4.7. This table has been reproduced from the Results section of Chapter 4 as a point of reference, and will not be further discussed in this chapter.

Table 5.9 - The number of patients experiencing each outcome in the primary analysis, and after the reclassification of mortality according to the presence of causes (cardiovascular or cancer) listed as being contributory.

Outcome	Primary Analysis	Reclassification as Cardiovascular Mortality	Reclassification as Cancer Mortality
Cardiovascular Mortality	5,574 (39.3%)	8,572 (60.4%)	5,299 (37.3%)
Cancer Mortality	3,455 (24.3%)	2,787 (19.6%)	3,998 (28.2%)
Other Mortality	5,172 (36.4%)	2,842 (20.0%)	4,904 (34.5%)
Total Mortality	14,201 (100%)	14,201 (100%)	14,201 (100%)

Cardiovascular Mortality

Table 5.10 shows the effect of reclassifying mortality on risk estimates for the effect of albuminuria on cardiovascular mortality. Reclassifying mortality as cardiovascular increased risk estimates for the effect of albuminuria on cardiovascular mortality across all models. Conversely, reclassifying mortality as cancer decreased risk estimates for the effect of albuminuria on cardiovascular mortality across models. The effects of reclassifying mortality were similar across all models, implying the similar effects on the CSHs and SDHs of cancer mortality. The reclassification of mortality as either cardiovascular or cancer did not affect the relationship between model estimates, with Cox-PH estimates being roughly equal to Lunn-McNeil estimates, and Fine-Gray estimates being lower.

Table 5.10 - Model estimates for the effect of albuminuria on cardiovascular mortality; from the primary analysis, and after the reclassification of mortality according to those listed as contributory causes. *Adjusted for age, gender, BMI, smoking status, SBP, HbA1c and Total : HDL cholesterol ratio.

Model	Univariable / Multivariable	Primary Analysis	Reclassification as Cardiovascular Mortality	Reclassification as Cancer Mortality
		HR / SHR (95% CI)	HR / SHR (95% CI)	HR / SHR (95% CI)
Cox-PH	Univariable	2.13 (2.01 - 2.25)	2.17 (2.08 - 2.27)	2.12 (2.00 - 2.24)
	Multivariable*	1.75 (1.63 - 1.87)	1.77 (1.68 - 1.87)	1.73 (1.62 - 1.86)
Lunn-McNeil	Univariable	2.14 (2.02 - 2.26)	2.17 (2.08 - 2.27)	2.12 (2.01 - 2.25)
	Multivariable*	1.75 (1.64 - 1.87)	1.77 (1.68 - 1.87)	1.74 (1.62 - 1.86)
Fine-Gray	Univariable	1.95 (1.85 - 2.07)	2.07 (1.98 - 2.16)	1.94 (1.83 - 2.05)
	Multivariable*	1.58 (1.48 - 1.69)	1.66 (1.57 - 1.75)	1.56 (1.45 - 1.67)

Cancer Mortality

Table 5.11 shows the effect of reclassifying mortality on risk estimates for the effect of albuminuria on cancer mortality. Reclassifying mortality as cardiovascular decreased risk estimates for the effect of albuminuria on cancer mortality across all models. Conversely, reclassifying mortality as cancer increased risk estimates for the effect of albuminuria on cancer mortality across models. The effects of reclassifying mortality were similar across all models, implying the similar effects on the CSHs and SDHs of cancer mortality. Reclassification of mortality as cardiovascular caused the effect of

albuminuria on the SDH of cancer mortality to become non-significant. The reclassification of mortality as either cardiovascular or cancer did not affect the relationship between model estimates, with Cox-PH estimates being roughly equal to Lunn-McNeil estimates, and Fine-Gray estimates being lower.

Table 5.11 - Model estimates for the effect of albuminuria on cancer mortality; from the primary analysis, and after the reclassification of mortality according to those listed as contributory causes. *Adjusted for age, gender, BMI, smoking status, SBP, HbA1c and Total : HDL cholesterol ratio.

Model	Univariable / Multivariable	Primary Analysis	Reclassification as Cardiovascular Mortality	Reclassification as Cancer Mortality
		HR / SHR (95% CI)	HR / SHR (95% CI)	HR / SHR (95% CI)
Cox-PH	Univariable	1.42 (1.31 - 1.53)	1.34 (1.23 - 1.47)	1.51 (1.41 - 1.62)
	Multivariable*	1.27 (1.16 - 1.39)	1.23 (1.11 - 1.36)	1.35 (1.24 - 1.47)
Lunn-McNeil	Univariable	1.42 (1.32 - 1.53)	1.35 (1.24 - 1.47)	1.51 (1.41 - 1.62)
	Multivariable*	1.28 (1.17 - 1.40)	1.24 (1.12 - 1.37)	1.36 (1.25 - 1.47)
Fine-Gray	Univariable	1.25 (1.16 - 1.35)	1.19 (1.09 - 1.29)	1.34 (1.25 - 1.43)
	Multivariable*	1.11 (1.01 - 1.21)	1.07 (0.97 - 1.18)	1.19 (1.09 - 1.29)

Albuminuria Variable Coding

The number and proportion of patients for whom it was possible to assign an albuminuria level using different sources of CPRD is presented in Table 4.11. This table has been reproduced from the Results section of Chapter 4 as a point of reference, and will not be further discussed in this chapter.

Table 5.12 - Patients classifiable using different CPRD data sources for albuminuria.

CPRD Source	Albuminuria Level, n (%)		
	Normoalbuminuria	Albuminuria	Total
Medical Codes Only	1,102 (18.4%)	4,873 (81.6%)	5,975 (100%)
Numerical Test Values Only	21,330 (74.0%)	7,498 (26.0%)	28,828 (100%)
Categorical Test Values Only	36,227 (91.9%)	3,183 (8.1%)	39,410 (100%)
All (Hierarchical Structure)	42,666 (77.9%)	12,135 (22.1%)	54,801 (100%)

Cardiovascular Mortality

In univariable analysis, the use of numerical test values, resulted in the highest cardiovascular mortality risk estimates, followed by medical codes, then categorical test values (Table 5.13). However, in multivariable analysis, the use of categorical test values resulted in the highest cardiovascular mortality risk estimates, followed by medical codes, then numerical test values. Comparisons between model estimates were less effected by the recoding of the albuminuria variable with Cox-PH estimates being roughly equal to Lunn-McNeil estimates, and Fine-Gray estimates being lower, for all albuminuria variable structures.

Table 5.13 - Model estimates for the effect of albuminuria on cardiovascular mortality using: the hierarchical variable structure from the primary analysis, medical diagnostic codes only, numerical test values only, and categorical test values only. *Adjusted for age, gender, BMI, smoking status, SBP, HbA1c and Total : HDL cholesterol ratio.

Model	Univariable / Multivariable	Primary Analysis	Medical Codes	Numerical Test Values	Categorical Test Values
		HR / SHR (95% CI)	HR / SHR (95% CI)	HR / SHR (95% CI)	HR / SHR (95% CI)
Cox-PH	Univariable	2.13 (2.01 - 2.25)	2.29 (1.83 - 2.88)	2.56 (2.38 - 2.76)	2.06 (1.87 - 2.25)
	Multivariable*	1.75 (1.63 - 1.87)	1.95 (1.46 - 2.61)	1.86 (1.70 - 2.03)	2.02 (1.81 - 2.26)
Lunn-McNeil	Univariable	2.14 (2.02 - 2.26)	2.29 (1.83 - 2.88)	2.57 (2.39 - 2.77)	2.07 (1.88 - 2.27)
	Multivariable*	1.75 (1.64 - 1.87)	1.95 (1.46 - 2.60)	1.86 (1.70 - 2.03)	2.04 (1.82 - 2.28)
Fine-Gray	Univariable	1.95 (1.85 - 2.07)	2.18 (1.73 - 2.73)	2.31 (2.14 - 2.48)	1.87 (1.70 - 2.05)
	Multivariable*	1.58 (1.48 - 1.69)	1.86 (1.39 - 2.48)	1.65 (1.51 - 1.81)	1.74 (1.55 - 1.94)

Cancer Mortality

In univariable analysis, the use of numerical test values, resulted in the highest cancer mortality risk estimates, followed by medical codes, then categorical test values (Table 5.14). However, in multivariable analysis, there was no consistent pattern between models. The use of medical codes alone resulted in the loss of statistical significance in risk estimates for cancer mortality from all multivariable models and the univariable Fine-Gray model. Comparisons between model estimates were less effected by the recoding of the albuminuria variable with Cox-PH estimates being roughly

equal to Lunn-McNeil estimates, and Fine-Gray estimates being lower, for all albuminuria variable structures.

Table 5.14 - Model estimates for the effect of albuminuria on cancer mortality using: the hierarchical variable structure from the primary analysis, medical diagnostic codes only, numerical test values only, and categorical test values only. *Adjusted for age, gender, BMI, smoking status, SBP, HbA1c and Total : HDL cholesterol ratio.

Model	Univariable / Multivariable	Primary Analysis	Medical Codes	Numerical Test Values	Categorical Test Values
		HR / SHR (95% CI)	HR / SHR (95% CI)	HR / SHR (95% CI)	HR / SHR (95% CI)
Cox-PH	Univariable	1.42 (1.31 - 1.53)	1.42 (1.08 - 1.86)	1.57 (1.42 - 1.73)	1.41 (1.23 - 1.61)
	Multivariable*	1.27 (1.16 - 1.39)	1.35 (0.94 - 1.94)	1.39 (1.24 - 1.56)	1.28 (1.09 - 1.50)
Lunn-McNeil	Univariable	1.42 (1.32 - 1.53)	1.42 (1.09 - 1.86)	1.58 (1.43 - 1.74)	1.41 (1.23 - 1.61)
	Multivariable*	1.28 (1.17 - 1.40)	1.36 (0.95 - 1.96)	1.40 (1.25 - 1.57)	1.29 (1.10 - 1.51)
Fine-Gray	Univariable	1.25 (1.16 - 1.35)	1.30 (0.99 - 1.70)	1.35 (1.23 - 1.49)	1.22 (1.06 - 1.39)
	Multivariable*	1.11 (1.01 - 1.21)	1.25 (0.87 - 1.80)	1.21 (1.07 - 1.35)	1.05 (0.90 - 1.23)

Missing Data

The recoding of all missing data as normoalbuminuria resulted in risk estimates which were consistently lower than those derived in the primary analysis for both outcomes, across all models (Table 5.15 and Table 5.16). The recoding of missing data did not affect the relationship between models estimates demonstrated by the primary analysis, i.e. estimates from the Cox-PH model were roughly equal to those of the Lunn-McNeil model, while Fine-Gray model estimates were lower.

Table 5.15 - Model estimates for the effect of albuminuria on cardiovascular mortality after recoding missing albuminuria values as normoalbuminuria. *Adjusted for age, gender, BMI, smoking status, SBP, HbA1c and Total : HDL cholesterol ratio.

Model	Univariable / Multivariable	Primary Analysis	Missing Data Analysis
		HR / SHR (95% CI)	HR / SHR (95% CI)
Cox-PH	Univariable	2.13 (2.01 - 2.25)	1.49 (1.43 - 1.56)
	Multivariable*	1.75 (1.63 - 1.87)	1.48 (1.40 - 1.56)
Lunn-McNeil	Univariable	2.14 (2.02 - 2.26)	1.49 (1.43 - 1.56)
	Multivariable*	1.75 (1.64 - 1.87)	1.49 (1.41 - 1.57)
Fine-Gray	Univariable	1.95 (1.85 - 2.07)	1.43 (1.37 - 1.49)
	Multivariable*	1.58 (1.48 - 1.69)	1.38 (1.30 - 1.45)

Table 5.16 - Model estimates for the effect of albuminuria on cancer mortality after recoding missing albuminuria values as normoalbuminuria. *Adjusted for age, gender, BMI, smoking status, SBP, HbA1c and Total : HDL cholesterol ratio.

Model	Univariable / Multivariable	Primary Analysis	Missing Data Analysis
		HR / SHR (95% CI)	HR / SHR (95% CI)
Cox-PH	Univariable	1.42 (1.31 - 1.53)	1.11 (1.05 - 1.17)
	Multivariable*	1.27 (1.16 - 1.39)	1.16 (1.09 - 1.25)
Lunn-McNeil	Univariable	1.42 (1.32 - 1.53)	1.11 (1.05 - 1.17)
	Multivariable*	1.28 (1.17 - 1.40)	1.17 (1.09 - 1.25)
Fine-Gray	Univariable	1.25 (1.16 - 1.35)	1.04 (0.98 - 1.10)
	Multivariable*	1.11 (1.01 - 1.21)	1.06 (0.99 - 1.14)

5.4. Discussion

5.4.1. Key Results

Albuminuria was significantly associated with the CSHs and SDHs of cardiovascular mortality and cancer mortality, in univariable and multivariable analyses. This suggested both aetiological and predictive associations between albuminuria and both outcomes.

For both outcomes, Cox-PH and Lunn-McNeil CSHs models produced near identical HRs for the effect of albuminuria. Thus, the incorporation of competing risks methodology had minimal impact on CSH risk estimates. For competing risk models, differences between the HRs of the Lunn-McNeil model and the SHRs Fine-Gray model were more pronounced than those demonstrated between the HRs of the Cox-PH and Lunn-McNeil models; with the Fine-Gray model yielding consistently lower risk estimates.

The Lunn-McNeil model and Fine-Gray models produced HRs and SHRs for the effect of albuminuria on cardiovascular mortality that would be have similar clinical implications, i.e. a 75% or 58% increase in cardiovascular mortality risk would be viewed similarly as a measure of either aetiological or prognostic association. However, clinical implications may differ in the case of cancer mortality, as the 28% increase in aetiological risk described by the Lunn-McNeil model is roughly 2.5 times the size of the 11% increase in prognostic risk described by the Fine-Gray model. Furthermore, the p-value for the latter is approaching statistical non-significance ($p = 0.026$).

5.4.2. Strengths and Limitations

Strengths and Limitations of the Dataset

Several adjustment variables in the multivariable models were poorly recorded in CPRD, with incorrectly specified units, or data entry errors, making it difficult to accurately define them. To counter this, GPs (HA, AF, and DM) were consulted regarding the biological plausibility of the ranges used for data cleaning, and individual data points were transformed where errors appeared to be

the result of unit misspecification, i.e. where secondary distributions of data existed. This allowed a greater confidence to be placed in the accuracy of the data, despite its recording quality.

It was not possible to adjust for the duration of diabetes in any of the multivariable models, as this quantity is not readily estimable in CPRD. This may have biased study analyses as duration of diabetes could be a potential confounder for the relationship between albuminuria and either outcome (i.e. duration of diabetes could be associated with the risk of developing albuminuria, and the risks of cardiovascular and cancer mortality).

Digit preference was exhibited by the distribution of the SBP study variable. It is not possible to remove or transform out this artefact from data. Moreover, the potential bias caused by digit preference remains largely unexplored in the literature. As time progresses, more and more GP practices will obtain blood pressure monitors that automatically transfer measurements to the practice computer system; removing the potential for this bias to occur. Thus, using more recent study cohorts may minimise any potential bias as more practices will be likely to have these automated devices.

Missing data were present to varying degrees (0.0% - 37.0%) across several study variables, with the albuminuria variable exhibiting the largest amount of missing data. As a result of the distribution of missing data across variables, complete case analyses could only be conducted on 41,408 (47.6%) patients from the 86,962 initial study population in multivariable analyses. The large quantities of missing data present in the dataset meant that all analyses may be susceptible to bias where data were either 'missing at random' or 'missing not at random'.

Strengths and Limitations of the Analyses

The findings of the primary analyses were robust to most sensitivity analyses performed, i.e. the use of alternative outcome structures in the Lunn-McNeil model, the reclassification of mortality, the use of different components of CPRD to define albuminuria, and the imputation of missing albuminuria data as normoalbuminuria. In these analyses, where HR or SHR values were not the same as those

described by the primary analyses, the patterns of association described between the models remained unchanged. Sensitivity analyses, in which the effect of albuminuria on the SDH of cancer mortality became non-significant, may largely be attributed to loss in statistical power resulting from a decrease in the analysable population. This occurred in the reclassification of mortality as cardiovascular and in the use of medical codes alone to define patient albuminuria.

The findings of the missing data sensitivity analyses implied that the data may not have been 'missing not at random' as a result of albuminuria values failing to be recorded in patients with normoalbuminuria. The failure to identify the mechanism of missingness for albuminuria meant that bias may have been present in risk estimates if data were either 'missing at random', or 'missing not at random' as a result of other unobserved variables. The effect this bias may have had on the risk estimates of each model is unknown, and poses questions for the reliability of the estimates obtained.

Graphical methods using Schoenfeld-type residuals constitute the best method to detect violations of the proportionality assumption, as unlike statistical tests, they are able to detect the presence of non-linear violations. However, the use of these methods added a large element of subjectivity because it was unclear what constituted a significant enough deviation from the line $y = 0$ to be classified as a violation of the proportionality assumption. Although no non-linear violations of the proportionality assumption were detected, the modelling of some variables as having a linear effect on the CSH or SDH of each outcome could be considered overly simplistic. 'U-shaped' relationships between BMI or SBP and cardiovascular or cancer mortality have been described in the literature[169,170]. The use of linear approximations may have imparted unknown bias on model risk estimates. More accurate modelling of the relationship between risk factors and study outcomes using reticulated cubic splines or fractional polynomials may help to mitigate this bias in future studies.

The use of time-interactions to correctly specify the relationship between albuminuria and the SDHs of cardiovascular and cancer mortality in the Fine-Gray model, (demonstrated in sensitivity analyses), meant that it was difficult to compare model estimates between CSH and SDH competing risk models. In the study primary analysis, where proportionality assumption violations were ignored,

SDH risk estimates were consistently lower than CSH ones for the effect of albuminuria. However, once time interactions were incorporated into the Fine-Gray models, the effect of albuminuria on the SDH of each outcome was no longer invariant on time. This resulted in estimates for the effect of albuminuria on the SDHs of both cardiovascular and cancer mortality that could be greater or smaller than those on the CSHs of the same outcomes, depending on the time point at which the SDH effect was estimated (Figure 5.2). However, the proportionality assumption is not always assessed when making comparisons between CSH and SDH models[17], as proportional hazards regression models, like Cox-PH and Fine-Gray, can still provide accurate summary analysis when such violations are present. In such circumstances, model outputs may be interpreted as time-averaged HRs or SHRs[22].

5.4.3. Implications

From a clinical perspective, the findings of this chapter suggest that albuminuria is aetiologically linked to both cardiovascular and cancer mortality in patients with T2DM, with stronger association demonstrated for cardiovascular mortality. Effect sizes described for cardiovascular mortality were similar to the published literature. However, effect sizes for cancer mortality have been less well described and warrant further research (Chapter 2).

The findings of this chapter suggest that the use of methods adjusted for competing risks had little to no impact on estimates for the effect of albuminuria on the CSHs of cardiovascular and cancer mortality in patients with T2DM. Comparisons between CSH and SDH competing risk model estimates suggest that the aetiological and prognostic associations are not interchangeable risk measures, and that model selection should dependent upon the nature of the research question posed, i.e. CSH models should be used to answer research questions in which the measurement of causal associations is desired, whereas SDH models should be used to answer research questions in which the measurement of predictive associations is desired.

The Lunn-McNeil model required additional parameterisation and data management than the Cox-PH and Fine-Gray model. Furthermore, it was also less flexible than either, requiring all present

competing CSHs be proportional to a joint baseline hazard distribution, and thus to each other. Therefore, implementation of the Lunn-McNeil model may prove to be problematic for some studies.

6. Discussion

6.1. Key Findings

Reviews conducted in Chapters 1 and 2 determined an absence of studies applying competing-risks-adjusted methods to the evaluation of the effect of albuminuria on cardiovascular and cancer mortality. Chapters 4 and 5 used the effect of albuminuria on cardiovascular and cancer mortality as a case study, through which to evaluate the effects of incorporating competing risks methodology on measures of absolute and relative risk. Throughout this Thesis, the 'cumulative risk' has been considered to be an absolute risk measure, while the 'cause-specific hazard ratio' (HR_{CS}) and the 'subdistribution hazard ratio' (HR_{SD}) have been considered to be relative risk measures.

In modern research, cumulative risk estimators are used less frequently than relative hazards models, and are generally only used to provide only summary analysis (Chapter 1). However, the findings of Chapter 4 indicate that substantial relative differences between standard- and competing-risks-adjusted cumulative risk estimators may be present for the outcomes of cardiovascular and cancer mortality in patients with type 2 diabetes mellitus (T2DM) (Table 6.1). For a given outcome, absolute and relative differences were greater for patients with albuminuria than normoalbuminuria. For a given albuminuria level, absolute differences were greater for cardiovascular mortality, while relative differences were greater for cancer mortality.

Table 6.1 - Absolute and relative differences between Kaplan-Meier and CICR estimators for the nine-year cumulative risks of cardiovascular and cancer mortality, from the results of Chapter 4.

Albuminuria Level	Cardiovascular Mortality		Cancer Mortality	
	Absolute Difference (%)	Relative Difference (%)	Absolute Difference (%)	Relative Difference (%)
Normoalbuminuria	0.9	8.8	0.8	11.1
Albuminuria	3.3	17.8	2.1	24.4

Risk estimates for the effect of albuminuria were near identical between standard and competing risks cause-specific hazards (CSHs) models for each outcome (Table 6.2). Thus, the incorporation of competing risks methodology had negligible impact on CSH measures of relative hazard in patients with T2DM. Differences were present between the estimates of CSH and SDH competing risks models in which the Fine-Gray model yielded consistently lower risk estimates than the Lunn-McNeil model. For the effect of albuminuria on cardiovascular mortality, these model risk estimates would have similar clinical implications, with HRs and SHRs representing 75% to 58% increases in risk (in multivariable models). Hence, measures of aetiological and prognostic association (in the sense defined in Chapter 1) would be largely interchangeable for this outcome. For the effect of albuminuria on cancer mortality, the 28% increase in aetiological risk described by the HR of the Lunn-McNeil model was roughly 2.5 times the magnitude of the 11% increase in prognostic risk described by the SHR of the Fine-Gray model. Furthermore, the p-value for prognostic association was approaching statistical non-significance ($p = 0.026$). The clinical implications of these risk measures would be very different. Hence, measures of aetiological and prognostic association would be largely non-interchangeable for this outcome.

Table 6.2 - Relative hazards model risk estimates for the effect of albuminuria on cardiovascular and cancer mortality, from the results of Chapter 5. *Adjusted for age, gender, smoking status, SBP, HbA1c and Total : HDL cholesterol ratio.

Model	Univariable / Multivariable	Cardiovascular Mortality	Cancer Mortality
		HR / SHR (95% CI)	HR / SHR (95% CI)
Cox-PH	Univariable	2.13 (2.01 - 2.25)	1.42 (1.31 - 1.53)
	Multivariable*	1.75 (1.63 - 1.87)	1.27 (1.16 - 1.39)
Lunn-McNeil	Univariable	2.14 (2.02 - 2.26)	1.42 (1.32 - 1.53)
	Multivariable*	1.75 (1.64 - 1.87)	1.28 (1.17 - 1.40)
Fine-Gray	Univariable	1.95 (1.85 - 2.07)	1.25 (1.16 - 1.35)
	Multivariable*	1.58 (1.48 - 1.69)	1.11 (1.01 - 1.21)

Point estimates for the effect of albuminuria on the CSH and SDH of cardiovascular mortality fell within the 95% confidence interval (CI) of the one described by the systematic review in Chapter 2. Consequently, these point estimates also fell within the wider 95% prediction interval for the effect of albuminuria on cardiovascular mortality. The 95% prediction interval gives the range that new studies (that would be eligible for inclusion in the review) would be projected to fall within on 95% of occasions. Literature searches identified only one study in which the effect of albuminuria on cancer mortality was assessed. As three or more studies are required to estimate the 95% prediction interval, this was not available as a point of comparison for the results in Chapter 5. Point estimates for the effect of albuminuria on the CSH and SDH of cancer mortality fell within the 95% CI of the estimate from the literature, however, the 95% CI from the study in the literature was wide.

In Chapter 4, significant differences were found between the survival functions of patients with and without albuminuria for cardiovascular and cancer mortality. Furthermore, all relative hazards models in Chapter 5 identified significant associations between the presence of albuminuria and the CSHs and SDHs of both outcomes. These findings suggest that albuminuria is an important factor in determining the risk of both cardiovascular and cancer mortality in patients with T2DM.

6.2. Strengths and Limitations

The systematic review in Chapter 2 was rigorously performed, ensuring a solid evidence base as a point of comparison for the survival model estimates in Chapter 5, whilst also providing the first literature-derived estimates for the effect of albuminuria on cardiovascular mortality in patients with T2DM.

The study in Chapter 3 was guided by the literature for the methods used to define T2DM in Clinical Practice Research Datalink (CPRD) data. This allowed confidence to be placed in the reliability of the methods used to classify T2DM in medical records, and subsequently also in establishing a population with T2DM for the analyses in Chapters 4 and 5.

The primary analyses carried out in Chapters 4 and 5 were protocol-driven research, approved by the Independent Scientific Advisory Committee (ISAC); protocol 15_011A (Appendix D.1). Thus, all findings stated as primary outcomes may be verified as being the primary purpose of the research, and not the result of spurious incidental associations or data mining.

All research in Chapters 3-5 was conducted using substantial CPRD datasets. CPRD is one of the largest primary health care databases in the world, with 75% of all UK practices and 58% of all English practices contributing patient data[128]. There was possibly no better database available in which to answer the research questions posed by this Thesis.

The findings of Chapters 4 and 5 were robust to many of the sensitivity analyses performed, while the effect of albuminuria on the relative hazards of cardiovascular and cancer mortality were consistent with those described in Chapter 2, allowing confidence to be placed in their findings.

The study of the effect of albuminuria on the risks of cancer and cardiovascular mortality in patients with T2DM represents a single case study through which to evaluate the effect of incorporating competing risks methodology. Furthermore, the study of the impact of competing risks methodology was limited to methods identified as being commonly used in the literature, and readily

implementable in statistical packages. Thus, the results described in Chapters 4 and 5 may not mirror the findings in other clinical areas, or from different models. However, the application of competing risks methods to the evaluation albuminuria's effect on cardiovascular and cancer mortality in T2DM was novel, and the use of the most frequently used and readily available survival models added relevance to the research.

In evaluating competing mortality, the studies in Chapters 4 and 5 accounted for the main competing outcomes present in patients with T2DM. However, non-fatal competing outcomes (like receiving a renal or pancreas transplant) were not accounted for as competing risks. This omission was intentional as, non-fatal competing outcomes are generally rare events, and the degree to which these may alter risk estimates has previously been shown to be low[10]. Furthermore, ascertaining the degree to which non-fatal outcomes alter the probability of occurrence of study primary outcomes is far from simple, and would require much clinical input. Thus, to simplify the analyses, only fatal competing events were modelled as competing risks.

6.3. Relationship to the Literature

In finding bias from competing risks in cumulative risk estimates from the Kaplan-Meier estimator, Chapter 4 confirmed the results of studies in clinical areas such as cancer progression in patients with head or neck cancer[1], cancer incidence in patients with chronic kidney disease (CKD)[70], cardiovascular mortality in patients who underwent renal replacement therapy[5], renal transplantation in patients on dialysis[8] and coronary artery bypass in patients who underwent cardiac catheterisation[10]. It also corroborated the results of the initial simulation work performed by Gooley et al.[1] in the derivation of the cumulative incidence competing risk (CICR) estimator[1]. The results of this study further add to the evidence of bias in the Kaplan-Meier estimator by demonstrating its presence in the novel clinical areas of cardiovascular and cancer mortality in patients with T2DM.

From the studies identified in the literature, those closest to the clinical areas evaluated in Chapter 4 were conducted by Wong et al.[70] and Verduijn et al.[5]. However, the populations or outcomes in these studies differed from those in Chapter 4, with Wong et al.[70] evaluating the cumulative risk of cancer in patients with diabetes mellitus, and Verduijn et al.[5] evaluating the cumulative risk of cardiovascular mortality in patients with renal replacement therapy. Verduijn et al.[5] did not stratify their cumulative risk estimates. Whereas, Wong et al.[70] stratified cumulative risk estimates by the presence or absence of CKD; analogous to stratification by albuminuria level in cumulative risk estimates of Chapter 4. The results of the study by Wong et al.[70] were similar to those described for the outcome of cancer mortality in Chapter 4, i.e. greater absolute and relative differences between the Kaplan-Meier and CICR estimators for the less common (higher risk) CKD strata. Comparisons for cardiovascular mortality were not possible as Wong et al.[70] did not explicitly model competing mortality.

Two studies were identified in the literature that explicitly modelled both primary and competing outcomes using Kaplan-Meier and CICR estimators[5,8]; making them methodologically similar to the analyses performed in Chapter 4. In these studies, differences between Kaplan-Meier and CICR estimators were evaluated for primary and competing outcomes with relatively equal event rates ("cardiovascular mortality" and "non-cardiovascular mortality") in patients who underwent renal replacement therapy in the study by Verduijn et al.[5], and primary and competing outcomes with relatively unequal event rates ("receiving a renal transplant" (rare) and "death before receiving a renal transplant" (common)) in patients on dialysis in the study by Noordzij et al.[8]. In the study by Verduijn et al.[5] absolute differences between the Kaplan-Meier and CICR estimators were 13.0% and 15.6%, for the outcomes of cardiovascular and non-cardiovascular mortality, respectively, while relative differences were 33.5% and 39.5%, respectively. In the study by Noordzij et al.[8] absolute differences between the Kaplan-Meier and CICR estimators were 9.0% and 9.0% for the outcomes of receiving a renal transplant and death before transplant, respectively, while relative differences were 37.5% and 17.6%, respectively. The results of Noordzij et al.[8] more closely resemble those of this study, and re-affirm the observation that unequal competing outcomes result in a greater relative difference between the Kaplan-Meier and CICR estimator for the less common outcome. However,

the identical absolute differences described between outcomes by Noordzij et al.[8] differed from the results obtained in this study, in which greater absolute differences were present for the more common outcome.

The Nelson-Aalen estimator is infrequently used in the literature, and is often overlooked as a standard survival analysis comparator to the CICR estimator. Some authors have stated that the Nelson-Aalen estimator is not biased by competing risks[14]. However, it has been shown to overestimate cumulative risk when compared to the CICR estimator, both in this study and in other studies in the literature[171].

Studies comparing the estimates of standard CSH survival models (like Cox-PH regression) and SDH competing risks models (like the Fine-Gray model) are common in the literature[7,8,17,20,140,172,173]. However, studies are rare in which comparisons have been made between standard and competing risks CSH models, or competing risks CSH and SDH models[17]. The work in Chapter 5 is the first to apply CSH and SDH competing risks models to the assessment of the association between albuminuria and the outcomes of cardiovascular and cancer mortality in patients with T2DM, and also the first to evaluate the differences between model estimates in this clinical area. Competing risks models are usually only used in the evaluation of rare outcomes, or outcomes for which there is a very strong competing risk, e.g. the evaluation of end-stage renal disease in patients on dialysis[17], or the evaluation of cancer incidence in patients with T2DM[140]. Hence, through the evaluation of outcomes that were not highly unbalanced, the work conducted in Chapter 5 may also be considered novel.

Comparisons to the literature for the work performed in Chapter 5 are problematic; partially owing to the novel nature of many attributes the work, but also because the estimates of SDH models may have lower transportability compared to those of CSH models[7]. Lau et al. note that, although not empirically shown in their data, the transportability of SDH model estimates "may be questionable if the distribution of competing events differs from the original population"[7]. This quality may be largely a result of the different types of risk associations being described by each model; with the

aetiological associations described by CSH models being independent of the effect of the risk factor on competing outcomes, and the prognostic relationships described by SDH models being dependent upon the effect of the risk factor on the primary outcome and all competing outcomes. Many of these effects may be population-specific. Thus, increasing the number of effects on which a risk metric is dependent may increase the nature to which it is population-specific, thereby decreasing its generalisability. It follows that as the risk estimates of CSH and SDH models may be likely to have different degrees of population specificity, so too may the differences between CSH and SDH model risk estimates.

For comparisons between standard and competing risks CSH models, a single study was found in the literature, by Lim et al.[17], in which the effects of model selection on risk estimates were assessed for the risk factors of age and gender on the outcomes of end-stage renal disease and death without end-stage renal disease in patients with diabetes. Despite similarities in the methodologies and study populations, the results of Lim et al.[17] contrasted with those of Chapter 5, as they found much greater differences between the estimates of Cox-PH and Lunn-McNeil models (whereas these models produced were near identical estimates in Chapter 5). Although it is possible that differences in the clinical areas may have had some impact on this finding, its consistency across risk factors and outcomes renders this unlikely. One possible explanation is that there were undetected violations of the proportional CSHs assumption for the relationship between some risk factors and outcomes. Lim et al.[17] state that they assumed proportionality towards both outcomes for all risk factors, rendering this a plausible explanation. The presence of violations of the proportional CSH assumption can pose problems for the accuracy of model estimates for the risk-factor-outcome combinations affected. This was demonstrated in the sensitivity analysis of Chapter 5, where the addition of time interactions for albuminuria led to albuminuria risk estimates that varied substantially across follow-up. An additional requirement of the Lunn-McNeil model (over the Cox-PH model) is that the CSHs of all risk factors must be proportional to a joint baseline hazard distribution shared by all outcomes in the model. Thus, it follows that the CSHs of all outcomes under assessment must also be proportional to each other for each risk factor. This additional requirement of the Lunn-McNeil model may amplify any bias present from individual violations of the proportional CSHs assumption, especially where the

violations are positive with respect to the CSH of one outcome in the model, and negative with respect to the CSH of another. A second possible reason for the greater differences found by Lim et al.[17] between Cox-PH and Lunn-McNeil models may be the use of ulterior specifications of the Lunn-McNeil model[174]. However, Lim et al.[17] cited the relevant publication for the same unstratified Lunn-McNeil model used for the analyses in Chapter 5[16].

For comparisons between CSH and SDH competing risks models the aforementioned study by Lim et al.[17] was the only study found in the literature to compare model estimates from the Lunn-McNeil and the Fine-Gray model[17]. In this study, the hazard ratios (HRs) of the Lunn-McNeil model were greater than those of the Fine-Gray model for all risk factors and outcomes, re-affirming the findings of Chapter 5. A further similarity between the findings of Chapter 5 and Lim et al.[17] was the overlapping 95% CIs present between CSH and SDH competing risk models. This finding was unexpected, as the primary outcome of end-stage renal disease evaluated by Lim et al.[17] is rarer in patients with diabetes than either of the outcomes evaluated in Chapter 5. However, the failure to determine a difference between model estimates for end-stage renal disease may have been the result of an insufficient population sample (N = 8,254).

Extrapolating a finding of Chapter 5, that the HRs of standard survival analysis appear to contain negligible bias from competing risks, to $HR_{Cox\text{-}PH} \approx HR_{Lunn\text{-}McNeil}$ allows for a much broader point of comparison to be made with the literature. With this assumption in mind, comparisons between the estimates of CSH and SDH survival models have previously been performed in patients with T2DM by Yu et al.[140]. In this study, risk estimates from Cox-PH models were compared to those from Fine-Gray models for the effect of estimated glomerular filtration rate and vascular adhesion protein on cancer incidence. The clinical areas described by this study may be considered similar to a subset of those present in Chapter 5, with both studies essentially assessing the effect of measures of kidney disease on cancer outcomes. For these assessments, risk estimates were lower from SDH models than CSH models. Unlike for the results in Chapter 5, the reduction in the magnitude of risk estimates moving from a CSH to SDH model rendered the SHR statistically non-significant. However, this difference in findings may be attributable to the small population sample (N = 568) used by Yu et

al.[140]. SDH models typically produce lower risk estimates than CSH models[7,8,17,20,140,172,173], and thus, tend to require larger population samples to determine statistical significance. The effect of reducing the population size was also demonstrated in some of the sensitivity analyses in Chapter 5, where SDH risk estimates for cancer mortality were rendered non-significant. The effects of reduced glomerular filtration rate on competing mortality were not explicitly modelled by Yu et al.[140], so no further comparisons could be made.

Aside from the study by Lim et al.[17], three studies were identified in the literature, in which the effect of risk factors were explicitly modelled on both the primary outcome and competing mortality, using CSH and SDH models[7,8,172]. However, the populations used by these studies differ from the one used in Chapter 5; consisting of patients on dialysis[8], patients with primary or secondary acute myeloid leukaemia[172], and patients with human immunodeficiency virus[7]. In each of these studies, larger decreases in the magnitude of effect of risk predictors were present for the less common outcome when transitioning from CSH to SDH models, mirroring the findings of Chapter 5. However, unlike the findings in Chapter 5, increases in the magnitude of effect of risk predictors were present for the effect of some risk factors on the more common outcome when transitioning from CSH to SDH models [7,8]. While it is true that the effect of a risk factor on the SDH of an outcome will usually be lower than the effect on a risk factor on the CSH of the same outcome, this is not an absolute rule, and it will usually only hold true when the effect of the risk factor on the CSH of the primary outcome is in the same direction as the effect on the risk factor on the CSH of competing outcomes, i.e. a HR_{CS} > 1 for both the primary outcome and competing outcomes[7]. Where the effect of risk factors on primary and competing outcomes diverges, i.e. a HR_{CS} > 1 for one for the primary outcome and a HR_{CS} < 1 for competing outcomes, HR_{SD} can be greater than HR_{CS}[7,8,173].

6.4. Further Research

In Chapter 4, there were approximately three times as many non-cancer deaths as cancer deaths, and one and a half times as many non-cardiovascular deaths as cardiovascular deaths. However, competing risks models are more often used in the evaluation of rare or highly unbalanced competing

outcomes. Extending the analyses of this Thesis to the investigation of more and less balanced outcomes may allow for a more in-depth evaluation of the effect of competing risks, and the degree to which these may be liable to effect risk estimates.

Non-parametric estimators, such as the Kaplan-Meier and CICR, provide one means from which to estimate cumulative risk. However, it is also possible to estimate cumulative risk from the hazard- and subhazard distribution of Cox-PH, Lunn-McNeil and Fine-Gray models[17]. The derivation of cumulative risk estimates through such means is uncommon, although studies in other medical fields have demonstrated that the use of Cox-PH models as a basis for these risk estimates may lead to the incorporation of bias from competing risks[17]. The application of Cox-PH, Lunn-McNeil and Fine-Gray models to the estimation of the cumulative risks of cardiovascular and cancer mortality in patients with T2DM would be a natural extension to the research conducted in Chapters 4 and 5.

Risk scores like QLung[175], QDScore[176] QRISK2[48] and QRISK Lifetime[177] typically use regression coefficients from CSH models as the basis for their risk prediction equations. However, the regression coefficients from SDH models may constitute a better means through which to quantify risk prediction, as they provide measures of prognostic association as opposed to aetiological association (in the sense defined in Chapter 1)[7,8]. Thus, the application of the Fine-Gray model to the derivation of risk scores may result in models that calibrate and discriminate better than their CSHs counterparts. The Fine-Gray model is not the sole means through which it is possible to calculate the effect of risk factors on the SDH of an outcome[178–180], and the development of new SDH models is a budding area of research. The further application of these models to the derivation or risk scores may result in risk scores with different levels of discrimination and calibration than those derived from Fine-Gray models.

6.5. Implications

When investigating absolute or cumulative risks, the choice between standard- and competing-risks-adjusted methods may be driven by the nature of the risk estimate desired. Cumulative hazards

derived from standard methods describe the cumulative risk of experiencing the primary outcome in CSH risk sets. As such, the cumulative hazard may prove some use in quantifying what the cumulative risk would have been in a population if patients hadn't succumbed to competing mortality. However, such inferences are rarely desired to be made, and the use of cumulative hazard to make such inferences would rely on the untestable assumption of independence between all competing outcomes[7]. Cumulative incidences derived from competing risks methods describe the cumulative risk of experiencing the primary outcome in SDH risk sets. As such, the cumulative incidence may be a more informative risk measure than the cumulative hazard as it's able to quantify the probability of experiencing an outcome that's adjusted for the occurrence of competing outcomes. Moreover, cumulative risk estimators are generally used to provide inferences on the real-world probability of experiencing study outcomes, for which the cumulative incidence is likely to be a much more representative measure than the cumulative hazard.

When investigating relative measures of risk, model selection should be primarily driven by the measure of association desired. For research questions where aetiological associations are desired, e.g. the investigation of epidemiological relationships between risk factors and outcomes, standard survival analysis CSH (Cox-PH) model estimates produced identical risk estimates to competing risk CSH (Lunn-McNeil) models. Although competing risks CSH models may be used, the benefits of their use are not immediately apparent, while their difficulty of implementation and their inflexibility render them problematic. For research questions where prognostic associations are desired, e.g. the investigation of predictive relationships between risk factors and outcomes, competing risks SDH (Fine-Gray) model estimates should be used in place of standard or competing risks CSH models, as they allow for the effect of risk factors on the cumulative incidence of an outcome to be quantified. CSH and SDH models produced risk estimates that would have similar clinical interpretations for the effect of albuminuria on cardiovascular mortality, but that would have different clinical interpretations for the effect of albuminuria on cancer mortality. Furthermore, the strength of association between albuminuria and each outcome was affected differently by model selection, i.e. albuminuria was much more strongly associated with cancer mortality in CSH models than SDH models, while albuminuria was strongly associated with cardiovascular mortality in both model types.

Thus, use of the incorrect relative hazards model type for describing prognostic relationships may have resulted in an overestimation of the incidence of cancer mortality associated with albuminuria.

The work in this Thesis has demonstrated that albuminuria is associated with the cumulative risk, CSH and SDH of both cardiovascular and cancer mortality in patients with T2DM. While this adds to a wealth of evidence for the outcome of cardiovascular mortality, only a single study was identified by the systematic review in Chapter 2 for which cancer mortality was evaluated as an outcome. The results from this study suggest that albuminuria may need to be included in future risk prediction models for cancer, and that these should consider the use of SDH models.

Appendices

Appendix A

Appendix A.1 - Review Search Strategy

MEDLINE

1. exp diabetes mellitus, type 2/

2. (((type 2 or type ii or noninsulin dependent or non-insulin dependent or adult onset) adj diabet*) or t2d or t2dm or niddm).ti,ab.

3. 1 or 2

4. diabetic nephropathies/

5. (proteinuria? or albuminuria? or m?croalbuminuria? or ((diabetic or diabetes) adj (nephropath* or glomeruloscleros?s or kidney disease? or renal disease? or nephrit?s)) or (kimmelstiel wilson adj (syndrome? or disease?)) or intercapillary glomerulonephrit?s).ti,ab.

6. 4 or 5

7. 3 and 6

8. cardiovascular diseases/ or exp heart diseases/ or exp vascular diseases/

9. (((cardiovascular or heart or cardiac or sinus node or coronary or myocardial or pulmonary or postphlebitic or vascula* or small vessel or aortic or arterial or artery or buerger* or pulseless or clarkson* or cerebrovascular or atherosclerotic or arteriosclerotic or binswanger* or chronic progressive subcortical or venoocclusive or veno occlusive) adj4 (disease? or disorder?)) or arrhythmia? or dysrhythmia? or ((atrial or auricular or ventricular) adj (fibrillation? or flutter?)) or ((sinus node or right bundle branch block or st segment elevation or sudden death or sudden unexplained nocturnal death or coronary or sick sinus or long qt or pre excitation or preexcitation or lown ganong levine or short pr normal qrs complex or wolff parkinson white or wpw or anomalous

ventricular excitation or auriculoventricular accessory pathway or av accessory pathway or a-v accessory pathway or oc?ulocraniosomatic or cardio renal or cardiorenal or reno cardiac or renocardial or blue toe or spinal artery or goldblatt* or leriche or susac* or capillary leak or claude* or weber* or millard gublar* or top of the basilar or benedict* or foville* or medullary or cerebellar artery or wallenberg* or lateral bulbar or cerebral artery or aca or pca or mca or lacunar or postphlebitic) adj2 syndrome?) or ((hear or caridiac or cardiopulmonary or cardio pulmonary or sudden) adj3 (arrest? or pause?)) or bradycardia? or bradyarrhythmia? or brugada or ectopic heartbeat? or extrasystole? or ((atrial or auricular or ventricular) adj ectopic beat?) or (premature adj4 (beat? or contraction? or complex* or complices)) or commontio cordis or cardiac concussion? or ((heart or atrioventricular or atrio-ventricular or a-v or av or atrioventricular conduction or atrio-ventricular conduction or bundle branch or fascicular or sinoatrial or sa) adj2 block?) or ((atrioventricular or atrio-ventricular or a-v or av) adj dissociation?) or adam? stokes or ((sinus node or ventricular) adj2 dysfunction?) or parasystole? or (mahaim type adj2 (pre excitation or preexcitation)) or tachycardia? or accelerated idioventricular rhythm? or torsade? de pointes or ((high or low) adj2 cardiac output?) or ((cardiac or heart or pericardial) adj tamponade?) or enlarged heart or ((cardiac or heart) adj (enlargement? or hypertryph*)) or cardiomegal* or cardiomyopath* or (ventricular adj hypertroph*) or myocardiopath* or ((arrhythmogenic or fibromuscular) adj4 dysplasia?) or (asymmetric* adj2 septal hypertroph*) or endo*cardial fibroelastos?s or endomyocardial fibros?s or kearns sayre or (kearn* adj4 syndrome) or ophthalmoplegia? or (myocardial adj2 reperfusion?) or cardit?s or myocardit?s or (endocardit?s adj3 (non infective or non bacterial or marantic)) or aneur?sm? or asystole? or ((heart or cardiac or myocardial) adj failure?) or (heart adj2 compensation?) or paroxysmal dyspn?ea? or (asthma adj2 cardiac) or ((cardiac or quincke* or angio or angioneurotic) adj (oedema? or edema?)) or ((heart or cardiac or free wall or ventricular or aortic) adj2 (rupture? or perforation?)) or ((heart or cardiac) adj (valve or valvular) adj prolapse?) or ((arterial or aortic or coronary or venous or valvular or subvalvular or supravalvular or subaortic or mitral or pulmonary or pulmonic or infundibular or tricuspid) adj2 (regurgitation? or incompetence? or insufficienc* or stenos?s or restenos?s or atresia?)) or obstructive subaortic conus or (absent adj2 (atrioventricular or av or a-v) adj connection?) or isch?emi* or angina? or stenocardia? or angor pectoris or (myocardial adj (preinfarction? or pre

infarction?)) or arterioscleros?s or arteriosclerotic or atherosclerosis or atherosclerotic or atherogenes?s or coronary occlusion? or subclavian steal or thrombos?s or thrombus or thrombophlebitis or thrombophlebitides or phlebitis or phlebitides or phlegmasia alba dolens or thromboembol* or atheroembolism? or embolism? or embolus or vasospasm? or angiospasm? or ((artery or arterial) adj spasm?) or (myocardial adj3 (stunn* or hibernation?)) or cardiogenic shock? or pericardial effusion? or h?emopericardium? or chylopericardium? or pericardit?s or pleuropercardit?s or (pick* adj4 heart) or pneumopericardium? or cor pulmonale or infarct* or (arter* adj3 dissection?) or pseudoaneur?sm? or endoleak? or (leak? adj3 perigraft) or angiodysplasia? or giant urticaria? or aortit?s or claudication or ((artery or arterial or vascular or vertibrobasilar or brachial basilar or carotid or vena cava or venous or vein) adj4 (stenos?s or narrowing? or plaque? or ulcer* or obstruction? or occlusion? or inflammation?)) or moyamoya? or moya moya? or progressive intracranial occlusive arteropath* or thromboangiitis obliterans or arteritis or arteritides or endarteritis or endarteritides or polyarteritis or polyarteritides or takayasu or ((cerebrovascular or vascular or arterial or artery) adj3 (disorder? or occlusion? or insufficienc*)) or vasculopath* or h?emorrhage? or lacunar dementia? or stroke? or dolichoectasia? or (subclavian adj4 steal) or angiopath* or m?croangiopath* or (microscopic adj (polyangitides or polyangiitis)) or (vascular adj (headache? or cephalgia?)) or ((vascular or atherosclerotic or arteriosclerotic) adj dementia?) or ((binswanger* or chronic progressive subcortical) adj encephalopath*) or subcortical leukoencephalopath* or h?ematoma? or apoplex* or sneddon or sneddon champion or angiitis or angiitides or vasculitis or vasculitides or ((vascular or vasculature) adj complication?) or ((diabetic or diabetes) adj2 (foot or feet or retinopath*)) or h?emorrhoid? or h?emostatic disorder? or cryoglobulinemia? or (hyperglobulinemic adj purpura?) or hyperemia? or (venous adj (engorgement? or congestion?)) or hypertens* or hypotens* or prehypertens* or prehypotens* or ((high or low) adj blood pressure) or optic neuropath* or erythromelalgia? or livedo reticularis or (reperfusion adj (injur* or damage?)) or h?ematomyelia? or telangiectas* or spider vein? or varicocele? or varicose vein? or varix or varices or ((varicose or venous or hypertens* or stasis or postphlebitic) adj ulcer?) or postplebitic or postthrombotic or sunds or aivr or ihss? or hocm? or cpeo? or chf? or chd? or ihd? or tia or tias or pvd? or pad? or sah or sahs).ti,ab.

