Ethnic inequalities in health and use of healthcare in the UK: how computerised health records can contribute substantively to the knowledge base

ROHINI MATHUR

Thesis submitted in accordance with the requirements for the degree of Doctor of Philosophy

University of London

OCTOBER 2015

Department of Non-Communicable Disease Epidemiology

Faculty of Epidemiology and Population Health

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by the Economic and Social Research Council

Research group affiliation: The Pathways Node of the National Centre for Research Methods

Declaration of Authorship

I, Rohini Mathur, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed Date

_____ _____

Rohini Mathur

Use of published work

Two papers have been published based on work undertaken for this thesis (Appendices A and B). All research for these papers was carried out as part of the PhD and took place during the period of registration of the PhD. The papers were based on drafts of Chapters 2 and 5 of this thesis and include some passages and tables and figures from these chapters. For both papers, Rohini Mathur (RM) was the lead and corresponding author, carried out the reviews and analysis and prepared all drafts of the paper. The co-authors' contributions to the manuscripts were restricted to providing comments on the drafts prepared by RM.

Abstract

Previous studies in the UK have established that minority ethnic groups as a whole experience more ill-health and onset of morbidity at younger ages or at lower levels of risk than the 'White British' population. Since the Race Relations Act of 1968, the official collection of ethnic group statistics by all government bodies has been mandated as a pre-requisite for identifying and tackling ethnic inequalities. The capture of ethnicity data in routine health records across the UK National Health Service forms part of this initiative. Although the facility to record ethnicity has been available in primary care since 1991 and in secondary care since 1995, until recently, unsystematic recording resulted in poor quality of the initial data, limiting the usefulness of these data for clinical care, commissioning and research. The incentivisation of ethnicity recording in 2006 as part of the Quality and Outcomes Framework has resulted in an improvement of the quality of these data, though their suitability for use in UK-wide population-based research, at the commencement of this PhD, had not yet been explored.

The studies reported in this thesis investigated the utility of electronic health records for research into ethnic differences in health and comprised three sub-studies. Firstly, the completeness, usability and generalisability of ethnicity data captured in primary and secondary care databases were assessed. Results showed that in 2012, valid ethnicity was recorded for 78.3% of patients in the Clinical Practice Research Datalink (CPRD), 79.4% of inpatients, and 50% of A&E patients and outpatients in the Hospital Episode Statistics for England (HES). Over 80% of patients with multiple ethnicities recorded had codes which either were identical or fell into the same five high-level ethnic group categorisation. The ethnic breakdown of the CPRD was found to be comparable to that of the combined censuses for England, Wales, Scotland and Northern Ireland, suggesting that studies of ethnic populations within the CPRD can be generalised to the UK population, particularly when using data from 2006 onwards, where completeness and consistency are highest.

Secondly, in collaboration with the UK Biobank study, a pragmatic and comprehensive definition of diabetes mellitus for use in electronic health databases was developed. Once applied to the CPRD, the algorithms identified 34,530 individuals with type 1 diabetes and 355,717 individuals with type 2 diabetes. The incidence of type 2 diabetes was almost doubled in South Asian compared with White groups (70.7 vs 42.0 events per 10,000 person years). After adjustment for gender and age group, the risk of type 2 diabetes was over three times higher in the South Asian group compared with White the group (Hazard Ratio 3.27 95%CI 3.19, 3.35).

Finally, a prospective cohort study of 860,000 patients registered with the CPRD was undertaken to quantify ethnic differences in the risk of incident coronary heart disease (CHD) and the extent to which this relationship is modified by the presence of type 2 diabetes. The presence of diabetes increased the risk of CHD by 40%, although this reduced to 22% after accounting for age, gender and deprivation (Hazard Ratio 1.22 CI95 1.20, 1.25). The excess risk associated with diabetes was markedly higher for ethnic minority groups, with an adjusted increase of 60% and 75% in South Asian and Black African/Caribbean groups respectively, compared with 28% in the White groups. Adjusted rates of CHD were consistently higher in South Asian groups and lower in Black African/Caribbean groups, with differences more pronounced amongst men than women. Ethnic differences in CHD risk were consistently more pronounced amongst patients without type 2 diabetes than in those with type 2 diabetes.

The studies have generated novel results which provide valuable information about the usability and generalisability of ethnicity data available in UK electronic health records. They have replicated findings from non-database studies of the prevalence and incidence of diabetes and extended our knowledge of the patterning of ethnic differences in heart disease outcomes. They represent the first ever use of UK routine electronic health records to answer these questions in relation to ethnicity. Together, the findings reported in this thesis provide a unique insight into the ways in which routinely recorded ethnicity data can be maximised for the purposes of epidemiological research into health inequalities across the UK.

Acknowledgements

Firstly, I would like to thank my supervisors, Professor Liam Smeeth and Professor Emily Grundy, for their guidance and encouragement throughout this project.

I am also thankful for the advice and direction received from my advisory committee members, Dr Krishnan Bhaskaran, Professor Nishi Chaturvedi and Professor David Leon.

Furthermore, I would like to acknowledge Dr Tjeerd van Staa, Professor Simon de Lusignan and Dr Louise Pealing, whose expertise and collaborative spirit helped shape the PhD throughout.

Finally, I would like to thank my husband Alex and my family for their encouragement and support over the last three years.

Funding

Rohini Mathur was supported by a three-year PhD fellowship from the Economic and Social Research Council, which funded the Pathways Node of the National Centre for Research Methods between 2011 and 2014.

This thesis is dedicated to my grandmother, Mrs Sushila Rani Mathur (1927–2015), and my father, Mr Rajul Purshottam Mathur (1954–2002), who both encouraged me to reach for the stars.

Table of Contents

List of Abbreviations

CALIBER	CArdiovascular disease research using LInked Bespoke studies and Electronic health Records
CCLG	Children's Cancer and Leukaemia Group
CHD	coronary heart disease
CPRD	Clinical Practice Research Datalink
CVD	cardiovascular disease
EMIS	Egton Medical Information System
GHS	General Household Survey
GPRD	General Practice Research Database
HES	Hospital Episode Statistics
IMD	Index of Multiple Deprivation
INPS	In Practice Systems
NICE	National Institute for Health and Care Excellence
NHS	National Health Service
QOF	Quality and Outcomes Framework
RCGP	Royal College of General Practitioners
THIN	The Health Improvement Network
T1DM	type 1 diabetes mellitus
T2DM	type 2 diabetes mellitus

List of Tables

List of Figures

Chapter 1
Introduction

1.1 Summary

This chapter presents the rationale for undertaking the programme of research, describes the aims and research questions and outlines the content of each of the thesis chapters.

1.2 Rationale for research

The social determinants of health are widely accepted as being integral to our understanding of why disease and mortality affect different population groups unequally. The concept of ethnicity is a vital tool with which to explore these differences, as it can provide valuable information about shared exposures for individuals with similar geographic origin, culture, language, beliefs about and access to health services. Across the UK, large-scale surveys have shown that minority ethnic groups experience higher rates of disease and poorer health-related outcomes than the 'white British' population.(1–6)

Health disadvantage accumulates over the life course, and is believed to contribute to the excess burden of disease amongst ethnic minority groups in the UK. As the ethnic diversity of the UK continues to grow, health outcomes for some white, black and south Asian ethnic minority groups remain worse than for the white British majority.(7–9)

Diabetes and cardiovascular disease are of particular concern and are responsible for a large number of healthcare consultations and healthcare expenditure across the National Health Service, with diabetes and diabetic complications alone accounting for 10% of all healthcare expenditures in the UK.(10) In the UK, as well as internationally, the burden of these diseases has been shown to be particularly increased amongst south Asian groups, with prevalence over two-fold and an incidence over six-fold higher than in White populations. Most

importantly, disease onset occurs at younger ages, and at lower levels of conventional risk factors. (11–14)

Though surveys throughout the UK have identified patterns by ethnic group in terms of disease risk, outcomes and mortality, no large-scale epidemiological studies have been undertaken to explore ethnic differences in cardio-diabetic prevalence, incidence and mortality at a national level. Large electronic health databases offer a pragmatic way of examining relationships between ethnicity and health across the UK, as they collate anonymised patient data from general practices and hospitals across the UK and have been shown to be representative of the general population. Furthermore, linkages to secondary care records, measures of social deprivation, disease registries and death records increase the utility of these databases for exploring epidemiological questions around the relationship between ethnicity and health outcomes.

These databases have been widely used to conduct observational epidemiological studies for a range of disease outcomes and population groups, but have not, to date, been used to examine ethnic differences in cardio-diabetic disease prevalence, incidence and outcomes. This is partly due to the historically poor quality and incompleteness of ethnicity data in electronic health records, which limited the suitability of these databases to investigate ethnic inequalities.

1.3 Aim and research questions

The aim of the research described in this thesis was to determine the extent to which ethnicity data in routine electronic health databases can be used for epidemiological research, and to utilise these data to examine how the burden of diabetes and coronary heart disease differs between ethnic groups across the UK. To this end, three observational studies utilising electronic health records from both primary and secondary care were undertaken.

1.3.1 Evaluating the quality of ethnicity data in electronic health records

The first study investigated the completeness and consistency of ethnicity recording in the Clinical Practice Research Datalink (CPRD) and Hospital Episode Statistics (HES). Improvements in completeness of recording from 1990 to 2013 and the consistency of coding for patients with multiple and linked records were assessed. The ethnic breakdown of the CPRD population was compared with that of the 2011 UK Census population and recommendations for the pragmatic use of these data for epidemiological and social research were established.

1.3.2 Improving classification of diabetes mellitus in electronic health records

The second study assessed the implementation of a series of algorithms to improve the classification and to reduce the misdiagnosis of patients diagnosed with type 1 and type 2 diabetes mellitus using routinely recorded diagnostic, therapeutic and clinical data. Ethnic differences in the incidence and prevalence of type 1 and type 2 diabetes using populations derived from the algorithms were quantified and the resulting estimates were compared with those from the original derivation cohort in the Welsh primary care database.

1.3.3 Exploring ethnic differences in the risk of coronary heart disease between patients with and without type 2 diabetes

The third and final study for this thesis examined ethnic differences in the incidence of first ever fatal and non-fatal coronary heart disease (CHD) for patients with and without type 2 diabetes. The excess risk of CHD in type 2 diabetics compared with non-diabetics was estimated overall and separately by ethnic group. Within the diabetic and non-diabetic populations, ethnic and gender differences in the crude and age-adjusted incidence of CHD and cardiac death were estimated. Finally, risk of CHD using multivariable Cox regression was estimated for South Asian and Black African/Caribbean groups compared with white, after accounting for time-varying measures of pharmacological treatment and clinical measures.

1.4 Outline of the thesis

The thesis comprises 12 chapters. The background, methods and findings relating to the first study, investigating the quality of ethnicity data in routine electronic health records, are presented in Chapters 3 to 5; those relating to the second study, on improving methods for the identification of diabetes mellitus, are presented in Chapters 6 and 7; the study comparing the incidence of CHD between patients with and without diabetics encompasses Chapters 8 to 11. A final discussion of all three studies is presented in Chapter 12. Two published papers, both reporting results from study 1, are included as Appendices, together with additional supplementary material.

Chapter 2 introduces the concept of ethnicity and describes the evolution of ethnicity recording in the UK and its subsequent inclusion into the National Health Service framework.

Chapter 3 describes the electronic health databases utilised throughout the research presented in the thesis and the ways in which ethnicity data are captured in each.

Chapter 4 presents a review of UK-based epidemiological studies which have utilised patient-level ethnicity data to examine health outcomes.

Chapter 5 details the study examining the quality, generalisability and usability of ethnicity data routinely recorded in primary and secondary care.

Chapter 6 begins with a brief overview of the literature on ethnic differences in diabetes and goes on to describe the development and implementation of a series of algorithms designed to improve the identification and classification of type 1 and type 2 diabetes mellitus.

Chapter 7 presents the ethnicity-specific prevalence and incidence of diabetes in the UK according to age, gender and calendar period.

Chapter 8 presents a review of the published literature on diabetes and CHD and describes the methods of defining cohorts of diabetic and non-diabetic patients between whom the incidence of CHD and cardiac death is compared.

Chapter 9 presents the results of the study of non-fatal CHD in the full CPRD cohort.

Chapter 10 presents the results of the study of fatal and non-fatal CHD combined in a subset of CPRD patients with linked mortality data from the Office for National Statistics (ONS).

Chapter 11 discusses the findings of the third study, in terms of ethnic differences, for both non-fatal CHD alone and fatal and non-fatal CHD combined.

In Chapter 12, the main findings of the three studies are summarised and discussed in the context of what is already known on the topics and what the studies presented herein contribute. The limitations of the data sources and study design are considered and the implications of the findings for future research and clinical practice are discussed.

Chapter 2
Ethnicity and health

2.1 Summary

This chapter briefly describes the current thinking around concepts of ethnicity and the key mechanisms by which ethnicity can influence health outcomes. This is followed by a description of the progress in ethnicity data collection in the UK. The chapter concludes with discussion of the challenges of classifying ethnicity and the ways in which the significance and interpretation of ethnicity as a concept evolve over time.

2.2 Definitions and concepts of ethnicity

Historically, the term ethnicity has been used synonymously with 'race' – the construction of humankind as being made up of biologically distinct subgroups. Developed in the 19th century, theories about 'race' were used as a means to justify the superiority of 'Caucasians' and policies supporting imperialism, eugenics and slavery.(15–18) However, the notion of race as a genetically immutable trait has been widely discredited – it is widely accepted that genetically distinct races are a myth, and that the genetic diversity within so-called 'races' is greater than that between races – and over the past 50 years definitions of race and ethnicity have diverged considerably. In 1951, the United Nations declared that:

"National, religious, geographic, linguistic and cultural groups do not necessarily coincide with racial groups: and the cultural traits of such groups have no demonstrated genetic connection with racial traits. Because serious errors of this kind are habitually committed when the term 'race' is used in popular parlance, it would be better when speaking of human races to drop the term 'race' altogether and speak of 'ethnic groups'." (19)

In contrast to 'race', which was defined by those in power and imposed upon others, ethnicity is now understood to reflect an individual's own self-identification, which encompasses a

broad range of socially constructed characteristics.(20) Bhopal posits that the process of allowing all individuals to choose their own ethnic identity can be empowering as it respects the primacy of self-identify and autonomy.(19) As such, ethnic self-identification can be fluid over time, responding to political and cultural forces.(21,22) The UK Department of Health now states that ethnicity is:

"complex, multifaceted and subjective, and defined by[:] a shared history, a common cultural tradition; a common geographical origin; descent from common ancestors; a common language; a common religion; and a distinct group within a larger community." (23)

It is now widely recommended that the concept of ethnicity replace the unscientific concept of 'race' in all spheres of research, as it is a more meaningful way of grouping individuals with some shared identity – encompassing, but not limited to, country of birth, religion, language, cultural practices and geography.(17,21)

Ethnic identity can evolve both within individuals over time and between generations. This is particularly salient for countries such as the UK, where the population has changed steadily with waves of large-scale immigration occurring over the past century. For example, first-generation migrants may identify ethnically with their home country, while their offspring may identify more strongly with newer social networks, though both generations may adapt over time. Similarly, children born to parents of different ethnic origins may identify with either or create a new identity for themselves.(1)

As such, there is no one universally accepted definition of ethnicity, as it is inherently tied to the social, cultural and political context in which it is used.

2.3 Ethnicity and health

When hypothesising about and interpreting the mechanisms through which ethnicity is related to health, it is essential to be clear that health outcomes are determined by factors associated

with ethnicity, not ethnicity itself. These factors include genetic influences, socio-economic deprivation, migration status, cultural practices and lifestyle. Their distribution is unequal in different population groups, and this gives rise to what can be conceptualised as ethnic differences.

2.3.1 Genetic influences

Defining how 'ethnicity' as a social construct impacts upon health and healthcare use is complicated; though current uses of 'ethnicity' focus on social determinants of health, this does not preclude the existence of some pertinent and biological variation. For example, the prevalence of the BRCA gene for breast cancer is higher amongst Ashkenazi Jewish populations, while the prevalence of the gene for sickle cell trait is highest amongst populations from Southern Europe, Africa and the Caribbean.(24,25)

Similarly, biological differences related to ethnicity are now being recognised and incorporated into evidence-based guidelines for clinical practice in the UK. Examples include the increased predisposition to type 2 diabetes and cardiovascular disease amongst South Asian groups, which has prompted the recommendation for the use of different thresholds for 'overweight' and 'obesity' in this population.(11,26)

2.3.2 Socio-economic deprivation

There is a well-established social gradient in health, with individuals of lowest socio-economic position experiencing poorer health outcomes than less deprived groups. Experiences of deprivation differ between ethnic groups, with individuals from some ethnic minority populations more likely to live in deprived neighbourhoods and be unemployed than the general population.(27–29) Early research into the relationship between ethnicity and deprivation often assumed that all ethnic groups experience disadvantage equally. However, research based on the UK Census shows that while Bangladeshi and African populations are more likely to live in deprived areas, Indian and Chinese groups live in more affluent areas and experience less material deprivation.(30) Ethnic disparities in health are inextricably linked to

socio-economic factors, with many ethnic minority groups, more likely to live in the most deprived boroughs of England.(31)

2.3.3 Migration

Although individuals mostly migrate to improve life for themselves and their families, the experience can often have negative influences on health. In addition to the physical and psycho-social stress of relocating to a place where limited language skills and lack of social network may limit economic, professional and social opportunities, migrant groups may also experience stigma and prejudice, increasing the likelihood of poor health outcomes among this population.(32) Furthermore, new immigrants are more likely than the host population to live in areas of high deprivation, such as inner cities, and have poor access to health and social services.(33,34)

These experiences may counter any initial advantage that new migrants may experience via the 'healthy migrant effect', which hypothesises that individuals who migrate long distances, particularly internationally, are healthier, better educated and less disadvantaged than those who remain in the home country.(35,36) This health selection effect is evidenced in lower mortality and chronic disease prevalence amongst first-generation migrants compared with those in the host country. However, over time, the process of 'acculturation' (the adoption of local lifestyle, dietary and cultural practices) causes the health profile of migrant populations and their descendants to converge with, and in some cases transcend, that of the host country.(37–39)

The timing of migration has a great impact on health due to the differing levels of acculturation experienced by different ethnic groups.(37,40) The UK has experienced several waves of immigration, notably since the end of the Second World War, when individuals living in the Commonwealth were encouraged to migrate to the UK. Prior to the war, the largest migrant group came from the Republic of Ireland. Immediately following the war, there was a large influx of individuals from Poland and India. This was followed by large-scale migration from

the Caribbean, starting in the 1950s. During the 1960s, a large increase in the Indian- and Pakistan-born population occurred. This was followed by an influx of residents from African Commonwealth countries in the 1970s. The UK's Bangladeshi population largely settled from the 1980s onwards. According to the 2011 Census, the proportion of individuals born outside the UK increased from 4.3% in 1950 to 13.4% in 2011.(41) In 2011, the largest foreign-born population was from India, with those from Poland second and those from Pakistan third.(41)

2.3.4 Cultural practices and lifestyle

Shared cultural norms around health-seeking behaviours, diet, exercise and religious practices may directly impact health in both positive and negative ways. For example, the fourth Health Survey for England found that while rates of smoking are higher in Bangladeshi men than other men in England, over 90% of all Bangladeshi adults are non-drinkers.(42) Targeting services and interventions at specific risk groups has demonstrated success in diffusing ideas and practices for improving health.(43)

2.4 History of ethnicity data collection in the UK

The UK is one of the few countries in Europe that emphasises the need for positive action to promote ethnic equality via the collection of official ethnic group statistics. In a recent review of practices of data collection in censuses across Europe, only 5 out of 35 countries surveyed collected ethnicity data in the most recent census, of which the UK was one.(44) Since the Race Relations Act of 1968, the official collection of ethnic group statistics has been mandated as an essential first step towards identifying and actively reducing ethnic inequalities.(45–50)

The origins of the modern system of ethnic group data collection in the UK can be traced to the Census Act of 1920, which recommended that 'race and nationality' statistics be collected as part of the Census for Great Britain in order to better understand the needs and circumstances of the population.(51) However, at that time it was felt that the existing question on country of birth, which had been used since 1841, was sufficient; the recommendation would not be acted upon until 70 years later.

In lieu of a national programme to capture ethnicity information for the whole population, the collection of ethnic statistics was incorporated into national surveys, the first of which was the General Household Survey (GHS) for Great Britain. Launched in 1971, the GHS captured information on aspects of family life, education, health and employment in order to inform social policy and resource allocation by the government. Contrary to principles widely held today, the GHS required that ethnicity be assigned by the interviewer, who was asked to record whether, upon visual inspection, the respondents appeared 'White', 'Coloured' or 'Unknown'. If a child was unseen, his or her ethnicity was 'imputed' from that of the parents. Self-classification of ethnic origin was introduced to the GHS in 1983 and used until the survey's close in 2012.(49,52,53)

2.4.1 The 1981 Census

The question of recording ethnicity in the Census was revisited in the years leading up to the 1981 Census when it became clear that country of birth was no longer sufficient for tracking the growth of the increasingly diverse UK population. As an interim measure, a question on parents' country of birth had been included in the 1971 Census, with the intention of deriving ethnicity indirectly.

However, large-scale immigration from both the Commonwealth and Europe dating back to the 1950s meant that questions on country of birth, and parents' country of birth, became increasingly irrelevant given the growth in second- and third-generation ethnic minority populations born in the UK. Despite extensive pilot testing and the development of a draft question, the plan to incorporate an ethnicity question was eventually abandoned due to strong opposition from civilian groups and the poor acceptability of the question to Afro-Caribbean populations following tensions in the 1970s.(1,54)

Following the 1981 Census, an enquiry by the House of Commons Home Affairs Committee on Race Relations concluded that questions on ethnicity and language should be included in

the 1991 Census. A House of Commons Select Committee Report set out the positive aspects of monitoring ethnic information, stating:

"The object of asking ethnic questions is, in conjunction with other indicators of general disadvantage, to assist Government and local authorities to identify and work against all aspects of racial disadvantage and racial discrimination." (55)

2.4.2 The 1991 Census

Bolstered by high levels of support from the 1989 post enumeration survey, a question on ethnicity was added to the 1991 Census for England, Wales and Scotland, but not Northern Ireland, where a question on ethnicity was not introduced until 2001.(56–58) The term 'ethnic group' was used instead of 'ethnic origin – which was found to be less acceptable to respondents, due to the implied meaning of historical background.(59)

The concept of ethnicity adopted by the Census was, and remains, that of self-classification, as recommended by the Commission for Racial Equality, ensuring that ethnicity refers to the individual's self-perception, rather than how they appear to others, recognising the fact that an individual's self-conceptualisation may change over time.[1] The 1991 Census ethnic group question for Great Britain consisted of the nine pre-coded categories visible on the household form plus 28 additional ethnic groups derived from any multi-ticking of the boxes and the written descriptions given in either of free text boxes under 'Black other' and 'other' (Table 2.1). Written descriptions which had the same or similar meaning to one of the pre-coded categories were assigned the relevant code between 0 and 6. Written responses for 'Black other' were assigned a code between 7 and 17, while written responses for "any other ethnic group" were allocated a code between 18 and 34. In the 1991 Census, 98.6% of respondents selected a pre-coded category while 1.4% specified their ethnic group as 'Any other ethnic group'. Output from the 1991 Census was collapsed into 10 categories for England and Wales. For Scotland, this was further reduced to 4 categories. Many outputs from the 1991 Census also included a 'born in Ireland' category derived from the country of birth question.

There is criticism that the 1991 Census managed only to separate out two broad ethnic minority groups, Black and South Asian, and grouped together "White" populations, despite the known social and health disparities.(18,60–62) Furthermore, the 1991 Census did not allow individuals to identify themselves as British, regardless of their geographic or ancestral origin. For example, 26% of the free text responses under the "Black- other" category stated "British".

2.4.3 The 2001 Census

In response to lobbying for the recognition of the poor social conditions for Irish populations born both inside and outside Great Britain, the 2001 Census expanded the White category to differentiate British, Irish and Other white. The most significant change from the 1991 Census was the addition of a mixed ethnicity category, which recognised the significant population of individuals born to parents of different ethnic groups.(60) This increased the number of ethnic group categories to 16 in England and Wales, 14 in Scotland and 12 in Northern Ireland.

Religion was added for the first time to the 2001 Census for Great Britain as an optional question. Pilot testing of the 2001 Census found that some respondents felt that religion was a more useful indicator, particularly for individuals from south Asia and Ireland, where religion may form the primary measure of identity, rather than ethnicity.(61)

According to the 2001 Census for England and Wales, 8% of the population identified themselves as members of minority ethnic groups. Furthermore, 8% of the population of England and Wales in 2001 was born out of the UK; half of this group classified themselves as non-white.(62)

Table 2.1 Evolution of census ethnic categories for England and Wales, 1991–2011

1991	2001	2011
1 White	*White*	*White*
	1 British	1 English/Welsh/Scottish/Northern Irish/British
	2 Irish	2 Irish
	3 Any other White background (write in)	3 Gypsy or Irish Traveller

		4 Any other white background (write in)
	Mixed	*Mixed/Multiple Ethnic Groups*
	4 White and Black Caribbean	5 White and Black Caribbean
	5 White and Black African	6 White and Black African
	6 White and Asian	7 White and Asian
	7 Any other mixed background (write in)	8 Any other Mixed/multiple ethnic background (write in)
	Asian or Asian British	*Asian or Asian British*
2 Indian	8 Indian	9 Indian
3 Pakistani	9 Pakistani	10 Pakistani
4 Bangladeshi	10 Bangladeshi	11 Bangladeshi
	11 Any other Asian background (write in)	12 Chinese
		13 Any other Asian background (write in)
	Black or Black British	*Black/African/Caribbean/Black British*
5 Black-Caribbean	12 Caribbean	14 African
6 Black-African	13 African	15 Caribbean
7 Black Other (write in)	14 Any other Black background (write in)	16 Any other Black/African/Caribbean background
	Chinese or other ethnic group	*Other ethnic group*
8 Chinese	15 Chinese	17 Arab
9 Any other ethnic group (write in)	16 Any other ethnic group (write in)	18 Any other ethnic group (write in)

2.4.4 The 2011 Census

In the 2011 Census, the White category was further expanded to incorporate Gypsy or Irish Traveller. Furthermore, Chinese ethnicity was reclassified from the 'Other' group to Asian, while a new category for Arab was added to the 'Other' group. In the 2001 Census, only people selecting "White" ethnicity could identify as "British". This problem was addressed in the 2011 Census by creating a distinct question for national identity; thus respondents could identify themselves as British (or any other nationality) independently of ethnicity.(63)

2.4.5 The validity of ethnic categories used in official statistics

The question on ethnicity has been refined over the past three waves of the census, making direct comparisons problematic. A reduced set of seven high-level categories has been found by Simpson and Akinwale to show stability over time (Table 2.2).(62)

Table 2.2 Census ethnicity groups, 1991 and 2001

1991 Categories	2001 Categories	Harmonised Groups
White	White + all subgroups	White
Black Caribbean	Black or Black British- Caribbean	Black Caribbean

Black African	Black or Black British- African	Black African
Indian	Asian or Asian British- Indian	Indian
Pakistani	Asian or Asian British- Pakistani	Pakistani
Bangladeshi	Asian or Asian British- Bangladeshi	Bangladeshi
Chinese	Chinese or Other- Chinese	Chinese
Black - Other	Black or Black British- Other Black	Other (not comparable over time)
Other- Asian	Asian or Asian British- Other Asian	
Other	Chinese or Other- Other	
Other	Mixed + all subgroups	

***Simpson and Akinwale (2006)**

2.5 Challenges in classifying ethnicity

One of the conceptual challenges of disaggregating populations by ethnicity is the way in which ethnic groups are defined and understood. Ethnic groups themselves should not be considered to be homogeneous, as it is well established that high-level groupings can conceal significant heterogeneity.(64–67)

Senior & Bhopal have suggested that, when utilising ethnicity as an epidemiological variable, it should be defined so as to allow the differentiation of populations in a way that is relevant to health and facilitates the generation of hypotheses about the aetiology of disease or explains observed differences in healthcare usage and outcomes.(67) However, there is no way to create a classification scheme that is valid and meaningful across all settings, as the relevance and validity of ethnic categories depend entirely upon the context in which they are used. This provides a challenge when conducting epidemiological research into ethnicity, where categories for explanatory variables must be meaningful, discrete and fixed in order to be interpretable. In order to investigate ethnicity effectively, it must be operationalised into practical analytical categories which are understood to encompass the rich variety of concepts, but are however constrained by the time and socio-cultural context in which they were developed.(60)

In both the US and the UK it has been acknowledged that the ethnic categories used in official statistics are, to some extent, arbitrary and have been selected primarily for pragmatic reasons.

A US directive for the collection of race and ethnicity data states *"These classifications should not be interpreted as being scientific or anthropological in nature."*(65) Indeed, one of the concerns about the original ethnicity question for 1981 was that:

"The question confused colour with ethnic and national origin and the category "white" in particular would be open to sensational and damaging treatment in the popular Press." (68)

Just as individuals can change their ethnic identity over time, so too does the classification of ethnicity evolve, with more categories being included in standard classifications, reflecting numerical growth in representation of certain groups on the one hand, and on the other more nuanced understanding of meaningful categories. This is of particular importance in countries such as the UK, where historic changes in migration patterns have increased the ethnic diversity of the country as well as created new ethnic categories for later generations born in the UK itself, with multiple and novel ethnic identities. Provided that researchers recognise the limitations of categories and approach them critically, the study of ethnic differences can nonetheless provide vital information about patterns of health and social indicators, and provide an essential foundation for tackling inequalities between different populations. As articulated by Mason,

"We should not be afraid to use categories which are not embraced by actors themselves if these can illuminate patterns of disadvantage and domination. We must, however, be clear when we are doing so and not imply that these categories coincide with the identities of those to whom they refer." (69)

As the ethnic diversity of the UK continues to grow, health outcomes for some White, Black and South Asian ethnic minority groups remain far worse than for the White British majority.There have been concerns that research on ethnicity and health has been of poor quality, with some researchers failing to explicitly report how their ethnic categories have been derived, which concepts they are intended to represent and what role they expect ethnicity to play in the phenomenon under study.(67,70,71)

White ethnicity is typically used as the reference group in studies of ethnic differences. In the UK, where the majority of the population identifies as "White British, Scottish, or Irish", this approach can be justified on the grounds that this ethnic group is the largest. However, this approach also implies that the experience of the White group represents the norm, to which other groups should aspire.(72)

2.6 Conclusions

From unscientific beginnings, the concept of ethnicity has evolved to encompass the spectrum of biological, social and cultural influences that are understood to form part of our individual identity. As the concept has evolved, so has the difficulty with which it can be operationalised into a single variable for research purposes. In the UK, and indeed worldwide, a pragmatic approach has been undertaken to create ethnic categories which are simple and meaningful, but still largely based on racial and geographic boundaries. Routinely collected national data are already widely used to examine health outcomes by gender, age, socio-economic status and country of birth. The introduction of standardised ethnic categories across the government and the National Health Service (NHS) has massively increased the potential for these same data to be used to examine ethnicity in great detail both across representative population samples, as found in the Clinical Practice Research Datalink, and in selected population samples, such as the Hospital Episode Statistics.

Chapter 3
Source electronic health databases

3.1 Summary

This chapter introduces the two primary electronic health databases used to conduct the three studies reported in this thesis: the Clinical Practice Research Datalink (CPRD) and the Hospital Episode Statistics for England (HES). A brief history of ethnicity recording across the UK National Health Service and the way in which ethnicity data are coded in each source is described.

3.2 The Clinical Practice Research Datalink

Across the UK there are numerous primary care databases which bring together electronic patient records; however, most of these cover small geographical areas, or small numbers of general practices. The Clinical Practice Research Datalink (CPRD, formerly GPRD) is one of three clinical research databases which provide patient data from across practices in the UK, allowing for research to be undertaken on samples generalisable to the whole population. The other two databases like the CPRD are the Health Improvement Network Database (THIN) and the QRESEARCH database.

The CPRD was initially set up in 1987 as a commercial databank by the company VAMP (Value Added Medical Products). Now run by the Medicines and Healthcare products Regulatory Agency (MHRA), the CPRD is the largest primary care database in the UK, covering just over 8% of the UK population.(73–75)

The UK has the advantage of near-universal registration with general practitioners, around 98% of the entire population. As such, analyses of the registered patient population are widely representative of the UK population, though notable exceptions include asylum seekers, the homeless, prison populations and those in the armed services, who are less likely to access

GP services.(76–78) Additional linkages to secondary care data, disease registries, surveys and vital statistics give these databases unique value for observational studies and increasingly for pragmatic clinical trials.(79–82)

The CPRD currently contains longitudinal primary care records for approximately 13.5 million patients, of whom 5.5 million are currently active. Continuous observational data have been collected in most practices for over six years, yielding over 30 million patient years of observation.[78] Patients contributing to the CPRD have been shown to be representative of the UK population in terms of age and gender, though in terms of regional representation the north of England is slightly under-represented.(74) Importantly, the validity of a wide range of diagnostic and clinical measures has been established, with a 2010 systematic review demonstrating a mean positive predictive value of 88% across a range of 183 diagnoses.(83–86) The distribution of general practices contributing to the CPRD compared with the distribution of all general practices in the UK in July 2012 is shown in table 3.1.

Table 3.1 Regional distribution of practices contributing to the July 2012 CPRD compared with the UK distribution

Region	CPRD July 2012	%	UK April 2012	%
England	483	77%	8123	82%
Scotland	69	11%	998	10%
Wales	50	8%	474	5%
NI	22	4%	354	4%
Total	624	100%	9949	100%

All patients contributing to the CPRD are registered with 624 practices which all use the Vision clinical software system. Vision is one of several clinical software systems recommended for use in primary care by the GP Systems of Choice (GPSoC) Initiative, which supplies information technology systems to general practices across the UK. Other software systems include Egton Medical Information Systems (EMIS) and TPP System One, amongst others.(87)

General practitioners and practice staff record data onto their clinical systems and send anonymised patient data every 6 weeks to the CPRD. These data are then appended to the continually growing database, which contains information on diagnoses, symptoms, referrals, test results, medications, consultations, demographics, and lifestyle factors. Fifty per cent of English practices contributing to the CPRD also allow linkage to other data sources, such as the Hospital Episode Statistics for England and the Office for National Statistics (ONS) Mortality Data.

Quality of research data is audited at both the patient and practice level by the CPRD team. Individual patient data are defined as being of 'Acceptable Research Quality' (ARQ) if they are free of gaps or inconsistencies which cast doubt on the accuracy of the data recorded. Practices are required to record a minimum of 95% of prescribing and relevant patient encounter events. Data from practices are routinely validated by internal checks. Practice-level data are defined as being Up to Standard' (UTS) if it conforms to set of 10 metrics, including having a high proportion of patients with ARQ data, and having rates of prescriptions, deaths, pregnancies and referrals comparable to other practices. The first practice to meet these quality criteria did so in 1987, with most other general practices reaching the same level of quality by 1991.

For research purposes, individual patient data is anonymised, with identifying information such as NHS number, name, date of birth, address and postcode removed. Information such as gender and year of birth are retained in order to conduct stratified analyses. In addition to these demographic data, researchers can access coded data pertaining to diagnoses, symptoms and processes of care. Free text entered by the primary care team are not routinely available to researchers, as these may contain identifiable information. Coded data are entered according to the Read clinical coding system, a hierarchical system of medical coding used across UK primary care.(88,89)

The CPRD is organised into 10 file types, each of which contains a subset of the patient record. For research purposes, information from these files can be joined using the anonymised patient or practice identifier. The file types are described in table 3.2.

Table 3.2 Description of Clinical Practice Research Datalink file types

File type	Contents	Example fields
Patient	Basic demographics and registration details	Anonymised identifier, year of birth, registration date, transfer out date, death date
Practice	Details for all participating practices	Practice identifier, geographical region, date of becoming "up to standard", date of last data collection
Staff	Practice staff details	Staff identifier, gender, role
Consultation	Information about consultation type as entered by the GP	Consultation identifier, consultation type, consultation date, staff identifier, consultation duration
Clinical	All medical history including symptoms signs and diagnoses	Date of clinical event, date of data entry, clinical code, episode type, additional details identifier
Additional	Details relating to events coded in the clinical file	Patient identifier, entity type, data fields (depends on entity type)
Referral	Information about referrals to external care centres	Referral method, referral specialty, referral type, attendance type, referral urgency
Immunisation	Details of immunisation records	Immunisation reason, type, stage, status, compound used, location, reason for immunisation, route of administration
Test	Test results linked to events coded in the clinical file	Type of test, result, normal range for result, unit of measure
Therapy	All prescriptions issued by the GP	CPRD product code, British National Formulary Code, product name, dosage, quantity, pack size, number of days prescribed

3.3 The Hospital Episode Statistics for England

Hospital Episode Statistics (HES) have been collected since 1989 on all patients in England receiving care in NHS or NHS-commissioned hospitals (such as private hospitals). Prior to the introduction of HES, only 10% of admitted patient data were recorded nationally. The primary purpose of HES is to monitor hospital workload and audit procedures undertaken across inpatient, outpatient and accident and emergency (A&E) settings. Data held in the HES database can be used to conduct research on the epidemiology of disease presenting in secondary care and also to examine differences in workload, waiting times and length of stay by socio-demographic factors and region. Data on inpatient episodes have been available

since 1989, while outpatient and accident and emergency data have been available since 2003 and 2007 respectively.(90)

HES data contain information falling into four main domains:

I. socio-demographic information, such as age, gender, ethnic group and socio-economic status
II. clinical information including up to 20 diagnoses coded using the International Classification of Disease, Version 10 (ICD-10) and up to 24 procedures coded using the Office of Population Censuses and Surveys, Version 4 (OPCS).(91)
III. administrative information, such as admission and discharge dates, time on waiting list, length of stay, and consultation information
IV. geographic information, such as region of the patients' residence, region of the hospital and NHS trust.

HES data are input by hospital-based clinical coders at the end of each admission and curated by the NHS Information Centre for Health and Social Care (NHS IC)

Overall completeness of HES data is high, with the number of admissions captured in the database closely reflecting the true number. Quality of HES data is audited against standards of completeness, duplication and consistency. Annual quality reports are produced detailing known problems and hospitals which are under-reporting admissions.

For the purposes of this study, the HES will be examined in two ways: firstly, as a standalone dataset in which completeness and patterns of ethnicity recording will be assessed for inpatients, outpatients and A&E admissions; and secondly, as a linked dataset for patients registered with the 308 of the CPRD practices (approximately 50%), in order to examine discrepancies between linked patient ethnicity data.

The linkage between the CPRD, HES and ONS Mortality datasets at the start of the research project in July 2012 is displayed in figure 3.1.

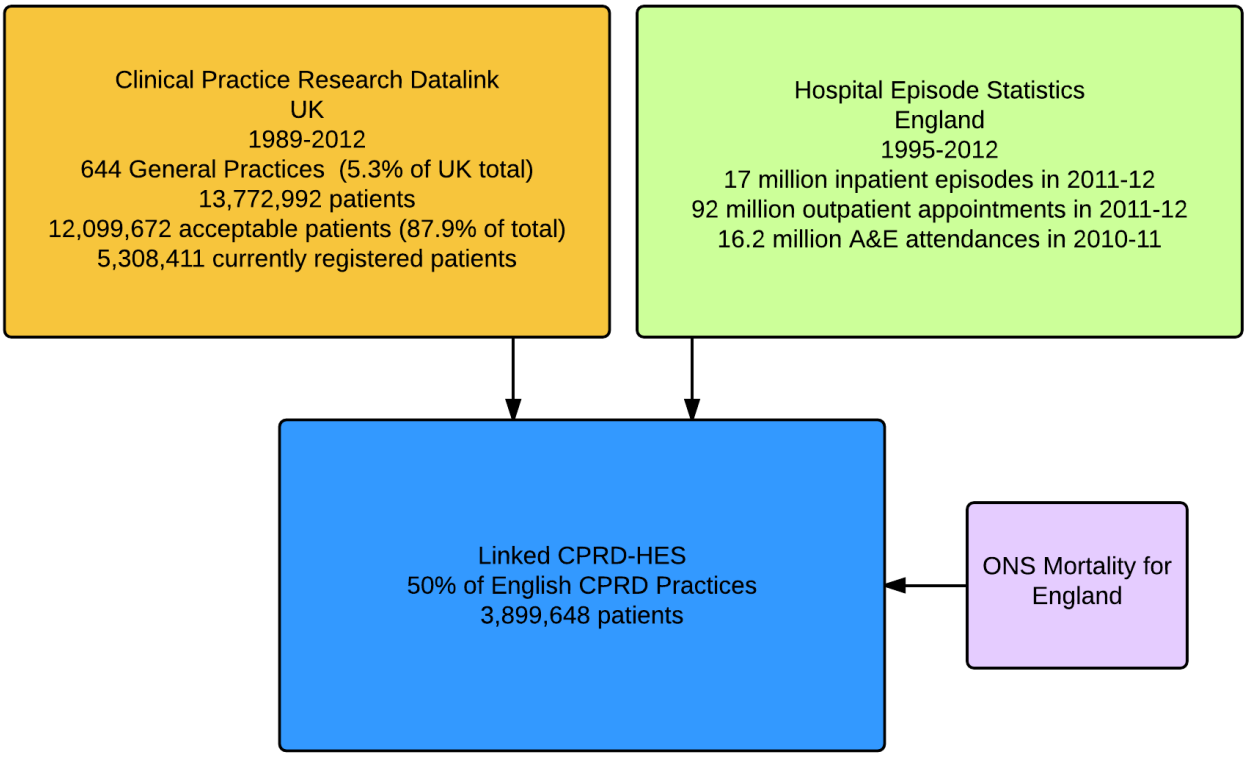


Figure 3.1 Schematic of data source linkage as of July 2012

3.4 Ethnicity recording in UK primary care

The capture of ethnic group information in routine health records is recognised in the UK as a pre-requisite to addressing inequalities in health service usage and health-related outcomes.(23,45,47,92,93) In line with the broader UK government strategy to promote equality across all departments, the Department of Health mandated the recording of ethnicity data to ensure equity of health service provision throughout the National Health Service.(94)

Although the facility to record ethnicity was introduced into primary care in 1991 and into the Hospital Episode Statistics for England in 1995, unsystematic implementation resulted in poor completeness and quality of the initial data, limiting their intended use for clinical care, commissioning and research.(52,95–97)

The 16 ethnic group categories defined by the 2001 Census for England and Wales currently form the national standard for mandatory ethnicity data collection across the National Health Service.(23) The Read system contains 88 unique codes for ethnicity, all of which map to the 16 categories for ethnic group based on the 2001 Census for England and Wales.(98) In 2011, Scotland became the first country in the world to record ethnicity on death certificates, however, ethnicity is still not routinely recorded on birth certificates anywhere in the UK.(99,100)

Though ethnicity recording was mandated across the National Health Service in 1991 alongside the Census, until recently, electronic health records have been of limited use for examining associations between ethnicity and health due to the poor completeness and quality of the data.(94,101) Instead, past studies exploring ethnicity have variously ascribed patient ethnicity indirectly, via name-recognition software or by estimating ethnic population size from Census data, both methods of questionable validity, particularly for individuals of mixed ethnicity and for descendants of migrants.(102–104)

Individual-level ethnicity data is becoming increasingly available in both primary and secondary care, providing a novel opportunity to conduct research across the whole of the UK population into ethnic differences in the healthcare usage and outcomes. The computerisation of healthcare records across both primary care (general practice) and secondary care (hospital settings) has generated enormous potential for population-based research on morbidity and the use of health services.

In order to improve the quality and completeness of ethnicity data available in primary care, the recording of ethnicity was incentivised under the Quality and Outcomes Framework (QOF) in 2004.(105,106) Under the scheme, general practices were awarded 1 QOF point per annum (equivalent to £125 for an average practice) for recording the ethnicity of 100% of all newly registered patients in each financial year.

In 2008 this scheme was replaced with a new enhanced service which remunerated practices 5.6 pence for every new patient whose ethnicity was captured. The recording of ethnicity was removed from the programme in April 2011, as it is now expected that general practices will record ethnicity, along with first language, routinely in order to meet the needs of their registered populations.(107) Data from the Quality and Outcomes Results database shows that 92% of general practices across the UK are now routinely recording ethnicity for 100% of their newly registered patients, up from 83% in 2007 (Figure 3.2).(106)

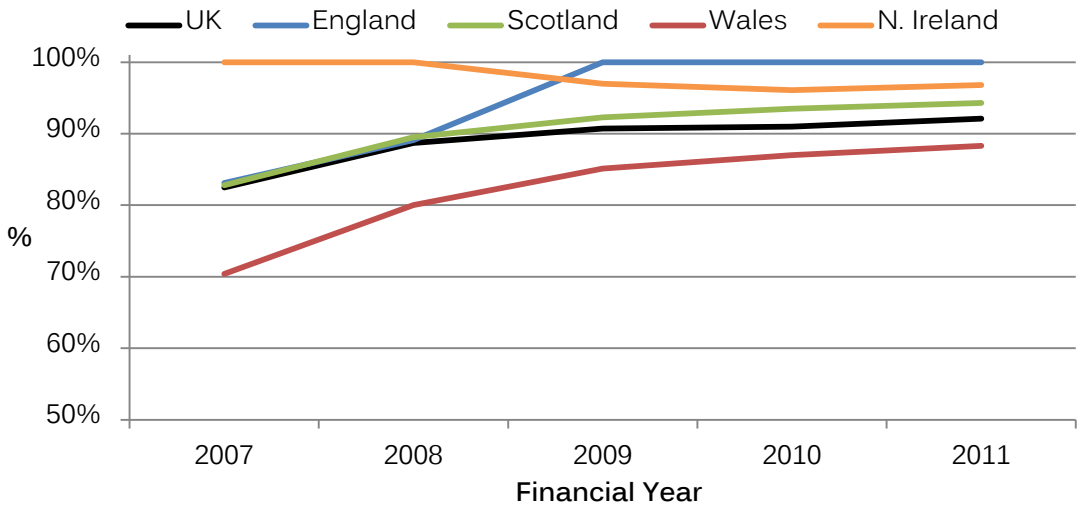


Figure 3.2 Proportion of UK practices achieving 100% ethnicity recording for all new patients, 2007-2011

3.5 Ethnicity recording in UK secondary care

Turning to hospital settings, a scheme was proposed in 1992 which required general practitioners to record patient ethnicity in all referral letters so that other NHS care providers need not repeatedly ask the patient.(108) This scheme was unsuccessful and recording of ethnic group data was later incorporated into the Hospital Episode Statistics in 1995.(96,109) In 2001, the ethnic codes first introduced in 1995 were updated to match the groupings of the 2001 Census.(97) Ethnic group data has been available for outpatients since 2003 and for A&E since 2007.(110) Hospital Episode Statistics are available for England only and cover all NHS admissions, as well as private patients. In 2010, valid ethnicity was being recorded for over 91% of all finished consultant episodes in England.(97)

3.6 How ethnicity data are captured

In primary care it is expected that ethnicity will be recorded in one of two ways: either the patient will be asked to fill in a drop-down checklist of similar categories to those of the 2001 Census, or the patient will be asked face to face by the general practitioner or other practice staff.(23) This can take place during registration or consultation. In hospital settings, guidelines state that ethnicity should be self-reported by patients whenever possible, with assistance from relatives, interpreters and advocates where necessary.(111) Qualitative studies exploring patient and practitioner perspectives have reported that the collection of ethnicity data is widely accepted as being important, though concerns around the need for repeated collection of ethnicity, time burden on practitioners, limitations of the 16 standard categories and lack of coordination between primary and secondary care have been expressed.(112,113)

Chapter 4
Current uses of routinely collected ethnicity data for population-based research in the UK: a literature review

4.1 Aim

The aim of the review was to ascertain how routinely available ethnicity data from electronic health databases in the UK is currently being used in observational studies.

4.2 Summary

This chapter summarises the current body of work that has made use of ethnicity data as routinely recorded in UK based primary and secondary care with the explicit purpose of examining health outcomes. The search strategy and inclusion and exclusion criteria are described. The results of the review and a discussion of the findings and rationale for the first of the three studies, of ethnicity recording (described in Chapter 5), conclude the chapter.

4.3 Search strategy

In February 2012, a keyword search in MEDLINE, EMBASE and Web of Knowledge was undertaken to identify all studies utilising data from four UK-based electronic health databases: the Clinical Practice Research Datalink (CPRD), The Health Improvement Network (THIN), the QRESEARCH database and the Hospital Episode Statistics (HES). Only three literature databases were interrogated as saturation was reached,- with the majority of articles appearing in all three databases. As my main interest was in the four selected health databases rather than the topic of electronic health records as a whole, no Medical Subject Headings (MESH) were required in the search. Inclusion and exclusion criteria are detailed in table 4.1.

Duplicates were removed by Endnote X5 and by hand searching. Abstracts were scanned to remove articles that did not describe primary observational research using the electronic health databases of interest. The remaining abstracts were finally reviewed in detail to determine whether patient-level ethnicity was utilised in the study.

Table 4.1 Search terms and inclusion and exclusion criteria for systematic review

Database	Text term
CPRD	"General Practice Research Database" or "GPRD" or "Clinical Practice Research Datalink" or "CPRD"
THIN	"Health Improvement Network"
QRESEARCH	"QRESEARCH"
HES	"Hospital Episode Statistics" or "HES"
Study Type	
Included:	Quantitative studies using the electronic health databases of interest Quantitative studies using patient-level ethnicity data as a variable in the analysis
Excluded:	Articles relating to other electronic health databases Studies that did not utilize patient-level ethnicity data derived from The database of interest (i.e. / used area level measures of ethnicity, Attributed Census derived ethnic distributions to the data) Studies which did not report patient level outcomes (i.e. / ecological Studies, reports about the technology of EHR, statistical methods, Data quality)
Publication Type	
Included:	Peer-reviewed articles published in journals
Excluded:	Review articles, opinion pieces, letters, editorials, summaries, conference abstracts, non-English publications

4.4 Results

From a total of 1,232 observational studies using UK-wide electronic health databases, 15 made use of the patient-level ethnicity data available (Figure 4.1).