10. exp neoplasms/ not cysts/

11. (tumo?r? or neoplas* or paraneoplas* or cancer* or malignan* or myos?s or metastas?s or micrometastas?s or carcinoid? or nodule? or cacinogenes?s or cocarcinogenes?s or oncogenes?s or tumo?rigenes?s or histiocytic disorder? or hamartoma* or sarcoma* or leuk?emi* or leucocyth?emia* or leuk?emic or chloroma* or lymphangioma* or endothelioma* or lymphangioendothelioma*).ti,ab. or lymphangioleiomyoma*.mp. or lymphangiomyoma*.ti,ab. or lymphangiosarcoma*.ti,ab. or lymphoma*.ti,ab. or reticulolymphosarcoma*.ti,ab. or germinoblastoma*.ti,ab. or hodgkin*.ti,ab. or lymphogranuloma*.ti,ab. or granuloma*.ti,ab. or nonhodgkin*.ti,ab. or reticulosarcoma*.ti,ab. or immunoblastoma*.ti,ab. or granulomatous slack skin.ti,ab. or lymphomatoid.ti,ab. or mycosis fungoides.ti,ab. or pagetoid reticulos?s.ti,ab. or woringer kolopp*.ti,ab. or ketron goodman*.ti,ab. or sezary*.ti,ab. or adenolymphoma*.ti,ab. or cystadenoma*.ti,ab. or adenoma*.ti,ab. or syringoma*.ti,ab. or adenomyoepithelioma*.ti,ab. or adenomyoma*.ti,ab. or adenosarcoma*.ti,ab. or carcinoma*.ti,ab. or carcinosarcoma*.ti,ab. or hepatoblastoma*.ti,ab. or mesenchymoma*.ti,ab. or myoepithelioma*.ti,ab. or nephroma*.ti,ab. or blastoma*.ti,ab. or thymoma*.ti,ab. or nephroblastoma*.ti,ab. or drash*.ti,ab. or angiolipoma*.ti,ab. or angiomyolipoma*.ti,ab. or hibernoma*.ti,ab. or liposarcoma*.ti,ab. or myelolipoma*.ti,ab. or chondroblastoma*.ti,ab. or chondroma*.ti,ab. or chondrosarcoma*.ti,ab. or mastocytos?s.ti,ab. or mast cell disease?.ti,ab. or mastocytoma*.ti,ab. or urticaria pigmentosa?.ti,ab. or myofibroma*.ti,ab. or myxoma*.ti,ab. or angiomyxoma*.ti,ab. or neurothekeoma*.ti,ab. or neurothec?oma*.ti,ab. or myxosarcoma*.ti,ab. or fibroma*.ti,ab. or osteoblastoma*.ti,ab. or osteoma*.ti,ab. or osteochondroma*.ti,ab. or chondrosteoma*.ti,ab. or exostos?s.ti,ab. or aclas?s.ti,ab. or osteoma*.ti,ab. or osteosarcoma*.ti,ab. or fibromatos?s.ti,ab. or desmoid?.ti,ab. or fibrosarcoma*.ti,ab. or dermatofibrosarcoma*.ti,ab. or neurofibrosarcoma*.ti,ab. or histiocytoma*.ti,ab. or dermatofibroma*.ti,ab. or h?emangioma*.ti,ab. or angioma*.ti,ab. or adenofibroma*.ti,ab. or cystadenofibroma*.ti,ab. or fibroadenoma*.ti,ab. or mesothelioma*.ti,ab. or submesothelioma*.ti,ab. or melanoma*.ti,ab. or synovioma*.ti,ab. or myoblastoma*.ti,ab. or myofibroblastoma*.ti,ab. or leiomyoma?.ti,ab. or fibroid?.ti,ab. or angiomyoma*.ti,ab. or angioleiomyoma*.ti,ab. or leiomyoblastoma*.ti,ab. or leiomyomatos?s.ti,ab. or leiomyosarcoma*.ti,ab. or myoma*.ti,ab. or rhabdomyoma*.ti,ab. or myosarcoma*.ti,ab. or rhabdomyosarcoma*.ti,ab. or pecoma*.ti,ab. or angiomyolipoma*.ti,ab. or lymphangioleiomyomatos?s.ti,ab. or lymphangiomyomatos?s.ti,ab. or adenosarcoma*.ti,ab. or

carcinosarcoma*.ti,ab. or chondrosarcoma*.ti,ab. or dermatofibrosarcoma*.ti,ab. or neurofibrosarcoma*.ti,ab. or h?emangiosarcoma*.ti,ab. or angiosarcoma*.ti,ab. or histiocytoma*.ti,ab. or lymphangioendothelioma*.ti,ab. or osteosarcoma*.ti,ab. or cystosarcoma*.ti,ab. or chordoma*.ti,ab. or germinoma*.ti,ab. or dysgerminoma*.ti,ab. or seminova*.ti,ab. or gonadoblastoma*.ti,ab. or mesonephroma*.ti,ab. or craniopharyngioma*.ti,ab. or spongioblastoma*.ti,ab. or glioma*.ti,ab. or astroblastoma*.ti,ab. or ependymoastrocytoma*.ti,ab. or ganglioneuroma*.ti,ab. or gangliocytoma*.ti,ab.
or astrocytoma*.ti,ab. or glioblastoma*.ti,ab. or ependymoma*.ti,ab. or glios?s.ti,ab. or ganglioglioma*.ti,ab. or gliosarcoma*.ti,ab. or medulloblastoma*.ti,ab. or oligodendroglioma*.ti,ab. or neurocytoma*.ti,ab. or medulloepithelioma*.ti,ab. or ependymoblastoma*.ti,ab. or neuroepithelioma*.ti,ab. or neuroblastoma*.ti,ab. or epithesioneuroblastoma*.ti,ab. or aesthesioneuroblastoma*.ti,ab. or pinealoma*.ti,ab. or ganglioneuroblastoma*.ti,ab. or pineoblastoma*.ti,ab. or pineocytoma*.ti,ab. or pinealocytoma*.ti,ab. or retinoblastoma*.ti,ab. or melanoameloblastoma*.ti,ab. or progonoma*.ti,ab. or apudoma*.ti,ab. or argentaffinoma*.ti,ab. or somatostatinoma*.ti,ab. or vipoma*.ti,ab. or ((verner morrison or watery diarrhea) adj syndrome?).ti,ab. or watery diarrhea with hypokalemic alkalos?s.ti,ab. or melanotic?.ti,ab. or lentigo maligna.ti,ab. or neurilem?oma*.ti,ab. or neurinoma*.ti,ab. or schwannoma*.ti,ab. or neurilem?oma*.ti,ab. or neurilemmosarcoma*.ti,ab. or neurofibroma*.ti,ab. or paraganglioma*.ti,ab. or chemodectoma*.ti,ab. or pheochromocytoma*.ti,ab. or teratocarcinoma*.ti,ab. or teratoma*.ti,ab. or dysembryoma*.ti,ab. or dermoid?.ti,ab. or struma ovarii.ti,ab. or choriocarcinoma*.ti,ab. or (trophoblastic gestational adj disease?).ti,ab. or mole.ti,ab. or moles.ti,ab. or molar.ti,ab. or cholangioma*.ti,ab. or chorioadenoma*.ti,ab. or nesidioblastoma*.ti,ab. or hepatoma*.ti,ab. or oncocytoma*.ti,ab. or syringadenoma*.ti,ab. or acrospiroma*.ti,ab. or hiradenoma*.ti,ab. or poroma*.ti,ab. or hidracanthoma*.ti,ab. or hidrocystoma*.ti,ab. or polyp?.ti,ab. or polypos*.ti,ab. or somatotrophinoma*.ti,ab. or prolactinoma*.ti,ab. or m?croprolactinoma*.ti,ab. or epithelioma*.ti,ab. or adenocarcinoma*.ti,ab. or lintis plastica.ti,ab. or cylindroma*.ti,ab. or paget*.ti,ab. or gastrinoma*.ti,ab. or glucagonoma*.ti,ab. or somatostatinoma*.ti,ab. or hypernephroma*.ti,ab. or cholangiocarcinoma*.ti,ab. or cystadenocarcinoma*.ti,ab. or porocarcinoma*.ti,ab. or ((bowen* or kahler*) adj disease?).ti,ab. or condyloma*.ti,ab. or pilomatrixoma*.ti,ab. or fibroadenos?s.ti,ab. or

acanthoma*.ti,ab. or papilloma*.ti,ab. or gynandroblastoma*.ti,ab. or luteoma*.ti,ab. or arrhenoblastoma*.ti,ab. or androblastoma*.ti,ab. or meningioma*.ti,ab. or myeloma*.ti,ab. or macroglobulin?emia?.ti,ab. or plasmacytoma*.ti,ab. or angiofibroma*.ti,ab. or angiokeratoma*.ti,ab. or glomangioma*.ti,ab. or h?emangioendothelioma*.ti,ab. or h?emangioblastoma*.ti,ab. or kasabach merritt*.ti,ab. or parkes*.ti,ab. or sturge*.ti,ab. or weber*.ti,ab. or phakoma*.ti,ab. or h?emangiopericytoma*.ti,ab. or ameloblastoma*.ti,ab. or cementoma*.ti,ab. or odontoma*.ti,ab. or fibroodontoma*.ti,ab. or adamantinoma*.ti,ab. or nesidioblastoma*.ti,ab. or insulinoma*.ti,ab. or conn.ti,ab. or conns.ti,ab. or conn's.ti,ab. or meig*.ti,ab. or thecoma*.ti,ab. or ((paraneoplastic endocrine or ectopic hormone or nelson*) adj syndrome?).ti,ab. or neuroma*.ti,ab. or muir torre*.ti,ab. or pancoat*.ti,ab. or coin lesion?.ti,ab. or wilm*.ti,ab. or ccmmt?.ti,ab. or pnet?.ti,ab. or wdha?.ti,ab. or wdhh?.ti,ab. or dcis?.ti,ab. or mpnst?.ti,ab. or ibc?.ti,ab.

12. 8 or 9 or 10 or 11

13. 7 and 12

14. cause of death/ or risk factors/ or risk/ or risk assessment/ or incidence/ or prevalence/ or prognosis/ or survival analysis/ or survival rate/ or life expectancy/ or proportional hazards models/ or mortality/ 15. risk*.ti. or (risk* adj factor?).ti,ab. or (risk* adj assessment).ti,ab. or (risk* adj marker?).ti,ab. or (risk adj scor*).ti,ab. or (predict or predictor? or prediction?).ti,ab. or incidence.ti,ab. or hazard.ti,ab. or epidemiol*.ti,ab. or cohort.ti,ab. or trial.ti,ab. or rct.ti,ab. or (prognosis or prognostic).ti,ab. or (mortality or death? or survival).ti,ab. or life expectancy.ti,ab. or follow up.ti,ab.

15. 14 or 15

16. 13 and 16

17. exp animal/ not human/

18. (malformation? or inherit* or familial or hereditary or infecti* or bacterial or viral or pathogenic or autoimmune or auto immune or injur* or trauma* or autosomal* or x link* or y link* or chromasomal or congenital or hiv or aids or tb or tuberculos?s or rheumatic or benign*).ti,ab.

19. 17 not 18

20. 20 not 19

EMBASE

1. exp *diabetes mellitus, type 2/

2. (((type 2 or type ii or noninsulin dependent or non-insulin dependent or adult onset) adj diabet*) or t2d or t2dm or niddm).ti,ab.

3. *diabetic nephropathies/

4. (proteinuria? or albuminuria? or m?croalbuminuria? or ((diabetic or diabetes) adj (nephropath* or glomeruloscleros?s or kidney disease? or renal disease? or nephrit?s)) or (kimmelstiel wilson adj (syndrome? or disease?)) or intercapillary glomerulonephrit?s).ti,ab.

5. 1 or 2

6. 3 or 4

7. 5 and 6

8. *cardiovascular diseases/ or exp *heart diseases/ or exp *vascular diseases/

9. (((cardiovascular or heart or cardiac or sinus node or coronary or myocardial or pulmonary or postphlebitic or vascula* or small vessel or aortic or arterial or artery or buerger* or pulseless or clarkson* or cerebrovascular or atherosclerotic or arteriosclerotic or binswanger* or chronic progressive subcortical or venoocclusive or veno occlusive) adj4 (disease? or disorder?)) or arrhythmia? or dysrhythmia? or ((atrial or auricular or ventricular) adj (fibrillation? or flutter?)) or ((sinus node or right bundle branch block or st segment elevation or sudden death or sudden unexplained nocturnal death or coronary or sick sinus or long qt or pre excitation or preexcitation or lown ganong levine or short pr normal qrs complex or wolff parkinson white or wpw or anomalous ventricular excitation or auriculoventricular accessory pathway or av accessory pathway or a-v accessory pathway or oc?ulocraniosomatic or cardio renal or cardiorenal or reno cardiac or renocardial or blue toe or spinal artery or goldblatt* or leriche or susac* or capillary leak or claude* or weber* or millard gublar* or top of the basilar or benedict* or foville* or medullary or cerebellar artery or wallenberg* or lateral bulbar or cerebral artery or aca or pca or mca or lacunar or postphlebitic) adj2 syndrome?) or ((hear or caridiac or cardiopulmonary or cardio pulmonary or sudden) adj3 (arrest? or pause?)) or bradycardia? or bradyarrhythmia? or brugada or ectopic heartbeat? or extrasystole? or ((atrial or auricular or ventricular) adj ectopic beat?) or (premature

adj4 (beat? or contraction? or complex* or complices)) or commontio cordis or cardiac concussion? or ((heart or atrioventricular or atrio-ventricular or a-v or av or atrioventricular conduction or atrio-ventricular conduction or bundle branch or fascicular or sinoatrial or sa) adj2 block?) or ((atrioventricular or atrio-ventricular or a-v or av) adj dissociation?) or adam? stokes or ((sinus node or ventricular) adj2 dysfunction?) or parasystole? or (mahaim type adj2 (pre excitation or preexcitation)) or tachycardia? or accelerated idioventricular rhythm? or torsade? de pointes or ((high or low) adj2 cardiac output?) or ((cardiac or heart or pericardial) adj tamponade?) or enlarged heart or ((cardiac or heart) adj (enlargement? or hypertryph*)) or cardiomegal* or cardiomyopath* or (ventricular adj hypertroph*) or myocardiopath* or ((arrhythmogenic or fibromuscular) adj4 dysplasia?) or (asymmetric* adj2 septal hypertroph*) or endo*cardial fibroelastos?s or endomyocardial fibros?s or kearns sayre or (kearn* adj4 syndrome) or ophthalmoplegia? or (myocardial adj2 reperfusion?) or cardit?s or myocardit?s or (endocardit?s adj3 (non infective or non bacterial or marantic)) or aneur?sm? or asystole? or ((heart or cardiac or myocardial) adj failure?) or (heart adj2 compensation?) or paroxysmal dyspn?ea? or (asthma adj2 cardiac) or ((cardiac or quincke* or angio or angioneurotic) adj (oedema? or edema?)) or ((heart or cardiac or free wall or ventricular or aortic) adj2 (rupture? or perforation?)) or ((heart or cardiac) adj (valve or valvular) adj prolapse?) or ((arterial or aortic or coronary or venous or valvular or subvalvular or supravalvular or subaortic or mitral or pulmonary or pulmonic or infundibular or tricuspid) adj2 (regurgitation? or incompetence? or insufficienc* or stenos?s or restenos?s or atresia?)) or obstructive subaortic conus or (absent adj2 (atrioventricular or av or a-v) adj connection?) or isch?emi* or angina? or stenocardia? or angor pectoris or (myocardial adj (preinfarction? or pre infarction?)) or arterioscleros?s or arteriosclerotic or atheroscleros?s or atherosclerotic or atherogenes?s or coronary occlusion? or subclavian steal or thrombos?s or thrombus or thrombophlebitis or thrombophlebitides or phlebitis or phlebitides or phlegmasia alba dolens or thromboembol* or atheroembolism? or embolism? or embolus or vasospasm? or angiospasm? or ((artery or arterial) adj spasm?) or (myocardial adj3 (stunn* or hibernation?)) or cardiogenic shock? or pericardial effusion? or h?emopericardium? or chylopericardium? or pericardit?s or pleuropercardit?s or (pick* adj4 heart) or pneumopericardium? or cor pulmonale or infarct* or (arter* adj3 dissection?) or pseudoaneur?sm? or endoleak? or (leak? adj3 perigraft) or

angiodysplasia? or giant urticaria? or aortit?s or claudication or ((artery or arterial or vascular or vertibrobasilar or brachial basilar or carotid or vena cava or venous or vein) adj4 (stenos?s or narrowing? or plaque? or ulcer* or obstruction? or occlusion? or inflammation?)) or moyamoya? or moya moya? or progressive intracranial occlusive arteropath* or thromboangiitis obliterans or arteritis or arteritides or endarteritis or endarteritides or polyarteritis or polyarteritides or takayasu or ((cerebrovascular or vascular or arterial or artery) adj3 (disorder? or occlusion? or insufficienc*)) or vasculopath* or h?emorrhage? or lacunar dementia? or stroke? or dolichoectasia? or (subclavian adj4 steal) or angiopath* or m?croangiopath* or (microscopic adj (polyangitides or polyangiitis)) or (vascular adj (headache? or cephalgia?)) or ((vascular or atherosclerotic or arteriosclerotic) adj dementia?) or ((binswanger* or chronic progressive subcortical) adj encephalopath*) or subcortical leukoencephalopath* or h?ematoma? or apoplex* or sneddon or sneddon champion or angiitis or angiitides or vasculitis or vasculitides or ((vascular or vasculature) adj complication?) or ((diabetic or diabetes) adj2 (foot or feet or retinopath*)) or h?emorrhoid? or h?emostatic disorder? or cryoglobulinemia? or (hyperglobulinemic adj purpura?) or hyperemia? or (venous adj (engorgement? or congestion?)) or hypertens* or hypotens* or prehypertens* or prehypotens* or ((high or low) adj blood pressure) or optic neuropath* or erythromelalgia? or livedo reticularis or (reperfusion adj (injur* or damage?)) or h?ematomyelia? or telangiectas* or spider vein? or varicocele? or varicose vein? or varix or varices or ((varicose or venous or hypertens* or stasis or postphlebitic) adj ulcer?) or postplebitic or postthrombotic or sunds or aivr or ihss? or hocm? or cpeo? or chf? or chd? or ihd? or tia or tias or pvd? or pad? or sah or sahs).ti,ab.

10. 8 or 9

11. exp *neoplasms/ not *cysts/

12. (tumo?r? or neoplas* or paraneoplas* or cancer* or malignan* or myos?s or metastas?s or micrometastas?s or carcinoid? or nodule? or cacinogenes?s or cocarcinogenes?s or oncogenes?s or tumo?rigenes?s or histiocytic disorder? or hamartoma* or sarcoma* or leuk?emi* or leucocyth?emia* or leuk?emic or chloroma* or lymphangioma* or endothelioma* or lymphangioendothelioma*).ti,ab. or lymphangioleiomyoma*.mp. or lymphangiomyoma*.ti,ab. or lymphangiosarcoma*.ti,ab. or lymphoma*.ti,ab. or reticulolymphosarcoma*.ti,ab. or germinoblastoma*.ti,ab. or hodgkin*.ti,ab. or lymphogranuloma*.ti,ab. or granuloma*.ti,ab. or nonhodgkin*.ti,ab. or reticulosarcoma*.ti,ab. or

immunoblastoma*.ti,ab. or granulomatous slack skin.ti,ab. or lymphomatoid.ti,ab. or mycosis fungoides.ti,ab. or pagetoid reticulos?s.ti,ab. or woringer kolopp*.ti,ab. or ketron goodman*.ti,ab. or sezary*.ti,ab. or adenolymphoma*.ti,ab. or cystadenoma*.ti,ab. or adenoma*.ti,ab. or syringoma*.ti,ab. or adenomyoepithelioma*.ti,ab. or adenomyoma*.ti,ab. or adenosarcoma*.ti,ab. or carcinoma*.ti,ab. or carcinosarcoma*.ti,ab. or hepatoblastoma*.ti,ab. or mesenchymoma*.ti,ab. or myoepithelioma*.ti,ab. or nephroma*.ti,ab. or blastoma*.ti,ab. or thymoma*.ti,ab. or nephroblastoma*.ti,ab. or drash*.ti,ab. or angiolipoma*.ti,ab. or angiomyolipoma*.ti,ab. or hibernoma*.ti,ab. or liposarcoma*.ti,ab. or myelolipoma*.ti,ab. or chondroblastoma*.ti,ab. or chondroma*.ti,ab. or chondrosarcoma*.ti,ab. or mastocytos?s.ti,ab. or mast cell disease?.ti,ab. or mastocytoma*.ti,ab. or urticaria pigmentosa?.ti,ab. or myofibroma*.ti,ab. or myxoma*.ti,ab. or angiomyxoma*.ti,ab. or neurothekeoma*.ti,ab. or neurothec?oma*.ti,ab. or myxosarcoma*.ti,ab. or fibroma*.ti,ab. or osteoblastoma*.ti,ab. or osteoma*.ti,ab. or osteochondroma*.ti,ab. or chondrosteoma*.ti,ab. or exostos?s.ti,ab. or aclas?s.ti,ab. or osteoma*.ti,ab. or osteosarcoma*.ti,ab. or fibromatos?s.ti,ab. or desmoid?.ti,ab. or fibrosarcoma*.ti,ab. or dermatofibrosarcoma*.ti,ab. or neurofibrosarcoma*.ti,ab. or histiocytoma*.ti,ab. or dermatofibroma*.ti,ab. or h?emangioma*.ti,ab. or angioma*.ti,ab. or adenofibroma*.ti,ab. or cystadenofibroma*.ti,ab. or fibroadenoma*.ti,ab. or mesothelioma*.ti,ab. or submesothelioma*.ti,ab. or melanoma*.ti,ab. or synovioma*.ti,ab. or myoblastoma*.ti,ab. or myofibroblastoma*.ti,ab. or leiomyoma?.ti,ab. or fibroid?.ti,ab. or angiomyoma*.ti,ab. or angioleiomyoma*.ti,ab. or leiomyoblastoma*.ti,ab. or leiomyomatos?s.ti,ab. or leiomyosarcoma*.ti,ab. or myoma*.ti,ab. or rhabdomyoma*.ti,ab. or myosarcoma*.ti,ab. or rhabdomyosarcoma*.ti,ab. or pecoma*.ti,ab. or angiomyolipoma*.ti,ab. or lymphangioleiomyomatos?s.ti,ab. or lymphangiomyomatos?s.ti,ab. or adenosarcoma*.ti,ab. or carcinosarcoma*.ti,ab. or chondrosarcoma*.ti,ab. or dermatofibrosarcoma*.ti,ab. or neurofibrosarcoma*.ti,ab. or h?emangiosarcoma*.ti,ab. or angiosarcoma*.ti,ab. or histiocytoma*.ti,ab. or lymphangioendothelioma*.ti,ab. or osteosarcoma*.ti,ab. or cystosarcoma*.ti,ab. or chordoma*.ti,ab. or germinoma*.ti,ab. or dysgerminoma*.ti,ab. or seminova*.ti,ab. or gonadoblastoma*.ti,ab. or mesonephroma*.ti,ab. or craniopharyngioma*.ti,ab. or spongioblastoma*.ti,ab. or glioma*.ti,ab. or astroblastoma*.ti,ab. or ependymoastrocytoma*.ti,ab. or ganglioneuroma*.ti,ab. or gangliocytoma*.ti,ab. or

astrocytoma*.ti,ab. or glioblastoma*.ti,ab. or ependymoma*.ti,ab. or glios?s.ti,ab. or ganglioglioma*.ti,ab. or gliosarcoma*.ti,ab. or medulloblastoma*.ti,ab. or oligodendroglioma*.ti,ab. or neurocytoma*.ti,ab. or medulloepithelioma*.ti,ab. or ependymoblastoma*.ti,ab. or neuroepithelioma*.ti,ab. or neuroblastoma*.ti,ab. or epithesioneuroblastoma*.ti,ab. or aesthesioneuroblastoma*.ti,ab. or pinealoma*.ti,ab. or ganglioneuroblastoma*.ti,ab. or pineoblastoma*.ti,ab. or pineocytoma*.ti,ab. or pinealocytoma*.ti,ab. or retinoblastoma*.ti,ab. or melanoameloblastoma*.ti,ab. or progonoma*.ti,ab. or apudoma*.ti,ab. or argentaffinoma*.ti,ab. or somatostatinoma*.ti,ab. or vipoma*.ti,ab. or ((verner morrison or watery diarrhea) adj syndrome?).ti,ab. or watery diarrhea with hypokalemic alkalos?s.ti,ab. or melanotic?.ti,ab. or lentigo maligna.ti,ab. or neurilem?oma*.ti,ab. or neurinoma*.ti,ab. or schwannoma*.ti,ab. or neurilem?oma*.ti,ab. or neurilemmosarcoma*.ti,ab. or neurofibroma*.ti,ab. or paraganglioma*.ti,ab. or chemodectoma*.ti,ab. or pheochromocytoma*.ti,ab. or teratocarcinoma*.ti,ab. or teratoma*.ti,ab. or dysembryoma*.ti,ab. or dermoid?.ti,ab. or struma ovarii.ti,ab. or choriocarcinoma*.ti,ab. or (trophoblastic gestational adj disease?).ti,ab. or mole.ti,ab. or moles.ti,ab. or molar.ti,ab. or cholangioma*.ti,ab. or chorioadenoma*.ti,ab. or nesidioblastoma*.ti,ab. or hepatoma*.ti,ab. or oncocytoma*.ti,ab. or syringadenoma*.ti,ab. or acrospiroma*.ti,ab. or hiradenoma*.ti,ab. or poroma*.ti,ab. or hidracanthoma*.ti,ab. or hidrocystoma*.ti,ab. or polyp?.ti,ab. or polypos*.ti,ab. or somatotrophinoma*.ti,ab. or prolactinoma*.ti,ab. or m?croprolactinoma*.ti,ab. or epithelioma*.ti,ab. or adenocarcinoma*.ti,ab. or lintis plastica.ti,ab. or cylindroma*.ti,ab. or paget*.ti,ab. or gastrinoma*.ti,ab. or glucagonoma*.ti,ab. or somatostatinoma*.ti,ab. or hypernephroma*.ti,ab. or cholangiocarcinoma*.ti,ab. or cystadenocarcinoma*.ti,ab. or porocarcinoma*.ti,ab. or ((bowen* or kahler*) adj disease?).ti,ab. or condyloma*.ti,ab. or pilomatrixoma*.ti,ab. or fibroadenos?s.ti,ab. or acanthoma*.ti,ab. or papilloma*.ti,ab. or gynandroblastoma*.ti,ab. or luteoma*.ti,ab. or arrhenoblastoma*.ti,ab. or androblastoma*.ti,ab. or meningioma*.ti,ab. or myeloma*.ti,ab. or macroglobulin?emia?.ti,ab. or plasmacytoma*.ti,ab. or angiofibroma*.ti,ab. or angiokeratoma*.ti,ab. or glomangioma*.ti,ab. or h?emangioendothelioma*.ti,ab. or h?emangioblastoma*.ti,ab. or kasabach merritt*.ti,ab. or parkes*.ti,ab. or sturge*.ti,ab. or weber*.ti,ab. or phakoma*.ti,ab. or h?emangiopericytoma*.ti,ab. or ameloblastoma*.ti,ab. or cementoma*.ti,ab. or odontoma*.ti,ab. or fibroodontoma*.ti,ab. or adamantinoma*.ti,ab. or

nesidioblastoma*.ti,ab. or insulinoma*.ti,ab. or conn.ti,ab. or conns.ti,ab. or conn's.ti,ab. or meig*.ti,ab. or thecoma*.ti,ab. or ((paraneoplastic endocrine or ectopic hormone or nelson*) adj syndrome?).ti,ab. or neuroma*.ti,ab. or muir torre*.ti,ab. or pancoat*.ti,ab. or coin lesion?.ti,ab. or wilm*.ti,ab. or ccmmt?.ti,ab. or pnet?.ti,ab. or wdha?.ti,ab. or wdhh?.ti,ab. or dcis?.ti,ab. or mpnst?.ti,ab. or ibc?.ti,ab.

13. 11 or 12

14. 10 or 13

15. 7 and 14

16. *cause of death/ or *risk factors/ or *risk/ or *risk assessment/ or *incidence/ or *prevalence/ or *prognosis/ or *survival analysis/ or *survival rate/ or *life expectancy/ or *proportional hazards models/ or *mortality/

17. risk*.ti. or (risk* adj factor?).ti,ab. or (risk* adj assessment).ti,ab. or (risk* adj marker?).ti,ab. or (risk adj scor*).ti,ab. or (predict or predictor? or prediction?).ti,ab. or incidence.ti,ab. or hazard.ti,ab. or epidemiol*.ti,ab. or cohort.ti,ab. or trial.ti,ab. or rct.ti,ab. or (prognosis or prognostic).ti,ab. or (mortality or death? or survival).ti,ab. or life expectancy.ti,ab. or follow up.ti,ab.

18. 16 or 17

19. 15 and 18

20. exp *animal/ not *human/

21. (malformation? or inherit* or familial or hereditary or infecti* or bacterial or viral or pathogenic or autoimmune or auto immune or injur* or trauma* or autosomal* or x link* or y link* or chromasomal or congenital or hiv or aids or tb or tuberculos?s or rheumatic or benign*).ti,ab.

22. 19 not 20

23. 22 not 21

CINAHL

S1. (MH Diabetes Mellitus, Type 2)

S2. TI (((type 2 OR type ii OR noninsulin dependent OR non-insulin dependent OR adult onset) n1 diabet*) OR t2d OR t2dm OR niddm) OR AB (((type 2 OR type ii OR noninsulin dependent OR non-insulin dependent OR adult onset) n1 diabet*) OR t2d OR t2dm OR niddm)

S3. S1 OR S2

S4. (MH Diabetic Nephropathies)

S5. TI (proteinuria? OR albuminuria? OR m?croalbuminuria? OR (diabet* n1 (nephropath* OR glomeruloscleros?s OR kidney disease? OR renal disease? OR nephrit?s)) OR ((kimmelstiel wilson) n1 (syndrome? OR disease?)) OR intercapillary glomerulonephrit?s) OR AB (proteinuria? OR albuminuria? OR m?croalbuminuria? OR (diabet* n1 (nephropath* OR glomeruloscleros?s OR kidney disease? OR renal disease? OR nephrit?s)) OR ((kimmelstiel wilson) n1 (syndrome? OR disease?)) OR intercapillary glomerulonephrit?s)

S6. S4 OR S5

S7. (S4 OR S5) AND (S3 AND S6)

S8. (MH "Heart Diseases+") OR (MH "Vascular Diseases+") OR (MM "Cardiovascular Diseases")

S9. TI (((cardiovascular OR heart OR cardiac OR sinus node OR coronary OR myocardial OR pulmonary OR postphlebitic OR vascula* OR small vessel OR aortic OR arterial OR artery OR buerger* OR pulseless OR clarkson* OR cerebrovascular OR atherosclerotic OR arteriosclerotic OR binswanger* OR chronic progressive subcortical OR venoocclusive OR veno occlusive) n4 (disease? OR disorder?)) OR arrhythmia? OR dysrhythmia? OR ((atrial OR auricular OR ventricular) n1 (fibrillation? OR flutter?)) OR ((sinus node OR right bundle branch block OR st segment elevation OR sudden death OR sudden unexplained nocturnal death OR coronary OR sick sinus OR long qt OR pre excitation OR preexcitation OR lown ganong levine OR short pr normal qrs complex OR wolff parkinson white OR wpw OR anomalous ventricular excitation OR auriculoventricular accessory pathway OR av accessory pathway OR a-v accessory pathway OR oc?ulocraniosomatic OR cardio renal OR cardiorenal OR reno cardiac OR renocardial OR blue toe OR spinal artery OR goldblatt* OR leriche OR susac* OR capillary leak OR claude* OR weber* OR millard gublar* OR top of the basilar OR benedict* OR foville* OR medullary OR cerebellar artery OR wallenberg* OR lateral bulbar OR cerebral artery OR aca OR pca OR mca OR lacunar OR postphlebitic) n2 syndrome?) OR ((hear OR caridiac OR cardiopulmonary OR cardio pulmonary OR sudden) n3 (arrest? OR pause?)) OR bradycardia? OR bradyarrhythmia? OR brugada OR ectopic heartbeat? OR extrasystole? OR ((atrial OR auricular OR ventricular) n1 ectopic beat?) OR (premature n4 (beat? OR contraction? OR complex* OR complices)) OR commontio cordis OR cardiac concussion? OR

((heart OR atrioventricular OR atrio-ventricular OR a-v OR av OR atrioventricular conduction OR atrio-ventricular conduction OR bundle branch OR fascicular OR sinoatrial OR sa) n2 block?) OR ((atrioventricular OR atrio-ventricular OR a-v OR av) n1 dissociation?) OR adam? stokes OR ((sinus node OR ventricular) n2 dysfunction?) OR parasystole? OR (mahaim type n2 (pre excitation OR preexcitation)) OR tachycardia? OR accelerated idioventricular rhythm? OR torsade? de pointes OR ((high OR low) n2 cardiac output?) OR ((cardiac OR heart OR pericardial) n1 tamponade?) OR enlarged heart OR ((cardiac OR heart) n1 (enlargement? OR hypertryph*)) OR cardiomegal* OR cardiomyopath* OR (ventricular n1 (hypertroph*)) OR myocardiopath* OR ((arrhythmogenic OR fibromuscular) n4 dysplasia?) OR (asymmetric* n2 (septal hypertroph*)) OR endo*cardial fibroelastos?s OR endomyocardial fibros?s OR kearns sayre OR (kearn* n4 syndrome) OR ophthalmoplegia? OR (myocardial n2 reperfusion?) OR cardit?s OR myocardit?s OR (endocardit?s n3 (non infective OR non bacterial OR marantic)) OR aneur?sm? OR asystole? OR ((heart OR cardiac OR myocardial) n1 failure?) OR (heart n2 compensation?) OR paroxysmal dyspn?ea? OR (asthma n2 cardiac) OR ((cardiac OR quincke* OR angio OR angioneurotic) n1 (oedema? OR edema?)) OR ((heart OR cardiac OR free wall OR ventricular OR aortic) n2 (rupture? OR perforation?)) OR ((heart OR cardiac) n1 (valve OR valvular) n1 prolapse?) OR ((arterial OR aortic OR coronary OR venous OR valvular OR subvalvular OR supravalvular OR subaortic OR mitral OR pulmonary OR pulmonic OR infundibular OR tricuspid) n2 (regurgitation? OR incompetence? OR insufficienc* OR stenos?s OR restenos?s OR atresia?)) OR obstructive subaortic conus OR (absent n2 (atrioventricular OR av OR a-v) n1 connection?) OR isch?emi* OR angina? OR stenocardia? OR angor pectoris OR (myocardial n1 (preinfarction? OR pre infarction?)) OR arterioscleros?s OR arteriosclerotic OR atheroscleros?s OR atherosclerotic OR atherogenes?s OR coronary occlusion? OR subclavian steal OR thrombos?s OR thrombus OR thrombophlebitis OR thrombophlebitides OR phlebitis OR phlebitides OR phlegmasia alba dolens OR thromboembol* OR atheroembolism? OR embolism? OR embolus OR vasospasm? OR angiospasm? OR ((artery OR arterial) n1 spasm?) OR (myocardial n3 (stunn* OR hibernation?)) OR cardiogenic shock? OR pericardial effusion? OR h?emopericardium? OR chylopericardium? OR pericardit?s OR pleuropercardit?s OR (pick* n4 heart) OR pneumopericardium? OR cor pulmonale OR infarct* OR (arter* n3 dissection?) OR pseudoaneur?sm? OR endoleak? OR (leak? n3 perigraft) OR angiodysplasia? OR giant urticaria? OR aortit?s OR claudication OR ((artery OR arterial OR

vascular OR vertibrobasilar OR brachial basilar OR carotid OR vena cava OR venous OR vein) n4 (stenos?s OR narrowing? OR plaque? OR ulcer* OR obstruction? OR occlusion? OR inflammation?)) OR moyamoya? OR moya moya? OR progressive intracranial occlusive arteropath* OR thromboangiitis obliterans OR arteritis OR arteritides OR endarteritis OR endarteritides OR polyarteritis OR polyarteritides OR takayasu OR ((cerebrovascular OR vascular OR arterial OR artery) n3 (disorder? OR occlusion? OR insufficienc*)) OR vasculopath* OR h?emorrhage? OR lacunar dementia? OR stroke? OR dolichoectasia? OR (subclavian n4 steal) OR angiopath* OR m?croangiopath* OR (microscopic n1 (polyangitides OR polyangiitis)) OR (vascular n1 (headache? OR cephalgia?)) OR ((vascular OR atherosclerotic OR arteriosclerotic) n1 dementia?) OR ((binswanger* OR chronic progressive subcortical) n1 encephalopath*) OR subcortical leukoencephalopath* OR h?ematoma? OR apoplex* OR sneddon OR sneddon champion OR angiitis OR angiitides OR vasculitis OR vasculitides OR ((vascular OR vasculature) n1 complication?) OR ((diabetic OR diabetes) n2 (foot OR feet OR retinopath*)) OR h?emorrhoid? OR h?emostatic disorder? OR cryoglobulinemia? OR (hyperglobulinemic n1 purpura?) OR hyperemia? OR (venous n1 (engorgement? OR congestion?)) OR hypertens* OR hypotens* OR prehypertens* OR prehypotens* OR ((high OR low) n1 blood pressure) OR optic neuropath* OR erythromelalgia? OR livedo reticularis OR (reperfusion n1 (injur* OR damage?)) OR h?ematomyelia? OR telangiectas* OR spider vein? OR varicocele? OR varicose vein? OR varix OR varices OR ((varicose OR venous OR hypertens* OR stasis OR postphlebitic) n1 ulcer?) OR postplebitic OR postthrombotic OR sunds OR aivr OR ihss? OR hocm? OR cpeo? OR chf? OR chd? OR ihd? OR tia OR tias OR pvd? OR pad? OR sah OR sahs) OR AB (((cardiovascular OR heart OR cardiac OR sinus node OR coronary OR myocardial OR pulmonary OR postphlebitic OR vascula* OR small vessel OR aortic OR arterial OR artery OR buerger* OR pulseless OR clarkson* OR cerebrovascular OR atherosclerotic OR arteriosclerotic OR binswanger* OR chronic progressive subcortical OR venoocclusive OR veno occlusive) n4 (disease? OR disorder?)) OR arrhythmia? OR dysrhythmia? OR ((atrial OR auricular OR ventricular) n1 (fibrillation? OR flutter?)) OR ((sinus node OR right bundle branch block OR st segment elevation OR sudden death OR sudden unexplained nocturnal death OR coronary OR sick sinus OR long qt OR pre excitation OR preexcitation OR lown ganong levine OR short pr normal qrs complex OR wolff parkinson white OR wpw OR anomalous ventricular excitation OR auriculoventricular

accessory pathway OR av accessory pathway OR a-v accessory pathway OR oc?ulocraniosomatic OR cardio renal OR cardiorenal OR reno cardiac OR renocardial OR blue toe OR spinal artery OR goldblatt* OR leriche OR susac* OR capillary leak OR claude* OR weber* OR millard gublar* OR top of the basilar OR benedict* OR foville* OR medullary OR cerebellar artery OR wallenberg* OR lateral bulbar OR cerebral artery OR aca OR pca OR mca OR lacunar OR postphlebitic) n2 syndrome?) OR ((hear OR caridiac OR cardiopulmonary OR cardio pulmonary OR sudden) n3 (arrest? OR pause?)) OR bradycardia? OR bradyarrhythmia? OR brugada OR ectopic heartbeat? OR extrasystole? OR ((atrial OR auricular OR ventricular) n1 ectopic beat?) OR (premature n4 (beat? OR contraction? OR complex* OR complices)) OR commontio cordis OR cardiac concussion? OR ((heart OR atrioventricular OR atrio-ventricular OR a-v OR av OR atrioventricular conduction OR atrio-ventricular conduction OR bundle branch OR fascicular OR sinoatrial OR sa) n2 block?) OR ((atrioventricular OR atrio-ventricular OR a-v OR av) n1 dissociation?) OR adam? stokes OR ((sinus node OR ventricular) n2 dysfunction?) OR parasystole? OR (mahaim type n2 (pre excitation OR preexcitation)) OR tachycardia? OR accelerated idioventricular rhythm? OR torsade? de pointes OR ((high OR low) n2 cardiac output?) OR ((cardiac OR heart OR pericardial) n1 tamponade?) OR enlarged heart OR ((cardiac OR heart) n1 (enlargement? OR hypertryph*)) OR cardiomegal* OR cardiomyopath* OR (ventricular n1 (hypertroph*)) OR myocardiopath* OR ((arrhythmogenic OR fibromuscular) n4 dysplasia?) OR (asymmetric* n2 (septal hypertroph*)) OR endo*cardial fibroelastos?s OR endomyocardial fibros?s OR kearns sayre OR (kearn* n4 syndrome) OR ophthalmoplegia? OR (myocardial n2 reperfusion?) OR cardit?s OR myocardit?s OR (endocardit?s n3 (non infective OR non bacterial OR marantic)) OR aneur?sm? OR asystole? OR ((heart OR cardiac OR myocardial) n1 failure?) OR (heart n2 compensation?) OR paroxysmal dyspn?ea? OR (asthma n2 cardiac) OR ((cardiac OR quincke* OR angio OR angioneurotic) n1 (oedema? OR edema?)) OR ((heart OR cardiac OR free wall OR ventricular OR aortic) n2 (rupture? OR perforation?)) OR ((heart OR cardiac) n1 (valve OR valvular) n1 prolapse?) OR ((arterial OR aortic OR coronary OR venous OR valvular OR subvalvular OR supravalvular OR subaortic OR mitral OR pulmonary OR pulmonic OR infundibular OR tricuspid) n2 (regurgitation? OR incompetence? OR insufficienc* OR stenos?s OR restenos?s OR atresia?)) OR obstructive subaortic conus OR (absent n2 (atrioventricular OR av OR a-v) n1 connection?) OR isch?emi* OR angina? OR stenocardia? OR angor pectoris OR (myocardial

n1 (preinfarction? OR pre infarction?)) OR arterioscleros?s OR arteriosclerotic OR atheroscleros?s OR atherosclerotic OR atherogenes?s OR coronary occlusion? OR subclavian steal OR thrombos?s OR thrombus OR thrombophlebitis OR thrombophlebitides OR phlebitis OR phlebitides OR phlegmasia alba dolens OR thromboembol* OR atheroembolism? OR embolism? OR embolus OR vasospasm? OR angiospasm? OR ((artery OR arterial) n1 spasm?) OR (myocardial n3 (stunn* OR hibernation?)) OR cardiogenic shock? OR pericardial effusion? OR h?emopericardium? OR chylopericardium? OR pericardit?s OR pleuropercardit?s OR (pick* n4 heart) OR pneumopericardium? OR cor pulmonale OR infarct* OR (arter* n3 dissection?) OR pseudoaneur?sm? OR endoleak? OR (leak? n3 perigraft) OR angiodysplasia? OR giant urticaria? OR aortit?s OR claudication OR ((artery OR arterial OR vascular OR vertibrobasilar OR brachial basilar OR carotid OR vena cava OR venous OR vein) n4 (stenos?s OR narrowing? OR plaque? OR ulcer* OR obstruction? OR occlusion? OR inflammation?)) OR moyamoya? OR moya moya? OR progressive intracranial occlusive arteropath* OR thromboangiitis obliterans OR arteritis OR arteritides OR endarteritis OR endarteritides OR polyarteritis OR polyarteritides OR takayasu OR ((cerebrovascular OR vascular OR arterial OR artery) n3 (disorder? OR occlusion? OR insufficienc*)) OR vasculopath* OR h?emorrhage? OR lacunar dementia? OR stroke? OR dolichoectasia? OR (subclavian n4 steal) OR angiopath* OR m?croangiopath* OR (microscopic n1 (polyangitides OR polyangiitis)) OR (vascular n1 (headache? OR cephalgia?)) OR ((vascular OR atherosclerotic OR arteriosclerotic) n1 dementia?) OR ((binswanger* OR chronic progressive subcortical) n1 encephalopath*) OR subcortical leukoencephalopath* OR h?ematoma? OR apoplex* OR sneddon OR sneddon champion OR angiitis OR angiitides OR vasculitis OR vasculitides OR ((vascular OR vasculature) n1 complication?) OR ((diabetic OR diabetes) n2 (foot OR feet OR retinopath*)) OR h?emorrhoid? OR h?emostatic disorder? OR cryoglobulinemia? OR (hyperglobulinemic n1 purpura?) OR hyperemia? OR (venous n1 (engorgement? OR congestion?)) OR hypertens* OR hypotens* OR prehypertens* OR prehypotens* OR ((high OR low) n1 blood pressure) OR optic neuropath* OR erythromelalgia? OR livedo reticularis OR (reperfusion n1 (injur* OR damage?)) OR h?ematomyelia? OR telangiectas* OR spider vein? OR varicocele? OR varicose vein? OR varix OR varices OR ((varicose OR venous OR hypertens* OR stasis OR postphlebitic) n1 ulcer?) OR postplebitic OR postthrombotic OR sunds OR

aivr OR ihss? OR hocm? OR cpeo? OR chf? OR chd? OR ihd? OR tia OR tias OR pvd? OR pad? OR sah OR sahs)

S10. ((MH "Neoplasms+") NOT (MH "Cysts"))

S11. TI (tumo?r? OR neoplas* OR paraneoplas* OR cancer* OR malignan* OR myos?s OR metastas?s OR micrometastas?s OR carcinoid? OR nodule? OR cacinogenes?s OR cocarcinogenes?s OR oncogenes?s OR tumo?rigenes?s OR histiocytic disorder? OR hamartoma* OR sarcoma* OR leuk?emi* OR leucocyth?emia* OR leuk?emic OR chloroma* OR lymphangioma* OR endothelioma* OR lymphangioendothelioma* OR lymphangioleiomyoma* OR lymphangiomyoma* OR lymphangiosarcoma* OR lymphoma* OR reticulolymphosarcoma* OR germinoblastoma* OR hodgkin* OR lymphogranuloma* OR granuloma* OR nonhodgkin* OR reticulosarcoma* OR immunoblastoma* OR granulomatous slack skin OR lymphomatoid OR mycosis fungoides OR pagetoid reticulos?s OR woringer kolopp* OR ketron goodman* OR sezary* OR adenolymphoma* OR cystadenoma* OR adenoma* OR syringoma* OR adenomyoepithelioma* OR adenomyoma* OR adenosarcoma* OR carcinoma* OR carcinosarcoma* OR hepatoblastoma* OR mesenchymoma* OR myoepithelioma* OR nephroma* OR blastoma* OR thymoma* OR nephroblastoma* OR drash* OR angiolipoma* OR angiomyolipoma* OR hibernoma* OR liposarcoma* OR myelolipoma* OR chondroblastoma* OR chondroma* OR chondrosarcoma* OR mastocytos?s OR mast cell disease? OR mastocytoma* OR urticaria pigmentosa? OR myofibroma* OR myxoma* OR angiomyxoma* OR neurothekeoma* OR neurothec?oma* OR myxosarcoma* OR fibroma* OR osteoblastoma* OR osteoma* OR osteochondroma* OR chondrosteoma* OR exostos?s OR aclas?s OR osteoma* OR osteosarcoma* OR fibromatos?s OR desmoid? OR fibrosarcoma* OR dermatofibrosarcoma* OR neurofibrosarcoma* OR histiocytoma* OR dermatofibroma* OR h?emangioma* OR angioma* OR adenofibroma* OR cystadenofibroma* OR fibroadenoma* OR mesothelioma* OR submesothelioma* OR melanoma* OR synovioma* OR myoblastoma* OR myofibroblastoma* OR leiomyoma? OR fibroid? OR angiomyoma* OR angioleiomyoma* OR leiomyoblastoma* OR leiomyomatos?s OR leiomyosarcoma* OR myoma* OR rhabdomyoma* OR myosarcoma* OR rhabdomyosarcoma* OR pecoma* OR angiomyolipoma* OR lymphangioleiomyomatos?s OR lymphangiomyomatos?s OR adenosarcoma* OR carcinosarcoma* OR chondrosarcoma* OR dermatofibrosarcoma* OR neurofibrosarcoma* OR h?emangiosarcoma* OR angiosarcoma* OR

histiocytoma* OR lymphangioendothelioma* OR osteosarcoma* OR cystosarcoma* OR chordoma* OR germinoma* OR dysgerminoma* OR seminova* OR gonadoblastoma* OR mesonephroma* OR craniopharyngioma* OR spongioblastoma* OR glioma* OR astroblastoma* OR ependymoastrocytoma* OR ganglioneuroma* OR gangliocytoma* OR astrocytoma* OR glioblastoma* OR ependymoma* OR glios?s OR ganglioglioma* OR gliosarcoma* OR medulloblastoma* OR oligodendroglioma* OR neurocytoma* OR medulloepithelioma* OR ependymoblastoma* OR neuroepithelioma* OR neuroblastoma* OR epithesioneuroblastoma* OR aesthesioneuroblastoma* OR pinealoma* OR ganglioneuroblastoma* OR pineoblastoma* OR pineocytoma* OR pinealocytoma* OR retinoblastoma* OR melanoameloblastoma* OR progonoma* OR apudoma* OR argentaffinoma* OR somatostatinoma* OR vipoma* OR ((verner morrison OR watery diarrhea) n1 syndrome?) OR watery diarrhea with hypokalemic alkalos?s OR melanotic? OR lentigo maligna OR neurilem?oma* OR neurinoma* OR schwannoma* OR neurilem?oma* OR neurilemmosarcoma* OR neurofibroma* OR paraganglioma* OR chemodectoma* OR pheochromocytoma* OR teratocarcinoma* OR teratoma* OR dysembryoma* OR dermoid? OR struma ovarii OR choriocarcinoma* OR ((trophoblastic gestational) n1 disease?) OR mole OR moles OR molar OR cholangioma* OR chorioadenoma* OR nesidioblastoma* OR hepatoma* OR oncocytoma* OR syringadenoma* OR acrospiroma* OR hiradenoma* OR poroma* OR hidracanthoma* OR hidrocystoma* OR polyp? OR polypos* OR somatotrophinoma* OR prolactinoma* OR m?croprolactinoma* OR epithelioma* OR adenocarcinoma* OR lintis plastica OR cylindroma* OR paget* OR gastrinoma* OR glucagonoma* OR somatostatinoma* OR hypernephroma* OR cholangiocarcinoma* OR cystadenocarcinoma* OR porocarcinoma* OR ((bowen* OR kahler*) n1 disease?) OR condyloma* OR pilomatrixoma* OR fibroadenos?s OR acanthoma* OR papilloma* OR gynandroblastoma* OR luteoma* OR arrhenoblastoma* OR androblastoma* OR meningioma* OR myeloma* OR macroglobulin?emia? OR plasmacytoma* OR angiofibroma* OR angiokeratoma* OR glomangioma* OR h?emangioendothelioma* OR h?emangioblastoma* OR kasabach merritt* OR parkes* OR sturge* OR weber* OR phakoma* OR h?emangiopericytoma* OR ameloblastoma* OR cementoma* OR odontoma* OR fibroodontoma* OR adamantinoma* OR nesidioblastoma* OR insulinoma* OR conn OR conns OR conn's OR meig* OR thecoma* OR ((paraneoplastic endocrine OR ectopic hormone OR nelson*) n1 syndrome?) OR

neuroma* OR muir torre* OR pancoat* OR coin lesion? OR wilm* OR ccmmt? OR pnet? OR wdha? OR wdhh? OR dcis? OR mpnst? OR ibc?) OR AB (tumo?r? OR neoplas* OR paraneoplas* OR cancer* OR malignan* OR myos?s OR metastas?s OR micrometastas?s OR carcinoid? OR nodule? OR cacinogenes?s OR cocarcinogenes?s OR oncogenes?s OR tumo?rigenes?s OR histiocytic disorder? OR hamartoma* OR sarcoma* OR leuk?emi* OR leucocyth?emia* OR leuk?emic OR chloroma* OR lymphangioma* OR endothelioma* OR lymphangioendothelioma* OR lymphangioleiomyoma* OR lymphangiomyoma* OR lymphangiosarcoma* OR lymphoma* OR reticulolymphosarcoma* OR germinoblastoma* OR hodgkin* OR lymphogranuloma* OR granuloma* OR nonhodgkin* OR reticulosarcoma* OR immunoblastoma* OR granulomatous slack skin OR lymphomatoid OR mycosis fungoides OR pagetoid reticulos?s OR woringer kolopp* OR ketron goodman* OR sezary* OR adenolymphoma* OR cystadenoma* OR adenoma* OR syringoma* OR adenomyoepithelioma* OR adenomyoma* OR adenosarcoma* OR carcinoma* OR carcinosarcoma* OR hepatoblastoma* OR mesenchymoma* OR myoepithelioma* OR nephroma* OR blastoma* OR thymoma* OR nephroblastoma* OR drash* OR angiolipoma* OR angiomyolipoma* OR hibernoma* OR liposarcoma* OR myelolipoma* OR chondroblastoma* OR chondroma* OR chondrosarcoma* OR mastocytos?s OR mast cell disease? OR mastocytoma* OR urticaria pigmentosa? OR myofibroma* OR myxoma* OR angiomyxoma* OR neurothekeoma* OR neurothec?oma* OR myxosarcoma* OR fibroma* OR osteoblastoma* OR osteoma* OR osteochondroma* OR chondrosteoma* OR exostos?s OR aclas?s OR osteoma* OR osteosarcoma* OR fibromatos?s OR desmoid? OR fibrosarcoma* OR dermatofibrosarcoma* OR neurofibrosarcoma* OR histiocytoma* OR dermatofibroma* OR h?emangioma* OR angioma* OR adenofibroma* OR cystadenofibroma* OR fibroadenoma* OR mesothelioma* OR submesothelioma* OR melanoma* OR synovioma* OR myoblastoma* OR myofibroblastoma* OR leiomyoma? OR fibroid? OR angiomyoma* OR angioleiomyoma* OR leiomyoblastoma* OR leiomyomatos?s OR leiomyosarcoma* OR myoma* OR rhabdomyoma* OR myosarcoma* OR rhabdomyosarcoma* OR pecoma* OR angiomyolipoma* OR lymphangioleiomyomatos?s OR lymphangiomyomatos?s OR adenosarcoma* OR carcinosarcoma* OR chondrosarcoma* OR dermatofibrosarcoma* OR neurofibrosarcoma* OR h?emangiosarcoma* OR angiosarcoma* OR histiocytoma* OR lymphangioendothelioma* OR osteosarcoma* OR cystosarcoma* OR chordoma* OR germinoma* OR dysgerminoma* OR seminova* OR

gonadoblastoma* OR mesonephroma* OR craniopharyngioma* OR spongioblastoma* OR glioma* OR astroblastoma* OR ependymoastrocytoma* OR ganglioneuroma* OR gangliocytoma* OR astrocytoma* OR glioblastoma* OR ependymoma* OR glios?s OR ganglioglioma* OR gliosarcoma* OR medulloblastoma* OR oligodendroglioma* OR neurocytoma* OR medulloepithelioma* OR ependymoblastoma* OR neuroepithelioma* OR neuroblastoma* OR epithesioneuroblastoma* OR aesthesioneuroblastoma* OR pinealoma* OR ganglioneuroblastoma* OR pineoblastoma* OR pineocytoma* OR pinealocytoma* OR retinoblastoma* OR melanoameloblastoma* OR progonoma* OR apudoma* OR argentaffinoma* OR somatostatinoma* OR vipoma* OR ((verner morrison OR watery diarrhea) n1 syndrome?) OR watery diarrhea with hypokalemic alkalos?s OR melanotic? OR lentigo maligna OR neurilem?oma* OR neurinoma* OR schwannoma* OR neurilem?oma* OR neurilemmosarcoma* OR neurofibroma* OR paraganglioma* OR chemodectoma* OR pheochromocytoma* OR teratocarcinoma* OR teratoma* OR dysembryoma* OR dermoid? OR struma ovarii OR choriocarcinoma* OR ((trophoblastic gestational) n1 disease?) OR mole OR moles OR molar OR cholangioma* OR chorioadenoma* OR nesidioblastoma* OR hepatoma* OR oncocytoma* OR syringadenoma* OR acrospiroma* OR hiradenoma* OR poroma* OR hidracanthoma* OR hidrocystoma* OR polyp? OR polypos* OR somatotrophinoma* OR prolactinoma* OR m?croprolactinoma* OR epithelioma* OR adenocarcinoma* OR lintis plastica OR cylindroma* OR paget* OR gastrinoma* OR glucagonoma* OR somatostatinoma* OR hypernephroma* OR cholangiocarcinoma* OR cystadenocarcinoma* OR porocarcinoma* OR ((bowen* OR kahler*) n1 disease?) OR condyloma* OR pilomatrixoma* OR fibroadenos?s OR acanthoma* OR papilloma* OR gynandroblastoma* OR luteoma* OR arrhenoblastoma* OR androblastoma* OR meningioma* OR myeloma* OR macroglobulin?emia? OR plasmacytoma* OR angiofibroma* OR angiokeratoma* OR glomangioma* OR h?emangioendothelioma* OR h?emangioblastoma* OR kasabach merritt* OR parkes* OR sturge* OR weber* OR phakoma* OR h?emangiopericytoma* OR ameloblastoma* OR cementoma* OR odontoma* OR fibroodontoma* OR adamantinoma* OR nesidioblastoma* OR insulinoma* OR conn OR conns OR conn's OR meig* OR thecoma* OR ((paraneoplastic endocrine OR ectopic hormone OR nelson*) n1 syndrome?) OR neuroma* OR muir torre* OR pancoat* OR coin lesion? OR wilm* OR ccmmt? OR pnet? OR wdha? OR wdhh? OR dcis? OR mpnst? OR ibc?)

S12. S8 OR S9 OR S10 OR S11

S13. S7 AND S12

S14. (MH "Animal Studies")

S15. TI (malformation? OR inherit* OR familial OR hereditary OR infecti* OR bacterial OR viral OR pathogenic OR autoimmune OR auto immune OR injur* OR trauma* OR autosomal* OR x link* OR y link* OR chromasomal OR congenital OR hiv OR aids OR tb OR tuberculos?s OR rheumatic OR benign*) OR AB (malformation? OR inherit* OR familial OR hereditary OR infecti* OR bacterial OR viral OR pathogenic OR autoimmune OR auto immune OR injur* OR trauma* OR autosomal* OR x link* OR y link* OR chromasomal OR congenital OR hiv OR aids OR tb OR tuberculos?s OR rheumatic OR benign*)

S16. S13 NOT (S14 OR S15)

S17. ((MH Cause of Death) OR (MH Risk Factors) OR (MH Risk Assessment) OR (MH Incidence) OR (MH Prognosis) OR (MH Survival Analysis) OR (MH Life Expectancy) OR (MH Cox Proportional Hazards Models) OR (MH Mortality))

S18. TI risk* OR TI ((risk* n1 factor?) OR (risk* n1 assessment) OR (risk* n1 marker?) OR (risk n1 scor*) OR (predict OR predictor? OR prediction?) OR incidence OR hazard OR epidemiol* OR cohort OR trial? OR rct OR (prognosis OR prognostic) OR (mortality OR death? OR survival) OR life expectancy OR follow up) OR AB ((risk* n1 factor?) OR (risk* n1 assessment) OR (risk* n1 marker?) OR (risk n1 scor*) OR (predict OR predictor? OR prediction?) OR incidence OR hazard OR epidemiol* OR cohort OR trial? OR rct OR (prognosis OR prognostic) OR (mortality OR death? OR survival) OR life expectancy OR follow up)

S19. S17 OR S18

S20. (S17 OR S18) AND (S16 AND S19)

Appendix B

Appendix B.1 - Systematic Review Search Strategies

MEDLINE

Type 2 Diabetes

[1] EXP non insulin dependent diabetes mellitus/

[2] (((type 2 OR type ii OR non insulin dependent OR noninsulin dependent OR adult onset OR late onset OR maturity onset) ADJ diabet*) OR t2d OR t2dm OR niddm).ti,ab.

[3] OR/1-2

Cardiovascular Mortality

[4] (cardiovascular diseases/ AND (mortality/ OR death/)) OR cardiovascular diseases/mo

[5] ((cv OR cardiovascular OR cardio vascular OR cvd) ADJ2 (death* OR mortalit* OR fatal* OR survival*)).ti,ab.

[6] OR/4-5

Cancer Mortality

[7] (neoplasms/ AND (mortality/ OR death/)) OR neoplasms/mo

[8] ((cancer* OR neoplas* OR malignan* OR tumo?r*) ADJ2 (death* OR mortalit* OR fatal* OR survival*)).ti,ab.

[9] OR/7-8

Combined RCT Filters

[10] randomized controlled trials as topic/ OR randomized controlled trial/ OR random allocation/ OR double blind method/ OR single blind method/ OR clinical trial/ OR EXP clinical trials as topic/ OR placebos/

[11] (clinical trial, phase i OR clinical trial, phase ii OR clinical trial, phase iii OR clinical trial, phase iv OR controlled clinical trial OR randomized controlled trial OR multicenter study OR clinical trial).pt.

[12] ((clinic* ADJ trial*) OR ((singl* OR doubl* OR treb* OR tripl*) ADJ (blind* OR mask*)) OR placebo* OR factorial* OR crossover* OR cross over* OR assign* OR allocat* or volunteer* OR randomi* control* trial* OR rct* OR (allocat* ADJ2 random*)).ti,ab.

[13] OR/10-12

Combined Cohort Study Filter

[14] epidemiologic studies/ OR EXP cohort studies/

[15] epidemiologic methods/

[16] limit 15 to yr="1971 - 1988"

[17] ((cohort ADJ (study OR studies OR analy*)) OR ((epidemiologic* OR follow up OR observational) ADJ (study OR studies)) OR longitudinal OR retrospective).ti,ab.

[18] OR/16-17, 14

Combined Irrelevant Study Types

[19] letter/ OR historical article/ OR EXP case control studies/ OR cross-sectional studies/

[20] (comment OR editorial OR meta-analysis OR practice-guideline OR review OR letter OR journal correspondence).pt.

[21] case report.tw.

[22] OR/19-21

Human Study Filter

[23] EXP animals/ NOT humans/

Compiling

[24] 6 OR 9 ~ {CVD or Cancer}

[25] 3 AND 24 ~ {Diabetes and CVD/Cancer}

[26] 13 OR 18 ~ {RCT or Cohort}

[27] 26 NOT 22 ~ {RCT/Cohort, not Irrelevant Study Types}

[28] 27 AND 25 ~ {Final Search}

[29] 28 NOT 23 ~ {Drop Animal Studies}

EMBASE

Type 2 Diabetes

[1] EXP *non insulin dependent diabetes mellitus/

[2] (((type 2 OR type ii OR non insulin dependent OR noninsulin dependent OR adult onset OR late onset OR maturity onset) ADJ diabet*) OR t2d OR t2dm OR niddm).ti,ab.

[3] OR/1-2

Cardiovascular Mortality

[4] (*cardiovascular disease/ AND (mortality/ OR death/)) OR cardiovascular mortality/

[5] ((cv OR cardiovascular OR cardio vascular OR cvd) ADJ2 (death* OR mortalit* OR fatal* OR survival*)).ti,ab.

[6] OR/4-5

Cancer Mortality

[7] (*neoplasm/ AND (mortality/ OR Death/)) OR cancer mortality/

[8] ((cancer* OR neoplas* OR malignan* OR tumo?r*) ADJ2 (death* OR mortalit* OR fatal* OR survival*)).ti,ab.

[9] OR/7-8

Combined RCT Filter

[10] clinical trial/ OR crossover-procedure/ OR double-blind procedure/ OR randomization/ OR randomized controlled trial/ OR single-blind procedure/ OR placebo/ OR prospective study/

[11] ((clinic* ADJ trial*) OR ((singl* OR doubl* OR treb* OR tripl*) ADJ (blind* OR mask*)) OR placebo* OR factorial* OR crossover* OR cross over* OR assign* OR allocat* or volunteer* OR randomi* control* trial* OR rct* OR (allocat* ADJ2 random*)).ti,ab.

[12] OR/10-11

Combined Observational Study Filter

[13] clinical study/ OR cohort analysis/ OR family study/ OR follow up/ OR longitudinal study/ OR retrospective study/ OR prospective study/

[14] ((cohort ADJ (study OR studies OR analy*)) OR ((epidemiologic* OR follow up OR observational) ADJ (study OR studies)) OR longitudinal OR retrospective).ti,ab.

[15] OR/13-14

Combined Irrelevant Study Types

[16] case study/ OR abstract report/ OR letter/

[17] (book OR conference paper OR editorial OR letter OR review).pt.

[18] case report.tw.

[19] OR/16-18

Human Study Filter

[20] animals/ NOT humans/

Compiling

[21] 6 OR 9 ~ {CVD or Cancer}

[22] 3 AND 21 ~ {Diabetes and CVD/Cancer}

[23] 12 OR 15 ~ {RCT or Cohort}

[24] 23 NOT 19 ~ {RCT/Cohort, not Irrelevant Study Types}

[25] 24 AND 22 ~ {Final Search}

[26] 25 NOT 20 ~ {Drop Animal Studies}

CINAHL

Type 2 Diabetes

[S1] MH ("diabetes mellitus, type 2")

[S2] TI (((type 2 OR type ii OR noninsulin dependent OR non insulin dependent OR adult onset OR late onset OR maturity onset) N1 diabet*) OR t2d OR t2dm OR niddm)

[S3] AB (((type 2 OR type ii OR noninsulin dependent OR non-insulin dependent OR adult onset OR late onset OR maturity onset) N1 diabet*) OR t2d OR t2dm OR niddm)

[S4] S1 OR S2 OR S3

Cardiovascular Mortality

[S5] MM "cardiovascular diseases" AND (MH "mortality" OR MH "death" OR MH "fatal outcome")

[S6] TI ((cv OR cardiovascular OR cardio vascular OR cvd) N2 (death* OR mortalit* OR fatal* OR survival*))

[S7] AB ((cv OR cardiovascular OR cardio vascular OR cvd) N2 (death* OR mortalit* OR fatal* OR survival*))

[S8] S5 OR S6 OR S7

Cancer Mortality

[S9] MM "neoplasms" AND (MH "mortality" OR MH "death" OR MH "fatal outcome")

[S10] TI ((neoplas* OR tumo#r* OR cancer* OR malignan*) N2 (death* OR mortalit* OR fatal* OR survival*))

[S11] AB ((neoplas* OR tumo#r* OR cancer* OR malignan*) N2 (death* OR mortalit* OR fatal* OR survival*))

[S12] S9 OR S10 OR S11

Combined RCT Filter

[S13] MH ("clinical trials+" OR "random assignment" OR "placebos" OR "quantitative studies")

[S14] PT (clinical trial)

[S15] TI ((clinic* N1 trial*) OR ((singl* OR doubl* OR treb* OR tripl*) N1 (blind* OR mask*)) OR placebo* OR factorial* OR crossover* OR cross over* OR assign* OR allocat* or volunteer* OR randomi* control* trial* OR rct* OR (allocat* N2 random*))

[S16] AB ((clinic* N1 trial*) OR ((singl* OR doubl* OR treb* OR tripl*) N1 (blind* OR mask*)) OR placebo* OR factorial* OR crossover* OR cross over* OR assign* OR allocat* or volunteer* OR randomi* control* trial* OR rct* OR (allocat* N2 random*))

[S17] S13 OR S14 OR S15 OR S16

Combined Cohort Study Filter

[S18] MH ("prospective studies" OR "correlational studies" OR "non-concurrent prospective studies")

[S19] TI ((cohort N1 (study OR studies OR analy*)) OR ((epidemiologic* OR follow up OR observational) N1 (study OR studies)) OR longitudinal OR retrospective)

[S20] AB ((cohort N1 (study OR studies OR analy*)) OR ((epidemiologic* OR follow up OR observational) N1 (study OR studies)) OR longitudinal OR retrospective)

[S21] S18 OR S19 OR S20

Human Study Filter

[S22] (MH "Animals+") NOT (MH "Human")

Compiling

[S23] S8 OR S12 ~ {CVD or Cancer}

[S24] S4 AND S23 ~ {Diabetes and CVD/Cancer}

[S25] S17 OR S21 ~ {RCT or Observational Study}

[S26] S24 AND S25 ~ {Final Search}

[S27] S26 NOT S22 ~ {Drop Animal Studies}

Appendix B.2 - Fixed-Effect Meta-Analyses

Studies Pooled via Fixed Effect Meta-Analysis for the Primary Analysis

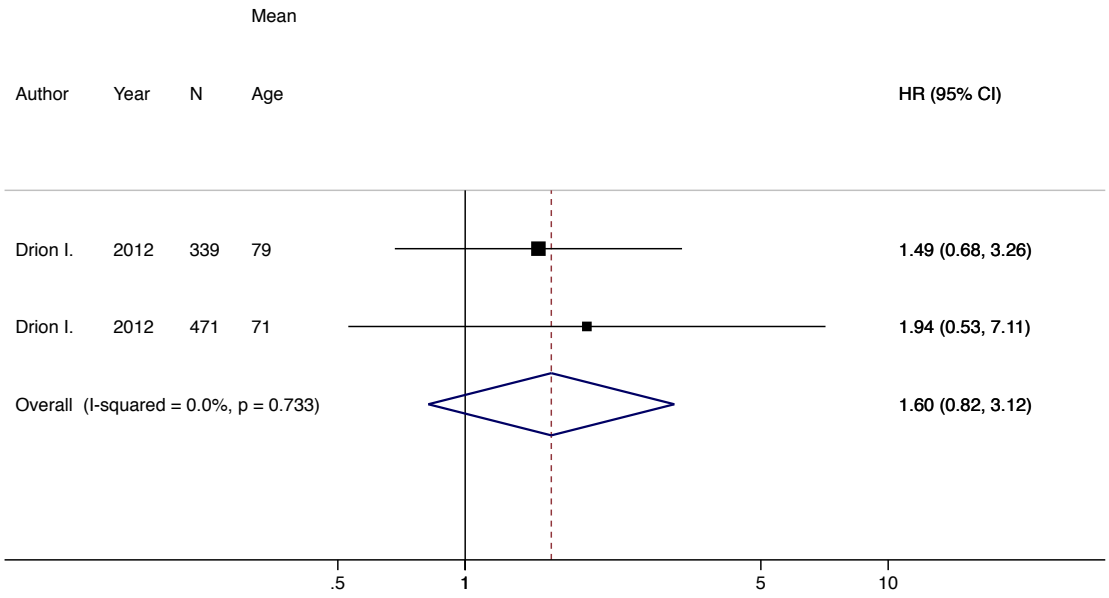


Figure B.2.a - Fixed effect meta-analysis to pool within-study age-stratified risk estimates for cardiovascular mortality associated with albuminuria for Drion I et al., 2012[94].

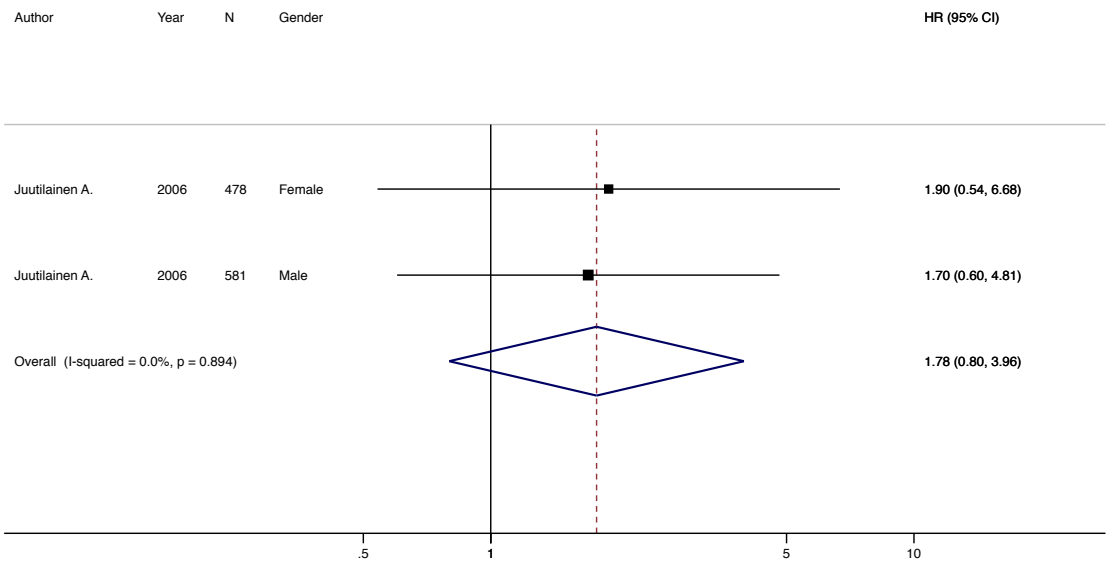


Figure B.2.b - Fixed effect meta-analysis pooling within-study gender-stratified risk estimates for cardiovascular mortality associated with albuminuria for Juutilainen A et al., 2006[98].

Studies Pooled via Fixed Effect Meta-Analysis for Secondary Analyses

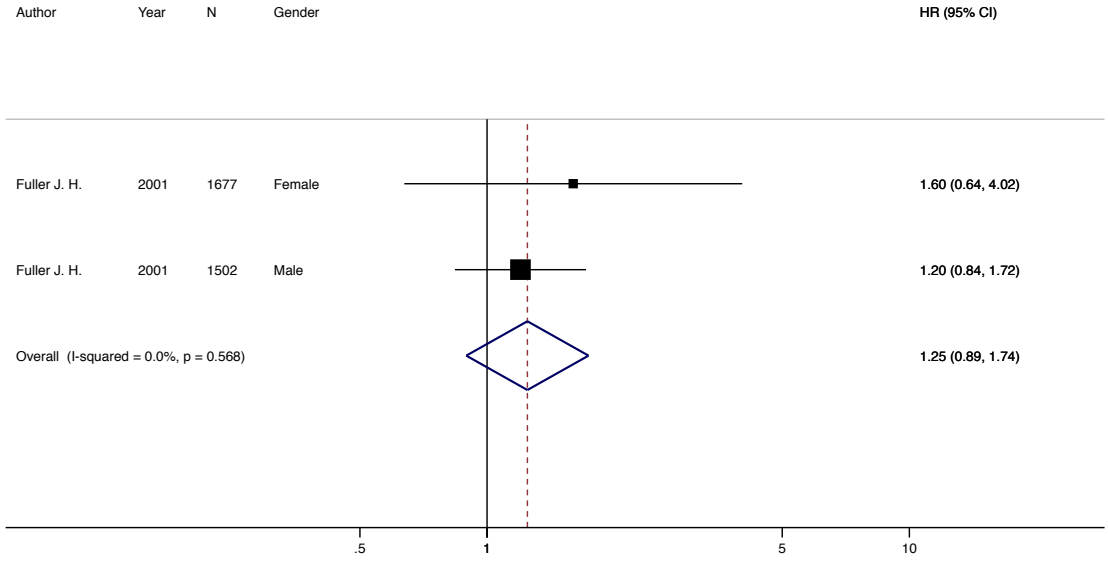


Figure B.2.c - Fixed effect meta-analysis pooling within-study gender-stratified risk estimates for cardiovascular mortality associated with microalbuminuria for Fuller J et al., 2001[113].

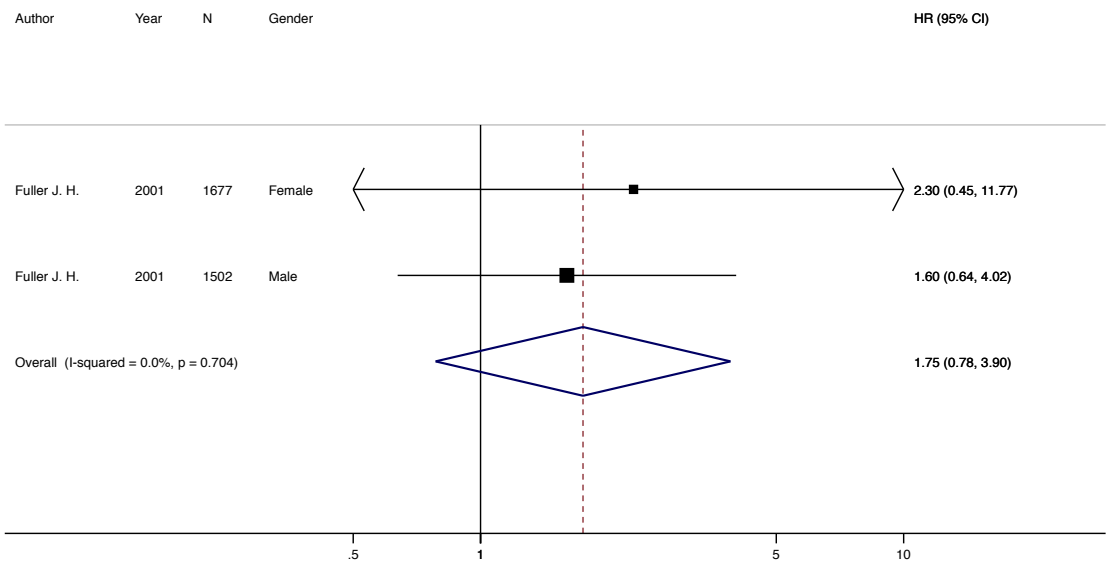


Figure B.2.d - Fixed effect meta-analysis pooling within-study gender-stratified risk estimates for cardiovascular mortality associated with macroalbuminuria for Fuller J et al., 2001[113].

Appendix C

Appendix C.1 - Medical Browser Search Terms

Diabetes

Search terms used under "Read Term":

diab or *insulin* or *bronze* or *haemochromatos*

Polycystic Ovary Syndrome

Search terms used under "Read Term":

pcos or *polycystic*ovar*

Appendix C.2 - Product Browser Search Terms

Anti-Diabetic Medication

Search terms used under "Drug Substance Name":

metform or *insulin* or *glitazone* or *acetohexamide* or *chlorpropamide* or *glibenclamide* or *glibornuride* or *gliclazide* or *glimepiride* or *glipizide* or *gliquidone* or *glymidine* or *tolazamide* or *tolbutamide* or *nateglinide* or *repaglinide* or *acarbose* or *exenatide* or *liraglutide* or *gliptin* or *dapagliflozin* or *guar gum*

Search terms used under "BNF Header":

biguanides or *antidiabetic* or *sulphonylureas* or *insulin* or *diabetes*

Search terms used under "Product Name":

metform or *insulin* or *glitazone* or *acetohexamide* or *chlorpropamide* or *glibenclamide* or *glibornuride* or *gliclazide* or *glimepiride* or *glipizide* or *gliquidone* or *glymidine* or *tolazamide* or *tolbutamide* or *nateglinide* or *repaglinide* or *acarbose* or *exenatide* or *liraglutide* or *gliptin* or *dapagliflozin* or *guar gum*

Appendix C.3 - Medical Codes for Diabetes Mellitus

Type 1 Diabetes

medcode	readcode	readterm	n
1549	C10E.00	Type 1 diabetes mellitus	1144
17858	C108.12	Type 1 diabetes mellitus	177
85660	66An.00	Diabetes type 1 review	52
10692	C10EM00	Type 1 diabetes mellitus with ketoacidosis	42
10418	C10ED00	Type 1 diabetes mellitus with nephropathy	24
30294	C10EL00	Type 1 diabetes mellitus with persistent microalbuminuria	24
30323	C10EK00	Type 1 diabetes mellitus with persistent proteinuria	21
24423	C108.13	Type I diabetes mellitus	9
18387	C10E700	Type 1 diabetes mellitus with retinopathy	7
35288	C10E800	Type 1 diabetes mellitus - poor control	7
55239	C10EQ00	Type 1 diabetes mellitus with gastroparesis	6
18683	C10E500	Type 1 diabetes mellitus with ulcer	2
21983	C108012	Type 1 diabetes mellitus with renal complications	2
22871	C10EP00	Type 1 diabetes mellitus with exudative maculopathy	2
40837	C10EN00	Type 1 diabetes mellitus with ketoacidotic coma	2
54008	C10EJ00	Type 1 diabetes mellitus with neuropathic arthropathy	2
104254	7L10011	Subcutaneous infusion with insulin pump	2
12455	C10E.11	Type I diabetes mellitus	1
18230	C108J12	Type 1 diabetes mellitus with neuropathic arthropathy	1
38161	C108711	Type I diabetes mellitus with retinopathy	1
39070	C10EE00	Type 1 diabetes mellitus with hypoglycaemic coma	1
43921	C10E400	Unstable type 1 diabetes mellitus	1
46850	C108811	Type I diabetes mellitus - poor control	1
47649	C10E100	Type 1 diabetes mellitus with ophthalmic complications	1
47650	C10E300	Type 1 diabetes mellitus with multiple complications	1
49554	C10EF00	Type 1 diabetes mellitus with diabetic cataract	1
49949	C10E411	Unstable type I diabetes mellitus	1
99231	C108B11	Type I diabetes mellitus with mononeuropathy	1
17545	C108F11	Type I diabetes mellitus with diabetic cataract	0

medcode	readcode	readterm	n
18642	C10EH00	Type 1 diabetes mellitus with arthropathy	0
40682	C10E900	Type 1 diabetes mellitus maturity onset	0
41049	C108712	Type 1 diabetes mellitus with retinopathy	0
42729	C108E11	Type I diabetes mellitus with hypoglycaemic coma	0
42831	C10E200	Type 1 diabetes mellitus with neurological complications	0
45914	C108812	Type 1 diabetes mellitus - poor control	0
46301	C10EC00	Type 1 diabetes mellitus with polyneuropathy	0
47582	C10E000	Type 1 diabetes mellitus with renal complications	0
49146	C108211	Type I diabetes mellitus with neurological complications	0
51957	C108511	Type I diabetes mellitus with ulcer	0
60107	C108411	Unstable type I diabetes mellitus	0
60208	C108J11	Type I diabetes mellitus with neuropathic arthropathy	0
61344	C108011	Type I diabetes mellitus with renal complications	0
61829	C108212	Type 1 diabetes mellitus with neurological complications	0
62209	C10EM11	Type I diabetes mellitus with ketoacidosis	0
62352	C108H11	Type I diabetes mellitus with arthropathy	0
62613	C10EA11	Type I diabetes mellitus without complication	0
63017	C108911	Type I diabetes mellitus maturity onset	0
66145	C10EN11	Type I diabetes mellitus with ketoacidotic coma	0
66872	C108D11	Type I diabetes mellitus with nephropathy	0
68105	C10EB00	Type 1 diabetes mellitus with mononeuropathy	0
68390	C108512	Type 1 diabetes mellitus with ulcer	0
68517	C10J.00	Insulin autoimmune syndrome	0
69043	ZC2C900	Dietary advice for type I diabetes	0
69676	C10EA00	Type 1 diabetes mellitus without complication	0
69993	C10E600	Type 1 diabetes mellitus with gangrene	0
70766	C108E12	Type 1 diabetes mellitus with hypoglycaemic coma	0
91942	C10E311	Type I diabetes mellitus with multiple complications	0
91943	C10EC11	Type I diabetes mellitus with polyneuropathy	0
93468	C10EG00	Type 1 diabetes mellitus with peripheral angiopathy	0
93878	C10E511	Type I diabetes mellitus with ulcer	0
95343	C10E711	Type I diabetes mellitus with retinopathy	0
95992	C108A11	Type I diabetes mellitus without complication	0
96235	C10E911	Type I diabetes mellitus maturity onset	0
97446	C108912	Type 1 diabetes mellitus maturity onset	0
97474	C108412	Unstable type 1 diabetes mellitus	0
97894	C10EP11	Type I diabetes mellitus with exudative maculopathy	0
99311	C10E111	Type I diabetes mellitus with ophthalmic complications	0
102112	C10E611	Type I diabetes mellitus with gangrene	0
102620	C10EL11	Type I diabetes mellitus with persistent microalbuminuria	0
102704	66At000	Type I diabetic dietary review	0
102740	C108112	Type 1 diabetes mellitus with ophthalmic complications	0

medcode	readcode	readterm	n
104453	66At011	Type 1 diabetic dietary review	0
105337	C10E811	Type I diabetes mellitus - poor control	0
108007	C108311	Type I diabetes mellitus with multiple complications	0
108724	C10EQ11	Type I diabetes mellitus with gastroparesis	0

Type 2 Diabetes

medcode	readcode	readterm	n
758	C10F.00	Type 2 diabetes mellitus	38645
1684	66A4.00	Diabetic on oral treatment	11732
7563	66A3.00	Diabetic on diet only	5714
4513	C109.00	Non-insulin dependent diabetes mellitus	5612
506	C100112	Non-insulin dependent diabetes mellitus	4881
17859	C109.12	Type 2 diabetes mellitus	3302
28769	66AV.00	Diabetic on insulin and oral treatment	955
83532	66Ao.00	Diabetes type 2 review	948
18390	C10FM00	Type 2 diabetes mellitus with persistent microalbuminuria	645
11047	66AH000	Conversion to insulin	600
1407	C10FJ00	Insulin treated Type 2 diabetes mellitus	476
93657	8Hj4.00	Referral to DESMOND diabetes structured education programme	324
26054	C10FL00	Type 2 diabetes mellitus with persistent proteinuria	297
5884	C109.11	NIDDM - Non-insulin dependent diabetes mellitus	269
18278	C109J00	Insulin treated Type 2 diabetes mellitus	193
18219	C109.13	Type II diabetes mellitus	185
12640	C10FC00	Type 2 diabetes mellitus with nephropathy	144
22884	C10F.11	Type II diabetes mellitus	78
25041	ZC2CA00	Dietary advice for type II diabetes	69
8403	C109700	Non-insulin dependent diabetes mellitus - poor control	63
25627	C10F700	Type 2 diabetes mellitus - poor control	63
95159	9NiD.00	Did not attend DESMOND diabetes structured education program	49
93529	9OLK.00	DESMOND diabetes structured education programme completed	39
18496	C10F600	Type 2 diabetes mellitus with retinopathy	36
32627	C10FN00	Type 2 diabetes mellitus with ketoacidosis	33
47954	C10F900	Type 2 diabetes mellitus without complication	27
35397	C1A..00	Insulin resistance	25
34450	C10FK00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus	20
18777	C10F000	Type 2 diabetes mellitus with renal complications	16
53392	C10F911	Type II diabetes mellitus without complication	12
58159	8I3k.00	Insulin therapy declined	10
63690	C10FR00	Type 2 diabetes mellitus with gastroparesis	10
36695	C10D.00	Diabetes mellitus autosomal dominant type 2	9
18425	C10FB00	Type 2 diabetes mellitus with polyneuropathy	8

medcode	readcode	readterm	n
25591	C10FQ00	Type 2 diabetes mellitus with exudative maculopathy	7
34268	C10F200	Type 2 diabetes mellitus with neurological complications	7
35385	C10FH00	Type 2 diabetes mellitus with neuropathic arthropathy	7
46917	C10FD00	Type 2 diabetes mellitus with hypoglycaemic coma	7
18264	C109J12	Insulin treated Type II diabetes mellitus	6
49074	C10F400	Type 2 diabetes mellitus with ulcer	6
44982	C10FE00	Type 2 diabetes mellitus with diabetic cataract	5
29979	C109900	Non-insulin-dependent diabetes mellitus without complication	4
34912	C109400	Non-insulin dependent diabetes mellitus with ulcer	4
45919	C109212	Type 2 diabetes mellitus with neurological complications	4
55123	66AO.00	Date diabetic treatment stopp.	4
62674	C10FA00	Type 2 diabetes mellitus with mononeuropathy	4
12736	C10F500	Type 2 diabetes mellitus with gangrene	3
17262	C109600	Non-insulin-dependent diabetes mellitus with retinopathy	3
24836	C109C12	Type 2 diabetes mellitus with nephropathy	3
37806	C10FF00	Type 2 diabetes mellitus with peripheral angiopathy	3
47315	C10F711	Type II diabetes mellitus - poor control	3
47321	C10F100	Type 2 diabetes mellitus with ophthalmic complications	2
48192	C109E11	Type II diabetes mellitus with diabetic cataract	2
52303	C109000	Non-insulin-dependent diabetes mellitus with renal comps	2
53238	66AG.00	Diabetic drug side effects	2
59365	C109C00	Non-insulin dependent diabetes mellitus with nephropathy	2
61071	C109D12	Type 2 diabetes mellitus with hypoglycaemic coma	2
100791	66Ar.00	Insulin treatment stopped	2
40962	C109H00	Non-insulin dependent d m with neuropathic arthropathy	1
42762	C109612	Type 2 diabetes mellitus with retinopathy	1
43785	C109D00	Non-insulin dependent diabetes mellitus with hypoglyca coma	1
45467	C109B00	Non-insulin dependent diabetes mellitus with polyneuropathy	1
45913	C109712	Type 2 diabetes mellitus - poor control	1
47409	C109B11	Type II diabetes mellitus with polyneuropathy	1
50429	C109100	Non-insulin-dependent diabetes mellitus with ophthalm comps	1
50609	L180600	Pre-existing diabetes mellitus, non-insulin-dependent	1
51939	ZV6DB00	[V]Admitted for conversion to insulin	1
54212	C109F00	Non-insulin-dependent d m with peripheral angiopath	1
54899	C109F11	Type II diabetes mellitus with peripheral angiopathy	1
55075	C109411	Type II diabetes mellitus with ulcer	1
55842	C109200	Non-insulin-dependent diabetes mellitus with neuro comps	1
59253	C10FG00	Type 2 diabetes mellitus with arthropathy	1
62146	C109300	Non-insulin-dependent diabetes mellitus with multiple comps	1
65267	C10F300	Type 2 diabetes mellitus with multiple complications	1
65704	C109412	Type 2 diabetes mellitus with ulcer	1
66965	C109H12	Type 2 diabetes mellitus with neuropathic arthropathy	1

medcode	readcode	readterm	n
98616	C10F211	Type II diabetes mellitus with neurological complications	1
18143	C109G11	Type II diabetes mellitus with arthropathy	0
18209	C109012	Type 2 diabetes mellitus with renal complications	0
24458	C109711	Type II diabetes mellitus - poor control	0
24693	C109G00	Non-insulin dependent diabetes mellitus with arthropathy	0
36633	C109K00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus	0
37648	C109J11	Insulin treated non-insulin dependent diabetes mellitus	0
37957	C10K.00	Type A insulin resistance	0
40401	C109500	Non-insulin dependent diabetes mellitus with gangrene	0
43139	C102100	Diabetes mellitus, adult onset, with hyperosmolar coma	0
43227	C10F311	Type II diabetes mellitus with multiple complications	0
44779	C109E12	Type 2 diabetes mellitus with diabetic cataract	0
46150	C109512	Type 2 diabetes mellitus with gangrene	0
47816	C109H11	Type II diabetes mellitus with neuropathic arthropathy	0
49655	C10F611	Type II diabetes mellitus with retinopathy	0
49869	C109G12	Type 2 diabetes mellitus with arthropathy	0
50225	C109011	Type II diabetes mellitus with renal complications	0
50527	C10FB11	Type II diabetes mellitus with polyneuropathy	0
50813	C109A11	Type II diabetes mellitus with mononeuropathy	0
51756	C10FP00	Type 2 diabetes mellitus with ketoacidotic coma	0
56268	C109D11	Type II diabetes mellitus with hypoglycaemic coma	0
56803	C107400	NIDDM with peripheral circulatory disorder	0
57278	C10F011	Type II diabetes mellitus with renal complications	0
58604	C109611	Type II diabetes mellitus with retinopathy	0
59725	C109111	Type II diabetes mellitus with ophthalmic complications	0
60699	C109F12	Type 2 diabetes mellitus with peripheral angiopathy	0
60796	C10FL11	Type II diabetes mellitus with persistent proteinuria	0
62107	C109511	Type II diabetes mellitus with gangrene	0
64571	C109C11	Type II diabetes mellitus with nephropathy	0
64668	C10FJ11	Insulin treated Type II diabetes mellitus	0
67905	C109211	Type II diabetes mellitus with neurological complications	0
69278	C109E00	Non-insulin depend diabetes mellitus with diabetic cataract	0
70316	C109112	Type 2 diabetes mellitus with ophthalmic complications	0
72320	C109A00	Non-insulin dependent diabetes mellitus with mononeuropathy	0
85991	C10FM11	Type II diabetes mellitus with persistent microalbuminuria	0
91646	C10F411	Type II diabetes mellitus with ulcer	0
93727	C10FE11	Type II diabetes mellitus with diabetic cataract	0
95093	8I83.00	Did not complete DESMOND diabetes structured educat program	0
95351	C10FA11	Type II diabetes mellitus with mononeuropathy	0
98723	C10FD11	Type II diabetes mellitus with hypoglycaemic coma	0
100964	C10F111	Type II diabetes mellitus with ophthalmic complications	0
101801	66At100	Type II diabetic dietary review	0

medcode	readcode	readterm	n
102201	C10FC11	Type II diabetes mellitus with nephropathy	0
102611	66At111	Type 2 diabetic dietary review	0
103902	C10FG11	Type II diabetes mellitus with arthropathy	0
104323	C10F511	Type II diabetes mellitus with gangrene	0
104639	C10FF11	Type II diabetes mellitus with peripheral angiopathy	0
105784	C109912	Type 2 diabetes mellitus without complication	0
106061	C10FP11	Type II diabetes mellitus with ketoacidotic coma	0
106528	C10FN11	Type II diabetes mellitus with ketoacidosis	0
107331	66AH100	Conversion to insulin in secondary care	0
107508	66AH200	Conversion to insulin by diabetes specialist nurse	0
107701	C10FK11	Hyperosmolar non-ketotic state in type II diabetes mellitus	0
108005	C109312	Type 2 diabetes mellitus with multiple complications	0
109103	C109911	Type II diabetes mellitus without complication	0
109197	C10FH11	Type II diabetes mellitus with neuropathic arthropathy	0

Gestational Diabetes

medcode	readcode	readterm	n
8446	L180811	Gestational diabetes mellitus	143
2664	L180900	Gestational diabetes mellitus	35
11359	L180.00	Diabetes mellitus during pregnancy/childbirth/puerperium	13
10278	L180800	Diabetes mellitus arising in pregnancy	3
94777	ZV13F00	[V]Personal history of gestational diabetes mellitus	2
46577	66AX.00	Diabetes: shared care in pregnancy - diabetol and obstet	1
49559	L180300	Diabetes mellitus during pregnancy - baby not yet delivered	1
34639	L180100	Diabetes mellitus during pregnancy - baby delivered	0
64384	L180z00	Diabetes mellitus in pregnancy/childbirth/puerperium NOS	0
67635	L180000	Diabetes mellitus - unspec whether in pregnancy/puerperium	0
96823	L180400	Diabetes mellitus in pueperium - baby previously delivered	0
104588	66Ay.00	Gestational diabetes mellitus annual review	0
108013	ZC2CB00	Dietary advice for gestational diabetes	0

Drug-Induced Diabetes

medcode	readcode	readterm	n
11551	C10B.00	Diabetes mellitus induced by steroids	69
32193	C11y000	Steroid induced diabetes	16
26108	C10B000	Steroid induced diabetes mellitus without complication	13
61122	C10H.00	Diabetes mellitus induced by non-steroid drugs	2

Bronzed Diabetes

medcode	readcode	readterm	n
2983	C350000	Haemochromatosis	336
23479	C350011	Bronzed diabetes	1
106557	Q47y100	Bronze baby	0

Monogenic Diabetes

medcode	readcode	readterm	n
46624	C10C.11	Maturity onset diabetes in youth	1
98392	C10C.12	Maturity onset diabetes in youth type 1	1
21472	Q441.00	Neonatal diabetes mellitus	0
59991	C10D.11	Maturity onset diabetes in youth type 2	0

Malnutrition Diabetes

medcode	readcode	readterm	n
33969	C10A100	Malnutrition-related diabetes mellitus with ketoacidosis	4
52236	C10A.00	Malnutrition-related diabetes mellitus	2
66675	C10A000	Malnutrition-related diabetes mellitus with coma	0
100347	C10A500	Malnutritn-relat diabetes melitus wth periph circul complctn	0
109133	L180700	Pre-existing malnutrition-related diabetes mellitus	0