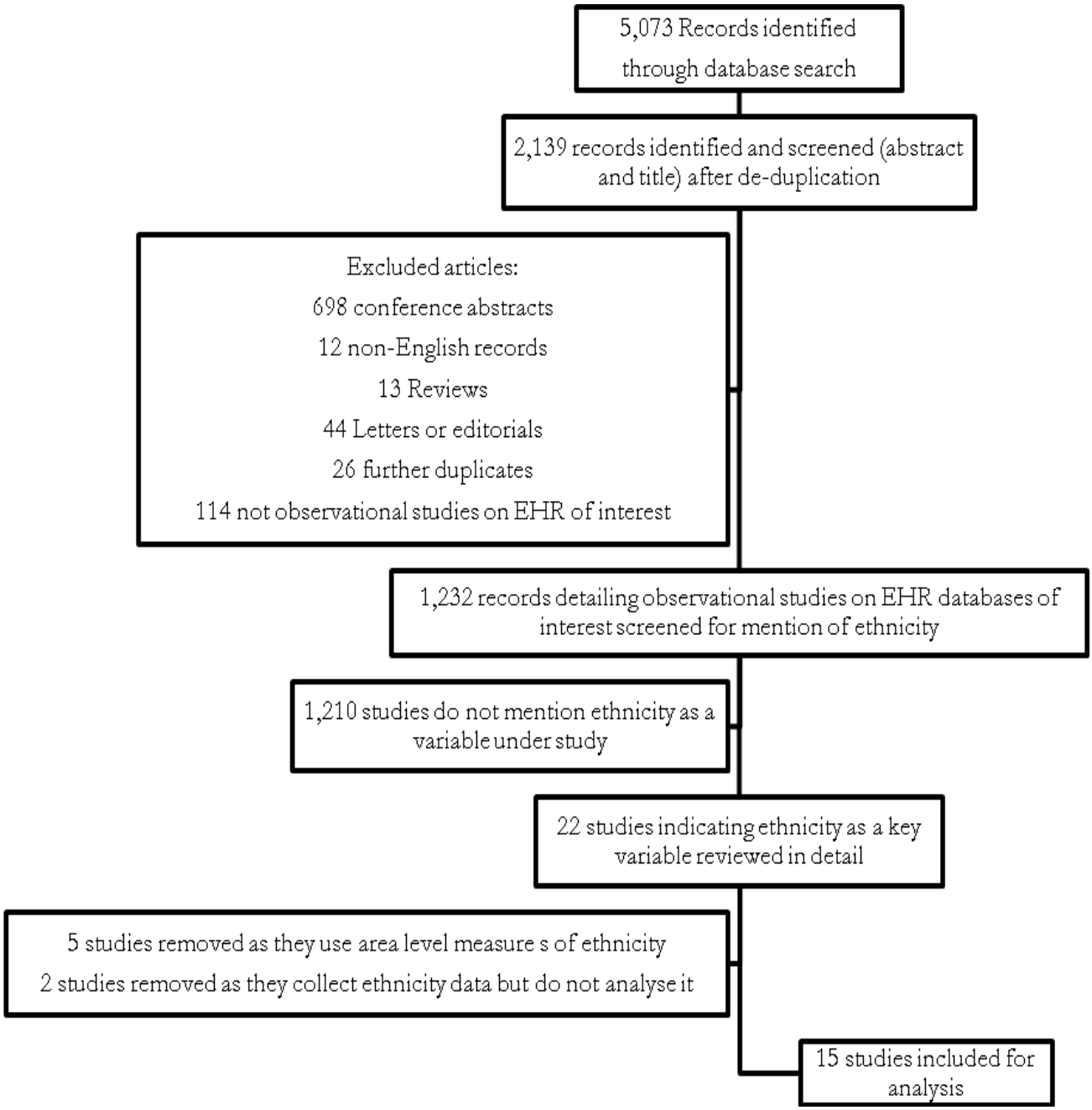


Figure 4.1 Systematic review search results

4.4.1 Study characteristics

The key characteristics of the 15 included papers are summarized in table 4.2. Briefly, 10 studies utilized the ethnicity data held within HES, 3 utilized the QRESEARCH database and 2 utilized THIN. None of the included studies drew on the CPRD. Articles were published between 2004 and 2012 and covered a range of study designs, geographical regions and age ranges.

Completeness of ethnicity data in secondary care was reported by eight studies as being between 65% and 95%. Missing ethnicity data was either considered as a distinct category for analysis, estimated using multiple imputation, collapsed together with the white ethnic groups or excluded from the analysis.

All five studies set in primary care made use of patient data across the UK. Of the 10 studies set in secondary care, five included patients from the whole of England, while the remainder selected patients from various regions of the country. In total, six studies incorporated ethnicity as a factor in the derivation and validation of risk prediction models,(114–119) three studies examined access to and use of secondary care services,(120,121) two examined in-patient procedures,(122,123) three studies focused on cancer,(124–126) and one on liver disease.(127)

Table 4.2 Key characteristics of studies included in literature review

#	Study	Title	Study population	Source of ethnicity data	% complete ethnicity	Ethnic categories for comparison	Treatment of Missing ethnicity
1	Cooper, 2009(124)	The influence of deprivation and ethnicity on the incidence of oesophageal cancer in England	4,252 oesophageal cancers in the West Midlands between 1977 and 2004	HES	Up to 89%	White, Asian, Black, Other/Mixed, Not known, Not HES matched	Treated as a category for analysis
2	Bottle, 2006(128)	Identifying patients at high risk of emergency hospital admissions: a logistic regression analysis	2,895,234 emergency admissions across England between 1999 and 2004	HES	Not reported	White, Black, Indian sub-continent, Chinese, Unknown, other	Treated as a category for analysis
3	Bragg, 2010(123)	Variation in rates of caesarean section among English NHS trusts after accounting for maternal and clinical risk: cross sectional study	620,604 women aged 15-44 with a singleton birth between January and December 2008	HES	89%	White, Afro-Caribbean, Asian, Other, Unknown	Treated as a category for analysis
4	Hacker, 2004(121)	Equity in waiting times for two surgical specialties: a case study at a hospital in the North West of England	4,306 waiting list patients in NW England in 2000/2001	HES	85%	White, Non-white	Excluded from analysis
5	Jack, 2007(125)	Testis and prostate cancer incidence in ethnic groups in South East England	194,590 male inpatients in SE England with diagnosed or suspected testis/prostate cancer	HES and Thames Cancer Registry	63% HES + 3% extra from TCR	Categories from the 1991 and 2001Census. White ethnic codes collapsed into a category for "All White"	Treated as a category for analysis
6	Mann, 2008(127)	Hepatitis C in ethnic minority populations in England	6,339 patients with hepatitis C related liver disease between 1997 and 2005	HES	65%	White, Black African, Black Caribbean, Pakistani, Bangladeshi, Other, Mixed, Indian, Chinese, Unknown	2 methods: 1. Unknown grouped with White 2. Unknown split between other ethnic groups
7	Mindell, 2008(122)	Using routine data to measure ethnic differentials in access to coronary revascularization	95,675 CHD and 19,282 coronary revascularization episodes for London residents between 2002 and 2004	HES	70%-95% from 2002-2004	17 Categories from the 2001 Census	Treated as a category for analysis
8	Billings, 2006(118)	Case finding for patients at risk of readmission to hospital: development of algorithm to identify high-risk patients	24,276 patients in England admitted to hospital between 1999 and 2003	HES	Not reported	White Black, Indian, Pakistani, Unknown/Not specified	Not reported
9	Shah, 2011(120)	Place of death and hospital care for children who died of cancer in England, 1999-2006	Children (<15 years) who were diagnosed with cancer and who died under the age of 20 years in England 1999–2006.	HES+CCLG	84%	White, South Asian, Black, Other, Missing/Unknown	Not reported

Table 4.2 Continued...

#	Study	Title	Study population	Source of ethnicity data	% complete ethnicity	Ethnic categories for comparison	Treatment of Missing ethnicity
10	Downing, 2011(126)	Using routinely collected health data to investigate the association between ethnicity and breast cancer incidence and survival: what is the impact of missing data and multiple ethnicities?	48,234 cases of female invasive breast cancer diagnosed in the Northern/Yorkshire and West Midlands cancer registries	HES	83%	White, South Asian, Black, Other, unknown	Filled in using multiple Imputation
11	Hippisley-Cox, 2008(119)	Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2	2.3 million Patients aged 35-74 in the QRESEARCH database between 1993 and 2008	QRESEARCH	27.1% women and 23.8% men	White/not recorded, South Asian, Black African, Black Caribbean, Chinese/Other	Unknown ethnicity grouped with White
12	Hippisley-Cox, 2010a(115)	Derivation, validation, and evaluation of a new QRISK model to estimate lifetime risk of cardiovascular disease: Cohort study using the QRESEARCH database	3,610,918 patients aged 30-80 from 563 general practices contributing to QRESEARCH database free from CVD and not taking statins between 1994 and 2010	QRESEARCH	Not reported	White/ not recorded, Indian, Pakistani, Bangladeshi, Other Asian, Black African, Black Caribbean, Chinese, Other/Mixed	Unknown ethnicity grouped with White
13	Hippisley-Cox, 2010b(114)	Predicting the risk of Chronic Kidney Disease in Men and Women in England and Wales: prospective derivation and external validation of the qkidney scores	2,363,069 patients aged 35-74 from general practices contributing to the QRESEARCH and THIN databases between 2002-2008	QRESEARCH and THIN	Not reported	White/ not recorded, Indian, Pakistani, Bangladeshi, Other Asian, Black Caribbean, Black African, Chinese, Other	Unknown ethnicity grouped with White
14	Collins, 2010(129)	An independent and external validation of QRISK2 cardiovascular disease risk score: A prospective open cohort study	1.58 million Patients 35-74 in the THIN database free from cardiovascular events between 1993 and 2008	THIN	Not reported	White/ not recorded, Indian, Pakistani, Bangladeshi, Other Asian, Black African, Black Caribbean, Chinese, other/mixed	Unknown ethnicity grouped with White
15	Collins, 2011(116)	External validation of QDSCORE (R) for predicting the 10-year risk of developing Type 2 diabetes	2.4 million Patients 25-79 in the THIN database free from diabetes between 1993 and 2008	THIN	Not reported	White/ not recorded, Indian, Pakistani, Bangladeshi, Other Asian, Black African, Black Caribbean, Chinese, other /mixed	Unknown ethnicity grouped with White

4.4.2 Study findings

All of the studies, bar one, reported a relationship between ethnicity and at least one outcome of interest. Studies focusing on cancer found that the incidences of oesophageal and testis cancer were highest in the White population while the incidence of prostate cancer was highest for ethnic minority men.(125,130) In a further study, the incidence of breast cancer was found to be lowest amongst South Asian women, though no ethnic differences in five-year survival were evident.(126)

Studies of secondary care usage found that ethnicity was a significant predictor of emergency hospital admission, or (any) readmission to hospital.(128) Patients of Black and South Asian ethnicity had lower use of coronary revascularisation surgery while Black and South Asian children with cancer were found to be less likely to die at home.(120,122) Only Hacker et al. found no ethnic differences in their study of waiting times to surgery.(121)

Studies of disease prevalence and risk uniformly reported increased risk in non-white groups, particularly South Asian groups for cardiovascular disease, chronic kidney disease and type 2 diabetes mellitus. Full results tables presented in table 4.3.

4.5 Discussion

Fifteen of the 1,232 (1.3%) published observational studies on national electronic health databases retrieved by the literature search utilized the patient-level ethnicity data available. The fact that all 15 were published after the financial incentivisation of ethnicity recording in primary care in 2004 may reflect increased confidence in data quality and completeness brought about by this top-down initiative. This is not to say that research into ethnicity using routine health records is not taking place at a local or regional level, but rather that for large-scale studies at a national level, this resource has yet to be capitalized on. As evidenced in several studies throughout the UK, primary care records are widely used to examine ethnicity at a regional level, particularly in large cities with greater ethnic minority populations and higher levels of deprivation relative to the national average.(2,131–135)

Table 4.3 Key findings of studies included in literature review

#	Study	Aims	What is known about ethnicity in this field	Main ethnicity related findings	Limitations of ethnicity data
1	Cooper, 2009(124)	To examine the role of sex, ethnicity, affluence and deprivation on the incidence of oesophageal cancer and to determine changes in the incidence of its morphological subtypes (EAC and ESCC) over time.	In the USA, ESCC is more common in the black population while EAC is most common among White males. A relationship between ethnicity and oesophageal cancer has not yet been studied in England.	EAC strongly associated with White ethnicity. ESCC more common black men compared to white men, but not statistically significant. The incidence of EAC is rising rapidly due to its strong association with male gender and White ethnicity. The incidence of ESCC has not changed over time.	None relating to ethnicity specified
2	Bottle, 2006(128)	To incorporate ethnicity as an explanatory variable into a predictive model to identify patients at high risk of emergency hospital admissions	Not stated	Ethnicity was found to be significantly associated with the likelihood of being a of high-impact user (someone with three or more emergency admissions within a 12 month period) and was included in the final regression model	None relating to ethnicity specified
3	Bragg, 2010(123)	To determine whether ethnicity, other maternal factors and clinical risk factors can explain variation in rates of caesarean section in England.	Not stated	Odds of having a C-section were significantly increased for Black African/Caribbean women compared to White independent of all other risk factors.	Incomplete and inaccurate coding of risk factors.
4	Hacker, 2004(121)	To examine equity in waiting times for secondary care by age, gender, ethnicity and deprivation	Patients from minority ethnic groups have been found to be less likely to receive the health services they require and are more likely to face discrimination in accessing services.	No significant differences in waiting time by ethnicity were found: Univariate analysis showed a non-significant increase in waiting only	Non-white ethnic group was small, even after combining ethnic minority groups
5	Jack, 2007(125)	To compare the incidence of prostate and testis cancer in different ethnic groups.	Studies in the US have shown that the incidence of testis cancer is five times higher in White compared to Black men and the incidence of prostate cancer is higher in Black men compared to White. Use of PSA testing is lower in Black men.	Incidence of testis cancer was significantly lower in all South Asian, Black African/Caribbean, and Chinese sub groups compared to the "All White" group. Incidence of prostate cancer significantly higher in Black, Indian, Pakistani, and Mixed White/Black groups and significantly lower in Bangladeshi and Chinese groups compared to White.	Residual confounding of deprivation may explain lower prostate cancer incidence in Bangladeshi men.
6	Mann, 2008(127)	To investigate ethnic differences in the prevalence of hepatitis-C related end-stage liver disease.	Ethnic minority groups may be less likely to be tested for Hepatitis C as they are not considered as being at high risk- they may be more likely to be admitted to hospital in end-stage disease or die without receiving successful treatment.	Crude morbidity and mortality rates for end-stage liver disease were higher for Black African/Caribbean, Pakistani, Bangladeshi and Other groups compared to White and lower for Indian, Chinese and Mixed groups. After age standardisation, all non-white groups had significantly higher rates of end-stage liver disease, death, and hepatocellular carcinoma compared to the White group.	Higher rates of liver disease in ethnic minority groups may be explained by higher prevalence of hepatitis C in host countries for new migrants.

Table 4.3 continued

7	Mindell, 2008(122)	To examine ethnic differences in coronary revascularization.	South Asian patients have been found to be less likely to received angioplasty or coronary bypass graft than White patients despite having equivalent levels of recommendations for revascularization.	Proportional ratios showed that revascularizations were less common in Bangladeshi and Black African/Caribbean groups relative to need (number of CHD admissions) - particularly for those with emergency admissions.	>20% of CHD patients did not have ethnicity coded in 2003/4. Grouping of South Asian disguises known heterogeneity in morbidity/mortality.
8	Billings, 2006(118)	To develop an algorithm to identify patients at high risk of readmission to hospital within 12 months.	Not stated	Ethnicity was found to be a significant predictor of readmission and was included in the final regression model.	None relating to ethnicity specified
9	Shah, 2011(120)	To describe patterns in hospital care and evaluate factors affecting place of death for children with cancer.	Not stated	Significant ethnic differences in place of death: 42% of White and 70% of Asian and Black children died in hospital. 79% of children who died in hospice were White. Odds of dying at home significantly reduced for South Asian and Black children after adjustment.	None relating to ethnicity specified
10	Downing, 2011(126)	To examine the relationship between ethnicity and breast cancer incidence/survival and to assess the impact of missing data/multiple ethnicities	In the UK, incidence of breast cancer has been found to be lower for ethnic minority groups than for White groups. Breast cancer survival has been found to be better for South Asian women, but worse for Black African/Caribbean women.	Incidence of breast cancer was similar for White and Black groups and significantly lower for South Asian groups. After adjustment for age, gender and stage of diagnosis, no ethnic differences in 5-year survival were found.	Study was unable to split ethnic groups into minor categories due to small numbers. Difficult to interpret results for "other" group due to heterogeneity.
11	Hippisley-Cox, 2008(119)	To revise an algorithm to predict the ten year risk of developing cardiovascular disease by incorporating ethnicity and relevant disease conditions (QRISK2).	Rates of CVD are known to vary by between ethnic groups. Existing CVD risk scores which have been prospectively derived do not include a variable for self-assigned ethnicity. NICE guidance recommends multiplying Framingham risk score by 1.4 for men, but this does not take into account heterogeneity between South Asian populations.	Incidence of CVD highest for South Asian groups and lowest for Black African/Caribbean and Chinese individuals. Prevalence of risk factors varied substantially between ethnic groups. The new risk model improves upon the older version and increases identification of ethnic minority groups at high risk.	Only 25% of patients had self-assigned ethnicity. Misclassification of patients with missing ethnicity would result in an underestimate of the effect of ethnicity on CVD risk.
12	Hippisley-Cox, 2010a(115)	To further refine the QRISK2 score so that it may be used across all age groups.	Not stated	South Asian patients were most likely to have a high lifetime risk of CVD compared to White- this was highest for Pakistani, followed by Bangladeshi and Indian groups. The new score will increase identification of high CVD risk amongst younger patients, men, ethnic minority groups and people with a family history of premature CHD.	None relating to ethnicity specified

Table 4.3 continued

13	Hippisley-Cox, 2010b(114)	To develop two risk scores to predict the 5 year risk of developing moderate/severe kidney disease and the 5 year risk of End Stage Kidney Failure.	Ethnicity is known to be associated with chronic kidney disease and thus relevant for inclusion in the risk scoring algorithm.	Risk of moderate/severe CKD was significantly increased for South Asian patients compared to the White/not recorded group; Black African/Caribbean women had the lowest risk. Risk of End Stage Kidney Failure was significantly increased for Pakistani, other Asian and Chinese women relative to White women and for Pakistani, Other Asian and Black Caribbean men.	None relating to ethnicity specified
14	Collins, 2010(129)	To evaluate the QRISK2 score in an independent UK cohort and compare it to QRISK1 and the modified Framingham equation.	Risk of CVD is known to vary between ethnic groups. To date, the choice of adjustment factor for the Framingham score has not been scientifically evaluated.	Incidence of CVD was highest in South Asian groups- particularly Pakistani. The performance of QRISK was superior to the modified Framingham equation.	None relating to ethnicity specified
15	Collins, 2011(116)	To validate the QDSCORE which estimates 10 year risk of developing diagnosed Type 2 diabetes.	Ethnic minority groups have an increased risk for type 2 diabetes. The prevalence of T2DM is 2-4x higher amongst South Asian groups and has an earlier onset than in the White European population.	Incidence of T2DM was highest for South Asian groups- particularly Bangladeshi. Large proportions of South Asians had a predicted risk of 10% or higher of developing T2DM in the next 10 years.	None relating to ethnicity specified

4.6 Rationale for study examining completeness and quality of ethnicity recording

There have been concerns that research on ethnicity and health has been of poor quality, with some researchers failing to explicitly report how their ethnic categories have been derived, which concepts they are intended to represent and what role they expect ethnicity to play in the phenomenon under study.(136,137)

The literature review identified several studies which both utilise datasets with high levels of ethnicity data completeness and also exemplify how electronic health records can be used to examine important ethnic differentials in usage of and access to healthcare services, prevalence and incidence of disease, and risk of outcomes, such as morbidity, mortality and hospital re-admission. Improving the quality and usability of the ethnicity data is therefore a necessary step towards increasing the use of these data for research into ethnic inequalities, and improving the validity and generalisability of the findings.

The UK differs from the US in that there is no legal requirement for the inclusion of ethnic minority groups in publicly funded research which can provide the evidence base for guidance and policy tailored to ethnically diverse populations.(8) Furthermore, as demonstrated by the small number of papers identified in the systematic review, much of this research does not consider ethnicity as an explanatory variable of interest, and thus cannot investigate otherwise invisible inequalities which may exist.(138)

One topic which is yet to be explored is the potential value of routinely collected healthcare data for examining ethnic disparities in healthcare usage and outcomes across whole populations. The computerisation of healthcare records across both primary care (general practice) and secondary care (hospital settings) has generated enormous potential for population-based research on morbidity and the use of health services.

Chapter 5
Completeness and usability of ethnicity data in UK-based primary care and hospital databases

5.1 Summary

This chapter describes the first study of this thesis – an audit of the quality, completeness and generalisability of routinely recorded ethnicity data in the Clinical Practice Research Database (CPRD) and the Hospital Episode Statistics (HES). A modified version of this chapter was published as a peer-reviewed article in the *Journal of Public Health* in 2013.(139)

5.2 Aims, objectives and research questions

Although the validity of morbidity indicators in both the CPRD and HES has been explored in depth, to date, no comprehensive audit of ethnicity data has been conducted.(85,140–143) Furthermore, even though improvements in ethnicity recording have been demonstrated nationally, very few epidemiological studies using these databases have utilised the patient-level ethnicity data held within.(139)

Confirming that the ethnicity data recorded in routine health databases are of sufficient quality to explore ethnic differentials in health is an essential first step towards maximising their use for research and clinical purposes. The aim of this study was to ascertain the completeness and validity of routinely collected ethnicity data held in electronic health databases and to demonstrate the feasibility of using such data to investigate ethnic differences in disease patterning and outcomes across the UK.

The objectives of this study were fourfold:

1. to chart progress in the completeness of ethnicity recording in the CPRD and HES over time
2. to assess consistency of ethnicity coding for individuals whose ethnicity is recorded multiple times

3. to develop a simple and pragmatic method for choosing an ethnic category for patients with multiple discrepant ethnicity codes either within the CPRD alone or across linked CPRD and HES

4. to determine whether the ethnic breakdown of the CPRD population is consistent with that of the 2011 UK Census.

5.3 Methods

5.3.1 Ethnicity data extraction

The 16 categories used for ethnicity data collection across the NHS were standardised to the five high-level groups used in the 2001 Census for England and Wales, allowing the ethnicity data in CPRD and HES to be compared with each other and with other sources. The 16 categories were not used in this analysis because the categorisation of ethnicity in the HES inpatient dataset prior to 2001 did not separate out the 'White' category into British, Irish and 'other White'.

Table 5.1 Mapping of 2001 census categories to high-level groups

16 groups	5 categories
1 British	1.White
2 Irish	
3 Any other White background (write in)	
4 White and Black Caribbean	2. Mixed
5 White and Black African	
6 White and Asian	
7 Any other mixed background (write in)	
8 Indian	3. Asian or Asian British
9 Pakistani	
10 Bangladeshi	
11 Any other Asian background (write in)	
12 Caribbean	4. Black or Black British
13 African	
14 Any other Black background (write in)	
15 Chinese	5. Chinese or Other Group
16 Any other ethnic group (write in)	

a) CPRD data

Information in the CPRD was entered using the Read system of alphanumeric codes.(88) Read codes for 'ethnic group' falling under the 9i (2001 Census) and 9S (1991 Census) hierarchies were extracted for all currently registered and past patients contributing to the July 2012 build of the CPRD database. All ethnic codes were collapsed into the 16 categories used in the 2001 Census (Table 5.1). Ethnicity codes were merged with the patient denominator file for July 2012. All patients registered in CPRD practices up to and including 31 December 2011 were included in the analysis. Usable ethnicity was considered to be any 9i or 9S Read code which was not 'unknown' (9SD, 9SE, 9SZ, 9ig), at too high a level to be interpreted (9i, 9S), or missing. Identical ethnicity codes entered on the same system date for a single patient were removed, as these yield no additional information.

b) HES data

In August 2012, all ethnicity and demographic data for patients contributing to the Hospital Episode Statistics between the 1997/98 and 2011/12 financial years was extracted. The two coding hierarchies relating to the 1991 and 2001 Census were collapsed into the 16 ethnic categories of the 2001 Census. Usable ethnicity was considered to be any ethnic code which was not 'unknown' (ethnos variable: codes 9, X, Z, Zn), or missing. Since the unique patient identifier "hesid" has only been used since 1997 onwards, inpatient data for the financial years 1997/98–2011/12 were included in the analysis. Outpatient data were available from 2003/04 onwards and A&E data from 2007/08 onwards (Table 5.2).

Table 5.2 Categorisation of ethnicity in the Hospital Episode Statistics for England

Inpatient 1995-2000	Inpatient 2001 onwards	Outpatients 2003 onwards and A&E 2008 onwards
0 White	A British (White)	An = British (White)
1 Black Caribbean	B Irish (White)	Bn = Irish (White)
2 Black African	C Any other White background	Cn = Any other White background
3 Black Other	D White and Black Caribbean (Mixed)	Dn = White and Black Caribbean (Mixed)
4 Indian	E White and Black African (Mixed)	En = White and Black African (Mixed)
5 Pakistani	F White and Asian (Mixed)	Fn = White and Asian (Mixed)
6 Bangladeshi	G Any other Mixed background	Gn = Any other Mixed background
7 Chinese	H Indian (Asian or Asian British)	Hn = Indian (Asian or Asian British)
8 Any other ethnic group	J Pakistani (Asian or Asian British)	Jn = Pakistani (Asian or Asian British)
9 Not given	K Bangladeshi (Asian or Asian British)	Kn = Bangladeshi (Asian or Asian British)
X Not known	L Any other Asian background	Ln = Any other Asian background
	M Caribbean (Black or Black British)	Mn = Caribbean (Black or Black British)
	N African (Black or Black British)	Nn = African (Black or Black British)
	P Any other Black background	Pn = Any other Black background
	R Chinese (other ethnic group)	Rn = Chinese (other ethnic group)
	S Any other ethnic group	Sn = Any other ethnic group
	Z Not stated	Zn = Not stated
	X Not known	X = Not known

5.3.2 Statistical analysis

a) Overall completeness of ethnicity recording

For both the CPRD and HES, the proportion of patients with ethnicity ever recorded was calculated. For the CPRD, completeness was compared between a) all patients present in the CPRD including those who have left or died, b) currently registered patients (that is, all patients who have not died or transferred out of their general practice), and c) patients registered after April 1st 2006 when incentivisation of ethnicity recording was introduced to primary care. For HES, completeness was assessed for all inpatients, outpatients and A&E attendees.

In the CPRD, ethnicity recording was further broken down by year of first ever registration. For HES, ethnicity recording was further broken down by year of first ever admission (inpatients and A&E) or appointment (outpatients).

b) Multiple ethnicity recording within each source

In both general practice and in hospital, patients can have their ethnicity recorded repeatedly over multiple consultations or visits. Discrepancies may arise if there are: mistakes while entering the data onto the system, different responses from the patient when asked by the service provider, or different codings of ethnicity information without consulting the patient.

In order to examine consistency of ethnicity coding, the proportion of patients with only one ethnicity code on their record was compared with the proportion of patients with multiple codes where these were:

1. truly matched (multiple ethnic codes which are identical)
2. categorically matched (multiple ethnic codes which are different but fall into the same five higher-level groups of ethnicity, namely White, Mixed, Asian/Asian British, Black/Black British, Chinese/Other)
3. truly mismatched (multiple ethnicities which span different higher-level groups).

5.3.3 Discrepant ethnicity recording between linked sources

Of the 624 general practices contributing to the CPRD in August 2013, 357 consented to further linkage with HES. Linkage using deterministic matching on NHS number, date of birth and gender was undertaken by a Trusted Third Party. The completeness of ethnicity recording in each database alone was compared with the completeness of the databases combined. For the subset of patients registered from April 1st 2006 onwards with a valid ethnicity recorded in *both* CPRD and HES, the most commonly recorded ethnicity code in each database was determined separately, and compared them to determine the proportion of patients with matched or mismatched ethnicity across databases. The degree of mismatch was further examined for each ethnic group in turn.

5.3.4 Comparison of the CPRD population with the 2011 UK Census population

The most recent census across the UK was undertaken on March 27th, 2011 with initial aggregate data released in September 2012. Ethnic breakdowns for the populations of England, Wales, Scotland and Northern Ireland were obtained from the relevant census

websites. Since the available categories for the ethnicity question vary slightly between the censuses for the constituent countries, categories were collapsed for comparison with the CPRD data.

For comparison between the Census and the CPRD, two populations were derived from the January 2012 build of the CPRD. For the first population, the January 2012 database was reduced to include only patients who were actively registered on the date of the Census (March 27th, 2011). All patients who had left the CPRD before this date or who had joined the CPRD after this date were removed. For patients who were present in the CPRD on March 27th, 2011, all ethnicity codes entered after this date were removed.

The second population was restricted to patients who registered after April 1st 2006 and were present in the CPRD March 27th 2011 population in order to examine distributions and consistency with other sources for a group for whom it is expected that ethnicity will have been recorded (and recording was incentivised). For both populations, the most recent ethnicity code prior to the census date was collapsed into the 16 categories and the 5 higher-level categories of the 2011 Census for analysis. The proportions of patients belonging to each ethnic group in the CPRD were then directly standardised against the age distribution of the 2011 Census.

5.4 Results

5.4.1 Overall completeness of ethnicity recording

The first objective of this study was to chart progress in the completeness of ethnicity recording in the CPRD and HES over time.

a) CPRD population

From a total of 12,099,672 patients contributing to the July 2012 build of the CPRD, (including patients who have died or left the practice), 3,544,589 patients (29.3%) had at least one Read code for ethnicity, including unusable codes (not stated, not known, high level). The proportion

of patients with at least one usable ethnicity code recorded ranged from 27% (n=3,282,739) for the whole of CPRD to 78% for patients registered from 2006 onwards (n=1,723,195) (Table 5.3).

Figure 5.1 illustrates the low levels of ethnicity completeness for new patient registrations between 1995, when ethnicity recording was still new, and 2006, when ethnicity recording became financially incentivised under the Quality and Outcomes Framework. The completeness of ethnicity recording approaches 90% in 2010 across general practices in the CPRD.

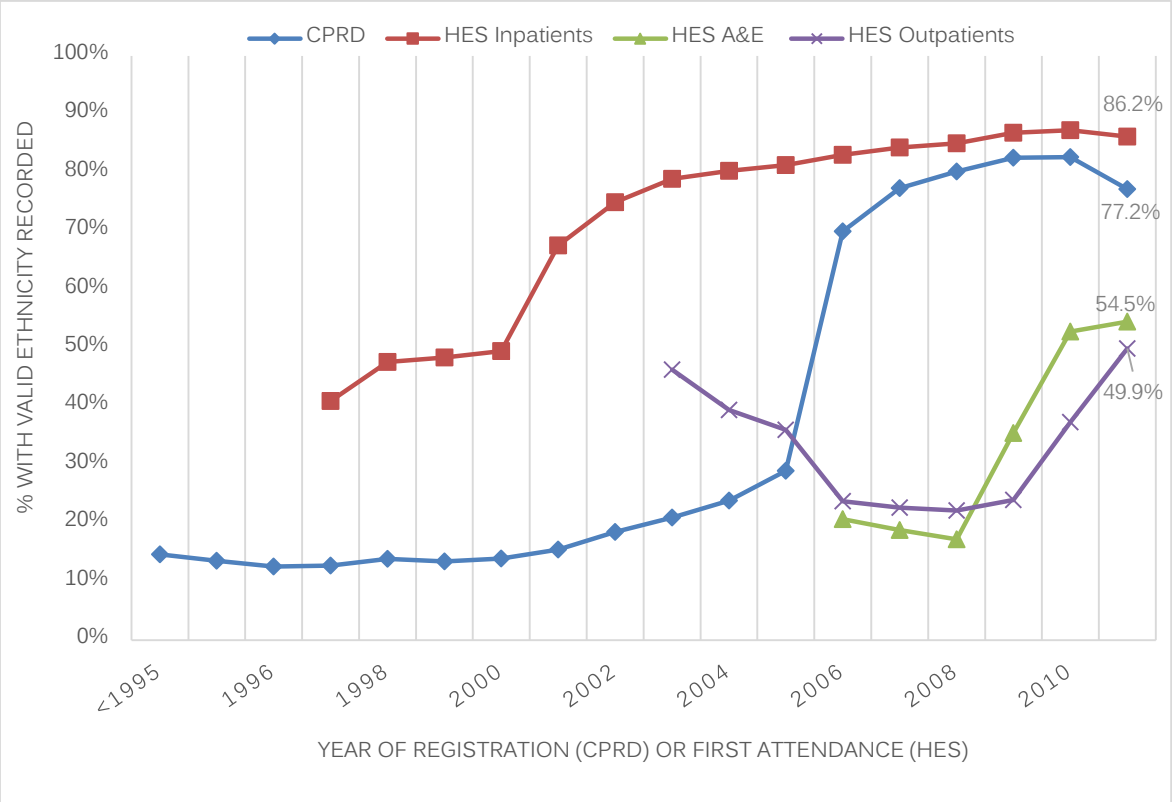


Figure 5.1 Proportion of patients with valid ethnicity recorded by financial year of registration (CPRD) or first attendance (HES)

Table 5.3 Overall completeness of ethnicity recording in CPRD (July 2012)

	All acceptable patients		Currently registered		Registered April 1st 2006 onwards	
	N	%	N	%	N	%
Total	12,099,672	100.0%	5,308,411	100.0%	2,201,065	100.0%
Any ethnicity recorded (including not stated, not known)	3,544,589	29.3%	2,605,232	49.1%	1,874,916	85.2%
% with usable ethnicity recorded (excluding not stated, not known)	3,282,739	27.1%	2,423,438	45.7%	1,723,195	78.3%
% with only 1 ethnicity on their record	2,802,284	23.2%	2,038,097	38.4%	1,481,112	67.3%
% with multiple ethnicities on their record	480,455	4.0%	385,341	7.3%	242,083	11.0%
% with multiple ethnicities which are truly identical	379,591	3.1%	306,771	5.8%	180,455	8.2%
% with multiple ethnicities which are categorically identical	62,413	0.5%	49,144	0.9%	35,208	1.6%
% with truly discrepant ethnicity	38,523	0.3%	39,426	0.7%	26,420	1.2%
% with no usable ethnicity on their record	8,816,933	72.9%	2,854,973	53.8%	486,870	22.1%

Table 5.4 Overall completeness of ethnicity recording in HES (April 2012)

	Inpatients		Outpatients		A&E	
	N	%	N	%	N	%
Total	51,965,028	100%	48,549,620	100%	31,860,530	100%
Any ethnicity recorded (including not stated, not known)	51,965,028	100%	48,549,620	100%	31,860,530	100%
% with usable ethnicity recorded (excluding not stated, not known)	41,281,350	79.4%	17,696,595	36.5%	8,531,890	26.8%
% with only 1 ethnicity on their record	16,354,201	31.5%	4,608,411	9.5%	5,860,016	18.4%
% with multiple ethnicities on their record	24,927,146	48.0%	13,088,173	27.0%	2,671,874	8.4%
% with multiple ethnicities which are truly identical	22,883,676	44.0%	2,589,948	5.3%	2,620,117	8.2%
% with multiple ethnicities which are categorically identical	652,246	1.3%	34,697	0.1%	7,740	0.2%
% with truly discrepant ethnicity	1,391,227	2.7%	10,463,828	21.6%	44,287	0.1%
% with no usable ethnicity on their record	10,683,678	20.6%	30,853,025	63.5%	23,328,640	73.2%

Considering differences in the completeness of ethnicity recording by gender and age, we find no notable differences by gender (Figure 5.2). By age, ethnicity recording is consistently highest for patients aged 40–79 in all years. Recording is markedly lower for patients aged 80 and over or 19 and under, though this gap diminishes over time (Figure 5.3).

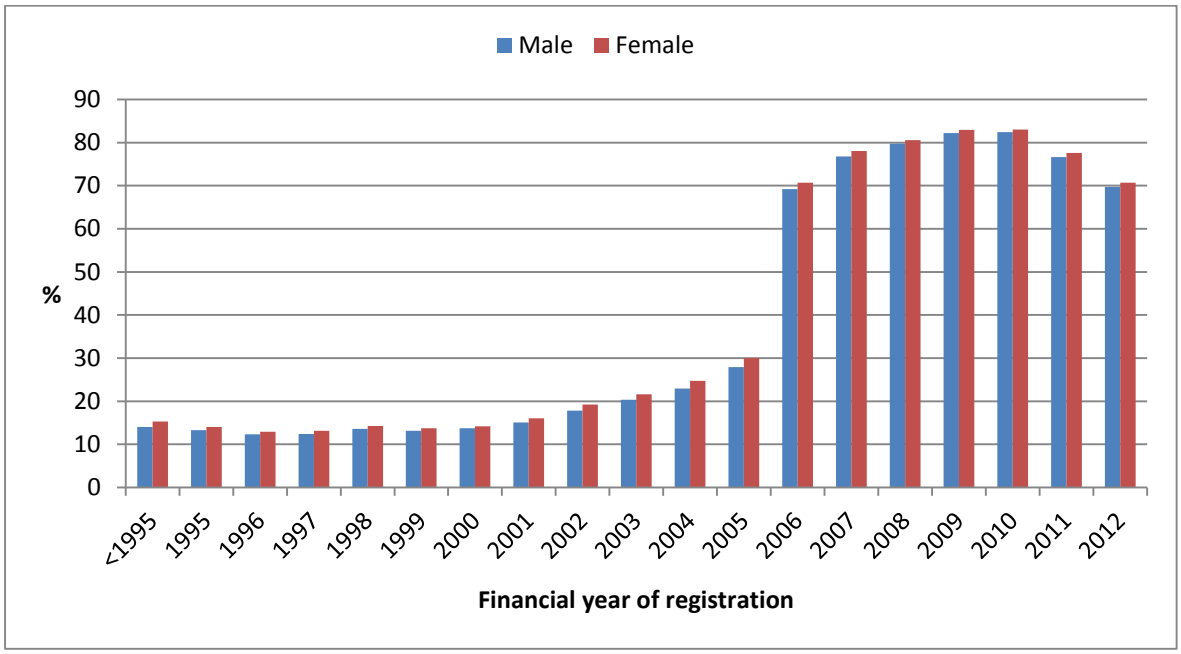


Figure 5.2 Proportion of patients in CPRD with any ethnicity code recorded in the same financial year as first registration with the general practitioner by gender

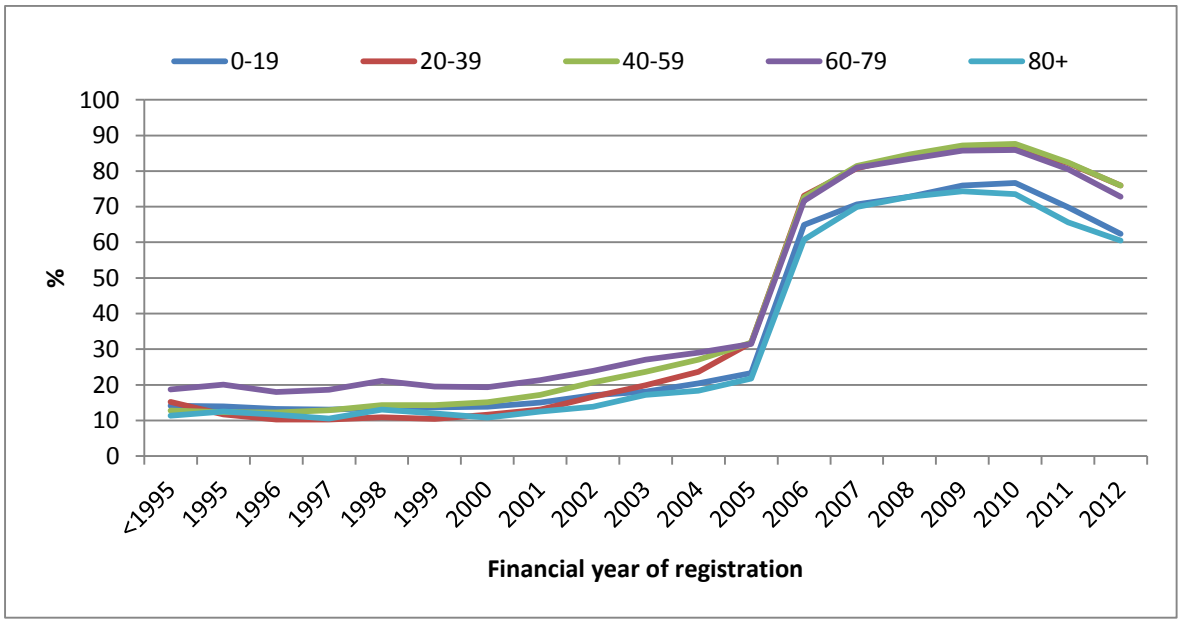


Figure 5.3 Proportion of patients in CPRD with any ethnicity code recorded in the same year as first registration with the general practitioner by age at registration

b) HES population

All patients in the HES Inpatient, A&E, and Outpatient databases have an ethnicity code attached to every attendance or episode. As such, the completeness of ethnicity recording, including unusable codes, is 100% across HES, though the proportion of patients with usable ethnicity ever recorded varies significantly between the three sub-sets (Table 5.4).

I. HES inpatient population

A total of 51,965,028 patients contributing 223,451,171 inpatient episodes across 16 years were available for analysis. 79.4% (n=41,281,350) of patients had at least one code which was usable. The proportion of patients with usable ethnicity recorded improved from 41% for patients who were first admitted in 1997, to 86% for patients who were first admitted in 2011 (Figure 5.1).

II. HES accident and emergency population

A total of 31,860,530 patients contributed 73,085,977 A&E visits across five years. 26.8% (n=8,531,890) of these patients had a code which was usable. The proportion of patients with usable ethnicity recorded improved from 20% for patients who were first admitted in 2008 to 53% in for patients who were first admitted in 2011 (Figure 5.1).

III. HES Outpatient Population

In total, 48,549,620 patients contributed 574,625,389 outpatient appointments over nine years. 36.5% (n=17,696,595) of patients had a code which was usable. The proportion of patients with usable ethnicity recorded did not improve over time, changing only from 46% in 2003 to 50% in 2011, though in the intervening years completeness dropped to a low of 22% (Figure 5.1).

5.4.2 Multiple ethnicity recording within each source

The second objective of this study was to assess consistency of ethnicity coding for individuals whose ethnicity is recorded multiple times, both within individual databases and across linked databases.

a) CPRD Population

Within the whole of CPRD (N=12,099,672), 4.3% of patients had had their ethnicity recorded multiple times. This increased to 7.3% for patients who were currently registered (n=5,308,411), and to 11.0% for patients registered from 2006 onwards (n=2,201,065). Amongst patients with multiple ethnicity codes on their record, the proportion of individuals whose ethnicity codes were either truly identical (at the 16-group level) or categorically identical (at the 5-category level) was consistently high, ranging from 89.1% for the 2006+ population, to 92.0% for the currently registered and total population (Table 5.3). Figure 5.4 illustrates the flow of data completeness and consistency in CPRD. Truly discrepant ethnicity is a problem for 24,253/11,801,879 patients (0.2%). Missing and not stated ethnicity is evident for 8,722,182 patients (73.9%).

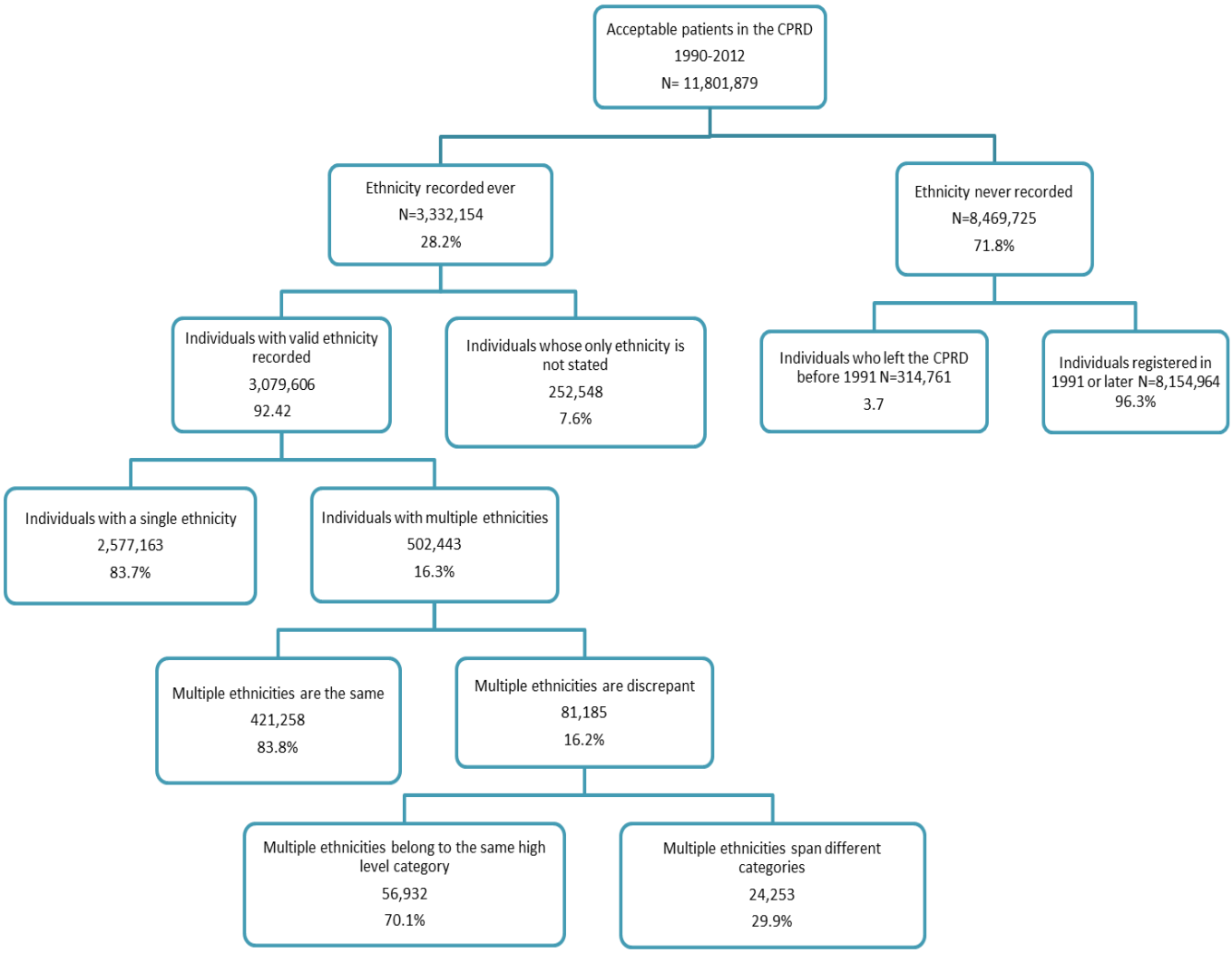


Figure 5.4 Ethnicity recording for all patients in the CPRD

b) HES Population

Multiple ethnicities were recorded for 48% of inpatients, 27% of outpatients and 8.4% of A&E patients. Amongst patients with multiple ethnicity codes on their record, ethnicity codes were either truly identical or categorically identical for 94.4% of inpatients and 98.4% of A&E patients. However, for outpatients, the comparable figure was only 20.1%, indicating that the majority of those with multiple ethnicity codes could not reliably be classified into a single ethnic group (Table 5.4).

I. HES Inpatients

Figure 5.5 illustrates the flow of data completeness and consistency in the HES inpatient database. Truly discrepant ethnicity is evident for 1,391,227 patients (2.7%). Missing and not stated ethnicity is evident for 10,683,687 patients (20.6%).

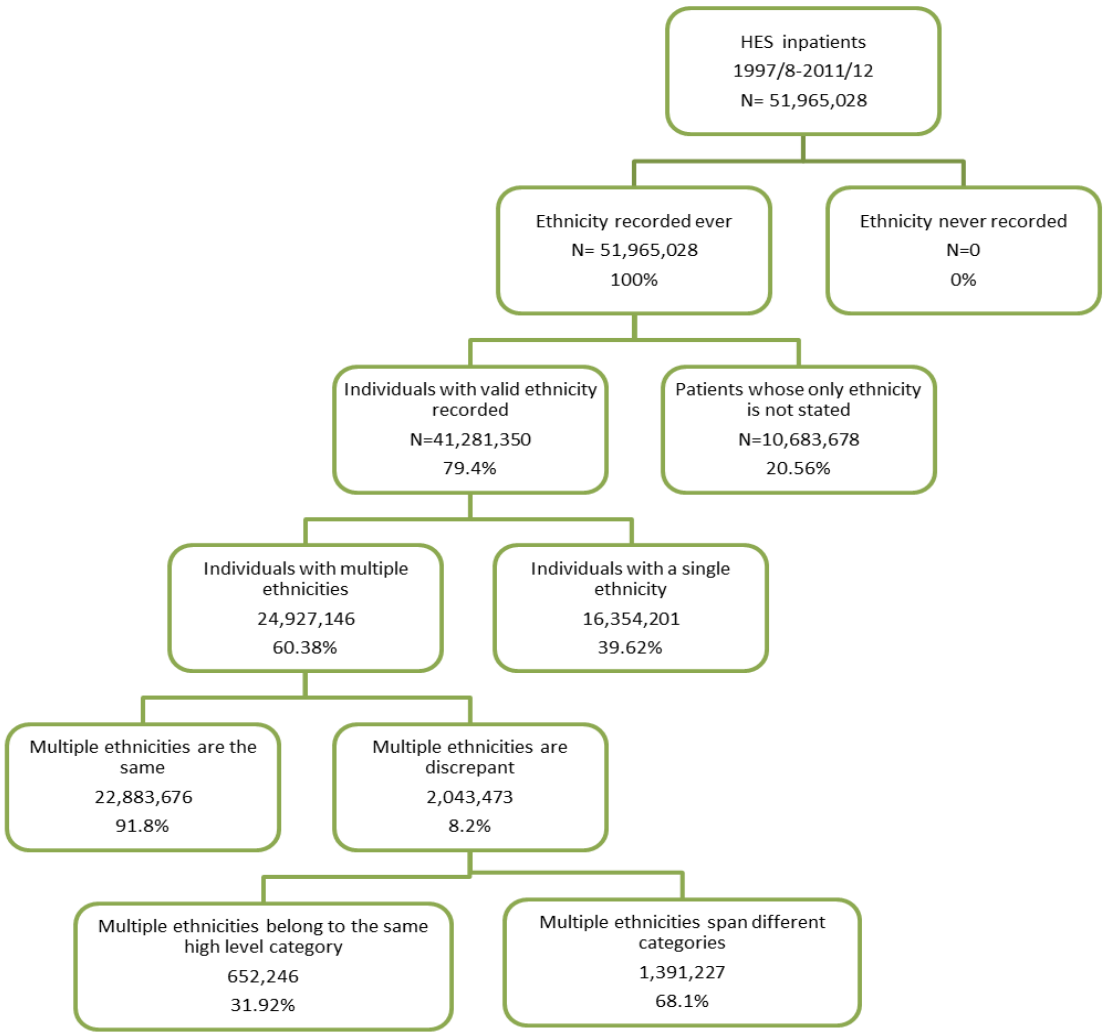


Figure 5.5 Ethnicity recording for all HES Inpatients

II. HES Outpatients

Figure 5.6 illustrates the flow of data completeness and consistency in the HES inpatients database. Truly discrepant ethnicity is evident for 10,463,828/48,549,620 (21.5%). Unknown ethnicity is evident for 30,853,025/48,549,620 patients (63.5%)

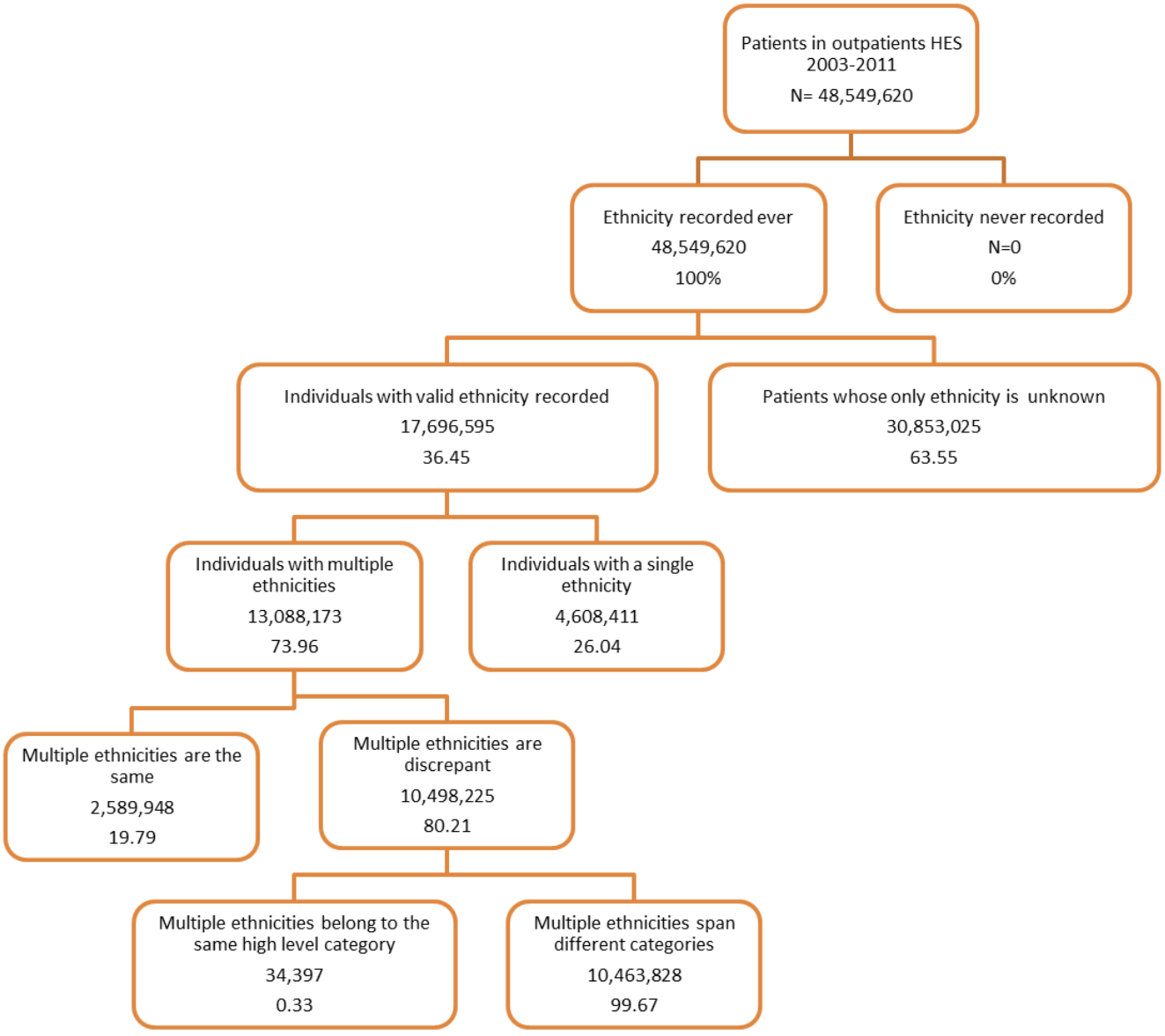


Figure 5.6 Ethnicity recording for all HES Outpatients

II. HES Accident and Emergency

Figure 5.7 illustrates the flow of data completeness and consistency in the HES A&E database. Truly discrepant ethnicity is evident for 44,287/31,860,530 patients (0.2%). Unknown ethnicity is a problem for 23,328,640 patients (7.0%).