Undefined Diabetes

medcode	readcode	readterm	n
9897	9OL..00	Diabetes monitoring admin	150027
3550	66A..00	Diabetic monitoring	124914
6125	66AS.00	Diabetic annual review	122651
2379	9N1Q.00	Seen in diabetic clinic	50947
13194	9OL4.00	Diabetes monitoring 1st letter	38354
711	C10..00	Diabetes mellitus	36678
608	66A2.00	Follow-up diabetic assessment	25648
11094	9NND.00	Under care of diabetic foot screener	21645
26666	2G5E.00	O/E - Right diabetic foot at low risk	20875
26667	2G5I.00	O/E - Left diabetic foot at low risk	20677
12506	66AP.00	Diabetes: practice programme	16358
13067	66AZ.00	Diabetic monitoring NOS	13204
18311	68A7.00	Diabetic retinopathy screening	11112
10977	66Ac.00	Diabetic peripheral neuropathy screening	9302
2378	66AJ.00	Diabetic - poor control	9008

medcode	readcode	readterm	n
95994	66Aq.00	Diabetic foot screen	8566
13197	9OL1.00	Attends diabetes monitoring	7619
13195	9OL5.00	Diabetes monitoring 2nd letter	7544
8836	66AR.00	Diabetes management plan given	6562
17817	7L19800	Subcutaneous injection of insulin	6212
1323	F420.00	Diabetic retinopathy	5961
31157	2G5F.00	O/E - Right diabetic foot at moderate risk	5823
31156	2G5J.00	O/E - Left diabetic foot at moderate risk	5611
7069	F420000	Background diabetic retinopathy	4806
13100	2BBJ.00	O/E - no right diabetic retinopathy	4232
13104	2BBK.00	O/E - no left diabetic retinopathy	4087
11348	9h42.00	Excepted from diabetes quality indicators: Informed dissent	3904
13192	9OLA.00	Diabetes monitor. check done	3904
28873	66Ai.00	Diabetic 6 month review	3652
11041	9h41.00	Excepted from diabetes qual indicators: Patient unsuitable	3459
11471	8B3I.00	Diabetes medication review	3247
8842	66A5.00	Diabetic on insulin	3246
12213	8BL2.00	Patient on maximal tolerated therapy for diabetes	3191
12030	9OL6.00	Diabetes monitoring 3rd letter	3176
9145	9N4I.00	DNA - Did not attend diabetic clinic	3089
13074	13B1.00	Diabetic diet	2928
13071	66AI.00	Diabetic - good control	2841
9974	9N1v.00	Seen in diabetic eye clinic	2629
22823	66Ab.00	Diabetic foot examination	2625
47032	8CS0.00	Diabetes care plan agreed	2551
11433	2BBP.00	O/E - right eye background diabetic retinopathy	2425
11129	2BBQ.00	O/E - left eye background diabetic retinopathy	2387
6430	9NM0.00	Attending diabetes clinic	2082
13196	66AD.00	Fundoscopy - diabetic check	2000
6813	1434	H/O: diabetes mellitus	1803
8414	8CA4100	Pt advised re diabetic diet	1720
13057	679L.00	Health education - diabetes	1715
50175	66AW.00	Diabetic foot risk assessment	1628
9958	42W..00	Hb. A1C - diabetic control	1594
31171	2G5G.00	O/E - Right diabetic foot at high risk	1570
31172	2G5K.00	O/E - Left diabetic foot at high risk	1534
38078	66A9.00	Understands diet - diabetes	1525
20900	9OLA.11	Diabetes monitored	1513
21689	13AB.00	Diabetic lipid lowering diet	1498
12675	66AQ.00	Diabetes: shared care programme	1493
10824	9N1i.00	Seen in diabetic foot clinic	1434
3837	F420400	Diabetic maculopathy	1337

medcode	readcode	readterm	n
11677	8H7r.00	Refer to diabetic foot screener	1302
12262	8I3X.00	Diabetic retinopathy screening refused	1240
22130	9OL3.00	Diabetes monitoring default	1216
30648	9N4p.00	Did not attend diabetic retinopathy clinic	1202
17236	14P3.00	H/O: insulin therapy	1068
13070	66A1.00	Initial diabetic assessment	987
1038	C100011	Insulin dependent diabetes mellitus	953
1647	C108.00	Insulin dependent diabetes mellitus	936
94647	9Oy..00	Diabetes screening administration	891
10642	ZC2C800	Dietary advice for diabetes mellitus	886
12247	8I6G.00	Diabetic foot examination not indicated	865
31141	9OL8.00	Diabetes monitor.phone invite	832
18167	66AT.00	Annual diabetic blood test	775
52237	9360	Patient held diabetic record issued	740
20696	66AA.11	Injection sites - diabetic	691
47144	2BBM.00	O/E - diabetic maculopathy absent both eyes	682
32619	66Af.00	Patient diabetes education review	642
18747	8I6F.00	Diabetic retinopathy screening not indicated	632
14889	C100111	Maturity onset diabetes	604
95124	9Oy0.00	Diabetes screening invitation	592
2342	F372.12	Diabetic neuropathy	576
13069	66A8.00	Has seen dietician - diabetes	543
12307	66AU.00	Diabetes care by hospital only	524
10755	F420600	Non proliferative diabetic retinopathy	522
16230	C106.00	Diabetes mellitus with neurological manifestation	522
18824	8I3W.00	Diabetic foot examination declined	519
7795	C106.12	Diabetes mellitus with neuropathy	504
47011	8Hj0.00	Referral to diabetes structured education programme	490
16490	66AH.00	Diabetic treatment changed	470
31241	9OLZ.00	Diabetes monitoring admin NOS	437
35383	9OLD.00	Diabetic patient unsuitable for digital retinal photography	423
3286	F420100	Proliferative diabetic retinopathy	364
19739	68A9.00	Diabetic retinopathy screening offered	363
31240	9OL7.00	Diabetes monitor.verbal invite	344
26664	2G5B.00	O/E - Left diabetic foot at risk	342
13191	9OL..11	Diabetes clinic administration	329
17095	2G5A.00	O/E - Right diabetic foot at risk	311
13078	13AC.00	Diabetic weight reducing diet	310
13102	2BBW.00	O/E - right eye diabetic maculopathy	293
12682	679R.00	Patient offered diabetes structured education programme	292
13108	2BBX.00	O/E - left eye diabetic maculopathy	275
11018	8HBG.00	Diabetic retinopathy 12 month review	271

medcode	readcode	readterm	n
28574	9h4..00	Exception reporting: diabetes quality indicators	267
93870	8Hj5.00	Referral to XPERT diabetes structured education programme	262
19381	8HTk.00	Referral to diabetic eye clinic	249
18066	8CE0.00	Diabetic leaflet given	247
14803	C100100	Diabetes mellitus, adult onset, no mention of complication	228
8306	8H7f.00	Referral to diabetes nurse	227
18505	C108.11	IDDM-Insulin dependent diabetes mellitus	217
7777	8H4F.00	Referral to diabetologist	214
103772	8CE0200	Insulin passport given	205
8618	ZLA2500	Seen by diabetic liaison nurse	203
64142	8HI1.00	Referral for diabetic retinopathy screening	199
2986	F420200	Preproliferative diabetic retinopathy	198
12225	8H7C.00	Refer, diabetic liaison nurse	190
26604	66AY.00	Diabetic diet - good compliance	181
9835	2BBL.00	O/E - diabetic maculopathy present both eyes	169
11599	7276	Pan retinal photocoagulation for diabetes	169
26605	9OLB.00	Attended diabetes structured education programme	167
54419	918T.00	Diabetes key contact	167
38986	C100.00	Diabetes mellitus with no mention of complication	164
1682	C101.00	Diabetes mellitus with ketoacidosis	160
32739	9N0n.00	Seen in community diabetes specialist clinic	160
2475	C104.11	Diabetic nephropathy	155
17067	F171100	Autonomic neuropathy due to diabetes	152
83485	66Am.00	Insulin dose changed	146
9881	M271200	Mixed diabetic ulcer - foot	143
103762	8BAi.00	Insulin passport completed	139
13099	2BBR.00	O/E - right eye preproliferative diabetic retinopathy	138
14050	42c..00	HbA1 - diabetic control	134
11977	ZL62500	Referral to diabetes nurse	131
82474	8HI4.00	Referral to community diabetes specialist nurse	131
13103	2BBS.00	O/E - left eye preproliferative diabetic retinopathy	123
11663	M271100	Neuropathic diabetic ulcer - foot	121
68818	ZRB5.11	DTSQ - Diabetes treatment satisfaction questionnaire	120
24327	M271000	Ischaemic ulcer diabetic foot	113
13097	2BBT.00	O/E - right eye proliferative diabetic retinopathy	107
38103	9N0m.00	Seen in diabetic nurse consultant clinic	96
46521	9N2d.00	Seen by diabetologist	94
13101	2BBV.00	O/E - left eye proliferative diabetic retinopathy	91
35399	C107.00	Diabetes mellitus with peripheral circulatory disorder	90
35316	2G5H.00	O/E - Right diabetic foot - ulcerated	86
35116	2G5L.00	O/E - Left diabetic foot - ulcerated	80
96010	66Ap.00	Insulin treatment initiated	77

medcode	readcode	readterm	n
94699	ZRB5.00	Diabetes treatment satisfaction questionnaire	70
26603	9OL2.00	Refuses diabetes monitoring	69
33254	C105.00	Diabetes mellitus with ophthalmic manifestation	69
11626	F420z00	Diabetic retinopathy NOS	63
25636	66Aa.00	Diabetic diet - poor compliance	63
27891	N030100	Diabetic Charcot arthropathy	61
16502	C104.00	Diabetes mellitus with renal manifestation	60
66475	66Ak.00	Diabetic monitoring - lower risk albumin excretion	55
13678	ZL62600	Referral to diabetic liaison nurse	54
38129	9N0o.00	Seen in community diabetic specialist nurse clinic	53
11930	9NN9.00	Under care of diabetes specialist nurse	51
12703	3881	Education score - diabetes	49
31790	F372.00	Polyneuropathy in diabetes	46
12507	9N2i.00	Seen by diabetic liaison nurse	44
93631	9OLL.00	XPERT diabetes structured education programme completed	44
7328	M037200	Cellulitis in diabetic foot	43
9013	66AJ.11	Unstable diabetes	43
96143	9kL..00	Insulin initiation - enhanced services administration	43
2478	66AJ100	Brittle diabetes	42
57389	93C4.00	Patient consent given for addition to diabetic register	41
94955	9NiE.00	Did not attend XPERT diabetes structured education programme	41
54601	9NN8.00	Under care of diabetologist	40
61470	66Al.00	Diabetic monitoring - higher risk albumin excretion	40
38130	ZRB6.00	Diabetes wellbeing questionnaire	38
24363	8A13.00	Diabetic stabilisation	37
103816	8CE0100	Insulin alert patient information booklet given	37
10099	F420300	Advanced diabetic maculopathy	35
61021	68AB.00	Diabetic digital retinopathy screening offered	35
17886	66AM.00	Diabetic - follow-up default	33
66274	66Ah.00	Insulin needles changed for each injection	33
22023	66AJz00	Diabetic - poor control NOS	31
2340	F381311	Diabetic amyotrophy	29
7059	8H2J.00	Admit diabetic emergency	29
37315	F3y0.00	Diabetic mononeuropathy	28
57723	8HHy.00	Referral to diabetic register	28
22573	C106z00	Diabetes mellitus NOS with neurological manifestation	27
27921	2G51000	Foot abnormality - diabetes related	27
18056	2G5C.00	Foot abnormality - diabetes related	26
16881	ZV65312	[V]Dietary counselling in diabetes mellitus	24
22967	2BBF.00	Retinal abnormality - diabetes related	24
35785	F372100	Chronic painful diabetic neuropathy	24
93854	9OLM.00	Diabetes structured education programme declined	23

medcode	readcode	readterm	n
47370	8HLE.00	Diabetology D.V. done	22
5002	F372.11	Diabetic polyneuropathy	21
17869	66AL.00	Diabetic-uncooperative patient	20
51697	C10G.00	Secondary pancreatic diabetes mellitus	19
38076	M21yC00	Insulin lipohypertrophy	18
94011	9OLG.00	Attended XPERT diabetes structured education programme	17
94186	9OLF.00	Diabetes structured education programme completed	17
6791	C108800	Insulin dependent diabetes mellitus - poor control	16
18662	8HBH.00	Diabetic retinopathy 6 month review	16
31310	C108900	Insulin dependent diabetes maturity onset	16
32770	44V3.00	Glucose tol. test diabetic	15
102768	9NiZ.00	Did not attend diabetes foot screening	15
18142	N030000	Diabetic cheiroarthropathy	14
2471	K01x100	Nephrotic syndrome in diabetes mellitus	13
10659	F464000	Diabetic cataract	12
29041	66AN.00	Date diabetic treatment start	11
43951	66AK.00	Diabetic - cooperative patient	11
103761	671F000	Insulin alert pat information booklet information discussed	11
32359	ZRbH.00	Perceived control of insulin-dependent diabetes	10
6509	C108700	Insulin dependent diabetes mellitus with retinopathy	9
7045	14F4.00	H/O: Admission in last year for diabetes foot problem	9
34152	G73y000	Diabetic peripheral angiopathy	9
36798	7L10000	Continuous subcutaneous infusion of insulin	9
102549	66Aw.00	Insulin dose	9
17247	F35z000	Diabetic mononeuritis NOS	8
35107	C104z00	Diabetes mellitus with nephropathy NOS	8
39809	C108J00	Insulin dependent diab mell with neuropathic arthropathy	8
47058	8Hg4.00	Discharged from care of diabetes specialist nurse	8
47328	2BBk.00	O/E - right eye stable treated prolif diabetic retinopathy	8
47341	8A12.00	Diabetic crisis monitoring	8
101455	9OLN.00	Diabetes monitor invitation by SMS (short message service)	8
24571	F372200	Asymptomatic diabetic neuropathy	7
35321	8H3O.00	Non-urgent diabetic admission	7
52041	2BBl.00	O/E - left eye stable treated prolif diabetic retinopathy	7
32403	C107.11	Diabetes mellitus with gangrene	6
45250	ZL22500	Under care of diabetic liaison nurse	6
46533	13Y1.00	Diabetic association member	6
49884	6761	Diabetic pre-pregnancy counselling	6
90301	66Ag.00	Insulin needles changed daily	6
100436	679L000	Education in self management of diabetes	6
21482	C102.00	Diabetes mellitus with hyperosmolar coma	5
22487	C10N.00	Secondary diabetes mellitus	5

medcode	readcode	readterm	n
30477	F420700	High risk proliferative diabetic retinopathy	5
36669	66b1.00	Diabetic monitoring not required	5
42505	C101z00	Diabetes mellitus NOS with ketoacidosis	5
43493	M21yC11	Insulin site lipohypertrophy	5
48078	F372000	Acute painful diabetic neuropathy	5
49640	2G5W.00	O/E - left chronic diabetic foot ulcer	5
58133	ZLD7500	Discharge by diabetic liaison nurse	5
63412	8CR2.00	Diabetes clinical management plan	5
93390	9OLH.00	Attended DAFNE diabetes structured education programme	5
94330	8H4e.00	Referral to diabetes special interest general practitioner	5
24490	C100000	Diabetes mellitus, juvenile type, no mention of complication	4
31053	R054300	[D]Widespread diabetic foot gangrene	4
32556	C107.12	Diabetes with gangrene	4
39317	C106100	Diabetes mellitus, adult onset, + neurological manifestation	4
51261	C10E.12	Insulin dependent diabetes mellitus	4
62384	2G5V.00	O/E - right chronic diabetic foot ulcer	4
17313	F440700	Diabetic iritis	3
53634	R054200	[D]Gangrene of toe in diabetic	3
58639	8I57.00	Patient held diabetic record declined	3
61523	C106y00	Other specified diabetes mellitus with neurological comps	3
69163	8HTi.00	Referral to multidisciplinary diabetic clinic	3
101177	66At.00	Diabetic dietary review	3
104669	8IF..00	Prof judgemnt not to engage pt wt insulin alert requirements	3
15690	C103.00	Diabetes mellitus with ketoacidotic coma	2
26855	C108400	Unstable insulin dependent diabetes mellitus	2
28856	8CP2.00	Transition of diabetes care options discussed	2
41716	C108C00	Insulin dependent diabetes mellitus with polyneuropathy	2
43857	C10M.00	Lipoatrophic diabetes mellitus	2
44033	F345000	Diabetic mononeuritis multiplex	2
48310	ZV6DA00	[V]Admitted for commencement of insulin	2
50937	8HTe.00	Referral to diabetes preconception counselling clinic	2
50972	C100z00	Diabetes mellitus NOS with no mention of complication	2
55431	L180X00	Pre-existing diabetes mellitus, unspecified	2
59903	C106.11	Diabetic amyotrophy	2
60499	C108600	Insulin dependent diabetes mellitus with gangrene	2
61210	TJ23z00	Adverse reaction to insulins and antidiabetic agents NOS	2
61557	8HKE.00	Diabetology D.V. requested	2
64357	C10zz00	Diabetes mellitus NOS with unspecified complication	2
65463	F420800	High risk non proliferative diabetic retinopathy	2
69152	66Aj.00	Insulin needles changed less than once a day	2
93491	9OLJ.00	DAFNE diabetes structured education programme completed	2
93704	8Hj3.00	Referral to DAFNE diabetes structured education programme	2

medcode	readcode	readterm	n
95094	8I81.00	Did not complete diabetes structured education programme	2
95813	9N1o.00	Seen in multidisciplinary diabetic clinic	2
101190	66AQ100	Declined consent for diabetes year of care programme	2
101728	66As.00	Diabetic on subcutaneous treatment	2
13279	C104y00	Other specified diabetes mellitus with renal complications	1
14049	42WZ.00	Hb. A1C - diabetic control NOS	1
16946	13L4.11	Diabetic child	1
33343	C10y.00	Diabetes mellitus with other specified manifestation	1
35105	C104100	Diabetes mellitus, adult onset, with renal manifestation	1
39420	F381300	Myasthenic syndrome due to diabetic amyotrophy	1
42567	C103000	Diabetes mellitus, juvenile type, with ketoacidotic coma	1
44260	C108F00	Insulin dependent diabetes mellitus with diabetic cataract	1
44440	C108E00	Insulin dependent diabetes mellitus with hypoglycaemic coma	1
44443	C108500	Insulin dependent diabetes mellitus with ulcer	1
45491	C10z.00	Diabetes mellitus with unspecified complication	1
47584	F420500	Advanced diabetic retinal disease	1
50960	L180500	Pre-existing diabetes mellitus, insulin-dependent	1
51066	9OLC.00	Family/carer attended diabetes structured education prog	1
52212	Cyu2.00	[X]Diabetes mellitus	1
52630	2BBo.00	O/E - sight threatening diabetic retinopathy	1
63357	C107100	Diabetes mellitus, adult, + peripheral circulatory disorder	1
63762	C10z100	Diabetes mellitus, adult onset, + unspecified complication	1
67664	ZRBa.00	Education score - diabetes	1
68928	TJ23.00	Adverse reaction to insulins and antidiabetic agents	1
94956	8I84.00	Did not complete XPERT diabetes structured education program	1
95553	9NiA.00	Did not attend diabetes structured education programme	1
97849	C10E912	Insulin dependent diabetes maturity onset	1
100422	8HgC.00	Discharged from diabetes shared care programme	1
100533	66AQ000	Unsuitable for diabetes year of care programme	1
103743	8IE2.00	Diabetes care plan declined	1
103817	8BAj.00	Informed dissent not to carry insulin passport	1
104287	8Hlc.00	Referral to community diabetes service	1
104858	8BAm.00	Insulin passport checked	1
10098	C10yy00	Other specified diabetes mellitus with other spec comps	0
16491	C106.13	Diabetes mellitus with polyneuropathy	0
24694	C108B00	Insulin dependent diabetes mellitus with mononeuropathy	0
30247	TJ23000	Adverse reaction to insulins	0
33807	C107200	Diabetes mellitus, adult with gangrene	0
34283	C105z00	Diabetes mellitus NOS with ophthalmic manifestation	0
34528	3882	Diabetes well being questionnaire	0
34541	8HVU.00	Private referral to diabetologist	0
38617	C101y00	Other specified diabetes mellitus with ketoacidosis	0

medcode	readcode	readterm	n
40023	C102000	Diabetes mellitus, juvenile type, with hyperosmolar coma	0
41389	C105100	Diabetes mellitus, adult onset, + ophthalmic manifestation	0
41686	Cyu2000	[X]Other specified diabetes mellitus	0
43453	C10C.00	Diabetes mellitus autosomal dominant	0
43542	43Gk.00	Insulin antibody level	0
45276	C10E312	Insulin dependent diabetes mellitus with multiple complicat	0
46290	C108y00	Other specified diabetes mellitus with multiple comps	0
46963	C108000	Insulin-dependent diabetes mellitus with renal complications	0
47377	C105y00	Other specified diabetes mellitus with ophthalmic complicatn	0
49276	C108100	Insulin-dependent diabetes mellitus with ophthalmic comps	0
52104	C108300	Insulin dependent diabetes mellitus with multiple complicatn	0
52283	C108200	Insulin-dependent diabetes mellitus with neurological comps	0
53200	C101000	Diabetes mellitus, juvenile type, with ketoacidosis	0
54600	C10E412	Unstable insulin dependent diabetes mellitus	0
54856	C101100	Diabetes mellitus, adult onset, with ketoacidosis	0
56448	C108A00	Insulin-dependent diabetes without complication	0
57333	N030011	Diabetic cheiropathy	0
57621	C108D00	Insulin dependent diabetes mellitus with nephropathy	0
59288	C103y00	Other specified diabetes mellitus with coma	0
61670	889A.00	Diab mellit insulin-glucose infus acute myocardial infarct	0
63337	43WQ.00	Insulin IgE antibody level	0
63364	U602312	[X] Adverse reaction to insulins	0
63371	C10y100	Diabetes mellitus, adult, + other specified manifestation	0
64283	C10zy00	Other specified diabetes mellitus with unspecified comps	0
64446	C108G00	Insulin dependent diab mell with peripheral angiopathy	0
64449	C108z00	Unspecified diabetes mellitus with multiple complications	0
65025	C107z00	Diabetes mellitus NOS with peripheral circulatory disorder	0
65062	C103z00	Diabetes mellitus NOS with ketoacidotic coma	0
65616	C108H00	Insulin dependent diabetes mellitus with arthropathy	0
65684	U602311	[X] Adverse reaction to insulins and antidiabetic agents	0
67853	C106000	Diabetes mellitus, juvenile, + neurological manifestation	0
68546	ZRB4.00	Diabetes clinic satisfaction questionnaire	0
68714	SL23.00	Insulins and antidiabetic poisoning	0
68792	C10z000	Diabetes mellitus, juvenile type, + unspecified complication	0
68843	C103100	Diabetes mellitus, adult onset, with ketoacidotic coma	0
69124	C107300	IDDM with peripheral circulatory disorder	0
69748	C105000	Diabetes mellitus, juvenile type, + ophthalmic manifestation	0
70448	C107000	Diabetes mellitus, juvenile +peripheral circulatory disorder	0
70821	C10yz00	Diabetes mellitus NOS with other specified manifestation	0
72333	8HME.00	Listed for Diabetology admissn	0
72345	C102z00	Diabetes mellitus NOS with hyperosmolar coma	0
72702	C10E812	Insulin dependent diabetes mellitus - poor control	0

medcode	readcode	readterm	n
91164	ZRB4.11	CSQ - Diabetes clinic satisfaction questionnaire	0
93380	C10N100	Cystic fibrosis related diabetes mellitus	0
93875	C10E712	Insulin dependent diabetes mellitus with retinopathy	0
93922	C104000	Diabetes mellitus, juvenile type, with renal manifestation	0
94098	43WR.00	Insulin IgG antibody level	0
94383	C10N000	Secondary diabetes mellitus without complication	0
95539	C10FS00	Maternally inherited diabetes mellitus	0
95636	C10ER00	Latent autoimmune diabetes mellitus in adult	0
96506	C10G000	Secondary pancreatic diabetes mellitus without complication	0
97809	8I82.00	Did not complete DAFNE diabetes structured education program	0
97824	ZRB6.11	DWBQ - Diabetes wellbeing questionnaire	0
98071	C10E112	Insulin-dependent diabetes mellitus with ophthalmic comps	0
98704	C10E512	Insulin dependent diabetes mellitus with ulcer	0
98954	3883	Diabetes treatment satisfaction questionnaire	0
99277	9NiC.00	Did not attend DAFNE diabetes structured education programme	0
99628	Kyu0300	[X]Glomerular disorders in diabetes mellitus	0
99716	C10EE12	Insulin dependent diabetes mellitus with hypoglycaemic coma	0
99719	C10EA12	Insulin-dependent diabetes without complication	0
100033	U60231E	[X] Adverse reaction to insulins and antidiabetic agents NOS	0
100292	Cyu2300	[X]Unspecified diabetes mellitus with renal complications	0
100770	C10EF12	Insulin dependent diabetes mellitus with diabetic cataract	0
101311	C10EC12	Insulin dependent diabetes mellitus with polyneuropathy	0
101456	8IAs.00	Diabetic dietary review declined	0
101735	C10E212	Insulin-dependent diabetes mellitus with neurological comps	0
101834	9h43.00	Excepted from diabetes qual indicators: service unavailable	0
101881	2BBr.00	Impaired vision due to diabetic retinopathy	0
102163	C10ED12	Insulin dependent diabetes mellitus with nephropathy	0
102434	66Au.00	Diabetic erectile dysfunction review	0
102490	66Av.00	Diabetic assessment of erectile dysfunction	0
102767	67IJ100	Pre-conception advice for diabetes mellitus	0
102946	C10E012	Insulin-dependent diabetes mellitus with renal complications	0
103798	9b92000	Diabetic medicine	0
103935	1IA..00	No evidence of diabetic nephropathy	0
105207	8HTE100	Referral to community diabetes clinic	0
105302	K08yA00	Proteinuric diabetic nephropathy	0
105446	679c.00	Insulin administration education	0
105585	8CMW700	Diabetes clinical pathway	0
105740	2G5d.00	O/E - Left diabetic foot at increased risk	0
105741	2G5e.00	O/E - Right diabetic foot at increased risk	0
105937	8IEQ.00	Referral to community diabetes specialist nurse declined	0
106218	9m0A.00	Declined diabetic retinopathy screening	0
106269	9m0..00	Diabetic retinopathy screening administrative status	0

medcode	readcode	readterm	n
106328	9m07.00	Excluded diabetc retinop screen as under care ophthalmolgist	0
106329	9m08.00	Excluded from diabetic retinopathy screening as blind	0
106332	9m00.00	Eligible for diabetic retinopathy screening	0
106350	9m05.00	Excluded from diabetic retinopathy screening as moved away	0
106352	9m06.00	Excluded from diabetic retinopathy screening as deceased	0
106360	K27y700	Erectile dysfunction due to diabetes mellitus	0
106441	9m01.00	Ineligible for diabetic retinopathy screening	0
106445	9m0E.00	Excluded from diabetic retinopathy screen physical disorder	0
106679	8OA3.00	Provision of written information about diabetes and driving	0
106722	9Oy0300	Diabetic foot screening invitation second letter	0
106723	9Oy0200	Diabetic foot screening invitation first letter	0
106738	9Oy0000	Diabetic foot screening invitation	0
106778	9m0C.00	Excluded frm diabetic retinopathy screen as terminal illness	0
106927	PKyP.00	Diab insipidus,diab mell,optic atrophy and deafness	0
106953	8IEa.00	Referral to DAFNE diabetes structured educn prog declined	0
107361	679L200	Education about diabetes and driving	0
107414	8I94.00	Diabetes structured education programme not available	0
107423	661N400	Diabetes self-management plan review	0
107452	66o..00	Further diabetic monitoring	0
107464	66AS000	Diabetes Year of Care annual review	0
107597	9m0D.00	Excluded from diabetic retinopthy screen as learn disability	0
107739	679L211	Advice about diabetes and driving	0
107793	9Oy0400	Diabetic foot screening invitation third letter	0
107881	K08yA11	Clinical diabetic nephropathy	0
108634	9NJy.00	In-house diabetic foot screening	0
108655	8BAp.00	Insulin passport not checked	0
108993	661M400	Diabetes self-management plan agreed	0
109051	C10E612	Insulin dependent diabetes mellitus with gangrene	0

Appendix C.4 - Product Codes for Anti-Diabetic Medication

Insulin-Based Therapies

prodcode	productname	n
5015	NovoFine hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/30gauge (Novo Nordisk Ltd)	32446
5142	BD Micro-Fine + hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/31gauge (Becton, Dickinson UK Ltd)	30476
5021	NovoRapid Penfill 100units/ml solution for injection 3ml cartridges (Novo Nordisk Ltd)	25643
7228	NovoMix 30 FlexPen 100units/ml suspension for injection 3ml pre-filled pen (Novo Nordisk Ltd)	25324
2454	Mixtard 30 penfill 100 100iu/ml Penfill (Novo Nordisk Ltd)	24079
5892	NovoRapid FlexPen 100units/ml solution for injection 3ml pre-filled pen (Novo Nordisk Ltd)	18671

prodcode	productname	n
1589	U100 Insulin syringe 0.5ml	17907
5421	NovoFine hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/31gauge (Novo Nordisk Ltd)	17618
1593	Insulatard penfill 100 100iu/ml Penfill (Novo Nordisk Ltd)	13344
36853	Lantus 100units/ml solution for injection 3ml pre-filled SoloStar pen (Sanofi)	12498
6061	Novomix 30 30/70 100units/ml Injection (Novo Nordisk Ltd)	12252
5873	BD Micro-Fine + hypodermic insulin needles for pre-filled / reusable pen injectors screw on 5mm/31gauge (Becton, Dickinson UK Ltd)	12222
1592	Actrapid penfill 100 100iu/ml Penfill (Novo Nordisk Ltd)	11213
6958	Levemir FlexPen 100units/ml solution for injection 3ml pre-filled pen (Novo Nordisk Ltd)	10987
2929	Mixtard 30 100iu/ml GE injection (Novo Nordisk Ltd)	10303
322	Humalog 100units/ml solution for injection 1.5ml cartridges (Eli Lilly and Company Ltd)	10266
4760	Humulin i 100unit/ml Injection (Eli Lilly and Company Ltd)	10209
7267	NovoMix 30 Penfill 100units/ml suspension for injection 3ml cartridges (Novo Nordisk Ltd)	10049
5953	Insulin glargine 100iu/ml Injection	9989
6057	Lantus 100iu/ml Injection (Aventis Pharma)	9596
1886	Insulatard 100iu/ml GE injection (Novo Nordisk Ltd)	9094
2221	Mixtard 30 NovoLet 100units/ml suspension for injection (Novo Nordisk Ltd)	8055
4198	Humulin m3 100unit/ml M3 injection (Eli Lilly and Company Ltd)	7896
4715	Humalog mix 25 25/75 100units/ml Injection (Eli Lilly and Company Ltd)	7663
7318	Humalog 100units/ml solution for injection 3ml cartridges (Eli Lilly and Company Ltd)	7590
7231	Mixtard 30 Penfill 100units/ml suspension for injection 3ml cartridges (Novo Nordisk Ltd)	7525
5966	Mylife Clickfine hypodermic insulin needles for pre-filled / reusable pen injectors snap on 8mm/31gauge (Ypsomed Ltd)	7480
7237	Lantus 100units/ml solution for injection 3ml pre-filled OptiSet pen (Sanofi)	7306
5845	Mixtard 30 InnoLet 100units/ml suspension for injection 3ml pre-filled pen (Novo Nordisk Ltd)	7098
1591	U100 Insulin syringe 1ml	6693
6965	Levemir Penfill 100units/ml solution for injection 3ml cartridges (Novo Nordisk Ltd)	6456
1588	Actrapid 100iu/ml Injection (Novo Nordisk Ltd)	6415
7266	Lantus 100units/ml solution for injection 3ml cartridges (Sanofi)	6205
14290	Insulatard Penfill 100units/ml suspension for injection 3ml cartridges (Novo Nordisk Ltd)	5669
5850	Mylife Clickfine hypodermic insulin needles for pre-filled / reusable pen injectors snap on 6mm/31gauge (Ypsomed Ltd)	5369
7400	Insulin glargine 100units/ml solution for injection 3ml pre-filled disposable devices	4847
7393	Insulin glargine 100units/ml solution for injection 3ml cartridges	4829
10243	Humalog Mix25 100units/ml suspension for injection 3ml cartridges (Eli Lilly and Company Ltd)	4377
1587	Monotard 100units/ml suspension for injection 10ml vials (Novo Nordisk Ltd)	4073
10277	Humulin M3 100units/ml suspension for injection 3ml cartridges (Eli Lilly and Company Ltd)	3731
5121	BD Micro-Fine + hypodermic insulin needles for pre-filled / reusable pen injectors screw on 12.7mm/29gauge (Becton, Dickinson UK Ltd)	3693
5022	NovoFine hypodermic insulin needles for pre-filled / reusable pen injectors screw on 12mm/28gauge (Novo Nordisk Ltd)	3510
1840	Humulin s 100unit/ml Injection (Eli Lilly and Company Ltd)	3271
5267	BD Micro-Fine + hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 8mm needle 0.3mm/30gauge (Becton, Dickinson UK Ltd)	2635
2459	Pork Mixtard 30 100units/ml suspension for injection 10ml vials (Novo Nordisk Ltd)	2625
1844	Ultratard 100units/ml suspension for injection 10ml vials (Novo Nordisk Ltd)	2603

prodcode	productname	n
7203	Unifine Pentips hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/31gauge (Owen Mumford Ltd)	2537
1595	Insulatard NovoLet 100units/ml suspension for injection (Novo Nordisk Ltd)	2485
4093	Humulin M2 100units/ml suspension for injection 3ml cartridges (Eli Lilly and Company Ltd)	2441
1843	Pork Insulatard 100units/ml suspension for injection 10ml vials (Novo Nordisk Ltd)	2364
14357	Humulin I 100units/ml suspension for injection 3ml cartridges (Eli Lilly and Company Ltd)	2333
5059	NovoPen 3 Classic hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 2-70 units (Novo Nordisk Ltd)	2288
10225	Lantus 100units/ml solution for injection 3ml OptiClik cartridges (Sanofi)	2100
10208	Insulatard InnoLet 100units/ml suspension for injection 3ml pre-filled pen (Novo Nordisk Ltd)	1943
7300	Mixtard 30 100units/ml suspension for injection 10ml vials (Novo Nordisk Ltd)	1842
5789	Hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 8mm needle 0.3mm/30gauge	1775
1806	Penmix 30/70 100iu/ml Penfill (Novo Nordisk Ltd)	1658
9363	U100 Insulin syringe 0.5ml	1652
14330	Insulin detemir 100units/ml solution for injection 3ml pre-filled disposable devices	1618
43489	BD Micro-Fine Ultra hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/32gauge (Becton, Dickinson UK Ltd)	1609
6209	NovoRapid 100units/ml solution for injection 10ml vials (Novo Nordisk Ltd)	1508
7402	Lantus 100units/ml solution for injection 10ml vials (Sanofi)	1471
5742	Unifine Pentips hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/30gauge (Owen Mumford Ltd)	1456
7164	Unifine Pentips hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/31gauge (Owen Mumford Ltd)	1424
6991	BD Micro-Fine + hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 12.7mm needle 0.33mm/29gauge (Becton, Dickinson UK Ltd)	1414
3551	Mixtard 20 penfill 100 100iu/ml Penfill (Novo Nordisk Ltd)	1402
4790	Mixtard 50 penfill 100 100iu/ml Penfill (Novo Nordisk Ltd)	1392
10001	Humalog Mix50 Pen 100units/ml suspension for injection 3ml pre-filled pen (Eli Lilly and Company Ltd)	1349
1839	INSULIN HUMULIN I (ISOPHANE) 100 I/U INJ	1347
14270	Humalog Mix25 Pen 100units/ml suspension for injection 3ml pre-filled pen (Eli Lilly and Company Ltd)	1297
9702	BD Micro-Fine + hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12.7mm needle 0.33mm/29gauge (Becton, Dickinson UK Ltd)	1288
7793	HumaJect M3 Pen 100units/ml suspension for injection (Eli Lilly and Company Ltd)	1280
1751	Hypodermic U100 insulin syringe glass reusable 1ml	1267
14301	Insulin detemir 100units/ml solution for injection 3ml cartridges	1266
9737	Insulatard innolet 100iu/ml Injection (Novo Nordisk Ltd)	1165
10259	Insulin glargine 100units/ml solution for injection 10ml vials	1155
14928	Insulatard 100units/ml suspension for injection 10ml vials (Novo Nordisk Ltd)	1152
36920	Apidra 100units/ml solution for injection 3ml pre-filled SoloStar pen (Sanofi)	1143
10264	Humalog Pen 100units/ml solution for injection 3ml pre-filled pen (Eli Lilly and Company Ltd)	1126
1594	Actrapid NovoLet 100units/ml solution for injection (Novo Nordisk Ltd)	1121
18593	Humalog Mix50 100units/ml suspension for injection 3ml cartridges (Eli Lilly and Company Ltd)	1113
35143	NovoFine Autocover hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/30gauge (Novo Nordisk Ltd)	1057
9578	Unifine Pentips hypodermic insulin needles for pre-filled / reusable pen injectors screw on 12mm/29gauge (Owen Mumford Ltd)	1056
12297	Hypurin bovine neutral 100unit/ml Injection (C P Pharmaceuticals Ltd)	1043
14944	Humulin S 100units/ml solution for injection 3ml cartridges (Eli Lilly and Company Ltd)	1042

prodcode	productname	n
39006	Humalog Mix25 KwikPen 100units/ml suspension for injection 3ml pre-filled pen (Eli Lilly and Company Ltd)	1029
3550	Mixtard 40 penfill 100 100iu/ml Penfill (Novo Nordisk Ltd)	1007
1805	Mixtard 30/70 100unit/ml Injection (Novo Nordisk Ltd)	1004
43991	Humulin M3 KwikPen 100units/ml suspension for injection 3ml pre-filled pen (Eli Lilly and Company Ltd)	952
10184	Insulin detemir 100 iu/ml Solution for injection	916
6228	BD Micro-Fine + hypodermic U100 insulin syringe sterile single use / single patient use 0.3ml with 8mm needle 0.3mm/30gauge (Becton, Dickinson UK Ltd)	850
5214	Insulin lispro 100units/ml solution for injection 1.5ml cartridges	824
5255	Mixtard 10 penfill 100 100iu/ml Penfill (Novo Nordisk Ltd)	824
6470	Autopen 24 hypodermic insulin injection pen reusable for 3ml cartridge 2 unit dial up / range 2-42 units (Owen Mumford Ltd)	765
11080	Insulin isophane human prb 100iu/ml Injection	755
7349	Actrapid 100units/ml solution for injection 10ml vials (Novo Nordisk Ltd)	744
35260	Levemir InnoLet 100units/ml solution for injection 3ml pre-filled pen (Novo Nordisk Ltd)	736
1842	Pork velosulin 100unit/ml Injection (Novo Nordisk Ltd)	728
19513	Humulin M3 100units/ml suspension for injection 10ml vials (Eli Lilly and Company Ltd)	724
10145	Humapen luxura insulin pen 3ml/1-60 units Insulin pen 3ml/1-60 units (Eli Lilly and Company Ltd)	721
9521	Pork Actrapid 100units/ml solution for injection 10ml vials (Novo Nordisk Ltd)	718
39086	Humalog Mix50 KwikPen 100units/ml suspension for injection 3ml pre-filled pen (Eli Lilly and Company Ltd)	714
6378	Hypodermic U100 insulin syringe sterile single use / single patient use 0.3ml with 8mm needle 0.3mm/30gauge	705
11056	Insulin biphasic isophane human pyr 30:70; 100 units/ml Injection	701
38986	Humalog KwikPen 100units/ml solution for injection 3ml pre-filled pen (Eli Lilly and Company Ltd)	690
7537	Humulin Zn 100units/ml suspension for injection 10ml vials (Eli Lilly and Company Ltd)	671
13516	Hypurin bovine isophane 100unit/ml Injection (C P Pharmaceuticals Ltd)	648
14918	Humulin I 100units/ml suspension for injection 10ml vials (Eli Lilly and Company Ltd)	614
13622	Hypurin porcine neutral 100unit/ml Injection (C P Pharmaceuticals Ltd)	579
5967	Hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with not less than 12mm needle 0.33mm/29gauge	571
11086	B-d u-100 0.5ml Insulin syringe (Becton, Dickinson UK Ltd)	541
6138	HumaPen Ergo hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units Burgundy (Eli Lilly and Company Ltd)	530
38774	NovoPen 4 hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units Blue (Novo Nordisk Ltd)	529
7771	Human protaphane penfill 100 100unit/ml Penfill (Novo Nordisk Ltd)	526
5933	Mixtard 50 NovoLet 100units/ml suspension for injection (Novo Nordisk Ltd)	504
5345	Monoject hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 12mm needle 0.36mm/28gauge (Covidien (UK) Commercial Ltd)	499
7772	Human protaphane 100unit/ml Injection (Novo Nordisk Ltd)	498
1649	Human actraphane 100iu/ml Injection (Novo Nordisk Ltd)	497
6724	U100 Insulin syringe 0.3ml	487
5891	Insulatard FlexPen 100units/ml suspension for injection (Novo Nordisk Ltd)	479
4706	Velosulin 100units/ml solution for injection 10ml vials (Novo Nordisk Ltd)	471
2373	INSULIN HUMAN VELOSULIN 100 I/U INJ	456
6781	Autopen Classic hypodermic insulin injection pen reusable for 3ml cartridge 2 unit dial up / range 2-42 units (Owen Mumford Ltd)	454
2455	Mixtard 20 NovoLet 100units/ml suspension for injection (Novo Nordisk Ltd)	451

prodcode	productname	n
7127	BD Micro-Fine + hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 8mm needle 0.3mm/30gauge (Becton, Dickinson UK Ltd)	449
13819	Hypurin Porcine Isophane 100units/ml suspension for injection 1.5ml cartridges (C P Pharmaceuticals Ltd)	423
4199	Humulin m1 100unit/ml M1 injection (Eli Lilly and Company Ltd)	405
8118	Humaject i 100iu/ml Pen (Eli Lilly and Company Ltd)	397
7861	INSULIN HUMULIN S (NEUTRAL) CARTRIDGE 100 I/U	395
5250	Insulin biphasic lispro human prb 25:75; 100 units/ml Injection	383
10229	Humulin I Pen 100units/ml suspension for injection 3ml pre-filled pen (Eli Lilly and Company Ltd)	379
8841	Humulin M5 100units/ml suspension for injection 10ml vials (Eli Lilly and Company Ltd)	378
43950	Humulin I KwikPen 100units/ml suspension for injection 3ml pre-filled pen (Eli Lilly and Company Ltd)	360
6447	Insulin aspart human pyr 100 iu/ml Injection	349
18224	Humalog 100units/ml solution for injection 10ml vials (Eli Lilly and Company Ltd)	348
16160	Humulin M3 Pen 100units/ml suspension for injection 3ml pre-filled pen (Eli Lilly and Company Ltd)	344
6730	Hypodermic U100 insulin syringe sterile single use / single patient use 1ml with not less than 12mm needle 0.33mm/29gauge	340
21235	Humulin S 100units/ml solution for injection 10ml vials (Eli Lilly and Company Ltd)	337
7412	U100 Insulin syringe 0.5ml	329
38808	NovoPen 4 hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units Silver (Novo Nordisk Ltd)	325
8895	Initard 50/50 100unit/ml Injection (Novo Nordisk Ltd)	315
14313	Insulin lispro 100units/ml solution for injection 3ml cartridges	306
10242	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/31gauge	303
9341	Insulin biphasic isophane human prb 30:70; 100 units/ml Injection	301
10175	Insulin isophane human 100units/ml suspension for injection 1.5ml cartridges	296
35454	Comfort Point hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/31gauge (Exel Int. (Europe) Ltd)	293
15199	Insuman comb 25 100iu/ml Injection (Aventis Pharma)	288
4896	Monoject hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm needle 0.36mm/28gauge (Covidien (UK) Commercial Ltd)	281
3076	NovoFine hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/30gauge (Novo Nordisk Ltd)	278
9834	Mylife Clickfine hypodermic insulin needles for pre-filled / reusable pen injectors snap on 12mm/29gauge (Ypsomed Ltd)	262
2456	Mixtard 10 NovoLet 100units/ml suspension for injection (Novo Nordisk Ltd)	261
11107	Humulin m4 100unit/ml M4 injection (Eli Lilly and Company Ltd)	256
24800	Hypurin Porcine 30/70 Mix 100units/ml suspension for injection 10ml vials (Wockhardt UK Ltd)	254
9565	HumaJect S Pen 100units/ml solution for injection (Eli Lilly and Company Ltd)	252
6233	Mylife Clickfine hypodermic insulin needles for pre-filled / reusable pen injectors snap on 10mm/29gauge (Ypsomed Ltd)	249
14642	Bd m 0.3ml Insulin syringe with 29gauge needle 12mm (Becton, Dickinson UK Ltd)	247
5164	NovoPen 3 Demi hypodermic insulin injection pen reusable for 3ml cartridge 0.5 unit dial up / range 1-35 units (Novo Nordisk Ltd)	246
11337	NovoRapid Novolet 100units/ml solution for injection (Novo Nordisk Ltd)	242
11878	Hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 8mm needle 0.3mm/30gauge	235
4247	Insulin isophane porcine 100units/ml suspension for injection 1.5ml cartridges	230
6554	Autopen 24 hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-21 units (Owen Mumford Ltd)	227
19877	Insulin aspart 100units/ml solution for injection 3ml pre-filled disposable devices	218

prodcode	productname	n
9618	Hypurin Porcine 30/70 Mix 100units/ml suspension for injection 1.5ml cartridges (C P Pharmaceuticals Ltd)	217
10258	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/31gauge	215
10133	U100 Insulin syringe 1ml	213
5557	NovoPen 3 Fun hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 2-70 units Blue (Novo Nordisk Ltd)	210
25786	Comfort Point hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/31gauge (Exel Int. (Europe) Ltd)	207
6060	Unifine Pentips hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/30gauge (Owen Mumford Ltd)	204
7075	Optipen Pro 1 hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units (Sanofi)	204
10067	Insulin biphasic aspart human pyr 30:70; 100 units/ml Injection	195
11271	B-d u-100 1ml Insulin syringe (Becton, Dickinson UK Ltd)	190
14345	Apidra 100units/ml solution for injection 3ml cartridges (Sanofi)	190
20995	Hypurin Porcine 30/70 Mix 100units/ml suspension for injection 3ml cartridges (Wockhardt UK Ltd)	179
13277	Mixtard 50 Penfill 100units/ml suspension for injection 3ml cartridges (Novo Nordisk Ltd)	177
6091	NovoPen 3 Fun hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 2-70 units Red (Novo Nordisk Ltd)	172
23099	Insulin aspart biphasic 30/70 100units/ml suspension for injection 3ml pre-filled disposable devices	166
2321	Insulin 0.5ml disposable syringe	165
12638	Insulin soluble human pyr 100unit/ml Injection	165
12654	Insulin soluble human prb 100unit/ml Injection	165
16142	Insulin aspart 100units/ml solution for injection 3ml cartridges	164
8203	Penmix 50/50 100iu/ml Penfill (Novo Nordisk Ltd)	160
10484	Penmix 20/80 Penfill (Novo Nordisk Ltd)	160
11346	B-d u-100 0.3ml Insulin syringe (Becton, Dickinson UK Ltd)	156
7783	INSULIN ISOPHANE (HUMAN) 100 I/U INJ	152
25133	Insuman Comb 25 100units/ml suspension for injection 3ml pre-filled OptiSet pen (Sanofi)	152
14930	Hypurin Porcine Neutral 100units/ml solution for injection 3ml cartridges (Wockhardt UK Ltd)	150
28183	Hypurin Porcine Isophane 100units/ml suspension for injection 10ml vials (Wockhardt UK Ltd)	148
24993	Insuman Comb 25 100units/ml suspension for injection 3ml cartridges (Sanofi)	147
2812	Mixtard 40 NovoLet 100units/ml suspension for injection (Novo Nordisk Ltd)	146
14299	Insulin glulisine 100units/ml solution for injection 3ml cartridges	142
10910	Humaject m2 100iu/ml M2 pen (Eli Lilly and Company Ltd)	138
10244	Mixtard 40 Penfill 100units/ml suspension for injection 3ml cartridges (Novo Nordisk Ltd)	137
35253	Insuman Comb 50 100units/ml suspension for injection 3ml cartridges (Sanofi)	133
6831	HumaPen Ergo hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units Teal (Eli Lilly and Company Ltd)	130
5620	U100 Insulin syringe 0.3ml	122
28185	Insulin lispro biphasic 25/75 100units/ml suspension for injection 3ml cartridges	121
17712	Hypurin Bovine Lente 100units/ml suspension for injection 10ml vials (Wockhardt UK Ltd)	120
26098	Hypurin Porcine Neutral 100units/ml solution for injection 10ml vials (Wockhardt UK Ltd)	115
14505	Insulin protamine zinc bovine 100units/ml suspension for injection 10ml vials	111
14887	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 5mm/31gauge	111
15484	Insulin isophane bovine 100units/ml suspension for injection 1.5ml cartridges	111
21583	Apidra 100units/ml solution for injection 3ml pre-filled OptiSet pen (Sanofi)	104

prodcode	productname	n
9503	Hypurin Bovine Protamine Zinc 100units/ml suspension for injection 10ml vials (Wockhardt UK Ltd)	102
5501	Insuman basal 100iu/ml Injection (Aventis Pharma)	99
10546	INSULIN HUMULIN M4 100 I/U INJ	98
12299	Semitard mc 100unit/ml Injection (Novo Nordisk Ltd)	98
7319	Mixtard 20 Penfill 100units/ml suspension for injection 3ml cartridges (Novo Nordisk Ltd)	96
10572	Insulin soluble bovine 100unit/ml Injection	93
11408	Autopen Classic hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-21 units (Owen Mumford Ltd)	92
4129	Insulin soluble porcine 100units/ml solution for injection 1.5ml cartridges	90
14339	Hypurin Bovine Neutral 100units/ml solution for injection 10ml vials (Wockhardt UK Ltd)	90
12840	B-d u-100 0.5ml Insulin syringe (Becton, Dickinson UK Ltd)	89
14340	Hypurin Bovine Isophane 100units/ml suspension for injection 10ml vials (Wockhardt UK Ltd)	86
35057	OptiClik hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-80 units Blue (Sanofi)	86
10545	INSULIN HUMULIN M4 CARTRIDGE 100 I/U	80
14644	Insulin biphasic isophane human prb 20:80; 100 units/ml Injection	77
27614	Penmix 30/70 100iu/ml Injection (Novo Nordisk Ltd)	76
30819	Insuman Comb 15 100units/ml suspension for injection 3ml pre-filled OptiSet pen (Sanofi)	76
45045	NovoTwist hypodermic insulin needles for pre-filled / reusable pen injectors screw on 5mm/32gauge (Novo Nordisk Ltd)	76
43833	ClikSTAR hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-80 units Blue (Sanofi)	75
11245	B-d u-100 0.3ml Insulin syringe (Becton, Dickinson UK Ltd)	72
15294	Innovo hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-70 units Green (Novo Nordisk Ltd)	72
21554	Insuman comb 50 100iu/ml Injection (Aventis Pharma)	72
10245	Mixtard 10 Penfill 100units/ml suspension for injection 3ml cartridges (Novo Nordisk Ltd)	71
14646	Hypodermic U100 insulin syringe sterile single use / single patient use 0.3ml with 12mm needle 0.33mm/29gauge	69
21395	Insulin biphasic isophane human pyr 40:60; 100 units/ml Injection	68
11345	Bd Ultra Pen 3ml Insulin pen (Becton, Dickinson UK Ltd)	66
35078	Optipen pro 1 blue insulin pen 3ml/1-60 units 3ml/1-60 units Insulin pen (Aventis Pharma)	66
8376	INSULIN ISOPHANE 100 I/U	65
44480	Insuman Comb 25 100units/ml suspension for injection 3ml pre-filled SoloStar pen (Sanofi)	64
8839	INSULIN SEMITARD 100 I/U INJ	63
26060	Insulin lispro 100units/ml solution for injection 10ml vials	63
14933	Hypurin Porcine Isophane 100units/ml suspension for injection 3ml cartridges (Wockhardt UK Ltd)	61
19271	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/30gauge	55
321	INSULIN HUMAN ACTRAPID (NEUTRAL) 40 I/U INJ	54
13009	Hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 12mm needle 0.36mm/28gauge	52
16152	Insulin isophane biphasic human 30/70 100units/ml suspension for injection 3ml cartridges	52
14619	Insulin isophane biphasic porcine 30/70 100units/ml suspension for injection 1.5ml cartridges	51
13837	Insulin biphasic isophane human prb 10:90; 100 units/ml Injection	50
14504	INSULIN HYPURIN PROTAMINE ZINC 100 I/U INJ	50
23636	Insulin 1ml disposable syringe	49
11521	Hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm needle 0.4mm/27gauge	48

prodcode	productname	n
21347	Penmix 40/60 100iu/ml Injection (Novo Nordisk Ltd)	48
16129	Insulin soluble human 100units/ml solution for injection 3ml cartridges	46
30236	Isophane insulin 100iu/ml Injection	46
10547	Humulin Lente 100units/ml suspension for injection 10ml vials (Eli Lilly and Company Ltd)	44
23992	Insuman Basal 100units/ml suspension for injection 3ml pre-filled OptiSet pen (Sanofi)	44
23993	Insuman Rapid 100units/ml solution for injection 3ml pre-filled OptiSet pen (Sanofi)	44
25812	Insulin isophane human 100units/ml suspension for injection 3ml pre-filled disposable devices	44
9079	INSULIN SOLUBLE 100 I/U INJ	43
18149	Monoject u100 insulin syringe 12mm(29g)0.5ml 12mm29G 0.5ml Insulin syringe (Covidien (UK) Commercial Ltd)	43
21374	Insulin biphasic isophane human prb 40:60; 100 units/ml Injection	43
10207	Insulin isophane human 100units/ml suspension for injection 3ml cartridges	42
21232	Insulin isophane biphasic human 30/70 100units/ml suspension for injection 10ml vials	40
13096	Autopen hypodermic insulin injection pen reusable for 1.5ml cartridge 2 unit dial up / range 2-32 units (Owen Mumford Ltd)	39
31258	Insulin lispro biphasic 25/75 100units/ml suspension for injection 3ml pre-filled disposable devices	38
17643	Autopen Special Edition hypodermic insulin injection pen reusable for 3ml cartridge 2 unit dial up / range 2-42 units (Owen Mumford Ltd)	37
46001	Insuman Basal 100units/ml suspension for injection 3ml pre-filled SoloStar pen (Sanofi)	36
27461	Insuman Basal 100units/ml suspension for injection 3ml cartridges (Sanofi)	35
22974	Hypodermic insulin needles for pre-filled / reusable pen injectors snap on 8mm/31gauge	34
2808	INSULIN LENTARD INJ	33
12818	Human Mixtard 50 100units/ml suspension for injection 10ml vials (Novo Nordisk Ltd)	32
37427	HumaPen Luxura HD hypodermic insulin injection pen reusable for 3ml cartridge 0.5 unit dial up / range 1-30 units (Eli Lilly and Company Ltd)	30
23231	Hypurin Bovine Neutral 100units/ml solution for injection 3ml cartridges (Wockhardt UK Ltd)	29
27402	Insulin soluble human 100units/ml solution for injection 10ml vials	29
39150	Hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm safety needle 0.33mm/29gauge	28
5769	Hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 12mm needle 0.4mm/27gauge	27
43568	ClikSTAR hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-80 units Silver (Sanofi)	27
18195	Hypodermic insulin needles for pre-filled / reusable pen injectors snap on 6mm/31gauge	26
8838	INSULIN SEMITARD 40 I/U INJ	25
24795	Insulin aspart biphasic 30/70 100units/ml suspension for injection 3ml cartridges	25
29567	Insulin aspart 100units/ml solution for injection 10ml vials	24
45158	Insuman Comb 15 100units/ml suspension for injection 3ml cartridges (Sanofi)	24
5649	Autopen hypodermic insulin injection pen reusable for 1.5ml cartridge 1 unit dial up / range 1-16 units (Owen Mumford Ltd)	23
9619	Bd Ultra Pen 1.5ml Insulin pen (Becton, Dickinson UK Ltd)	23
15710	Insulin soluble human emp 100unit/ml Injection	23
43670	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/32gauge	23
29953	Apidra 100units/ml solution for injection 3ml OptiClik cartridges (Sanofi)	22
40555	HumaPen Memoir hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units (Eli Lilly and Company Ltd)	22
36959	Comfort Point hypodermic insulin needles for pre-filled / reusable pen injectors screw on 12mm/29gauge (Exel Int. (Europe) Ltd)	21
13416	Insulin biphasic 100 units/ml Injection	20
20422	Insuman comb 15 100iu/ml Injection (Aventis Pharma)	20

prodcode	productname	n
22697	Insulin isophane biphasic human 50/50 100units/ml suspension for injection 1.5ml cartridges	19
15895	Innovo hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-70 units Orange (Novo Nordisk Ltd)	18
18590	Insulin isophane bovine 100units/ml suspension for injection 10ml vials	17
28588	Hypurin Bovine Isophane 100units/ml suspension for injection 3ml cartridges (Wockhardt UK Ltd)	17
8483	MONOJECT INSULIN NEEDLES	16
9376	Insulin zinc suspension crystalline human pyr 100unit/ml long acting Injection	16
21110	Insulin biphasic isophane human prb 50:50; 100 units/ml Injection	16
25422	Hypoguard u100 click/count 1ml Insulin syringe (Hypoguard Ltd)	16
10915	Humaject m1 100iu/ml M1 pen (Eli Lilly and Company Ltd)	15
18645	INSULIN NEUTRAL (PURIFIED) 100 I/U INJ	15
4248	INSULIN NOVO ULTRATARD MC 100 I/U INJ	14
35218	Optipen pro 1 yellow insulin pen 3ml/1-60 units 3ml/1-60 units Insulin pen (Aventis Pharma)	14
35701	Insulin lispro biphasic 50/50 100units/ml suspension for injection 3ml pre-filled disposable devices	14
3396	Penmix 10/90 Penfill (Novo Nordisk Ltd)	13
4784	Lentard mc 100unit/ml Injection (Novo Nordisk Ltd)	13
5634	Hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm needle 0.36mm/28gauge	13
14649	Insulin biphasic isophane human pyr 10:90; 100 units/ml Injection	13
17809	Humaject m4 100iu/ml M4 pen (Eli Lilly and Company Ltd)	13
4163	Rapitard MC 100unit/ml Injection (Novo Nordisk Ltd)	12
8322	Insulin zinc suspension mixed human pyr 100unit/ml Injection	12
14925	Insulin isophane human vial 100unit/ml Sterile suspension injection	12
22060	Insulin 0.5 0.5ml Syringe	12
34713	Insulin 1 ml syringe	12
6753	Novopen Junior hypodermic insulin injection pen reusable for 3ml cartridge 0.5 unit dial up / range 1-35 units Green (Novo Nordisk Ltd)	11
19878	Insulin isophane biphasic human 30/70 100units/ml suspension for injection 3ml pre-filled disposable devices	11
22945	Insuman rapid 100iu/ml Injection (Aventis Pharma)	11
27177	Insulin biphasic lispro human prb 50:50; 100 units/ml Injection	11
36513	Velosulin cartridge 100unit/ml Injection (Novo Nordisk Ltd)	11
11055	Insulin biphasic isophane human pyr 20:80; 100 units/ml Injection	10
13274	Clinipak u100 single use ins syr+12mm need28g 1ml [rand] 1ml Insulin syringe with 28gauge needle 12mm (Rand Rocket Ltd)	10
35081	OptiClik hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-80 units Grey (Sanofi)	10
7765	INSULIN NEUTRAL (HUMAN) 100 I/U INJ	9
12892	Clinipak u100 single use ins syr+12mm need28g 0.5ml [rand] 0.5ml Insulin syringe with 28gauge needle 12mm (Rand Rocket Ltd)	9
13729	Insulin isophane human emp 100unit/ml Injection	9
14362	Insulin lispro 100units/ml solution for injection 3ml pre-filled disposable devices	9
22983	Insuman Rapid 100units/ml solution for injection 3ml cartridges (Sanofi)	9
23437	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 12mm/28gauge	9
28101	Insulin glulisine 100units/ml solution for injection 10ml vials	9
33966	Insulatard 100unit/ml Injection (Novo Nordisk Ltd)	9
2220	Penmix 20/80 Pen (Novo Nordisk Ltd)	8

prodcode	productname	n
17731	Penmix 50/50 100iu/ml Injection (Novo Nordisk Ltd)	8
35017	Optipen pro 1 white insulin pen 3ml/1-60 units 3ml/1-60 units Insulin pen (Aventis Pharma)	8
27280	Insulin isophane biphasic porcine 30/70 100units/ml suspension for injection 10ml vials	7
13108	Autopen Special Edition hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-21 units (Owen Mumford Ltd)	6
13969	U100 Insulin syringe 0.3ml	6
21422	Insulin isophane biphasic human 40/60 100units/ml suspension for injection 3ml cartridges	6
6981	Novopen Junior hypodermic insulin injection pen reusable for 3ml cartridge 0.5 unit dial up / range 1-35 units Yellow (Novo Nordisk Ltd)	5
13474	U100 Insulin syringe Sp.36 [2A] 1ml	5
16866	Unifine hypodermic U100 insulin syringe sterile single use / single patient use 0.3ml with 12mm needle 0.33mm/29gauge (Owen Mumford Ltd)	5
10566	INSULIN HUMULIN M CARTRIDGE 100 I/U	4
15951	Autopen Junior hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-21 units (Owen Mumford Ltd)	4
19491	Apidra 100units/ml solution for injection 10ml vials (Sanofi)	4
21590	Insulin glulisine 100units/ml solution for injection 3ml pre-filled disposable devices	4
42395	Humalog Mix25 100units/ml suspension for injection 10ml vials (Eli Lilly and Company Ltd)	4
6009	Myjector hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 12mm needle 0.4mm/27gauge (Terumo UK Ltd)	3
8646	INSULIN ZINC CRYSTALLINE susp 100 I/U INJ	3
24002	Insuman Comb 25 100units/ml suspension for injection 5ml vials (Sanofi)	3
28442	Insulin glulisine 100unit/ml Solution for injection	3
12035	Insulin zinc mixed bovine 100units/ml suspension for injection 10ml vials	2
13036	Unifine hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 12mm needle 0.33mm/29gauge (Owen Mumford Ltd)	2
13550	INSULIN BP 100 I/U	2
15961	Insulin isophane human crb 100iu/ml Injection	2
18446	Unifine hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm needle 0.33mm/29gauge (Owen Mumford Ltd)	2
20195	INSULIN BOVINE PROTAMINE ZINC 40 I/U INJ	2
20672	INSULIN HUM/ACTRAPID	2
26338	Injex starter set 60500 (Ocon Chemicals Ltd)	2
33101	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 12.7mm/29gauge	2
14506	INSULIN BOVINE PROTAMINE ZINC 100 I/U INJ	1
16389	Myjector hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm needle 0.33mm/29gauge (Terumo UK Ltd)	1
16959	Hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm needle 0.3mm/30gauge	1
18931	Insulin zinc crystalline human 100units/ml suspension for injection 10ml vials	1
21223	U100 Insulin syringe 0.3ml	1
21945	INSULIN PORK INSULATARD	1
22328	Unifine single use Insulin syringe with 30gauge needle 8mm 0.5ml (Owen Mumford Ltd)	1
22823	INSULIN ISOPHANE (PURIFIED) 100 I/U INJ	1
26621	Insulin soluble human crb 100iu/ml Injection	1
30686	Insulin isophane porcine 100units/ml suspension for injection 3ml cartridges	1
47751	TS-A Safety hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm safety needle 0.33mm/29gauge (Minsmed)	1
1643	INSULIN NOVO MONOTARD MC 100 I/U INJ	0
1645	INSULIN NOVO ACTRAPID MC 100 I/U INJ	0

prodcode	productname	n
3439	Penmix 10/90 Pen (Novo Nordisk Ltd)	0
7062	Myjector hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm needle 0.4mm/27gauge (Terumo UK Ltd)	0
7350	Insulin isophane porcine 100units/ml suspension for injection 10ml vials	0
7757	INSULIN NEULENTE (ZINC SUSP)(PURIFIED) 100 I/U INJ	0
7763	INSULIN NEUPHANE (ISOPHANE)(PURIFIED) 100 I/U INJ	0
7764	INSULIN NEUSULIN (NEUTRAL)(PURIFIED) 100 I/U INJ	0
7959	INSULIN MIXTARD 30/70 40 I/U INJ	0
8354	INSULIN ISOPHANE 70%/NEUTRAL 30% 100 I/U INJ	0
10691	INSULIN ISOPHANE (NPH) 100 I/U INJ	0
10887	Penmix 40/60 100iu/ml Penfill (Novo Nordisk Ltd)	0
12060	INSULIN QUICKSOL (SOLUBLE NEUTRAL) 100 I/U INJ	0
12244	INSULIN ZINC BOVINE susp 100 I/U INJ	0
12300	SYRINGE INSULIN (BS1619/1) 2ML	0
14191	Autopen Junior hypodermic insulin injection pen reusable for 3ml cartridge 2 unit dial up / range 2-42 units (Owen Mumford Ltd)	0
14938	Insulin soluble bovine cartridge 100unit/ml Solution for injection	0
15040	INSULIN MONOPHANE (ISOPHANE) 100 I/U INJ	0
15624	INSULIN ISOPHANE (HIGHLY PURIFIED) 100 I/U INJ	0
16209	INSULIN HYPURIN SOLUBLE 100 I/U INJ	0
16682	Tempulin 100unit/ml Injection (Knoll Ltd)	0
16700	Insulin zinc mixed bovine vial 100unit/ml Sterile suspension injection	0
17076	Omnican 100 hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 8mm needle 0.3mm/30gauge (B.Braun Medical Ltd)	0
17336	Novopen 100unit/ml Injection device (Novo Nordisk Ltd)	0
18208	Hypodermic insulin needles for pre-filled / reusable pen injectors snap on 10mm/29gauge	0
18301	INSULIN SOLUBLE INJ I/U^2	0
18461	Insulin zinc mixed human 100units/ml suspension for injection 10ml vials	0
18592	Insulin soluble bovine 100units/ml solution for injection 10ml vials	0
19029	SYRINGE PRE-SET INSULIN FOR BLIND 2ML	0
19707	INSULIN HUMULIN S (NEUTRAL SOLUBLE)	0
19829	INSULIN NOVO MONOTARD MC	0
19977	Omnican 50 hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 12mm needle 0.3mm/30gauge (B.Braun Medical Ltd)	0
20196	INSULIN SOLUBLE 40 I/U INJ	0
20671	INSULIN HUM/ACTRAPHANE	0
21459	SYRINGE INSULIN U100 S/U+8MM NEEDLE	0
22058	Pur-in mix 15/85 Injection (C P Pharmaceuticals Ltd)	0
22094	INSULIN HUMULIN M2 VIAL	0
22155	Humaject m5 100iu/ml M5 pen (Eli Lilly and Company Ltd)	0
22161	INSULIN HUMULIN M1 VIAL	0
22496	INSULIN ZINC LENTE PURIFIED SUSPENSION	0
22806	INSULIN PORK ACTRAPID	0
22946	Omnican 100 hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm needle 0.3mm/30gauge (B.Braun Medical Ltd)	0
22987	Hypodermic insulin needles for pre-filled / reusable pen injectors snap on 12mm/29gauge	0

prodcode	productname	n
23003	INSULIN ISOPHANE (NPH) 40 I/U	0
24485	INSULIN ZINC ANIMAL SUSPENSION	0
24554	Unifine single use Insulin syringe with 30gauge needle 8mm 0.3ml (Owen Mumford Ltd)	0
24593	Neutral insulin bovine 100unit/ml Injection	0
24722	INSULIN ISOPHANE 50%/NEUTRAL 50% 100 I/U INJ	0
24845	INSULIN PUR-IN ISOPHANE 100 I/U INJ	0
24846	Pur-in neutral 100unit/ml Injection (C P Pharmaceuticals Ltd)	0
24866	INSULIN INSULATARD (LEO RETARD) 40 I/U INJ	0
25006	INSULIN HUMAN ACTRAPID (NEUTRAL)	0
25479	Insulin soluble porcine 100units/ml solution for injection 3ml cartridges	0
25735	Insulin isophane biphasic human 20/80 100units/ml suspension for injection 3ml cartridges	0
25736	Insulin isophane biphasic human 10/90 100units/ml suspension for injection 3ml cartridges	0
26403	Pur-in mix 25/75 Injection (C P Pharmaceuticals Ltd)	0
26498	Insulin zinc suspension mixed bovine and porcine 100unit/ml Injection	0
26784	INSULIN ZINC SEMILENTE SUSP BP 100 I/U INJ	0
26795	SYRINGE INSULIN DISPOSABLE	0
27151	SYRINGE INSULIN U100 S/U+8MM NEEDLE	0
27396	Insulin soluble porcine 100units/ml solution for injection 10ml vials	0
27911	INSULIN HUMAN ACTRAPID PENFILL	0
28096	Insulin isophane biphasic human 50/50 100units/ml suspension for injection 3ml cartridges	0
28666	Omnican 50 hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 8mm needle 0.3mm/30gauge (B.Braun Medical Ltd)	0
28723	INSULIN ZINC BOVINE SUSPENSION	0
28978	INSULIN PUR-IN MIX 15/85 100 I/U INJ	0
29837	Insulin biphasic isophane human prb 25:75; 100 units/ml Injection	0
30209	Actrapid mc 100unit/ml Injection (Arun Products Ltd)	0
30305	Hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 12mm needle 0.3mm/30gauge	0
30861	INSULIN ZINC HUMAN SUSPENSION	0
30918	ABCare hypodermic U100 insulin syringe glass reusable 0.5ml (Rand Rocket Ltd)	0
31205	Insuman Comb 50 100units/ml suspension for injection 3ml pre-filled OptiSet pen (Sanofi)	0
31267	INSULIN PUR-IN MIX 50/50 100 I/U INJ	0
31465	Exubera 1mg inhalation powder blisters (Pfizer Ltd)	0
31466	Exubera insulin release unit (Pfizer Ltd)	0
31467	Exubera 3mg inhalation powder blisters (Pfizer Ltd)	0
31699	ABCare hypodermic U100 insulin syringe glass reusable 1ml (Rand Rocket Ltd)	0
32053	INSULIN HUMALOG MIX 25	0
33167	Insulin biphasic isophane human crb 25:75; 100 units/ml Injection	0
33232	Insulin isophane biphasic human 50/50 100units/ml suspension for injection 5ml vials	0
33356	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 12mm/29gauge	0
34031	Monotard mc 100unit/ml Injection (Novo Nordisk Ltd)	0
34097	Human initard 50/50 100unit/ml Injection (Novo Nordisk Ltd)	0
35468	Insuman Basal 100units/ml suspension for injection 5ml vials (Sanofi)	0
36031	Insulin isophane biphasic porcine 30/70 100units/ml suspension for injection 3ml cartridges	0

prodcode	productname	n
36066	Insulin isophane bovine 100units/ml suspension for injection 3ml cartridges	0
36146	Insulin lispro biphasic 50/50 100units/ml suspension for injection 3ml cartridges	0
36194	Insulin isophane biphasic human 25/75 100units/ml suspension for injection 3ml cartridges	0
36355	Insulin human 1mg inhalation powder blisters	0
36356	Insulin human 3mg inhalation powder blisters	0
36430	Insulin soluble human 100units/ml solution for injection 3ml pre-filled disposable devices	0
37055	Comfort Point hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 12mm needle 0.33mm/29gauge (Exel Int. (Europe) Ltd)	0
38093	Insulin 1ml pre-set syringe	0
38236	Comfort Point hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm needle 0.33mm/29gauge (Exel Int. (Europe) Ltd)	0
38422	Isophane 100iu/ml Injection (Celltech Pharma Europe Ltd)	0
40085	Insulin 2ml syringe	0
41120	Insulin isophane biphasic human 50/50 100units/ml suspension for injection 3ml pre-filled disposable devices	0
41834	Insulin zinc suspension lente 100iu/ml Injection (Celltech Pharma Europe Ltd)	0
41959	Penject 100unit/ml Injection device (Hypoguard Ltd)	0
42305	Injex 10ml vial adaptor pack 60502 (Ocon Chemicals Ltd)	0
42797	Injex ampoule pack 60501 (Ocon Chemicals Ltd)	0
42954	Insulin isophane biphasic human 25/75 100units/ml suspension for injection 5ml vials	0
43953	Insulin lispro biphasic 25/75 100units/ml suspension for injection 10ml vials	0
44251	Insulin zinc suspension mixed porcine 100unit/ml Injection	0
44378	Insulin isophane biphasic human 25/75 100units/ml suspension for injection 3ml pre-filled disposable devices	0
44601	Injex 4 monthly refill pack 60504 (Ocon Chemicals Ltd)	0
44810	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 5mm/32gauge	0
45639	IME-FINE hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/31gauge (Arctic Medical Ltd)	0
46666	NovoRapid FlexTouch 100units/ml solution for injection 3ml pre-filled pen (Novo Nordisk Ltd)	0
47360	Neutral insulin 100unit/ml Injection (Celltech Pharma Europe Ltd)	0
47588	Insulin 1ml click count	0
47856	Neuphane 100unit/ml Injection (Wellcome Medical Division)	0
48342	HumaPen Luxura hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units Burgundy (Eli Lilly and Company Ltd)	0
48435	BD AutoShield hypodermic insulin needles for pre-filled / reusable pen injectors screw on 5mm/29gauge (Becton, Dickinson UK Ltd)	0
48501	HumaPen Luxura hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units Champagne (Eli Lilly and Company Ltd)	0
48576	BD AutoShield Duo hypodermic insulin needles for pre-filled / reusable pen injectors screw on 5mm/30gauge (Becton, Dickinson UK Ltd)	0
48633	Comfort Point hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/31gauge (Exel Int. (Europe) Ltd)	0
48765	Unifine Pentips hypodermic insulin needles for pre-filled / reusable pen injectors screw on 5mm/31gauge (Owen Mumford Ltd)	0
48771	Unifine Pentips Plus hypodermic insulin needles for pre-filled / reusable pen injectors screw on 5mm/31gauge (Owen Mumford Ltd)	0
48811	Unifine Pentips Plus hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/31gauge (Owen Mumford Ltd)	0
48829	NovoPen Echo hypodermic insulin injection pen reusable for 3ml cartridge 0.5 unit dial up / range 0.5-30 units Blue (Novo Nordisk Ltd)	0
49052	Unifine Pentips Plus hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/31gauge (Owen Mumford Ltd)	0
49108	NovoRapid Penfill 100units/ml solution for injection 3ml cartridges (Necessity Supplies Ltd)	0

prodcode	productname	n
49172	NovoPen Echo hypodermic insulin injection pen reusable for 3ml cartridge 0.5 unit dial up / range 0.5-30 units Red (Novo Nordisk Ltd)	0
49307	Mylife Clickfine hypodermic insulin needles for pre-filled / reusable pen injectors snap on 4.5mm/31gauge (Ypsomed Ltd)	0
49451	Mylife Clickfine AutoProtect hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/29gauge (Ypsomed Ltd)	0
49479	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/31gauge	0
49506	Insujet needle free starter kit Starter kit (European Pharma Group BV)	0
49507	InsuJet nozzle pack 012070GB15 (Spirit Healthcare Ltd)	0
49508	InsuJet 10ml vial adaptor pack 01207GB1510 (Spirit Healthcare Ltd)	0
49509	InsuJet starter set 012005GBSP Blue (Spirit Healthcare Ltd)	0
49831	Lantus 100units/ml solution for injection 3ml pre-filled SoloStar pen (Necessity Supplies Ltd)	0
50633	Lantus 100units/ml solution for injection 3ml cartridges (Necessity Supplies Ltd)	0
50691	Human Mixtard 20 Penfill 100units/ml suspension for injection 1.5ml cartridges (Novo Nordisk Ltd)	0
50798	Hypodermic insulin injection pen reusable for 3ml cartridge 0.5 unit dial up / range 1-35 units	0
51107	Hypodermic insulin injection pen reusable for 3ml cartridge 2 unit dial up / range 2-42 units	0
51182	InsuJet starter set 012003GBSP Lime (Spirit Healthcare Ltd)	0
51612	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 5mm/29gauge	0
51650	Omnican Fine hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/31gauge (B.Braun Medical Ltd)	0
51743	NovoRapid Penfill 100units/ml solution for injection 3ml cartridges (Sigma Pharmaceuticals Plc)	0
51836	InsuJet 3ml cartridge adaptor pack 012071GB1503 (Spirit Healthcare Ltd)	0
51881	Omnican Fine hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/31gauge (B.Braun Medical Ltd)	0
52232	Hypodermic insulin injection pen reusable for 3ml cartridge 0.5 unit dial up / range 1-30 units	0
52319	Hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-80 units	0
52522	Humalog Mix50 KwikPen 100units/ml suspension for injection 3ml pre-filled pen (Doncaster Pharmaceuticals Ltd)	0
52722	Human Mixtard 30 Penfill 100units/ml suspension for injection 1.5ml cartridges (Novo Nordisk Ltd)	0
52748	Insulatard Penfill 100units/ml suspension for injection 3ml cartridges (Waymade Healthcare Plc)	0
53118	NovoRapid FlexPen 100units/ml solution for injection 3ml pre-filled pen (Mawdsley-Brooks & Company Ltd)	0
53148	Hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units	0
53251	NovoRapid Penfill 100units/ml solution for injection 3ml cartridges (Doncaster Pharmaceuticals Ltd)	0
53437	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 5mm/30gauge	0
53710	Insulin human 500units/ml solution for injection 20ml vials	0
54027	Hypodermic insulin needles for pre-filled / reusable pen injectors snap on 4.5mm/31gauge	0
54028	Unifine Pentips Plus hypodermic insulin needles for pre-filled / reusable pen injectors screw on 12mm/29gauge (Owen Mumford Ltd)	0
54462	Insulin biphasic isophane human emp 25:75; 100 units/ml Injection	0
54573	Hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-21 units	0
54629	Insupen hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/31gauge (Spirit Healthcare Ltd)	0
54885	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/32gauge	0
54886	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/32gauge	0
55234	Tresiba FlexTouch 200units/ml solution for injection 3ml pre-filled pen (Novo Nordisk Ltd)	0
55462	Tresiba FlexTouch 100units/ml solution for injection 3ml pre-filled pen (Novo Nordisk Ltd)	0
55517	Insulin isophane human 100units/ml suspension for injection 10ml vials	0

prodcode	productname	n
55521	Insupen hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/32gauge (Spirit Healthcare Ltd)	0
55603	Humalog KwikPen 100units/ml solution for injection 3ml pre-filled pen (Doncaster Pharmaceuticals Ltd)	0
55618	Levemir FlexPen 100units/ml solution for injection 3ml pre-filled pen (Waymade Healthcare Plc)	0
55627	Insupen hypodermic insulin needles for pre-filled / reusable pen injectors screw on 5mm/31gauge (Spirit Healthcare Ltd)	0
55687	Insulin degludec 100units/ml solution for injection 3ml pre-filled disposable devices	0
55746	Insupen hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/32gauge (Spirit Healthcare Ltd)	0
55907	Insulin degludec 100units/ml solution for injection 3ml cartridges	0
55910	Tresiba Penfill 100units/ml solution for injection 3ml cartridges (Novo Nordisk Ltd)	0
56115	Human Actrapid Penfill 100units/ml solution for injection 1.5ml cartridges (Novo Nordisk Ltd)	0
56352	HumaPen Savvio hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units Green (Eli Lilly and Company Ltd)	0
56489	NovoMix 30 Penfill 100units/ml suspension for injection 3ml cartridges (Waymade Healthcare Plc)	0
56495	Lantus 100units/ml solution for injection 3ml pre-filled OptiSet pen (Waymade Healthcare Plc)	0
56502	Actrapid Penfill 100units/ml solution for injection 3ml cartridges (Novo Nordisk Ltd)	0
56624	Kendall Magellan hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 12mm safety needle 0.3mm/30gauge (Covidien (UK) Commercial Ltd)	0
56639	HumaPen Savvio hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units Blue (Eli Lilly and Company Ltd)	0
56642	Microdot Droplet hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/31gauge (Cambridge Sensors Ltd)	0
56656	Kendall Magellan hypodermic U100 insulin syringe sterile single use / single patient use 0.3ml with 8mm safety needle 0.3mm/30gauge (Covidien (UK) Commercial Ltd)	0
56691	Insulin degludec 200units/ml solution for injection 3ml pre-filled disposable devices	0
56785	HumaPen Savvio hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units Red (Eli Lilly and Company Ltd)	0
56808	Omnican Fine hypodermic insulin needles for pre-filled / reusable pen injectors screw on 12mm/29gauge (B.Braun Medical Ltd)	0
56857	Insulin isophane biphasic human 15/85 100units/ml suspension for injection 3ml cartridges	0
56879	HumaPen Savvio hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units Graphite (Eli Lilly and Company Ltd)	0
56939	HumaPen Savvio hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units Pink (Eli Lilly and Company Ltd)	0
56983	Insupen hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/31gauge (Spirit Healthcare Ltd)	0
57153	Kendall Magellan hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 8mm safety needle 0.3mm/30gauge (Covidien (UK) Commercial Ltd)	0
57243	Omnican Fine hypodermic insulin needles for pre-filled / reusable pen injectors screw on 10mm/30gauge (B.Braun Medical Ltd)	0
57387	Kendall Magellan hypodermic U100 insulin syringe sterile single use / single patient use 0.3ml with 12mm safety needle 0.33mm/29gauge (Covidien (UK) Commercial Ltd)	0
57388	Kendall Magellan hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 8mm safety needle 0.3mm/30gauge (Covidien (UK) Commercial Ltd)	0
57451	Microdot Droplet hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/31gauge (Cambridge Sensors Ltd)	0
57493	HumaPen Savvio hypodermic insulin injection pen reusable for 3ml cartridge 1 unit dial up / range 1-60 units Silver (Eli Lilly and Company Ltd)	0
57529	Humalog 100units/ml solution for injection 10ml vials (Dowelhurst Ltd)	0
57564	Humalog KwikPen 100units/ml solution for injection 3ml pre-filled pen (Waymade Healthcare Plc)	0
57620	Humulin M3 100units/ml suspension for injection 10ml vials (Mawdsley-Brooks & Company Ltd)	0
57622	Humalog Mix50 KwikPen 100units/ml suspension for injection 3ml pre-filled pen (Waymade Healthcare Plc)	0
57744	Omnican Fine hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/31gauge (B.Braun Medical Ltd)	0
58449	Unifine Pentips hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/32gauge (Owen Mumford Ltd)	0

prodcode	productname	n
58578	Kendall Magellan hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 12mm safety needle 0.33mm/29gauge (Covidien (UK) Commercial Ltd)	0
58579	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/32.5gauge	0
58581	GlucoRx FinePoint hypodermic insulin needles for pre-filled / reusable pen injectors screw on 5mm/31gauge (GlucoRx Ltd)	0
58745	GlucoRx FinePoint hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/31gauge (GlucoRx Ltd)	0
58754	BD SafetyGlide hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm safety needle 0.33mm/29gauge (Becton, Dickinson UK Ltd)	0
58798	GlucoRx FinePoint hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/31gauge (GlucoRx Ltd)	0
58801	GlucoRx FinePoint hypodermic insulin needles for pre-filled / reusable pen injectors screw on 10mm/29gauge (GlucoRx Ltd)	0
58817	GlucoRx FinePoint hypodermic insulin needles for pre-filled / reusable pen injectors screw on 12mm/29gauge (GlucoRx Ltd)	0
58878	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 10mm/29gauge	0
58884	GlucoRx FinePoint hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/31gauge (GlucoRx Ltd)	0
58961	Mylife Penfine Classic hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/32gauge (Ypsomed Ltd)	0
58995	BD SafetyGlide hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 12mm safety needle 0.33mm/29gauge (Becton, Dickinson UK Ltd)	0
59004	Mylife Penfine Classic hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/32gauge (Ypsomed Ltd)	0
59005	Mylife Penfine Classic hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/31gauge (Ypsomed Ltd)	0
59133	BD SafetyGlide hypodermic U100 insulin syringe sterile single use / single patient use 0.5ml with 8mm safety needle 0.3mm/30gauge (Becton, Dickinson UK Ltd)	0
59243	BD SafetyGlide hypodermic U100 insulin syringe sterile single use / single patient use 0.3ml with 8mm safety needle 0.25mm/31gauge (Becton, Dickinson UK Ltd)	0
59311	Kendall Magellan hypodermic U100 insulin syringe sterile single use / single patient use 1ml with 12mm safety needle 0.33mm/29gauge (Covidien (UK) Commercial Ltd)	0
59475	Nanopass hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/32.5gauge (Terumo UK Ltd)	0
59500	Insulin isophane human 100units/ml suspension for injection 5ml vials	0
59533	NovoRapid FlexPen 100units/ml solution for injection 3ml pre-filled pen (Sigma Pharmaceuticals Plc)	0
59793	Microdot Droplet hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/32gauge (Cambridge Sensors Ltd)	0
59846	Mylife Clickfine hypodermic insulin needles for pre-filled / reusable pen injectors snap on 4mm/32gauge (Ypsomed Ltd)	0
60028	Hypodermic insulin needles for pre-filled / reusable pen injectors snap on 4mm/32gauge	0
60609	Insupen hypodermic insulin needles for pre-filled / reusable pen injectors screw on 6mm/32gauge (Spirit Healthcare Ltd)	0
60621	Hypodermic insulin injection pen reusable for 3ml cartridge 0.5 unit dial up / range 0.5-30 units	0
60626	InsuJet starter set 012007GBSP Grey (Spirit Healthcare Ltd)	0
60750	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 4mm/33gauge	0
60933	Humulin M3 100units/ml suspension for injection 10ml vials (Sigma Pharmaceuticals Plc)	0
60938	Mixtard 30 100units/ml suspension for injection 10ml vials (Waymade Healthcare Plc)	0
60951	Insulin human 100units/ml solution for injection 10ml vials	0
60967	Nanopass hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/32.5gauge (Terumo UK Ltd)	0
61562	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/29gauge	0
61780	Insupen hypodermic insulin needles for pre-filled / reusable pen injectors screw on 12mm/29gauge (Spirit Healthcare Ltd)	0
61845	NovoRapid PumpCart 100units/ml solution for injection 1.6ml cartridges (Novo Nordisk Ltd)	0
62098	Hypodermic insulin needles for pre-filled / reusable pen injectors screw on 8mm/32.5gauge	0
62180	Insulin aspart 100units/ml solution for injection 1.6ml cartridges	0
62276	Humulin R 500units/ml solution for injection 20ml vials (Imported (United States))	0

Metformin Mono-Therapies

prodcode	productname	n
23	Metformin 500mg tablets	612775
93	Metformin 850mg tablets	195384
7048	Metformin 500mg modified-release tablets	73244
39598	Metformin 1g modified-release tablets	10063
16044	Glucophage SR 500mg tablets (Merck Serono Ltd)	7162
7166	Glucophage 500mg tablets (Merck Serono Ltd)	4218
38355	Metformin 750mg modified-release tablets	4145
7610	Glucophage 850mg tablets (Merck Serono Ltd)	1734
39729	Glucophage SR 1000mg tablets (Merck Serono Ltd)	806
11990	Metformin 500mg/5ml oral solution sugar free	697
39988	Metformin 500mg oral powder sachets sugar free	445
735	Metformin 100mg/ml Oral solution	236
38400	Glucophage SR 750mg tablets (Merck Serono Ltd)	215
40233	Metformin 1g oral powder sachets sugar free	193
3252	METFORMIN HCl 500 MG TAB	124
40110	Glucophage 500mg oral powder sachets (Merck Serono Ltd)	76
33087	Metformin 500mg tablets (Actavis UK Ltd)	73
2928	METFORMIN HCl 850 MG TAB	58
39560	Bolamyn SR 500mg tablets (Teva UK Ltd)	25
34504	Metformin 500mg tablets (Wockhardt UK Ltd)	23
40007	Glucophage 1000mg oral powder sachets (Merck Serono Ltd)	16
34323	Metformin 500mg tablets (A A H Pharmaceuticals Ltd)	12
31146	Metsol 500mg/5ml oral solution (Kappin Ltd)	5
45581	Metabet SR 500mg tablets (Morningside Healthcare Ltd)	3
7815	METFORMIN 800 MG TAB	0
16213	METFORMIN 250 MG TAB	0
20810	METFORMIN	0
25678	Glucamet 500mg Tablet (Opus Pharmaceuticals Ltd)	0
26258	Glucamet 850mg Tablet (Opus Pharmaceuticals Ltd)	0
27501	Orabet 500mg Tablet (Lagap)	0
33674	Metformin 850mg tablets (A A H Pharmaceuticals Ltd)	0
34004	Metformin 500mg tablets (IVAX Pharmaceuticals UK Ltd)	0
34020	Metformin 850mg tablets (IVAX Pharmaceuticals UK Ltd)	0
34135	Metformin 500mg Tablet (M & A Pharmachem Ltd)	0
34598	Metformin 500mg tablets (Generics (UK) Ltd)	0
34697	Metformin 850mg tablets (Wockhardt UK Ltd)	0
34742	Metformin 850mg tablets (Teva UK Ltd)	0
34836	Metformin 850mg tablets (Actavis UK Ltd)	0
34917	Metformin 500mg tablets (Teva UK Ltd)	0

prodcode	productname	n
42161	Orabet 500mg Tablet (Sandoz Ltd)	0
43270	Metformin 500mg/5ml oral solution sugar free (Rosemont Pharmaceuticals Ltd)	0
44250	Metformin 500mg/5ml Oral solution (Hillcross Pharmaceuticals Ltd)	0
46989	Metabet SR 1000mg tablets (Morningside Healthcare Ltd)	0
47939	Glucient SR 500mg tablets (Consilient Health Ltd)	0
48149	Metformin 500mg tablets (Almus Pharmaceuticals Ltd)	0
49502	Glucophage SR 500mg tablets (Mawdsley-Brooks & Company Ltd)	0
49738	Metformin 1g modified-release tablets (A A H Pharmaceuticals Ltd)	0
50570	Glucophage SR 500mg tablets (Lexon (UK) Ltd)	0
50821	Metformin 850mg tablets (Pfizer Ltd)	0
50970	Metformin 500mg tablets (Bristol Laboratories Ltd)	0
51080	Metabet SR 1000mg tablets (Actavis UK Ltd)	0
51135	Metformin 500mg modified-release tablets (A A H Pharmaceuticals Ltd)	0
51527	Metformin 500mg tablets (Boston Healthcare Ltd)	0
52221	Diagemet XL 500mg tablets (Genus Pharmaceuticals Ltd)	0
52442	Metformin 500mg tablets (Pfizer Ltd)	0
52634	Glucophage SR 500mg tablets (Doncaster Pharmaceuticals Ltd)	0
53478	Metformin 500mg modified-release tablets (Kent Pharmaceuticals Ltd)	0
53774	Metabet SR 500mg tablets (Actavis UK Ltd)	0
53867	Metformin 500mg tablets (Zentiva)	0
54442	Metformin (roi) 1000mg Tablet	0
54898	Metformin 850mg tablets (Almus Pharmaceuticals Ltd)	0
55270	Duformin 500mg Tablet (Dumex Ltd)	0
55711	Metformin 500mg tablets (Alliance Healthcare (Distribution) Ltd)	0
55739	Metformin 500mg tablets (Tillomed Laboratories Ltd)	0
57147	Bolamyn SR 1000mg tablets (Teva UK Ltd)	0
57457	Metformin 500mg tablets (Aurobindo Pharma Ltd)	0
58051	Metformin 500mg/5ml oral solution	0
58607	Metformin 500mg/5ml oral solution sugar free (Zentiva)	0
59620	Glucophage SR 500mg tablets (Waymade Healthcare Plc)	0
60074	Metformin 1g modified-release tablets (Waymade Healthcare Plc)	0
60286	Metformin 500mg/5ml oral suspension	0
60968	Metformin 500mg modified-release tablets (Actavis UK Ltd)	0
61043	Sukkarto SR 1000mg tablets (Morningside Healthcare Ltd)	0
61559	Sukkarto SR 500mg tablets (Morningside Healthcare Ltd)	0
62144	Metformin 500mg modified-release tablets (DE Pharmaceuticals)	0
62265	Metformin 500mg modified-release tablets (Mawdsley-Brooks & Company Ltd)	0