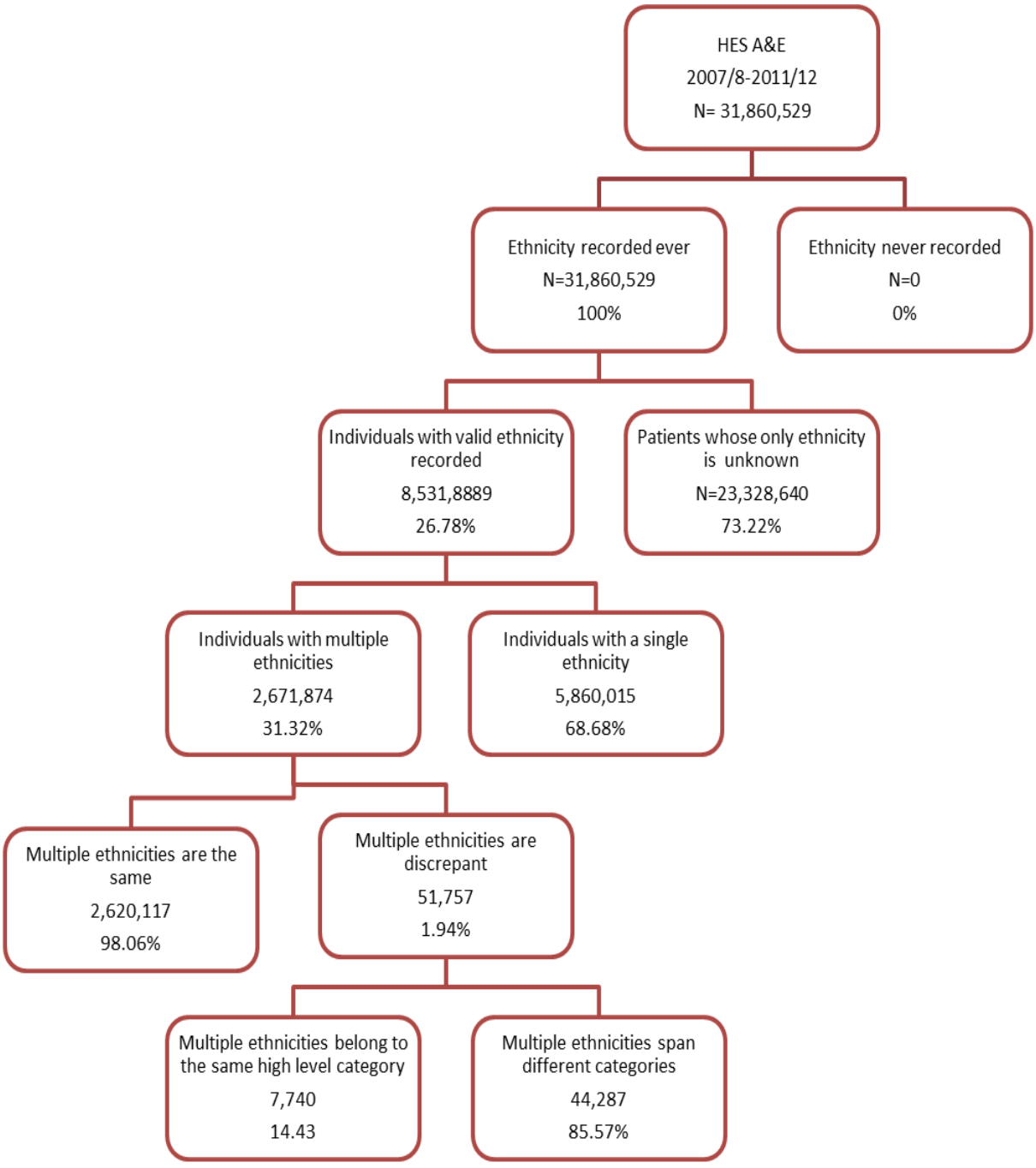


Figure 5.7 Ethnicity recording for all HES A&E Patients

5.4.3 Missing and discrepant ethnicity in linked sources

In the July 2012 build of the CPRD database, 3,899,648 patients also had linked HES data available. Completeness of valid ethnicity was 35.2% for CPRD, 86.4 for HES and 88.9% for the combined database. For the 827,753 patients registered from April 1st 2006 onwards, completeness was 78.7% for CPRD, 86.3% for HES and 97.1% for the combined database.

To examine matched and mismatched ethnicity between linked CPRD and HES, the analysis was restricted to patients registered from 2006 onwards in order to maximise the proportion of patients with a valid ethnicity code in both databases. From the 827,753 patients registered from April 1st 2006 onwards, 561,502 (67.8%) met this criterion.

When comparing the most commonly recorded ethnicity, 72.7% of patients (408,046/561,502) had a matching ethnicity code which belonged to the same 16-level category in both databases. This proportion increased to 85.0% (477,364/561,502) when collapsing the 16 categories into 5 groups (Table 5.5).

Table 5.5 Comparison of matched ethnicity between CPRD and HES using 16 vs. 5 categories (N=561,502)

Most common ethnic group in CPRD and HES	16 categories		5 categories	
	N	%	N	%
Mismatched	153,456	27.3	84,148	15.0
Matched	408,046	72.7	477,364	85.0

When using matrices to compare how individuals have had their ethnicity recoded in both datasets, the direction of mismatch was similar going from CPRD→ HES and from HES→CPRD. Over 96% of individuals coded as White in one database were White in the linked database (453,244/470,535). Almost half of all individuals coded as South Asian in CPRD were South Asian in HES (17,636/35,708). Going in the other direction, 71% of South Asians in HES were coded the same in CPRD (17,636/25,011). Patients coded as Black in CPRD were most commonly coded as Mixed in HES. Individuals coded as Mixed in CPRD were most commonly White in HES; however, most individuals coded as Mixed in HES were coded as Black in CPRD. Individuals coded as

Other in CPRD were most commonly coded as South Asian or Other in HES; most individuals coded as Other in HES were coded as White in CPRD (Table 5.6).

Table 5.8 Proportion of patients with matching ethnicity in linked CPRD and HES (N=561,502)

	Most common HES Ethnic Group						
Most common CPRD Ethnic Group	White	South Asian	Black	Other	Mixed	Equally common	Total
White	453,244	1,294	2,082	9,549	2,095	2,271	470,535
Row %	96.33	0.28	0.44	2.03	0.45	0.48	100
Column%	97.66	5.17	30.76	45.47	5.36	40.88	83.8
South Asian	1,545	17,636	498	3,256	11,619	1,154	35,708
Row %	4.33	49.39	1.39	9.12	32.54	3.23	100
Column%	0.33	70.51	7.36	15.5	29.74	20.77	6.36
Black	1,447	452	1,136	1,645	21,873	822	27,375
Row %	5.29	1.65	4.15	6.01	79.9	3	100
Column%	0.31	1.81	16.78	7.83	55.99	14.8	4.88
Other	2,804	4,192	271	3,392	677	495	11,831
Row %	23.7	35.43	2.29	28.67	5.72	4.18	100
Column%	0.6	16.76	4	16.15	1.73	8.91	2.11
Mixed	3,487	770	2,418	2,237	1,754	621	11,287
Row %	30.89	6.82	21.42	19.82	15.54	5.5	100
Column%	0.75	3.08	35.73	10.65	4.49	11.18	2.01
Equally common	1,572	667	363	923	1,049	192	4,766
Row %	32.98	13.99	7.62	19.37	22.01	4.03	100
Column%	0.34	2.67	5.36	4.39	2.69	3.46	0.85
Total	464,099	25,011	6,768	21,002	39,067	5,555	561,502
Row %	82.65	4.45	1.21	3.74	6.96	0.99	100
Column%	100	100	100	100	100	100	100

5.4.4 Developing a pragmatic method of identifying ethnicity in the CPRD

The third objective of this study was to develop a simple and pragmatic method for assigning ethnicity for patients with multiple discrepant ethnicity codes either within the CPRD alone or across linked CPRD and HES.

Based on the results preceding, an algorithm for assigning an individual a single ethnic category based on their CPRD record and utilising HES ethnicity when available has been developed. The details of the algorithm are displayed in figure 5.8.

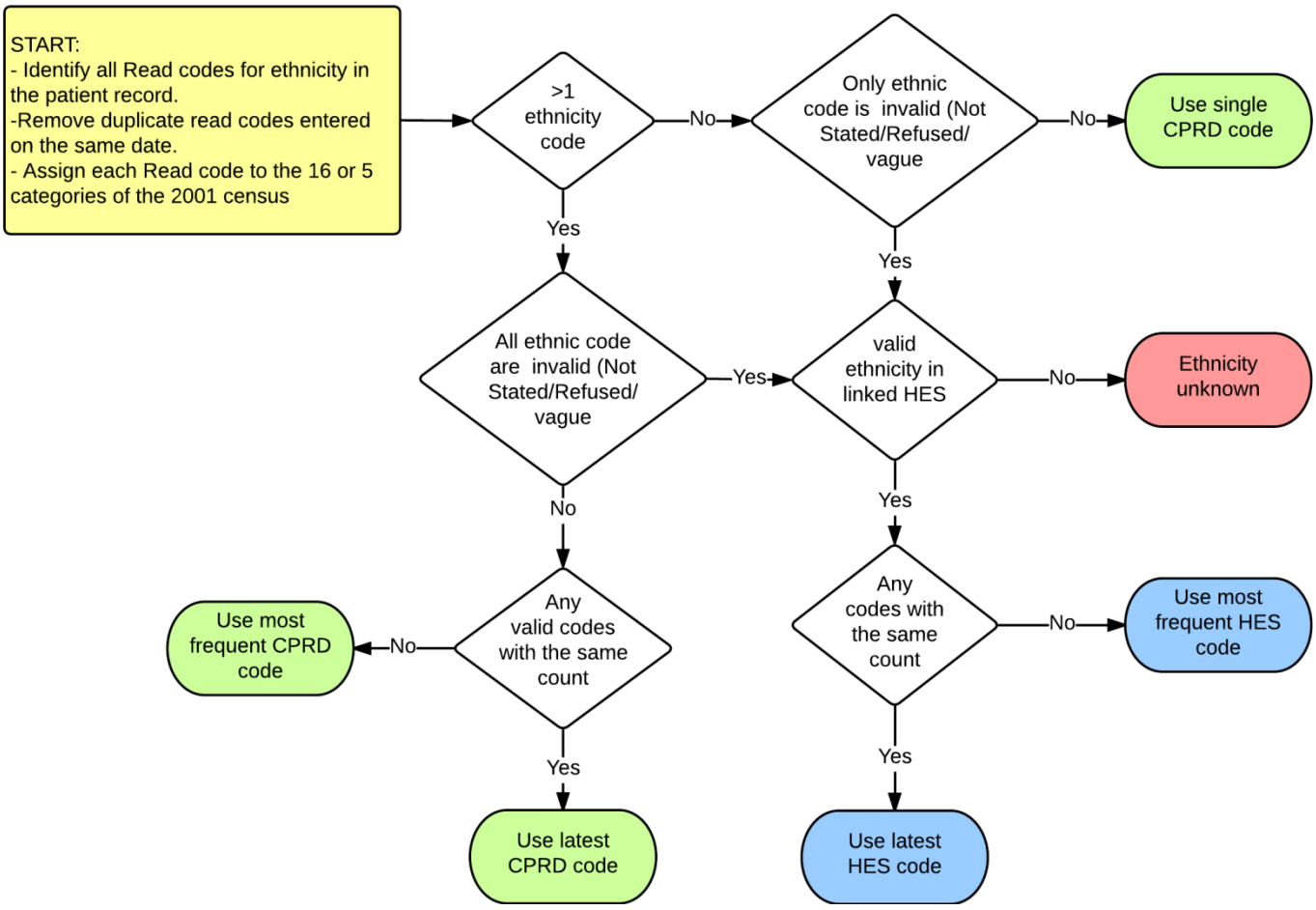
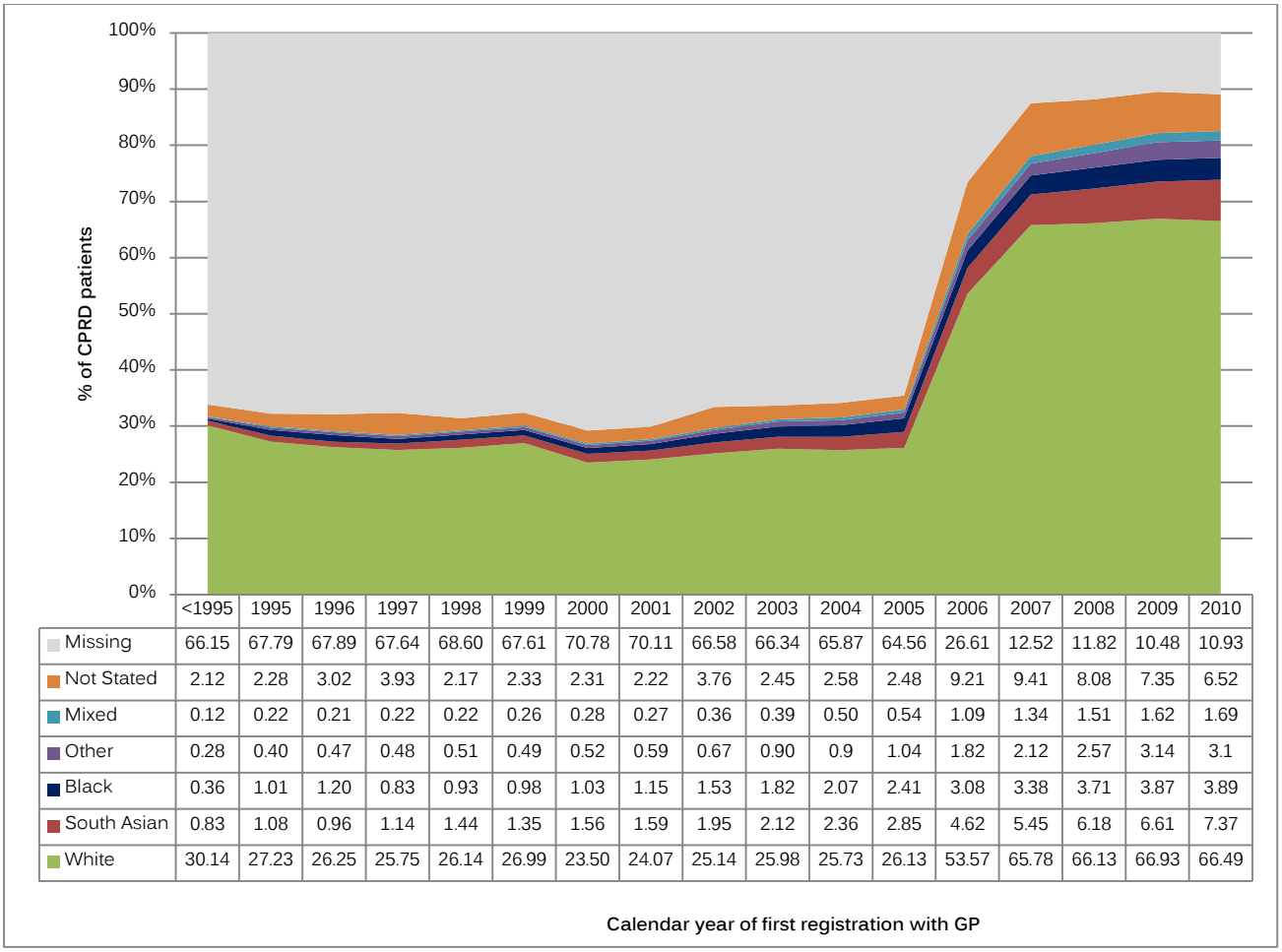


Figure 5.8 Pragmatic classification of ethnicity for patients in the CPRD

Estimates of the ethnic population in the CPRD using this method are shown in figure 5.9. Between 1995 and 2010, the proportion of patients with missing ethnicity in their first year of registration dropped from 66.15% to 10.93%. This was accompanied by a commensurate increase in all coded ethnicities, including "Not stated", which increased three-fold, from 2.12% in 1995 to 6.52% in 2010. The proportion of patients coded as being of White ethnicity doubled from 30.14% in 1995 to 66.49% in 2010. The proportion of patients coded as being of South Asian ethnicity increased eight-fold, from 0.93% to 7.37%. The proportion of patients coded as being of Black African/Caribbean and Other ethnicity both increased 11-fold, from 0.36% to 3.99% and 0.29% to 3.10% respectively. The proportion of patients coded as being of Mixed ethnicity increased 14-fold, from 0.12% to 1.69%.



	<1995	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Missing	66.15	67.79	67.89	67.64	68.60	67.61	70.78	70.11	66.58	66.34	65.87	64.56	26.61	12.52	11.82	10.48	10.93
Not Stated	2.12	2.28	3.02	3.93	2.17	2.33	2.31	2.22	3.76	2.45	2.58	2.48	9.21	9.41	8.08	7.35	6.52
Mixed	0.12	0.22	0.21	0.22	0.22	0.26	0.28	0.27	0.36	0.39	0.50	0.54	1.09	1.34	1.51	1.62	1.69
Other	0.28	0.40	0.47	0.48	0.51	0.49	0.52	0.59	0.67	0.90	0.9	1.04	1.82	2.12	2.57	3.14	3.1
Black	0.36	1.01	1.20	0.83	0.93	0.98	1.03	1.15	1.53	1.82	2.07	2.41	3.08	3.38	3.71	3.87	3.89
South Asian	0.83	1.08	0.96	1.14	1.44	1.35	1.56	1.59	1.95	2.12	2.36	2.85	4.62	5.45	6.18	6.61	7.37
White	30.14	27.23	26.25	25.75	26.14	26.99	23.50	24.07	25.14	25.98	25.73	26.13	53.57	65.78	66.13	66.93	66.49

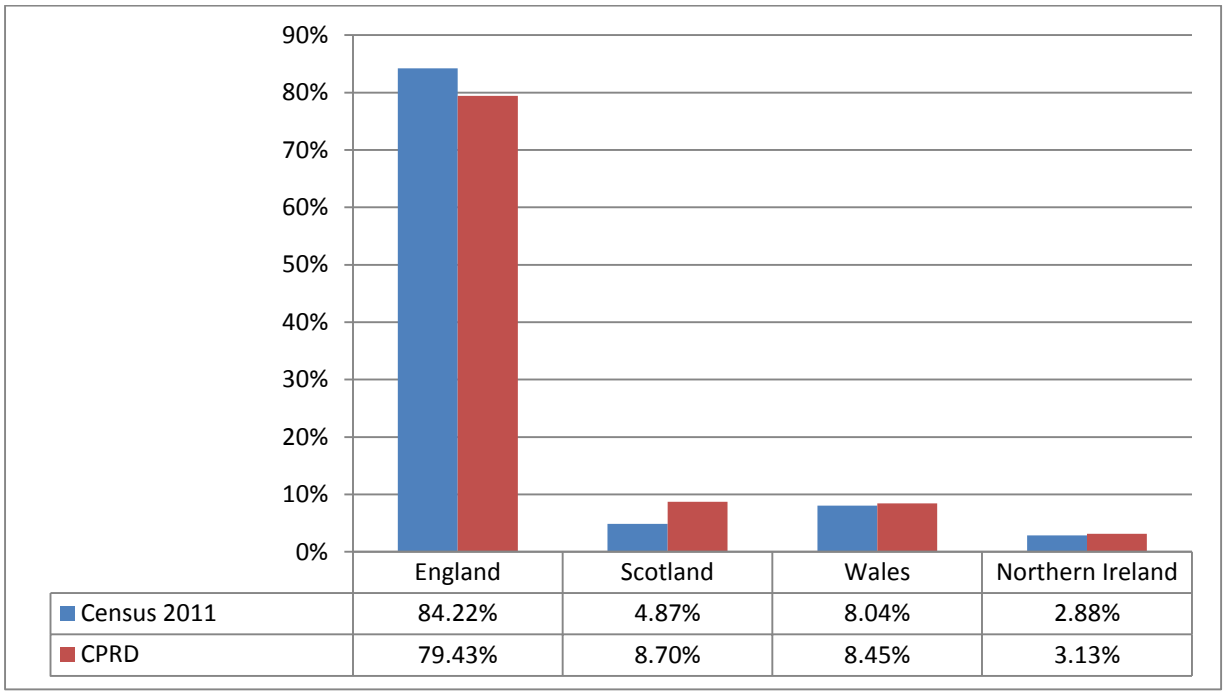
Figure 5.9 Ethnic makeup of the CPRD population, 1995–2010

5.4.5 Comparison of the CPRD Population with the 2011 UK Census Population

The final objective of this study was to determine whether the ethnic breakdown of the CPRD population is consistent with that of the 211 UK Census population. From the 12,099,672 patients contributing to the July 2012 build of the CPRD database, 5,219,411 were active on census day, March 27th, 2011. Within this population, 1,446,254 had registered on or after April 1st, 2006.

a) Regional comparison

Compared with the 2011 Census, the CPRD census day population has a slightly higher proportion of individuals from Scotland, Wales and Northern Ireland, and a smaller proportion of individuals residing in England (Figure 5.8).



	England	Scotland	Wales	Northern Ireland
Census 2011	84.22%	4.87%	8.04%	2.88%
CPRD	79.43%	8.70%	8.45%	3.13%

Figure 5.10 Regional breakdown of patients contributing to CPRD and the 2011 UK Census on March 27th 2011

b) Age comparison

The age structure of the active CPRD population on census day is virtually identical to that of the UK census population, indicating that patients contributing to CPRD are representative of the UK population in terms of age. Registrations from 2006 onwards relate to a much younger population, as would be expected as this population excludes some older individuals, but includes all children born from April 2006 onward. Figures 5.11 through 5.13 illustrate the

similarities between the age structures of the CPRD and Census populations by age group, gender and ethnic group.

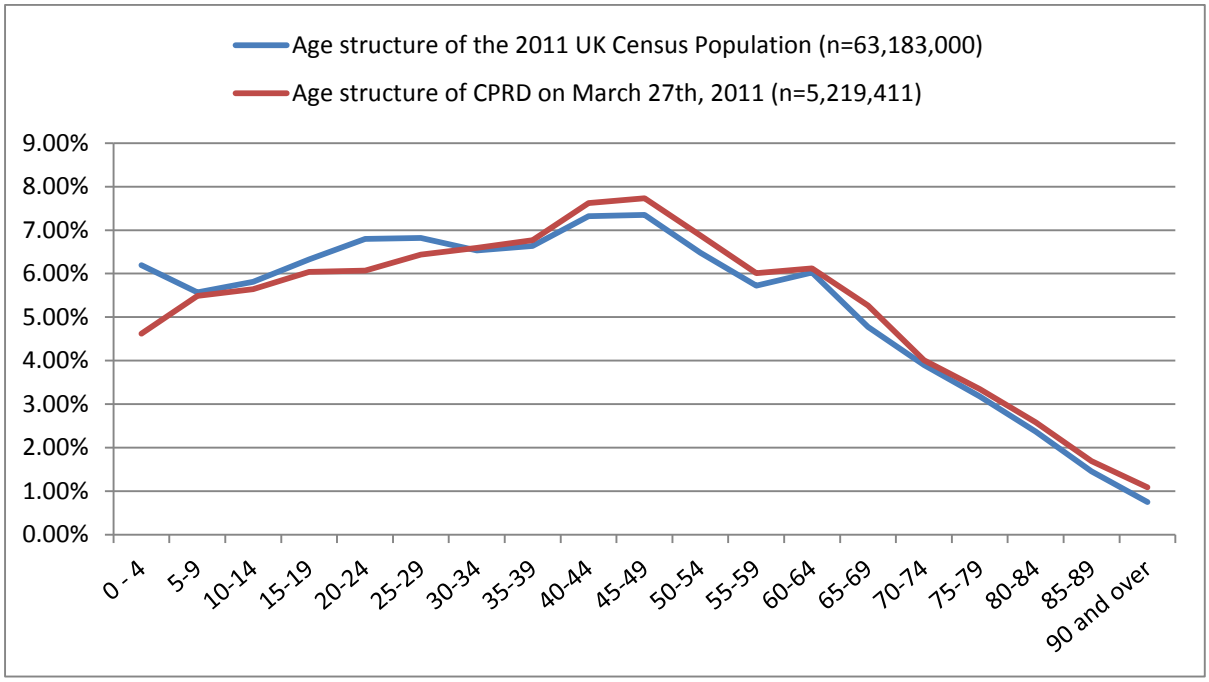


Figure 5.11 Comparison of age structure in the 2011 UK Census Population and 2011 CPRD Population on March 27th, 2011

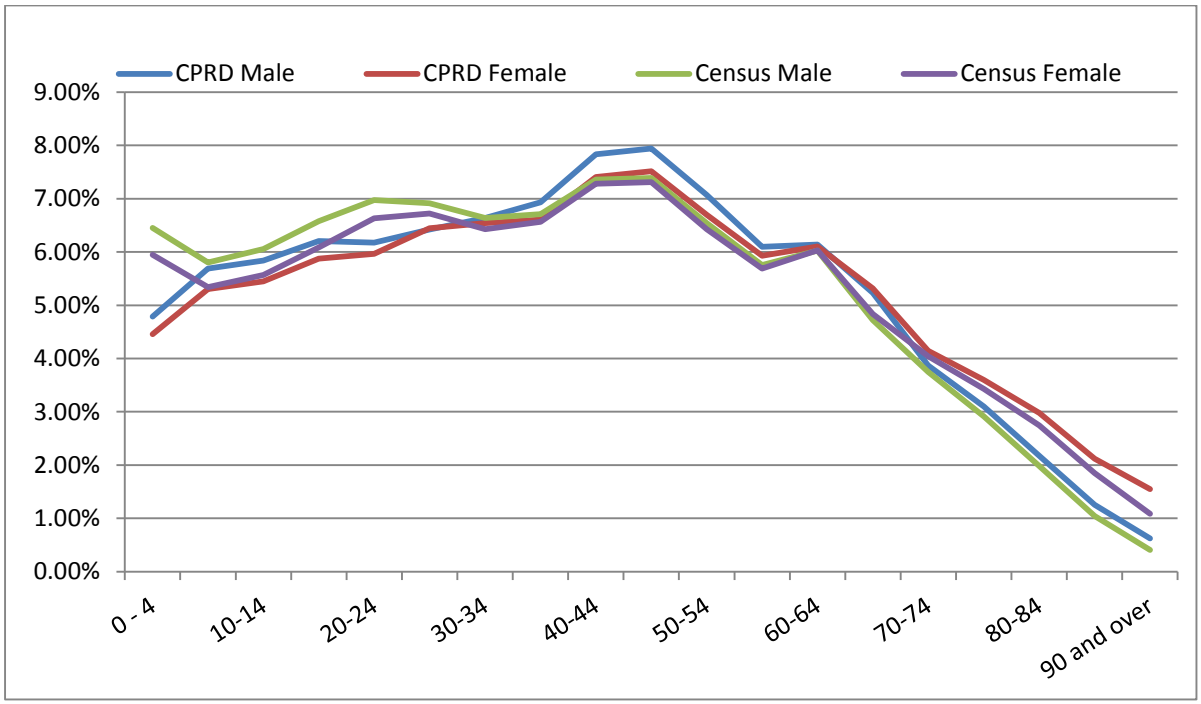


Figure 5.12 Age structure in the 2011 UK Census Population and 2011 CPRD Population by gender on March 27th, 2011

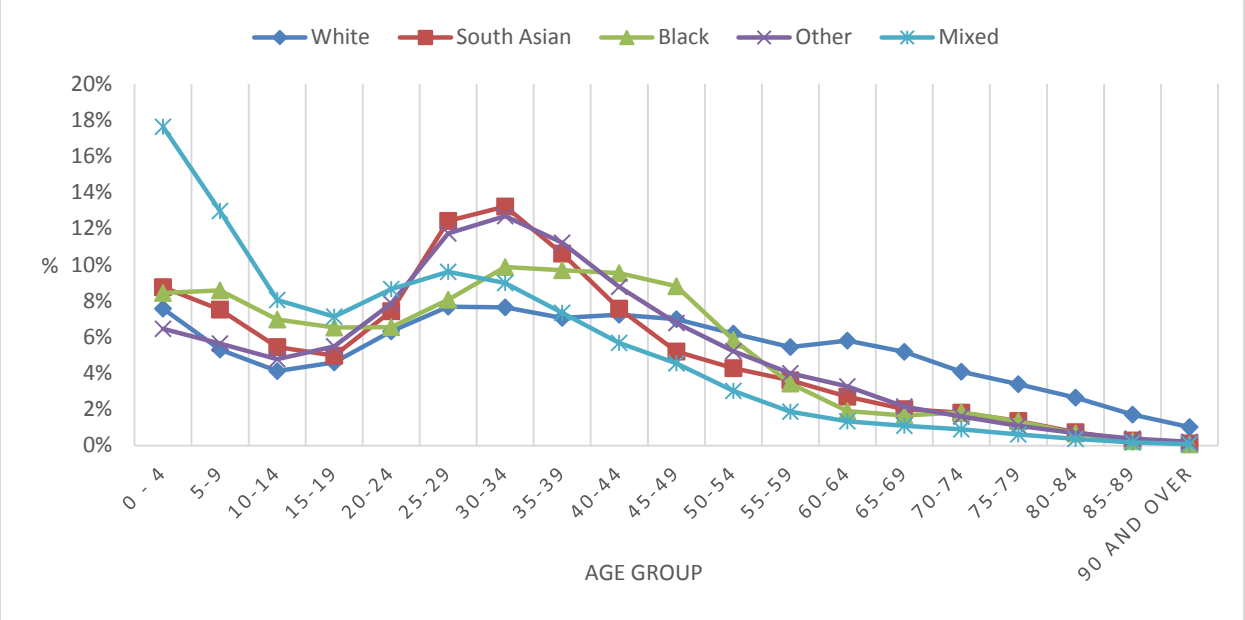


Figure 5.13 Age Structure of Census day CPRD population by ethnic group on March 27th, 2011

c) Ethnicity Comparison

Of the 5,219,411 patients in the CPRD who were active on March 27th, 2011, 42% had ever had their ethnicity recorded. The ethnic breakdown of the UK population in the 2011 census is very similar to the whole CPRD population on that date, both before and after age standardisation. Once restricted to registrations from 2006 onwards, the proportion of non-white individuals in the CPRD is slightly higher than in the Census (Figure 5.14).

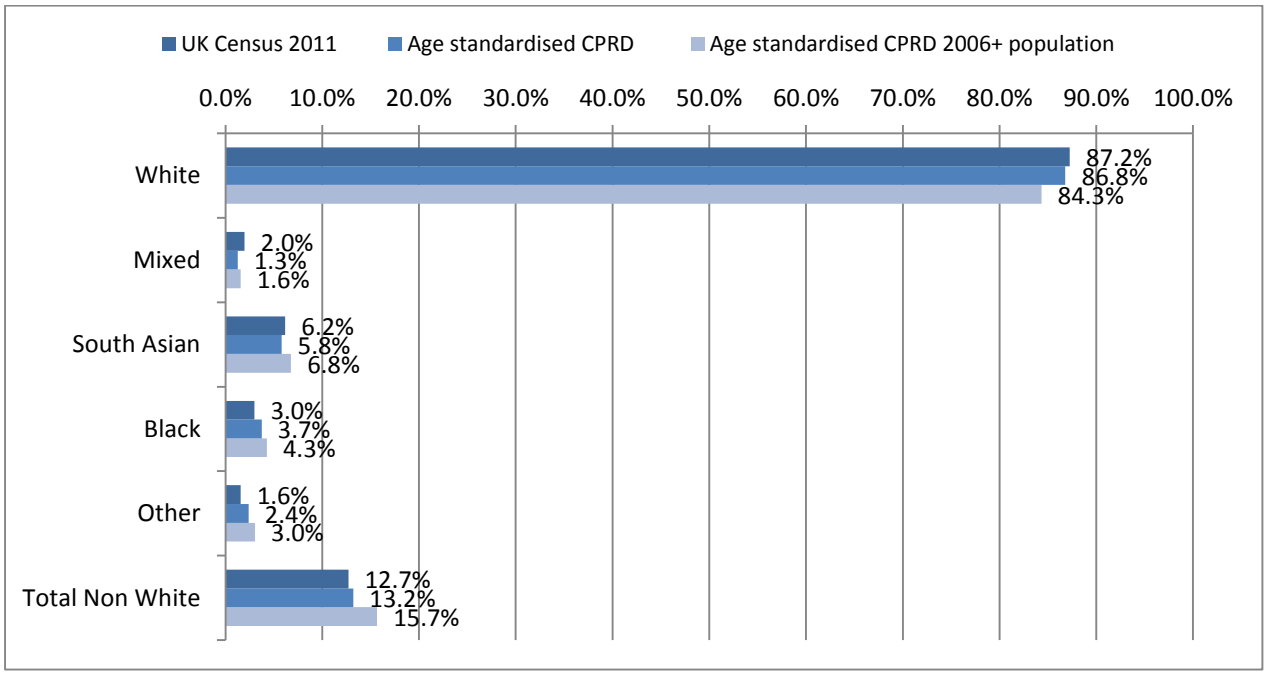


Figure 5.14 Ethnic breakdown of the CPRD and UK Populations on March 27th 2011

5.5 Discussion

5.5.1 Main findings of this study

This study compared the completeness of self-reported ethnicity recording in a sample of UK-wide primary care patients (via the CPRD) and complete records from English secondary care via HES. It further investigated generalisability of the primary care data by comparing the ethnic breakdown in CPRD with that in the UK census, and explored issues of multiple and discrepant recording both within each resource and across linked databases. The study showed that, as of 2012, valid ethnicity is now being recorded for 86% of newly registered patients in primary care, 77% of HES inpatients, and 50% of both HES A&E patients and outpatients.

Over 80% of patients in the CPRD and 90% of HES inpatients and A&E patients with multiple ethnicities had codes which were either truly identical or fell into the same five high-level groups. However, the comparable figure for HES outpatients was only 10%, indicating a high degree of instability, limiting the usefulness of this particular dataset.

The ethnic breakdown of the CPRD, which has already been shown to be representative of the UK population in terms of age and gender, was found to be comparable to that of the combined censuses for England, Wales, Scotland and Northern Ireland.

The overall completeness of ethnicity recording in the whole of CPRD shown here (29.3%) is comparable to that of other UK-based primary care databases such as QRESEARCH (33.5%) and The Health Improvement Network (33.5%)117, though the latter figure excludes patients with discrepant ethnicity. Furthermore, the completeness of usable ethnicity coding shown here for HES inpatients is comparable to that found by Jack et al. in 2002/3 (81.1%) and Mindell et al. in 2003/4 (79%).(122,125)

Linkage of the CPRD and HES inpatient data increased the proportion of complete records to 88.9% overall and to 97.1% for those registered from 2006 onwards. The ability of routine health database linkage to reduce missing data has been detailed in a study linking the UK Renal Registry (UKRR) to HES inpatient data, and Office for National Statistics Mortality data. Similar to

the present study, the authors found completeness of ethnicity recording improved from 75.5% in the UKRR to 98.9% after linkage.(119)

Finally, though agreement between HES and CPRD were found to be high overall, this was driven primarily by patients coded as being of White ethnicity. For patients of South Asian ethnicity, the correlation was only 50%, and it was weaker still for other ethnic groups. The findings here mirror those of a recent study which compared ethnicity recorded in HES to the "gold standard" of self-reported ethnicity as captured in the 2010 Cancer Patient Experience Survey in England.(120) The study reported high accuracy of HES coding for patients of White British ethnicity, but far weaker agreement for all other ethnic groups.

5.5.2 What is already known on this topic

We know that routinely collected ethnicity data in UK-based healthcare databases is under-utilised for observational epidemiological studies. Primary reasons for this include perceptions of poor completeness and quality of these data. National programmes targeting the improvement of this measure have been implemented, but the completeness and usability of these data are not currently audited. Completeness has been shown to be higher in regional databases such as the Wandsworth primary care database and the east London primary care database, which, as a result of local schemes targeting ethnicity recording, both have reported completeness of ethnicity recording over 90% for the past five years.(144,145)

5.5.3 What this study adds

This study has demonstrated that ethnicity data is being captured for the majority of the population in electronic healthcare databases, and that these data are largely complete and comparable to those for the general population. Linkage of datasets yields completeness of almost 100%, with high levels of agreement for patients of White and South Asian ethnicity.

Previous studies have ascribed patient ethnicity indirectly, via name-recognition software or by estimating ethnic population size from census data. Both these methods are of questionable validity, particularly for individuals of mixed ethnicity and for descendants of migrants.(68-71)

Though these methods have been useful for certain situations in the past, they are increasingly less useful now, especially in countries such as the UK, where large proportions of current ethnic minority groups are UK born. Looking forward, there is little alternative to routine recording if we wish to study ethnicity in the long run.

With respect to primary care, this study shows that the recording of valid ethnicity for new patients registering with general practice across the UK has improved dramatically following incentivisation under the Quality and Outcomes Framework. For secondary care, we have shown that the overall completeness of valid ethnicity for HES inpatients has been high for over a decade. As such, we recommend that the ethnicity data held in primary care databases, particularly for patients registered from 2006 onwards, and for hospital inpatients, can and should be more widely used for research, commissioning and audit purposes, particularly given the potential of linking to other resources, such as disease registries, birth and death registers, laboratory results and medication databases.

5.5.4 Limitations

The trade-off for maximising completeness of ethnicity is age selection. While the population from 2006 onwards includes all children and is thus representative for studies examining ethnicity and care patterns in this group, for older patients the study population will be partially missing, though the level of bias this introduces will depend greatly on the research question and study design.

Potential problems with the remaining HES datasets include poor completeness (particularly for A&E) and poor consistency of ethnic group coding (for outpatients). The trends shown in the analysis above suggest that completeness of valid coding in these sources is improving. It is possible that prioritising the recording of valid ethnicity in these settings, perhaps via financial incentivisation as in primary care, may facilitate this process.

For researchers interested in using routinely recorded ethnicity data, it is important to be aware of the biases that may arise from using incomplete data. The likelihood of having missing

ethnicity may not be random, and instead be related to factors such as the circumstances in which patients are admitted, pressures on the available staff and lack of time or opportunity to ask the patient about their ethnicity. As none of these service-level factors are recorded in routine health databases, we cannot estimate the impact these may have on the completeness of ethnicity data.

Turning to biases from inconsistent data, though patients with discrepant ethnicity accounted for a very small proportion of all those included in our analysis, utilising the most commonly recorded ethnic group for these patients proved a reasonable approach in our analysis. In our study of linked datasets, individuals with more than one ethnicity recorded equally commonly represented less than 1% of the sample. For this group, using the latest ethnicity recorded instead of excluding them entirely would maximize usable patient data.

Further bias may have arisen from the possibility that the quality of ethnicity recording varies between general practices. The proportion of patients with usable ethnicity may have varied between the practices contributing to the study and will warrant future investigation. In general, the data quality of practices that choose to contribute to the CPRD is higher than the average nationally, and thus this variability may have been minimal.

This study did not explore ethnicity beyond what was recorded in routine databases. Though the collection of self-reported ethnicity for all patients is the gold standard across the NHS, it is likely that data are collected in a range of non-standardised ways across a variety of non-standardised situations. This is a fundamental challenge of using routinely recorded health data for observational studies. Furthermore, due to the observational nature of this study, we are not in a position to explore the reasons why ethnicity is not recorded, or recorded inconsistently over time and between data sources.

Relying on observational data meant that we were unable to “validate” whether the ethnicity data held in the databases was correct, according to how patients would actually classify themselves.

Furthermore, this study was unable to account for any actual changes in the self-perceived ethnicity that the patient may have experienced. Any differences in ethnicity were classified as discrepancies rather than true differences, as this would be impossible to ascertain without contacting the patient directly.

5.6 Conclusions

The importance of ethnicity in explaining differences in patterns of disease, healthcare usage and outcomes is widely recognised. Previous research has been hampered by deficits in the quality of routine data and insufficiency of estimation methods. This study has demonstrated that the completeness and consistency of routinely recorded ethnicity data in UK-based primary and secondary care have largely improved over time and, with certain caveats, can be usefully incorporated into health research. We have highlighted dramatic improvements in the quality of primary care ethnicity data, particularly since the incentivisation of ethnicity recording in 2006.

Completeness of ethnicity information also appears to have been consistently high over the last decade in hospital inpatient settings, but there is still much room for improvement in A&E and outpatient settings. The concept of ethnicity is a moving target, with the relevance and meaning of ethnic categories perpetually evolving over time. To maximise the value of routinely recorded ethnicity data, both researchers and healthcare professionals must work in tandem to continuously improve both the quality of these data and their impact via timely research.

Chapter 6
Improving the identification of diabetes mellitus in electronic health records

6.1 Summary

This chapter describes the background and methods for the second study of the thesis, exploring methods of improving the classification of diabetes and identifying ethnic differences in incidence and prevalence. The chapter begins with a brief overview of the current literature on diabetes mellitus and ethnicity, outlining the most up to date definitions of diabetes subtypes, and describing the global burden of diabetes and known risk factors. Challenges of accurately classifying diabetes type from electronic health records and initiatives being undertaken by the UK Biobank study to tackle this, the methods pertaining to the execution of the diabetes classification algorithms and resulting analyses of prevalence and incidence are described. The results of the study and discussion of the findings are presented in chapter 7.

6.2 Diabetes Mellitus Classification

The term diabetes mellitus refers to a cluster of metabolic diseases characterized by chronic hyperglycaemia due to deficiency in either the production or action of the hormone insulin. The body uses insulin to regulate blood glucose levels and maintains normal levels by balancing the hormone's production and secretion. The pathophysiology of diabetes is complex, with genetic, lifestyle, early life and environmental exposures all playing a role in its aetiology.(146–148) There are four main subtypes of diabetes, broadly classified according to aetiology: type 1 diabetes; type 2 diabetes; diabetes due to specific mechanisms, such as disease; and gestational diabetes.(149)

Type 1 diabetes mellitus (T1DM, previously known as insulin-dependent or childhood-onset/juvenile diabetes) is characterized by a complete lack of insulin production caused by

the autoimmune destruction of pancreatic β-cells. T1DM accounts for between 5% and 10% of all diabetic cases and onset typically occurs in childhood or young adulthood, though it can occur at any age.(149) Management of T1DM requires continuous insulin treatment in order to simulate the function of the β-cell.(150)

Type 2 diabetes mellitus (T2DM, formerly referred to as non-insulin-dependent or adult-onset diabetes) accounts for 90–95% of the total diabetes burden worldwide. T2DM has a complex and multifactorial pathogenesis resulting in two main deficiencies, decreased insulin secretion and resistance to insulin action, resulting in inadequate removal of excess blood glucose. Primarily caused by excess body weight, particularly around the abdomen, T2DM tends to develop later in life. Recent studies have demonstrated increased incidence in teenagers and young adults, most commonly amongst those who are obese.(149) Most Individuals with T2DM do not require insulin to survive. Less severe forms of the disease can be treated with weight reduction, dietary modification and exercise, while therapeutic management with metformin and other antidiabetic drugs is used in more severe cases. At the most severe end of the treatment spectrum, surgery can be used to assist weight loss and metabolic regulation.(151)

Further types of diabetes include gestational diabetes, maturity-onset diabetes of the young (MODY), drug- or chemical-induced diabetes, genetic diabetes and idiopathic diabetes (of unknown aetiology). Pre-diabetic states, where glucose levels are above normal but not high enough to be classified as diabetes, include impaired glucose tolerance (IGT) and impaired fasting glucose (IFG).(152)

Chronic diabetes mellitus causes abnormalities in the metabolism of protein, fat and carbohydrates, resulting in long-term damage to the kidneys (nephropathy), eyes (retinopathy) and nerves (neuropathy), and increases the risk of cardiovascular disease and premature death by up to four-fold compared with individuals without diabetes.(153–158)

6.3 The global burden of type 2 diabetes

Diabetes presents one of the largest public health concerns globally, with prevalence increasing rapidly, particularly in low- and middle-income countries, as a result of rapid urbanization and the adoption of western lifestyles and diets.(159–167) In 2014, the International Diabetes Federation reported that T2DM affects approximately 381.8 million people globally, with the number projected to increase to 592 million by 2035. Globally, 80% of people with diabetes live in developing countries.(168) Prevalence of diabetes is highest amongst individuals of originating from the Pacific Island regions of Melanesia, Micronesia and Polynesia, in both native and migrant populations. In 2010, the World Health Organization reported that the prevalence of obesity across 10 Pacific Island nations was over 30%, with prevalence of non-communicable disease at 40%.(169) Contemporary studies estimate the prevalence of diabetes to be 40% in Pacific Island communities and 18% in the North American diaspora community.(170,171)

India and China host the largest number of diabetic cases worldwide. Given that these two countries are both experiencing both rapid population and economic growth, it is estimated that they will continue to do so through to 2035.(172,173)

6.4 Risk factors for type 2 diabetes

Obesity is the primary driver of the epidemic of type 2 diabetes worldwide. Over 80% of people with type 2 diabetes are overweight or obese.(174) Increased consumption of calorie-dense foods and the normalization of sedentary lifestyles resulting from rural to urban migration are largely responsible for increases in obesity. Obesity can trigger abnormalities in β-cell function, which increase the body's resistance to the actions of insulin, resulting in hyperglycaemia and diabetes.(175) It is not generalized obesity, but centralized obesity, which has metabolic consequences.(175–178) Individuals who tend to store more fat in the abdominal area are considered to be at high risk of metabolic and cardiovascular disease, regardless of whether or not they are obese according to standard criteria.(152)

Epigenetic factors such as the intrauterine and postnatal environment have been shown to be strongly linked to future risk of type 2 diabetes via a mechanism known as fetal programming.(179)

Family history of diabetes is associated with a two- to four-fold increase in type 2 diabetes risk, indicating a hereditary component to disease pathophysiology.(180,181) Genome studies have identified over 50 genetic abnormalities associated with type 2 diabetes, which, even when combined, explain less than 15% of type 2 diabetes risk.(182,183) Though these abnormalities can be broadly correlated to ethnicity, currently there is no clear understanding of how they contribute to the pathophysiology of the disease, nor whether continued research into genetic markers will yield any benefit in terms of treatment or outcomes.(184)(185) Diabetes and cardiovascular disease interact synergistically, with the presence of one increasing the risk of the other.(148)

The two primary goals of pharmacological treatment for diabetes are to control levels of blood glucose and prevent weight gain.(186–188) The glucose-lowering medications recommended for glycaemic control in type 2 diabetes can increase weight and negate the beneficial effects of reduced hyperglycaemia.(188,189) Weight loss can substantially reduce the incidence of type 2 diabetes and related complications.(190,191)

6.5 Mechanisms of Increased Risk amongst South Asian Populations

The prevalence of type 2 diabetes varies significantly by ethnic group, with the prevalence highest for individuals of South Asian ethnicity, both in native and migrant populations. Ethnic disparities in type 2 diabetes can be traced to differences in the underlying disease process; while insulin resistance is the primary driver of type 2 diabetes in South Asians populations, in other groups diabetes is predominantly caused by β-cell dysfunction.(192,193)

Amongst South Asian groups, onset of type 2 diabetes occurs, on average, a decade earlier than in White/European groups, and at lower levels of conventional risk factors such as age and body mass index (BMI).(192,194,195) Furthermore, South Asian populations have a higher incidence of adverse events and mortality than other ethnic groups. In particular, the age-adjusted incidence of death from CHD amongst those with type 2 diabetes is increased by 50% in South Asian groups.(196)

Several mechanisms which put South Asian populations at increased risk of type 2 diabetes compared with other ethnic groups are discussed below.

6.5.1 Greater insulin resistance

Studies in India, the UK and the United States have demonstrated that South Asian populations exhibit higher levels of insulin resistance than the White population, regardless of age, gender or BMI.(197–201) The prevalence of insulin resistance is also increasing at a faster rate amongst children and adolescents of South Asian ethnicity compared with White ethnicity, resulting in young-onset type 2 diabetes.(202,203)

6.5.2 Body composition

Though South Asian people tend to have a lower BMI than other ethnic groups, they tend to store more fat in the abdominal region than other ethnic groups. As a result, the prevalence of type 2 diabetes in the South Asian population is higher than would be anticipated by the degree of obesity as measured by BMI.(204,205) In order to recognize that diabetes and cardiovascular risk is heightened at lower levels of BMI in this population, the World Health Organization has recommended revised cut-offs of 23 kg/m^2 and 25 kg/m^2 be used to delineate overweight and obesity respectively, revised down from the standard cut-offs of 25kg/m^2 for overweight and 30 kg/m^2 for obesity.(11,206)

6.5.3 Thrifty Genotype and Thrifty Phenotype

The thrifty genotype hypothesis posits that evolutionary traits which allowed individuals to survive in famine conditions by storing excess energy during historic periods of “feast or

famine" have now turned into an evolutionary disadvantage by predisposing individuals to diabetes in the current period of "continuous feasting". According to Bhopal & Rafnsson, this may affect South Asian groups disproportionately, via the biological mechanism of 'mitochondrial efficiency'.(207) The hypothesis suggests that South Asian populations, and other groups adapted to living in warm climates, may be at increased susceptibility to diabetes and cardiovascular disease due the energy-conserving actions of mitochondria in response to environmental and nutritional pressures.