Other Anti-Diabetic Therapies

prodcode	productname	n
32	Gliclazide 80mg tablets	413917

prodcode	productname	n
1254	Glibenclamide 5mg tablets	45844
5627	Gliclazide 30mg modified-release tablets	36265
469	Rosiglitazone 4mg tablets	34699
1965	Tolbutamide 500mg tablets	31084
9699	Pioglitazone 30mg tablets	28150
5636	Glipizide 5mg tablets	26002
548	Pioglitazone 15mg tablets	21971
35022	Sitagliptin 100mg tablets	20567
5227	Rosiglitazone 8mg tablets	17081
10051	Pioglitazone 45mg tablets	16039
5316	Glimepiride 4mg tablets	13262
5353	Glimepiride 2mg tablets	12672
2219	Glibenclamide 2.5mg tablets	12126
5276	Glimepiride 1mg tablets	9892
479	Acarbose 50mg tablets	9834
6337	Glimepiride 3mg tablets	9747
1964	Diamicron 80mg tablets (Servier Laboratories Ltd)	5004
547	Glipizide 2.5mg tablets	4520
18220	Pioglitazone 15mg / Metformin 850mg tablets	3811
5174	Acarbose 100mg tablets	3590
7375	Rosiglitazone 4mg / Metformin 1g tablets	3185
43065	Gliclazide 40mg tablets	3009
11717	Rosiglitazone 2mg / Metformin 1g tablets	2819
35149	Exenatide 10micrograms/0.04ml solution for injection 2.4ml pre-filled disposable devices	2652
40693	Liraglutide 6mg/ml solution for injection 3ml pre-filled disposable devices	2307
14164	Avandamet 2mg/1000mg tablets (GlaxoSmithKline UK Ltd)	2281
9748	Repaglinide 2mg tablets	2114
9707	Repaglinide 1mg tablets	1957
9865	Repaglinide 500microgram tablets	1678
11695	Diamicron 30mg MR tablets (Servier Laboratories Ltd)	1646
7325	Avandamet 4mg/1000mg tablets (GlaxoSmithKline UK Ltd)	1607
11483	Nateglinide 60mg tablets	1478
35150	Byetta 10micrograms/0.04ml solution for injection 2.4ml pre-filled disposable devices (Bristol-Myers Squibb Pharmaceuticals Ltd)	1475
37875	Vildagliptin 50mg tablets	1465
5621	Glucobay 50mg tablets (Bayer Plc)	1438
41204	Saxagliptin 5mg tablets	1417
11601	Rosiglitazone 2mg / Metformin 500mg tablets	1412
7744	Daonil 5mg tablets (Sanofi)	1282
6855	Avandamet 2mg/500mg tablets (GlaxoSmithKline UK Ltd)	1226
5678	Nateglinide 120mg tablets	1155
9662	Avandia 4mg tablets (GlaxoSmithKline UK Ltd)	1075
12513	Glibenese 5mg tablets (Pfizer Ltd)	1036

prodcode	productname	n
40642	Victoza 6mg/ml solution for injection 3ml pre-filled pen (Novo Nordisk Ltd)	920
21832	Diabetamide 5mg tablets (Ashbourne Pharmaceuticals Ltd)	876
35462	Januvia 100mg tablets (Merck Sharp & Dohme Ltd)	876
13331	Euglucon 5mg tablets (Sanofi)	831
37902	Vildagliptin 50mg / Metformin 1g tablets	666
30316	Metformin with pioglitazone 850mg + 15mg Tablet	654
11284	Amaryl 4mg tablets (Zentiva)	616
35251	Exenatide 5micrograms/0.02ml solution for injection 1.2ml pre-filled disposable devices	584
11760	Metformin with rosiglitazone 1000mg + 2mg Tablet	575
7284	Amaryl 2mg tablets (Zentiva)	563
8168	Diabinese 250mg Tablet (Pfizer Ltd)	533
7409	Amaryl 3mg tablets (Zentiva)	518
9105	Glucobay 100mg tablets (Bayer Plc)	514
1847	Chlorpropamide 250mg tablets	506
7332	Amaryl 1mg tablets (Zentiva)	501
1253	Chlorpropamide 100mg tablets	495
5989	Nateglinide 180mg tablets	490
7912	Semi-Daonil 2.5mg tablets (Sanofi)	421
11610	Metformin with rosiglitazone 500mg + 2mg Tablet	403
17698	Minodiab 5mg tablets (Pfizer Ltd)	370
45775	Saxagliptin 2.5mg tablets	367
43619	Metformin 1g / Sitagliptin 50mg tablets	332
35144	Byetta 5micrograms/0.02ml solution for injection 1.2ml pre-filled disposable devices (Bristol-Myers Squibb Pharmaceuticals Ltd)	326
4862	Diabetamide 2.5mg tablets (Ashbourne Pharmaceuticals Ltd)	265
46665	Linagliptin 5mg tablets	256
11737	Metformin with rosiglitazone 1000mg + 4mg Tablet	255
31077	Competact 15mg/850mg tablets (Takeda UK Ltd)	243
8976	Euglucon 2.5mg tablets (Aventis Pharma)	228
8390	Gliquidone 30mg tablets	216
11604	Rosiglitazone 1mg / Metformin 500mg tablets	199
21892	Diaglyk 80mg tablets (Ashbourne Pharmaceuticals Ltd)	196
15232	Avandia 8mg tablets (GlaxoSmithKline UK Ltd)	193
15374	Gliclazide 40mg/5ml oral suspension	177
37874	Vildagliptin 50mg / Metformin 850mg tablets	176
17580	Avandamet 1mg/500mg tablets (GlaxoSmithKline UK Ltd)	173
20889	Actos 30mg tablets (Takeda UK Ltd)	173
15955	Starlix 120mg tablets (Novartis Pharmaceuticals UK Ltd)	149
8034	Diabinese 100mg Tablet (Pfizer Ltd)	146
38551	Eucreas 50mg/1000mg tablets (Novartis Pharmaceuticals UK Ltd)	134
43684	Janumet 50mg/1000mg tablets (Merck Sharp & Dohme Ltd)	107
11316	NovoNorm 500microgram tablets (Novo Nordisk Ltd)	95
11321	NovoNorm 1mg tablets (Novo Nordisk Ltd)	95

prodcode	productname	n
22145	Tolanase 100mg Tablet (Pharmacia Ltd)	92
23945	Starlix 60mg tablets (Novartis Pharmaceuticals UK Ltd)	91
19472	Actos 45mg tablets (Takeda UK Ltd)	90
20287	Actos 15mg tablets (Takeda UK Ltd)	85
46469	Bydureon 2mg powder and solvent for suspension for injection vials (Bristol-Myers Squibb Pharmaceuticals Ltd)	84
17706	Minodiab 2.5mg tablets (Pfizer Ltd)	82
11366	NovoNorm 2mg tablets (Novo Nordisk Ltd)	77
12455	Rastinon 500mg Tablet (Hoechst Marion Roussel)	75
41431	Onglyza 5mg tablets (Bristol-Myers Squibb Pharmaceuticals Ltd)	72
10427	Tolazamide 250mg Tablet	71
11609	Metformin with rosiglitazone 500mg + 1mg Tablet	58
29939	Gliclazide 80mg tablets (Generics (UK) Ltd)	47
17343	Gliclazide 80mg tablets (A A H Pharmaceuticals Ltd)	34
31212	Gliclazide 80mg tablets (Actavis UK Ltd)	33
16602	Calabren 2.5mg Tablet (Berk Pharmaceuticals Ltd)	31
46458	Exenatide 2mg powder and solvent for suspension for injection vials	29
4426	CHLORPROPAMIDE 500 MG TAB	25
19336	Tolazamide 100mg Tablet	25
26218	Calabren 5mg Tablet (Berk Pharmaceuticals Ltd)	25
34802	Glipizide 5mg tablets (IVAX Pharmaceuticals UK Ltd)	25
46716	Trajenta 5mg tablets (Boehringer Ingelheim Ltd)	24
7695	Guarem Sachets (Shire Pharmaceuticals Ltd)	23
36774	Prandin 1mg tablets (Novo Nordisk Ltd)	21
12897	Guar gum 5g granules sachets sugar free	17
35561	Prandin 2mg tablets (Novo Nordisk Ltd)	17
27125	Starlix 180mg tablets (Novartis Pharmaceuticals UK Ltd)	14
36948	Prandin 0.5mg tablets (Novo Nordisk Ltd)	10
39149	Galvus 50mg tablets (Novartis Pharmaceuticals UK Ltd)	9
39203	Eucreas 50mg/850mg tablets (Novartis Pharmaceuticals UK Ltd)	8
3740	Guar gum 90% granules	7
4307	Guarina Sachets (Norgine Pharmaceuticals Ltd)	1
9108	TOLBUTAMIDE 250 MG TAB	1
16211	TOLBUTAMIDE 100 MG TAB	1
19658	Glurenorm 30mg tablets (Sanofi)	1
7745	GUAR GUM TAB	0
11946	Tolbutamide 50mg/ml Injection	0
12245	Glutril 25mg Tablet (Roche Products Ltd)	0
12259	Glibornuride 25mg Tablet	0
13628	Romozin 400mg Tablet (Glaxo Wellcome UK Ltd)	0
21424	Glibenclamide 5mg/5ml oral suspension	0
21489	Tolanase 250mg Tablet (Pharmacia Ltd)	0
21564	Gliclazide 80mg tablets (Wockhardt UK Ltd)	0

prodcode	productname	n
22636	TOLBUTAMIDE 1 GM TAB	0
22858	Acetohexamide 500mg tablets	0
24848	Glymidine sodium 500mg Tablet	0
25636	Libanil 2.5mg Tablet (Approved Prescription Services Ltd)	0
26118	Dimelor 500mg Tablet (Eli Lilly and Company Ltd)	0
27969	Glymese 250mg Tablet (DDSA Pharmaceuticals Ltd)	0
28708	Malix 2.5mg Tablet (Lagap)	0
29326	Glipizide 5mg tablets (Generics (UK) Ltd)	0
30460	Malix 5mg Tablet (Lagap)	0
31474	Libanil 5mg Tablet (Approved Prescription Services Ltd)	0
33562	Duclazide 80mg Tablet (Dumex Ltd)	0
33673	Tolbutamide 500mg tablets (Actavis UK Ltd)	0
34399	Gliclazide 80mg tablets (IVAX Pharmaceuticals UK Ltd)	0
34507	Glibenclamide 2.5mg tablets (Wockhardt UK Ltd)	0
34563	Glibenclamide 5mg tablets (Wockhardt UK Ltd)	0
34676	Glibenclamide 2.5mg tablets (A A H Pharmaceuticals Ltd)	0
34706	Glibenclamide 2.5mg tablets (IVAX Pharmaceuticals UK Ltd)	0
34932	Gliclazide 80mg tablets (Genus Pharmaceuticals Ltd)	0
34957	Tolbutamide 500mg tablets (A A H Pharmaceuticals Ltd)	0
36856	Gliclazide 80mg tablets (Sandoz Ltd)	0
37617	Rosiglitazone 2mg tablet	0
40365	Glimepiride 1mg tablets (Actavis UK Ltd)	0
40425	Nazdol MR 30mg tablets (Teva UK Ltd)	0
41558	Glibenclamide 5mg tablets (Teva UK Ltd)	0
41559	Glibenclamide 5mg tablets (A A H Pharmaceuticals Ltd)	0
41593	Glibenclamide 2.5mg tablets (Teva UK Ltd)	0
41898	GLIBENCLAMIDE	0
42790	Gliclazide 80mg Tablet (Merck Generics (UK) Ltd)	0
43465	Zicron 40mg tablets (Bristol Laboratories Ltd)	0
44304	Glyconon 500mg Tablet (DDSA Pharmaceuticals Ltd)	0
44473	Edicil MR 30mg tablets (Ratiopharm UK Ltd)	0
44738	Niddaryl 1mg tablets (Dee Pharmaceuticals Ltd)	0
45215	Gliclazide 80mg Tablet (Neo Laboratories Ltd)	0
45821	Onglyza 2.5mg tablets (Bristol-Myers Squibb Pharmaceuticals Ltd)	0
45831	Dacadis MR 30mg tablets (Generics (UK) Ltd)	0
46927	Tolbutamide 500mg tablets (Teva UK Ltd)	0
47074	Gliclazide 80mg/5ml oral suspension	0
47894	Nazdol MR 30mg tablets (Consilient Health Ltd)	0
48056	Gliclazide 80mg tablets (Sovereign Medical Ltd)	0
48120	Avandia 2mg Tablet (GlaxoSmithKline UK Ltd)	0
48139	Pioglitazone 30mg tablets (A A H Pharmaceuticals Ltd)	0
48401	Sitagliptin 50mg tablets	0

prodcode	productname	n
48533	Sitagliptin 25mg tablets	0
50087	Januvia 50mg tablets (Merck Sharp & Dohme Ltd)	0
50124	Januvia 25mg tablets (Merck Sharp & Dohme Ltd)	0
50682	Jentadueto 2.5mg/1000mg tablets (Boehringer Ingelheim Ltd)	0
51955	Gliclazide 80mg tablets (Accord Healthcare Ltd)	0
52203	Enyglid 0.5mg tablets (Consilient Health Ltd)	0
52445	Linagliptin 2.5mg / Metformin 1g tablets	0
52449	Linagliptin 2.5mg / Metformin 850mg tablets	0
53288	Gliclazide 30mg modified-release tablets (A A H Pharmaceuticals Ltd)	0
54150	Jentadueto 2.5mg/850mg tablets (Boehringer Ingelheim Ltd)	0
54182	Dapagliflozin 10mg tablets	0
54203	Forxiga 10mg tablets (Bristol-Myers Squibb Pharmaceuticals Ltd)	0
54265	Dapagliflozin 5mg tablets	0
54480	Forxiga 5mg tablets (Bristol-Myers Squibb Pharmaceuticals Ltd)	0
54764	Gliclazide 80mg tablets (Arrow Generics Ltd)	0
54891	Saxagliptin 2.5mg / Metformin 1g tablets	0
54973	Saxagliptin 2.5mg / Metformin 850mg tablets	0
55413	Lixisenatide 20micrograms/0.2ml solution for injection 3ml pre-filled disposable devices	0
55459	Lixisenatide 10micrograms/0.2ml solution for injection 3ml pre-filled disposable devices	0
55723	Lixisenatide 10micrograms/0.2ml solution for injection 3ml pre-filled disposable devices and Lixisenatide 20micrograms/0.2ml solution for injection 3ml pre-filled disposable devices	0
55728	Lyxumia 10micrograms/0.2ml solution for injection 3ml pre-filled pen (Sanofi)	0
55729	Lyxumia 20micrograms/0.2ml solution for injection 3ml pre-filled pen (Sanofi)	0
55767	Lyxumia 10micrograms/0.2ml solution for injection 3ml pre-filled pen and Lyxumia 20micrograms/0.2ml solution for injection 3ml pre-filled pen (Sanofi)	0
55862	Gliclazide Oral solution	0
56008	Gliclazide 80mg tablets (Almus Pharmaceuticals Ltd)	0
56208	Pioglitazone 15mg tablets (A A H Pharmaceuticals Ltd)	0
56376	Rosiglitazone 4mg with glimepiride 4mg tablet	0
56437	Gliclazide 60mg modified-release tablets	0
56831	Troglitazone 200mg Tablet	0
56965	Komboglyze 2.5mg/1000mg tablets (Bristol-Myers Squibb Pharmaceuticals Ltd)	0
57601	Daonil 5mg tablets (Dowelhurst Ltd)	0
57659	Pioglitazone 30mg tablets (Actavis UK Ltd)	0
57830	Gliclazide 30mg modified-release tablets (Alliance Healthcare (Distribution) Ltd)	0
58865	Komboglyze 2.5mg/850mg tablets (AstraZeneca UK Ltd)	0
58882	Gliclazide 120mg/5ml oral suspension	0
59177	Alogliptin 25mg tablets	0
59385	Vipdomet 12.5mg/1000mg tablets (Takeda UK Ltd)	0
59809	Alogliptin 6.25mg tablets	0
60012	Dapagliflozin 5mg / Metformin 1g tablets	0
60066	Invokana 100mg tablets (Janssen-Cilag Ltd)	0
60073	Canagliflozin 100mg tablets	0
60211	Canagliflozin 100mg tablets	0

prodcode	productname	n
60328	Alogliptin 12.5mg tablets	0
60379	Invokana 300mg tablets (Janssen-Cilag Ltd)	0
60386	Canagliflozin 300mg tablets	0
60430	Invokana 100mg tablets (Janssen-Cilag Ltd)	0
60495	Gliclazide 80mg tablets (Teva UK Ltd)	0
60497	Alogliptin 12.5mg / Metformin 1g tablets	0
60643	Xigduo 5mg/1000mg tablets (AstraZeneca UK Ltd)	0
60681	Vipidia 12.5mg tablets (Takeda UK Ltd)	0
60682	Vipidia 25mg tablets (Takeda UK Ltd)	0
61311	Glimepiride 4mg tablets (Sigma Pharmaceuticals Plc)	0
61756	Empagliflozin 10mg tablets	0
61925	NovoNorm 500microgram tablets (Waymade Healthcare Plc)	0
61957	Gliclazide 40mg tablets (A A H Pharmaceuticals Ltd)	0
62014	Glimepiride 2mg tablets (Accord Healthcare Ltd)	0
62034	Laaglyda MR 60mg tablets (Consilient Health Ltd)	0
62172	Empagliflozin 25mg tablets	0

Appendix C.5 - Medical Codes for Polycystic Ovary Syndrome

medcode	readcode	readterm	n
1466	C164.00	Polycystic ovaries	561
11347	C165.00	Polycystic ovarian syndrome	42
16103	C164.12	Stein - Leventhal syndrome	1

Appendix C.6 - Medical Codes for Terminology Traditionally Associated with Type of Diabetes

Insulin-Dependent Diabetes

medcode	readcode	readterm	n
1038	C100011	Insulin dependent diabetes mellitus	953
1647	C108.00	Insulin dependent diabetes mellitus	936
18505	C108.11	IDDM-Insulin dependent diabetes mellitus	217
6791	C108800	Insulin dependent diabetes mellitus - poor control	16
31310	C108900	Insulin dependent diabetes maturity onset	16
6509	C108700	Insulin dependent diabetes mellitus with retinopathy	9

medcode	readcode	readterm	n
39809	C108J00	Insulin dependent diab mell with neuropathic arthropathy	8
51261	C10E.12	Insulin dependent diabetes mellitus	4
26855	C108400	Unstable insulin dependent diabetes mellitus	2
41716	C108C00	Insulin dependent diabetes mellitus with polyneuropathy	2
60499	C108600	Insulin dependent diabetes mellitus with gangrene	2
44260	C108F00	Insulin dependent diabetes mellitus with diabetic cataract	1
44440	C108E00	Insulin dependent diabetes mellitus with hypoglycaemic coma	1
44443	C108500	Insulin dependent diabetes mellitus with ulcer	1
97849	C10E912	Insulin dependent diabetes maturity onset	1
24694	C108B00	Insulin dependent diabetes mellitus with mononeuropathy	0
45276	C10E312	Insulin dependent diabetes mellitus with multiple complicat	0
49276	C108100	Insulin-dependent diabetes mellitus with ophthalmic comps	0
52104	C108300	Insulin dependent diabetes mellitus with multiple complicatn	0
54600	C10E412	Unstable insulin dependent diabetes mellitus	0
57621	C108D00	Insulin dependent diabetes mellitus with nephropathy	0
64446	C108G00	Insulin dependent diab mell with peripheral angiopathy	0
65616	C108H00	Insulin dependent diabetes mellitus with arthropathy	0
69124	C107300	IDDM with peripheral circulatory disorder	0
72702	C10E812	Insulin dependent diabetes mellitus - poor control	0
93875	C10E712	Insulin dependent diabetes mellitus with retinopathy	0
98071	C10E112	Insulin-dependent diabetes mellitus with ophthalmic comps	0
98704	C10E512	Insulin dependent diabetes mellitus with ulcer	0
99716	C10EE12	Insulin dependent diabetes mellitus with hypoglycaemic coma	0
99719	C10EA12	Insulin-dependent diabetes without complication	0
100770	C10EF12	Insulin dependent diabetes mellitus with diabetic cataract	0
101311	C10EC12	Insulin dependent diabetes mellitus with polyneuropathy	0
101735	C10E212	Insulin-dependent diabetes mellitus with neurological comps	0
102163	C10ED12	Insulin dependent diabetes mellitus with nephropathy	0
102946	C10E012	Insulin-dependent diabetes mellitus with renal complications	0
109051	C10E612	Insulin dependent diabetes mellitus with gangrene	0

Non-Insulin-Dependent Diabetes

medcode	readcode	readterm	n
4513	C109.00	Non-insulin dependent diabetes mellitus	5612
506	C100112	Non-insulin dependent diabetes mellitus	4881
5884	C109.11	NIDDM - Non-insulin dependent diabetes mellitus	269
8403	C109700	Non-insulin dependent diabetes mellitus - poor control	63
29979	C109900	Non-insulin-dependent diabetes mellitus without complication	4
34912	C109400	Non-insulin dependent diabetes mellitus with ulcer	4
17262	C109600	Non-insulin-dependent diabetes mellitus with retinopathy	3
52303	C109000	Non-insulin-dependent diabetes mellitus with renal comps	2

medcode	readcode	readterm	n
59365	C109C00	Non-insulin dependent diabetes mellitus with nephropathy	2
40962	C109H00	Non-insulin dependent d m with neuropathic arthropathy	1
43785	C109D00	Non-insulin dependent diabetes mellitus with hypoglyca coma	1
45467	C109B00	Non-insulin dependent diabetes mellitus with polyneuropathy	1
50429	C109100	Non-insulin-dependent diabetes mellitus with ophthalm comps	1
50609	L180600	Pre-existing diabetes mellitus - non-insulin-dependent	1
54212	C109F00	Non-insulin-dependent d m with peripheral angiopath	1
55842	C109200	Non-insulin-dependent diabetes mellitus with neuro comps	1
62146	C109300	Non-insulin-dependent diabetes mellitus with multiple comps	1
24693	C109G00	Non-insulin dependent diabetes mellitus with arthropathy	0
37648	C109J11	Insulin treated non-insulin dependent diabetes mellitus	0
40401	C109500	Non-insulin dependent diabetes mellitus with gangrene	0
56803	C107400	NIDDM with peripheral circulatory disorder	0
69278	C109E00	Non-insulin depend diabetes mellitus with diabetic cataract	0
72320	C109A00	Non-insulin dependent diabetes mellitus with mononeuropathy	0

Dose-Adjustment for Normal Eating

medcode	readcode	readterm	n
93390	9OLH.00	Attended DAFNE diabetes structured education programme	5
93491	9OLJ.00	DAFNE diabetes structured education programme completed	2
93704	8Hj3.00	Referral to DAFNE diabetes structured education programme	2
97809	8I82.00	Did not complete DAFNE diabetes structured education program	0
99277	9NiC.00	Did not attend DAFNE diabetes structured education programme	0
106953	81Ea.00	Referral to DAFNE diabetes structured educn prog declined	0

Diabetes Education and Self-Management for Ongoing and Newly Diagnosed

medcode	readcode	readterm	n
93657	8Hj4.00	Referral to DESMOND diabetes structured education programme	324
95159	9NiD.00	Did not attend DESMOND diabetes structured education program	49
93529	9OLK.00	DESMOND diabetes structured education programme completed	39
95093	8I83.00	Did not complete DESMOND diabetes structured educat program	0

Diabetic Ketoacidosis

medcode	readcode	readterm	n
1682	C101.00	Diabetes mellitus with ketoacidosis	160
42505	C101z00	Diabetes mellitus NOS with ketoacidosis	5

medcode	readcode	readterm	n
15690	C103.00	Diabetes mellitus with ketoacidotic coma	2
42567	C103000	Diabetes mellitus - juvenile type with ketoacidotic coma	1
38617	C101y00	Other specified diabetes mellitus with ketoacidosis	0
53200	C101000	Diabetes mellitus - juvenile type with ketoacidosis	0
54856	C101100	Diabetes mellitus - adult onset with ketoacidosis	0
65062	C103z00	Diabetes mellitus NOS with ketoacidotic coma	0
68843	C103100	Diabetes mellitus - adult onset with ketoacidotic coma	0

Hyperosmolar Nonketotic Syndrome

medcode	readcode	readterm	n
21482	C102.00	Diabetes mellitus with hyperosmolar coma	5
40023	C102000	Diabetes mellitus - juvenile type with hyperosmolar coma	0
43139	C102100	Diabetes mellitus - adult onset with hyperosmolar coma	0
72345	C102z00	Diabetes mellitus NOS with hyperosmolar coma	0

Juvenile-Onset Diabetes

medcode	readcode	readterm	n
24490	C100000	Diabetes mellitus - juvenile type without complication	4
28856	8CP2.00	Transition of diabetes care options discussed	2
16946	13L4.11	Diabetic child	1
42567	C103000	Diabetes mellitus - juvenile type with ketoacidotic coma	1
40023	C102000	Diabetes mellitus - juvenile type with hyperosmolar coma	0
53200	C101000	Diabetes mellitus - juvenile type with ketoacidosis	0
67853	C106000	Diabetes mellitus - juvenile with neurological manifestation	0
68792	C10z000	Diabetes mellitus - juvenile type with unspecified complication	0
69748	C105000	Diabetes mellitus - juvenile type with ophthalmic manifestation	0
70448	C107000	Diabetes mellitus - juvenile onset with peripheral circulatory disorder	0
93922	C104000	Diabetes mellitus - juvenile type with renal manifestation	0

Appendix C.7 - Type-Explicit Medical Codes for Diabetes

Type 1 Diabetes

medcode	readcode	readterm	n
1549	C10E.00	Type 1 diabetes mellitus	1144
17858	C108.12	Type 1 diabetes mellitus	177
85660	66An.00	Diabetes type 1 review	52
10692	C10EM00	Type 1 diabetes mellitus with ketoacidosis	42

medcode	readcode	readterm	n
10418	C10ED00	Type 1 diabetes mellitus with nephropathy	24
30294	C10EL00	Type 1 diabetes mellitus with persistent microalbuminuria	24
30323	C10EK00	Type 1 diabetes mellitus with persistent proteinuria	21
24423	C108.13	Type I diabetes mellitus	9
18387	C10E700	Type 1 diabetes mellitus with retinopathy	7
35288	C10E800	Type 1 diabetes mellitus - poor control	7
55239	C10EQ00	Type 1 diabetes mellitus with gastroparesis	6
18683	C10E500	Type 1 diabetes mellitus with ulcer	2
21983	C108012	Type 1 diabetes mellitus with renal complications	2
22871	C10EP00	Type 1 diabetes mellitus with exudative maculopathy	2
40837	C10EN00	Type 1 diabetes mellitus with ketoacidotic coma	2
54008	C10EJ00	Type 1 diabetes mellitus with neuropathic arthropathy	2
12455	C10E.11	Type I diabetes mellitus	1
18230	C108J12	Type 1 diabetes mellitus with neuropathic arthropathy	1
38161	C108711	Type I diabetes mellitus with retinopathy	1
39070	C10EE00	Type 1 diabetes mellitus with hypoglycaemic coma	1
43921	C10E400	Unstable type 1 diabetes mellitus	1
46850	C108811	Type I diabetes mellitus - poor control	1
47649	C10E100	Type 1 diabetes mellitus with ophthalmic complications	1
47650	C10E300	Type 1 diabetes mellitus with multiple complications	1
49554	C10EF00	Type 1 diabetes mellitus with diabetic cataract	1
49949	C10E411	Unstable type I diabetes mellitus	1
99231	C108B11	Type I diabetes mellitus with mononeuropathy	1
17545	C108F11	Type I diabetes mellitus with diabetic cataract	0
18642	C10EH00	Type 1 diabetes mellitus with arthropathy	0
40682	C10E900	Type 1 diabetes mellitus maturity onset	0
41049	C108712	Type 1 diabetes mellitus with retinopathy	0
42729	C108E11	Type I diabetes mellitus with hypoglycaemic coma	0
42831	C10E200	Type 1 diabetes mellitus with neurological complications	0
45914	C108812	Type 1 diabetes mellitus - poor control	0
46301	C10EC00	Type 1 diabetes mellitus with polyneuropathy	0
47582	C10E000	Type 1 diabetes mellitus with renal complications	0
49146	C108211	Type I diabetes mellitus with neurological complications	0
51957	C108511	Type I diabetes mellitus with ulcer	0
60107	C108411	Unstable type I diabetes mellitus	0
60208	C108J11	Type I diabetes mellitus with neuropathic arthropathy	0
61344	C108011	Type I diabetes mellitus with renal complications	0
61829	C108212	Type 1 diabetes mellitus with neurological complications	0
62209	C10EM11	Type I diabetes mellitus with ketoacidosis	0
62352	C108H11	Type I diabetes mellitus with arthropathy	0
62613	C10EA11	Type I diabetes mellitus without complication	0
63017	C108911	Type I diabetes mellitus maturity onset	0

medcode	readcode	readterm	n
66145	C10EN11	Type I diabetes mellitus with ketoacidotic coma	0
66872	C108D11	Type I diabetes mellitus with nephropathy	0
68105	C10EB00	Type 1 diabetes mellitus with mononeuropathy	0
68390	C108512	Type 1 diabetes mellitus with ulcer	0
69043	ZC2C900	Dietary advice for type I diabetes	0
69676	C10EA00	Type 1 diabetes mellitus without complication	0
69993	C10E600	Type 1 diabetes mellitus with gangrene	0
70766	C108E12	Type 1 diabetes mellitus with hypoglycaemic coma	0
91942	C10E311	Type I diabetes mellitus with multiple complications	0
91943	C10EC11	Type I diabetes mellitus with polyneuropathy	0
93468	C10EG00	Type 1 diabetes mellitus with peripheral angiopathy	0
93878	C10E511	Type I diabetes mellitus with ulcer	0
95343	C10E711	Type I diabetes mellitus with retinopathy	0
95992	C108A11	Type I diabetes mellitus without complication	0
96235	C10E911	Type I diabetes mellitus maturity onset	0
97446	C108912	Type 1 diabetes mellitus maturity onset	0
97474	C108412	Unstable type 1 diabetes mellitus	0
97894	C10EP11	Type I diabetes mellitus with exudative maculopathy	0
99311	C10E111	Type I diabetes mellitus with ophthalmic complications	0
102112	C10E611	Type I diabetes mellitus with gangrene	0
102620	C10EL11	Type I diabetes mellitus with persistent microalbuminuria	0
102704	66At000	Type I diabetic dietary review	0
102740	C108112	Type 1 diabetes mellitus with ophthalmic complications	0
104453	66At011	Type 1 diabetic dietary review	0
105337	C10E811	Type I diabetes mellitus - poor control	0
108007	C108311	Type I diabetes mellitus with multiple complications	0
108724	C10EQ11	Type I diabetes mellitus with gastroparesis	0

Type 2 Diabetes

medcode	readcode	readterm	n
758	C10F.00	Type 2 diabetes mellitus	38645
17859	C109.12	Type 2 diabetes mellitus	3302
83532	66Ao.00	Diabetes type 2 review	948
18390	C10FM00	Type 2 diabetes mellitus with persistent microalbuminuria	645
1407	C10FJ00	Insulin treated Type 2 diabetes mellitus	476
26054	C10FL00	Type 2 diabetes mellitus with persistent proteinuria	297
18278	C109J00	Insulin treated Type 2 diabetes mellitus	193
18219	C109.13	Type II diabetes mellitus	185
12640	C10FC00	Type 2 diabetes mellitus with nephropathy	144
22884	C10F.11	Type II diabetes mellitus	78
25627	C10F700	Type 2 diabetes mellitus - poor control	63

medcode	readcode	readterm	n
18496	C10F600	Type 2 diabetes mellitus with retinopathy	36
32627	C10FN00	Type 2 diabetes mellitus with ketoacidosis	33
47954	C10F900	Type 2 diabetes mellitus without complication	27
34450	C10FK00	Hyperosmolar non-ketotic state in type 2 diabetes mellitus	20
18777	C10F000	Type 2 diabetes mellitus with renal complications	16
53392	C10F911	Type II diabetes mellitus without complication	12
63690	C10FR00	Type 2 diabetes mellitus with gastroparesis	10
36695	C10D.00	Diabetes mellitus autosomal dominant type 2	9
18425	C10FB00	Type 2 diabetes mellitus with polyneuropathy	8
25591	C10FQ00	Type 2 diabetes mellitus with exudative maculopathy	7
34268	C10F200	Type 2 diabetes mellitus with neurological complications	7
35385	C10FH00	Type 2 diabetes mellitus with neuropathic arthropathy	7
46917	C10FD00	Type 2 diabetes mellitus with hypoglycaemic coma	7
18264	C109J12	Insulin treated Type II diabetes mellitus	6
49074	C10F400	Type 2 diabetes mellitus with ulcer	6
44982	C10FE00	Type 2 diabetes mellitus with diabetic cataract	5
45919	C109212	Type 2 diabetes mellitus with neurological complications	4
62674	C10FA00	Type 2 diabetes mellitus with mononeuropathy	4
12736	C10F500	Type 2 diabetes mellitus with gangrene	3
24836	C109C12	Type 2 diabetes mellitus with nephropathy	3
37806	C10FF00	Type 2 diabetes mellitus with peripheral angiopathy	3
47315	C10F711	Type II diabetes mellitus - poor control	3
47321	C10F100	Type 2 diabetes mellitus with ophthalmic complications	2
48192	C109E11	Type II diabetes mellitus with diabetic cataract	2
61071	C109D12	Type 2 diabetes mellitus with hypoglycaemic coma	2
42762	C109612	Type 2 diabetes mellitus with retinopathy	1
45913	C109712	Type 2 diabetes mellitus - poor control	1
47409	C109B11	Type II diabetes mellitus with polyneuropathy	1
54899	C109F11	Type II diabetes mellitus with peripheral angiopathy	1
55075	C109411	Type II diabetes mellitus with ulcer	1
59253	C10FG00	Type 2 diabetes mellitus with arthropathy	1
65267	C10F300	Type 2 diabetes mellitus with multiple complications	1
65704	C109412	Type 2 diabetes mellitus with ulcer	1
66965	C109H12	Type 2 diabetes mellitus with neuropathic arthropathy	1
98616	C10F211	Type II diabetes mellitus with neurological complications	1
18143	C109G11	Type II diabetes mellitus with arthropathy	0
18209	C109012	Type 2 diabetes mellitus with renal complications	0
24458	C109711	Type II diabetes mellitus - poor control	0
43227	C10F311	Type II diabetes mellitus with multiple complications	0
44779	C109E12	Type 2 diabetes mellitus with diabetic cataract	0
46150	C109512	Type 2 diabetes mellitus with gangrene	0
47816	C109H11	Type II diabetes mellitus with neuropathic arthropathy	0

medcode	readcode	readterm	n
49655	C10F611	Type II diabetes mellitus with retinopathy	0
49869	C109G12	Type 2 diabetes mellitus with arthropathy	0
50225	C109011	Type II diabetes mellitus with renal complications	0
50527	C10FB11	Type II diabetes mellitus with polyneuropathy	0
50813	C109A11	Type II diabetes mellitus with mononeuropathy	0
51756	C10FP00	Type 2 diabetes mellitus with ketoacidotic coma	0
56268	C109D11	Type II diabetes mellitus with hypoglycaemic coma	0
57278	C10F011	Type II diabetes mellitus with renal complications	0
58604	C109611	Type II diabetes mellitus with retinopathy	0
59725	C109111	Type II diabetes mellitus with ophthalmic complications	0
60699	C109F12	Type 2 diabetes mellitus with peripheral angiopathy	0
60796	C10FL11	Type II diabetes mellitus with persistent proteinuria	0
62107	C109511	Type II diabetes mellitus with gangrene	0
64571	C109C11	Type II diabetes mellitus with nephropathy	0
64668	C10FJ11	Insulin treated Type II diabetes mellitus	0
67905	C109211	Type II diabetes mellitus with neurological complications	0
70316	C109112	Type 2 diabetes mellitus with ophthalmic complications	0
85991	C10FM11	Type II diabetes mellitus with persistent microalbuminuria	0
91646	C10F411	Type II diabetes mellitus with ulcer	0
93727	C10FE11	Type II diabetes mellitus with diabetic cataract	0
95351	C10FA11	Type II diabetes mellitus with mononeuropathy	0
98723	C10FD11	Type II diabetes mellitus with hypoglycaemic coma	0
100964	C10F111	Type II diabetes mellitus with ophthalmic complications	0
101801	66At100	Type II diabetic dietary review	0
102201	C10FC11	Type II diabetes mellitus with nephropathy	0
102611	66At111	Type 2 diabetic dietary review	0
103902	C10FG11	Type II diabetes mellitus with arthropathy	0
104323	C10F511	Type II diabetes mellitus with gangrene	0
104639	C10FF11	Type II diabetes mellitus with peripheral angiopathy	0
105784	C109912	Type 2 diabetes mellitus without complication	0
106061	C10FP11	Type II diabetes mellitus with ketoacidotic coma	0
106528	C10FN11	Type II diabetes mellitus with ketoacidosis	0
107701	C10FK11	Hyperosmolar non-ketotic state in type II diabetes mellitus	0
108005	C109312	Type 2 diabetes mellitus with multiple complications	0
109103	C109911	Type II diabetes mellitus without complication	0
109197	C10FH11	Type II diabetes mellitus with neuropathic arthropathy	0

Appendix C.8 - Medical Codes for Diabetes Education Programmes

Dose-Adjustment for Normal Eating

medcode	readcode	readterm	n
93390	9OLH.00	Attended DAFNE diabetes structured education programme	5
93491	9OLJ.00	DAFNE diabetes structured education programme completed	2
93704	8Hj3.00	Referral to DAFNE diabetes structured education programme	2
97809	8I82.00	Did not complete DAFNE diabetes structured education program	0
99277	9NiC.00	Did not attend DAFNE diabetes structured education programme	0
106953	8IEa.00	Referral to DAFNE diabetes structured educn prog declined	0

Diabetes Education and Self-Management for Ongoing and Newly Diagnosed

medcode	readcode	readterm	n
93657	8Hj4.00	Referral to DESMOND diabetes structured education programme	324
95159	9NiD.00	Did not attend DESMOND diabetes structured education program	49
93529	9OLK.00	DESMOND diabetes structured education programme completed	39
95093	8I83.00	Did not complete DESMOND diabetes structured educat program	0

X-PERT

medcode	readcode	readterm	n
93870	8Hj5.00	Referral to XPERT diabetes structured education programme	262
93631	9OLL.00	XPERT diabetes structured education programme completed	44
94955	9NiE.00	Did not attend XPERT diabetes structured education programme	41
94011	9OLG.00	Attended XPERT diabetes structured education programme	17
94956	8I84.00	Did not complete XPERT diabetes structured education program	1

Other

medcode	readcode	readterm	n
47011	8Hj0.00	Referral to diabetes structured education programme	490
12682	679R.00	Patient offered diabetes structured education programme	292
26605	9OLB.00	Attended diabetes structured education programme	167
93854	9OLM.00	Diabetes structured education programme declined	23
94186	9OLF.00	Diabetes structured education programme completed	17

medcode	readcode	readterm	n
95094	8I81.00	Did not complete diabetes structured education programme	2
51066	9OLC.00	Family/carer attended diabetes structured education prog	1
95553	9NiA.00	Did not attend diabetes structured education programme	1
107414	8I94.00	Diabetes structured education programme not available	0

Appendix D

Appendix D.1 - Independent Scientific Advisory Committee Protocol

The document in the following section is a copy of the protocol sent to the Independent Scientific Advisory Committee (ISAC) on 16th January 2015. The protocol was accepted on 16th March 2015. Copies of the ISAC application form and the protocol annexes are not enclosed.

Diabetic Kidney Disease and the Competing Risks of Cardiovascular and Cancer Mortality in Type 2 Diabetes: A Methodological Study of Standard- vs. Competing Risks Survival Analyses Methods

Summary of Research in Plain English

Patients with type 2 diabetes and diabetic kidney disease are at increased risk of multiple medical conditions, including heart disease or strokes (cardiovascular diseases) and cancers[69]. This study aims to estimate the risks diabetic kidney disease poses to death from these outcomes, and to compare different statistical methods for this kind of research, as discussed below.

Many current methods of analyzing follow up studies consider only one cause of death at a time. The occurrence of death from other ("competing") causes can distort the predictions of risk made using these methods. The causes of death we will consider are death from cardiovascular disease (CVD) and cancer, which are common in people with type 2 diabetes and diabetic kidney disease[69]. We will compare risk estimates from standard statistical methods to risk estimates from "competing risks" methods.

A population of patients with type 2 diabetes will be identified in CPRD. This will be analyzed using standard- and competing-risks methods to determine effect of diabetic kidney disease on the 10-year risks of cardiovascular and cancer death.

This study has the potential to demonstrate the impact on risk estimates of incorporating competing risks methodology, with application to diabetic kidney disease in type 2 diabetes.

Background

Diabetic Kidney Disease

Diabetic kidney disease is a complication of diabetes mellitus that manifests in 25% of patients within 10 years of type 2 diabetes onset[29]. It is typically characterized by an initial onset of albuminuria, proceeded by a decline in glomerular filtration rate (GFR)[181]. Diabetes mellitus[182], albuminuria[183] and reduced GFR[184] are strong and well-established independent risk factors for a range of serious cardiovascular outcomes, and have all been linked to the occurrence of cancer[71,183,184].

Standard Survival Analysis

Prognostic studies typically use standard time-to-event analyses, such as Kaplan-Meier curves[3] or Cox regression[4] to generate risk estimates. These models assume that at any given moment, censoring is not associated with the probability of the study outcome occurring; i.e. the probability of being censored for any subject at time *t* does not depend upon that subject's prognosis for failure at time *t*[185]. However, studies often censor individuals who have succumbed to competing causes of failure.

This violates the non-informative censoring assumption, as those who have died from competing causes will be unable to go on and experience the main study outcome. Thus standard methods may be liable to incorporate bias in the presence of competing risks. Estimates from such models are particularly liable to bias in frail populations, or populations at high risk of competing causes of failure. In particular, we hypothesize that such bias may affect studies of disease outcomes in people with type 2 diabetes and diabetic kidney disease, as at any one time they are at high risk of mortality from numerous cardiovascular, cancerous or diabetes-related causes. e.g. when investigating the progression of nephropathy in type 2 diabetes, Adler et al., 2003 demonstrated that once nephropathy was established, the risk of mortality was always greater than that of progressing to either more severe nephropathy or end-stage renal disease[29].

Brief Literature Overview

Models have been developed that are able to account for the presence of competing risks in their analysis. These include:

1. ***The Cumulative Incidence Competing Risk (CICR) model***[5] ~ for cumulative incidence estimates.
2. ***The Lunn-McNeil model***[16] ~ for estimates of covariate effects on the relative hazard.
3. ***The Fine-Gray model***[20] ~ for estimates of covariate effects on the cumulative incidence function.

I conducted a systematic review to determine how widely used competing risks methodologies were in large studies of people with type 2 diabetes and diabetic kidney disease, for cardiovascular and cancer outcomes. Of the 29 studies included in review: 8 performed no analyses susceptible to competing risks, while the remaining 21 used either Kaplan-Meier curves or Cox regression at some point within their analyses. The use of competing risks methodology was completely absent in the included studies.

Objectives, Specific Aims and Rationale

Primary Objective

To evaluate how competing risks methodology affects estimates of cardiovascular and cancer mortality risk in diabetic kidney disease in type 2 diabetes.

Specific Aims

- To compare estimates of cumulative incidence of cardiovascular and cancer mortality by albuminuria status, from competing risks methodology [CICR] vs. standard methods [Kaplan-Meier].
- To compare estimates of relative hazard of cardiovascular and cancer mortality by albuminuria status from competing risks [Lunn-McNeil] vs. standard methodology [Cox regression]
- To compare the relative effect of albuminuria status on the cause-specific-hazard [Lunn-McNeil] and subdistribution-hazard [Fine-Gray] of cardiovascular and cancer mortality.

Rationale

Patients with type 2 diabetes and diabetic kidney disease are at increased risk of multiple causes of death[69]. CVD and cancer are relatively common outcomes in people with type 2 diabetes and diabetic kidney disease; CVD representing the more common outcome (81.4 vs. 22.4 events per 1,000 person-years). Thus, cancer mortality poses a weak competing risk to cardiovascular mortality, and cardiovascular mortality poses a strong competing risk to cancer mortality.

Typically, study investigators use Kaplan-Meier or Nelson-Aalen curves to derive absolute risk estimates in terms of cumulative incidence; and Cox or parametric-model-type regression to derive relative risk estimates in terms of hazards ratios. Both of these may be liable to bias in the presence of competing risks. Through the use of competing risks methodology, it may be possible to produce less biased absolute and relative risk estimates. Furthermore, the comparison of risk estimates from cause-specific- and subdistribution-hazards models will allow for the comparison of aetiological and prognostic relationships between diabetic kidney disease and both outcomes.

Study Type

Methodological study.

Study Design

Cohort study.

Data Linkage

As the competing event of interest will either be cancer or cardiovascular mortality, Office for National Statistics (ONS) mortality data linkage will be required for accurate ascertainment of the cause of death.

A scoping search has revealed there to be 280,728 patients registered within CPRD on 1st January 2005 at up-to-standard practices, who have either been prescribed anti-diabetic medication or assigned diagnostic codes conferring the presence of diabetes mellitus. Of these, 150,604 are eligible for ONS-linkage. Using an identification strategy piloted on data from a previous CPRD project (protocol 12_091R), it is estimated the 91% of individuals within this population will have type 2 diabetes, (i.e. 137,050 patients). This figure is less than the proposed study sample size of 150,000. However, the updated versions of the ONS-linked data are due imminently and should raise the available eligible population within CPRD above the sample size requirement. Failing this, we will require all eligible individuals.

Study Population

This is an open cohort study of type 2 diabetes patients (≥35 years of age) registered at "up-to-standard" (UTS) CPRD practices. Subjects will be excluded who have:

- Ambiguously recorded gender.
- Monogenetic-, bronzed- (haemochromatosis), gestational- or drug-induced-diabetes.
- Polycystic ovary syndrome.

- Renal or pancreatic transplantation at any time prior to study entry.

The study start date will be 1st January 2005. Eligible patients will be age 35 or over on this date, and registered with the practice for a minimum of 2 years prior to study entry (to ensure adequate recording of baseline covariates). Eligibility will be defined using all available UTS data prior to entry date. The index date for each patient will be the latest of the following dates: the study start date, practice up-to-standard date or date of registration with the practice plus 2 years.

Study end date will be 1st January 2015 (or date of last available linked data). Follow-up ends at this date or (if earlier) at time of death or transfer out of CPRD, or (where applicable) date of undergoing renal or pancreatic transplantation. Patient records will be censored at the earliest of the following dates: study end date, date of last upload of practice data, date of death, transfer out date, and date of incident record of renal or pancreatic transplantation within study period.

Diabetes mellitus will be identified in CPRD using a combination of medical codes conveying the disease presence (see Annex A) and product codes for anti-diabetic prescriptions (see Annex B). For detailed information regarding the identification strategy, please contact Benjamin Feakins.

Study Primary Comparisons

The main comparison of this study will be that of competing risks methodology vs. standard survival analyses.

- For estimates of **cumulative incidence,** outputs from Kaplan-Meier and the CICR[5] model will be compared.
- For **method-specific relative hazard** metrics, estimates from the Cox and Lunn-McNeil[16] models will be compared to provide inference on the effect of incorporating competing risks methodology. Estimates from the Lunn-McNeil and Fine-Gray[20] models will be compared to gain insight onto the differential effects of albuminuria on the cause-specific hazards and cumulative incidence functions of all outcomes, within a competing risks framework.

Clinical Exposures, Outcomes and Covariates

Exposures

Albuminuria Status - [Binary; No units]

Albuminuria status will be defined using medical codes indicating status (Annex B1) or by categorizing test results as shown below (Table 1). Test results will be extracted from entity codes for urine microalbumin (code 435) and from general testing entity codes (urinalysis protein (287) and urine biochemistry (340) when these are associated with read codes indicating albuminuria/proteinuria testing (Annex B2).

Table 1 - Classification of albuminuria, (adapted from Kidney Disease Improving Global Outcomes guidelines)[28]. Abbreviations: AER = Albumin Excretion Rate; ACR = Albumin : Creatinine Ratio; PER = Protein Excretion Rate; PCR = Protein : Creatinine Ratio.

			Categories	
			A1	**≥A2**
Biomarkers & Tests	**Albuminuria**	**AER (mg/day)**	<30	≥30
		ACR (mg/mmol)	<3	≥3
		ACR (mg/g)	<30	≥30
	Proteinuria	**PER (mg/day)**	<150	≥150
		PCR (mg/mmol)	<15	≥15
		PCR (mg/g)	<150	≥150

To establish baseline values, the nearest recorded values to index date will be used up to 2 years before index date. If individuals possess no values within this time window then the omissions will be treated as missing data.

Outcomes

All outcomes will be established using a combination of UK ONS death certificates and CPRD. Outcomes will be defined as:

- **Cardiovascular Mortality** - Death from cardiovascular causes - defined as the occurrence of death, with ICD-10 codes: I10 - I79 listed as the "underlying cause" on the death certificate.
- **Cancer Mortality** - Death from neoplastic causes - defined as occurrence of death, with ICD-10 codes: C00 - C97 listed as the "underlying cause" on the death certificate.

Covariates

Current Age - [Continuous; Years]

Estimated through taking the midpoint in the year represented by the "birthyear" variable within the patient file (i.e. June 1st), and subtracting the date from the individual's cohort entry date.

Duration of Diabetes - [Continuous; Years]

Estimated by taking the date of first occurrence of either: a medical code from the clinical file relating to the presence of diabetes mellitus (see Annex A), or a product code from the therapy file relation to a prescription for anti-diabetic medication (see Annex A). This date is then subtracted from the individual's cohort entry date.

Gender - [Binary; No units]

Gender will be obtained from the patient file, where the "gender" variable indexes it as 1 for male and 2 for female. It will be recoded as 0 for male and 1 for female.

Body Mass Index (BMI) - [Continuous; kg/m²]

Estimated through the use of values linked to entity codes 13 *[Weight]* and 14 *[Height]*. It is calculated through dividing a patient's weight by their height². The closest values from up to two years preceding an individual's cohort entry date will be used to calculate a representation of their BMI. Where no values are found within this time window, values will be treated as missing data.

Smoking Status - [Categorical; "Never", "Current", "Ex"]

Estimated through the use of medical codes indicating the presence of smoking, and values linked to entity code 4 *[Smoking]*. Patients will be classified as follows:

- Where no codes indicating the presence of smoking are found at any point in a patient's history they will be classified as "**Never**" having smoked. (Sensitivity analyses will evaluate the significance of this assumption).
- Where patients possess codes to indicate the presence of smoking at any point prior to cohort entry, but within two years of this date possess either: no codes to indicate the presence of smoking; a most recent code indicating smoking cessation, they will be classified as "**Ex**" smokers.
- Where the most recent codes present within two years of cohort entry indicate the presence of smoking, the patient will be classified as a "**Current**" smoker.

Systolic blood pressure (SBP) - [Continuous; mg Hg]

Estimated through the use of values linked to entity code 1 *[Blood Pressure]*. The closest values from up to two years preceding an individual's cohort entry date will be used to represent their SBP. Where no values are found within this time window, values will be treated as missing data.

Glycated Haemoglobin (HbA1c) - [Continuous; mmol/mol]

Estimated through the use of values relating to entity code 275 *[HbA1c - Diabetic Control]*. All values will be converted into International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) units (mmol/mol). The closest value from up to two years preceding an individual's cohort entry date will be used as a representation of their HbA1c. Where no values are found within this time window, values will be treated as missing data.

Cholesterol Ratio (Ratio of "Total Serum Cholesterol: High Density Lipoprotein (HDL) Cholesterol") - [Continuous; No units]

Values linked to entity code 163 *[Serum Cholesterol]*, 175 *[High Density Lipoprotein]* and 214 *[Blood Lipids]* will be used to calculate the cholesterol ratio. The closest value from up to two years preceding an individual's cohort entry date will be used as a representation of their cholesterol ratio. Where no values are found within this time window, values will be treated as missing data.

Sample Size

For this methodological comparison, standard sample size methods are not applicable. Previous studies have suggested that the difference between standard- and competing risks estimates may not be large, but confidence intervals were wide and the standard error (0.11) was larger than the effect size (0.07)[70]. Within these studies the standard error estimates around the regression coefficients (β's) demonstrated proportionality towards the total number of events as opposed to the event rate[70,140]. Thus, scaling up the expected number of events to represent a population of 150,000 people with type 2 diabetes and 10 years of follow up would reduce the standard errors of the regression coefficients of previous studies by a factor of 10[70]. Conservative hazard ratio estimates from the literature suggest the effect of albuminuria on major cardiovascular events and incident cancer to be 1.84 (1.46-2.31)[141] and 1.53 (0.76-3.08)[140] respectively. In 150,000 people with type 2 diabetes, the regression coefficient standard errors would be reduced to 0.017 and 0.022. This would yield projected 95% confidence intervals around the hazard ratio estimates of 1.81-1.87 and 1.47-1.58 respectively, giving the best estimates to date of the importance of competing risks methodology in diabetes outcomes research.

Statistical Analysis

All analyses will be carried out using a combination a Stata[91] and R[186]. For time-to-event analyses, "time since cohort entry" will be used as the underlying time function. The following table gives an overview of all analyses models that will be employed within this study.

Table 2 - Overview of analyses to be used.

	Model ID #	Models	Analysis Type	Output Risk Metric	Adjusted / Unadjusted	Encapsulation of Albuminuria	Outcomes
Absolute Risk Metrics	1	Kaplan-Meier	Standard	Cumulative Incidence	Unadjusted	Stratification Variable	CVD Mortality
							Cancer Mortality
	2	Cumulative Incidence	Competing Risks	Cumulative Incidence	Unadjusted	Stratification Variable	[1] CVD Mortality, [2] Cancer Mortality, [3] Other Mortality
Relative Risk Metrics	3	Cox Proportional Hazards Regression	Standard	Hazard Ratio	Unadjusted & Adjusted	Covariate	CVD Mortality
							Cancer Mortality
	4	Lunn-McNeil	Competing Risks	Hazard Ratio	Unadjusted & Adjusted	Covariate	[1] CVD Mortality, [2] Cancer Mortality, [3] Other Mortality
	5	Fine-Gray	Competing Risks	Subhazard Ratio	Unadjusted & Adjusted	Covariate	[1] CVD Mortality, [2] Non-CVD Mortality
							[1] Cancer Mortality, [2] Non-Cancer Mortality

To compare estimates of cumulative incidence of cardiovascular and cancer mortality by albuminuria status, from competing risks methodology vs. standard methods.

Standard estimates for the 10-year cumulative incidence of cardiovascular and cancer mortality will be derived through the use of Kaplan-Meier[3] curves (Table 2; Model ID 1). Competing-Risks-adjusted estimates for the 10-year cumulative incidence of cardiovascular and cancer mortality will be derived through the use of CICR curves[5] (Table 2 Model ID 2). Plots will be stratified by albuminuria status for all outcomes.

To compare estimates of relative hazard of cardiovascular and cancer mortality for albuminuria status from competing risks vs. standard methodology.

Standard relative hazard estimates for the effect of albuminuria status on the 10-year risks of cardiovascular and cancer mortality will be obtained using Cox regression[4] (Table 2; Model ID 3). Competing-risks-adjusted estimates for the effect of albuminuria status on the cause-specific relative hazards of cardiovascular and cancer mortality will be obtained through the use of the Lunn-McNeil model[16] (Table 2; Model ID 4).

Prior to performing analyses, the proportionality assumptions will be checked for each predictor in the models with respect to each outcome using diagnostic plots based on weighted Schoenfeld residuals, developed by Therneau and Grambsch[167].

To compare the relative effect of albuminuria status on the cause-specific hazard and subdistribution hazard of cardiovascular and cancer mortality.

The Fine-Gray subdistribution hazards model will be used to generate subdistribution hazard ratios, for the 10-year risk of each outcome. The competing outcome structure will differ from the cause-specific hazards models; i.e. competing failure will be modelled as "all mortality not related to the incident event under investigation", as represented by Table 2; Model ID 5. The subdistribution hazard ratios obtained will be compared to the cause-specific hazard ratios estimates from the Lunn-McNeil Model (Table 2; Model ID 4).

Prior to performing subdistribution hazards model analyses, the proportional subdistribution hazards assumption will be checked for each predictor in the models with respect to each outcome using diagnostic plots based on weighted Schoenfeld-like residuals, developed by Zhou, Fine and Laird[168].

Sensitivity Analyses

Diabetic Kidney Disease Metrics

Analyses using other categorizations of diabetic kidney disease will be explored, (e.g. eGFR, calculated from creatinine clearance using the 4-part modification of diet in renal disease equation).

Death Certificates Where both CVD and Cancer Listed as Causes

For primary study analyses the cause of death is unambiguously defined as that which is listed as being the "underlying cause" on the death certificate. However, up to 15 other cause of death may also be listed on death certificates as "contributing causes". Scoping searches performed on the ONS data from a previous ISAC application (protocol 12_091R) demonstrated substantial overlap between the number of individuals possessing ICD-10 codes for cardiovascular and cancer mortality on their death certificates. Where both causes of death are present on the death certificate,

sensitivity analyses will investigate the effect on risk estimates of assuming failure from either cancer or cardiovascular causes.

Other Standard Survival Analysis Methods

The Kaplan-Meier method is not the sole means of calculating cumulative incidence in standard survival analysis and other means exist through which this may be achieved. Nelson-Aalen plots[9] are also commonly used for this purpose, yet this method has remained largely un-scrutinized to determine whether it may be biased by the presence of competing risks. Thus, this section of the analysis will compare cumulative incidence estimates derived by Nelson-Aalen methods with those derived from both standard (Kaplan-Meier) and competing risks (CICR) methods.

Missing Data

As this study is primarily interested in methodological issues surrounding competing risks in survival analyses, complete case analysis will be used in all models. Sensitivity analyses will use multiple imputation to compare obtained risk estimates with those in the literature.

Limitations of the Study Design, Data Sources, and Analytic Methods

When using CPRD data, limitations inherent to primary health care data sets (e.g. loss to follow up and misclassification) are well documented. For the proposed study, additional limitations may also need to be considered:

A large issue in the use of CPRD data is the difficulty in establishing type of diabetes. I have spent the past few months working on this issue. During this time I have conducted a systematic review to identify the pragmatic approaches adopted by researchers to identify type of diabetes in CPRD. I have subsequently used these to devise my own coding strategies, which I have piloted on data sourced from a previous CPRD project (protocol 12_091R).

Many of the competing risks statistical approaches outlined within this protocol are relatively new, and are infrequently used within biomedical research. Therefore, these models have not been subject to the same degree of scrutiny, and our study will contribute to better understanding of these models.

One specific limitation to the Fine-Gray model is that it is only able to model two outcomes at once; the second of which must be "all mortality not occurring as the primary outcome". Thus, the model must be run twice, once for each outcome of cardiovascular and cancer mortality.

Plans for Disseminating and Communicating Study Results, Including the Presence or Absence of Any Restrictions on the Extent and Timing of Publication

Upon completion, papers will be submitted to medical and/or statistical journals. Abstracts will also be forwarded to relevant medical and/or statistical.

Amendments

Study Population

This is a closed cohort study of type 2 diabetes patients (≥35 years of age) registered at "up-to-standard" (UTS) CPRD practices. Subjects will be excluded who have:

- Ambiguously recorded gender.
- Monogenic-, bronzed- (haemochromatosis), gestational- or drug-induced-diabetes.
- Polycystic ovary syndrome.
- Renal or pancreatic transplantation at any time prior to study entry.

The study start date will be 1st January 2005. Eligible patients will be aged 35 years and over, possess medical or product codes indicating the presence of type 2 diabetes prior to this date, and have been registered with the practice for a minimum of two years (to ensure adequate recording of baseline covariates). Eligibility will be defined using all available UTS data prior to the study start date.

Appendix D.2 - Medical Codes for Albuminuria

Normoalbuminuria

medcode	readcode	readterm	n
14091	4672	Urine protein test negative	2976
95175	1Z1E.00	Chronic kidney disease stage 3A without proteinuria	703
95123	1Z1C.00	Chronic kidney disease stage 3 without proteinuria	522
17106	46W1.00	Urine microalbumin negative	391
95177	1Z1G.00	Chronic kidney disease stage 3B without proteinuria	237
95406	1Z1J.00	Chronic kidney disease stage 4 without proteinuria	77
95121	1Z1A.00	Chronic kidney disease stage 2 without proteinuria	76
9430	4679	Urine dipstick for protein	22
95572	1Z18.00	Chronic kidney disease stage 1 without proteinuria	6
95405	1Z1L.00	Chronic kidney disease stage 5 without proteinuria	5
95188	1Z1C.11	CKD stage 3 without proteinuria	2
14382	46N1.00	Urine protein normal	0
54369	46O1.00	Urine protein electr. normal	0
95176	1Z1E.11	CKD stage 3A without proteinuria	0
97587	1Z1J.11	CKD stage 4 without proteinuria	0
97683	1Z1L.11	CKD stage 5 without proteinuria	0
97978	1Z1A.11	CKD stage 2 without proteinuria	0
100633	1Z1G.11	CKD stage 3B without proteinuria	0
103935	1IA..00	No evidence of diabetic nephropathy	0
104677	2126A00	Proteinuria resolved	0

Albuminuria

medcode	readcode	readterm	n
10924	R110300	[D]Microalbuminuria	2612
11248	R110.00	[D]Proteinuria	1815
1802	4678	Proteinuria	1081
18390	C10FM00	Type 2 diabetes mellitus with persistent microalbuminuria	645
5451	R110000	[D]Albuminuria	475
13590	4674	Urine protein test = +	368
26054	C10FL00	Type 2 diabetes mellitus with persistent proteinuria	297
28180	46W0.00	Urine microalbumin positive	195
2475	C104.11	Diabetic nephropathy	155
12640	C10FC00	Type 2 diabetes mellitus with nephropathy	144
13612	4673	Urine protein test = trace	137

medcode	readcode	readterm	n
94793	1Z1B.00	Chronic kidney disease stage 3 with proteinuria	124
11873	K03..12	Nephropathy, unspecified	101
95408	1Z1D.00	Chronic kidney disease stage 3A with proteinuria	76
95178	1Z1F.00	Chronic kidney disease stage 3B with proteinuria	70
13611	4675	Urine protein test = ++	60
95122	1Z1H.00	Chronic kidney disease stage 4 with proteinuria	52
38284	R110z00	[D]Proteinuria NOS	47
11875	K02..12	Nephropathy - chronic	44
16465	K190X00	Persistent proteinuria, unspecified	30
13621	4676	Urine protein test = +++	28
95146	1Z19.00	Chronic kidney disease stage 2 with proteinuria	27
10418	C10ED00	Type 1 diabetes mellitus with nephropathy	24
30294	C10EL00	Type 1 diabetes mellitus with persistent microalbuminuria	24
95508	1Z1K.00	Chronic kidney disease stage 5 with proteinuria	23
30323	C10EK00	Type 1 diabetes mellitus with persistent proteinuria	21
13600	4677	Urine protein test = ++++	12
33580	K03..00	Nephritis and nephropathy unspecified	12
25573	R105800	[D]Abnormality of albumin	10
35107	C104z00	Diabetes mellitus with nephropathy NOS	8
4850	K03..11	Nephritis and nephropathy unspecified	4
13613	46N2.00	Urine protein abnormal	3
24836	C109C12	Type 2 diabetes mellitus with nephropathy	3
94789	1Z17.00	Chronic kidney disease stage 1 with proteinuria	3
41159	K0C1.00	Nephropathy induced by other drugs meds and biologl substncs	2
59365	C109C00	Non-insulin dependent diabetes mellitus with nephropathy	2
36243	K136.11	Orthostatic proteinuria	1
44270	K0A5200	Hereditry nephropathy NEC,difus membran glomerulnephritis	1
95145	1Z1B.11	CKD stage 3 with proteinuria	1
95571	1Z1D.11	CKD stage 3A with proteinuria	1
97980	1Z17.11	CKD stage 1 with proteinuria	1
2482	D011100	Vit B12 defic anaemia due to malabsorption with proteinuria	0
14901	K136.00	Benign postural proteinuria	0
31969	4I3B.11	Albumin in sample	0
36205	K0A5.00	Hereditary nephropathy not elsewhere classified	0
41239	K0A5100	Hereditary nephropathy NEC,focal+segmnt glomerular lesion	0
43611	K0A4.00	Isolated proteinuria with specified morphological lesion	0
45499	K01x111	Kimmelstiel - Wilson disease	0
50893	K0C4.00	Toxic nephropathy, not elsewhere classified	0
51113	K0A5000	Hereditary nephropathy NEC, minor glomerular abnormality	0
56736	46O2.00	Urine protein electr. abnormal	0
57621	C108D00	Insulin dependent diabetes mellitus with nephropathy	0
57784	K0C2.00	Nephropathy induced by unspec drug medicament or biol subs	0

medcode	readcode	readterm	n
59992	K0A4W00	Isolated proteinuria, with unspecified morpholog changes	0
60796	C10FL11	Type II diabetes mellitus with persistent proteinuria	0
62980	K0A5X00	Hereditary nephropathy, unspecif morphological changes	0
64030	Kyu5G00	[X]Persistent proteinuria, unspecified	0
64571	C109C11	Type II diabetes mellitus with nephropathy	0
66872	C108D11	Type I diabetes mellitus with nephropathy	0
72478	Kyu1400	[X]Nephropathy induced by other drugs+biological substances	0
85991	C10FM11	Type II diabetes mellitus with persistent microalbuminuria	0
91738	K0A5600	Hereditary nephropathy, NEC, dense deposit disease	0
95180	1Z1F.11	CKD stage 3B with proteinuria	0
97979	1Z19.11	CKD stage 2 with proteinuria	0
99160	1Z1K.11	CKD stage 5 with proteinuria	0
99312	1Z1H.11	CKD stage 4 with proteinuria	0
101572	K0A4X00	Isolated proteinuria, with oth specif morpholog changes	0
102163	C10ED12	Insulin dependent diabetes mellitus with nephropathy	0
102201	C10FC11	Type II diabetes mellitus with nephropathy	0
102210	7L1f000	Extracorporeal albumin haemodialysis	0
102620	C10EL11	Type I diabetes mellitus with persistent microalbuminuria	0
105302	K08yA00	Proteinuric diabetic nephropathy	0
107216	Kyu0F00	[X]Hereditary nephropathy, unspecif morphological changes	0
107881	K08yA11	Clinical diabetic nephropathy	0

Appendix D.3 - Medical Codes for Urine Creatinine

medcode	readcode	readterm	n
14411	46M7.00	Urine creatinine	49

Appendix E

Appendix E.1 - Medical Codes for Smoking Status

Never Smoked

medcode	readcode	readterm	n
33	1371	Never smoked tobacco	512912
60	137L.00	Current non-smoker	79214
11788	1371.11	Non-smoker	28662
98177	9kn..00	Non-smoker annual review - enhanced services administration	102
101878	9kn..11	Non-smoker annual review	0

Ex-Smoker

medcode	readcode	readterm	n
90	137S.00	Ex smoker	391881
776	137K.00	Stopped smoking	10612
12946	137F.00	Ex-smoker - amount unknown	9895
12955	1379	Ex-moderate smoker (10-19/day)	8184
12956	137A.00	Ex-heavy smoker (20-39/day)	5381
12957	1378	Ex-light smoker (1-9/day)	4857
12878	137T.00	Date ceased smoking	1714
12959	137B.00	Ex-very heavy smoker (40+/day)	1358
34127	13p1.00	Smoking status at 4 weeks	1334
12961	1377	Ex-trivial smoker (<1/day)	1187
26470	137N.00	Ex pipe smoker	423
34374	13p2.00	Smoking status between 4 and 52 weeks	348
19488	137O.00	Ex cigar smoker	260
10898	13p4.00	Smoking free weeks	220
97210	137j.00	Ex-cigarette smoker	41
6359	E023.00	Nicotine withdrawal	37
99838	137K000	Recently stopped smoking	32
41405	13p3.00	Smoking status at 52 weeks	22

medcode	readcode	readterm	n
100963	9km..11	Ex-smoker annual review	9
100495	137I.00	Ex roll-up cigarette smoker	5
72700	ZV11600	[V]Personal history of tobacco abuse	0
72706	E251300	Tobacco dependence in remission	0
98447	9km..00	Ex-smoker annual review - enhanced services administration	0
103208	13p7.00	Smoking status at 12 weeks	0
105710	8HBP.00	Smoking cessation 12 week follow-up	0

Current Smoker

medcode	readcode	readterm	n
93	137P.00	Cigarette smoker	204295
7622	8CAL.00	Smoking cessation advice	183575
54	137..00	Tobacco consumption	134685
2111	6791	Health ed. - smoking	126087
10558	137R.00	Current smoker	16443
1878	1374	Moderate smoker - 10-19 cigs/d	13405
11356	9N2k.00	Seen by smoking cessation advisor	9395
1823	137P.11	Smoker	9330
12240	137G.00	Trying to give up smoking	9060
3568	1375	Heavy smoker - 20-39 cigs/day	8723
12944	1373	Light smoker - 1-9 cigs/day	8463
10211	13p..00	Smoking cessation milestones	5124
12945	137M.00	Rolls own cigarettes	3645
7130	9OO..12	Stop smoking monitoring admin.	3561
12941	1372.11	Occasional smoker	3409
12947	137H.00	Pipe smoker	3155
18573	8H7i.00	Referral to smoking cessation advisor	3089
12943	137J.00	Cigar smoker	3085
12953	9OO1.00	Attends stop smoking monitor.	2936
9833	8B2B.00	Nicotine replacement therapy	2712
10742	8HTK.00	Referral to stop-smoking clinic	2705
34126	13p0.00	Negotiated date for cessation of smoking	2178
12942	137..11	Smoker - amount smoked	1915
12958	1372	Trivial smoker - < 1 cig/day	1794
74907	745H.00	Smoking cessation therapy	1721
42722	9OO4.00	Stop smoking monitor 1st lettr	1615
28834	9OO..00	Anti-smoking monitoring admin.	1074
30762	137d.00	Not interested in stopping smoking	988
11527	9N4M.00	DNA - Did not attend smoking cessation clinic	950
1822	1376	Very heavy smoker - 40+cigs/d	939
12619	9hG1.00	Excepted from smoking quality indicators: Informed dissent	909

medcode	readcode	readterm	n
38112	13p5.00	Smoking cessation programme start date	903
18926	67H1.00	Lifestyle advice regarding smoking	723
21637	9OOZ.00	Stop smoking monitor admin.NOS	697
12964	137C.00	Keeps trying to stop smoking	672
12967	137a.00	Pipe tobacco consumption	668
19485	9OOA.00	Stop smoking monitor.chck done	668
32083	9OO..11	Stop smoking clinic admin.	517
31114	137b.00	Ready to stop smoking	475
12951	137Q.11	Smoking restarted	448
12963	137Y.00	Cigar consumption	415
40418	9OO2.00	Refuses stop smoking monitor	360
100099	8IAj.00	Smoking cessation advice declined	328
32687	E251.00	Tobacco dependence	325
9045	ZG23300	Advice on smoking	257
98137	67H6.00	Brief intervention for smoking cessation	249
12960	137Z.00	Tobacco consumption NOS	242
30423	137c.00	Thinking about stopping smoking	231
12952	137Q.00	Smoking started	225
12966	137V.00	Smoking reduced	218
81440	745H000	Nicotine replacement therapy using nicotine patches	207
25106	8B3f.00	Nicotine replacement therapy provided free	197
98154	8HkQ.00	Referral to NHS stop smoking service	161
12965	137X.00	Cigarette consumption	134
104185	8IEM.00	Smoking cessation drug therapy declined	123
62686	137h.00	Minutes from waking to first tobacco consumption	117
16717	H310100	Smokers' cough	110
104230	8IEK.00	Smoking cessation programme declined	77
94958	745H400	Smoking cessation drug therapy	74
40417	9OO3.00	Stop smoking monitor default	72
60720	9OO5.00	Stop smoking monitor 2nd lettr	68
58597	9OO8.00	Stop smoking monitor phone inv	60
97029	137k.00	Refusal to give smoking status	58
32572	8B3Y.00	Over the counter nicotine replacement therapy	57
101338	137m.00	Failed attempt to stop smoking	44
68658	E251z00	Tobacco dependence NOS	43
85247	745H200	Nicotine replacement therapy using nicotine inhalator	43
41979	137e.00	Smoking restarted	41
90522	745Hz00	Smoking cessation therapy NOS	37
101764	13p5000	Practice based smoking cessation programme start date	34
103507	8CdB.00	Stop smoking service opportunity signposted	28
102361	9NS0200	Referral for smoking cessation service offered	26
101851	9Ndg.00	Declined consent for follow-up by smoking cessation team	22

medcode	readcode	readterm	n
26096	13cA.00	Smokes drugs	21
89464	745H300	Nicotine replacement therapy using nicotine lozenges	21
41042	8CAg.00	Smoking cessation advice provided by community pharmacist	20
53101	9OO7.00	Stop smoking monitor verb.inv.	20
103955	1782	Asthma trigger - tobacco smoke	17
66387	9OO6.00	Stop smoking monitor 3rd lettr	16
85975	745H100	Nicotine replacement therapy using nicotine gum	16
10184	67A3.00	Pregnancy smoking advice	13
24529	8I39.00	Nicotine replacement therapy refused	13
98245	8HBM.00	Stop smoking face to face follow-up	13
104310	9ko..11	Current smoker annual review	12
91708	745Hy00	Other specified smoking cessation therapy	11
12954	ZV4K000	[V]Tobacco use	8
35055	ZV6D800	[V]Tobacco abuse counselling	5
46654	137D.00	Admitted tobacco cons untrue ?	5
98493	9kc0.00	Smoking cessatn monitor template complet - enhanc serv admin	5
104086	9OOB000	Stop smoking invitation first SMS text message	5
70746	E251100	Tobacco dependence, continuous	4
96992	9kc..00	Smoking cessation - enhanced services administration	4
67178	8BP3.00	Nicotine replacement therapy provided by community pharmacis	2
101385	9Ndf.00	Consent given for follow-up by smoking cessation team	2
35475	J087300	Leukokeratosis nicotina palati	1
46321	137f.00	Reason for restarting smoking	1
56144	Eu17100	[X]Mental and behav dis due to use of tobacco: harmful use	1
63666	ZRBm200	Fagerstrom test for nicotine dependence	1
97643	38DH.00	Fagerstrom test for nicotine dependence	1
98347	9ko..00	Current smoker annual review - enhanced services admin	1
101634	9NdV.00	Consent given follow-up after smoking cessation intervention	1
47273	ZRaM.00	Motives for smoking scale	0
49418	ZRh4.11	RFS - Reasons for smoking scale	0
59866	ZRh4.00	Reasons for smoking scale	0
61905	Eu17.00	[X]Mental and behavioural disorder due to use of tobacco	0
63299	ZRBm211	FTND - Fagerstrom test for nicotine dependence	0
66409	8I2l.00	Nicotine replacement therapy contraindicated	0
91513	ZRao.00	Occasions for smoking scale	0
95610	E251000	Tobacco dependence, unspecified	0
98283	9kf2.00	COPD structured smoking assessment declined - enh serv admin	0
98284	9kf1.00	Refer COPD structured smoking assessment - enhanc serv admin	0
101210	9NdW.00	Consent given for smoking cessation data sharing	0
101325	9NdY.00	Declin cons follow-up evaluation after smoking cess interven	0
101519	Eu17300	[X]Mental and behav dis due to use tobacco: withdrawal state	0
102951	13p8.00	Lost to smoking cessation follow-up	0

medcode	readcode	readterm	n
103400	9kf1.11	Referred for COPD structured smoking assessment	0
103760	9kf2.11	COPD structured smoking assessment declined	0
105501	137o.00	Waterpipe tobacco consumption	0
105572	9OOB.00	Stop smoking invitation short message service text message	0
105999	1V08.00	Smokes drugs in cigarette form	0
106359	8T08.00	Referral to smoking cessation service	0
106384	9OOB100	Stop smoking invitation second SMS text message	0
106385	9OOB200	Stop smoking invitation third SMS text message	0
106391	8IEo.00	Referral to smoking cessation service declined	0
107504	ZRaM.11	MFS - Motives for smoking scale	0
107593	1V09.00	Smokes drugs through a pipe	0
107792	Eu17200	[X]Mental and behav dis due to use tobacco: dependence syndr	0
108835	E251200	Tobacco dependence, episodic	0
108966	9kc0.11	Smoking cessation ESA monitoring template completed	0

Appendix E.2 - Baseline Variable Distributions

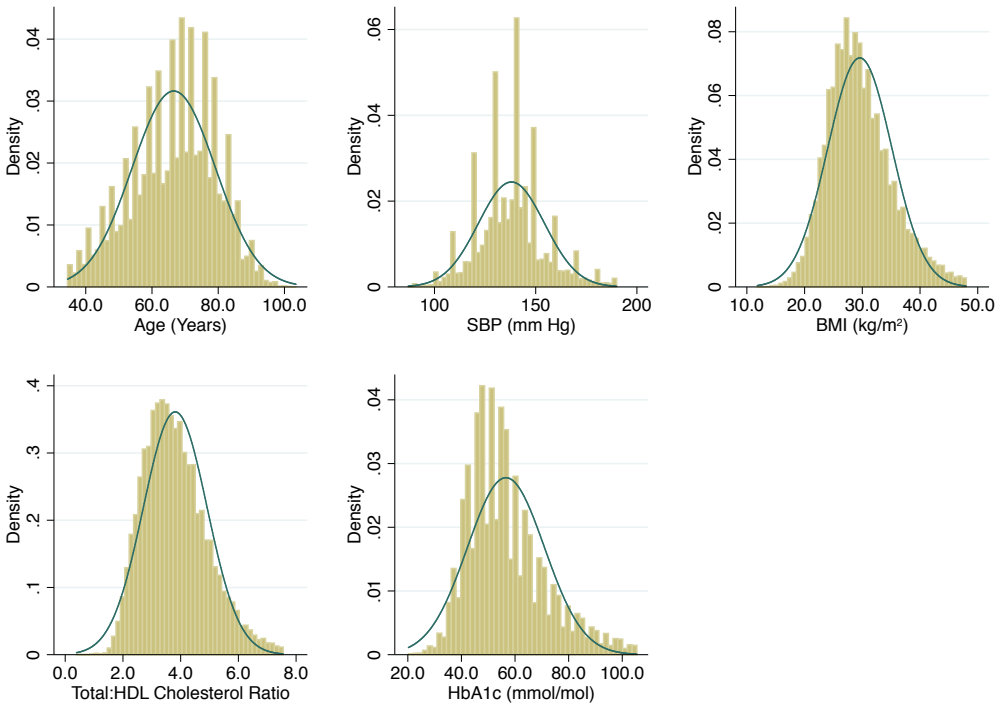


Figure E.2.a - Histograms of baseline values for continuous study variables. Each histogram depicts the mean + three standard deviations of the variable. Reference normal distributions are shown as black lines.

Appendix E.3 - Multivariable Survival Model Adjustment Variables

Cardiovascular Mortality

Table E.3.a - Adjusted estimates for the effect of albuminuria on cardiovascular mortality from Cox-PH, Lunn-McNeil and Fine-Gray Models.

Variable	Cox-PH Model	Lunn-McNeil Model	Fine-Gray Model
	HR (95% CI)	**HR (95% CI)**	**SHR (95% CI)**
Albuminuria (vs normoalbuminuria)	1.75 (1.63 - 1.87)	1.75 (1.64 - 1.87)	1.58 (1.48 - 1.69)
Male Gender (vs Female Gender)	1.36 (1.28 - 1.45)	1.36 (1.28 - 1.45)	1.30 (1.22 - 1.39)
Age (per Year)	1.11 (1.11 - 1.11)	1.11 (1.11 - 1.11)	1.09 (1.09 - 1.10)
BMI (per kg/m²)	1.02 (1.01 - 1.02)	1.02 (1.01 - 1.02)	1.02 (1.01 - 1.02)
Total : HDL Cholesterol	1.06 (1.03 - 1.09)	1.06 (1.03 - 1.09)	1.05 (1.02 - 1.08)
HbA1c (per mmol/mol)	1.01 (1.01 - 1.02)	1.01 (1.01 - 1.02)	1.01 (1.01 - 1.01)
SBP (per mmHg)	1.00 (1.00 - 1.00)	1.00 (1.00 - 1.00)	1.00 (1.00 - 1.00)
Ex-Smoker (vs Never Smoked)	1.18 (1.08 - 1.28)	1.18 (1.09 - 1.28)	1.16 (1.07 - 1.26)
Current Smoker (vs Never Smoked)	1.55 (1.41 - 1.71)	1.56 (1.42 - 1.72)	1.46 (1.32 - 1.61)

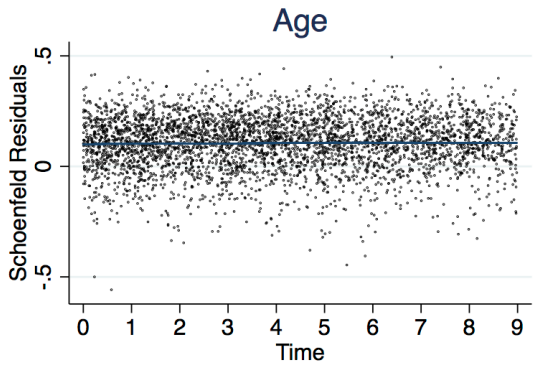
Cancer Mortality

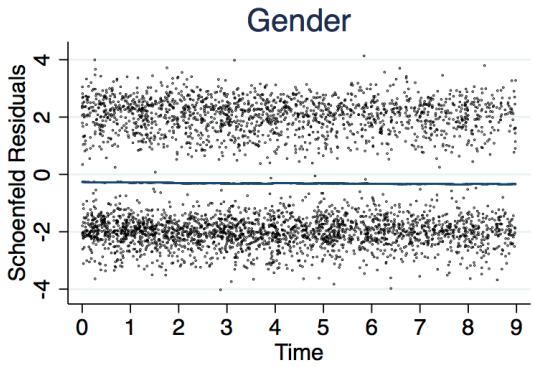
Table E.3.b - Adjusted estimates for the effect of albuminuria on cancer mortality from Cox-PH, Lunn-McNeil and Fine-Gray Models.

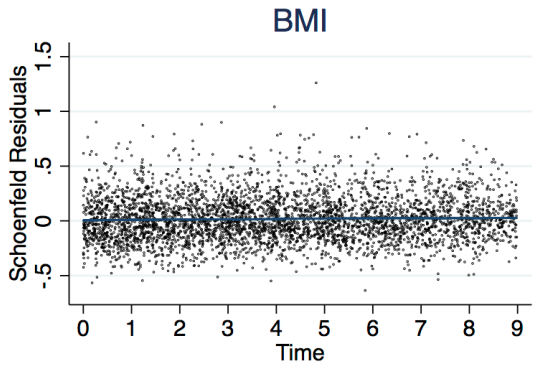
Variable	Cox-PH Model	Lunn-McNeil Model	Fine-Gray Model
	HR (95% CI)	**HR (95% CI)**	**SHR (95% CI)**
Albuminuria (vs normoalbuminuria)	1.27 (1.16 - 1.39)	1.28 (1.17 - 1.40)	1.11 (1.01 - 1.21)
Male Gender (vs Female Gender)	1.44 (1.32 - 1.56)	1.44 (1.33 - 1.56)	1.40 (1.29 - 1.52)
Age (per Year)	1.07 (1.07 - 1.08)	1.07 (1.07 - 1.08)	1.05 (1.05 - 1.06)
BMI (per kg/m²)	1.01 (1.00 - 1.02)	1.01 (1.00 - 1.02)	1.01 (1.00 - 1.02)
Total : HDL Cholesterol	1.02 (0.98 - 1.06)	1.02 (0.98 - 1.06)	1.01 (0.97 - 1.05)
HbA1c (per mmol/mol)	1.00 (1.00 - 1.00)	1.00 (1.00 - 1.00)	1.00 (1.00 - 1.00)
SBP (per mmHg)	1.00 (1.00 - 1.00)	1.00 (1.00 - 1.00)	1.00 (1.00 - 1.00)
Ex-Smoker (vs Never Smoked)	1.42 (1.27 - 1.59)	1.42 (1.27 - 1.59)	1.40 (1.25 - 1.56)
Current Smoker (vs Never Smoked)	2.21 (1.95 - 2.49)	2.22 (1.96 - 2.51)	2.08 (1.85 - 2.36)

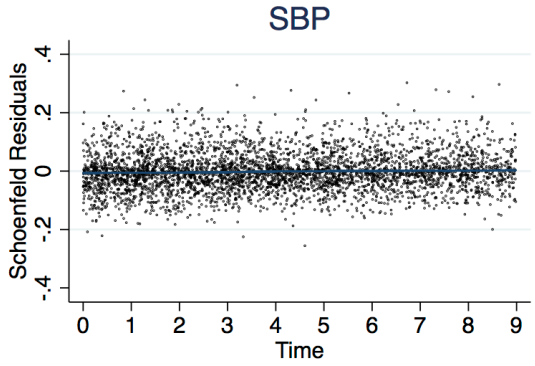
Appendix E.4 - Proportional Hazards Assumption Tests

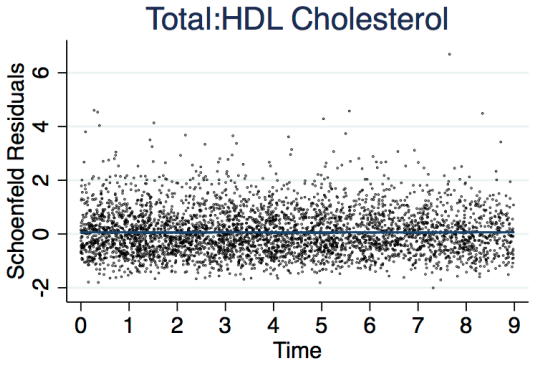
Cardiovascular Mortality

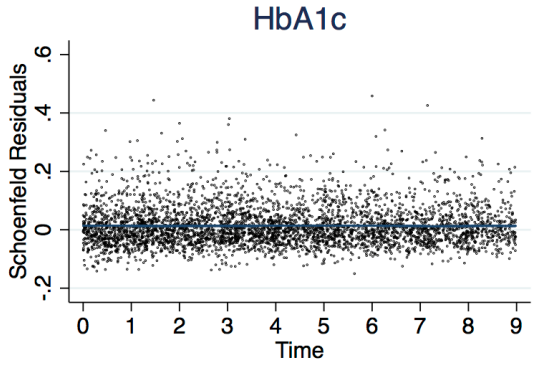


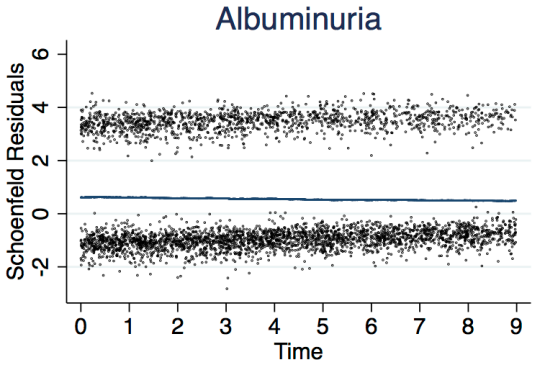


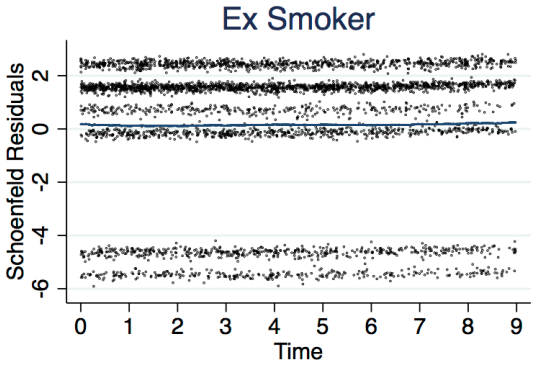


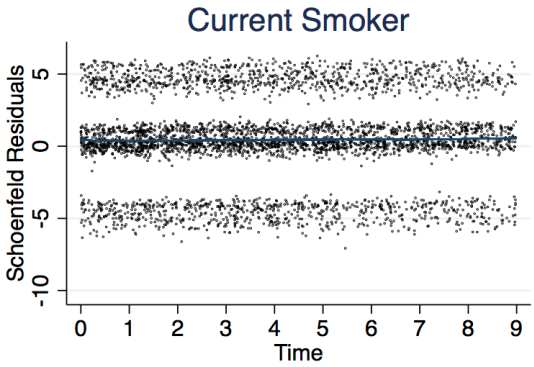




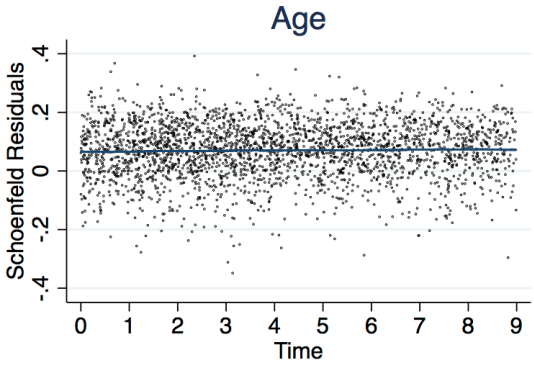


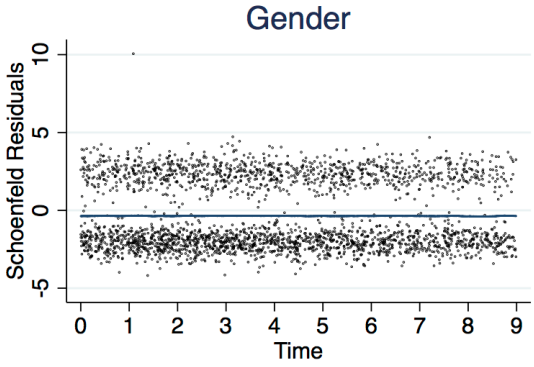


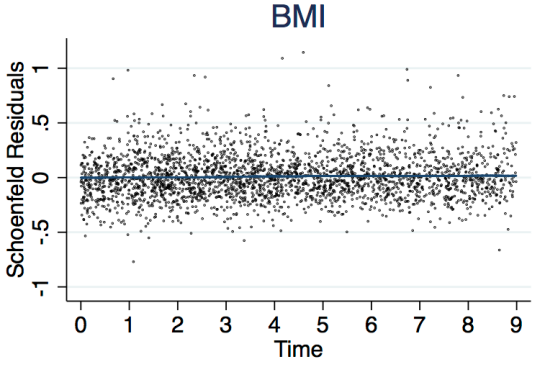


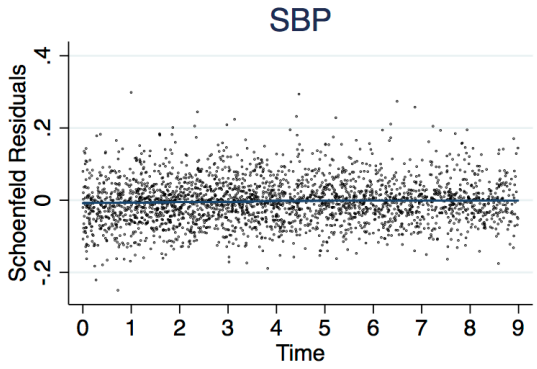


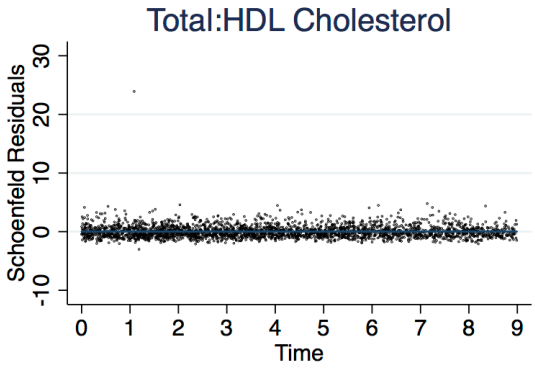
Cancer Mortality

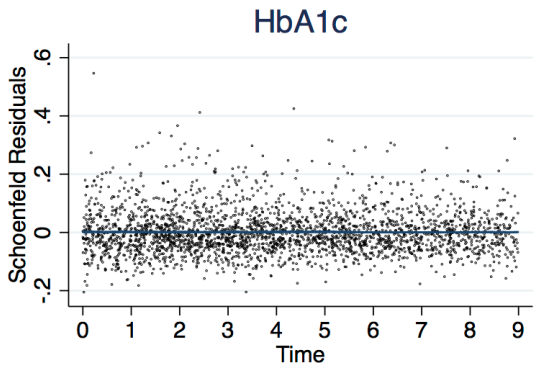


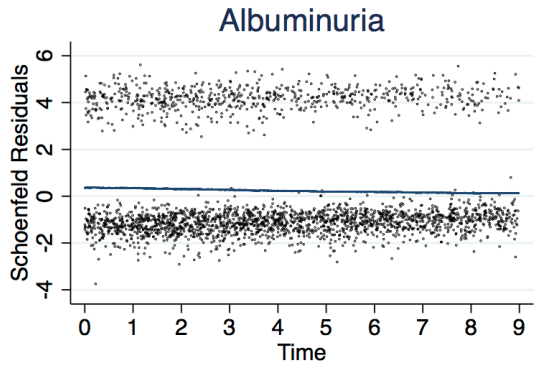


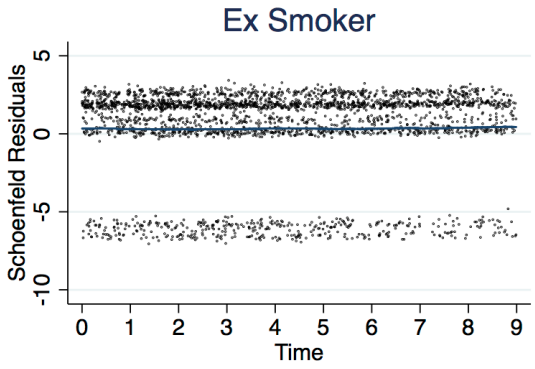


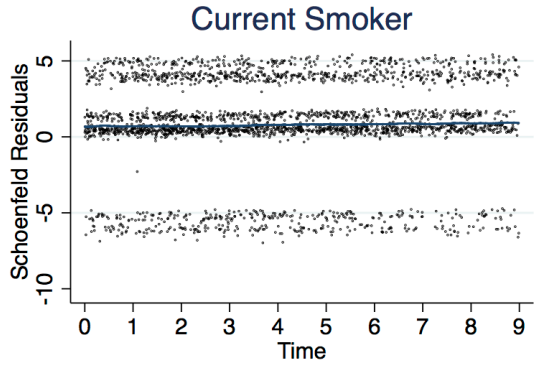

SBP
Schoenfeld Residuals
-.2
0
.2
.4
0 1 2 3 4 5 6 7 8 9
Time


Total:HDL Cholesterol
Schoenfeld Residuals
-10
0
10
20
30
0 1 2 3 4 5 6 7 8 9
Time

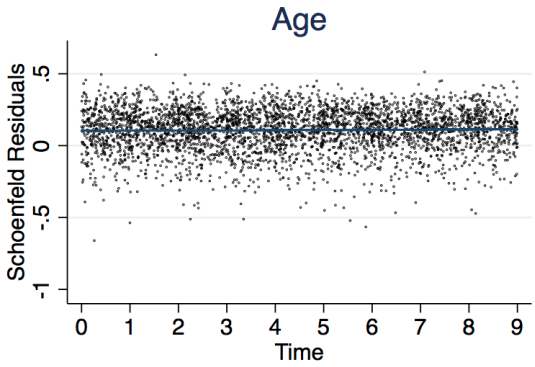

HbA1c
Schoenfeld Residuals
-.2
0
.2
.4
.6
0 1 2 3 4 5 6 7 8 9
Time

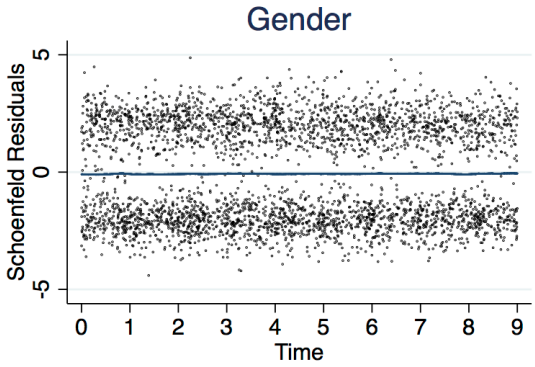

Albuminuria
Schoenfeld Residuals
-4
-2
0
2
4
6
0 1 2 3 4 5 6 7 8 9
Time

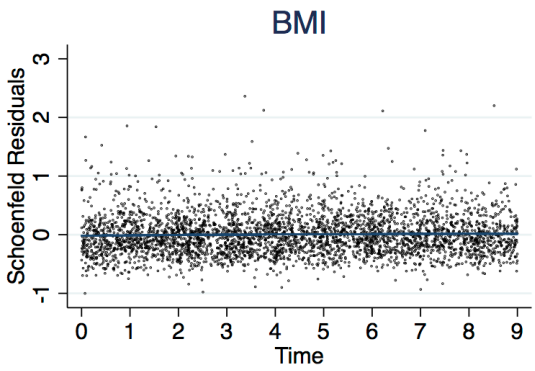

Ex Smoker
Schoenfeld Residuals
-10
-5
0
5
0 1 2 3 4 5 6 7 8 9
Time

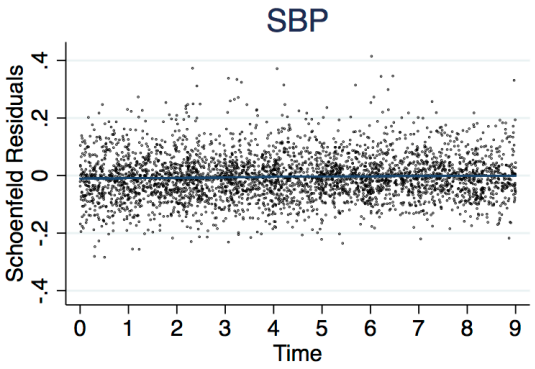

Current Smoker
Schoenfeld Residuals
-10
-5
0
5
0 1 2 3 4 5 6 7 8 9
Time

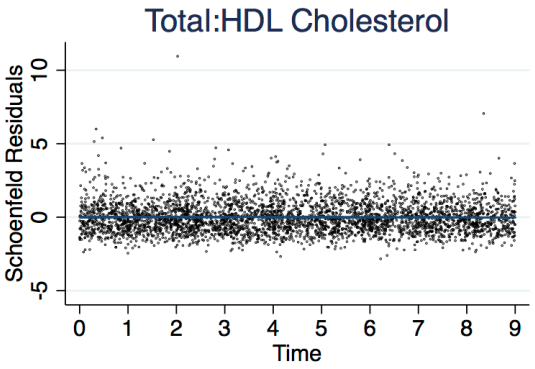
Non-Cardiovascular/Cancer Mortality

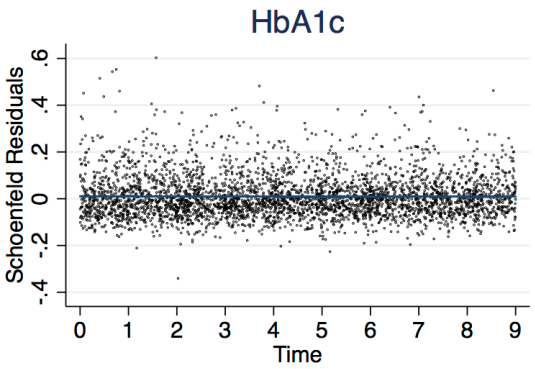


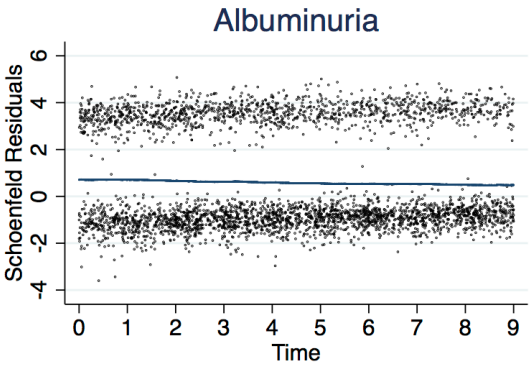


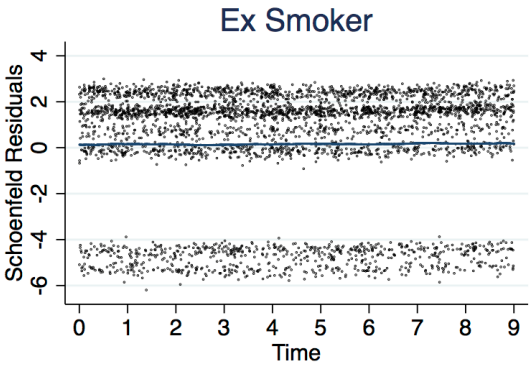


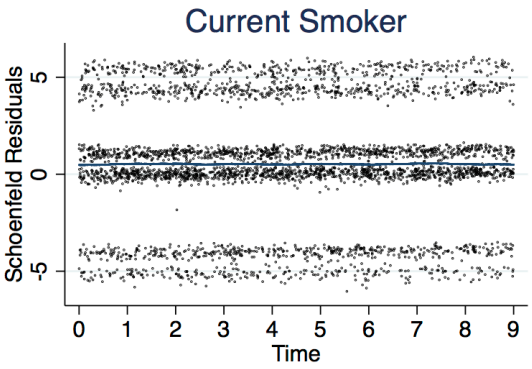






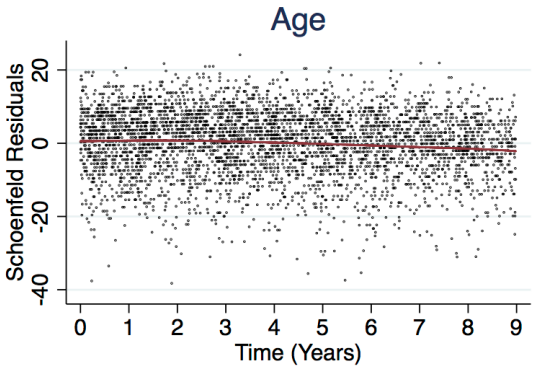

Albuminuria
Schoenfeld Residuals
-4
-2
0
2
4
6
0
1
2
3
4
5
6
7
8
9
Time

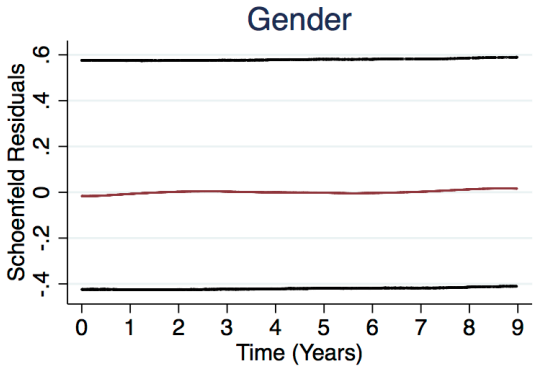

Ex Smoker
Schoenfeld Residuals
-6
-4
-2
0
2
4
0
1
2
3
4
5
6
7
8
9
Time

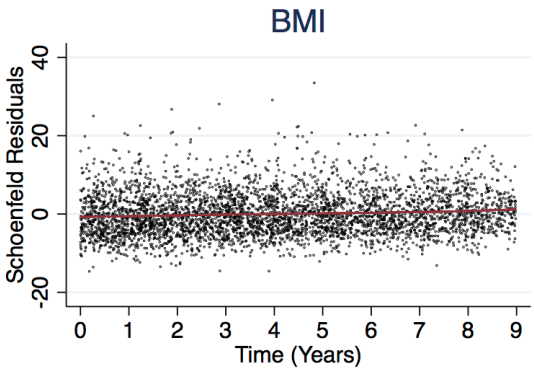

Current Smoker
Schoenfeld Residuals
-5
0
5
0
1
2
3
4
5
6
7
8
9
Time

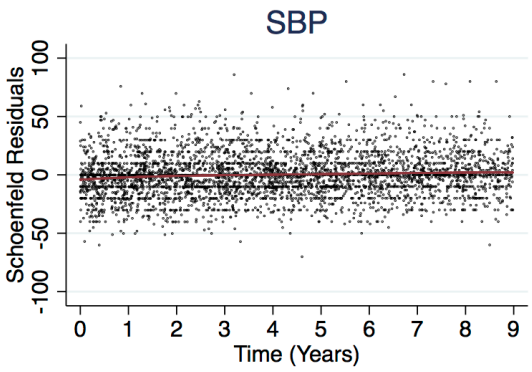
Appendix E.5 - Proportional Subdistribution Hazards Assumption Tests