The thrifty phenotype hypothesis suggests that diabetes risk may be increased amongst individuals who experience an unfavourable intrauterine environment as a result of maternal malnutrition or maternal hyperglycaemia. This nutritional scarcity may lead to changes in metabolism which protect the body from food shortages. While these changes may be beneficial in the neonatal and early life period, they may become detrimental in a nutrition-rich environment.(208)

6.6 Type 2 Diabetes in the UK

The combination of a genetic predisposition to diabetes, the stresses of migration and the immersion in western environment greatly increases the risk of diabetes amongst migrants to more affluent countries such as the UK, and serves to outweigh the benefits of the healthy-migrant effect.(36,209–213) Currently in the UK, migrants comprise 13% of the UK population, the majority of whom are of South Asian origin, and at the highest risk of developing type 2 diabetes and related complications.(214)

In the UK 3.2 million people are currently diagnosed with diabetes mellitus, contributing to an overall prevalence of 6.0.%(215) Type 2 diabetes accounts for roughly 90% of all diagnosed cases, and has a prevalence of 5.4%. The prevalence of diabetes increases dramatically with age, ranging from 0.24% in those aged 0–9 to 26.3% in those aged 60–69. The prevalence is higher in men than women, with 56% of cases occurring in men and 44% in women in the UK.(216)

Compared with the White UK population, the prevalence of type 2 diabetes is 6 times higher in South Asian and 3 times higher in Black African/Caribbean populations.(26)(217)(218) Type 2 diabetes is also increasingly prevalent amongst children in the UK, with girls of South Asian ethnicity being at highest risk. According to the National Paediatric Diabetes Audit, children of South Asian ethnicity are 8.7 times more likely to have type 2 diabetes than White children. Correspondingly, children of Black African/Caribbean ethnicity are 6.2 times more likely to have the condition than White children.(219)

6.7 Identifying diabetes mellitus from electronic health records

Accurate diagnostic coding of disease is an essential first step towards identifying patterns of disease and targeting interventions and resources appropriately. This is of particular importance for diabetes mellitus, which has a range of types and subtypes with differing aetiology but similar presentation.(152,220) Though diabetes subtypes are classified according to aetiological process, methods to identify the pathogenesis of diabetes are not yet satisfactory. As such, it is not always possible to accurately identify the correct diabetic type at initial presentation.

Identifying diabetic disease and assigning diabetic type largely depend on the information available at the time of diagnosis. Because different diabetic types can often present in a similar fashion, further investigation over time can reveal mistakes in the initial diagnosis. In routine primary care records, where historical data cannot be removed, this can lead to the presence of multiple contradictory diagnostic codes, vague codes where the diabetic type cannot be determined, or incorrect diagnoses being carried forward.

All of these problems fall under the three main types of diagnostic error: misclassification, miscoding and misdiagnosis. Misclassification is when the patient is coded as the wrong diabetic type; miscoding is when the diagnostic code is too vague or at too high a level to

allow identification of diabetes type; and misdiagnosis occurs when a patient who is not diabetic is coded as having the disease.

In UK primary care, diagnostic codes for diabetes fall under the Read hierarchy of C10% (Table 6.1). When electronic health records were introduced in primary care in the early 1990s, no guidelines on the coding of diabetes were available. Several studies have demonstrated great variation in coding of diabetes. These variations stem from the use of different clinical systems which utilize different modes of data entry and the evolution of terminology for diabetes – most importantly the move away from insulin-dependent/non-insulin-dependent to type 1 and type 2 diabetes mellitus.(221)

The Quality and Outcomes Framework has made the accurate coding of diabetes a payment trigger, and thus streamlined the coding of diabetes in primary care records. When first introduced in 2004, the use of any code under the C10% hierarchy was adequate to trigger payment. In 2006, this was restricted to the use of the codes under the C10E hierarchy for type 1 diabetes and codes under the C10F hierarchy for type 2 diabetes. The rationale for this change was to make QOF indicators more comparable to the NICE guidance, which distinguishes between the type 1 and type 2 diabetes.(222) Miscoding stemming from the use of vague codes has also decreased because these patients will not appear on practice diabetic registers and thus practices cannot be paid for them.

Table 6.1 High level classification of type 1 and type 2 diabetes mellitus

Read code	Description
C10	Diabetes mellitus
C100	Diabetes mellitus with no complications
C108-1	Insulin dependent diabetes mellitus
C109-1	Non-insulin dependent diabetes mellitus
C108-2	Type 1 diabetes mellitus
C109-2	Type 2 diabetes mellitus
C108-3	Type I diabetes mellitus
C109-3	Type II diabetes mellitus
C1000	Diabetes mellitus of juvenile onset with no complications
C1001	Diabetes mellitus of adult onset with no complications
C10E	Type 1 diabetes mellitus
C10F	Type 2 diabetes mellitus

The consequences of improper classification of diabetes have been summarised in a 2010 systematic review of articles describing diabetic coding.(220) These include inappropriate or delayed treatment, failure to identify correct risk factors, psychological effects on the individual and their family, unnecessary treatment resulting in resource wastage, and poor validity of primary care data for audit and research purposes.(220)

6.8 The UK Biobank Study

The UK Biobank is a prospective cohort study established by the Medical Research Council and Wellcome Trust. Between 2006 and 2010, half a million men and women aged 40–69 were recruited for the purpose of investigating genetic, lifestyle and environmental risk factors for major chronic diseases such as cancer, diabetes, heart disease and stroke.(223) This study population will be followed up prospectively over the next several decades and all data collected as part of the study will be freely available for use by any researchers wishing to conduct health-related research that is of public benefit.(224)

Linkage to national electronic health databases such as the Hospital Episode Statistics for England (HES) and the National Cancer Registry is being used to identify incident disease occurring after the baseline interview for all Biobank participants. Since improper or imprecise coding of disease can affect case ascertainment and subsequent attempts to characterise the relationships between risk factors and outcomes, a key objective of the Biobank programme is to improve the identification and classification of diagnoses in routine medical records.(224) This is being piloted across several disease domains, including diabetes, via the use of outcomes adjudication groups, whose role it is to identify and validate disease events by cross-referencing reported outcomes with other sources of supporting data.(225) It is envisaged that incident case identification for diabetes will occur via linkage to primary care records, a process which has been piloted using Welsh primary care data, and is described in section 6.9.

Improving the ascertainment of diabetic type is of particular importance. In addition to the 27,564 cases of diabetes present at baseline in the Biobank cohort, it was projected that 20,000 further cases of diabetes would have been identified by 2015.(226)(223) As such, the Biobank cohort is designed to provide great statistical power to studies looking to typify relationships between diabetes and its risk factors and outcomes.

The algorithms developed for the Biobank study are the most current iteration of a diagnostic toolkit designed to counteract the most common
diagnostic errors as identified studies in London and the southeast of England. The algorithms for classification of diabetes follow the general principles for outcomes adjudication as defined by the Biobank study(225):

1. to use a staged approach to ascertain, confirm and sub-classify disease
2. to avoid false positive cases, but tolerate some false negative cases
3. to be geographically generalizable, scalable, cost-effective and future proof.

6.9 Rationale

In November 2014, Medline was completed by myself to identify published articles utilizing the CPRD to examine ethnic differences in diabetes mellitus. The search showed that no study had been undertaken to quantify ethnic differences in the incidence and prevalence of diabetes mellitus in the CPRD. The following study set out to examine whether estimates of diabetes burden by ethnic group using the observational data held in the CPRD are comparable to those found in cohort and interventional studies undertaken throughout the UK.

Diabetes mellitus is the largest public health problem facing the UK. Its increasing prevalence means that significant numbers of cases are available in the CPRD. Since the disease is predominantly managed in primary care we are likely to pick up the vast majority of cases in our database. Furthermore, findings from this research can be translated into action to address inequalities since the majority of diabetes risk is attributable to modifiable risk factors. Should the data in the CPRD be found usable to examine ethnic inequalities in outcome, the scope for research into ethnicity and diabetes and opportunities for timely and high-impact research at a low cost are expanded.

The purpose of adjudicating disease outcomes is two-fold: firstly, to enhance power by improving identification of disease cases; and secondly, to increase specificity of classification

by differentiating between conditions of similar presentation but different aetiology.(225) Though linkage of primary care records for a proportion of Biobank participants has been achieved in Wales and Scotland, no linkage to English, Scottish or Irish primary care yet exists.

The algorithms designed to adjudicate diabetes cases were developed by a team at the University of Surrey led by Simon de Lusignan using the Welsh data held in the Secure Anonymised Information Linkage Databank (SAIL) at the University of Swansea. Though the algorithms as designed required the use of ethnicity data to help adjudicate diabetes type, no ethnicity data were available at the time of development and thus the results from the initial derivation cohort may underestimate the prevalence of type 2 diabetes in ethnic minority subgroups in the SAIL database population.

The present study sought to improve on the original implementation by utilizing the patient-level ethnicity data available in the CPRD. Additionally, the performance of these algorithms in the CPRD was examined, as improving classification of diabetes by ethnic group is a critical first step towards validating these algorithms for use in other, more diverse settings. The increased specificity of diabetes diagnoses across the CPRD population will greatly improve power to identify ethnic differences in patterns of diabetes prevalence and incidence in large-scale epidemiological studies using this resource.

6.10 Study objectives and hypotheses

The objectives of this study were to:

1. Implement three algorithms designed to improve the classification of type 1 and type 2 diabetes mellitus and develop a simplified version of the algorithms which does not require the use of linked HES data.
2. Compare the performance of the algorithms with respect to improving the coding and classification of diabetes type between the Welsh SAIL database, where the algorithm was derived, and the CPRD.

3. Determine ethnic differences in the prevalence and incidence of type 1 and type 2 diabetes over time using populations derived from the algorithms.
4. Examine ethnic differences age at onset of type 1 and type 2 diabetes, and BMI value at diagnosis.
5. Quantify differences in the incidence of type 1 and type 2 diabetes for South Asian and Black/Caribbean subgroups.

The hypotheses of this study were:

a) The use of algorithms incorporating routinely recorded information on prescribing, clinical measures and competing diagnoses will improve the classification of diabetic type over the use of diagnostic Read codes alone.
b) The prevalence and incidence of type 2 diabetes will increase over time, with incidence highest for South Asian followed by Black African/Caribbean and White groups.
c) Onset of type 2 diabetes will be earlier in South Asian groups compared with White and Black African/Caribbean groups.
d) The BMI value closest to the date of diabetes diagnosis will not vary by ethnicity for T1DM, but be lower for the South Asian in comparison with the White group for T2DM.

6.11 Methods

All clinical and therapeutic data were extracted from the August 2013 build of the CPRD for all patients with at least one diagnostic Read code for diabetes mellitus (see Appendix).

6.11.1 Data Extraction

a) Diagnostic Read codes for diabetes mellitus

Each individual Read code was assigned to one of seven categories (type 1 definite/probable/possible, type 2 definite/probable/possible, or other). If a patient had multiple Read codes falling into any one category, the earliest recorded code of each type was retained for analysis (Table 7.1).

Table 6.2 Categorization of Read codes for diabetes mellitus

	Type 1 Diabetes	Type 2 Diabetes	Other
Definite	Type 1 DM: C10E Not contradicted/ceased/superseded	Type 2 DM: C10F Not contradicted/ceased/superseded	Gestational L180 Genetic C10c-C10D Other/Secondary C10G-J, L-N, C11y0 Insulin resistance: C10K, C1098, C10F8 Ceased: 21263, 212H
Probable	IDDM: C108 Adult onset: C1073 Gestational: L1805 Not contradicted/ceased/superseded	NIDDM: C109 Gestational: L1806 Gestational: L180X Not contradicted/ceased/superseded	
Possible	Diabetes mellitus, adult onset: C10z1 C10y0 C110 Not contradicted/ceased/superseded	Diabetes mellitus, adult onset: C10%, C112 (z), L180x Not contradicted/ceased/superseded	

b) Clinical Measures

For each clinical measure, the value closest to the date of diabetes diagnosis was retained for analysis.

III. Body Mass Index

Body mass index values were either taken directly from the value associated with the clinical Read code for BMI (22K) or calculated using height and weight values recorded in the CPRD additional file using the equation (weight/height^2). Implausible values for height (below 1.37 m and over 2.3m) were removed. Values over 100 which could reasonably be considered to have been measured in centimetres were converted to metres. Similarly, implausible values for weight (<25.4kg or >254kg) were removed. If a patient had multiple height or weight values recorded on the same day, the difference between the smallest and largest values was calculated; if the difference was less than or equal to 5 cm for height or 2 kg for weight, the average of the values was taken. If the difference was any larger, the records were dropped. BMI values were collapsed in a dichotomous variable for obesity, defined as a BMI >30 kg/m^2.

For the purposes of examining BMI at disease onset, the BMI value closest to the date of diagnosis for both type 1 diabetes and type 2 diabetes was selected. This value could have been before or after the diagnosis date.

IV. Hyperglycaemia

Glycated haemoglobin (HbA1c) and blood glucose values were obtained from the test results file of the CPRD. Hba1c values were converted to % if recorded as mmol/mol and. HbA1c values greater than 20% and blood glucose values greater than 50 were removed due to being outside of the plausible value range. A dichotomous variable was constructed for analysis, where patients were considered to have hyperglycaemia if they had an HbA1c value over 6.5% or a blood glucose value of greater than 11.1 mmol/L at the date of diabetes diagnosis.

c) Medications

All medications falling under the British National Formulary (BNF) headings of "6.1.1 Insulins" and "6.1.2 Antidiabetic drugs" were extracted from the Therapy file of the CPRD. The antidiabetic drugs were split into two groups of "Metformin" and "Other antidiabetic drugs" to create three drug categories in total. Insulin was considered to be "non-continuous" if there was a gap of 6 months or more between any two prescriptions between the date of diabetes diagnosis and August 2013.

6.11.2 Description of the algorithms

a) Flowchart 1: Assigning initial diabetic classification

The first of the three Biobank algorithms provided an initial diabetes classification for each patient, based on diagnostic Read codes alone. Patients with type 1 or type 2 diabetes were classified using a hierarchical method starting with definite codes and stepping down to probable codes and then possible codes. Patients whose only diabetes codes were subtype specific were classified separately (Figure 6.1).

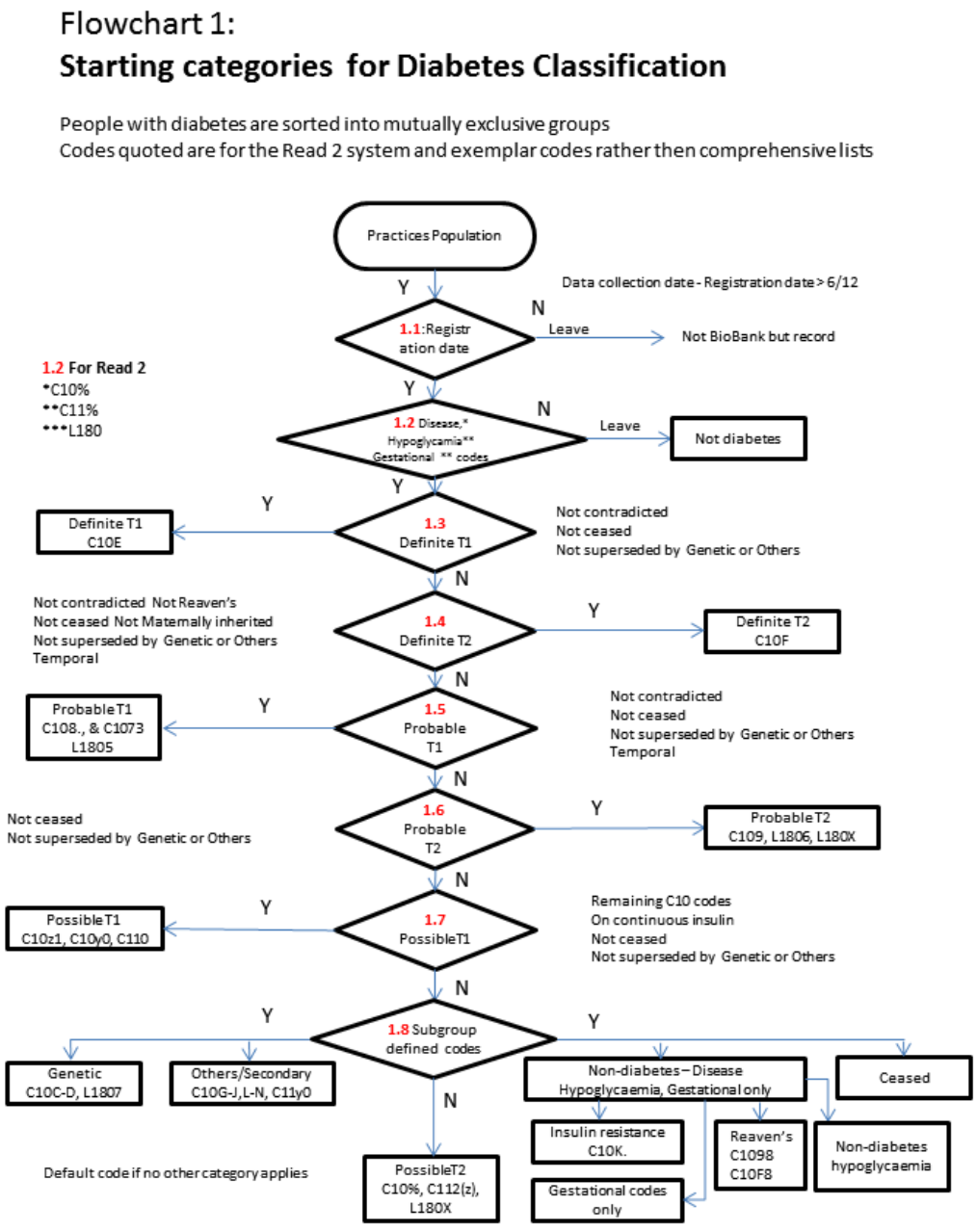


Figure 6.1 Flowchart 1: Initial Sort and Classification of diabetes type

b) Flowchart 2: Improving classification of type 1 diabetes

The second algorithm further refined the classification of patients initially identified as having type 1 diabetes by utilizing supporting information which allowed patients to retain their classification as type 1 diabetics, or to be reclassified as having either type 2 diabetes or gestational diabetes. Supporting information – including prescriptions of diabetic medications (insulin, metformin and other antidiabetic drugs), pregnancy, hospitalizations, body mass index, age, gender and ethnicity – was used to detect potential errors in the classification of type 1 diabetics.

The RCGP algorithm at step 2.8 classified patients as probable type 2 diabetics if they were: aged 35 or over at date of diagnosis if White; aged 30 or over at date of diagnosis if non-White; obese at date of diagnosis (BMI >30) (Figure 6.2). In the original derivation cohort, the RCGP algorithm considered only age and obesity, as ethnicity data were not available.

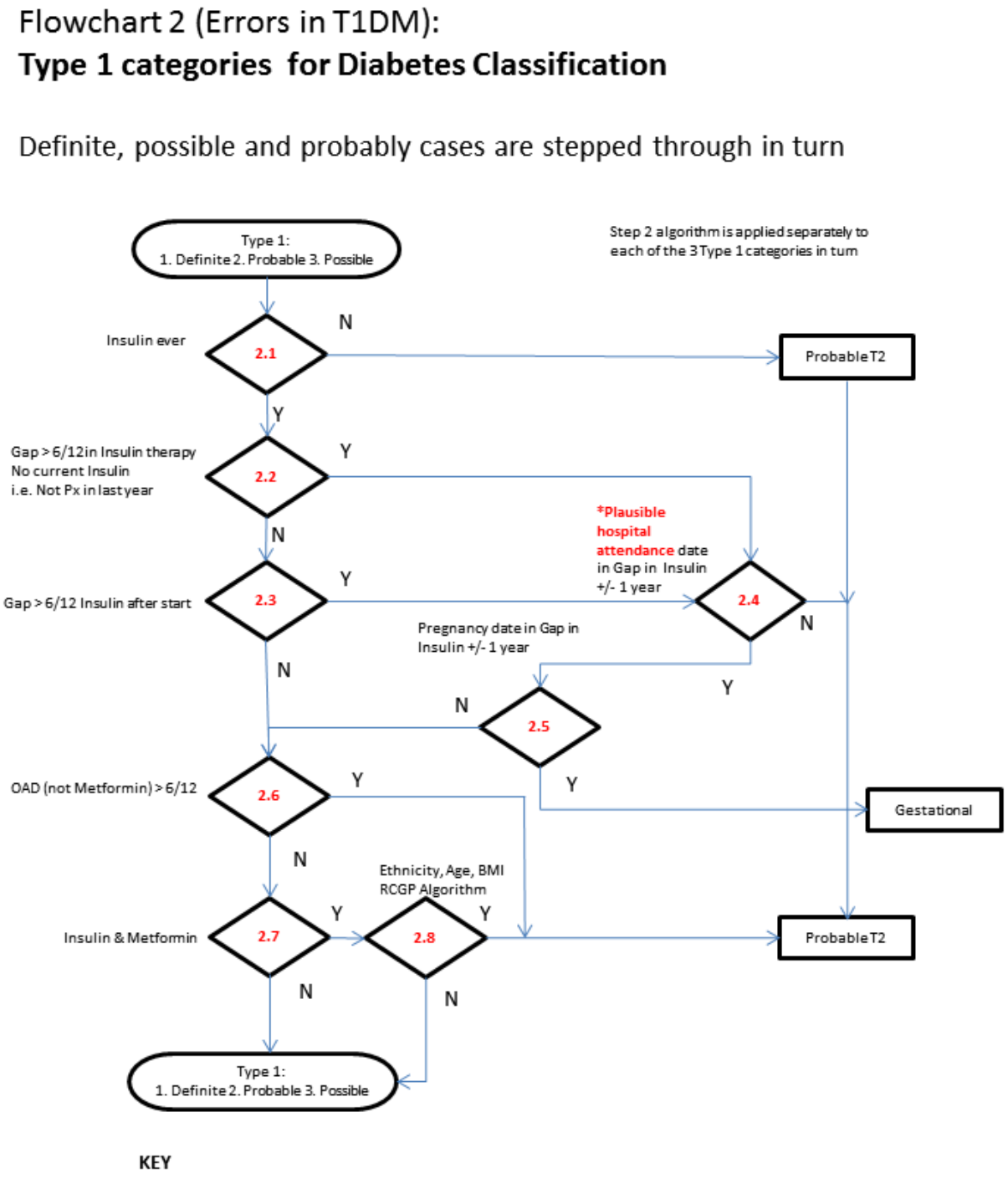


Figure 6.2 Flowchart 2: Detecting errors in T1DM

c) Flowchart 3: Improving classification of type 2 diabetes

The third algorithm further refined the classification of patients initially identified as having type 2 diabetes by utilizing supporting information which allowed patients to retain their classification as type 2 diabetics, or to be reclassified as having type 1 diabetes, polycystic ovary syndrome, gestational diabetes, or not diabetes (Figure 6.3). In addition to the supporting information used in flowchart 2, diabetic hyperglycaemia defined as HbA1c >6.5% or raised blood sugar of >11.1 mmol/L was used in step 3.5.

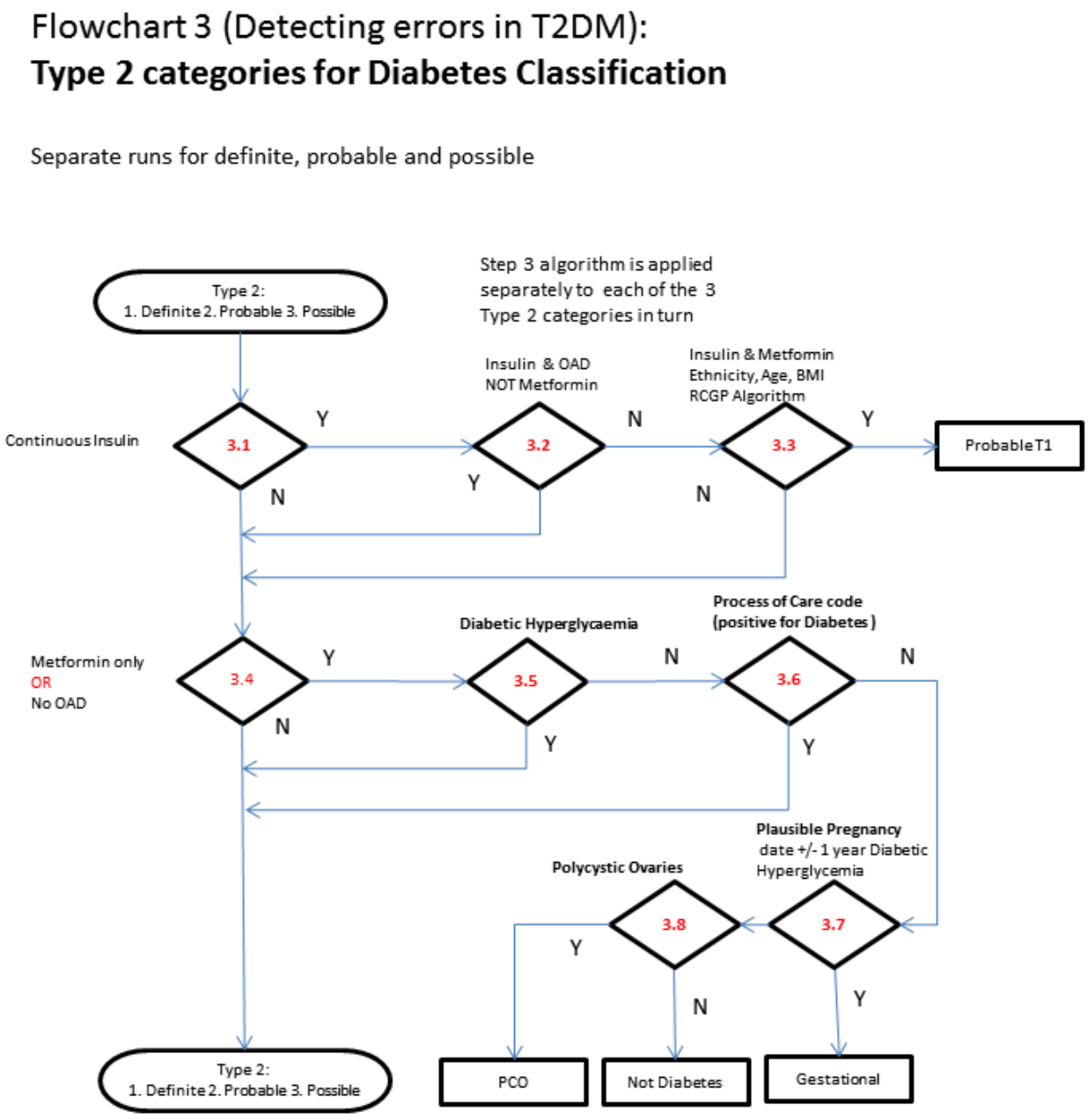


Figure 6.3 Flowchart 3 Detecting errors in T2DM

6.11.3 Description of the simplified algorithms

a) Flowchart 2: Improving classification of type 1 diabetes

In the full Biobank algorithm for detecting errors in T1DM (Flowchart 2), information on hospital admissions is required at step 2.5 as a possible explanation for a gap of 6 months or more in insulin prescription in the GP record. Though hospital data were available for all participants in the original derivation cohort in the SAIL database, hospital admission data were available for only half of the practices contributing to the CPRD that consented to additional linkages. In order to tailor the algorithms for use in the CPRD, a simplified version of the algorithm removing this step was tested, and results were compared between the two versions.

b) Flowchart 3: Improving classification of type 2 diabetes

In the full Biobank algorithm for detecting errors in T2DM (Flowchart 3), women with a pregnancy recorded within 12 months of hyperglycaemia are categorized as having gestational diabetes at step 3.8. After discussion with the Biobank team, it was agreed that this combination does not rule out the possibility that the individual could have type 2 diabetes. In the simplified version of Flowchart 3, step 3.8 was removed and results between the two versions were compared.

6.11.4 Calculating prevalence and incidence of T1DM and T2DM

A prospective cohort study was conducted to examine the prevalence and incidence of type 1 and type 2 diabetes in the CPRD.

a) Participants

All patients identified as having type 1 diabetes at the end of Flowcharts 2 and 3 were included in the analysis of the prevalence and incidence of type 1 diabetes if they had a diagnosis code for T1DM with a date attached. Similarly, all patients identified as having type 2 diabetes at the end of Flowcharts 2 and 3 were included in the analysis of incidence and prevalence of type 2 diabetes if they had a diagnosis code for T2DM with a date attached.

b) Prevalence analysis

For the study of disease prevalence, the outcome was all individuals with a diagnosis of diabetes at the midpoint of each calendar year from 1990 to 2013. Point prevalence was calculated by dividing the number of individuals with diabetes by the number of acceptable patients in the CPRD on July 1st of each year. Prevalence of both type 1 and type 2 diabetes was presented both as an overall proportion and broken down by calendar year, five-year age band, gender and ethnic group. Indirect standardisation against the European standard population was used for the 2012 population in order to compare crude and standardised prevalences for the whole study population, and by ethnic group and gender for 2012, the latest full calendar year of data available.(227)

c) Incidence analysis

For the study of disease incidence, the outcome was first diagnosis of type 1 or type 2 diabetes between January 1990 and August 2013. Individuals with a first diagnosis prior to 1990 were excluded from the analysis. Incidence was calculated by dividing the number of newly diagnosed patients by the number of person-years of follow-up of all eligible patients contributing to the CPRD.

Crude incidence rates of diabetes per 10,000 person years of follow-up time were calculated for patients with either T1DM or T2DM. For the analysis, the start of follow-up was defined as the date at which the practice became up to standard or 6 months after the registration date, whichever was the later.

Follow-up time ended at the date of first diagnosis of type 1 or type 2 diabetes. For patients not diagnosed with diabetes, follow-up time was censored at the earliest of date of transferring out of the practice, date of latest data collection, death, or August 1st 2013.

To examine trends by calendar year, Cox proportional hazards regression with time since study entry as the timescale was used to evaluate the risk of diabetes in all patients between January 1990 and August 2013, broken down by five-year age band, gender and ethnic group. To evaluate trends by age group, Cox proportional hazards using age as the timescale was used.

Chapter 7
Quantifying ethnic differences in incidence and prevalence of diabetes mellitus

7.1 Summary

This chapter outlines the implementation of the series of three algorithms described in chapter 6, designed to improve the classification and reduce misdiagnosis of patients diagnosed with type 1 and type 2 diabetes mellitus using routinely recorded diagnostic, therapeutic and clinical data. The resulting estimates of the incidence and prevalence of type 1 and type 2 diabetes according to calendar period, ethnic group, age and gender are then reported and the utility of the CPRD for examining ethnic differences in diabetes is discussed.

7.2 Results of the diabetes adjudication algorithms

The first objective of this study was to implement the three algorithms and to develop a simplified version of the algorithms which doesn't require the use of linked HES data.

7.2.1 Numbers of patient identified by algorithms

From a total of 15,000,986 patients contributing to the August 2013 build of the CPRD, 10,641,428 were registered for a minimum follow-up period of 6 months and had data that was of acceptable research quality (Figure 7.1).

a) Flowchart 1: Assigning initial diabetic classification

From the total of 10,641,428 acceptable patients, 391,994 had at least one diagnostic Read code for diabetes. At the end of Flowchart 1, which assigns each patient an initial diabetic type based on Read codes alone (page 95), 33,575 patients were classified as having type 1 diabetes (8.6%), 343,047 were classified as having type 2 diabetes (87.5%) and 15,002 patients were classified as having any other sub-group of diabetes (3.8%) (Figure 7.2).

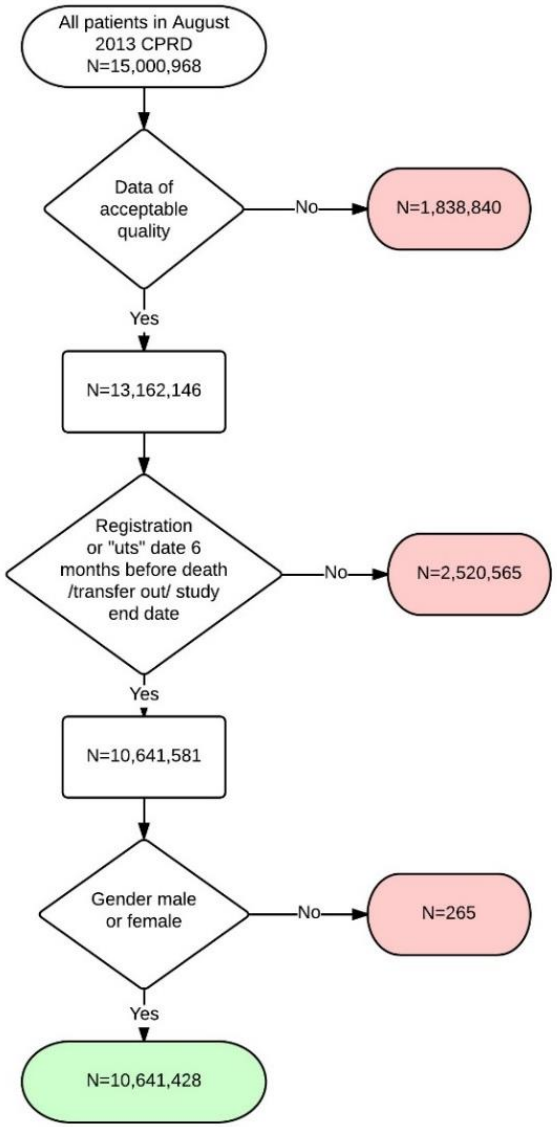


Figure 7.1 Cleaning of the CPRD Denominator

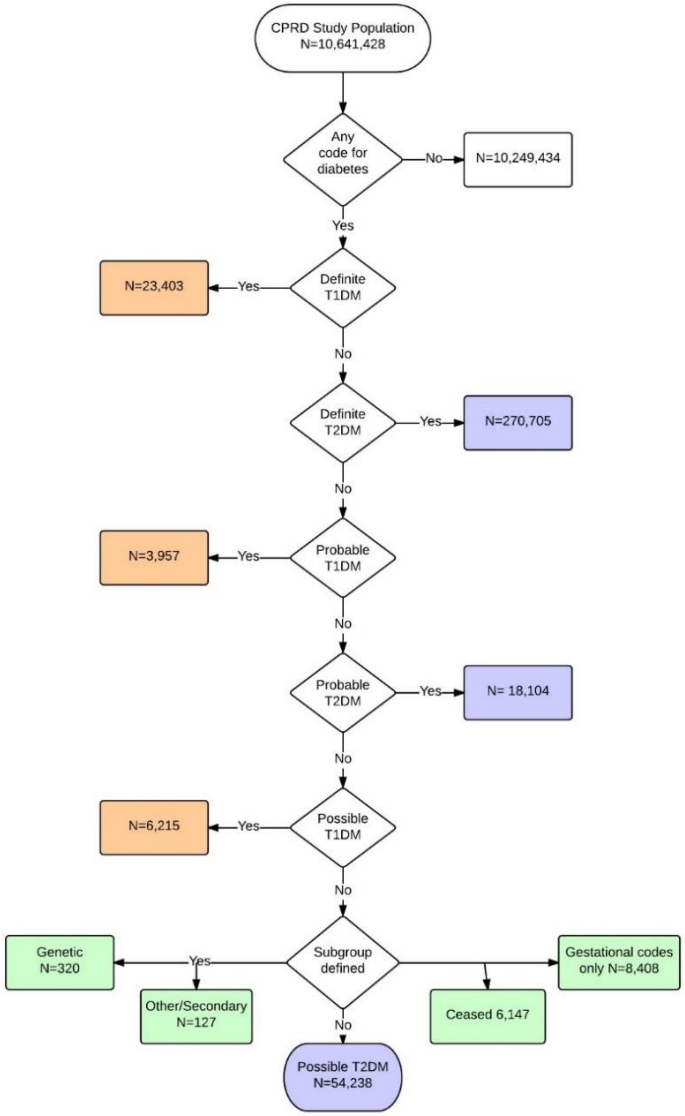


Figure 7.2 Initial Sort and Classification Results

b) Flowchart 2: Improving the classification of type 1 diabetes

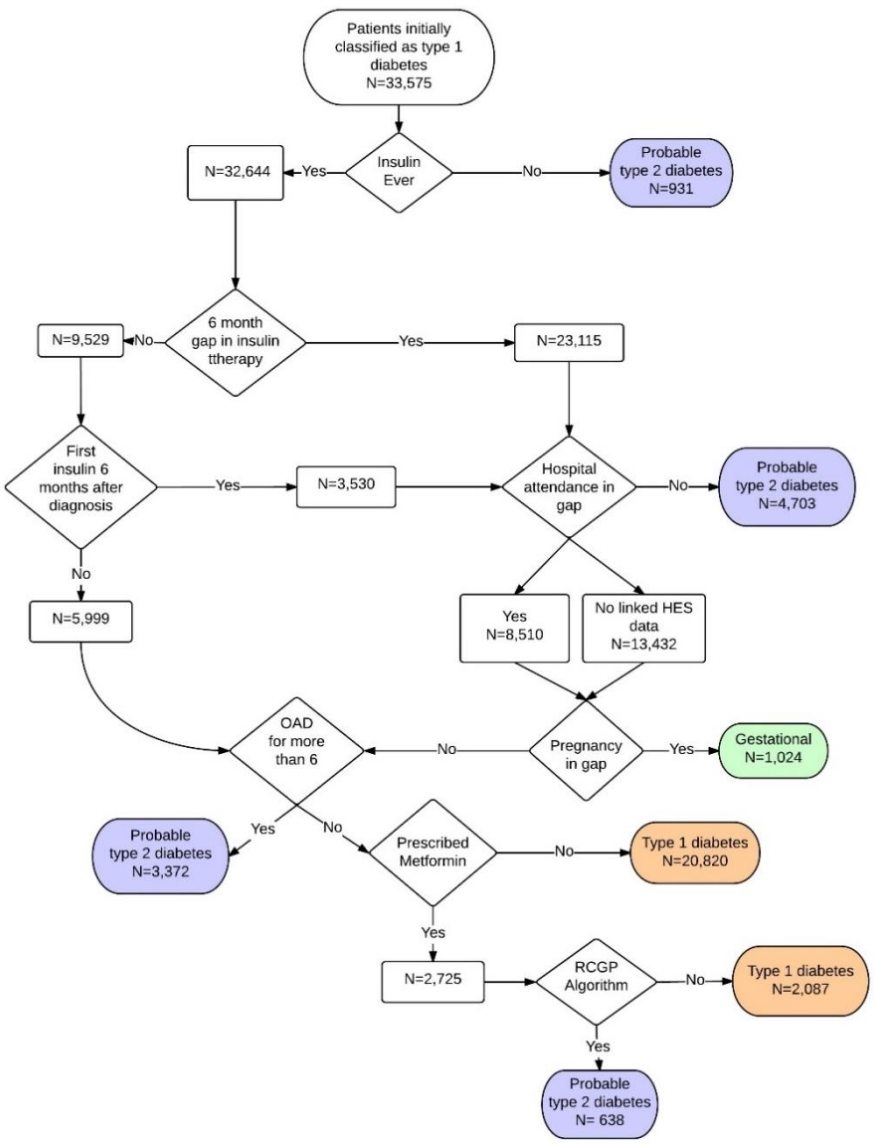
Flowchart 2 was used to refine the classification of type 1 diabetes using supporting information such as medications, hospital admissions and clinical measures (page 96). The results of the second algorithm are displayed in Figure 7.3. A total of 33,575 patients initially classified as type 1 diabetics were entered into Flowchart 2. Using the full version of the algorithm, 68.2% of the initial cohort remained as type 1 diabetics at the end (n=22,907). Using the simplified version of the algorithm, 83.6% of type 1 diabetics remained so at the end (n=28,075). In the full algorithm, 28.7% of patients were re-classified as having probable type 2 diabetes (n=9,644), compared with 16.4% in the simplified algorithm (n=5,500).

Of the 2,725 patients eligible to enter the RCGP algorithm in the main algorithm, 146 (5.4%) were of non-White ethnicity (South Asian, Black African/Caribbean, Other, or Mixed). Employing an age cut-off of 30 instead of 35 for date of first diabetes diagnosis in the non-white group resulted in an additional 20 individuals initially classified as having type 1 diabetes being re-classified as having probable type 2 diabetes. In the simplified analysis, 5.4% of patients eligible to enter the RCGP algorithm were of non-White ethnicity (185/3,417), of whom 22 were classified as probable T2DM by using the lower age cut-off.

c) Flowchart 3: Improving the classification of type 2 diabetes

Flowchart 3 was used to refine the classification of type 2 diabetes by utilising information on medications, clinical measures, and diagnostic codes pregnancy and PCOS (page 97). The results of the third algorithm detecting errors in type 2 diabetes are displayed in Figure 7.4. A total of 343,047 patients initially classified as type 2 diabetics were entered into Flowchart 3. Using the full version of the algorithm, 95.1% of the initial cohort remained as type 2 diabetics at the end (n=326,135). In the simplified algorithm, an additional 201 women classified as having gestational diabetes previously were included in the group of type 2 diabetics, increasing the proportion marginally.

Of the 2,193 patients eligible to enter the RCGP algorithm in both the main and simplified algorithms, 12.3% were of non-White ethnicity (n=269), of whom 20 retained their classification as having type 2 diabetes as a result of the lower age cut-off.



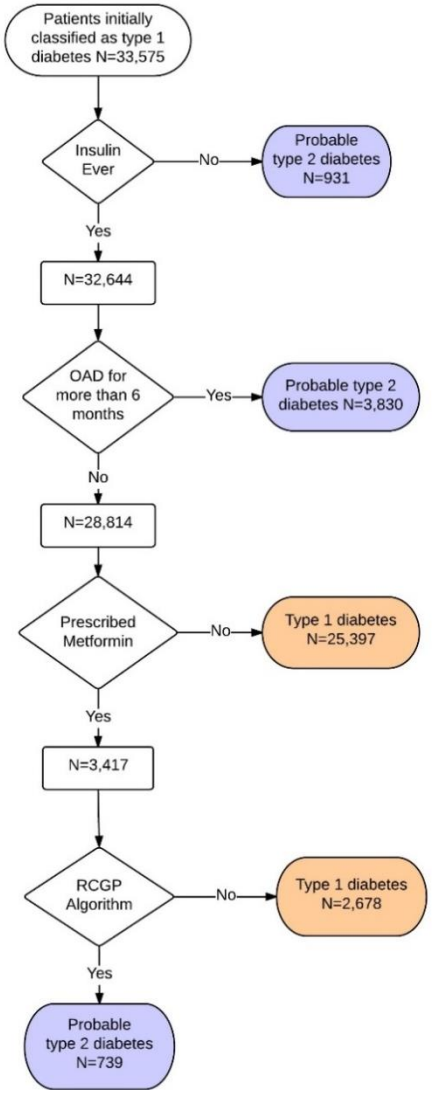


Figure 7.3 Detecting errors in type 1 diabetes: Results for full and simplified algorithms

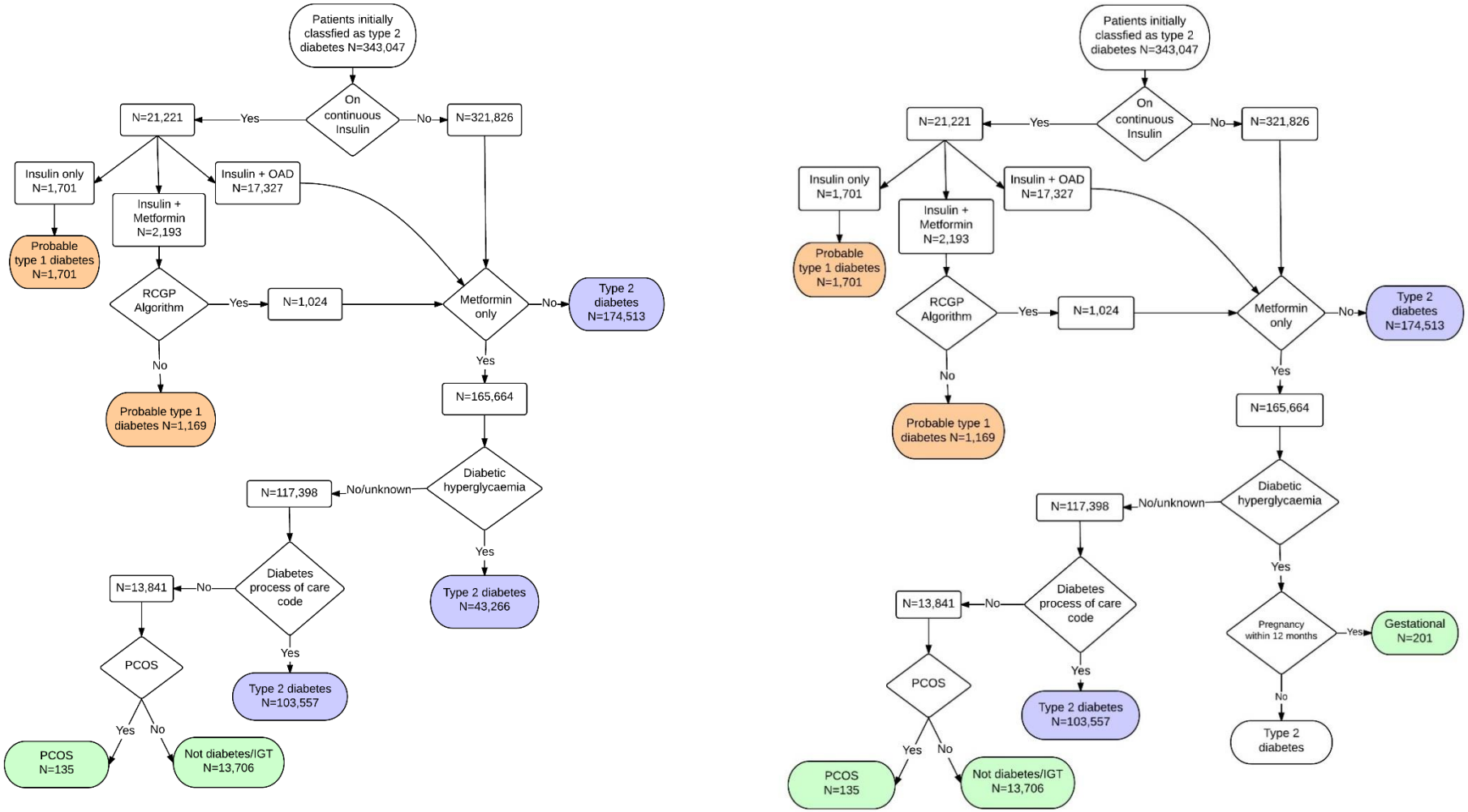


Figure 7.4 Detecting errors in type 2 diabetes: Results for the full and simplified algorithms

7.2.2 Comparison with derivation cohort

The second objective of this study was to compare the performance of the algorithms with respect to improving the coding and classification of diabetes type between the Welsh SAIL database, where the algorithm was derived, and the CPRD.

The comparison of the results from the full and simplified algorithms applied to the CPRD with those of the SAIL database is summarized in Table 7.1. The number of patients comprising the initial cohort of type 1 and type 2 diabetics was 4.5 times larger in the CPRD than in the SAIL database (376,622 vs. 34,596). After initial classification, the CPRD cohort had a greater proportion of patients classified as having type 1 diabetes (8.9% vs. 7.1%) and a smaller proportion of patients classified as having type 2 diabetes (91.1% vs. 92.9%) than the SAIL database.

Table 7.1 Proportion of patients classified as having type 1 or type 2 diabetes in the CPRD and SAIL databases

	CPRD (N=376,622)				SAIL (N=84,596)			
	T1DM		T2DM		T1DM		T2DM	
	N	Row %	N	Row %	N	Row %	N	Row %
Initial Classification (Flowchart 1)	33,575	8.9	343,047	91.1	5,965	7.1	78,631	92.9
Full Algorithms								
T1DM Verification (Flowchart 2)	22,907		9,644					
T2DM Verification (Flowchart 3)	2,870		326,135					
Total	25,777	6.8	335,779	89.1	3,855	4.6	74,628	88.2
% retaining initial classification	77.0		97.9		64.6		94.9	
Simplified Algorithms								
T1DM Verification (Flowchart 2)	28,075		5,500					
T2DM Verification (Flowchart 3)	2,870		326,336					
Total	30,945	8.2	331,836	88.1				
% retaining initial classification	92.2		96.7					

After both the full and simplified versions of the algorithms were run, the proportion of patients retaining their initial classification was higher in the CPRD than in the SAIL database. Comparing the results from applying the full algorithms in the CPRD and SAIL databases, 77% of those initially classified as having type 1 diabetes (25,777/33,575) and 97.9% of those initially classified as type 2 diabetes retained their classification in CPRD (335,779/343,047). In SAIL,

the equivalent figures were 64.6% for type 1 diabetes (3,955/5,695) and 94.9% for type 2 diabetes (74,628/79,631).

Using the simplified algorithms in the CPRD increased the proportion retaining their type 1 classification from 77% in the full algorithm to 92.2% in the simplified algorithm (30,945/33,575). In comparison, the proportion retaining their type 2 classification reduced slightly, from 97.9% in the full algorithm to 96.7% in the simplified algorithm (331,936/343,047).

7.3 Prevalence of type 1 and type 2 diabetes

The third objective of this study was to determine ethnic differences in the prevalence and incidence of type 1 and type 2 diabetes over time using populations derived from the algorithms.

7.3.1 Study population

Patients with type 1 diabetes identified from the simplified versions of the adjudication algorithms were used to calculate the prevalence and incidence of diabetes in the CPRD. From a cohort of 30,945 patients classified as having type 1 diabetes, 28,938 (93.5%) had a date associated with their diagnostic Read code and were included in the analysis. From a cohort of 331,836 patients classified as having type 2 diabetes from the simplified adjudication algorithms, 305,916 (92.2%) had a date associated with their diagnostic Read code and were included in the analysis.

In order to determine whether patients without a diabetes date recorded differed from those with a date recorded, the breakdown of diabetes type, ethnic group and gender in both populations was compared. Table7.2 compares the gender and ethnic breakdown of patients with and without a date for diabetes recorded. Diabetic patients without a diagnosis date recorded had a greater proportion of individuals without ethnicity recorded (58.7% vs 37.3%), a higher proportion of individuals classified as having type 2 diabetes and a higher proportion of females.

Table 7.2 Comparison of patients with and without date of diabetes diagnosis recorded

	No diagnosis date (N=27,927)		Diagnosis date (n=334,854)	
	N	Col %	N	Col %
Diabetes type				
Type 1	2,007	7.2	28,938	8.6
Type 2	25,920	92.8	305,916	91.4
Ethnic Group				
White	10,392	37.2	189,835	56.7
South Asian	545	2.0	11,031	3.3
Black	351	1.3	5,005	1.5
Other	189	0.7	3,091	0.9
Mixed	49	0.2	951	0.3
Unknown	16,401	58.7	124,941	37.3
Gender				
Male	14,938	53.5	185,053	55.3
Female	12,989	46.5	149,801	44.7

7.3.2 Overall Prevalence

The demographic characteristics of the study populations are described in table 7.3. During the study period there were a total of 334,584 cases of diabetes, giving an overall prevalence of 0.26% for type 1 diabetes (28,938/10,641,428) and 2.87% for type 2 diabetes (305,916/10,641,428) across all age groups. Indirectly age standardising the overall prevalence figures for 2012 against the EU standard population did not alter the overall prevalence of type 1 diabetes (0.30% CI95% 0.29, 0.30) but did increase the overall prevalence of type 2 diabetes, from 3.73% to 3.88% (CI95% 3.86, 3.90).

The overall prevalence of type 1 diabetes was highest for White groups (0.36%) and lowest for South Asian groups (0.15%). The prevalence for all other ethnic groups was 0.20%. The overall prevalence of type 1 diabetes was 0.32% for males and 0.23% for females. The most recent prevalence figures (for 2012) are comparable to the overall prevalence for all years. Indirectly age standardising these figures against the EU standard population did not change the overall prevalence of 0.30% but did increase the prevalence slightly in each ethnic group except 'unknown'.

The overall prevalence of type 2 diabetes was highest for South Asian groups (4.7%) followed by White (3.9%) and Black African/Caribbean groups (3.2%). The overall prevalence of type 2 diabetes higher for males than for females. The prevalence of type 2 diabetes in 2012 was higher in all ethnic and gender groups. Age standardisation results in a doubling of the prevalence in South Asian and Black African/Caribbean groups, to 10.32% and 7.42%, respectively.

.

Table 7.3 Gender and ethnic breakdown the study population

	Type 1 Diabetes		Type 2 Diabetes		Denominator population		T1DM prevalence crude all years	T1DM prevalence crude 2012	T1DM prevalence age standardized 2012*	T2DM prevalence crude all years	T2DM prevalence crude 2012	T2DM prevalence age standardized 2012*
	N	Col %	N	Col %	N	%	%	%	%	%	%	%
Total	28,938	100	305,916	100	10,641,428	100	0.26	0.30	0.30 (0.29, 0.30)	2.87	3.73	3.88 (3.86, 3.90)
Ethnic Group												
White	16,143	55.8	173,692	59.8	4,504,565	42.3	0.36	0.36	0.37 (0.37, 0.38)	3.86	4.20	4.08 (4.06, 4.11)
South Asian	347	1.2	10,684	3.5	226,570	2.1	0.15	0.14	0.17 (0.15, 0.20)	4.72	5.48	10.32 (10.08, 10.56)
Black	298	1.0	4,707	1.5	148,320	1.4	0.20	0.18	0.21 (0.17, 0.24)	3.17	3.65	7.42 (7.14, 7.69)
Other	210	0.7	2,881	0.9	11,803	1.1	0.18	0.17	0.19 (0.15, 0.22)	2.44	3.15	5.40 (5.15, 5.65)
Mixed	118	0.4	833	0.3	59,959	0.6	0.20	0.19	0.24 (0.18, 0.31)	1.39	1.55	5.88 (5.35, 6.41)
Unknown	11,822	40.9	113,119	37.0	5,583,971	52.5	0.21	0.24	0.23 (0.22, 0.23)	2.03	3.04	3.23 (3.21, 3.26)
Gender												
Male	16,456	56.9	168,597	55.1	5,179,187	48.7	0.32	0.35	0.34 (0.34, 0.35)	3.26	4.21	4.61 (4.58, 4.64)
Female	12,482	43.1	137,319	44.9	5,462,241	51.3	0.23	0.25	0.25 (0.25, 0.26)	2.51	3.27	3.23 (3.21, 3.25)

***Prevalence figures standardized against the EU standard population**(227)

7.3.3 Ethnic differences in age and body mass index at diabetes onset

The fourth objective of this study was to examine ethnic differences age at onset of type 1 and type 2 diabetes and BMI at diagnosis. The mean and median age and BMI values at onset by ethnic group are displayed in table 7.4

Table 7.4 Age and BMI at onset of type 1 and type 2 diabetes

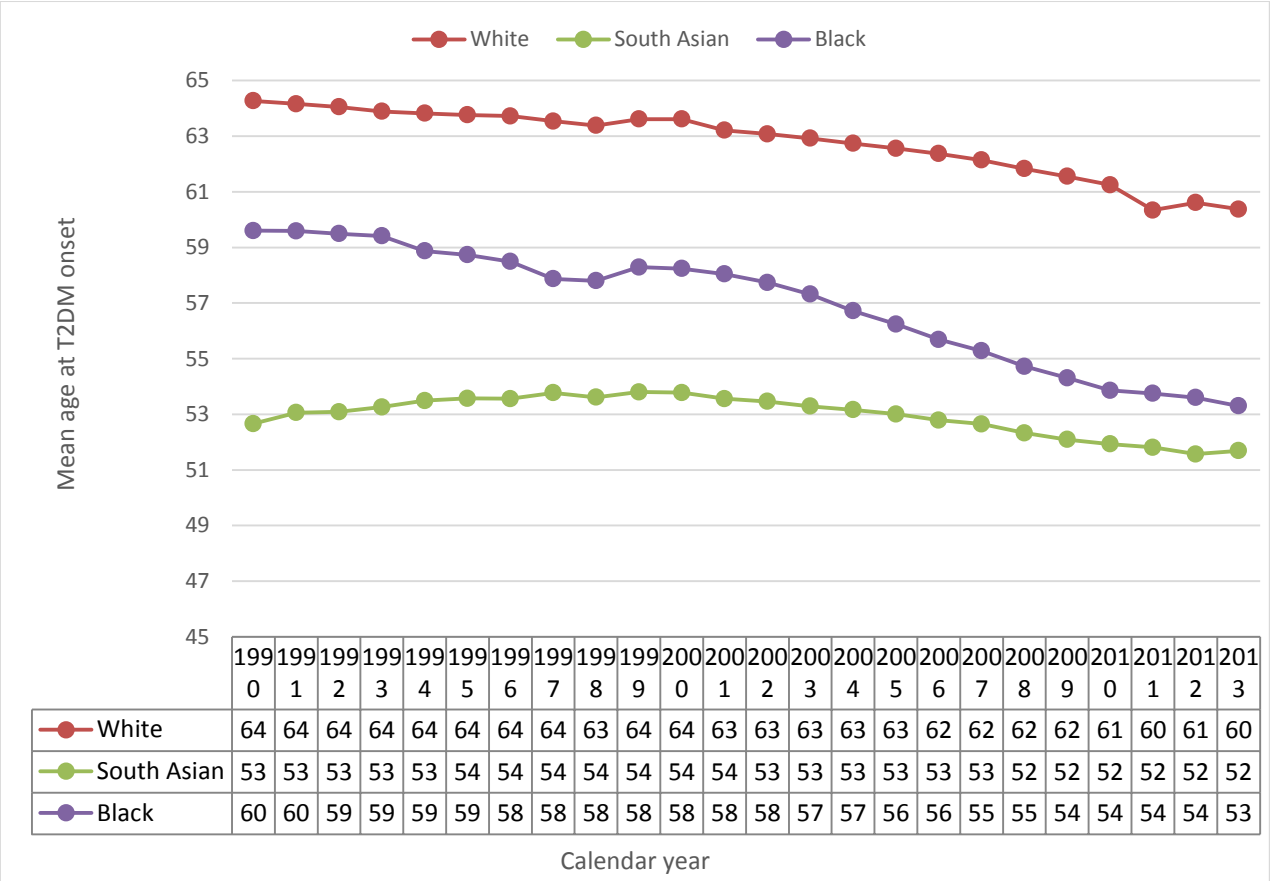
	Type 1 diabetes		Type 2 diabetes		Denominator population	
	N	Col %	N	Col %	N	%
Age at diabetes diagnosis/study entry						
0-9	39, 77	13.7	168	0.1	2,008,865	18.9
10-19	6,856	23.7	471	0.2	1,013,174	9.5
20-29	5,480	18.9	3,156	1.0	1,989,653	18.7
30-39	4,518	15.6	14,818	4.8	1,812,336	17.0
40-49	2,830	9.8	41,380	13.5	1,263,129	11.9
50-59	2,084	7.2	71,006	23.2	951,273	8.9
60-69	1,519	5.3	83,059	27.2	719,843	6.7
70-79	1,155	4.0	63,662	20.8	518,147	4.9
80+	519	1.8	28,196	9.2	365,008	3.4
Median age at DM diagnosis (SD)						
Total	26	(20.3)	62	(13.7)		
White	25	(16.7)	63	(13.5)		
South Asian	27	(19.2)	52	(13.1)		
Black	28	(18.7)	53	(13.5)		
Other	28	(17.8)	55	(13.3)		
Mixed	17	(15.3)	53	(13.6)		
Unknown	27	(21.2)	62	(13.7)		
Mean age at DM diagnosis (CI95)						
Total	30.2	(29.9, 30.4)	61.5	(61.4, 61.5)		
White	29.1	(28.7, 29.3)	62.1	(62.1, 62.2)		
South Asian	29.9	(27.8, 31.9)	52.0	(51.8, 52.3)		
Black	29.8	(27.7, 32.0)	54.2	(51.8, 52.3)		
Other	29.4	(27.0, 31.8)	56.1	(55.6, 56.6)		
Mixed	20.8	(18.0, 23.6)	53.5	(52.6, 54.5)		
Unknown	31.9	(31.5, 32.3)	61.9	(61.9, 62.0)		
Median BMI value closest to DM diagnosis date in kg/m^2 (SD)						
Total	24.1	(4.8)	29.1	(5.55)		
White	24.1	(4.8)	29.4	(5.6)		
South Asian	23.5	(4.3)	27.1	(4.8)		
Black	24.8	(5.3)	29.1	(5.5)		
Other	23.3	(4.8)	27.7	(5.3)		
Mixed	23.2	(5.0)	28.3	(5.8)		
Unknown	24.2	(4.7)	29.1	(5.4)		

The median age of T1DM diagnosis was 26 (SD 20.3). When stratified by ethnic group, median age of T1DM diagnosis was similar for White, South Asian and Black African/Caribbean groups. The mean age of T1DM diagnosis was higher than the median, at 30.2 (CI95% 29.9, 30.4). The mean age of T1DM diagnosis was similar for all ethnic groups except for the Mixed group, whose mean age of onset was nine years lower than that for the whole population, at 20.8 years of age (CI95% 18.0, 23.6). The mean and median age of T1DM onset for patients with unknown ethnicity were both higher than for the White population.

The median age of T2DM diagnosis was 62 (SD 13.7). When stratified by ethnic group, South Asian, Black African/Caribbean and Mixed groups were diagnosed with T2DM a decade earlier than the White group (Table 7.3). Ethnic differences in the mean age of T2DM onset mirror those of the median, with onset in ethnic minority groups 8–10 years earlier than in the White population.

The mean age of T2DM onset stratified by ethnic group is shown for each calendar year between 1990 and 2013 in figure 7.5. The graph highlights large differences between ethnic groups in 1990, which diminish over time. Though a reduction in mean age at onset was observed in all ethnic groups, the decline was greatest in the Black African/Caribbean population, for whom the mean age reduced from 59.6 in 1990 to 53.3 in 2013. The difference in mean age of onset over the study period was 3.9 years for the White population and 1.0 years for the South Asian population.

Turning to BMI value closest to date of diagnosis, the mean value for the whole T1DM population was 24.1 kg/m^2, and highest for the Black African/Caribbean population, at 24.8 kg/m^2, and lowest for the South Asian population, at 23.5 kg/m^2. Mean BMI in the T2DM population was higher, at 29.1 kg/m^2. Mean BMI was higher and equivalent in the White and Black African/Caribbean populations, at 29.4 kg/m^2 and 29.1 kg/m^2, respectively, and lower in the South Asian, and Other ethnic groups, at 27.1 kg/m^2 and 27.7 kg/m^2, respectively.



	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
White	64	64	64	64	64	64	64	64	63	64	64	63	63	63	63	63	62	62	62	62	61	60	61	60
South Asian	53	53	53	53	53	54	54	54	54	54	54	54	53	53	53	53	53	53	52	52	52	52	52	52
Black	60	60	59	59	59	59	58	58	58	58	58	58	58	57	57	56	56	55	55	54	54	54	54	53

Figure 7.5 Mean age of type 2 diabetes diagnosis by calendar year and ethnic group

7.3.4 Prevalence by calendar year

Figure 7.6 illustrates the crude prevalence of type 1 and type 2 diabetes at the midpoint of each calendar year between 1990 and 2012 in the CPRD. The prevalence of type 1 diabetes rose slightly over the study period, from 0.2% in 1990 to 0.3% in 2012. Standardisation against the EU population did not alter the crude prevalence of type 1 diabetes in 2012

The prevalence of type 2 diabetes rose six-fold over the study period, from 0.6% in 1990 to 3.7% in 2012. The prevalence of type 2 diabetes in 2012 increased after standardisation against the EU standard population to 3.88% (CI95 3.86, 3.90). When stratified by gender, the prevalence both type 1 and type 2 diabetes was higher for males than for females in all years. Amongst type 2 diabetics, prevalence increased over time, slightly more so in males than in females.

Ethnic differences in diabetes prevalence over time are illustrated in Figure 7.7. The prevalence of type 1 diabetes remains consistently highest in the White group across all study years; it was lower prevalence for the South Asian group and for the Black African/Caribbean group it was similar to that of the White group. The prevalence of type 2 diabetes is consistently highest in the South Asian group in all calendar years.

Between 1990 and 2007, prevalence was lowest for the White population. From 2008 onwards, the prevalence in the White population overtook that in the Black African/Caribbean population. In the non-White ethnic groups, prevalence peaked in the year 2005, with a midyear prevalence of 4.4% in the Black African/Caribbean group and 6.7% in the South Asian group. In the White ethnic group, prevalence increased annually, to peak in 2012, at 4.2%.

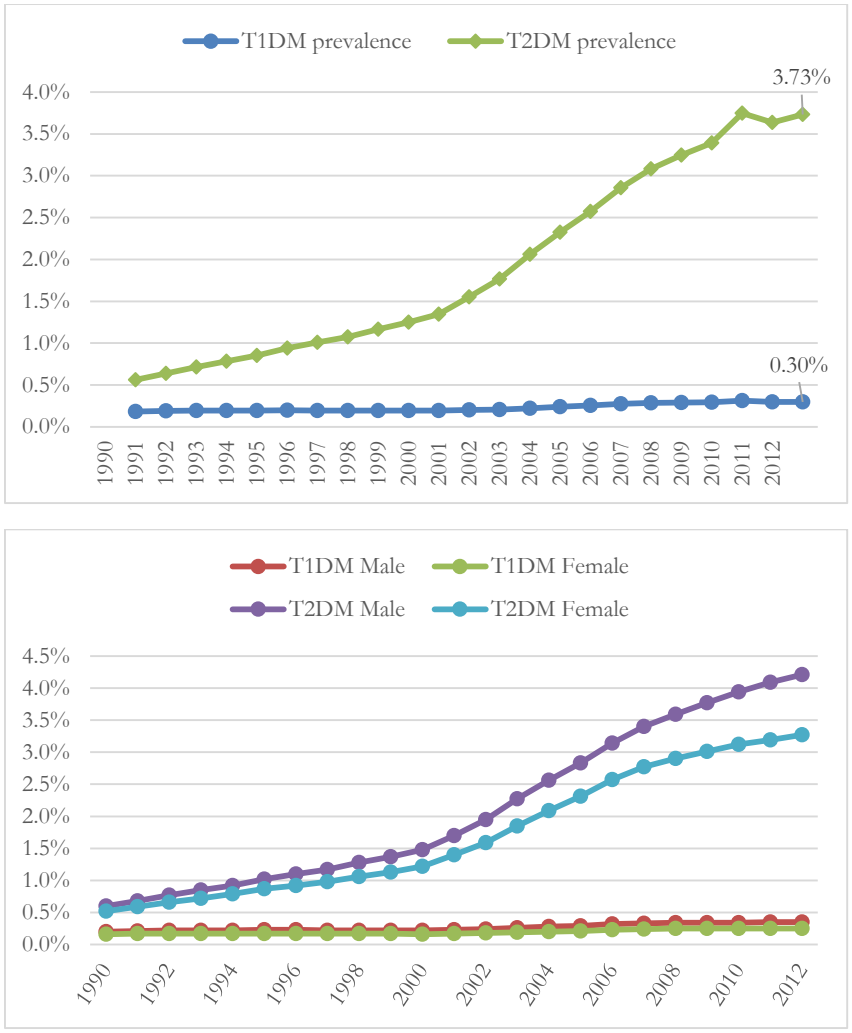


Figure 7.6 Prevalence of type 1 and type 2 diabetes by calendar year and gender

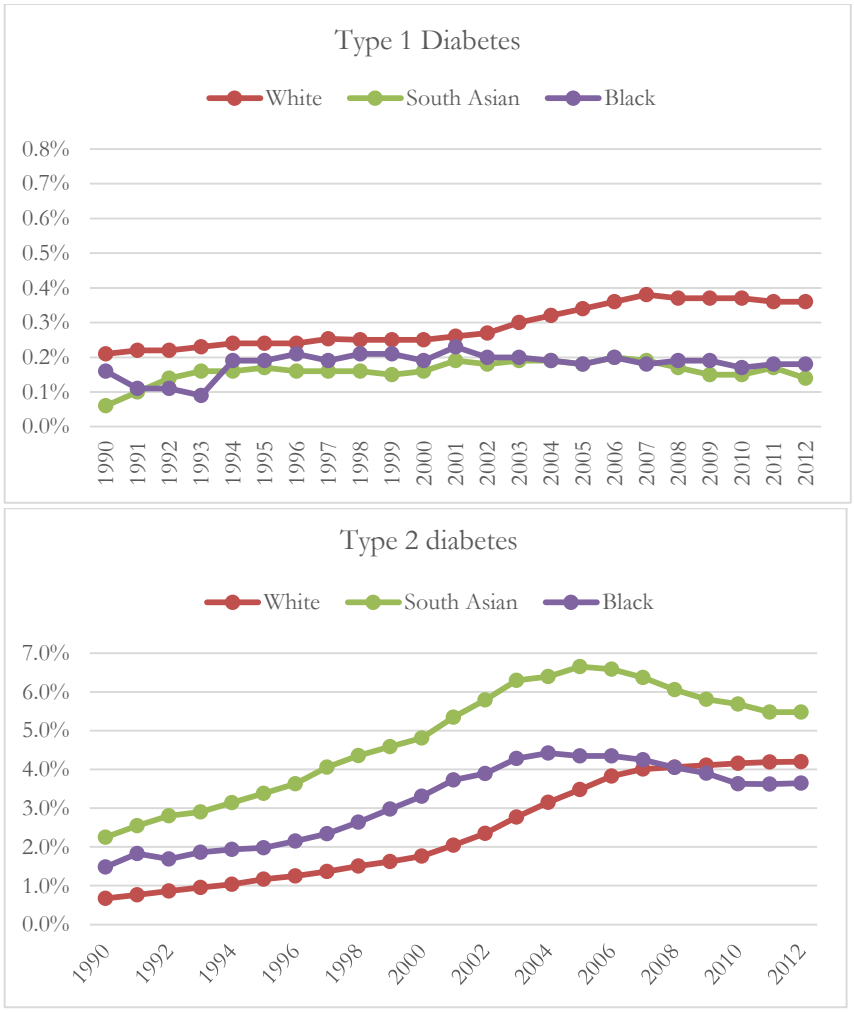


Figure 7.7 Prevalence of type 1 and type 2 diabetes by calendar year and ethnic group

7.3.5 Prevalence by Age group

The most recent prevalence estimates for the mid-2012 CPRD population were used to examine age and gender differences in diabetes prevalence within ethnic groups. The overall and gender specific prevalence of type 1 and type 2 diabetes by five-year age band is shown in figure 7.8. While the prevalence of type 1 diabetes remained constant across all age groups, the prevalence of type 2 diabetes increased rapidly from age 30 onwards and peaked in the age group 75–79 with a prevalence of 14.3%.

When stratified by gender, the prevalence of type 1 diabetes was raised slightly in men compared with women in all age groups. The prevalence of type 2 diabetes was equivalent between males and females until age 35, after which the prevalence was higher in males. Prevalence of type 2 diabetes peaked at in the age group 75–79, with a prevalence of 16.6% in males and 12.4% in females.

The age-specific prevalence of type 1 and type 2 diabetes by ethnic group is shown in figure 7.9. The prevalence of type 1 diabetes was highest in the White group at all ages until age 85, after which the prevalence was highest in the Black African/Caribbean group. No clear differences between South Asian and Black African/Caribbean groups were apparent.

The prevalence of type 2 diabetes was highest in the South Asian population at all ages. Prevalence peaked earlier for South Asian groups than for Black African/Caribbean and White groups. Peak prevalence occurred at age 75–79 in the South Asian group at 31.6% and at age 80–84 in the other groups with a prevalence of 14.7% in White and 29.4% in Black African/Caribbean patients. The decline in prevalence observed amongst in the oldest age groups was related to deaths in the diagnosed population.

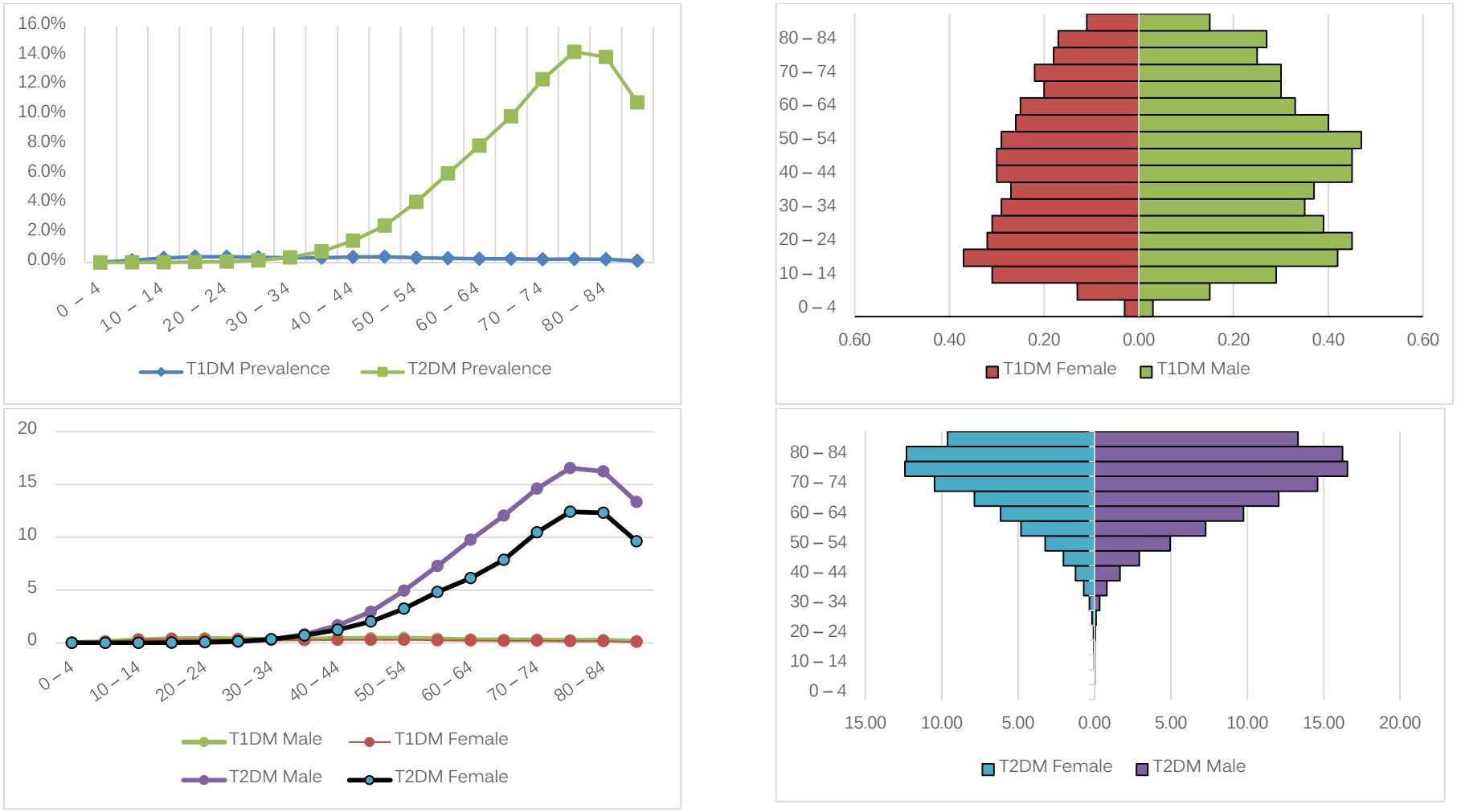


Figure 7.8 Prevalence of type 1 and type 2 diabetes by age at diagnosis and gender in 2012

Figure 7.9 Prevalence of type 1 and type 2 diabetes by age at diagnosis and ethnic group in 2012

7.4 Incidence of type 1 and type 2 diabetes

7.4.1 Overall incidence

The analysis presented below describes the crude incidence and crude risk of type 1 and type 2 diabetes for all patients in the CPRD. A total of 11,972 incident type 1 diabetes events occurred between January 1990 and August 2013, giving a crude incidence rate of 1.58 per 10,000 person-years (CI95% 1.56, 1.61). A total of 227,483 incident type 2 diabetes events occurred over the study period, giving an overall crude incidence rate of 30.75 per 10,000 person years (CI95% 30.63, 30.88).

Table 7.5 presents the crude incidence rates per 10,000 person-years, and age- and gender-adjusted hazard ratios for type 1 and type 2 diabetes by ethnic group, gender and age group.

For type 1 diabetes, the overall rate was highest in the White population, followed by Black African/Caribbean and South Asian groups. In comparison with the White population, the risk of T1DM was reduced by 27% in the Black African/Caribbean population (HR 0.73, CI95% 0.60, 0.89) and by 45% in the South Asian population (HR 0.55, CI95% 0.46, 0.63). The incidence rate was lower in women than in men, with a relative hazard of 0.74 for women (CI95% 1.29, 1.37). Incidence peaked in the 10–19 age group, with a crude rate of 2.96 per 10,000 person-years. Relative to the 0–9 age group, the risk of T1DM was increased by 43% in this group (HR 1.43, CI95% 1.34, 1.52).

For type 2 diabetes, relative to the White group, the incidence was raised two-fold in the Black African/Caribbean group (HR 2.05, CI95% 1.97, 2.13) and three-fold in the South Asian group (HR3.04, CI95% 3.00, 3.11). The incidence was reduced by 29% in women compared with men (HR 0.71, CI95% 0.71, 0.72). The incidence peaked in the 70–79 age group, with a crude rate of 95.47 per 10,000 person-years. Relative to those aged 30–39, the risk of T2DM was increased 11-fold in this group (HR 11.37, CI95% 11.12, 11.62).

Table 7.5 Crude incidence rates and adjusted hazard ratios for type 1 and type 2 diabetes by ethnic group, gender and age

	Type 1 diabetes					Type 2 diabetes				
	Events	Crude rate	[CI95%]	Adjusted HR	[CI95%]	Events	Crude rate	[CI95%]	Adjusted HR	[CI95%]
Overall	11,972	1.58	[1.56, 1.61]			227,483	30.75	[30.63, 30.88]		
Ethnic Group										
White	6,469	1.92	[1.87, 1.96]	1	--	129,770	39.56	[39.35, 39.78]	1	--
South Asian	126	1.18	[0.99, 1.41]	0.58	[0.48, 0.69]	6,148	61.13	[59.62, 62.68]	3.05	[2.97, 3.13]
Black	102	1.59	[1.32, 1.94]	0.78	[0.64, 0.95]	2,589	42.08	[40.49, 43.74]	2.10	[2.01, 2.18]
Other	68	1.29	[1.02, 1.64]	0.64	[0.51, 0.82]	1,784	34.89	33.30, 36.54]	1.38	[132, 1.45]
Mixed	47	1.97	[1.48, 2.62]	0.87	0.65, 1.16]	485	20.60	[18.85, 22.52]	1.59	[1.46, 1.75]
Unknown	5,158	1.31	[1.27, 1.34]	0.60	[0.57, 0.62]	86,707	22.35	[22.20, 22.50]	0.69	[0.68, 0.70]
Gender										
Male	6,902	1.84	[1.80, 1.88]	1	--	125,472	34.31	[34.13, 34.50]	1	--
Female	5,067	1.33	[1.29, 1.37]	0.72	[0.769 0.75]	102,011	27.27	[27.11, 27.44]	0.70	[0.69, 0.71]
Age group										
0-9	1,703	2.06	[1.96, 2.16]	1	--	78	0.09	[0.08, 0.11]	0.01	[0.09, 0.014]
10-19	2,546	2.96	[2.85, 3.08]	1.46	[1.37, 1.55]	261	0.30	[0.27, 0.35]	0.04	[0.03, 0.05]
20-29	1,614	1.79	[1.71, 1.88]	0.89	[0.83, 0.95]	1,852	2.05	[1.96, 2.15]	0.25	[0.23, 0.26]
30-39	1,835	1.65	[1.58, 1.73]	0.79	[0.74, 0.85]	9,416	8.47	[8.30, 8.64]	1	--
40-49	1,427	1.25	[1.19, 1.32]	0.60	[0.56, 0.64]	29,053	25.73	[25.44, 26.03]	2.99	[2.92, 3.06]
50-59	1,055	1.06	[1.00, 1.13]	0.49	[0.46, 0.53]	51,399	53.42	[52.96, 83.88]	6.16	[6.03, 6.30]
60-69	822	1.03	[0.96, 1.10]	0.47	[0.44, 0.52]	63,374	84.33	[83.67, 84.99]	9.66	[9.45, 9.88]
70-79	687	1.19	[1.11, 1.29]	0.55	[0.51, 0.60]	50,298	95.47	[94.64, 96.31]	11.16	[10.91, 11.41]
80+	280	0.80	[0.72, 0.90]	0.39	[0.35, 0.44]	21,752	67.32	[66.43, 68.22]	8.35	[8.15, 8.55]

***Rates per 10,000 person years, hazard ratios adjusted for age and gender**

7.4.2 Incidence by calendar year

Figure 7.10 illustrates the crude incidence of type 1 and type 2 diabetes in each calendar year between 1990 and 2013 in the CPRD. The incidence of type 1 remained largely constant over time, with a slight increase in 2006 only. The incidence of type 2 diabetes rose 2.5-fold between 1990 and 2003. Apart from a potentially artefactual increase in 2006, likely resulting from increased ascertainment associated with changes in the Quality and Outcomes Framework scheme, the incidence decreased between 2003 and 2013, from 39.29 per 10,000 person-years in 2003 (CI95% 38.69, 39.89) to 32.67 per 10,000 person-years in 2013 (CI95% 31.92, 33.43).

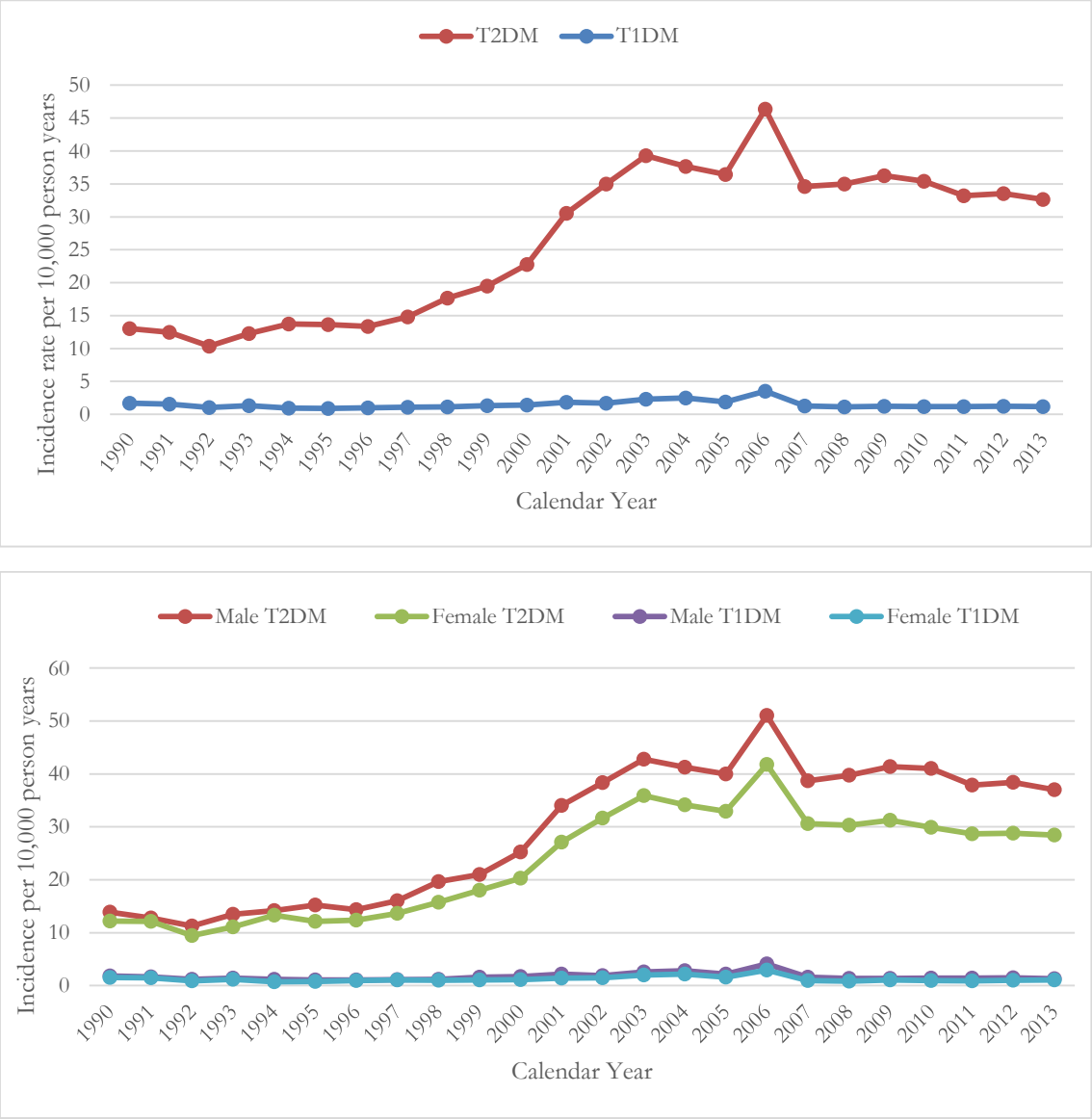


Figure 7.10 Incidence of type 1 and type 2 diabetes by calendar year and gender

The crude cumulative incidence curves for type 1 and type 2 diabetes by years of follow-up are shown in figure 7.11. For type 1 diabetics, the incidence was highest in the White population over the entire duration of follow-up for males. For females, the cumulative incidence in the Black African/Caribbean group overtook that of the White group after 15 years of follow-up. Amongst type 2 diabetics, cumulative incidence was consistently highest for the South Asian group and lowest for the White groups throughout the duration of follow-up.

7.4.3 Incidence by age group

The crude cumulative incidence curves for type 1 and type 2 diabetes by age are shown in figure 7.12. In the T1DM population, cumulative incidence was higher in the White population in both males and females at all ages. The difference between the White and non-White populations was more pronounced amongst females. In the T2DM population, cumulative incidence was highest for the South Asian population, followed by Black African/Caribbean and White populations for both genders. Disparities between ethnic groups increased with age, with incidence rising fastest in the South Asian population and slowest in the White population.

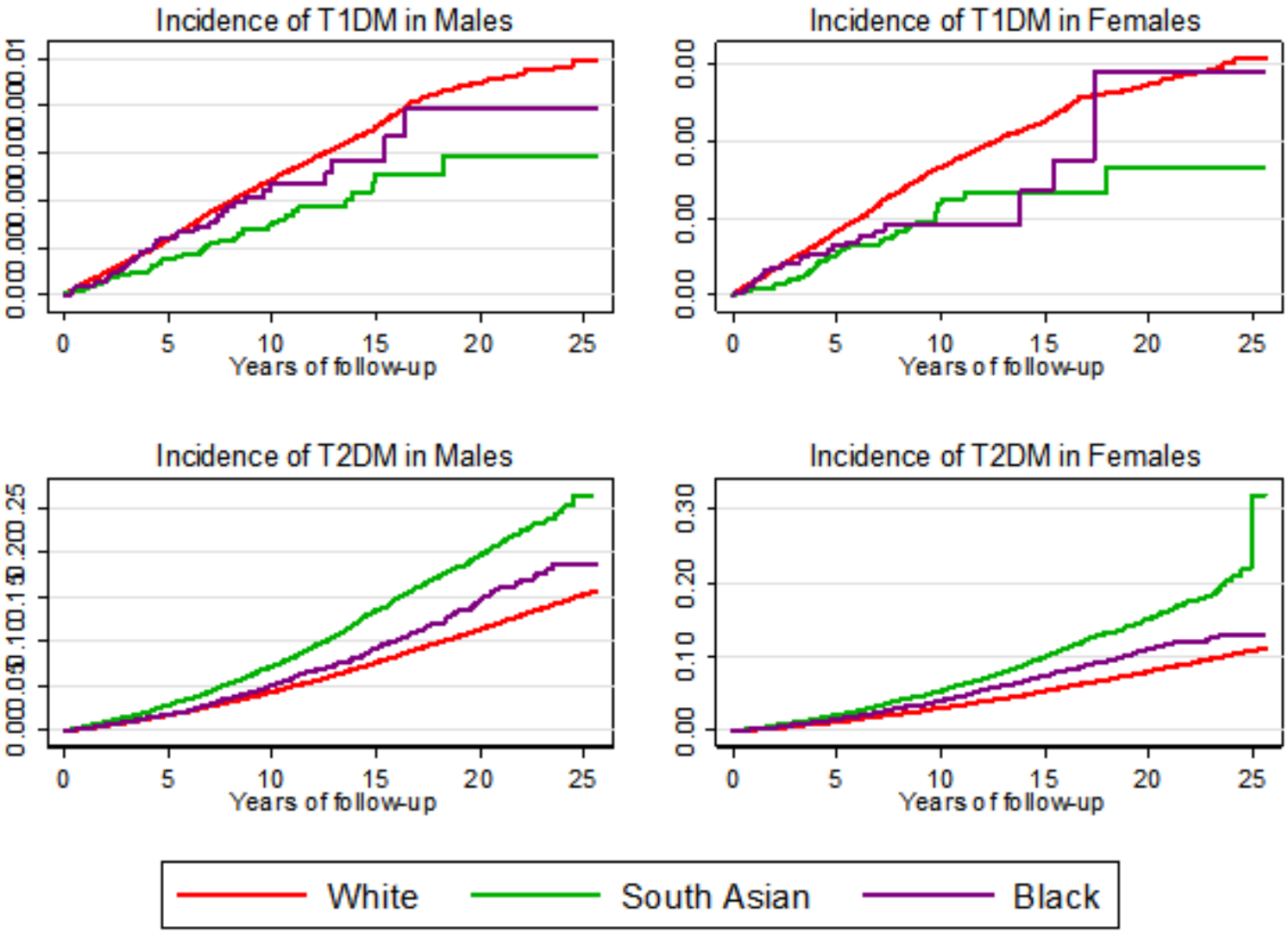


Figure 7.11 Cumulative incidence of type 1 and type 2 diabetes by calendar year and ethnic group

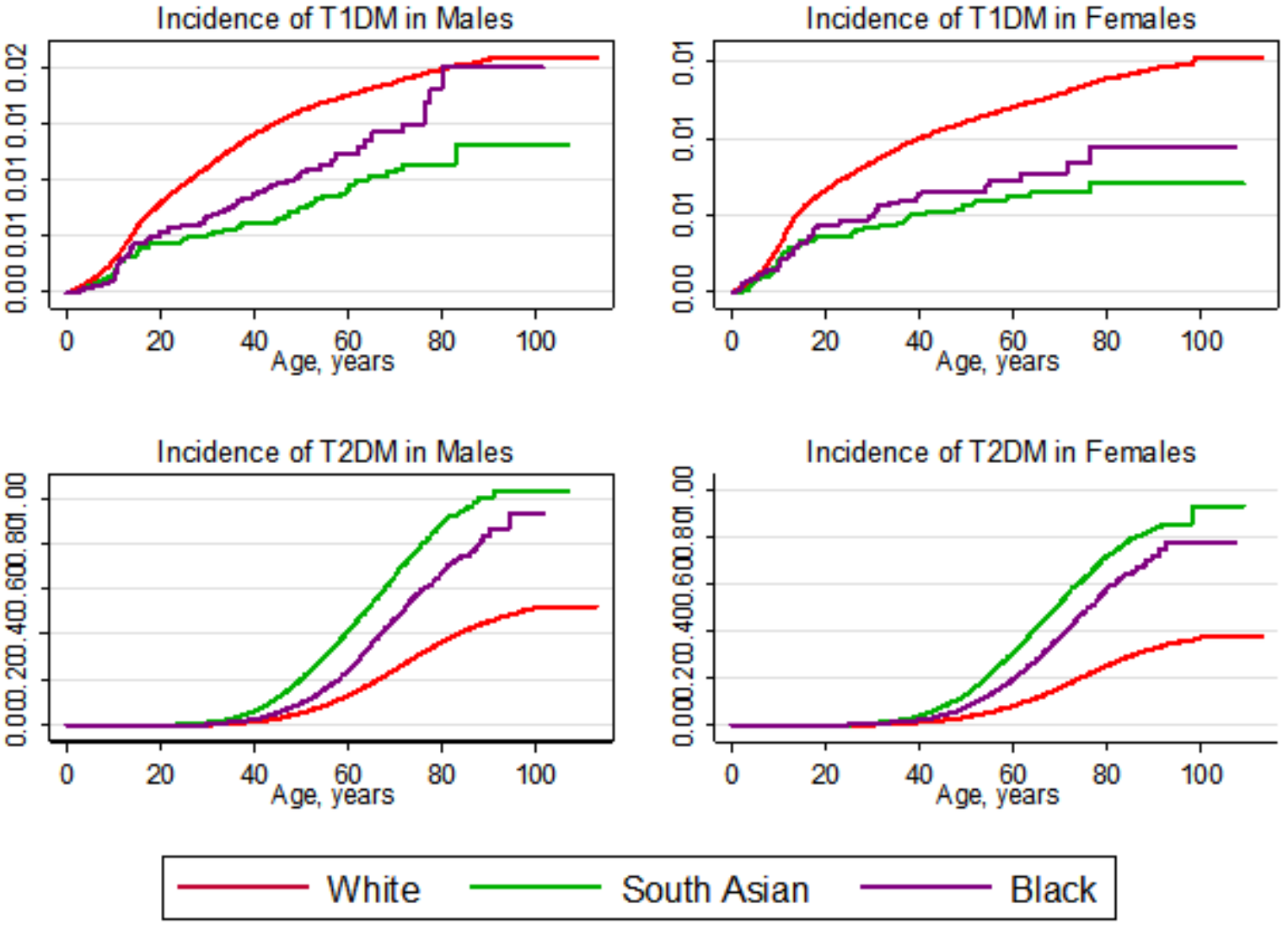


Figure 7.12 Cumulative incidence of type 1 and type 2 diabetes by age and ethnic group

7.5 Incidence in ethnic minority subgroups

The final objective of this study was to examine differences in the incidence of type 1 and type 2 diabetes for South Asian and Black/Caribbean subgroups. In order to identify heterogeneity in the risk of type 1 and type 2 diabetes between ethnic minority subgroups, crude rates and adjusted hazard ratios exploring risk in Indian, Pakistani, Bangladeshi, African and Caribbean groups were calculated and compared with the White population. The results are presented in table 7.6. Compared with the White population, the risk of type 1 diabetes was reduced in the Indian group by 40% (HR 0.60, CI95% 0.46, 0.78) and in the Pakistani groups by half (HR 0.52 CI95% 0.36, 0.75) after adjusting for age and gender. No differences in T1DM incidence between the White and Black African/Caribbean subgroups were evident.

For type 2 diabetes, risk was increased in all ethnic minority subgroups relative to the White population. Relative risk was highest in the Bangladeshi population, who were found to have an almost 6-fold increase in T2DM risk relative to the White population (CI95% 5.61, 6.28). This was followed by the Pakistani group, who had a 4-fold increase in risk relative to the White population (CI95% 3.84, 4.43), and the Indian group, who had a 2.6-fold increase in risk (CI95% 2.46, 2.66). Risk of T2DM was doubled in the African and Caribbean groups relative to the White population (African HR 2.00, CI95% 1.88, 2.13; Caribbean HR 2.15, CI95% 2.03, 2.27).

Table 7.6 Crude incidence rates and adjusted hazard ratios by ethnic minority subgroup

	Type 1 diabetes					Type 2 diabetes				
	Events	Crude rate	[CI95%]	Adj. HR	[CI95%]	Events	Crude rate	[CI95%]	Adj. HR	[CI95%]
Ethnic Group										
White	6,469	1.92	[1.87, 1.96]	1	--	129,770	39.56	[39.35, 39.78]	1	--
Indian	54	1.18	[0.90, 1.54]	0.60	[0.46, 0.78]	2.709	62.98	[60.65, 65.39]	2.56	[2.46, 2.66]
Pakistani	29	1.15	[0.80, 1.66]	0.52	[0.36, 0.75]	1.442	60.40	[57.36, 63.60]	4.05	[3.84, 4.43]
Bangladeshi	10	1.28	[0.69, 2.38]	0.59	[0.32, 1.10]	544	74.45	[68.49, 81.02]	5.77	[5.31, 6.28]
African	50	1.64	[1.24, 2.16]	0.79	[0.60, 1.04]	984	33.01	[31.01, 35.14]	2.00	[1.88, 2.13]
Caribbean	34	1.64	[1.18, 2.29]	0.84	[0.60, 1.17]	1194	61.42	[58.04, 65.01]	2.15	[2.03, 2.27]

***Rates per 10,000 person years, hazard ratios adjusted for age and gender**

The cumulative incidence of type 1 and type 2 diabetes for ethnic minority subgroups by age is shown in figure 7.13. In the T1DM population, cumulative incidence was highest in the Indian

population, followed by Pakistani and then Bangladeshi populations until age 70, after which point rates were highest in the Bangladeshi group. The pattern was reversed amongst women with type 1 diabetes, with rates highest in the Bangladeshi population and lowest in the Indian population. In the T2DM population, cumulative incidence was highest in the Bangladeshi population, followed by the Pakistani and Indian populations for both genders. Disparities between South Asian groups increase with age, with incidence rising fastest in the Bangladeshi population and slowest in the Indian population.

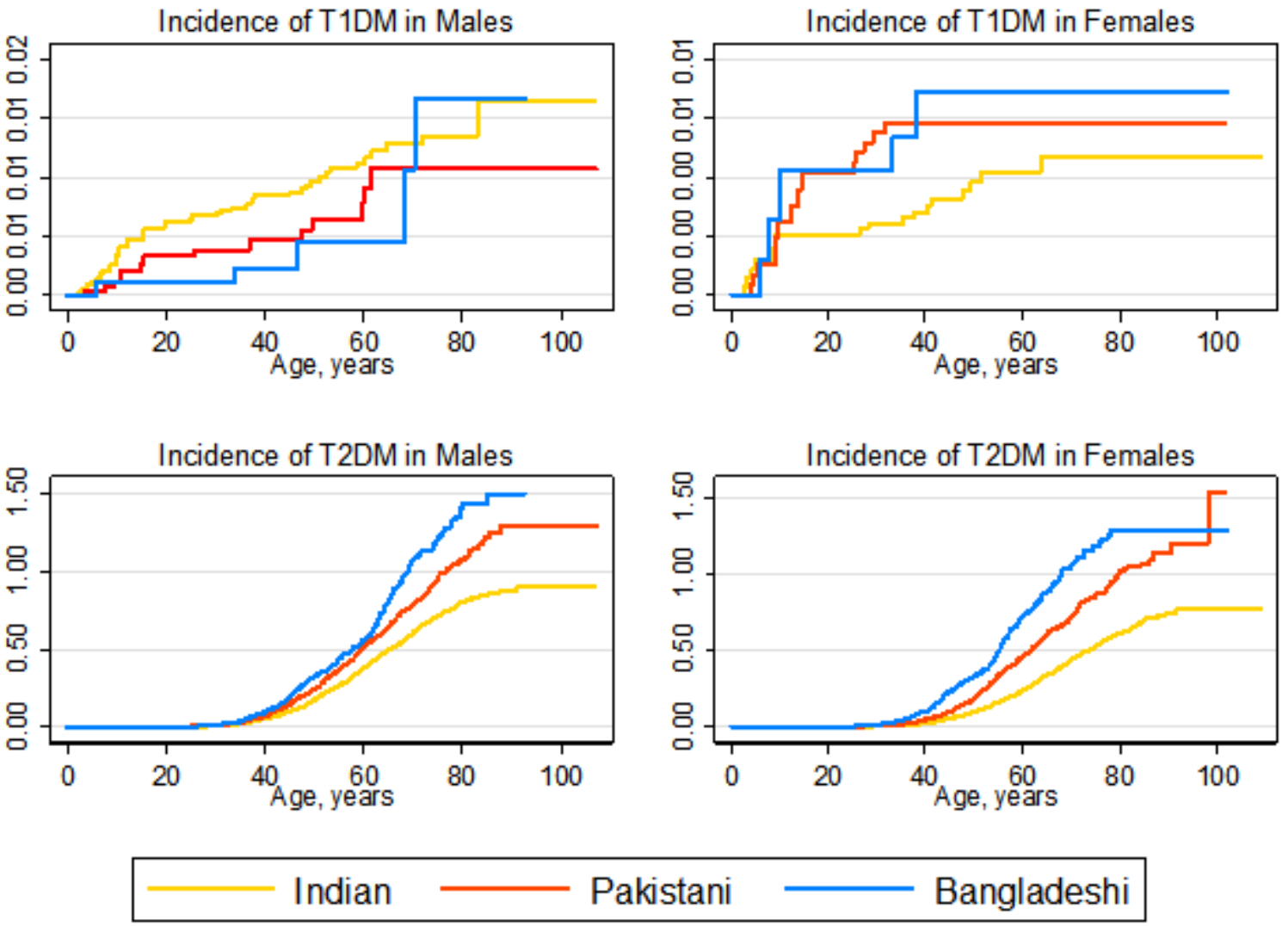


Figure 7.13 Cumulative incidence of type 1 and type 2 diabetes for ethnic minority subgroups

7.6 Discussion

The aim of this study was to explore how algorithms improving the classification of diabetes type perform in the CPRD and to quantify ethnic differences in the incidence and prevalence of type 1 and type 2 diabetes using these algorithms. This study represents the first implementation of these algorithms in a large database representative of the UK population as a whole, and also the first assessment of ethnic differences in diabetes burden for the entire cohort of patients contributing to the CPRD. The results show that it is feasible to implement the algorithms and that the resulting estimates of diabetes prevalence and incidence are comparable to those found in other UK-based studies.

7.6.1 Improving the coding and classification of diabetes mellitus

This study improved upon the original implementation of the algorithms in the Welsh SAIL database by utilizing the patient-level ethnicity data available in the CPRD to improve the identification of patients with type 2 diabetes. Were ethnicity data to be available in the SAIL database, it is possible that ethnic differences in diabetes prevalence would still be underestimated, as the Welsh database covers a much smaller and less ethnically diverse population than the CPRD. According to the 2011 Census, 14.5% of the English population identified themselves as non-white, compared with 4.4% of the Welsh population.

Compared with the original derivation cohort, a higher proportion of diabetics retained their initial classification. This difference may be attributable to the lack of full HES linkage in CPRD. In Flowchart 2, HES linked data is used to determine whether type 1 diabetics with a 6-month gap in their insulin prescription dates could have plausibly received their prescription during a visit to hospital. Individuals without a hospital visit should be reclassified as having T2DM. Since 50% of CPRD patients have no hospital data, they all retain their T1DM classification, as we cannot definitively say whether or not they attended hospital. Were HES linkage available for 100% of the CPRD population, we would identify a greater proportion of patients initially classified as T1DM who should be reclassified as T2DM due to having a gap of 6 months or more in their insulin prescription.

Since the completion of this analysis, a similar study of diagnostic coding error has been published regarding a cross-sectional on five general practices in Leicester.(228) Compared with our study sample of 391,994 patients diagnosed with diabetes between 1990 and 2013, the Leicester study identified 2,324 adults aged 17 and over diagnosed with diabetes in 2009/10. The overall prevalence of diabetes was 6.9% in 2011 in Leicester compared with 4.0% in 2012 in the CPRD, potentially reflecting the greater ethnic diversity of the study area in relation to the UK as a whole.

7.6.2 Type 1 diabetes

The study demonstrated a stable prevalence and incidence of type 1 diabetes over time, with a slight increase in 2006. Notable differences by ethnicity were evident, with prevalence in the White group more than double that in all other ethnic groups. Ethnic differences in incidence continued in the same direction, with the adjusted risk of T1DM reduced significantly in both the South Asian and Black African/Caribbean groups relative to the White group.

The UK has the fifth highest incidence of T1DM amongst children under 14 globally. A report by the International Diabetes Federation found that in 2013 the incidence of T1DM in this age group was 2.5 per 10,000 person years.(229) This is comparable to the findings of this study, where the incidence of T1DM was 2.06 per 10,000 person-years in the age group 0–9 and 2.96 per 10,000 person-years in the age group 10–19, average a rate of 2.51 across the two age groups.

7.6.3 Type 2 diabetes

The study identified substantial differences in the incidence and prevalence of type 2 diabetes by ethnic group, and considerable heterogeneity in risk between ethnic minority subgroups. Our findings of a 3-fold risk increase in the South Asian population and a 2-fold increase in the Black African/Caribbean population echo the results of several other UK-based studies, confirming that the CPRD population provides a substantial cohort to explore a range of research questions relating to population-wide trends in the differential burden of diabetes by ethnicity.(194,230,231)

The incidence of type 2 diabetes in the CPRD has been examined previously. In 2010 Holden et al. reported an overall incidence rate of 51.5 per 10,000 person-years, with significant increases over time. The study confirmed a higher incidence in males in all calendar years. Their study did not explore ethnic differences, but did report a secular increase in the proportion of patients being diagnosed with T2DM under the age of 40, likely attributable in part to the increasing proportion of South Asian and Black African/Caribbean individuals amongst diabetic cases.(232)

Our study has identified significant differences in the age of diabetes onset, confirming the hypothesis that onset will be earliest in the South Asian population and latest in the White population. The 10-year difference in median age at onset is comparable to that found in non-database studies of diabetes in the UK.