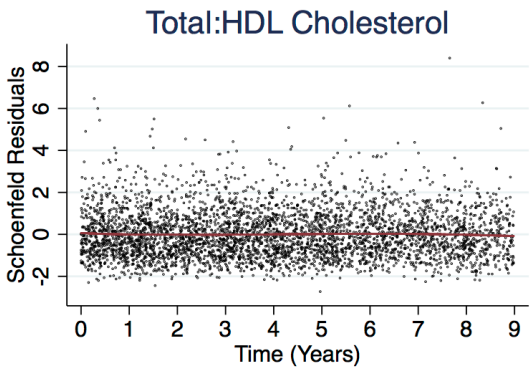
Cardiovascular Mortality

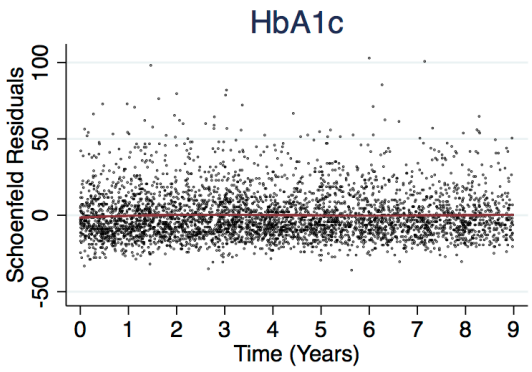


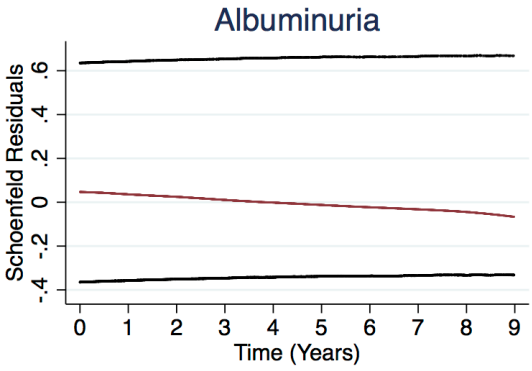


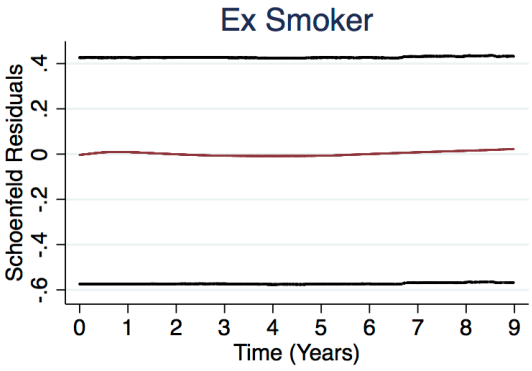


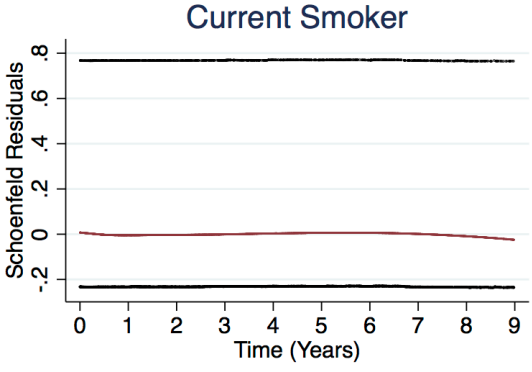

SBP
Schoenfeld Residuals
-100 -50 0 50 100
0 1 2 3 4 5 6 7 8 9
Time (Years)


Total:HDL Cholesterol
Schoenfeld Residuals
-2 0 2 4 6 8
0 1 2 3 4 5 6 7 8 9
Time (Years)

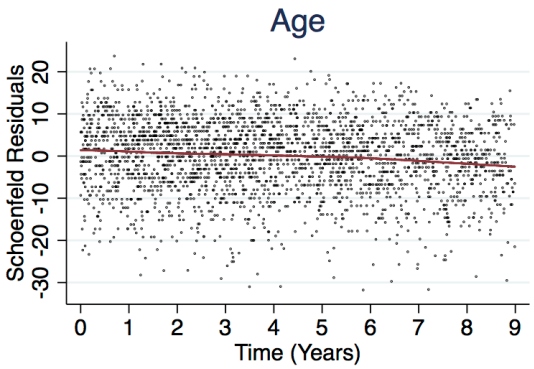

HbA1c
Schoenfeld Residuals
-50 0 50 100
0 1 2 3 4 5 6 7 8 9
Time (Years)

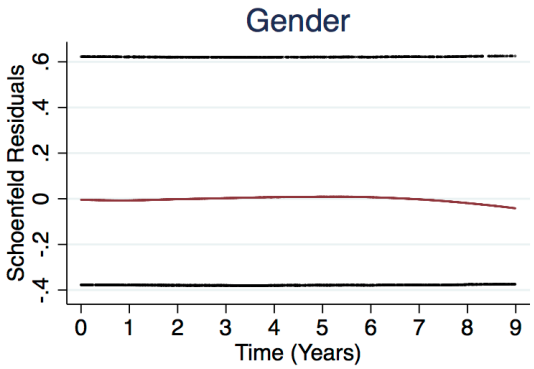

Albuminuria
Schoenfeld Residuals
-.4 -.2 0 .2 .4 .6
0 1 2 3 4 5 6 7 8 9
Time (Years)

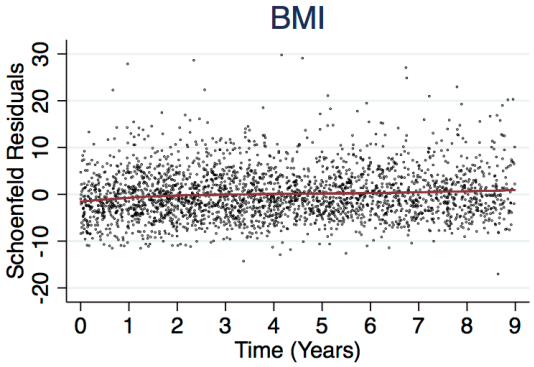

Ex Smoker
Schoenfeld Residuals
-.6 -.4 -.2 0 .2 .4
0 1 2 3 4 5 6 7 8 9
Time (Years)

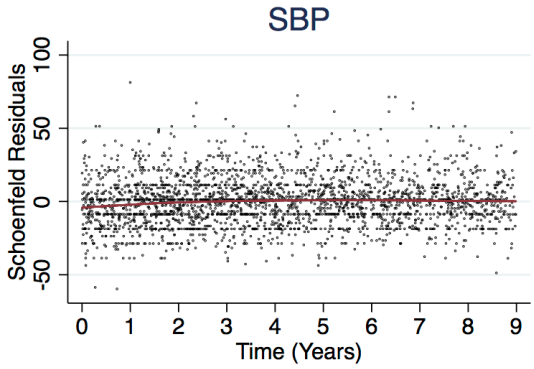

Current Smoker
Schoenfeld Residuals
-.2 0 .2 .4 .6 .8
0 1 2 3 4 5 6 7 8 9
Time (Years)

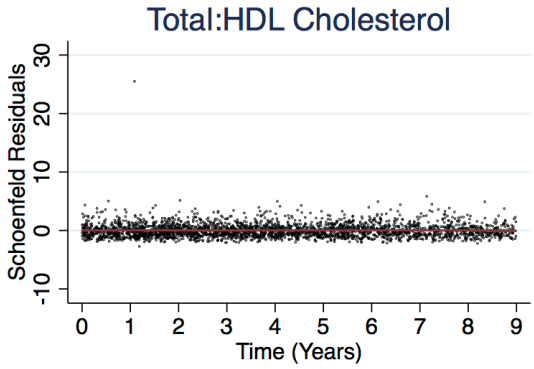
Cancer Mortality

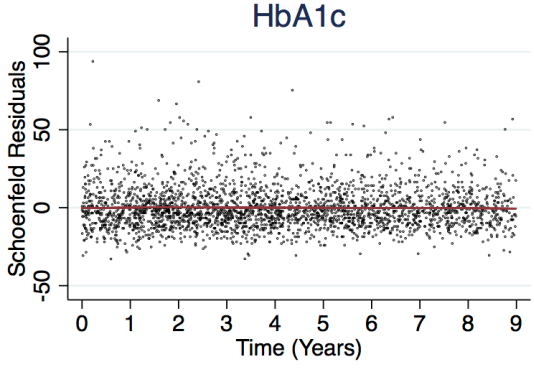


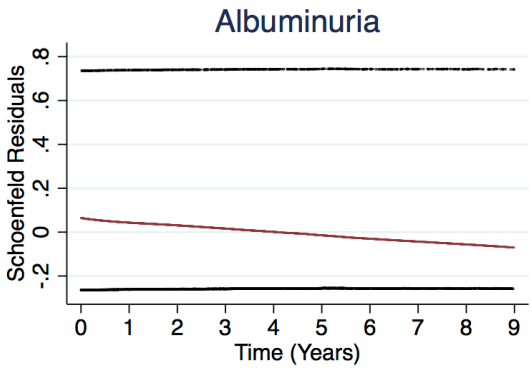


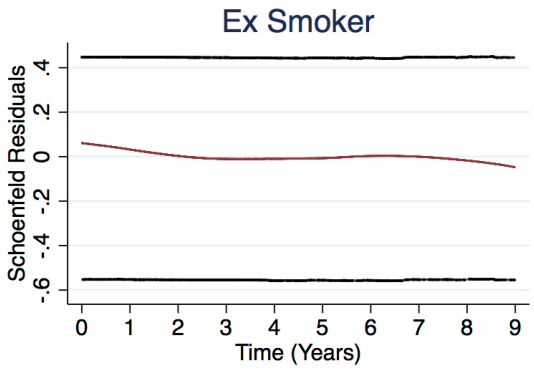


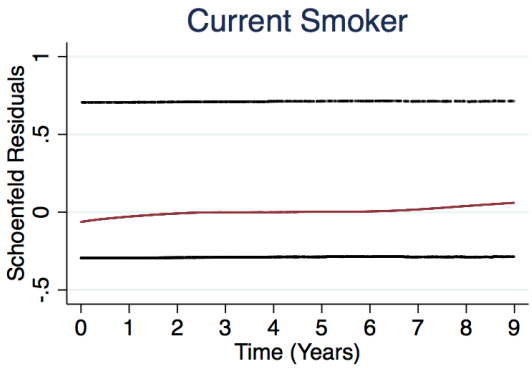







Albuminuria
Schoenfeld Residuals
-.2 0 .2 .4 .6 .8
0 1 2 3 4 5 6 7 8 9
Time (Years)


Ex Smoker
Schoenfeld Residuals
-.6 -.4 -.2 0 .2 .4
0 1 2 3 4 5 6 7 8 9
Time (Years)


Current Smoker
Schoenfeld Residuals
-.5 0 .5 1
0 1 2 3 4 5 6 7 8 9
Time (Years)

Bibliography

1. Gooley, T. A., Leisenring, W., Crowley, J. & Storer, B. E. Estimation of Failure Probabilities in the Presence of Competing Risks: New Representations of Old Estimators. *Stat Med* **18**, 695–706 (1999).
2. Messerli, F. H., Bangalore, S., Torp-Pedersen, C., Staessen, J. A. & Kostis, J. B. Cardiovascular Drugs and Cancer: of Competing Risk, Smallpox, Bernoulli, and d'Alembert. *Eur Hear. J* **34**, 1095–1098 (2013).
3. Kaplan, E. L. & Meier, P. Nonparametric Estimation from Incomplete Observations. *JASA* **53**, 457–481 (1958).
4. Cox, D. R. Regression Models and Life-Tables. *J R Stat Soc* **34**, 187–220 (1972).
5. Verduijn, M., Grootendorst, D. C., Dekker, F. W., Jager, K. J. & le Cessie, S. The Analysis of Competing Events Like Cause-Specific Mortality - Beware of the Kaplan-Meier Method. *Nephrol Dial Transpl.* **26**, 56–61 (2011).
6. Haller, B., Schmidt, G. & Ulm, K. Applying Competing Risks Regression Models: An Overview. *Lifetime Data Anal* **19**, 33–58 (2013).
7. Lau, B., Cole, S. R. & Gange, S. J. Competing Risk Regression Models for Epidemiologic Data. *Am J Epidemiol* **170**, 244–256 (2009).
8. Noordzij, M. *et al*. When do we Need Competing Risks Methods for Survival Analysis in Nephrology? *Nephrol Dial Transpl*. (2013). doi:10.1093/ndt/gft355
9. Aalen, O. O. Nonparametric Inference for a Family of Counting Processes. *Ann Stat.* **6**, 701–726 (1978).
10. Southern, D. A. *et al*. Kaplan-Meier Methods Yielded Misleading Results in Competing Risk Scenarios. *J Clin Epidemiol* **59**, 1110–1114 (2006).

11. Lambert, P. C., Dickman, P. W., Nelson, C. P. & Royston, P. Estimating the Crude Probability of Death due to Cancer and Other Causes Using Relative Survival Models. *Stat Med* **29**, 885–895 (2010).

12. David, A. P. Causal Inference without Counterfactuals. *JASA* **95**, 407–424 (2000).

13. Koller, M. T., Raatz, H., Steyerberg, E. W. & Wolbers, M. Competing Risks and The Clinical Community: Irrelevance or Ignorance? *Stat Med* **31**, 1089–1097 (2012).

14. Wolkewitz, M., Cooper, B. S., Bonten, M. J. M., Barnett, A. G. & Schumacher, M. Interpreting and Comparing Risks in the Presence of Competing Events. *BMJ (Clin Res Ed)* **349**, g5060 (2014).

15. Kalbfleisch, J. D. & Prentice, R. L. Estimation of the Average Hazard Ratio. *Biometrika* **68**, 105–112 (1981).

16. Lunn, M. & McNeil, D. Applying Cox Regression to Competing Risks. *Biometrics* **51**, 524–532 (1995).

17. Lim, H. J., Zhang, X., Dyck, R. & Osgood, N. Methods of Competing Risks Analysis of End-Stage Renal Disease and Mortality Among People with Diabetes. *BMC Med Res Methodol* **10**, 97 (2010).

18. Persson, I. & Khamis, H. Bias of the Cox Model Hazard Ratio. *J Mod Appl Stat Meth* **4**, 90–99 (2005).

19. Allignol, A., Schumacher, M., Wanner, C., Drechsler, C. & Beyersmann, J. Understanding Competing Risks: a Simulation Point of View. *BMC Med Res Methodol* **11**, 86 (2011).

20. Fine, J. P. & Gray, R. J. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *JASA* **94**, 496–509 (1999).

21. Prentice, R. L. *et al*. The Analysis of Failure Times in the Presence of Competing Risks. *Biometrics* **34**, 541–554 (1978).

22. Grambauer, N., Schumacher, M. & Beyersmann, J. Proportional Subdistribution Hazards Modeling Offers a Summary Analysis, Even if Misspecified. *Stat Med* **29**, 875–884 (2010).

23. Diabetes UK. *Facts and Stats*. (2015). at <https://www.diabetes.org.uk/About_us/What-we-say/Statistics/>

24. Health & Social Care Information Centre. *National Diabetes Audit 2013-2014 and 2014-*

2015. Report 1: Care Processes and Treatment Targets. (2016). at <http://www.hscic.gov.uk/catalogue/PUB19900/nati-diab-rep1-audi-2013-15.pdf>

25. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* **26,** S5–S20 (2003).

26. Thunander, M. *et al*. Incidence of Type 1 and Type 2 Diabetes in Adults and Children in Kronoberg, Sweden. *Diab Res Clin Pr*. **82,** 247–255 (2008).

27. Stratton, I. M. *et al*. Association of Glycaemia with Macrovascular and Microvascular Complications of Type 2 diabetes (UKPDS 35): Prospective Observational Study. *BMJ* **321,** 405–412 (2000).

28. KDIGO. KDIGO 2012 Clinical Practice Guideline for the Evaluation and Management of Chronic Kidney Disease. *Kidney Int Suppl* **3,** 136–150 (2013).

29. Adler, A. I. *et al*. Development and Progression of Nephropathy in Type 2 Diabetes: The United Kingdom Prospective Diabetes Study (UKPDS 64). *Kidney Int* **63,** 225–232 (2003).

30. Giovannucci, E. *et al*. Diabetes and Cancer: A Consensus Report. *Diabetes Care* **33,** 1674–1685 (2010).

31. Elton B Stephens Company. Cumulative Index to Nursing and Allied Health Literature (CINAHL). (1977). at <https://www.ebscohost.com/nursing/products/cinahl-databases/the-cinahl-database>

32. United States National Library of Medicine. Medical Literature Analysis and Retrieval System Online (MEDLINE). (1964). at <https://www.nlm.nih.gov/pubs/factsheets/medline.html>

33. Elsevier. Excerpta Medica dataBASE (EMBASE). (1946). at <https://www.elsevier.com/solutions/embase-biomedical-research>

34. National Center for Biotechnology Information. PubMed. (1996). at <https://www.ncbi.nlm.nih.gov/pubmed/>

35. University of York Centre Centre for Reviews & Dissemination. PROSPERO: International Prospective Register of Systematic Reviews. (2013). at <http://www.crd.york.ac.uk/prospero/>

36. Sasaki, A., Uehara, M., Horiuchi, N. & Hasagawa, K. A Long-Term Follow-Up Study of

Japanese Diabetic Patients: Mortality and Causes of Death. *Diabetologia* **25**, 309–312 (1983).

37. Fuller, J., Stevens, L., Wang, S.-L. & The WHO Multinational Study Group. Risk Factors for Cardiovascular Mortality and Morbidity: The WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia* **44**, S54–S64 (2001).

38. Sievers, M. L., Nelson, R. G. & Bennett, P. H. Sequential Trends in Overall and Cause-Specific Mortality in Diabetic and Nondiabetic Pima Indians. *Diabetes Care* **19**, 107–111 (1996).

39. Wang, S.-L., Head, J., Stevens, L., Fuller, J. H. & The WHO Multinational Study Group. Excess Mortality and its Relation to Hypertension and Proteinuria in Diabetic Patients. The World Health Organization Multinational Study of Vascular Disease in Diabetes. *Diabetes Care* **19**, 305–312 (1996).

40. Davis, T. M. E. *et al*. Risk Factors for Stroke in Type 2 Diabetes Mellitus: United Kingdom Prospective Diabetes Study (UKPDS) 29. *Arch Intern Med* **159**, 1097–1103 (1999).

41. Colhoun, H. M. *et al*. Risk Factors for Renal Failure: The WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia* **44**, S46–S53 (2001).

42. Keen, H. *et al*. The Appearance of Retinopathy and Progression to Proliferative Retinopathy: the WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia* **44**, S22–S30 (2001).

43. Miki, E. *et al*. The Incidence of Visual Impairment and its Determinants in the WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia* **44**, S31–S36 (2001).

44. Morrish, N. *et al*. Mortality and Causes of Death in the WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia* **44**, S14–S21 (2001).

45. Pavkov, M. E., Sievers, M. L., Knowler, W. C., Bennett, P. H. & Nelson, R. G. An Explanation for the Increase in Heart Disease Mortality Rates in Diabetic Pima Indians: Effect of Renal Replacement Therapy. *Diabetes Care* **27**, 1132–1136 (2004).

46. Bruno, G. *et al*. Fibrinogen and AER are Major Independent Predictors of 11-Year Cardiovascular Mortality in Type 2 Diabetes: The Casale Monferrato Study. *Diabetologia* **48**, 427–434 (2005).

47. The FIELD Study Investigators. Effects of Long-Term Fenofibrate Therapy on Cardiovascular

Events in 9795 People with Type 2 Diabetes Mellitus (the FIELD Study): Randomised Controlled Trial. *Lancet* **366**, 1849–1861 (2005).

48. Hippisley-Cox, J. *et al*. Predicting Cardiovascular Risk in England and Wales: Prospective Derivation and Validation of QRISK2. *BMJ* **336**, 1475–1482 (2008).

49. The ADVANCE Collaborative Group. Intensive Blood Glucose Control and Vascular Outcomes in Patients with Type 2 Diabetes. *N Engl J Med* **358**, 2560–2572 (2008).

50. Yang, X. *et al*. Additive Interaction of Hyperglycemia and Albuminuria on Risk of Ischemic Stroke in Type 2 Diabetes: Hong Kong Diabetes Registry. *Diabetes Care* **31**, 2294–2300 (2008).

51. Yang, X. *et al*. Thresholds of Risk Factors for Ischemic Stroke in Type 2 Diabetic Patients with and without Albuminuria: A Non-Linear Approach. *Clin Neurol Neurosurg* **110**, 701–709 (2008).

52. Eeg-Olofsson, K. *et al*. Risk of Cardiovascular Disease and Mortality in Overweight and Obese Patients with Type 2 Diabetes: An Observational Study in 13,087 Patients. *Diabetologia* **52**, 65–73 (2009).

53. Ma, R. C. W. *et al*. Effects of Systolic and Diastolic Blood Pressures on Incident Coronary Heart Disease and All-Cause Death in Chinese Women with Type 2 Diabetes: The Hong Kong Diabetes Registry. *J Diabetes* **1**, 90–98 (2009).

54. Nilsson, P. M. *et al*. Smoking as an Independent Risk Factor for Myocardial Infarction or Stroke in Type 2 Diabetes: A Report from the Swedish National Diabetes Register. *Eur J Cardiovasc Prev Rehabil* **16**, 506–512 (2009).

55. Burgess, D. C. *et al*. Incidence and Predictors of Silent Myocardial Infarction in Type 2 Diabetes and the Effect of Fenofibrate: An Analysis from the Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) Study. *Eur Hear. J* **31**, 92–99 (2010).

56. Lin, E. H. B. *et al*. Depression and Advanced Complications of Diabetes: A Prospective Cohort Study. *Diabetes Care* **33**, 264–269 (2010).

57. Drury, P. L. *et al*. Estimated Glomerular Filtration Rate and Albuminuria are Independent Predictors of Cardiovascular Events and Death in Type 2 Diabetes Mellitus: The Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) study. *Diabetologia* **54**, 32–43 (2011).

58. Vupputuri, S., Nichols, G. A., Lau, H., Joski, P. & Thorp, M. L. Risk of Progression of Nephropathy in a Population-Based Sample with Type 2 Diabetes. *Diab Res Clin Pr*. **91**, 246–252 (2011).

59. Girman, C. J. *et al*. Risk of Acute Renal Failure in Patients with Type 2 Diabetes Mellitus. *Diabet Med* **29**, 614–621 (2012).

60. Morton, J. *et al*. Low HDL Cholesterol and the Risk of Diabetic Nephropathy and Retinopathy: Results of the ADVANCE Study. *Diabetes Care* **35**, 2201–2206 (2012).

61. Yang, X. *et al*. Synergistic Effects of Low LDL Cholesterol with Other Factors for the Risk of Cancer in Type 2 Diabetes: The Hong Kong Diabetes Registry. *Acta Diabetol* **49 (Suppl**, S185–S193 (2012).

62. Hippisley-Cox, J., Coupland, C. & Brindle, P. Derivation and Validation of QStroke Score for Predicting Risk of Ischaemic Stroke in Primary Care and Comparison with Other Risk Scores: A Prospective Open Cohort Study. *BMJ (Clin Res Ed)* **346**, f2573 (2013).

63. Ting, R. Z. W. *et al*. High Risk for Cardiovascular Disease in Chinese Type 2 Diabetic Patients with Major Depression - A 7-Year Prospective Analysis of the Hong Kong Diabetes Registry. *J Affect Disord*. **149**, 129–135 (2013).

64. Brenner, B. M. *et al*. The RENAAL Study Investigation. *Clin Diabetes* **20**, 19–20 (2002).

65. Fox, C. S. *et al*. Associations of Kidney Disease Measures with Mortality and End-Stage Renal Disease in Individuals with and without Diabetes: A Meta-Analysis. *Lancet* **380**, 1662–1673 (2012).

66. Savarese, G. *et al*. Reduction of Albumin Urinary Excretion is Associated with Reduced Cardiovascular Events in Hypertensive and/or Diabetic Patients. A Meta-Regression Analysis of 32 Randomized Trials. *Int J Cardiol* **172**, 403–410 (2014).

67. Toyama, T. *et al*. The Impacts of Albuminuria and Low eGFR on the Risk of Cardiovascular Death, All-Cause Mortality, and Renal Events in Diabetic Patients: Meta-Analysis. *PLoS One* **8**, e71810 (2013).

68. Dineen, S. F. & Gerstein, H. C. The Association of Microalbuminuria and Mortality in Non-Insulin-Dependent Diabetes Mellitus. A Systematic Overview of the Literature. *Arch Intern Med* **157**, 1413–1418 (1997).

69. Hartmann, A., Jenssen, T. & Holdaas, H. Diabetes, Chronic Kidney Disease and Cancer Risk. *Nephrol Dial Transpl.* **27**, 3018–3020 (2012).

70. Wong, G. *et al.* The Risk of Cancer in People with Diabetes and Chronic Kidney Disease. *Nephrol Dial Transpl.* **27**, 3337–3344 (2012).

71. Jørgensen, L., Heuch, I., Jenssen, T. & Jacobsen, B. K. Association of Albuminuria and Cancer incidence. *J Am Soc Nephrol* **19**, 992–998 (2008).

72. Yu, T.-T. *et al.* Proteinuria Predicts 10-Year Cancer-Related Mortality in Patients with Type 2 Diabetes. *J Diabetes Complicat.* **27**, 201–207 (2013).

73. Parving, H.-H., Chaturvedi, N., Viberti, G. & Mogensen, C. E. Does Microalbuminuria Predict Diabetic Nephropathy? *Diabetes Care* **24**, 1560–1566 (2001).

74. The Cochrane Collaboration. The Cochrane Collaboration Systematic Review Search Filters. (2014). at <http://handbook.cochrane.org/chapter_6/6_4_11_search_filters.htm>

75. The British Medical Journal Evidence Centre. The British Medical Journal (BMJ) Study Design Search Filters. (2014). at <http://clinicalevidence.bmj.com/x/set/static/ebm/learn/665076.html>

76. Health Improvement Scotland. The Scottish Intercollegiate Guidelines Network (SIGN) Systematic Review Filters. (2014). at <http://www.sign.ac.uk/methodology/filters.html>

77. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. (2009). at <http://www.prisma-statement.org/statement.htm>

78. Hayden, J. A., van der Windt, D. A., Cartwright, J. L., Côté, P. & Bombardier, C. Assessing Bias in Studies of Prognostic Factors. *Ann Intern Med* **158**, 280–286 (2013).

79. Hayden, J. A., Côté, P., Steenstra, I. A. & Bombardier, C. Identifying Phases of Investigation Helps Planning, Appraising, and Applying the Results of Explanatory Prognosis Studies. *J Clin Epidemiol* **61**, 552–560 (2008).

80. Hayden, J. A., Côté, P. & Bombardier, C. Evaluation of the Quality of Prognosis Studies in Systematic Reviews. *Ann Intern Med* **144**, 427–438 (2006).

81. Ubink-Veltmaat, L. J. *et al.* Prevalence, Incidence and Mortality of Type 2 Diabetes Mellitus Revisited: A Prospective Population-Based Study in The Netherlands (ZODIAC-1). *Eur J*

Epidemiol **18**, 793–800 (2003).

82. Riley, R. D., Higgins, J. P. T. & Deeks, J. J. Interpretation of Random Effects Meta-Analyses. *BMJ* **342**, d549 (2011).

83. DerSimonian, R. & Laird, N. M. Meta Analysis in Clinical Trials. *Con Clin Trials* **7**, 177–188 (1986).

84. Chen, H., Manning, A. K. & Dupuis, J. A Method of Moments Estimator for Random Effect Multivariate Meta-Analysis. *Biometrics* **68**, 1278–1284 (2012).

85. Kyzas, P. A., Loizou, K. T. & Ioannidis, J. P. A. Selective Reporting Biases in Cancer Prognostic Factor Studies. *J Natl Cancer Inst* **97**, 1043–1055 (2005).

86. Egger, M., Davey Smith, G., Schneider, M. & Minder, C. Bias in Meta-Analysis Detected by a Simple, Graphical Test. *BMJ (Clin Res Ed)* **315**, 629–634 (1997).

87. Light, R. J. & Pillemer, D. B. *Summing Up: The Science of Reviewing Research*. (Harvard University Press, 1984).

88. Harris, R. J. *et al*. Metan: Fixed- and Random-Effects Meta-Analysis. *Stata J* **8**, 3–28 (2008).

89. Harbord, R. M., Harris, R. J. & Sterne, J. A. C. Updated Tests for Small-Study Effects in Meta-Analyses. *Stata J* **9**, 197–201 (2009).

90. Sterne, J. A. C. & Harbord, R. M. Funnel Plots in Meta-Analysis. *Stata J* **4**, 127–141 (2004).

91. StataCorp. Stata Statistical Software: Release 12. (2011).

92. Bo, S. *et al*. Mortality within the First 10 Years of the Disease in Type 2 Diabetic Patients. *Nutr Metab Cardiovas Dis* **16**, 8–12 (2006).

93. Bruno, G. *et al*. N-Terminal Probrain Natriuretic Peptide is a Stronger Predictor of Cardiovascular Mortality than C-Reactive Protein and Albumin Excretion Rate in Elderly Patients with Type 2 Diabetes: The Casale Monferrato Population-Based Study. *Diabetes Care* **36**, 2677–2682 (2013).

94. Drion, I. *et al*. Chronic Kidney Disease and Mortality Risk Among Older Patients with Type 2 Diabetes Mellitus (ZODIAC-24). *Age Ageing* **41**, 345–350 (2012).

95. Isomaa, B. *et al*. Cardiovascular Morbidity and Mortality Associated with the Metabolic Syndrome. *Diabetes Care* **24**, 683–689 (2001).

96. Jager, A. *et al*. Microalbuminuria and Peripheral Arterial Disease Are Independent Predictors

of Cardiovascular and All-Cause Mortality, Especially Among Hypertensive Subjects: Five-Year Follow-Up of the Hoon Study. *Arter. Thromb Vasc Biol* **19**, 617–624 (1999).

97. Jimenez-Corona, A. *et al*. Electrocardiographic Abnormalities Predict Deaths from Cardiovascular Disease and Ischemic Heart Disease in Pima Indians with Type 2 Diabetes. *Am Hear. J* **151**, 1087–1093 (2006).

98. Juutilainen, A., Lehto, S., Rönnemaa, T., Pyörälä, K. & Laakso, M. Proteinuria and Metabolic Syndrome as Predictors of Cardiovascular Death in Non-Diabetic and Type 2 Diabetic Men and Momen. *Diabetologia* **49**, 56–65 (2006).

99. Ko, G. T.-C. *et al*. Prediction of Cardiovascular and Total Mortality in Chinese Type 2 Diabetic Patients by the WHO Definition for the Metabolic Syndrome. *Diabetes Obes Metab* **8**, 94–104 (2006).

100. Lin, C.-C. *et al*. Joint Relationship Between Renal Function and Proteinuria on Mortality of Patients with Type 2 Diabetes: The Taichung Diabetes Study. *Cardiovas Diabetol* **11**, 1–11 (2012).

101. Nelson, R. G., Pettitt, D. J., Carraher, M. J., Baird, H. R. & Knowler, W. C. Effect of Proteinuria on Mortality in NIDDM. *Diabetes* **37**, 1499–1504 (1988).

102. Pavkov, M. E. *et al*. Predominant Effect of Kidney Disease on Mortality in Pima Indians with or without Type 2 Diabetes. *Kidney Int* **68**, 1267–1274 (2005).

103. Sievers, M. L., Bennett, P. H., Roumain, J. & Nelson, R. G. Effect of Hypertension on Mortality in Pima Indians. *Circulation* **100**, 33–40 (1999).

104. Svensson, M., Cederholm, J., Eliasson, B., Zethelius, B. & Gudbjörnsdottir, S. Albuminuria and Renal Function as Predictors of Cardiovascular Events and Mortality in a General Population of Patients with Type 2 Diabetes: A Nationwide Observational Study From the Swedish National Diabetes Register. *Diab Vasc Dis Res* **10**, 520–529 (2013).

105. Yu, T. Y. *et al*. Proteinuria Predicts 10-year Cancer-Related Mortality in Patients with Type 2 Diabetes. *J Diabetes Complicat*. **27**, 201–207 (2013).

106. Beilin, J., Stanton, K. G., McCann, V. J., Knuiman, M. W. & Divitini, M. L. Microalbuminuria in Type 2 diabetes: An Independent Predictor of Cardiovascular Mortality. *Aust N Z J Med* **26**, 519–525 (1996).

107. Bruce, D. G., Davis, W. A., Starkstein, S. E. & Davis, T. M. E. A Prospective Study of Depression and Mortality in Patients with Type 2 Diabetes: The Fremantle Diabetes Study. *Diabetologia* **48**, 2532–2539 (2005).

108. Bruno, G. *et al*. Fibrinogen and AER are Major Independent Predictors of 11-year Cardiovascular Mortality in Type 2 Diabetes: The Casale Monferrato Study. *Diabetologia* **48**, 427–434 (2005).

109. Casiglia, E. *et al*. Cardiovascular Mortality in Non-Insulin-Dependent Diabetes Mellitus. A Controlled Study Among 683 Diabetics and 683 Age- and Sex-Matched Normal Subjects. *Eur J Epidemiol* **16**, 677–684 (2000).

110. de Fine Olivarius, N. *et al*. Predictors of Mortality of Patients Newly Diagnosed with Clinical Type 2 Diabetes: A 5-year Follow Up Study. *BMC Endocr Dis* **10**, 14 (2010).

111. Drury, P. L. *et al*. Estimated Glomerular Filtration Rate and Albuminuria are Independent Predictors of Cardiovascular Events and Death in Type 2 Diabetes Mellitus: The Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) Study. *Diabetologia* **54**, 32–43 (2011).

112. Ekinci, E. I. *et al*. Dietary Salt Intake and Mortality in Patients with Type 2 Diabetes. *Diabetes Care* **34**, 703–709 (2011).

113. Fuller, J. H., Stevens, L. K., Wang, S.-L. & The WHO Multinational Study Group. Risk Factors for Cardiovascular Mortality and Morbidity: The WHO Multinational Study of Vascular Disease in Diabetes. *Diabetologia* **44**, S54–S64 (2001).

114. McEwen, L. N. *et al*. Predictors of Mortality Over 8 Years in Type 2 Diabetic Patients: Translating Research Into Action for Diabetes (TRIAD). *Diabetes Care* **35**, 1301–1309 (2012).

115. Ninomiya, T. *et al*. Albuminuria and Kidney Function Independently Predict Cardiovascular and Renal Outcomes in Diabetes. *J Am Soc Nephrol* **20**, 1813–1821 (2009).

116. Nobakhthaghighi, N. *et al*. Relationship Between Urinary Albumin Excretion and Left Ventricular Mass with Mortality in Patients with Type 2 Diabetes. *Clin J Am Soc Nephrol* **1**, 1187–1190 (2006).

117. Targher, G. *et al*. Glomerular Filtration Rate, Albuminuria and Risk of Cardiovascular and All-Cause Mortality in Type 2 Diabetic Individuals. *Nutr Metab Cardiovas Dis* **21**, 294–301 (2011).

118. Valmadrid, C. T., Klein, R., Moss, S. E. & Klein, B. E. The Risk of Cardiovascular Disease Mortality Associated with Microalbuminuria and Gross Proteinuria in Persons with Older-Onset Diabetes Mellitus. *Arch Intern Med* **160**, 1093–1100 (2000).

119. Zoppini, G. *et al*. Usefulness of the Triglyceride to High-Density Lipoprotein Cholesterol Ratio for Predicting Mortality Risk in Type 2 Diabetes: Role of Kidney Dysfunction. *Atherosclerosis* **212**, 287–291 (2010).

120. Scargle, J. D. Publication Bias: The 'File-Drawer Problem' in Scientific Inference. *J Sci Explor* **14**, 94–106 (2000).

121. Koopman, R. J., Mainous III, A. G., Diaz, V. A. & Geesey, M. E. Changes in Age at Diagnosis of Type 2 Diabetes Mellitus in United States, 1988 to 2000. *Ann Fam Med* **3**, 60–63 (2005).

122. Hayes, A. J., Leal, J., Gray, A. M., Holman, R. R. & Clarke, P. M. UKPDS Outcomes Model 2: A New Version of a Model to Simulate Lifetime Health Outcomes of Patients with Type 2 Diabetes Mellitus Using Data from the 30 Year United Kingdom Prospective Diabetes Study: UKPDS 82. *Diabetologia* **56**, 1925–1933 (2013).

123. UK Diabetes. *Diabetes in the UK 2012 - Key Statistics on Diabetes*. (2012). at <https://www.diabetes.org.uk/About_us/What-we-say/Statistics/Diabetes-in-the-UK-2012/>

124. National Diabetes Data Group. Classification and Diagnosis of Diabetes Mellitus and Other Categories of Glucose Intolerance. *Diabetes* **28**, 1038–1057 (1979).

125. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. WHO Expert Committee on Diabetes Mellitus: Second Report. *World Heal. Organ Tech Rep Ser* **646**, 1–80 (1980).

126. Diabetes Mellitus. Report of a WHO Study Group. *World Heal. Organ Tech Rep Ser* **727**, 1–113 (1985).

127. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Report of the Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care* **20**, 1183–1197 (1997).

128. Herrett, E. *et al*. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol* **44**, 827–836 (2015).

129. National Institute for Health Research. The Clinical Practice Research Datalink. (2014). at <http://www.cprd.com/home/>

130. de Lusignan, S. *et al*. A Method of Identifying and Correcting Miscoding, Misclassification and Misdiagnosis in Diabetes: A Pilot and Validation Study of Routinely Collected Data. *Diabet Med* **27**, 203–209 (2010).

131. Health & Social Care Information Centre. *Quality and Outcomes Framework (QOF) - 2014-15*. (2015). at <http://www.hscic.gov.uk/catalogue/PUB18887>

132. de Lusignan, S. *et al*. Miscoding, Misclassification and Misdiagnosis of Diabetes in Primary Care. *Diabet Med* **29**, 181–189 (2012).

133. Bagheri, A., Sadek, A., Chan, T., Khunti, K. & de Lusignan, S. Using Surrogate Markers in Primary Electronic Patient Record Systems to Confirm or Refute the Diagnosis of Diabetes. *Inf. Prim Care* **17**, 121–129 (2009).

134. The Cochrane Collaboration. Cochrane Library. (1993). at <http://www.cochranelibrary.com/>

135. Health and Social Care Information Centre. *National Quality and Outcomes Framework Statistics for England - 2004-05*. (2004). at <http://www.hscic.gov.uk/catalogue/PUB01946>

136. Rollason, W., Khunti, K. & de Lusignan, S. Variation in the Recording of Diabetes Diagnostic Data in Primary Care Computer Systems: Implications for the Quality of Care. *Inf. Prim Care* **17**, 113–119 (2009).

137. Sadek, A.-R., van Vlymen, J., Khunti, K. & de Lusignan, S. Automated Identification of Miscoded and Misclassified Cases of Diabetes from Computer Records. *Diabet Med* **29**, 410–414 (2012).

138. NHS Diabetes & Royal College of General Practitioners. *Coding, Classification and Diagnosis of Diabetes*. (2011).

139. National Institute for Health and Care Excellence. *Chronic Kidney Disease: Early Identification and Management of Chronic Kidney Disease in Adults in Primary and Secondary Care*. (2014). at <http://guidance.nice.org.uk/cg182>

140. Yu, T.-Y. *et al*. Serum Vascular Adhesion Protein-1 Level Predicts Risk of Incident Cancers in

Subjects with Type II Diabetes. *Cancer Epidemiol Biomarkers Prev* **23**, 1366–1373 (2014).

141. Gerstein, H. C., Hoogwerf, B. & Rashkow, A. Albuminuria and Risk of Cardiovascular Events, Death, and Heart Failure. *JAMA* **286**, 421–426 (2001).

142. Student. The Probable Error of a Mean. *Biometrika* **6**, 1–25 (1908).

143. Fisher, R. A. On the interpretation of χ2 from contingency tables, and the calculation of P. *J R Stat Soc* **85**, 87–94 (1922).

144. Bonferroni, C. E. *Teoria Statistica delle Classi e Calcolo delle Probabilità*. (R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 1936).

145. Dunn, O. J. Estimation of the Medians for Dependent Variables. *Ann Math. Stat* **30**, 192–197 (1959).

146. Dunn, O. J. Multiple Comparisons Among Means. *JASA* **56**, 52–64 (1961).

147. Mantel, N. Evaluation of Survival Data and Two New Rank-Order Statistics Arising in its Consideration. *Cancer Chemo Rep* **50**, 163–170 (1966).

148. Mantel, N. Chi-Square Tests with One Degree of Freedom; Extensions of the Mantel-Haenszel Procedure. *JASA* **58**, 690–700 (1963).

149. Mantel, N. & Haenszel, W. Statistical Aspects of the Analysis of Data from Retrospective Studies of Disease. *J Natl Cancer Inst* **22**, 719–748 (1959).

150. Savage, I. R. Contributions to the Theory of Rank-Order Statistics - The Two-Sample Case. *Ann Math. Stat* **27**, 590–615 (1956).

151. Gray, R. J. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *Ann Stat.* **16**, 1141–1154 (1988).

152. Coviello, V. & Boggess, M. Cumulative Incidence Estimation in the Presence of Competing Risks. *Stata J* **4**, 103–112 (2004).

153. Fleiss, J. L., Levin, B. & Paik, M. C. *Statistical Methods for Rates and Proportions*. (Wiley).

154. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas* **20**, 37–46 (1960).

155. Naresh, C. N., Hayen, A., Weening, A., Craig, J. C. & Chadban, S. J. Day-to-Day Variability in Spot Urine Albumin-Creatinine Ratio. *Am J Kidney Dis* **62**, 1095–1101 (2013).

156. Oke, J. *et al*. Monitoring Young People with Type 1 Diabetes for Diabetic Nephropathy:

Potential Errors of Annual ACR Testing. *Diabetes Res Clin Pr.* **99**, 307–314 (2013).

157. NICE. *Chronic Kidney Disease in Adults: Assessment and Management.* (2014). at <http://nice.org.uk/guidance/cg182>

158. Gustafson, P. *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments.* (Chapman and Hall, 2003).

159. Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. *Measurement Error in Nonlinear Models: A Modern Perspective.* (Chapman and Hall, 2006).

160. Lauer, M. S., Blackstone, E. H. & Young, J. B. Cause of Death in Clinical Research: Time for a Reassessment? *J Am Coll. Cardiol* **34**, 618–620 (1999).

161. Engel, L. W., Strauchen, J. A., Chiazze Jr, L. & Heid, M. Accuracy of Death Certification in an Autopsied Population with Specific Attention to Malignant Neoplasms and Vascular Diseases. *Am J Epidemiol* **111**, 99–112 (1980).

162. Lloyd-Jones, D. M., Martin, D. O., Larson, M. G. & Levy, D. Accuracy of Death Certificates for Coding Coronary Heart Disease as the Cause of Death. *Ann Intern Med* **129**, 1020–1026 (1998).

163. StataCorp LP. stcurve - Plot Survivor, Cumulative Hazard, or Cumulative Incidence Function. 1–12 (2015). at <http://www.stata.com/manuals14/ststcurve.pdf>

164. Framinhan Heart Study. Framingham Risk Score. (2013). at <http://www.framinghamheartstudy.org/>

165. Stevens, R. J., Kothari, V., Adler, A. I. & Stratton, I. M. The UKPDS Risk Engine: A Model for the Risk of Coronary Heart Disease in Type II Diabetes (UKPDS 56). *Clin. Sci.* **101**, 671 (2001).

166. Centers for Disease Control and Prevention. National Center for Health Statistics. (2009). at <http://www.cdc.gov/growthcharts/percentile_data_files.htm>

167. Grambsch, P. M. & Therneau, T. M. Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika* **81**, 515–526 (1994).

168. Zhou, B., Fine, J. P. & Laird, G. Goodness-of-Fit Test for Proportional Subdistribution Hazards Model. *Stat Med* **32**, 3804–3811 (2013).

169. Mulnier, H. E. *et al.* Mortality in People with Type 2 Diabetes in the UK. *Diabet Med* **23**, 516–

521 (2006).

170. Kontopantelis, E. *et al*. Glucose, Blood Pressure and Cholesterol Levels and their Relationships to Clinical Outcomes in Type 2 Diabetes: A Retrospective Cohort Study. *Diabetologia* **58**, 505–518 (2014).

171. Andersen, P. K., Geskus, R. B., de Witte, T. & Putter, H. Competing Risks in Epidemiology: Possibilities and Pitfalls. *Int J Epidemiol* **41**, 861–870 (2012).

172. Latouche, A., Boisson, V., Chevret, S. & Porcher, R. Misspecified Regression Model for the Subdistribution Hazard of a Competing Risk. *Stat Med* **26**, 965–974 (2007).

173. Wolbers, M., Koller, M. T., Witteman, J. C. M. & Steyerberg, E. W. Prognostic Models with Competing Risks: Methods and Application to Coronary Risk Prediction. *Epidemiology* **20**, 555–561 (2009).

174. Boggess, M. Competing Risk Models. (2004). at <http://www.stata.com/statalist/archive/2004-05/msg00506.html>

175. Hippisley-Cox, J. & Coupland, C. Identifying Patients with Suspected Lung Cancer in Primary Care: Derivation and Validation of an Algorithm. *Br J Gen Pr.* **61**, e715–e723 (2011).

176. Hippisley-Cox, J., Coupland, C., Robson, J., Sheikh, A. & Brindle, P. Predicting Risk of Type 2 Diabetes in England and Wales: Prospective Derivation and Validation of QDScore. *BMJ* **338**, b880–b880 (2009).

177. Hippisley-Cox, J., Coupland, C., Robson, J. & Brindle, P. Derivation, Validation, and Evaluation of a New QRISK Model to Estimate Lifetime Risk of Cardiovascular Disease: Cohort Study Using QResearch Database. *BMJ* **341**, c6624–c6624 (2010).

178. Scheike, T. H., Zhang, M.-J. & Gerds, T. A. Predicting Cumulative Incidence Probability by Direct Binomial Regression. *Biometrika* **95**, 205–220 (2008).

179. Klein, J. P. & Andersen, P. K. Regression Modeling of Competing Risks Data Based on Pseudovalues of the Cumulative Incidence Function. *Biometrics* **61**, 223–229 (2005).

180. Checkley, W., Brower, R. G. & Muñoz, A. Inference for Mutually Exclusive Competing Events Through a Mixture of Generalized Gamma Distributions. *Epidemiology* **21**, 557–565 (2010).

181. Farmer, A. J. *et al*. Optimal Strategies for Identifying Kidney Disease in Diabetes: Properties of Screening Tests, Progression of Renal Dysfunction and Impact of Treatment - Systematic

Review and Modelling of Progression and Cost-Effectiveness. *Heal. Technol Assess* **18**, 1–128 (2014).

182. Collins, G. S. & Altman, D. G. Predicting the 10 year Risk of Cardiovascular Disease in the United Kingdom: Independent and External Validation of an Updated Version of QRISK2. *BMJ* **344**, e4181 (2011).

183. Klausen, K. *et al*. Very Low Levels of Microalbuminuria are Associated with Increased Risk of Coronary Heart Disease and Death Independently of Renal Function, Hypertension, and Diabetes. *Circulation* **110**, 32–35 (2004).

184. Di Angelantonio, E. *et al*. Chronic Kidney Disease and Risk of Major Cardiovascular Disease and Non-Vascular Mortality: Prospective Population Based Cohort Study. *BMJ (Clin Res Ed)* **341**, c4986 (2010).

185. Rebolj Kodre, A. & Pohar Perme, M. Informative Censoring in Relative Survival. *Stat Med* **32**, 4791–4802 (2013).

186. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2015). at <https://www.r-project.org/>