The study has confirmed that significant differences in diabetes risk exist between ethnic minority subgroups. If numbers allow, using a higher level of granularity to explore ethnic differences can reveal hidden heterogeneity between ethnic groups and contribute to a more nuanced understanding of the pathways linking ethnicity and health. Reasons for differences between South Asian subgroups are manifold. Indian, Pakistani and Bangladeshi groups migrated to the UK in different waves; as such, their current level of acculturation and subsequent health risk may differ.(37) Because the Indian population came to the UK first, on average, this group has a higher level of education and tends to live in more affluent areas. In contrast, because the Bangladeshi population is largely made up of much more recent migrants, this group tends to have a lower educational and professional level, and clusters in more deprived areas, such as inner east London.(28,233)

Furthermore, religious and cultural factors may play a role; Pakistani and Bangladeshi populations are largely Muslim, and this has an impact on dietary practices and acceptance of smoking and alcohol use.(40,234–236) As discussed earlier, the Health Survey for England has shown that while rates of smoking are higher in Bangladeshi men compared with other

men in England, over 90% of all Bangladeshi adults are non-drinkers.(237) Turning to alcohol use, a large proportion of Muslim adults report abstaining from alcohol use, though those who do drink alcohol tend to have higher consumption than other groups.(40)

7.6.4 Reasons for observed increases in incidence and prevalence in type 2 diabetes

Firstly, the apparent rise in incidence may be due to an actual increase in the number of people developing diabetes in the UK population. This in turn could be due to rising BMI in the population and a growth of ethnic minority populations who are contributing disproportionately to the diabetes burden. This study has demonstrated that the incidence of type 2 diabetes is higher in ethnic minority groups than in the White population. Additionally, the incidence may have increased due to changes in the diagnostic criteria for identifying type 2 diabetes adopted in the UK, which have lowered the threshold for fasting plasma glucose, resulting a larger cohort of T2DM patients based on a broader inclusion criteria, reflecting newer evidence that diabetic complications occur at lower levels of glucose impairment.(238–240)

Secondly, the increasing prevalence and incidence of type 2 diabetes may stem from the combined influences of the lowering in age at first diagnosis, as demonstrated in this study, and improved survival after diagnosis. Metformin, which is recommended as the first-line pharmacological treatment for type 2 diabetes, has been shown to improve survival for patients with type 2 diabetes, primarily by reducing the risk of cardiovascular outcomes.(241–243)

Thirdly, improved case ascertainment via initiatives such as the NHS health checks programme, local enhanced services and the Quality & Outcomes Framework may have contributed to the increase in the reported prevalence and incidence rates of type 2 diabetes. Such initiatives improve opportunities to detect disease by increasing the number of patient–GP encounters and by increasing the amount of targeted activity by practice staff.(244–247) This increase in ascertainment serves to reduce the number of people with undiagnosed

diabetes in the population. The combined influence of these concurrent initiatives may have contributed to the smaller ethnic differences reported here compared with earlier UK-based studies due to improving equity of care for all patients across the UK.

Finally, temporal changes in the diagnostic coding of diabetes mellitus in primary care may also contribute to the increase. Though the recording of any diabetes mellitus was incentivised under the Quality and Outcomes Framework in 2004, this was changed in 2006 to reward only the coding of type 1 and type 2 diabetes mellitus. More specific diagnostic coding, as evidenced by the spike in the incidence of both type 1 and type 2 diabetes observed in 2006.

7.6.5 Limitations

The primary limitation of this study is the reliance on date of first recorded diagnostic code to approximate the date of diabetes onset. It has been shown that approximately 7 years can elapse between the onset of disease and its identification by health professionals; thus duration of disease and number of incident events is likely underestimated in our sample, due to the presence of undiagnosed diabetes in our study population.(248) In the UK it is estimated that 1% of the population, or one in three diabetics, has disease that is either undiagnosed or unrecorded in primary care.(249,250) The prevalence of undiagnosed diabetes in the UK has reduced in recent years, likely via the dual actions of increased awareness and concentrated efforts by GPs to create registers of diabetic patients to ensure appropriate access to care and treatment. (251)

Secondly, this study did not attempt to identify patients in the CPRD without diagnostic codes for diabetes. It is possible that patients may have evidence on their primary care record which supports a diagnosis of diabetes, such as prescriptions, raised blood glucose or 'process of care' codes, but do not have an appropriate diagnostic code which would identify them as being on a general practice's diabetic register. A study in the QRESEARCH database found that it is possible to identify undiagnosed diabetes using routine primary care databases via

the recording of blood glucose data, and it reported that 30% of all people aged 40 and over had a blood glucose value recorded in the preceding two years.(252)

Thirdly, secular improvements in the recording of ethnicity and covariate information such as glycaemia and BMI mean that individuals registered more recently are more likely to have their diabetes diagnosed in a timelier manner. Furthermore, ascertainment of correct diabetic type may be more accurate in these individuals due to the greater completeness of supporting information required to adjudicate diabetic outcomes. Historic patients contributing to the study may be more likely to have errors or missing data, particularly if diagnosed prior to the incentivisation of diabetes coding.

Finally, the lack of HES linkage for all patients registered with the CPRD means that we are able to determine whether gaps in insulin prescriptions are plausible, thus limiting our ability to appropriately retain T1DM classification or reassign those with non-continuous insulin to T2DM.

Chapter 8
Ethnic and gender differences in risk of coronary heart disease amongst patients with and without type 2 diabetes: Methods

8.1 Summary

This chapter describes the methods and background of the third and final study reported in the thesis, that examining differences in the incidence of first non-fatal coronary heart disease (CHD) by ethnic group, gender, deprivation and diabetic status. All patients with type 2 diabetes identified in the previous chapter free from CHD at diagnosis were eligible for inclusion. A random sample of CPRD patients free from both CHD and type 2 diabetes aged 30 and above was selected as the comparison group. Ethnic and gender differences in incidence of CHD and cardiac death were estimated for individuals with and without diabetes using multivariable Cox regression. The results for non-fatal CHD are presented in chapter 9. The results for fatal and non-fatal CHD combined are presented in chapter 10. A discussion of both chapters is presented in chapter 11.

8.2 Study objectives and hypotheses

The main objectives were to:

1. Identify how much excess CHD risk is conferred by having diabetes and whether there is an interaction with age, gender, deprivation and ethnic group.
2. Quantify ethnic and gender differences in crude incidence of fatal and non-fatal CHD amongst patients with and without type 2 diabetes
3. Quantify ethnic and gender differences in the adjusted risk of fatal and non-fatal CHD amongst patients with and without type 2 diabetes
4. Break down estimates of adjusted risk further by South Asian subgroup (Indian, Pakistani and Bangladeshi) and Black subgroup (African, Caribbean) in comparison with the White population as a whole.

The hypotheses of this study were:

1. That the South Asian group will have a higher risk of incident fatal and non-fatal CHD and the Black African/Caribbean group a lower risk than the White group
2. That the excess risk for the South Asian group will be greater in the diabetic population than in the non-diabetic population.
3. The risk of CHD will be higher in diabetics compared with non-diabetics both overall and by age, gender, ethnic group and deprivation.
4. In the ethnic subgroup analysis, CHD risk will be raised for all South Asian subgroups, and, in particular, be highest for the Pakistani group in comparison with the White group.

8.3 Background

The term coronary heart disease (CHD) refers to diseases of the arteries which supply oxygenated blood to the heart. CHD develops via the process of atherosclerosis, whereby plaques composed primarily of fat and calcium build up in the blood vessels. As these harden over time, they can narrow the arteries and restrict blood flow to the heart, resulting in angina or myocardial infarction.(253)

The causal pathway towards coronary heart disease is multifactorial. The majority of CHD risk stems from modifiable risk factors.(254) The key risk factors for atherosclerosis are dyslipidaemia, type 2 diabetes mellitus, hypertension and obesity.(255,256)

CHD and stroke combined form the largest grouping of cardiovascular disease (CVD), which is the leading cause of death in the UK and worldwide. CVD contributes to over half of the entire global burden of non-communicable disease and in 2010 accounted for 29.6% of all deaths globally.(257),(258) Data from 2012 show that 46% of all deaths in Europe were CVD related, 20% of which were attributable to CHD.(259)

Coronary heart disease was first described by Scottish clinicians Heberden, in 1772, and Jenner in 1788.(260) Throughout the late 1800s and early 1900s, clinical evidence of myocardial infarction and arterial occlusion contributed to the understanding of the atherosclerotic process resulting in CHD, and in 1930 CHD was added to the International Classification of Diseases (ICD).(261),(262)

The study of CHD heralded the beginnings of modern non-communicable disease epidemiology, which recognised that chronic disease risk is mediated by the interplay of multiple risk factors rather than a single aetiological element.(263) To this end, longitudinal cohort studies and epidemiological surveillance programmes were set up to monitor global changes in CHD prevalence and the contributions of prevention and treatment to risk factor reduction. Prominent examples of these include the Monitoring Trends and Determinants in Cardiovascular Disease (MONICA) Project, which was established by the World Health Organization in 1979, and the Framingham Heart Study, which was established in 1948.(263,264)

By the mid-20th century, the prevalence of CHD had risen to epidemic proportions, primarily across the United States and northern Europe, quickly overtaking infectious diseases such as tuberculosis.(265) In the United States, the epidemic peaked in the 1960's, at which time 1 in 3 deaths was caused by CHD.(265) In the UK, the epidemic reached its zenith ten years later; Mortality rates increased exponentially between 1921 and 1970, peaking at 730 deaths/100,000 in England and Wales, and 960/100,000 in Scotland for men aged 55-64.(261)

Since 1970, CHD mortality rates have declined dramatically across the developed world. Between 1980 and 2006, the UK had a 41% reduction in mortality rates from myocardial infarction, the largest across all of Europe. In 2010, the age standardized mortality rate for CHD was 111.1/100,000 in men and 49.4/100,000 in women and is predicted to continue to decline through 2016.(259,266)

Since the 1980's a substantial epidemiological transition has taken place. As CHD mortality has declined in developed countries, it has increased rapidly in low- and middle-income countries, where now 80% of all CHD-related deaths occur.(257,258) CHD death in these regions occur approximately 10 years earlier than in Europe and north America, due in part to earlier onset of disease and poorer healthcare infrastructure.(254)

This shift has been driven concurrently by increasing life expectancy and the adoption of western life styles, increasing the prevalence cardiovascular risk factors, and the lack of adequate population-based strategies to reduce the impact of these risk factors on incidence and mortality.(258) In 2012 the World Health Assembly introduced a target to reduce global mortality from non-communicable diseases by 25% by the year 2025, with the prevention of cardiovascular disease at the heart of the programme.(258) The proposed strategy focuses on the key risk factors of obesity, hypertension, hypercholesterolaemia, tobacco smoking, alcohol consumption and dietary fat and salt intake.(258)

8.3.1 Coronary heart disease and diabetes mellitus

A key driver of the global rise in CHD is diabetes mellitus. 75% of the global burden of diabetes stems from low- and middle-income countries. Diabetes and CHD share many of the core risk factors, and are suspected to have a common genetic basis.(267) Diabetes mellitus triggers changes in endothelial function and energy metabolism which promote atherosclerosis; The atherosclerotic process tends to initiate earlier in diabetic individuals and progress faster, resulting in a 2- to 4-fold increase in CHD events and CHD-related mortality in this group.(268–271) CHD is responsible for 80% of all deaths amongst individuals with diabetes and is the leading cause of morbidity in this population. Individuals with hyperglycaemia, diabetes or metabolic syndrome have up to three times the risk of cardiovascular mortality compared with those without metabolic abnormalities.(268,269) Correspondingly, diabetes is considered to be a CHD equivalent and thought to confer a cardiovascular risk equal to that of having established CHD.(272)

8.3.2 Coronary heart disease and ethnicity

It has been established that traditional risk factors for heart disease such as smoking, hypertension, diabetes and obesity increase risk in all populations, but that the magnitude of effect and interactions between the risk factors differ according to ethnic group.(273) Both prevalence and severity of CHD are known to vary by ethnicity, with marked differences in incidence and mortality demonstrated globally. In the UK, incidence and mortality from CHD are increased amongst South Asians but decreased amongst the Black African/Caribbean population compared with the White population.

Broadly speaking, South Asian groups as a whole carry a higher burden of cardiovascular risk factors and benefit from fewer protective factors.(192,193,236,274) Incidence of CHD is higher in South Asian populations, both in the host countries and in the diaspora.(275) In India, CHD is the leading cause of death. A meta-analysis of 9 studies conducted in 2013 demonstrated an increase in CHD incidence of 35% in South Asian populations compared with White populations (HR 1.35, CI95% 1.30,1.40).(276,277)

a) Mechanisms by which diabetes increases CHD risk amongst South Asians

Ethnic differences in CHD are exacerbated by diabetes, which has been shown to have a greater impact on CHD risk amongst South Asians compared to White populations. Though conventional risk factors account for over 80% of all CHD risk in all ethnic groups, they do not explain the 40% excess of CHD observed amongst South Asians.(278)

Insulin resistance is well established as the underlying mechanism responsible for the increased rates of CHD amongst South Asians globally, and has been found to account for 70% of the excess risk of CHD in South Asians compared to Europeans.(269) Two prospective studies of CHD mortality based in London (SABRE and LOLIPOP) have confirmed that the prevalence and incidence of CHD are doubled in South Asian groups, even after accounting for cardiovascular risk factors, further supporting the independent effect of diabetes and metabolic disturbances on CHD risk.(279)

Though South Asian groups have a higher CHD incidence, they have been found to have better survival. As such, increased mortality amongst South Asians has been attributed to a greater burden of modifiable risk factors rather than poor management after a coronary event, suggesting that the most effective way to reduce ethnic inequalities in coronary disease outcomes is to focus public health efforts on primary care.(276,280)

Despite having higher rates of stroke, Black African/Caribbean groups in the UK have lower rates of CHD and CHD-related mortality than White and South Asian groups.(281) Though both South Asian and Black African/Caribbean populations have higher levels of insulin resistance than the White majority population, lipid profiles and levels of obesity and central adiposity are more favourable in Black African/Caribbean groups compared to South Asians.(282)

Amongst South Asian groups, CHD tends to develop earlier and at lower levels of risk compared to White populations. In India, 50% of CHD-related deaths occur before age 70, compared to 6% in high-income countries.(283) Furthermore, as a result of higher levels of central adiposity, South Asian groups tend to develop CHD at lower levels of BMI and smaller waist circumference; thus, CHD events are increasing amongst people without obesity as defined by BMI.(284)

b) Differences between South Asian subgroups

Combining data for South Asian groups can be misleading – ethnic and racial differences are almost never demonstrably genetic and therefore genetic and social environment are likely to be crucial. As highlighted in the introduction of this thesis, ethnicity serves as a useful concept which encompasses a range of factors which cannot be reduced simplistically to genetic differences. In the UK context, it has been demonstrated that Pakistani populations are at high risk of CVD compared with other South Asian ethnic groups, with studies demonstrating up to 60% increased risk of first myocardial infarction (MI).(281,285)

CHD prevalence and incidence have both been found to be higher in Pakistani and Bangladeshi groups and lower in Indians.(26,286,287) A study by Bhopal et al. found that Bangladeshi groups have the highest risk profile for CHD compared to the other South Asian subgroups, and that South Asians as a whole have a higher risk profile for CHD compared to Europeans.(236) A Scottish study of acute MI demonstrated that incidence was highest for Pakistani men and lower other South Asian subgroups, White and Chinese groups. These differences were not present amongst women.(280)

While smoking and lack of exercise are more common amongst Bangladeshi and Pakistani men, alcohol use is highest amongst Indians.(236,280,288) Beneficially, Pakistani and Indian groups are more likely to eat fruits and vegetables daily. Obesity is less common in the Bangladeshi subgroup than in the other two subgroups. There are no significant differences between South Asian subgroups with respect to hypertension.(236)

8.3.3 Coronary heart disease and gender

The incidence of CHD has been shown to be lower in women than in men, with onset on average 10 years later.(289–291) CHD mortality in women is reduced 5-fold, though gender differences diminish after age 70.(292) As with ethnicity, the relationship between gender and CHD is also modified by the presence of diabetes mellitus; the increase in CHD risk associated with diabetes has been shown to be higher for women than men, cancelling out any cardioprotection evident amongst non-diabetics.(293),(294)

8.4 Study Design

A prospective cohort study was undertaken to examine:

a) Incidence of non-fatal CHD in:

i) All patients with T2DM

ii) A random sample of patients without T2DM

iii) A sample of age matched patients without T2DM

b) Incidence of fatal and non-fatal CHD in

i) All patients with T2DM whose records were linked to ONS data, enabling ascertainment of date and cause of death

ii) A random sample of patients without T2DM whose records were linked to ONS data.

8.5 Participants

8.5.1 Patients with Type 2 diabetes

From the total cohort of patients defined as having type 2 diabetes in chapter 6, a subset meeting the following criteria were retained for analysis:

a) Aged ≥30 at date of T2DM diagnosis

b) Diagnosed between January 1st 1990 and August 1st 2013

c) Free from CHD at time of T2DM diagnosis

d) Registered for a minimum of 12 months prior to T2DM diagnosis in order to capture incident diabetics diagnosed during registration with the CPRD, for whom onset date and classification of diabetes type is likely to be more accurate and to exclude prevalent diabetics with unknown type and duration of diabetes.(295)

8.5.2 Patients without Type 2 diabetes

a) Random sample

A random sample of 1 million patients aged 30 or over in August 2013 was extracted from the CPRD. From this sample, all patients with a diagnosis of T2DM were removed. A subset of patients free from CHD at age of 30 was retained. Information on the day and month of birth

was not provided. Therefore, each patient's date of birth was estimated as 1st July of their year of birth, allowing a maximum error margin of six months for age.

b) Age matched sample

A second sample of age matched patients without type 2 diabetes was extracted from the CPRD. Up to four controls for each case were identified in the database. A maximum difference in age of 1 year was tolerated in order to maximize the number of successful matches. As previously, each patient's date of birth was estimated as 1st July of their year of birth, allowing a maximum error margin of six months for age. The start of follow-up for the non-diabetic sample was matched to that of the diabetic population.

The matched sample of patients without type 2 diabetes was used for in the primary analysis examining the incidence of non-fatal CHD in the whole CPRD population, and not for the secondary analysis of combined fatal and non-fatal CHD in the subset of patients with linked ONS mortality data.

8.6 Outcome definition

The outcome of the primary study was incidence of first non-fatal coronary heart disease between January 1990 and August 2013, defined using the Quality and Outcomes Framework definition of a clinical diagnosis of ischaemic heart disease, myocardial infarction or angina pectoris (Table 8.1). The first incident non-fatal CHD during the study period was identified from the patient's CPRD clinical file. Individuals with a CHD event prior to January 1990 were excluded from the analysis.

For the secondary analysis, the outcome of the study was combined fatal and non-fatal CHD for the subset of CPRD patients with linkage to ONS mortality data. Codes for fatal CHD were identified from the ONS record using the ICD-9 and ICD-10 classification systems (Table 8.1). The code list was based on those used by the CALIBER study group and modified to add

several additional codes identified in recent studies.(296) In this population the date of first fatal or non-fatal CHD event during the study period was identified.

Table 8.1 Medical codes for non-fatal CHD from CPRD and fatal CHD from ONS Mortality data

Read Codes for non-fatal CHD found in the CPRD	
G3...00	Ischaemic heart disease
G30..00	Acute myocardial infarction
G31..00	Other acute and subacute ischaemic heart disease
G32..00	Old myocardial infarction
G33..00	Angina pectoris
G34..00	Other chronic ischaemic heart disease
G35..00	Subsequent myocardial infarction
G38..00	Postoperative myocardial infarction
G39..00	Coronary microvascular disease
G3y..00	Other specified ischaemic heart disease
G3z..00	Ischaemic heart disease NOS
Gyu3.00	[X]Ischaemic heart diseases
ICD Codes for Fatal CHD found in ONS Mortality data	
I20	Angina pectoris
I21	Acute myocardial infarction
I22	Subsequent myocardial infarction
I23	Certain current complications following acute myocardial infarction
I24	Other acute ischaemic heart diseases
I25	Chronic ischaemic heart disease
I46	Sudden cardiac death
I49	Death due to ventricular fibrillation
410	Acute myocardial infarction
411	Other acute and subacute forms of ischemic heart disease
412	Old myocardial infarction
413	Angina pectoris
414	Other forms of chronic ischaemic heart disease
4292	Ill-defined descriptions and complications of heart disease; Cardiovascular disease, unspecified
4295	Ill-defined descriptions and complications of heart disease; Rupture of chordae tendineae
4296	Ill-defined descriptions and complications of heart disease; Rupture of papillary muscle
4297	Ill-defined descriptions and complications of heart disease; Certain sequelae of MI

8.7 Covariate definition

Both time-constant and time-varying covariates were extracted for all subjects meeting the inclusion criteria.

8.7.1 Time constant covariates

Ethnicity, gender and deprivation were constant for each patient. For the main analysis ethnicity was categorized using the high-level Census groupings of White, South Asian, Black,

Mixed, Other and Not Stated. In this chapter, results for White, South Asian and Black groups only are reported.

For the secondary analysis South Asian and Black groups were further subdivided into Indian, Pakistani, Bangladeshi, African and Caribbean, with rates and risk of CHD compared to the White population as a whole.

Practice-level Index of Multiple Deprivation (IMD) data were available for all patients and used in the primary analysis of non-fatal CHD. Individual-level IMD was available for the subset of patients belonging to practices with linkage to deprivation data and used only in the analysis of fatal and non-fatal CHD for patients with linked ONS Mortality data.

8.7.2 Time-varying covariates

a) Medications

For both diabetic and non-diabetic cohorts, the first ever prescription of antihypertensives, lipid-lowering medications and aspirin were extracted. For diabetic patients, first prescription of insulin, metformin or other antidiabetic drugs were additionally extracted. Exposure status was denoted by "0" for unexposed and "1" for exposed. Follow-up time prior to first prescription of each drug was denoted using "0". This value switched to "1" on the date of first prescription. Patients whose first prescription occurred before the start of follow-up were classified as "1" throughout.

b) Clinical measures

V. Body mass index

Body mass index values were either taken directly from the value associated with the clinical Read code for BMI (22K) or calculated using height and weight values recorded in the CPRD additional file using the equation (weight/height^2). Implausible values for height (below 1.37 m and over 2.3m) were removed. Values over 100 which could reasonably be considered to

have been measured in centimetres were converted to metres. Similarly, implausible values for weight (<25.4kg or >254kg) and duplicate values recorded on the same day were removed. If a patient had multiple height or weight values recorded on the same day, the difference between the smallest and largest values was calculated; if the difference was less than or equal to 5 cm for height or 2 kg for weight, the average of the values was taken. If the difference was any larger, the records were dropped.

BMI was calculated for all dates when both height and weight values were available. For dates when BMI could not be calculated, BMI values as recorded directly into the CPRD were substituted, resulting in a longitudinal record for each patient indicating all BMI values recorded during their time in the CPRD. BMI values below 15 or over 50 and duplicate values recorded on the same day were dropped.

For analysis, the latest BMI recorded prior to follow-up and all values during follow-up were retained. BMI values were grouped into four categories of underweight, normal weight, overweight and obese, as defined by the World Health Organization:(297)

I. Underweight (15–18.4 kg/m^2)
II. Normal weight (18.5–24.9 kg/m^2)
III. Overweight (25–29.9 kg/m^2)
IV. Obese (≥30 kg/m^2)

VI. Glycated haemoglobin

Glycated haemoglobin (HbA1c) values were obtained from the Test results file of the CPRD and converted to % if recorded as mmol/mol. All HbA1c values greater than 20% and duplicate values recorded on the same day were removed. For analysis, the latest HbA1c recorded prior to follow-up and all values during follow-up were retained. HbA1c values were grouped into three categories of <5.4%, 5.5-6.4% and 6.5-20%. According to the UK National Institute for Health and Care Excellence (NICE), the target value for blood glucose control is 6.5%.(298)

VII. Total Serum Cholesterol

Total serum cholesterol values were obtained from the Test results file of the CPRD. Implausible values (<2 mmol/L or >10 mmol/L) and duplicate values recorded on the same day were removed. For analysis, the latest cholesterol value recorded prior to follow-up and all values during follow-up were retained. Cholesterol values were dichotomized into two groups indicated by "0" for values ≤5mmol/L and "1" for values >5mmol/L.

VIII. Hypertension

All values for systolic and diastolic blood pressure were obtained from the additional file of the CPRD. Implausible values for systolic blood pressure (<50 mmHg or >250 mmHg) and diastolic blood pressure (<35 mmHg or >140 mmHg) and duplicate values recorded on the same day were removed. For each date of blood pressure recording, an indicator for clinical hypertension was created to equal "0" if blood pressure was ≤140/90mmHg and "1" if blood pressure was >140/90 mmHg, as defined by the NICE guidelines.(299)

c) Lifestyle measures

IX. Tobacco Consumption

Clinical codes for tobacco consumption falling under the Read code hierarchy of 137% were categorized into three groups of "Non-smoker", "Current smoker" and "Ex-smoker". Patient's smoking status changed over follow-up each time their category changed. For example, several patients with smoking status recorded alternated between smoker and non-smoker over the course of follow-up.

X. Alcohol consumption

Alcohol usage was derived either from the Read codes for alcohol intake or values for "Alcohol units per day" from the additional file of the CPRD. Values over 1,000 and duplicate values recorded on the same date were removed. Read codes and alcohol values were organised into three categories of "Non-drinker", "Moderate drinker" and "Heavy drinker", as used by the CALIBER study group.(300)

8.8 Statistical Analysis

All analyses were performed using Stata MP, release 13.(301)

8.8.1 Descriptive analyses

Characteristics of patients, including demographics and clinical characteristics, at baseline and 'ever usage of medications' was summarised for patients with and without type 2 diabetes separately. For categorical variables, proportions were calculated; for continuous variables, means and standard deviations were calculated. Medians were calculated for continuous variables with a skewed distribution.

8.8.2 Multivariable survival analysis

a) Differences in CHD risk by diabetic status

In order to quantify differences in the risk of first non-fatal CHD between patients with and without type 2 diabetes, the diabetic and non-diabetic denominator populations were combined. Interactions between diabetic status and age, gender, deprivation quintile and ethnic group were explored.

b) Differences in CHD risk by ethnic group

In order to compare the incidence of first CHD event by ethnic group, follow-up time was limited to each patient's period of "Up to standard" (UTS) observation, as described in chapter 3. For the diabetic cohort, the start of observation time was defined as being the latest of:

i) date of T2DM diagnosis

ii) 1 year after registration date

iii) 'practice up to standard' date.

For the non-diabetic random sample, the start of observation time was defined as being the latest of:

i) turning age 30

ii) registration date

iii) 'practice up to standard' date.

For the non-diabetic matched sample, the start of observation time was equal to that of the matched individual with type two diabetes.

In all cohorts, follow-up time ended at the date of first CHD event. In the secondary analysis of fatal and non-fatal CHD in patients with linked data, follow-up time ended at the date of first CHD event or cardiac death. For patients not experiencing a CHD event, follow-up time was censored at the earliest of transfer out date, last collection date and death from all causes or August 1st 2013. Follow-up time for those not experiencing the event of interest was censored at the earliest of transfer out date, last collection date, death from other causes, or August 1st 2013.

c) Analysis of crude incidence

Crude incidence rates of first CHD per 10,000 person years of follow-up time were calculated for patients with and without T2DM. All analyses were stratified by gender and ethnic group. Cox proportional hazards regression using time since study entry as the timescale was used to evaluate the risk of developing a first CHD event in:

I. all patients with T2DM diagnosed between January 1990 and August 2013
II. a random sample of patients free from T2DM who were aged 30 or older between January 1990 and August 2013.

d) Analysis of adjusted risk

Multivariable Cox proportional hazards regression was used to examine the relative risk of fatal and non-fatal CHD in all ethnic groups relative to the White reference population. Four pre-specified survival models with increasing levels of covariate adjustment were compared for each population of interest:

1) Crude model for risk of CHD by ethnic group
2) Additionally adjusted for age

3) Additionally adjusted for component factors of the Framingham Risk Score(302) (Blood pressure, smoking status, and total cholesterol). Component measures from the Framingham Risk score were selected because they were well recorded in the database. More current risk models such as the QRISK score for ten year cardiovascular risk were not used due to poorer completeness of key variables such as family history of cardiovascular disease, and cholesterol:HDL ratio.

4) Additionally adjusted for index of multiple deprivation, medication use and additional confounders well characterized in the CPRD. Models for both diabetic and non-diabetic cohorts included adjustment for BMI, alcohol consumption and first ever prescription of antihypertensives, lipid lowering medications or aspirin. Models for patients in the diabetic cohort further adjusted for HbA1c category and first prescription of insulin, metformin and other antidiabetic drugs. Models for the primary analysis of non-fatal CHD in the full CPRD cohort adjust for practice level IMD score, while those for the analysis of fatal and non-fatal CHD in the subset of patients with linked data adjust for patient-level IMD score.

8.8.3 Clustering by practice

In order to improve the accuracy of the statistical inference, regression analyses for all three study populations accounted for the effects of clustering by practice by using cluster robust standard errors. This accounts for the fact that the individuals contributing to the analysis are not independent of one another, but rather share some variation related to the practice that they attend. For example, health care delivery may differ between practices depending on local population needs, local financial incentives, and local care delivery schemes. Therefore patients within a single practice may share experiences which differ from patients attending different practices.

Failure to account for clustering can result in biased estimates of standard error and narrow confidence intervals. Accounting for clustering increase the standard errors and widens the 95% confidence intervals to better reflect the non-independence of observations.(303) Clustering of the errors does not change the estimate of the hazard ratios, but allows for the errors to be correlated within clusters of practice and uncorrelated across practices.

The use of hierarchical, or multilevel modelling methods using a random effects term allowing for patients to be clustered within practices was not suitable for this study as the shared frailty model (which models random effects In survival analyses) is not allowed when patients enter the study at different times (delayed study entry), as is the case with this analysis. Furthermore, adjusting for practice as a categorical, time independent variable in the regression models was not possible due to computational limits on matrix size in Stata, which would not allow for the addition of 640 indicators to the analyses.

8.8.4 Approaches to handling missing data

Apart from the demographic variables of year of birth and gender, all other Read-coded variables are missing to some degree in the CPRD. Missing data compromise the ability to make causal inferences by creating bias, causing a loss of statistical power, or by reducing the precision of estimates of effect size when restricting analyses to include patients with complete data (complete case analysis).

In routine electronic health records, the input of information onto the patient record is dependent on a combination of factors relating to the patient, practitioner and service provision infrastructure. In databases such as the CPRD, where large patient populations are available for study, the missing data can potentially lead to information bias. The impact of information loss and bias depends both on the research question of interest and the ways in which the observed data are selected – known as missing data mechanisms. In 1976, Rubin identified three main missing data mechanisms:(304)

1. Missing completely at random: The missing values and observed values do not differ systematically and the reason for the data being missing is unrelated to the outcome of interest.
2. Missing at random: Missing values differ systematically from observed values, but the reasons for this difference can be explained by the other variables.
3. Missing not at random: Missing values differ systematically from observed values and the reason is related to the outcome of interest.

Two methods for dealing with missing data include using complete case analysis or replacing missing values with imputed values. Multiple imputation methods are appropriate for situations in which the mechanism is missing at random, or approaching missing at random. For the purposes of this study, multiple imputation was found to be inappropriate because several of the variables included in the study are likely to be missing not at random, and because reasons for 'missingness' are likely to differ depending on the covariate of interest. While measures such blood pressure and BMI are likely to be measured regularly as part of a health check, such health checks may occur more frequently for groups such as older adults and pregnant women.

Measures such as HbA1c and cholesterol are likely to be measured more frequently in individuals with established or suspected health conditions, limiting the ability to make accurate comparisons between those with and without disease. Furthermore, variables such as tobacco consumption and alcohol use are subject to reporting bias which may affect an individual's propensity to visit the GP or to report habits candidly, most often leading to an underestimate of the true exposure.(305) Finally, ethnicity may be better recorded for patients who are perceived to be at high risk, or have established disease conditions.

a) Sensitivity analyses

For each of the regression models included in the study, sensitivity analysis using complete cases only was undertaken to examine the robustness of estimates of ethnic differences when restricting the analysis to individuals with complete covariate recording.

For the main models describing differences between White, South Asian, and Black African/Caribbean ethnic groups, an additional sensitivity analysis was conducted using the non-diabetic population who were also free of type 1 diabetes at the time of the study.

Chapter 9
Ethnic and gender differences in CHD risk amongst patients with and without type 2 diabetes: Results for non-fatal CHD

9.1 Summary

This chapter reports the findings of the prospective cohort study identifying ethnic and gender differences in the risk of first non-fatal CHD for patients with and without type 2 diabetes. Section 9.2 describes the derivation of the study populations and section 9.3 their demographic and baseline clinical characteristics. The next section presents the findings of: the primary analyses, which identified ethnic differences in CHD risk amongst patients with T2DM and compared CHD risk to those without T2DM. The third section presents the results from secondary analyses which explored heterogeneity of risk between ethnic minority subgroups. The final section explores differences in risk of non-fatal CHD by diabetic status.

9.2 Study Populations

A total of 862,184 patients were included in the study. Participants were separated into two denominator populations: one of patients with T2DM diagnosed between 1990 and 2013, and one of patients free from T2DM over the same period. The diabetic cohort was derived from the 331,836 patients with T2DM identified in chapter 7. After applying inclusion criteria and logic checks, 196,254 were included in the main analysis of incident non-fatal CHD. Forty-four per cent of this population had linked ONS mortality data and were included in the secondary analysis of incident fatal and non-fatal CHD combined, presented in chapter 10.

The first non-diabetic cohort was derived from a random sample of 1 million patients aged 30 and over on January 1st 1990. From this sample, 967,810 were found to be free from diabetes as of August 1st 2013, and after applying the inclusion criteria and logic checks, 665,930 individuals were included in the main analysis of incident non-fatal CHD.

The second non-diabetic cohort was derived from an age matched sample of 915,305 individuals. The requirement that study entry date precede study exit date was incorporated into the matching program syntax. After removing patients with prevalent CHD, 858,054 individuals were included in the main analysis of incident non-fatal CHD (Figure 9.1).

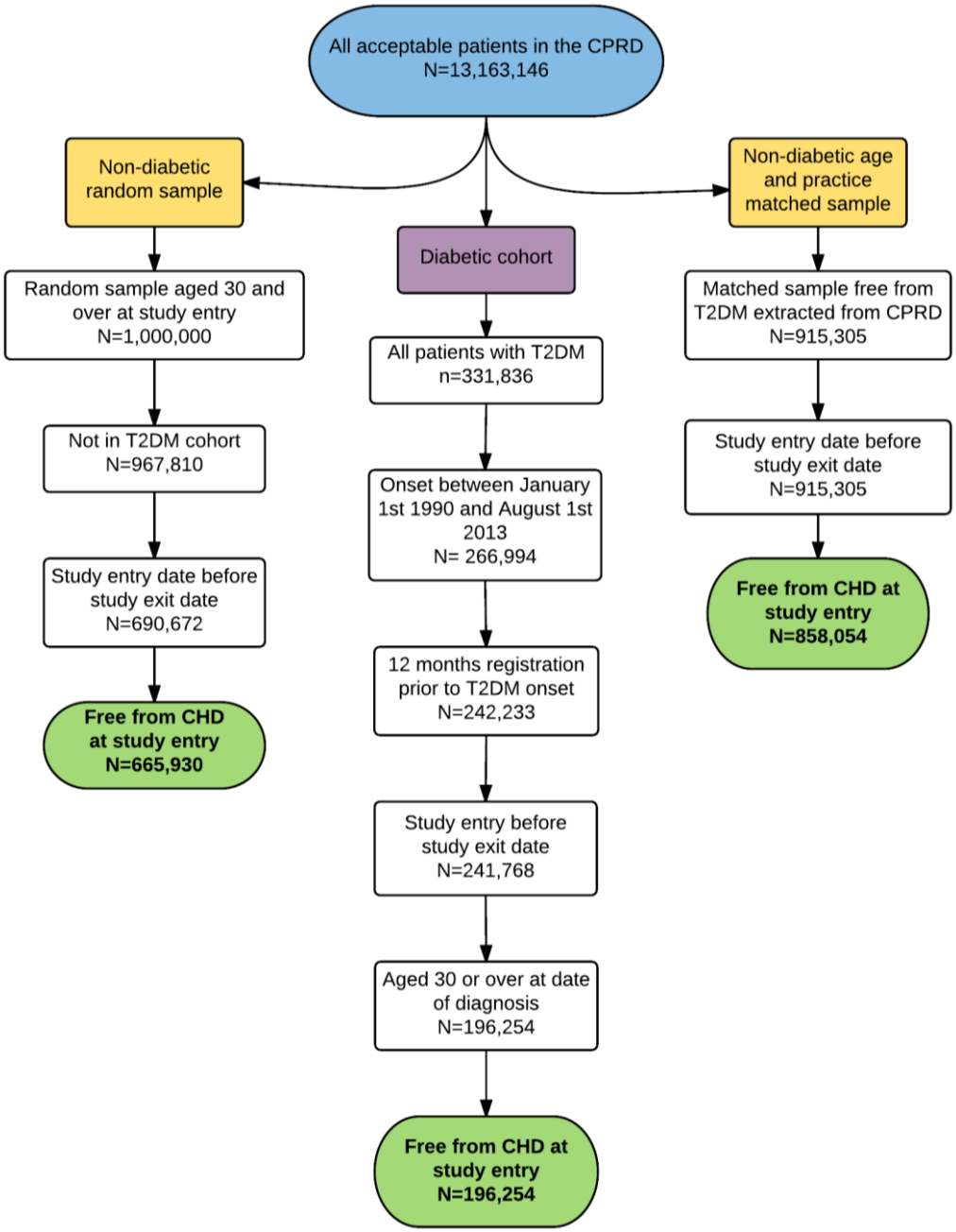


Figure 9.1 Derivation of study cohorts

The gender and ethnic breakdown of the study populations is illustrated in figure 9.2.

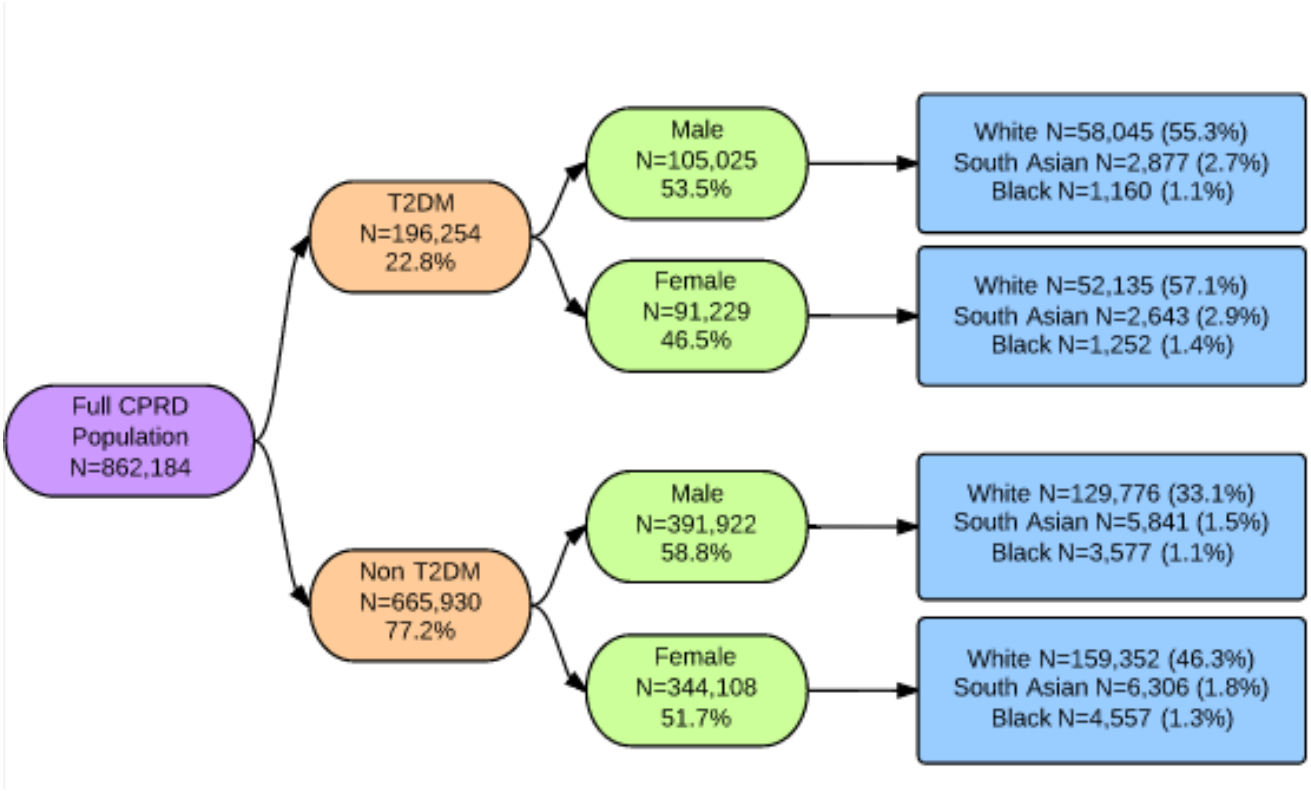


Figure 9.2 Breakdown of study populations by diabetic status, gender and ethnic group

9.3 Demographic and baseline characteristics for main analysis of incident non-fatal CHD in the whole CPRD

9.3.1 Demographic characteristics

Table 9.1 shows the demographic and clinical characteristics of the subjects according to diabetic status. The two denominator populations for the main analysis of incident non-fatal CHD were 196,254 patients with type 2 diabetes free from CHD at the time of study entry and 665,930 non-diabetic patients free from CHD at the age of 30. Ethnicity was recorded for 61.3% of the diabetic population and 47.8% of the non-diabetic population. The diabetic population comprised 56.1% White, 2.8% South Asian, 1.2% Black African/Caribbean and 39.8% of other or unknown ethnicity.

The non-diabetic random sample comprised 43.4% White, 1.8% South Asian, 1.2% Black African/Caribbean and 53.5% of other or unknown ethnicity. The non-diabetic matched sample comprised 46.4% White, 1.2% South Asian, 0.8% Black African/Caribbean and 51.6% of other or unknown ethnicity.

In the diabetic population, 13,331 patients (6.8%) experienced a non-fatal CHD event over a total of 1,052,717 person-years of follow-up. Median follow-up time was 4.6 years in the diabetic cohort (SD 4.0)

In the non-diabetic random sample, 23,248 patients (3.5%) experienced a non-fatal CHD event over a total of 4,731,546 person-years of follow-up. Median follow-up time was 5.5 years in the non-diabetic random sample (SD 6.0).

In the non-diabetic matched sample, 62,976 patients (7.3%) experienced a non-fatal CHD event over a total of 7,435,005 person-years of follow-up. Median follow-up time was 7.8 years in the non-diabetic matched sample (SD 6.5).

In the diabetic cohort, the proportion of patients experiencing a non-fatal CHD event was highest in the White population (7.2%) followed by South Asian (7.0%) and Black African/Caribbean (3.3%). In the non-diabetic population, events were most common in the White group (6.2%) followed by South Asian (4.5%) and Black African/Caribbean (4.4%).

The median age at baseline was 62.5 for diabetic participants (SD 13.0) and 42.0 for the non-diabetic random sample (SD 16.4) and 56 for the non-diabetic matched sample (SD 14.1). In both the diabetic and non-diabetic cohorts, the median age at study entry was over 10 years lower for South Asian than for White patients. Compared with the non-diabetic population, the diabetic population had a greater proportion of males (53.5% vs. 48.3%).

9.3.2 Medications

Recording of medication usage can be thought of as being 100% complete, as a code is added to the medical record only when a prescription is issued; thus, the absence of a code indicates that no prescription was made. Over the course of follow-up, insulin was prescribed for 13.2% of those with diabetes, metformin for 70.9% and other antidiabetic drugs for 50.2%. Prescription of insulin and other antidiabetic drugs (OAD) was comparable between ethnic groups while metformin prescription was 10% higher for the South Asian group (82.8%) compared with the White group (72.3%). The prescription of cardiovascular medications was 3–6 times higher in the diabetic compared with the non-diabetic cohort. In both cohorts, the prescribing of cardiovascular medication was highest in the White group.

9.3.3 Clinical covariates

Completeness of covariate recording was consistently higher in the diabetic cohort than in the non-diabetic cohort. A baseline blood pressure value was available for 99.8% of diabetics and 86.0% of non-diabetics, with hypertension recorded in 42% of diabetics and 19.6% of non-diabetics. Hypertension was most prevalent in the White group, followed by Black and South Asian groups in both diabetic and non-diabetic cohorts.

Total cholesterol values were available for 97.5% of diabetics and 40.7% of non-diabetics, with raised cholesterol at baseline evident in 31% of diabetics and 26% of non-diabetics. The proportion of patients with raised baseline cholesterol did not vary by ethnic group amongst diabetics, but was raised in the White group compared with South Asian and Black groups in the non-diabetic cohort.

Smoking status was recorded for 97.8% of diabetics and 88.2% of non-diabetics. At baseline, fewer diabetics had their latest smoking status recorded as “current smoker” compared with non-diabetics (15.6% vs. 26.1%). In both the diabetic and non-diabetic cohorts, the proportion of current and ex-smokers was highest in the White group and lowest in the South Asian group.

Alcohol usage was recorded for 94.9% of diabetics and 76.4% of non-diabetics. The majority of South Asians in both diabetic and non-diabetics had their baseline alcohol status recorded as “non-drinker” (66.5% diabetic, 54.4% non-diabetic). In both diabetic cohorts, heavy drinking was 4–8 times more prevalent in the White group than in the South Asian and Black groups.

Body mass index was recorded for 97.9% of diabetics and 77.1% of non-diabetics. While 37.8% of non-diabetics were overweight or obese, 81.8% of diabetics were in that weight category. In the diabetic cohort, the majority of White and Black groups were obese, while the majority of South Asian patients fell into the overweight category. In the non-diabetic cohort, the majority of patients of all ethnic groups fell into the normal weight category.

Amongst the 97.5% of diabetics with HbA1c values recorded, 69.2% had a baseline value greater than 6.5%. The proportion of patients with raised HbA1c was highest for South Asian patients (78.8%) followed by Black (72.4%) and White (68.8%) patients.

Table 9.1 Demographic, clinical, and therapeutic characteristics of CPRD patients with and without T2DM

	T2DM N=196,254										Non diabetic random sample N=665,930									
	White		South Asian		Black		All other		Total		White		South Asian		Black		All other		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Total	110,180	100	5,520	100	2,412	100	78142	100	196,254	100	289,128	100	12,147	100	8,134	100	356,521	100	665,930	100
Incident CHD	7,928	7.2	385	7.0	80	3.3	4,938	6.3	13,331	6.8	11,553	4.0	208	1.7	51	0.6	11,436	3.2	23,248	3.5
Median follow-up (SD)	4.9	±4.1	4.3	±4.2	4.0	±4.1	4.2	±4.0	4.6	±4.0	6.0	±6.2	3.0	±4.5	3.1	±4.4	5.2	±5.9	5.5	±6.0
Gender																				
Male	58,045	52.7	2,877	52.1	1,160	48.1	42,943	55.0	105,025	53.5	129,776	44.9	5,841	48.1	3,577	44	182,628	51.2	321,822	48.3
Female	52,135	47.3	2,643	47.9	1,252	51.9	35,199	45.0	91,229	46.5	159,352	55.1	6,306	51.9	4,557	56	173,893	48.8	344,108	51.7
Age at study entry																				
30-39	4,183	3.8	763	13.8	206	8.5	3,379	4.3	8,531	4.3	122,295	42.3	7,907	65.1	4,808	59.1	162,349	45.5	297,359	44.7
40-49	14,023	12.7	1,519	27.5	603	25.0	10,985	14.1	27,130	13.8	57,129	19.8	2,166	17.8	1,977	24.3	69,485	19.5	130,757	19.6
50-59	25,684	23.3	1,628	29.5	639	26.5	19,027	24.3	46,978	23.9	43,883	15.2	1,079	8.9	710	8.7	46,097	12.9	91,769	13.8
60-69	31,482	28.6	1,020	18.5	564	23.4	21,104	27.0	54,170	27.6	31,444	10.9	600	4.9	365	4.5	31,788	8.9	64,197	9.6
70+	34,808	31.6	590	10.7	400	16.6	23,647	30.3	59,445	30.3	34,377	11.9	395	3.3	274	3.4	46,802	13.1	81,848	12.3
Median age (SD)	63.2	±12.7	52.3	±12.0	55.5	±12.5	62.3	±13.1	62.5	±13.0	43	±16.2	35	±11.6	37	±11.3	41	±16.8	42	±16.4
Median age male (SD)	61.7	±12.2	51.4	±11.9	54.9	±12.6	60.4	±12.5	60.8	±12.4	43	±15.0	36	±11.4	38	±11.0	40	±15.2	41	±15.1
Median age fem. (SD)	65.1	±13.2	53.4	±121	56.1	±12.4	64.9	±13.5	64.6	±13.4	43	±17.1	35	±11.8	36	±11.5	44	±18.1	43	±17.5
Practice level IMD score																				
1	14,880	13.5	511	9.3	116	4.8	15,826	20.3	31,333	16.0	46,217	16	1,590	13.1	449	5.5	69,780	19.6	118,036	17.7
2	23,760	21.6	779	14.1	265	11.0	10,911	14.0	35,715	18.2	64,394	22.3	1,903	15.7	925	11.4	62,836	17.6	130,058	19.5
3	22,734	20.6	1,180	21.4	501	20.8	18,062	23.1	42,477	21.6	62,274	21.5	2,818	23.2	1,538	18.9	75,596	21.2	142,226	21.4
4	24,636	22.4	1,394	25.3	874	36.2	17,597	22.5	44,501	22.7	62,093	21.5	2,813	23.2	2,815	34.6	82,026	23	149,747	22.5
5	24,170	21.9	1,656	30.0	656	27.2	15,746	20.2	42,228	21.5	54,150	18.7	3,023	24.9	2,407	29.6	66,283	18.6	125,863	18.9
Medications ever prescribed+																				
Insulin	15,412	14	684	12.4	373	15.5	9,477	12.1	25,946	13.2										
Metformin	79,629	72.3	4,571	82.8	1,921	79.6	53,102	68	139,223	70.9										
OAD	55,426	50.3	2,833	51.3	1,195	49.5	39,076	50	98,530	50.2										
Antihypertensives	90,985	82.6	3,811	69	1,783	73.9	60,994	78.1	157,573	80.3	96,912	33.5	2,093	17.2	1,755	21.6	89,449	25.1	190,209	28.6
Lipid lowering	87,817	79.7	4,225	76.5	1,659	68.8	56,446	72.2	150,147	76.5	47,351	16.4	1,281	10.5	646	7.9	32,024	9	81,302	12.2
Aspirin	58,665	53.2	2,511	45.5	1,000	41.5	37,646	48.2	99,822	50.9	43,303	15	842	6.9	495	6.1	37,047	10.4	81,687	12.3
With linked ONS mortality data																				
	71,405	64.8	3,120	56.5	1,500	62.2	10,755	13.8	86,780	44.2	165,794	57.3	4,872	40.1	3,833	47.1	35,063	9.8	209,562	31.5

+ Medications for diabetes extracted only for patients with diagnosed type 2 diabetes

Table 9.1 Continued…

	T2DM N=196,254										Non diabetic random sample N=665,930									
	White		South Asian		Black		All other		Total		White		South Asian		Black		All other		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Baseline clinical measures																				
Blood Pressure																				
≤140/90 mmHg	63,040	57.2	3,495	63.3	1,450	60.1	44,805	57.3	112,790	57.5	204944	70.9	9423	77.6	5976	73.5	221907	62.2	442250	66.4
>140/90 mmHg	47,058	42.7	2,018	36.6	958	39.7	32,962	42.2	82,996	42.3	61913	21.4	1338	11	1428	17.6	66064	18.5	130743	19.6
Unknown	82	0.1	7.0	0.1	4.0	0.2	375	0.5	468	0.2	22271	7.7	1386	11.4	730	9	68550	19.2	92937	14
Cholesterol																				
≤5 mmol/L	75,420	68.5	3,839	69.5	1,629	67.5	49,924	63.9	130,812	66.7	52485	18.2	2765	22.8	1913	23.5	42359	11.9	99522	14.9
>5 mmol/L	33,347	30.3	1,649	29.9	754	31.3	24,690	31.6	60,440	30.8	94132	32.6	2763	22.7	1575	19.4	72768	20.4	171238	25.7
Unknown	1,413	1.3	32	0.6	29	1.2	3528	4.5	5002	2.5	142511	49.3	6619	54.5	4646	57.1	241394	67.7	395170	59.3
Smoking Status																				
Non	56,626	51.4	4,255	77.1	1,705	70.7	42,709	54.7	105,295	53.7	154150	53.3	9229	76	5858	72	177154	49.7	346391	52
Current	17,812	16.2	561	10.2	255	10.6	11,961	15.3	30,589	15.6	84645	29.3	1716	14.1	1454	17.9	86231	24.2	174046	26.1
Ex	35,075	31.8	685	12.4	441	18.3	21,745	27.8	57,946	29.5	37919	13.1	769	6.3	539	6.6	29384	8.2	68611	10.3
Unknown	667	0.6	19	0.3	11	0.5	1727	2.2	2424	1.2	12414	4.3	433	3.6	283	3.5	63752	17.9	76882	11.5
Alcohol consumption																				
Non	26,412	24	3,671	66.5	1,037	43	20,091	25.7	51,211	26.1	43833	15.2	6610	54.4	3329	40.9	51955	14.6	105727	15.9
Moderate	70,304	63.8	1,495	27.1	1,221	50.6	46,855	60	119,875	61.1	173676	60.1	3287	27.1	3331	41	175163	49.1	355457	53.4
High	9,056	8.2	120	2.2	55	2.3	5,854	7.5	15,085	7.7	24414	8.4	140	1.2	152	1.9	22725	6.4	47431	7.1
Unknown	4,408	4.0	234	4.2	99	4.1	5,342	6.8	10,083	5.1	47205	16.3	2110	17.4	1322	16.3	106678	29.9	157315	23.6
Body Mass Index																				
Underweight	908	0.8	33	0.6	6	0.2	580	0.7	1527	0.8	6095	2.1	479	3.9	147	1.8	6499	1.8	13220	2
Normal	16,163	14.7	1,301	23.6	340	14.1	12,181	15.6	29,985	15.3	117111	40.5	5386	44.3	2669	32.8	124740	35	249906	37.5
Overweight	35,638	32.3	2,297	41.6	886	36.7	25,932	33.2	64,753	33	83120	28.7	3388	27.9	2552	31.4	83110	23.3	172170	25.9
Obese	55,915	50.7	1,853	33.6	1,156	47.9	36,919	47.2	95,843	48.8	40420	14	1210	10	1703	20.9	35608	10	78941	11.9
Unknown	1,556	1.4	36.0	0.7	24	1.0	2,530	3.2	4,146	2.1	42382	14.7	1684	13.9	1063	13.1	106564	29.9	151693	22.8
HbA1c+																				
<5.5%	7,708	7.0	286	5.2	181	7.5	5037	6.4	13,212	6.7										
5.5-6.4%	24,952	22.6	831	15.1	433	18	16,068	20.6	42,284	21.5										
≥5.5%	75,845	68.8	4,348	78.8	1,746	72.4	53,936	69.0	135,875	69.2										
Unknown	1,675	1.5	55.0	1.0	52	2.2	3,101	4.0	4,883	2.5										

+Hba1C values collected and analysed for patients with diagnosed type 2 diabetes only

Table 9.1 continued...

	Non diabetic matched sample N=858,054									
	White		South Asian		Black		All other		Total	
	N	%	N	%	N	%	N	%	N	%
Total		100.0	9,974	100.0	6,535	100.0	443,009	100.0	858,054	100.0
Incident CHD	34,332	8.6	691	6.9	188	2.9	27,765	6.3	62,976	7.3
Median follow-up (SD)	9.1	±6.6	4.8	±5.7	4.5	±5.5	6.9	±6.4	7.8	±6.5
Gender										
Male	181,564	45.6	4,754	47.7	2,900	44.4	217,766	49.2	406,984	47.4
Female	216,972	54.4	5,220	52.3	3,635	55.6	225,226	50.8	451,053	52.6
Age at study entry										
30-39	42,877	10.8	2,136	21.4	1,420	21.7	69,072	15.6	115,505	13.5
40-49	72,320	18.1	2,356	23.6	1,771	27.1	93,360	21.1	169,807	19.8
50-59	105,680	26.5	2,419	24.3	1,373	21.0	109,040	24.6	218,512	25.5
60-69	98,319	24.7	1,947	19.5	1,219	18.7	93,550	21.1	195,035	22.7
70+	79,340	19.9	1,116	11.2	752	11.5	77,987	17.6	159,195	18.6
Median age (SD)	58	±13.8	52	±13.5	50	±13.5	55	±14.3	56	±14.1
Median age male (SD)	56	±13.0	51	±13.4	50	±13.2	53	±13.7	55	±13.5
Median age fem. (SD)	58	±14.3	52	±13.8	50	±13.8	57	±14.7	57	±14.5
Practice level IMD score										
1	55,904	14.0	923	9.3	280	4.3	84,203	19.0	141,310	16.5
2	85,607	21.5	1,199	12.0	694	10.6	69,185	15.6	156,685	18.3
3	83,666	21.0	2,266	22.7	1,376	21.1	95,544	21.6	182,852	21.3
4	87,975	22.1	2,514	25.2	2,152	32.9	100,538	22.7	193,179	22.5
5	85,384	21.4	3,072	30.8	2,033	31.1	93,539	21.1	184,028	21.4
Medications ever prescribed+										
Antihypertensives	213,996	53.7	4,437	44.5	2,993	45.8	174,709	39.4	396,135	46.2
Lipid lowering	135,652	34.0	3,524	35.3	1,715	26.2	84,866	19.2	225,757	26.3
Aspirin	118,087	29.6	2,522	25.3	1,243	19.0	86,084	19.4	207,936	24.2

Table 9.1 continued...

	Non diabetic matched sample N=858,054									
	White		South Asian		Black		All other		Total	
	N	%	N	%	N	%	N	%	N	%
Baseline clinical measures										
Blood Pressure										
≤140/90 mmHg	245,240	61.5	6,907	69.3	4,162	63.7	241,376	54.5	497,685	58.0
>140/90 mmHg	135,967	34.1	2,478	24.8	2,031	31.1	127,171	28.7	267,647	31.2
Unknown	17,329	4.3	589	5.9	342	5.2	74,462	16.8	92,722	10.8
Cholesterol										
≤5 mmol/L	83,847	21.0	3,082	30.9	1,896	29.0	56,144	12.7	144,969	16.9
>5 mmol/L	192,220	48.2	4,129	41.4	2,422	37.1	132,598	29.9	331,369	38.6
Unknown	122,469	30.7	2,763	27.7	2,217	33.9	254,267	57.4	381,716	44.5
Smoking Status										
Non	217,998	54.7	7,729	77.5	4,631	70.9	218,052	49.2	448,410	52.3
Current	106,023	26.6	1,295	13.0	1,079	16.5	103,026	23.3	211,423	24.6
Ex	58,582	14.7	621	6.2	566	8.7	42,510	9.6	102,279	11.9
Unknown	15,933	4.0	329	3.3	259	4.0	79,421	17.9	95,942	11.2
Alcohol consumption										
Non	70,649	17.7	5,961	59.8	2,630	40.2	72,766	16.4	152,006	17.7
Moderate	240,433	60.3	2,517	25.2	2,862	43.8	215,097	48.6	460,909	53.7
High	34,163	8.6	183	1.8	162	2.5	29,913	6.8	64,421	7.5
Unknown	53,291	13.4	1,313	13.2	881	13.5	125,233	28.3	180,718	21.1
Body Mass Index										
Underweight	5,822	1.5	241	2.4	76	1.2	5,822	1.3	11,961	1.4
Normal	139,959	35.1	3,831	38.4	1,749	26.8	136,896	30.9	282,435	32.9
Overweight	134,607	33.8	3,328	33.4	2,274	34.8	119,091	26.9	259,300	30.2
Obese	69,414	17.4	1,489	14.9	1,685	25.8	55,971	12.6	128,559	15.0
Unknown	48,734	12.2	1,085	10.9	751	11.5	125,229	28.3	175,799	20.5

9.4 Comparison of incident non-fatal CHD risk according to diabetic status

The first objective of this study was to compare the risk of incident CHD between patients with and without type 2 diabetes and explore interactions between diabetic status and ethnic group, age, gender and deprivation

9.4.1 Comparison with random sample of non-diabetics

The combined study population of 862,184 patients was used to compare the incidence of non-fatal CHD between individuals with type 2 diabetes and the random sample without type 2 diabetes. Table 9.2 displays the crude and adjusted hazard ratios for the entire study population as a whole and stratified by ethnic group. Within the whole study population, 36,579 incident events of non-fatal CHD were recorded. The crude risk of non-fatal CHD was 39% higher in the diabetic population compared with the non-diabetic reference population (HR 1.39, CI 95 1.36, 1.42). After adjustment for age, gender and practice-level deprivation, the excess risk in the diabetic population was reduced to 22% (HR 1.22, CI95% 1.20, 1.25). By ethnic group, the adjusted risk of non-fatal CHD was increased in the Black population by 75% (HR 1.75, CI95% 1.18, 2.58), in the South Asian population by 60% (HR 1.60, CI95% 1.34, 1.91) and in the White population by 28% (HR 1.28, CI95% 1.24, 1.31).

Table 9.2 Incidence of non-fatal CHD according to diabetic status: random sample of non-diabetics

	White		South Asian		Black		Other/unknown		Total	
N	399,308		17,667		10,546		434,663		862,184	
Incident CHD	19,481		593		131		16,374		36,579	
	HR	[CI 95]	HR	[CI 95]	HR	[CI 95]	HR	[CI 95]	HR	[CI 95]
Non T2DM (ref)	1	--	1	--	1	--	1	--	1	--
T2DM crude	1.44	[1.40, 1.48]	1.67	[1.40, 1.99]	1.75	[1.19, 2.58]	1.32	[1.27, 1.36]	1.39	[1.36, 1.42]
T2DM adjusted	1.28	[1.24, 1.31]	1.60	[1.34, 1.91]	1.75	[1.18, 2.58]	1.15	[1.11, 1.19]	1.22	[1.20, 1.25]

The results from the Cox regression examining interactions between diabetic status demographic factors are displayed in table 9.3. Tests for effect modification by gender, age and deprivation were found to be statistically significant. No significant interaction between diabetic status and ethnic group was found. The interaction between diabetic status and gender indicates that the relative risk of CHD for females compared with males differs between diabetic and non-diabetic patients. The reduction in CHD risk for females was greater for non-diabetics than diabetics (HR 0.58, CI95% 0.57, 0.59 vs. HR 0.67, CI95% 0.64, 0.69).

Similarly, the increase in CHD risk by ten-year age band was significantly greater for non-diabetic patients than for diabetic patients. In the diabetic cohort, CHD risk increased by 46% in the oldest group compared with the youngest group (CI95% 1.16, 1.84); In the non-diabetic cohort, CHD risk increased by almost 6 times in the oldest group compared with the youngest (HR 5.81, CI95% 5.19, 6.52).

The increase in CHD risk for those in the least affluent deprivation quintile compared with those in the most affluent quintile was larger for non-diabetics than for diabetics (HR 1.48, CI95% 1.42, 1.55 vs. HR 1.16, CI95% 1.10, 1.23). The test for interaction between ethnic group and diabetic status was non-significant, as exemplified by the overlapping confidence intervals for diabetic/non-diabetic South Asian patients and diabetic/non-diabetic Black patients. Thus, the difference in CHD risk for ethnic minority groups compared with White does not depend on diabetic status.

Table 9.3 Incidence of non-fatal CHD: Interactions between diabetic status and demographic factors

	NON T2DM random sample		T2DM		*P* value for interaction with diabetic status
	HR	[CI 95]	HR	[CI 95]	
Ethnic Group					
White (ref)	1	--	1	--	
South Asian	1.52	[0.33, 1.75]	1.32	[1.19, 1.47]	0.297
Black	0.56	[0.43, 0.75]	0.58	[0.47, 0.72]	0.370
Gender					
Male (ref)	1	--	1	--	
Female	0.58	[0.57, 0.59]	0.67	[0.64, 0.69]	0.007
Age Group					
30–39 (ref)	1	--	1	--	
40–49	1.48	[1.37, 1.62]	1.04	[0.86, 1.27]	<0.001
50–59	2.21	[2.01, 2.43]	1.19	[0.96, 1.47]	<0.001
60–69	3.56	[3.21, 3.96]	1.26	[1.01, 1.58]	<0.001
70+	5.81	[5.19, 6.52]	1.46	[1.16, 1.84]	<0.001
Deprivation Quintile					
IMD 1 (most affluent) (ref)	1	--	1	--	
IMD 5 (least affluent)	1.48	[1.42, 1.55]	1.16	[1.10, 1.23]	<0.001

9.4.2 Comparison with non-diabetic matched sample

The combined study population of 1,054,308 patients was used to compare the incidence of non-fatal CHD between individuals with type 2 diabetes and the matched sample without type 2 diabetes. Table 9.4 displays the crude and adjusted hazard ratios for the entire study population as a whole and stratified by ethnic group. Within the whole study population, 76,307 incident events of non-fatal CHD were recorded.

The influence of type 2 diabetes on the risk of incident non-fatal CHD was more modest in this analysis compared to the previous analysis utilizing the random sample of non-diabetics.

The crude risk of non-fatal CHD was 23% higher in the diabetic population compared with the non-diabetic reference population (HR 1.23, CI 95 1.21, 1.26). After adjustment for age, gender and practice-level deprivation, the excess risk in the diabetic population was reduced to 7% (HR 1.07, CI95% 1.05, 1.10). By ethnic group, the adjusted risk of non-fatal CHD was increased in all groups by 7%, except for the Black African/Caribbean population, for whom no difference in CHD risk between individuals with and without diabetes was found.

Table 9.4 Incidence of non-fatal CHD according to diabetic status: non-diabetics matched sample

	White		South Asian		Black		Other/unknown		Total	
N	508,716		15,494		8,971		521,151		1,054,308	
Incident CHD	42,260		1,076		268		32,703		76,307	
	HR	[CI 95]	HR	[CI 95]	HR	[CI 95]	HR	[CI 95]	HR	[CI 95]
Non T2DM (ref)	1	--	1	--	1	--	1	--	1	--
T2DM crude	1.21	[1.18, 1.24]	1.13	[0.99, 1.28]	1.05	[0.81, 1.37]	1.24	[1.21, 1.27]	1.23	[1.21, 1.26]
T2DM adjusted	1.07	[1.05, 1.10]	1.07	[0.94, 1.21]	0.97	[0.75, 1.27]	1.07	[1.04, 1.10]	1.08	[1.06, 1.10]

The results from the Cox regression examining interactions between diabetic status demographic factors are displayed in table 9.5.

As with the comparison using the random non-diabetic sample, tests for effect modification indicated a significant difference in the relationship between diabetes and CHD by gender, age and deprivation in the matched sample. No notable interaction between diabetic status and ethnic group was found.

Table 9.5 Incidence of non-fatal CHD: Interactions between diabetic status and demographic factors

	NON T2DM matched sample		T2DM		*P* value for interaction with diabetic status
	HR	[CI 95]	HR	[CI 95]	
Ethnic Group					
White (ref)	1	--	1	--	
South Asian	1.62	[0.33, 1.75]	1.32	[1.19, 1.47]	0.301
Black	0.71	[0.43, 0.75]	0.58	[0.47, 0.72]	0.333
Gender					
Male (ref)	1	--	1	--	
Female	0.57	[0.56, 0.57]	0.67	[0.64, 0.69]	<0.001
Age Group					
30–39 (ref)	1	--	1	--	
40–49	1.52	[1.42, 1.63]	1.04	[0.86, 1.27]	<0.001
50–59	2.37	[2.21, 2.55]	1.19	[0.96, 1.47]	<0.001
60–69	3.68	[3.41 3.97]	1.26	[1.01, 1.58]	<0.001
70+	5.68	[5.25 6.16]	1.46	[1.16, 1.84]	<0.001
Deprivation Quintile					
IMD 1 (most affluent) (ref)	1	--	1	--	
IMD 5 (least affluent)	1.39	[1.36, 1.43]	1.16	[1.10, 1.23]	<0.001

9.5 Crude incidence of first non-fatal CHD event

The second objective of this study was to quantify ethnic and gender differences in crude incidence rate of CHD for patients with and without type 2 diabetes. The analysis presented below describes the crude incidence and adjusted risk (section 9.6) of CHD within the diabetic and non-diabetic cohorts separately, stratified by gender and ethnic group. Crude incidence rates are displayed in table 9.6. The crude overall rate of incident non-fatal CHD was 126.5 per 10,000 person-years in the diabetic cohort (CI95% 124.5, 128.8), 49.1 per 10,000 person-years

in the non-diabetic random sample (CI95% 48.5, 49.7), and 84.7 per 10,000 person-years in the non-diabetic matched sample (CI95 84.0, 85.4)

9.5.1 By ethnic group

The crude cumulative incidence curves for non-fatal CHD by ethnic group and gender are shown in figure 9.3. In the diabetic population, overall rates for males were highest for South Asians followed by White and Black African/Caribbean. Amongst diabetic women, rates were similar for White and South Asian groups and lower for Black African/Caribbean. In the non-diabetic random sample, crude incidence rates in both men and women were highest in the White group, followed by South Asian and Black African/Caribbean.

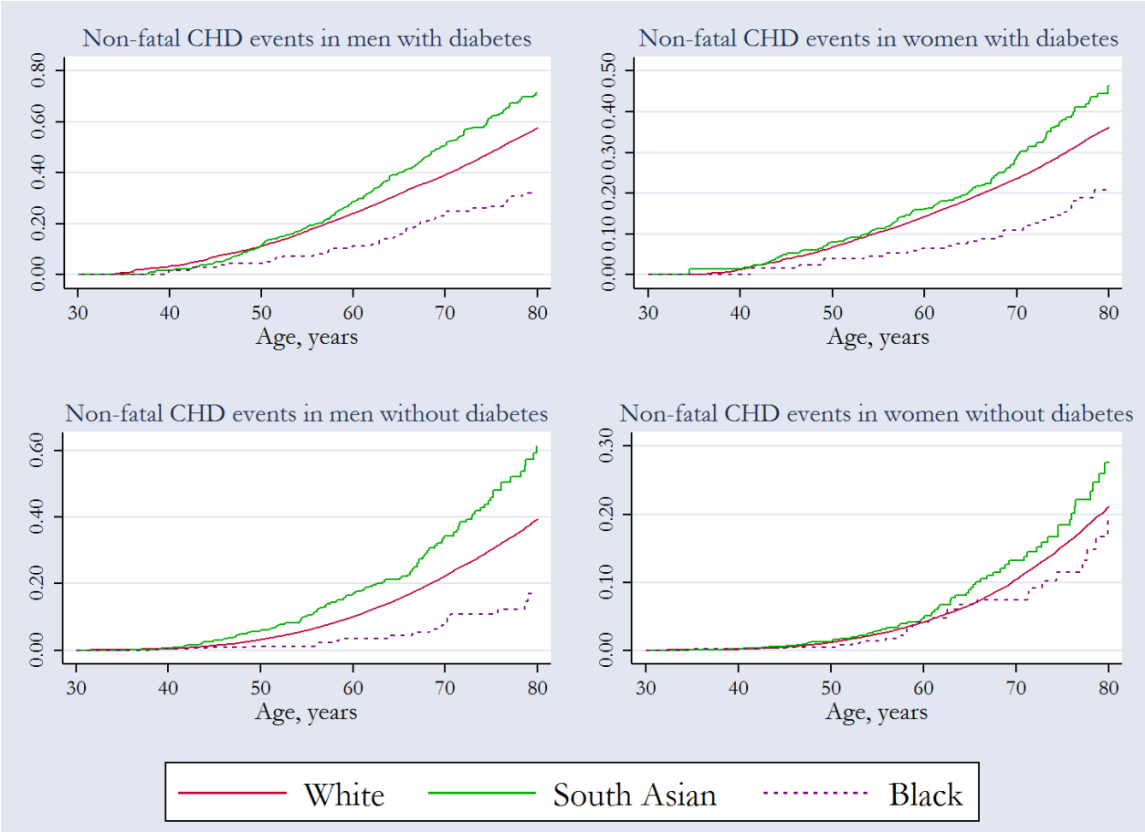


Figure 9.3 Cumulative incidence of non-fatal coronary heart disease

9.5.2 By deprivation quintile

Crude incidence of non-fatal CHD was higher for patients in the most deprived quintile of IMD score relative to the least deprived quintile for all patients regardless of ethnicity, gender, or diabetic status, save for Black diabetic females, where the trend was reversed. The difference

by deprivation quintile was greatest for White diabetic males, for whom CHD incidence was increased 9.5-fold in the most deprived compared with the least deprived group.

9.5.3 By baseline clinical values and medication use

Hypertension was associated with a 2- to 3-fold increase in CHD incidence amongst non-diabetics, and a more modest increase amongst diabetics, except for Black African/Caribbean men, who had a slightly lower incidence rate of hypertension. Across all populations, total serum cholesterol over 5 mmol/L was associated with a higher crude incidence of non-fatal CHD except for amongst South Asian men without diabetes, for whom high cholesterol was associated with a lower incidence. No consistent relationship with BMI category, smoking status, alcohol consumption or HbA1c was evident in the crude analysis.

Amongst diabetics only, prescriptions of insulin and other antidiabetic drugs were associated with increased CHD incidence, while those prescribed metformin had lower, higher or equivalent incidence across the three major ethnic groups. In all groups, the prescription of antihypertensives, lipid-lowering medications and statins were associated with an increase in the CHD incidence by 3- to 11-fold, though the increases were broadly more modest in the diabetic group compared with the non-diabetic group.

Table 9.6 Crude incidence rate of non-fatal CHD per 10,000 person-years

	T2DM N=196,254						Non T2DM random sample N=665,930					
	Males			Females			Males			Females		
	White	South Asian	Black	White	South Asian	Black	White	South Asian	Black	White	South Asian	Black
Number of subjects	*58,045*	*2,877*	*1,160*	*52,135*	*2,643*	*1,252*	*129,776*	*5,841*	*3,577*	*159,352*	*6,306*	*4,557*
Number of CHD events	*4,867*	*245*	*46*	*3,056*	*140*	*34*	*6,521*	*139*	*22*	*5,032*	*69*	*29*
Person-years of follow-up	*319,966*	*14,984*	*5,718*	*293,928*	*13,959*	*6,430*	*966,176*	*24,845*	*15,618*	*1,239,146*	*29,511*	*20,897*
Overall Rate/10,000 pyrs	*152.3*	*163.5*	*80.4*	*104.0*	*100.3*	*52.9*	*67.5*	*55.9*	*14.1*	*40.6*	*23.4*	*13.9*
Practice level deprivation												
1 (least deprived)	15.0	158.8	147.8	98.0	114.2	71.9	61.8	55.8	0	32.3	22.2	9.4
5 (most deprived)	142.9	201.3	134.7	122.7	123.7	56.3	79.6	57.9	24.2	52.7	28.7	20.1
Hypertension												
BP <=140/90 mmHg	141.3	152.6	89.9	95.6	92.8	41.7	53.7	45.1	9.6	25.8	15.4	9.9
BP >140/90 mmHg	156.1	176.6	70.1	108.7	104.7	69.5	118.6	142.0	32.0	86.4	77.4	35.8
Cholesterol												
Chol =<5 mmol/L	129.2	152.3	65.5	83.4	86.2	47.4	99.2	91.5	21.2	45.5	27.0	14.8
Chol >5 mmol/L	173.2	171.3	100.0	117.0	114.4	67.0	82.5	80.6	26.1	53.0	38.1	28.3
Smoking Status												
Non smoker	136.3	150.2	67.2	99.9	97.3	55.5	57.2	52.5	14.1	36.9	22.2	16.0
Smoker	144.3	144.5	87.1	103.7	119.8	0	73.7	58.8	14.6	38.3	14.3	4.3
Ex-smoker	154.7	182.0	63.4	98.3	77.2	71.0	87.0	57.3	11.1	44.9	20.7	7.2
Alcohol Consumption												
Non-drinker	158.8	173.3	71.1	89.4	66.5	67.2	87.2	65.8	21.0	57.8	23.1	10.5
Moderate alcohol drinker	146.7	150.5	77.0	85.4	0	36.4	69.2	47.1	11.6	33.9	9.5	22.3
Heavy drinker	118.7	104.3	0	96.6	67.2	0	57.9	86.8	19.4	26.3	38.6	0

Table 9.6 Continued...

	T2DM N=196,254						Non T2DM random sample N=665,930					
	Males			Females			Males			Females		
	White	South Asian	Black	White	South Asian	Black	White	South Asian	Black	White	South Asian	Black
BMI Category												
Underweight	151.6	0.0	0.0	137.5	0.0	0.0	65.5	18.5	49.3	53.4	-	-
Normal	155.4	155.2	118.1	108.3	112.4	66.4	58.9	60.9	19.0	30.4	10.6	6.7
Overweight	156.8	176.1	65.0	113.4	93.8	59.1	75.5	61.0	11.0	43.2	30.4	30.7
Obese	139.5	142.3	81.4	91.0	86.1	48.7	81.7	47.3	12.0	43.9	49.5	6.6
Hba1c Category												
<5.4	131.8	131.1	116.7	108.8	252.6	0.0						
5.5-6.4	124.7	125.3	86.3	86.2	87.0	39.3						
>=6.5	154.4	169.3	78.3	106.5	95.9	59.3						
Medications												
Antihypertensives												
No	100.0	110.0	52.5	58.0	35.6	7.1	41.4	32.7	7.8	19.1	10.1	4.4
Yes	170.0	200.0	92.4	110.0	130.0	65.8	159.0	201.1	42.5	88.3	78.7	43.3
Statins												
No	162.0	135.8	60.7	115.2	91.1	43.5	55.6	40.1	13.0	33.9	13.8	9.1
Yes	147.8	176.8	93.7	99.4	105.3	58.1	185.6	223.9	30.4	118.8	159.5	92.5
Aspirin												
No	111.7	88.2	42.2	81.0	60.3	26.5	48.8	43.7	10.7	27.3	15.9	8.6
Yes	195.9	259.5	133.2	131.0	158.5	90.4	276.6	282.1	91.9	182.1	141.5	100.1
Insulin												
No	148.0	151.8	85.0	97.2	95.8	54.6						
Yes	189.9	296.0	51.1	155.3	129.7	42.7						
Metformin												
No	165.4	162.0	99.8	109.2	97.6	40.5						
Yes	145.0	164.0	72.2	101.1	101.1	56.6						
OAD												
No	143.8	137.3	70.6	88.5	73.3	43.7						
Yes	161.5	187.8	89.2	121.6	126.1	60.8						

Table 9.6 continued…

	Non T2DM Matched sample N= 858,054					
	Males			Females		
	White	South Asian	Black	White	South Asian	Black
Number of subjects	*181,564*	*4,754*	*2,900*	*216,972*	*5,220*	*3,635*
Number of CHD events	*19,854*	*439*	*99*	*14,478*	*252*	*89*
Person-years of follow-up	*1,667,779*	*29,468*	*17,335*	*2,082,204*	*34,979*	*22,772*
Overall Rate/10,000 pyrs	*119.1*	*149.0*	*57.1*	*69.5*	*72.0*	*39.1*
Practice level deprivation						
1 (least deprived)	107.4	101.8	66.9	60.8	72.5	25.6
5 (most deprived)	132.9	163.3	59.3	83.5	88.1	43.7
Hypertension						
BP <=140/90 mmHg	99.3	120.4	48.1	50.3	53.7	28.2
BP >140/90 mmHg	17201	235.9	79.2	108.3	133.5	61.3
Cholesterol						
Chol =<5 mmol/L	142.0	184.3	47.4	83.0	82.9	53.5
Chol >5 mmol/L	164.5	200.5	103.1	89.7	110.9	55.9
Smoking Status						
Non smoker	103.2	131.2	52.0	63.0	71.0	42.5
Smoker	135.5	171.0	75.2	77.7	41.5	9.1
Ex-smoker	145.4	206.5	53.2	78.8	78.5	39.7
Alcohol Consumption						
Non-drinker	159.6	173.1	51.9	91.9	80.0	50.8
Moderate alcohol drinker	120.9	127.1	51.2	61.5	27.5	27.1
Heavy drinker	104.7	164.4	68.8	52.3	79.0	28.4

Table 9.6 continued...

	Non T2DM Matched sample N= 858,054					
	Males			Females		
	White	South Asian	Black	White	South Asian	Black
BMI Category						
Underweight	110.6	61.9	0	72.8	22.3	0
Normal	103.0	128.9	47.0	53.7	56.4	41.8
Overweight	131.2	161.9	52.7	74.9	72.3	33.4
Obese	148.3	197.9	59.4	85.1	117.7	37.3
Medications						
Antihypertensives						
No	77.6	81.1	24.2	35.9	28.1	7.9
Yes	21.6	296.2	108.3	116.6	148.0	76.3
Statins						
No	102.5	118.8	46.6	59.1	55.5	28.5
Yes	227.5	300.0	125.3	143.7	183.2	115.7
Aspirin						
No	89.4	111.2	43.2	50.0	46.6	24.2
Yes	321.2	409.5	176.4	205.1	274.5	157.5

9.5.4 By age group

The age-specific incidence rates stratified by gender are shown in Figure 9.4. The incidence of CHD increased with age in both diabetic population and non-diabetic random sample, with all rates higher in males than females. Crude rates in each age band stayed highest in the diabetic population until age 70, at which point the rates for diabetic and non-diabetic groups converge.

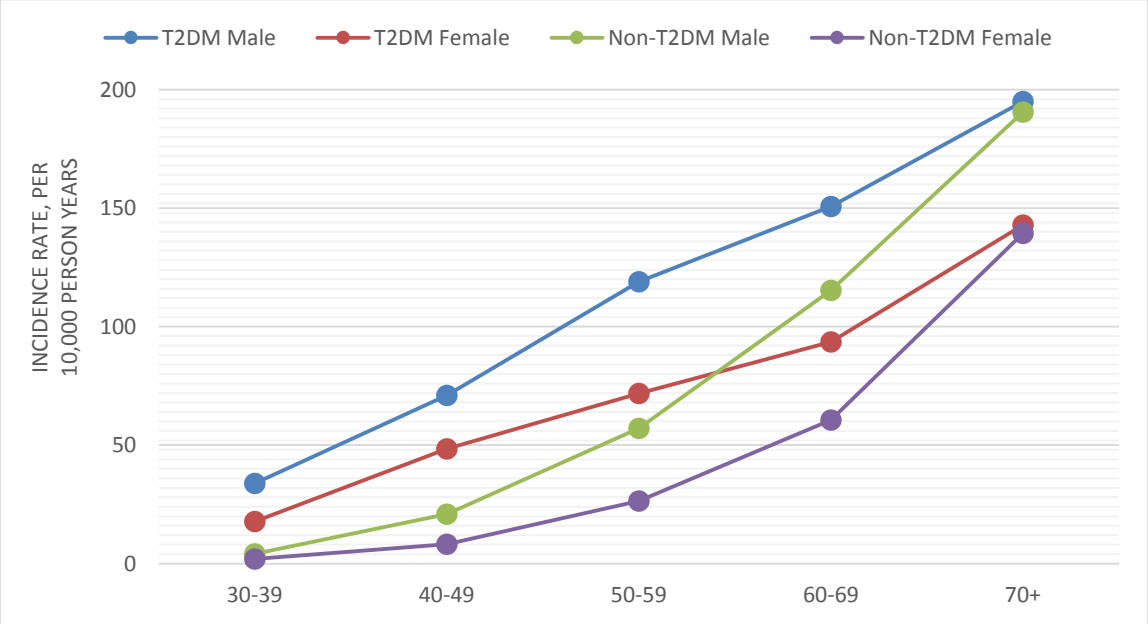


Figure 9.4 Age-specific incidence of non-fatal CHD by gender

The age-specific incidence rates stratified by ethnic group are shown in Table 9.7. In all groups, the age-specific rates were highest in South Asians, followed by White and Black African/Caribbean groups. Crude rates in the youngest age group were between 4 and 15 times higher in the diabetic population than in the non-diabetic population, with the largest difference evident for White females (crude event rate of 1.4/10,000 person-years in non-diabetics and 21.4/10,000 person-years in diabetics). Rates increased uniformly with age, except for Black African/Caribbean males, where the rate in those aged 70 and older was lower than in those aged 60–69. In the black male population, the amount of time at risk contributed by the group aged 70 or more was greater than the time at risk in all other age groups. In the group aged 60–69 years, 20 incident CHD events occurred over 1,471 person-years of follow-up. In the group aged 70 year or more, 13 incident CHD events occurred over 1,547 person-years of follow-up.

Table 9.7 Age-specific incidence rates of non-fatal CHD according to gender, ethnic group and diabetic status, per 10,000 person-years

	Person-years of follow-up	Men									Women								
		White			South Asian			Black			White			South Asian			Black		
		Rate	CI 95%		Rate	CI 95%		Rate	CI 95%		Rate	CI 95%		Rate	CI 95%		Rate	CI 95%	
T2DM																			
Events		4,867			245			46			3,056			140			34		
30–39	23, 902	43.5	[29.6,	63.9]	26.3	[8.5,	81.7]	41.7	[5.9,	295.8]	21.4	[12.4,	36.8]	11.1	[1.6,	78.7]	0.0	.	.
40–49	107,703	81.6	[72.3,	92.2]	106.7	[77.3,	147.3]	27.5	[8.9,	85.2]	57.3	[48.6,	67.6]	68.0	[43.9,	105.4]	36.4	[13.7,	96.9]
50–59	239,923	130.5	[122.4,	139.1]	170.1	[136.4,	212.1]	65.9	[34.3,	126.6]	75.8	[68.6,	83.7]	81.7	[58.9,	113.2]	26.4	[9.9,	70.4]
60–69	302,362	151.1	[143.6,	159.0]	223.8	[178.5,	280.7]	136.0	[87.8,	210.9]	93.5	[87.0,	100.4]	120.5	[89.0,	163.0]	45.4	[22.7,	90.7]
70+	394,758	193.0	[185.0,	201.4]	214.3	[162.9,	282.0]	84.1	[48.8,	144.8]	133.6	[127.5,	140.0]	184.8	[136.0,	250.9]	94.4	[58.7,	151.9]
Non- T2DM random sample																			
Events		6,521			139			22			5,032			69			29		
30–39	1,206,198	4.9	[4.0,	5.9]	6.2	[3.0,	13.1]	3.4	[0.8,	13.4]	1.4	[1.0,	1.9]	1.5	[0.4,	6.0]	2.2	[0.6,	8.9]
40–49	1,175,799	27.0	[25.0,	29.1]	51.8	[37.2,	72.2]	9.1	[3.8,	21.9]	9.8	[8.7,	11.0]	10.3	[5.1,	20.5]	3.0	[0.8,	12.0]
50–59	948,068	67.9	[64.4,	71.5]	100.0	[74.7,	140.0]	17.6	[6.6,	46.8]	29.5	[27.4,	31.7]	35.1	[21.5,	57.3]	32.7	[17.1,	62.9]
60–69	684,908	120.0	[120.0,	130.0]	160.0	[110.0,	230.0]	37.6	[14.1,	100.0]	61.9	[58.5,	65.4]	85.8	[56.0,	130.0]	34.5	[14.4,	82.9]
70+	716,573	180.0	[170.0,	180.0]	220.0	[150.0,	320.0]	83.8	[40.0,	180.0]	120.0	[120.0,	130.0]	150.0	[99.8,	230.0]	110.0	[58.7,	190.0]
Non- T2DM matched sample																			
Events			19,854			439			99			14,478			252			89	
30–39	486,124	9.4	[7.6,	11.7]	21.0	[10.5,	41.9]	5.1	[0.7,	35.9]	5.0	[3.8,	6.5]	8.6	[3.2,	22.8]	3.2	[0.5,	22.7]
40–49	1,212,263	41.1	[38.7,	43.7]	82.7	[64.3,	110.0]	10.8	[4.5,	26.1]	16.5	[15.1,	18.1]	24.3	[15.7,	37.7]	11.6	[5.5,	24.3]
50–59	1,922,600	95.4	[92.6,	98.3]	140.0	[120.0,	170.0]	59.1	[40.3,	86.8]	41.6	[39.8,	43.4]	70.2	[55.5,	88.9]	32.5	[20.5,	51.6]
60–69	1,994,820	150.0	[140.0,	150.0]	230.0	[200.0,	270.0]	97.3	[70.5,	130.0]	76.3	[74.1,	78.6]	110.0	[91.2,	140.0]	65.7	[46.5,	92.9]
70+	1,819,198	190.0	[180.0 ,	190.0]	290.0	[240.0,	340.0]	120.0	[82.3,	170.0]	120.0	[120.0,	120.0]	160.0	[130.0,	200.0]	97.2	[68.4,	140.0]

Rates are per 10,000 person-years [95% confidence interval]

9.6 Adjusted risk of non-fatal CHD by ethnic group

The third objective of this study was to examine ethnic differences in the adjusted risk of CHD for patients with and without type 2 diabetes.

Table 9.8a presents the results of the Cox proportional hazards models for ethnic differences in non-fatal CHD risk stratified by diabetic status and gender. For both the diabetic and non-diabetic cohorts, no ethnic differences were apparent in the crude model. After adjustment or matching for age, South Asian ethnicity was consistently associated with a higher risk of CHD compared with White ethnicity. Conversely, the risk of CHD in the Black African/Caribbean population was reduced in comparison with the White population.

9.6.1 Diabetic Cohort

In the fully adjusted models for diabetic males, risk of incident non-fatal CHD was increased by 42% for South Asian males compared with White males (HR 1.42, CI95% 1.21, 1.68) and reduced by 42% for Black African/Caribbean males compared with White males (HR 0.58, CI95% 0.41,0.84). This pattern was mirrored for diabetic females, though the final differences were smaller than for the male cohort. In the fully adjusted models, the risk of incident non-fatal CHD was increased by 31% for South Asian women compared with White women (HR 1.31, CI95% 1.05, 1.63) and reduced by 37% for Black African/Caribbean women compared with White women, with the confidence interval crossing unity (HR 0.63, CI95% 0.38, 1.03).

9.6.2 Non-Diabetic Cohorts

In the fully adjusted models, the ethnic differences in the non-diabetic random samples were comparable to the fully adjusted models for the diabetic population. In the random sample, the risk of incident non-fatal CHD was increased by 52% for South Asian males compared with White males (HR 1.52, CI95% 1.21, 1.91) and reduced by 66% for Black African/Caribbean males compared with White males (HR 0.33, CI95% 0.71, 0.64). In the matched sample, ethnic differences were more modest than in the random sample, but in the same direction. The risk of incident non-fatal CHD was increased by 36% for South Asian males compared with White

males (HR 1.36, CI95% 1.18, 1.57) and reduced by 55% for Black African/Caribbean males compared to White males (HR 0.45, CI95% 0.32, 0.63).

The excess risk of CHD for South Asian females was comparable in both the random and matched non-diabetic samples, with stronger evidence for increased risk in the matched sample than in the random sample (HR 1.40, CI95% 1.17, 1.66 vs HR 1.39, CI95% 0.99, 1.96). Evidence for a reduction in risk for African/Caribbean women compared to White women was found for the matched non-diabetic population only (HR 0.73, CI95% 0.55, 0.99).

9.6.3 Sensitivity Analysis

Sensitivity analyses replicating all of the models described above using only the complete cases available for the fully adjusted models are presented in table 9.8b. In diabetics and non-diabetics of both genders, the point estimates and confidence intervals are comparable to those derived in the main analysis, with no difference in the interpretation of the relative risk of incident non-fatal CHD by ethnic group.

Sensitivity analyses replicating the analysis of non-fatal CHD in patients without both T2DM and T1DM are presented in table 9.6c. The patterning of differences in CHD risk by ethnic group closely matches those of the main analysis.

Table 9.8a Incidence of non-fatal CHD amongst subjects with and without type 2 diabetes

		Males								Females							
		Crude		Age		Framingham		Full		Crude		Age		Framingham		Full	
T2DM	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
	South Asian	1.10	[0.97,1.25]	1.38	[1.21,1.58]	1.49	[1.29,1.73]	1.42	[1.21,1.68]	1.00	[0.85,1.18]	1.37	[1.16,1.62]	1.51	[1.25,1.82]	1.31	[1.05,1.63]
	Black	0.54	[0.41,0.72]	0.60	[0.45,0.80]	0.57	[0.40,0.80]	0.58	[0.41,0.83]	0.53	[0.37,0.76]	0.62	[0.44,0.88]	0.73	[0.51,1.06]	0.63	[0.38,1.03]
Subjects		105,025		105, 025		99,917		93,787		91,525		91,525		86,684		80,587	
Non T2DM	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
Random	South Asian	0.98	[0.82,1.18]	1.77	[1.49,2.11]	1.63	[1.30,2.03]	1.52	[1.20,1.93]	0.71	[0.54,0.93]	1.61	[1.22,2.11]	1.65	[1.20,2.29]	1.39	[0.99,1.96]
Sample	Black	0.25	[0.16,0.37]	0.44	[0.29,0.67]	0.34	[0.18,0.62]	0.33	[0.17,0.65]	0.42	[0.29,0.61]	1.00	[0.69,1.44]	1.03	[0.65,1.64]	0.91	[0.57,1.46]
Subjects		321,822		321,822		123,953		107,470		344,108		344,108		140,495		124,581	
Non T2DM	White (ref)	1	--	n/a		1	--	1	--	1	--	n/a		1	--	1	--
Matched	South Asian	1.66	[1.48, 1.85]	Already matched		1.59	[1.39,1.83]	1.36	[1.18,1.57]	1.56	[1.37,1.77]	Already matched			[1.44,1.98]	1.40	[1.17,1.66]
Sample	Black	0.62	[0.50,0.76]	on age		0.54	[0.40,0.72]	0.45	[0.32,0.63]	0.85	[0.68,1.08]	on age			[0.69,1.16]	0.73	[0.55,0.99]
Subjects		406984		--		200374		174560		451053		--		234678		208839	

***All models account for clustering by practice**

Table 9.6b Incidence of non-fatal CHD amongst subjects with and without type 2 diabetes – sensitivity analysis using complete cases only

		Males								Females							
		Crude		Age		Framingham		Full		Crude		Age		Framingham		Full	
T2DM	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
	South Asian	1.17	[1.00,1.36]	1.47	[1.25,1.72]	1.52	[1.30,1.78]	1.42	[1.21,1.68]	1.02	[0.83,1.25]	1.38	[1.12,1.71]	1.49	[1.20,1.85]	1.31	[1.05,1.63]
	Black	0.56	[0.39,0.79]	0.60	[0.42,0.86]	0.62	[0.43,0.88]	0.58	[0.41,0.83]	0.57	[0.35,0.94]	0.66	[0.41,1.08]	0.69	[0.42,1.12]	0.63	[0.38,1.03]
Subjects		93,787		93,787		93,787		93,787		80,587		80,587		80,587		80,587	
Non T2DM	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
Random	South Asian	1.07	[0.85,1.34]	1.59	[1.26,2.00]	1.69	[1.34,2.13]	1.52	[1.20,1.93]	0.83	[0.60,1.17]	1.44	[1.03,2.03]	1.66	[1.18,2.33]	1.39	[0.99,1.96]
Sample	Black	0.23	[0.12,0.45]	0.33	[0.17,0.65]	0.35	[0.18,0.68]	0.33	[0.17,0.65]	0.65	[0.41,1.03]	1.07	[0.67,1.69]	1.16	[0.73,1.85]	0.91	[0.57,1.46]
Subjects		107,470		107,470		107,470		107,470		124581		124581		124581		124581	
Non T2DM	White (ref)	1	--	n/a		1	--	1	--	1	--	n/a		1	--	1	--
Matched	South Asian	1.49	[1.29,1.72]	Already matched		1.60	[1.38,1.86]	1.36	[1.18,1.57]	1.45	[1.22,1.71]	Already matched		1.66	[1.41,1.96]	1.40	[1.17,1.66]
Sample	Black	0.47	[0.34,0.66]	on age		0.49	[0.35,0.69]	0.45	[0.32,0.63]	0.81	[0.60,1.09]	on age		0.88	[0.65,1.18]	0.73	[0.55,0.99]
Subjects		176560		--		176560		176560		208839		--		208839		208839	

***All models account for clustering by practice**

Table 9.c Incidence of non-fatal CHD amongst subjects without type 2 diabetes – sensitivity analysis excluding patients with T1DM

		Males								Females							
		Crude		Age		Framingham		Full		Crude		Age		Framingham		Full	
Non T2DM	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
Non T1DM	South Asian	0.99	[0.83,1.18]	1.79	[1.50,2.13]	1.64	[1.32,2.05]	1.52	[1.20,1.93]	0.71	[0.54,0.93]	1.61	[1.21,2.12]	1.65	[1.18,2.30]	1.38	[0.97,1.96]
	Black	0.25	[0.16,0.37]	0.45	[0.30,0.68]	0.34	[0.19,0.63]	0.33	[0.17,0.65]	0.41	[0.27,0.61]	0.98	[0.66,1.45]	1.05	[0.66,1.67]	0.95	[0.60,1.50]
Subjects		320,667		320,667		122,961		106,572		324,281		324,281		139,806		123,946	

9.7 Adjusted risk of non-fatal CHD by ethnic minority subgroup

The fourth objective of this study was to examine differences in the adjusted risk of CHD for ethnic minority subgroups in comparison with the White majority ethnic group.

Table 9.9a presents the results of the Cox proportional hazards models for ethnic differences in risk of non-fatal CHD between the main South Asian and Black subgroups and the White group as a whole, stratified by diabetic status and gender. In the diabetic population, risk was significantly increased for all South Asian subgroups, and significantly reduced for Caribbean groups in relation to the White group. No differences between African and White groups were evident in the adjusted analysis.

In both non-diabetic populations, risk was increased for Pakistani and Indian groups compared to white, with no evidence for a difference between the Bangladeshi and White populations. Strong evidence was found for a reduction in risk for Caribbean men compared to White men, with no differences between African/Caribbean and White groups for women after full adjustment.

9.7.1 Diabetic Cohort

In the fully adjusted models for diabetic males, risk of incident non-fatal CHD was almost doubled for Pakistani and Bangladeshi groups compared with White (Pakistani HR 1.90, CI95% 1.40, 2.59, Bangladeshi HR 1.90, CI95% 1.07, 3.36). The risk for Indian males was increased by 32% in comparison with White males (HR 1.32, CI95% 1.06, 1.64) and reduced by just under half for Caribbean males (HR 0.52, CI95% 0.31, 0.86).

Amongst diabetic females, all risk differences were smaller than for males. The largest risk difference observed was for Indian women compared with White (HR 1.52, CI95% 1.15, 2.00), followed by Pakistani (HR 1.43, CI95% 0.92, 2.24) and Bangladeshi groups (HR 1.24 CI95% 0.58, 2.62). The reduction in the Caribbean group was slightly less than in the male diabetic population (HR 0.60, CI95% 0.34, 1.06).

9.7.2 Non-diabetic Cohort

a) Random sample

The patterning of risk differences in the non-diabetic cohort differed from that of the diabetic cohort. Compared with diabetic males, the risk difference (relative to the White group) for non-diabetic males was larger for Indians (HR 1.42, CI95% 1.03, 1.95) and smaller for Pakistanis (HR 1.79, CI95% 1.07, 3.02). No evidence for a difference between Bangladeshi and White groups was found. The difference between Caribbean and White males was much larger in the non-diabetic random sample compared with the diabetic cohort, with a 92% reduction in risk in the former (HR 0.08, CI95% 0.01, 0.58).

Though the crude and age-adjusted models for female non-diabetics suggested an increase in risk for Pakistani and Indian groups, no ethnic differences for any South Asian or Black subgroups were evident in the Framingham and fully adjusted models.

b) Matched sample

Ethnic differences in the matched non-diabetic sample were comparable in direction and magnitude to those of the random non-diabetic sample after adjusting for all confounders and controlling for the effects of clustering by practice.

9.7.3 Sensitivity Analysis

Sensitivity analyses replicating all of the models described above using only the complete cases available for the fully adjusted models are presented in table 9.9b. In all subgroups the point estimates were increased slightly for South Asians and decreased slightly for all other groups. No differences in the interpretation of the relative risk of incident non-fatal CHD by ethnic sub group were evident.

Table 9.9a Incidence of non-fatal CHD amongst subjects with and without type 2 diabetes – ethnic subgroup breakdown

		Male								Female							
		Crude		Age		Framingham		Full		Crude		Age		Framingham		Full	
T2DM	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
	Indian	0.97	[0.81,1.16]	1.17	[0.98,1.40]	1.34	[1.10,1.65]	1.32	[1.06,1.64]	1.15	[0.92,1.44]	1.48	[1.18,1.85]	1.66	[1.30,2.13]	1.52	[1.15,2.00]
	Pakistani	1.61	[1.31,1.99]	2.20	[1.76,2.76]	2.31	[1.74,3.07]	1.90	[1.40,2.59]	1.07	[0.77,1.47]	1.53	[1.10,2.13]	1.79	[1.25,2.57]	1.43	[0.92,2.24]
	Bangladeshi	1.56	[1.03,2.36]	2.03	[1.31,3.13]	2.18	[1.30,3.64]	1.90	[1.07,3.36]	0.80	[0.39,1.64]	1.34	[0.65,2.77]	1.22	[0.57,2.62]	1.24	[0.58,2.62]
	African	0.51	[0.31,0.85]	0.65	[0.40,1.07]	0.72	[0.39,1.35]	0.74	[0.39,1.43]	0.37	[0.12,1.14]	0.50	[0.17,1.53]	0.68	[0.23,2.05]	0.72	[0.20,2.56]
	Caribbean	0.57	[0.37,0.87]	0.57	[0.37,0.89]	0.50	[0.31,0.81]	0.52	[0.31,0.86]	0.57	[0.36,0.92]	0.62	[0.40,0.99]	0.69	[0.42,1.13]	0.60	[0.34,1.06]
Subjects		105,455		105,455		100,009		94,439		91,525		91,525		86,684		80,587	
Non	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
T2DM	Indian	1.02	[0.79,1.31]	1.61	[1.25,2.06]	1.49	[1.10,2.01]	1.42	[1.03,1.95]	0.71	[0.49,1.03]	1.41	[0.96,2.06]	1.47	[0.95,2.28]	1.38	[0.85,2.23]
Random	Pakistani	1.01	[0.72,1.41]	2.28	[1.63,3.19]	1.92	[1.21,3.05]	1.79	[1.07,3.02]	0.73	[0.42,1.29]	2.06	[1.14,3.72]	1.82	[0.85,3.87]	1.23	[0.53,2.88]
sample	Bangladeshi	0.86	[0.43,1.74]	1.75	[0.86,3.57]	1.48	[0.52,4.22]	1.08	[0.32,3.68]	0.22	[0.03,1.59]	0.69	[0.09,4.99]	0.98	[0.14,6.98]	--¥	--
	African	0.24	[0.13,0.45]	0.49	[0.26,0.94]	0.48	[0.21,1.08]	0.54	[0.24,1.22]	0.32	[0.17,0.61]	0.94	[0.52,1.71]	1.36	[0.69,2.70]	1.26	[0.63,2.50]
	Caribbean	0.21	[0.10,0.46]	0.28	[0.13,0.60]	0.15	[0.04,0.56]	0.08	[0.01,0.58]	0.60	[0.37,0.97]	1.14	[0.70,1.86]	1.00	[0.52,1.90]	0.84	[0.44,1.61]
Subjects		315,215		315,215		122,288		106,124		337,324		337,324		138,802		123,160	
Non	White (ref)	1	--			1	--	1	--	1	--			1	--	1	--
T2DM	Indian	1.54	[1.33,1.78]	n/a		1.39	[1.15,1.68]	1.20	[1.00,1.46]	1.54	[1.32,1.80]	n/a		1.64	[1.35,1.99]	1.39	[1.12,1.72]
Matched	Pakistani	1.97	[1.61,2.42]	Already matched		1.94	[1.48,2.54]	1.63	[1.22,2.17]	2.35	[1.76,3.15]	Already matched on age		2.70	[1.99,3.67]	1.96	[1.39,2.77]
sample	Bangladeshi	2.96	[2.17,4.04]	On age		2.54	[1.53,4.21]	1.68	[0.88,3.23]	1.17	[0.56,2.46]			0.84	[0.29,2.42]	0.74	[0.27,2.02]
	African	0.69	[0.49,0.97]			0.57	[0.35,0.93]	0.50	[0.29,0.85]	0.82	[0.53,1.28]			0.80	[0.47,1.38]	0.68	[0.39,1.19]
	Caribbean	0.57	[0.42,0.78]			0.48	[0.31,0.73]	0.37	0.85	0.85	[0.64,1.13]			0.90	[0.64, 1.28]	0.74	[0.51,1.08]
Subjects		406,984				200,374		176,560		451,053				234,678		208,839	

¥Unable to calculate hazard ratios and point estimates due to zero counts in some cells

Table 9.9b Incidence of non-fatal CHD amongst subjects with and without type 2 diabetes – ethnic subgroup breakdown, complete case analysis

		Male								Female							
		Crude		Age		Framingham		Full		Crude		Age		Framingham		Full	
T2DM	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
	Indian	1.13	[0.92,1.39]	1.35	[1.09,1.68]	1.41	[1.14,1.75]	1.31	[1.06,1.63]	1.21	[0.93,1.58]	1.58	[1.21,2.06]	1.70	[1.29,2.23]	1.51	[1.15,1.99]
	Pakistani	1.48	[1.10,1.99]	2.04	[1.49,2.78]	2.09	[1.53,2.85]	1.87	[1.37,2.54]	1.10	[0.71,1.69]	1.56	[1.00,2.42]	1.70	[1.10,2.65]	1.40	[0.90,2.18]
	Bangladeshi	1.52	[0.89,2.61]	2.08	[1.19,3.64]	2.12	[1.22,3.70]	1.88	[1.05,3.35]	0.79	[0.38,1.67]	1.31	[0.61,2.82]	1.44	[0.67,3.09]	1.19	[0.56,2.52]
	African	0.56	[0.28,1.10]	0.72	[0.37,1.39]	0.74	[0.38,1.43]	0.74	[0.38,1.42]	0.50	[0.14,1.77]	0.68	[0.19,2.40]	0.72	[0.20,2.54]	0.71	[0.20,2.54]
	Caribbean	0.56	[0.34,0.93]	0.55	[0.33,0.91]	0.55	[0.33,0.92]	0.52	[0.31,0.86]	0.60	[0.33,1.10]	0.64	[0.36,1.15]	0.67	[0.37,1.20]	0.59	[0.33,1.06]
Subjects		94,439		94,439		94,439		94,439		80,587		80,587		80,587		80, 587	
Non	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
T2DM	Indian	1.05	[0.78,1.42]	1.42	[1.05,1.92]	1.54	[1.14,2.09]	1.42	[1.03,1.95]	0.88	[0.54,1.41]	1.36	[0.85,2.17]	1.56	[0.98,2.50]	1.38	[0.85,2.23]
Random	Pakistani	1.10	[0.69,1.76]	2.09	[1.32,3.32]	2.10	[1.31,3.37]	1.79	[1.07,3.02]	0.76	[0.35,1.64]	1.53	[0.68,3.46]	1.77	[0.78,4.01]	1.23	[0.53,2.88]
sample	Bangladeshi	0.76	[0.24,2.40]	1.40	[0.42,4.70]	1.33	[0.40,4.48]	1.08	[0.32,3.68]	--¥	--	--	--	--	--	--	--
	African	0.32	[0.15,0.71]	0.52	[0.23,1.18]	0.56	[0.25,1.26]	0.54	[0.24,1.22]	0.65	[0.33,1.30]	1.34	[0.68,2.65]	1.52	[0.77,3.00]	1.26	[0.63,2.50]
	Caribbean	0.076	[0.01,0.54]	0.084	[0.01,0.61]	0.086	[0.01,0.62]	0.080	[0.01,0.58]	0.78	[0.42,1.46]	1.06	[0.56,2.01]	1.14	[0.60,2.15]	0.84	[0.44,1.61]
Subjects		106,124		106,124		106,124		106,124		123,160		123,160		123,160		123,160	
Non	White (ref)	1	--			1	--	1	--	1	--			1	--	1	--
T2DM	Indian	1.25	[1.02,1.52]	n/a		1.35	[1.11,1.66]	1.20	[1.00,1.46]	1.42	[1.16,1.74]	n/a		1.63	[1.34,2.00]	1.39	[1.12,1.72]
Matched	Pakistani	1.99	[1.52,2.60]	Already matched		2.09	[1.59,2.76]	1.63	[1.22,2.17]	2.31	[1.65,3.25]	Already matched		2.64	[1.88,3.73]	1.96	[1.39,2.77]
sample	Bangladeshi	2.20	[1.25,3.88]	On age		2.29	[1.27,4.12]	1.68	[0.88,3.23]	0.82	[0.29,2.31]	On age		0.95	[0.34,2.71]	0.74	[0.27,2.02]
	African	0.49	[0.29,0.83]			0.53	[0.31,0.91]	0.50	[0.29,0.85]	0.77	[0.43,1.35]			0.84	[0.47,1.48]	0.68	[0.39,1.19]
	Caribbean	0.40	[0.25,0.65]			0.41	[0.25,0.66]	0.37	[0.22,0.60]	0.83	[0.57,1.22]			0.89	[0.60,1.30]	0.74	[0.51,1.08]
Subjects		176,560				176,560		176,560		208,839				208,839		208,839	

¥Unable to calculate hazard ratios and point estimates due to zero counts in some cells

9.8 Testing the assumptions underlying Cox proportional hazards regression

Cox regression modelling makes no assumptions about the form of the baseline hazard. The advantage of cox models over Poisson is the use of infinitesimal time bands, which allows the relationship between rates and time to be modelled as finely as possible. The method does not require a particular survival model but it is not truly non-parametric because it assumes that the effects of the predictor variables upon survival are constant over time. Cox regression is considered a 'semiparametric' procedure because the baseline hazard function, h0 (t), (and the probability distribution of the survival times) does not have to be specified. Since the baseline hazard is not specified and the hazard function is not restricted to a specific form, the semi-parametric model is widely used.

A key assumption underlying the Cox regression model is that of proportional hazards. The assumption requires that, while the individual and baseline hazard rates may vary with time, their ratio must remain constant. This assumption implies that the effect of the predictor variables included in the Cox model do not change during the time covered by the study.
The assumptions can be tested both graphically and inferentially.

Cumulative incidence plots (also known as Nelson-Aalen plots) were used to test the proportionality of the effect of ethnicity on incident non-fatal CHD.

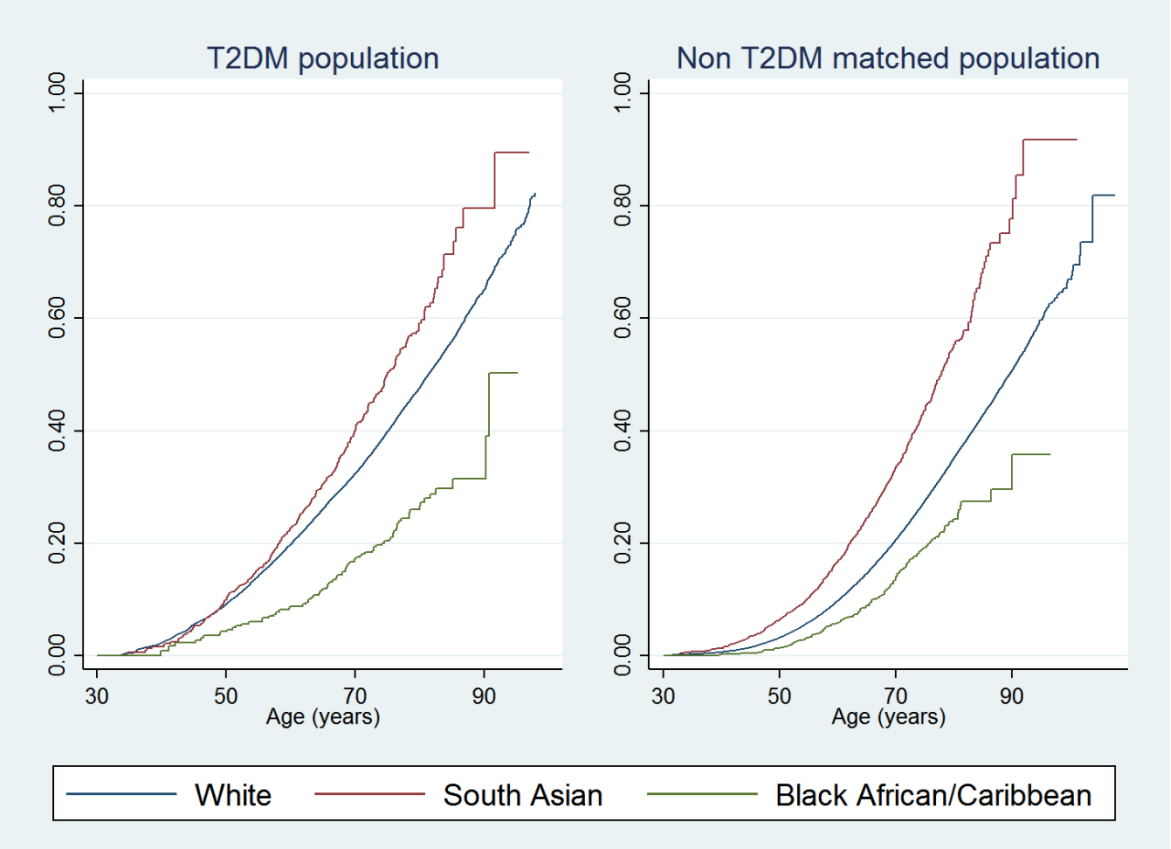


Figure 9.5 Nelson-Aalen cumulative incidence plots for patients with and without type 2 diabetes

The plots show that the curves for each ethnic group do not cross, except for at the very beginning of follow-up for the diabetic population. This may be due to the fact that the time scale used in the Cox regression analysis allowed for delayed entry, meaning that patients entered the study at different points in time. In this situation graphical methods may be less useful as very few subjects may be in follow-up when the time scale beings, and thus cumulative incidence rates may be unstable at that point in time.

Chapter 10
Ethnic and gender differences in CHD risk amongst patients with and without type 2 diabetes: Results for fatal and non-fatal CHD combined

10.1 Summary

This chapter reports on the findings of the cohort study identifying ethnic and gender differences in incidence of fatal and non-fatal CHD combined, in a subsample of patients with linked ONS mortality data. Section 10.2 describes the derivation of the study populations and section 10.3 their demographic and baseline clinical characteristics. The next section presents the findings of the primary analyses, which examine ethnic differences in CHD risk amongst patients with and without T2DM. Sections 10.6 and 10.7 presents the results from secondary analyses which explored heterogeneity of risk between ethnic minority subgroups. Section 10.8 presents a comparison of results from the main and ONS linked studies.

10.2 Study Populations

The denominator populations required for this secondary study of fatal and non-fatal CHD were drawn from the samples identified in chapter 9. Of the 196,254 individuals with type 2 diabetes described in the main analysis, 86,780 (44%) had linked ONS mortality data. Of the 665,930 non-diabetic individuals included in the previous analysis, 209,543 (31%) had linked ONS mortality data and were included. A final sample of 296,342 individuals was included in the study of fatal and non-fatal CHD (Figure 10.1).

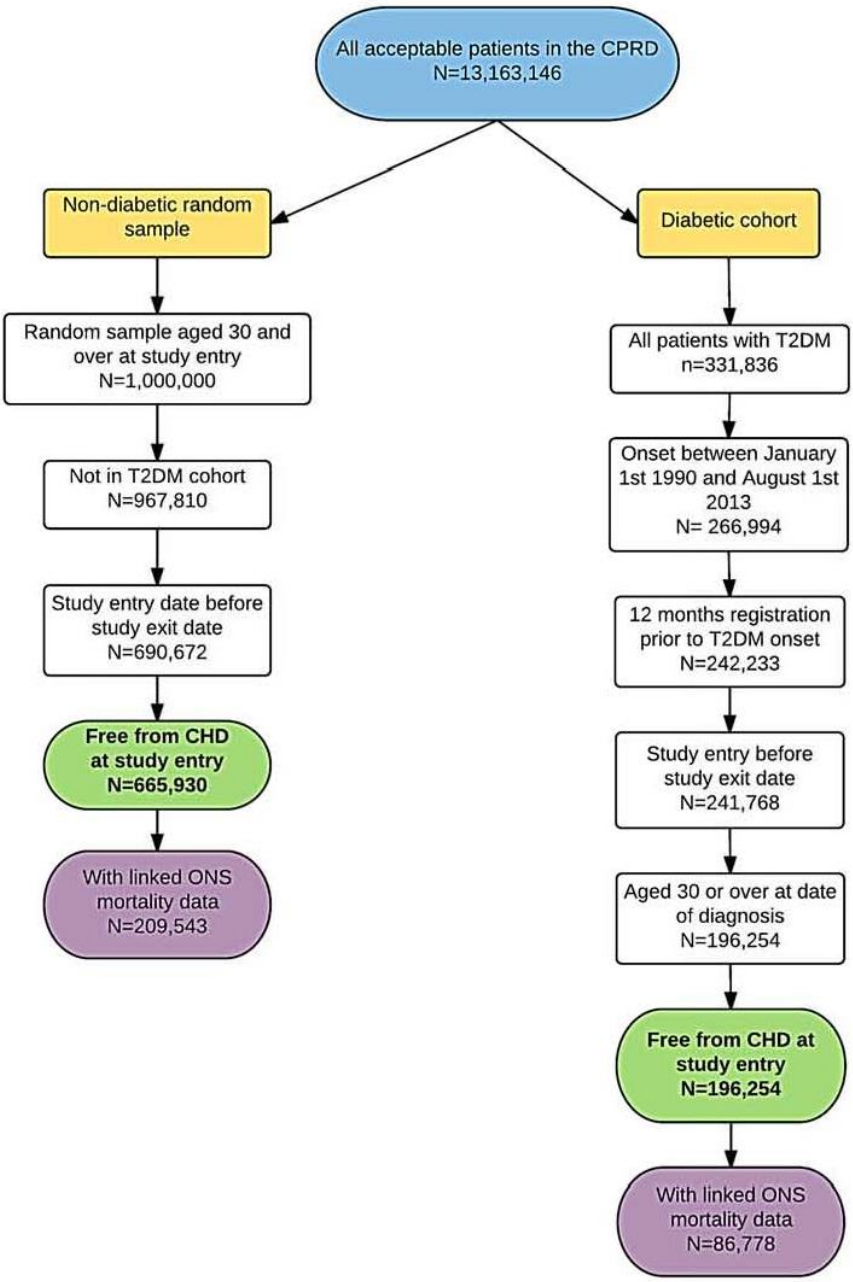


Figure 10.1 Derivation of study cohorts

The gender and ethnic breakdown of the CPRD and ONS linked study populations is illustrated in figure 10.2.

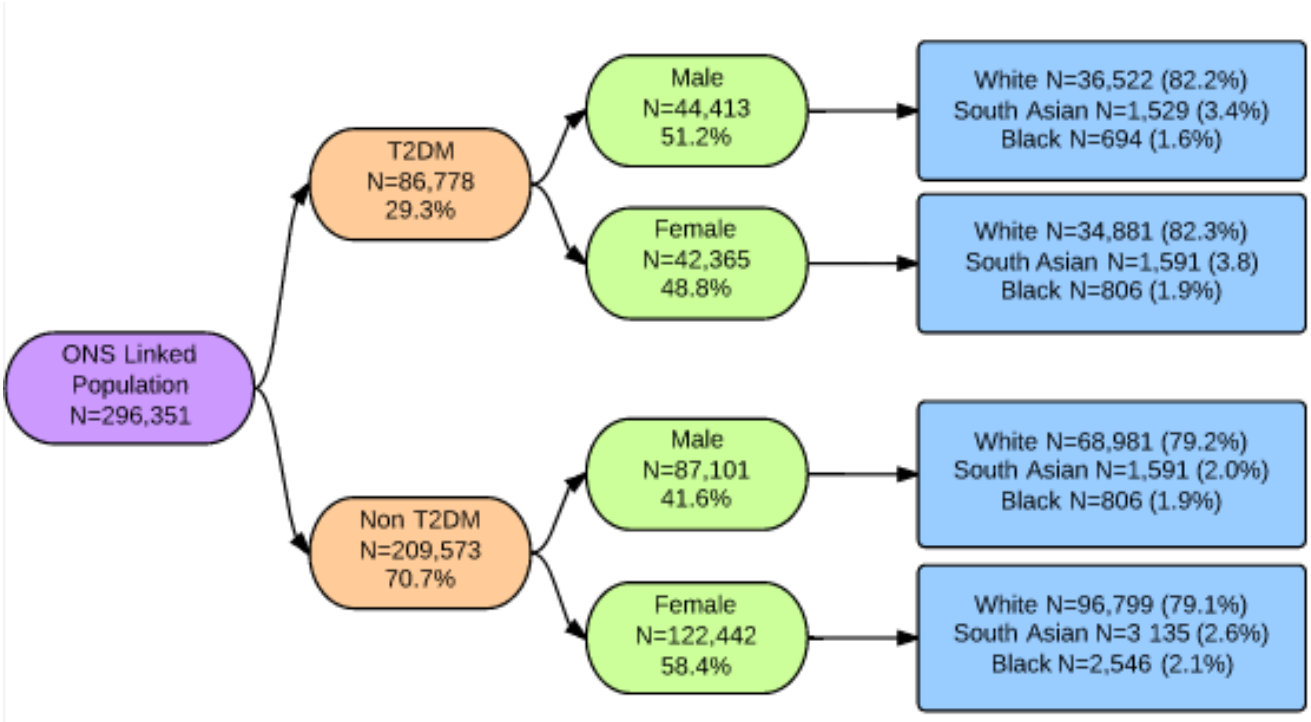


Figure 10.2 Breakdown of study populations by diabetic status, gender and ethnic group

10.3 Demographics and Baseline Characteristics

10.3.1 Demographics

Table 10.1 shows the demographic and clinical characteristics of the subjects according to diabetic status. In the diabetic cohort, a total of 9,157 incident fatal and non-fatal CHD events occurred, over a total of 480,030 person-years of follow-up. In the non-diabetic cohort, a total of 13,854 incident fatal and non-fatal CHD events occurred over a total of 1.7 million person-years of follow-up. Median follow-up time was 4.8 years for the diabetic cohort (SD 4.1 years) and 6.7 years for the non-diabetic cohort (SD 6.2 years).

In the diabetic cohort, the proportion of patients experiencing a fatal and non-fatal CHD event was highest in the White population (10.7%), followed by South Asian (9.8%) and Black African/Caribbean (4.6%). In the non-diabetic population, events were most common in the

White group (6.9%), followed by South Asian (3.7%) and Black African/Caribbean (1.2%) groups.

The median age at baseline was 63.8 for diabetic participants (SD 13.1) and 44.0 for non-diabetic participants (SD 17.1). In both the diabetic and non-diabetic cohorts, the median age at study entry was 10 years lower for South Asian and 8 years lower for Black patients compared with White patients.

10.3.2 Medications

The prescription of all medications was slightly higher in the ONS-inked subset compared to the main study sample. The largest difference was for antihypertensive medications, which were prescribed for 36.5% of the ONS linked subset and 28.6% of the main study sample. Over the course of follow-up, insulin was prescribed for 15.2% of diabetics, metformin for 71.9% and other antidiabetic drugs for 52%. Prescription of OAD and insulin was comparable between ethnic groups, while metformin prescription was 11% higher for the South Asian group (83.0%) compared with the White group (71.63%). The prescription of cardiovascular medications was 3–6 times higher in the diabetic compared with the non-diabetic cohort and highest in the White group compared with all others.

10.3.3 Clinical covariates

Completeness of all clinical covariates was higher in the ONS-linked subset compared with the main study sample. A baseline blood pressure value was available for 99.9% of diabetics and 93.4% of non-diabetics, with hypertension recorded in 42.9% of diabetics and 22.3% of non-diabetics. Prevalence of hypertension at baseline was higher in the ONS-linked subset compared with the main study sample. Hypertension was most prevalent in the White group, followed by Black and South Asian groups in both diabetic and non-diabetic cohorts.

Total cholesterol values were available for 98.0% of diabetics and 50.1% of non-diabetics, with raised cholesterol at baseline evident in 30.7% of diabetics and 31.8% of non-diabetics. Raised cholesterol was more prevalent amongst non-diabetics in the ONS subset compared to the

main study sample. The proportion of patients with raised baseline cholesterol did not vary by ethnic group amongst diabetics, but in the non-diabetic cohort was highest in the White group, followed by South Asian and Black groups.

Smoking status was recorded for 99.0% of diabetics and 94.1% of non-diabetics. At baseline, fewer diabetics than non-diabetics had their latest smoking status recorded as "current smoker", (15.6% vs. 26.1%). In both the diabetic and non-diabetic cohorts, the proportion of current and ex-smokers was highest in the White group and lowest in the South Asian group.

Alcohol usage was recorded for 95.2% of diabetics and 83.2% of non-diabetics. The majority of South Asians in both diabetic and non-diabetics had their baseline alcohol status recorded as "non-drinker" (65.3% diabetic, 55.4% non-diabetic). In both diabetic cohorts, heavy drinking was 4–8 times more prevalent in the White group than in the South Asian and Black groups.

Body mass index was recorded for 98.1% of diabetics and 83.9% of non-diabetics. While 81.0% of diabetics were overweight or obese, only 41.0% of non-diabetics were so. In the diabetic cohort, the majority of White and Black groups were obese, while the majority of South Asian patients fell into the overweight category. In the non-diabetic cohort, the majority of patients of all ethnic groups fell into the normal weight category.

A total of 98.2% of diabetics had HbA1c values recorded at baseline, of whom 72.9% had a baseline value greater than 6.5%. The proportion of patients with HbA1c greater than or equal to 6.5% was highest for South Asian patients (77.5%), followed by Black (72.9%) and White (68.3%) patients.

Table 10.1 Demographic, clinical, and therapeutic characteristics of CPRD patients with and without T2DM

	T2DM N=86,778										Non T2DM N=209,543									
	White		South Asian		Black		Other		Total		White		South Asian		Black		Other		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Total	71,403	100	3,120	100	1,500	100	10,755	100	86,778	100	165,780	100	4,872	100	3,833	100	35,058	100	209,543	100
Male	36,522	51.1	1,529	49	694	46.3	5,668	52.7	44,413	51.2	68,981	41.6	1,737	35.7	1,287	33.6	15,096	43.1	87,101	41.6
Female	34,881	48.9	1,591	51	806	53.7	5,087	47.3	42,365	48.8	96,799	58.4	3,135	64.3	2,546	66.4	19,962	56.9	122,442	58.4
Incident fatal/non-fatal CHD	7,651	10.7	307	9.8	69	4.6	1,128	10.5	9,155	10.5	11,367	6.9	182	3.7	46	1.2	2,259	6.4	13,854	6.6
Median follow-up (SD)	7.8	±4.3	7.8	±4.4	8.0	±4.5	7.4	±4.2	7.8	±4.3	12.5	±6.2	8.2	±5.7	7.8	±5.8	11.5	±6.1	12.2	±6.2
Age at study entry																				
30-39	2,531	3.5	364	11.7	125	8.3	461	4.3	3,481	4	66,640	40.2	3,088	63.4	2,306	60.2	15,321	43.7	87,355	41.7
40-49	8,306	11.6	789	25.3	339	22.6	1,436	13.4	10,870	12.5	30,062	18.1	800	16.4	869	22.7	63, 84	18.2	38,115	18.2
50-59	15,436	21.6	918	29.4	387	25.8	2,454	22.8	19,195	22.1	24,749	14.9	473	9.7	318	8.3	4,427	12.6	29,967	14.3
60-69	20,142	28.2	675	21.6	378	25.2	2,784	25.9	23,979	27.6	19,667	11.9	300	6.2	182	4.7	3,331	9.5	23,480	11.2
70+	24,988	35	374	12	271	18.1	3,620	33.7	29,253	33.7	24,662	14.9	211	4.3	158	4.1	5,595	16	30,626	14.6
Median age (SD)	62.9	±12.0	54.0	±11.3	58.8	±11.8	61.5	±12.2	62.5	±12.1	5.5	±15.2	41.5	±13.4	40.3	±12.7	48.1	±15.6	49.9	±15.3
Median age Male (SD)	63.0	±12.3	54.0	±11.9	58.4	±12.6	61.8	±12.6	62.5	±12.4	45.5	±15.9	37.9	±13.5	39.2	±12.7	42.7	±16.3	44.6	±15.9
Median age Female (SD)	66.1	±13.4	53.9	±12.2	56.0	±12.5	65.3	±14.1	65.4	±13.6	43.7	±17.8	33.8	±12.0	35.6	±11.3	42.6	±18.6	42.7	±17.9
Individual level IMD score																				
1 (most affluent)	13,303	18.6	456	14.6	76	5.1	2,415	22.5	16,250	18.7	37,038	22.3	842	17.3	264	6.9	8,756	25	46,900	22.4
2	16,411	23	493	15.8	130	8.7	2,370	22	19,404	22.4	39,254	23.7	850	17.4	346	9	8,000	22.8	48,450	23.1
3	14,289	20	673	21.6	228	15.2	2,289	21.3	17,479	20.1	32,846	19.8	1,010	20.7	563	14.7	7,114	20.3	41,533	19.8
4	14,402	20.2	653	20.9	471	31.4	2,104	19.6	17,630	20.3	30,688	18.5	1,065	21.9	1,234	32.2	6,341	18.1	39,328	18.8
5 (least affluent)	12,710	17.8	828	26.5	591	39.4	1,533	14.3	15,662	18	24,690	14.9	1,058	21.7	1,389	36.2	4,607	13.1	31,744	15.1
Unknown	288	0.4	17	0.5	4	0.3	44	0.4	353	0.4	1,264	0.8	47	1.0	37	1	240	0.7	1,588	0.8
Medications ever prescribed																				
Insulin	11,082	15.5	454	14.6	279	18.6	1,398	13	13,213	15.2										
Metformin	51,117	71.6	2,589	83	1,186	79.1	7,527	70	62,419	71.9										
OADs	37,141	52	1,692	54.2	762	50.8	5,533	51.4	45,128	52										
Antihypertensives	59,886	83.9	2,294	73.5	1,142	76.1	8,529	79.3	71,851	82.8	63,142	38.1	1,156	23.7	985	25.7	11,140	31.8	76,423	36.5
Lipid lowering	56,095	78.6	2,448	78.5	1,050	70	7,673	71.3	67,266	77.5	29,201	17.6	660	13.5	372	9.7	4,036	11.5	34,269	16.4
Aspirin	39,872	55.8	1,638	52.5	695	46.3	5,282	49.1	47,487	54.7	30,461	18.4	543	11.1	316	8.2	5,205	14.8	36,525	17.4

Table 10.1 Continued...

	T2DM N=86,778										Non T2DM N=209,543									
	White		South Asian		Black		Other		Total		White		South Asian		Black		Other		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Blood Pressure																				
≤140/90 mmHg	40,531	56.8	1,914	61.3	877	58.5	6,121	56.9	49,443	57	117629	71	3982	81.7	2950	77	24412	69.6	148973	71.1
>140/90 mmHg	30,809	43.1	1,206	38.7	621	41.4	4,605	42.8	37,241	42.9	37981	22.9	558	11.5	659	17.2	7553	21.5	46751	22.3
Unknown	63	0.1	-	0	2	0.1	29	0.3	94	0.1	10170	6.1	332	6.8	224	5.8	3093	8.8	13819	6.6
Cholesterol																				
≤5 mmol/L	48,394	67.8	2,200	70.5	1,016	67.7	6,886	64	58,496	67.4	30622	18.5	1307	26.8	1000	26.1	5344	15.2	38273	18.3
>5 mmol/L	21,826	30.6	910	29.2	465	31	3,443	32	26,644	30.7	55797	33.7	1220	25	795	20.7	8923	25.5	66735	31.8
Unknown	1,183	1.7	10	0.3	19	1.3	426	4	1,638	1.9	79361	47.9	2345	48.1	2038	53.2	20791	59.3	104535	49.9
Smoking Status																				
Non	36,743	51.5	2,423	77.7	1,058	70.5	5,908	54.9	46,132	53.2	88171	53.2	3773	77.4	2684	70	19210	54.8	113838	54.3
Current	11,162	15.6	294	9.4	161	10.7	1,592	14.8	13,209	15.2	48723	29.4	638	13.1	770	20.1	9210	26.3	59341	28.3
Ex	22,887	32.1	386	12.4	273	18.2	3,015	28	26,561	30.6	20005	12.1	261	5.4	225	5.9	3468	9.9	23959	11.4
Unknown	611	0.9	17	0.5	8	0.5	240	2.2	876	1	8881	5.4	200	4.1	154	4	3170	9	12405	5.9
Alcohol consumption																				
Non	17,229	24.1	2,037	65.3	640	42.7	2,884	26.8	22,790	26.3	25137	15.2	2701	55.4	1543	40.3	5996	17.1	35377	16.9
Moderate	45,332	63.5	888	28.5	760	50.7	6,313	58.7	53,293	61.4	99956	60.3	1306	26.8	1628	42.5	19120	54.5	122010	58.2
High	5,714	8	70	2.2	28	1.9	755	7	6,567	7.6	14228	8.6	71	1.5	86	2.2	2467	7	16852	8
Unknown	3,128	4.4	125	4	72	4.8	803	7.5	4,128	4.8	26459	16	794	16.3	576	15	7475	21.3	35304	16.8
BMI																				
Underweight	675	0.9	19	0.6	5	0.3	91	0.8	790	0.9	3680	2.2	230	4.7	70	1.8	793	2.3	4773	2.3
Normal	11,171	15.6	767	24.6	220	14.7	1,903	17.7	14,061	16.2	67479	40.7	2226	45.7	1273	33.2	14043	40.1	85021	40.6
Overweight	23,112	32.4	1,240	39.7	537	35.8	3,656	34	28,545	32.9	47107	28.4	1295	26.6	1238	32.3	8901	25.4	58541	27.9
Obese	35,226	49.3	1,074	34.4	720	48	4,747	44.1	41,767	48.1	22300	13.5	519	10.7	807	21.1	3852	11	27478	13.1
Unknown	1,219	1.7	20	0.6	18	1.2	358	3.3	1,615	1.9	25214	15.2	602	12.4	445	11.6	7469	21.3	33730	16.1
HbA1c																				
<5.5%	5,202	7.3	178	5.7	110	7.3	773	7.2	6,263	7.2										
5.5-6.4%	16,207	22.7	500	16	268	17.9	2,258	21	19,233	22.2										
≥6.5%	48,760	68.3	2,419	77.5	1,094	72.9	7,422	69	59,695	68.8										
Unknown	1,234	1.7	23	0.7	28	1.9	302	2.8	1,587	1.8										

10.4 Comparison of incident fatal and non-fatal CHD risk according to diabetic status

The entire ONS-linked study population of 296,342 patients was used to compare the incidence of fatal and non-fatal CHD between individuals with and without type 2 diabetes. Table 10.2 displays the crude and adjusted hazard ratios for the entire study population and stratified by ethnic group. The crude risk of non-fatal CHD was 36% higher in the diabetic population than in the non-diabetic reference population (HR 1.36, CI 95 1.32, 1.39).

After adjustment for age gender and individual-level deprivation, the excess risk in the diabetic population was reduced to 20% (HR 1.20, CI95% 117, 1.23). The adjusted risk of fatal and non-fatal CHD was increased in the White diabetic population by 22% compared with the White non-diabetic population (HR 1.22, CI95% 1.18, 1.26). The point estimates for South Asian and Black groups suggested an increased risk in the diabetic population, though the 95% confidence intervals crossed unity.

Table 10.2 Incidence of non-fatal CHD according to diabetic status

	White		South Asian		Black		Other/unknown		Total	
N	237,199		7,992		5,333		45,818		296,341	
Incident CHD	19,034		489		115		3,392		23,030	
	HR	[CI 95]	HR	[CI 95]	HR	[CI 95]	HR	[CI 95]	HR	[CI 95]
Non-T2DM (ref)	1	--	1	--	1	--	1	--	1	--
T2DM crude	1.37	[1.32, 1.41]	1.31	[1.08, 1.58]	1.40	[0.92, 2.12]	1.31	[1.21, 1.40]	1.36	[1.32, 1.39]
T2DM adjusted	1.22	[1.18, 1.26]	1.20	[0.99, 1.45]	1.46	[0.95, 2.26]	1.14	[1.06, 1.23]	1.20	[1.17, 1.23]
N	237,199		7,992		5,333		45,818		296,342	

The results from the Cox regression examining interactions between diabetic status demographic factors are displayed in table 10.3. Tests for effect modification by gender, age and deprivation were found to be statistically significant. No significant interaction between diabetic status and ethnic group was found.

Table 10.3 Risk of non-fatal CHD: Interactions between diabetic status and demographic factors

	NON T2DM		T2DM		*P* value for interaction with diabetic status
	HR	[CI 95]	HR	[CI 95]	
Gender					
Male (ref)	1	--	1	--	
Female	0.54	[0.52, 0.56]	0.64	[0.61, 0.67]	<0.001
Age Group					
30-39 (ref)	1	--	1	--	
40-49	1.40	[1.24, 1.58]	1.14	[0.87, 1.50]	<0.001
50-59	1.78	[1.55, 2.04]	1.27	[0.951.71]	<0.001
60-69	2.46	[2.12, 2.86]	1.37	[1.01, 1.87]	<0.001
70+	3.50	[3.00, 4.12]	1.53	[1.11, 2.11]	<0.001
Individual Level Deprivation Quintile					
IMD 1 (most affluent) (ref)	1	--	1	--	
IMD 5 (least affluent)	1.43	[1.35, 1.53]	1.13	[1.05, 1.22]	<0.001
Ethnic Group					
White (ref)	1	--	1	--	
South Asian	1.64	[1.39, 1.92]	1.37	[1.22, 1.56]	0.993
Black	0.96	[0.91, 1.02]	0.56	[0.44, 0.73]	0.855

The interaction between diabetic status and gender indicates that the relative risk of CHD for females compared with males differs between diabetic and non-diabetic patients. The reduction in CHD risk for females was greater in the non-diabetic cohort (HR 0.54, CI95% 0.52, 0.56 vs. HR 0.64, CI95% 0.61, 0.67). Similarly, the increase in CHD risk by 10-year age band was significantly greater for diabetic patients than for non-diabetic patients. In the non-diabetic cohort, CHD risk was increased by 3.5 times in the oldest group compared with the youngest group (CI95% 3.00, 4.12). In the diabetic cohort, CHD risk was 1.5 times higher in the oldest group than in the youngest group (CI95% 1.11, 2.11).

The increase in CHD risk for those in the least affluent quintile compared with those in the most affluent quintile was larger for non-diabetics than for diabetics (HR 1.43, CI95% 1.35, 1.53 vs. HR 1.13, CI95% 1.05, 1.22). The test for interaction between ethnic group and diabetic status was non-significant, as indicated by the overlapping confidence intervals for diabetic/non-diabetic South Asian patients and diabetic/non-diabetic Black patients. Thus, the difference in CHD risk for ethnic minority groups compared with White does not depend on diabetic status.

10.5 Crude incidence of first fatal and non-fatal CHD event

The analysis presented below describes the crude incidence rates and adjusted hazard ratios for fatal and non-fatal CHD within the diabetic and non-diabetic cohorts separately, stratified by gender and ethnic group. Crude incidence rates are displayed in table 10.4. The crude overall rate of incident fatal and non-fatal CHD was 190.3 per 10,000 person-years in the diabetic cohort (CI95% 186.4, 194.2) and 82.8 per 10,000 person-years in the non-diabetic cohort (CI95% 81.4, 84.1). Compared with the main study sample, overall rates in the ONS-linked subset were raised by 50% in the diabetic cohort and 68% in the non-diabetic cohort.

10.5.1 By ethnic group

In both diabetic and non-diabetic men, crude rates were highest for South Asians, followed by White and Black African/Caribbean groups. This pattern differs from the main study cohort, where the rate of non-fatal CHD for non-diabetic men was highest in the White group. In diabetic and non-diabetic females, rates were highest for the White group, followed by South Asian and Black/African Caribbean (Figure 10.2).

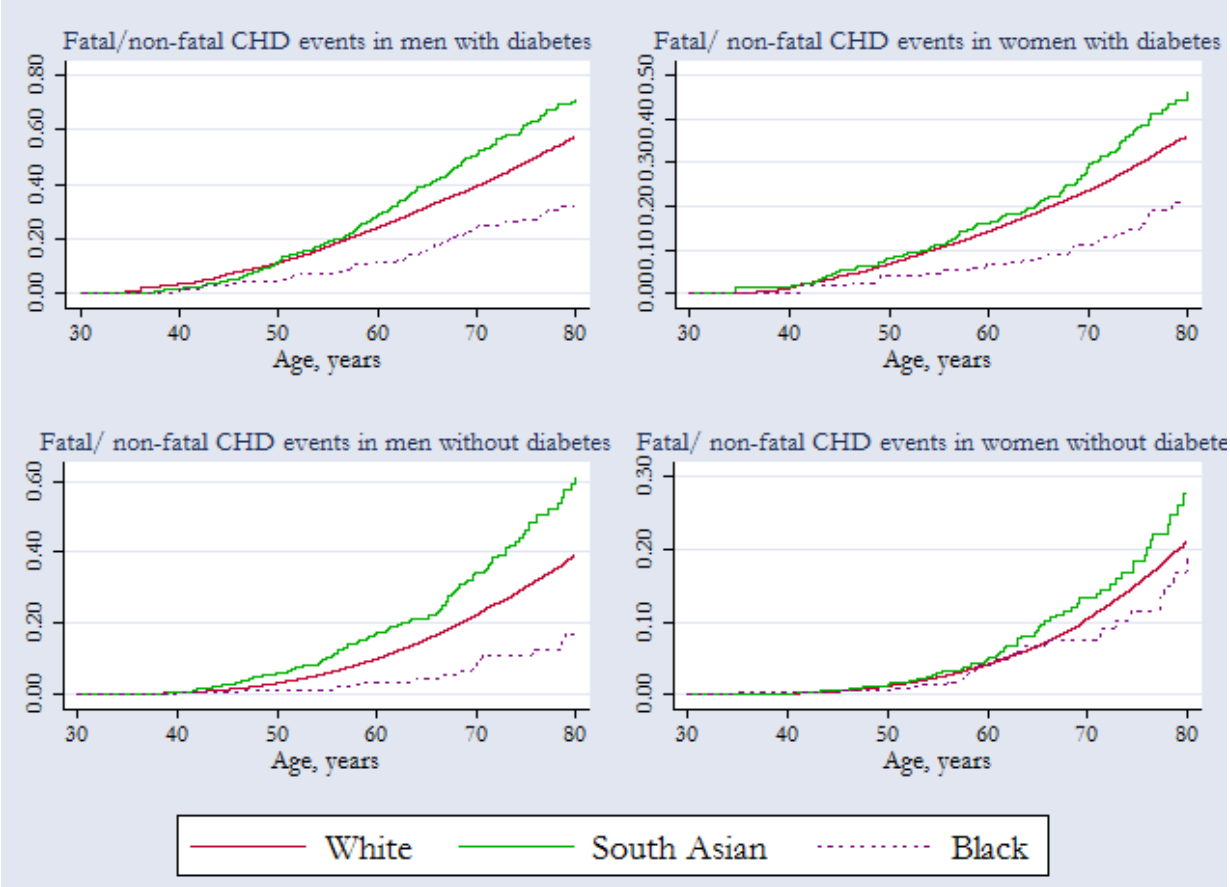


Figure 10.3 Cumulative incidence of non-fatal coronary heart disease

10.5.2 By deprivation quintile

In the diabetic cohort, crude incidence was higher in the most deprived quintile for White males and females and lower in the most deprived quintile for South Asian males and females. Crude incidence rates were higher in the most deprived quintile of all non-diabetic groups. In contrast to the main study sample, differences in rates by deprivation quintile were modest across all groups. Differences could not be calculated for Black female diabetics, who had no events in the most affluent group.

10.5.3 By baseline clinical values and medication use

Incidence was higher for patients with hypertension at baseline compared to those without hypertension in all groups except for Black African/Caribbean men, who had a slightly lower incidence rate if hypertensive. Hypertension was associated with a 2- to 3-fold increase in CHD incidence amongst non-diabetics and a more modest increase amongst diabetics. Across all populations, total serum cholesterol over 5 mmol/L was associated with a higher crude

incidence of non-fatal CHD except for Black African/Caribbean women without diabetes, for whom high cholesterol was associated with a lower incidence.

Turning to body mass index, amongst diabetics, an inverse relationship between BMI category and incidence rate was found for White males and females with diabetes. In most ethnic groups the highest incidence rates occurred in those who were underweight at baseline, excepting all Black females and South Asian female non-diabetics. No consistent relationship with smoking status or alcohol consumption was evident. Hba1c over 6.5% at baseline was associated with the highest incidence of fatal and non-fatal CHD in all diabetic males and Black African/Caribbean females. Amongst White and South Asian females, incidence was highest in those with a baseline HbA1c <5.4%.

Amongst diabetics only, prescriptions of insulin and other antidiabetic drugs were associated with increased CHD incidence, with no clear pattern for those prescribed metformin. In all groups, the prescription of antihypertensives, lipid-lowering medications and statins were associated with an increase in the CHD incidence by 3- to 11-fold, though the increases were broadly more modest in the diabetic group compared with the non-diabetic group.

Table 10.4 Crude incidence rate of fatal and non-fatal CHD according to diabetic status, gender, and ethnic group, per 10,000 person-years

	T2DM N=86,778						Non T2DM N=209,543					
	Males			Females			Males			Females		
	White	South Asian	Black	White	South Asian	Black	White	South Asian	Black	White	South Asian	Black
Number of subjects	36,522	1,529	694	34,881	1,591	806	68,981	1,737	1,287	96,799	3,135	2,546
Number of CHD events	4,507	194	40	3,144	113	29	6,179	123	20	5,188	59	26
Person years of follow-up	203,870	8,239	3,640	196,910	8,562	4,319	566,227	9,667	6,957	800,488	17,072	12,966
Overall Rate	221.1	235.5	109.9	159.7	132.0	67.1	109.1	127.2	28.7	64.8	34.6	20.1
Patient-level IMD												
1 (least deprived)	201.2	260.8	84.3	130.2	162.8	0	94.9	118.3	17.4	48.5	41.8	9.6
5 (most deprived)	252.5	247.9	80.8	192.5	149.7	61.9	129.0	123.4	24.5	88.6	25.3	20.4
Hypertension												
BP <=140/90	211.9	227.3	115.0	147.9	110.0	54.7	89.1	110.3	19.5	42.4	23.5	14.8
BP >140/90	232.0	243.4	101.2	175.1	151.0	84.1	174.3	244.6	54.3	124.7	101.6	45.3
Cholesterol												
=<5 mmol/L	196.8	217.9	94.5	134.7	114.9	57.9	143.2	154.3	36.3	74.6	43.4	39.7
>5 mmol/L	248.3	256.8	127.6	175.7	146.0	87.3	151.9	170.2	50.5	89.0	69.6	25.3
Smoking Status												
Non smoker	210.1	232.5	105.0	159.6	122.0	73.9	94.9	120.0	29.7	60.1	34.4	21.3
Smoker	219.1	212.4	140.9	158.2	149.6	0	110.0	130.0	21.7	55.0	12.1	6.2
Ex-smoker	226.0	245.4	70.2	154.1	249.5	78.0	140.0	150.0	37.7	66.3	55.9	11.1
Alcohol Consumption												
Non-drinker	259.5	262.1	101.3	189.1	130.9	93.6	153.0	159.6	56.1	95.7	41.4	13.7
Moderate alcohol drinker	214.3	215.2	103.2	137.4	56.1	55.0	107.8	100.7	13.8	50.3	13.4	27.2
Heavy drinker	185.5	184.0	0.0	143.4	566.5	0	90.4	90.7	55.7	39.0	0.0	0.0

Table 10.4 Continued...

	T2DM N=86,778						Non T2DM N=209,543					
	Males			Females			Males			Females		
	White	South Asian	Black	White	South Asian	Black	White	South Asian	Black	White	South Asian	Black
BMI Category												
Underweight	261.5	243.7	155.3	199.1	151.2	68.4	144.5	89.9	122.5	96.4	16.4	0.0
Normal	227.1	234.9	87.4	170.1	114.1	63.1	98.8	148.8	28.3	48.7	18.6	13.7
Overweight	201.0	222.5	147.4	148.0	95.7	62.1	115.3	125.1	20.6	62.3	46.1	33.0
Obese	183.3	174.2	0.0	120.2	133.3	82.9	123.3	89.3	37.8	61.4	57.2	8.2
Hba1c Category												
<5.4	209.1	171.5	79.5	168.1	425.8	0						
5.5-6.4	188.4	188.6	90.1	134.9	100.8	67.3						
>=6.5	226.9	243.9	119.6	166.5	123.3	72.6						
Medications												
Antihypertensives												
No	143.4	178.0	70.0	77.6	47.2	10.4	67.1	81.2	16.5	29.0	16.2	6.2
Yes	248.0	265.2	125.0	178.2	167.8	83.4	236.8	313.6	72.5	137.9	104.8	61.9
Statins												
No	232.7	183.8	95.2	182.2	117.8	58.6	94.2	102.1	28.2	55.9	22.5	14.7
Yes	215.3	258.7	118.7	149.7	139.5	71.9	252.0	314.3	34.5	171.9	212.1	109.0
Aspirin												
No	159.1	139.6	65.9	119.8	80.6	16.5	78.4	102.9	19.8	43.3	24.6	14.8
Yes	284.6	332.9	161.8	205.9	195.1	131.9	393.4	386.7	173.9	269.0	164.5	97.8
Insulin												
No	212.4	219.2	108.8	149.2	131.0	70.2						
Yes	290.8	385.9	115.3	232.4	138.0	52.8						
Metformin												
No	233.4	217.1	149.5	165.1	154.5	68.6						
Yes	214.0	243.0	93.4	156.6	125.9	66.7						
OAD												
No	203.3	199.6	89.3	130.4	98.1	56.2						
Yes	239.4	267.1	127.5	191.6	161.4	76.2						

10.5.4 By age group

The age-specific incidence rates stratified by gender are shown in Figure 10.3. The incidence of CHD increased with age in both diabetic and non-diabetic groups, with all rates higher in males than females. Crude rates in each age band stayed highest in the diabetic population until age 70, at which point the rates for the non-diabetic group overtook those for the diabetic group.

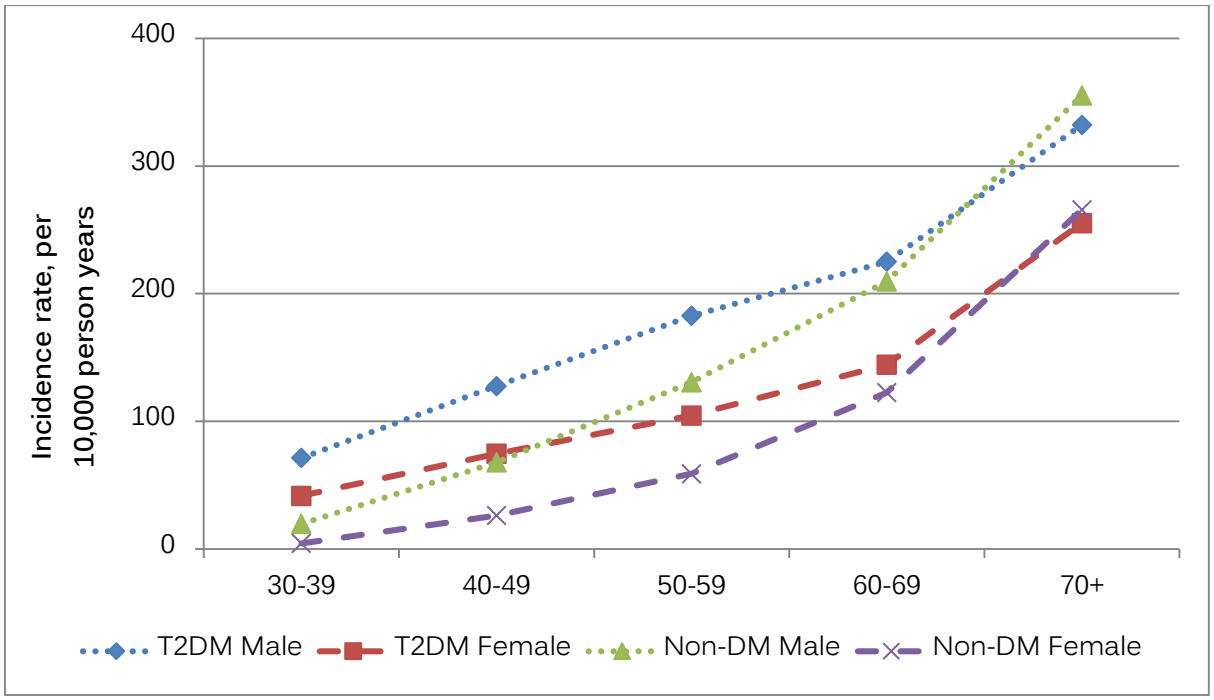


Figure 10.4 Age-specific incidence of fatal and non-fatal CHD by gender

The age-specific incidence rates stratified by ethnic group are shown in Table 10.5. In the diabetic cohort, the crude incidence rate increased with age band for White men and women. In South Asian men and women, rates peaked in those aged 60–69 years. In the non-diabetic cohort group, the crude incidence rate increased steadily with age band regardless of ethnicity or gender. Rates at all age strata were higher in the diabetic cohort than the non-diabetic cohort until age 70+, where rates for South Asian and Black African/Caribbean men and women overtook those in the diabetic cohort.

Table 10.5 Age-specific incidence rates of fatal and non-fatal CHD combined according to gender, ethnic group, and diabetic status

	Person-years of follow-up	Men									Women								
		White			South Asian			Black			White			South Asian			Black		
		Rate	CI 95%		Rate	CI 95%		Rate	CI 95%		Rate	CI 95%		Rate	CI 95%		Rate	CI 95%	
T2DM																			
CHD events		4,507			194			40			3,144			113			29		
30-39	9,944	59.7	[38.5,	92.6]	46.7	[11.7,	186.7]	0.0	.	.	22.2	[11.6,	42.7]	18.4	[2.6,	130.8]	0.0	.	.
40-49	44,238	96.7	[83.5,	112.1]	157.7	[106.6,	233.4]	17.1	[2.4,	121.7]	65.4	[54.0,	79.3]	65.3	[36.2,	117.9]	41.4	[13.4,	128.4]
50-59	93,194	167.0	[155.1,	179.8]	239.4	[185.9,	308.3]	91.4	[43.6,	191.8]	93.2	[83.2,	104.4]	104.1	[71.4,	151.8]	20.1	[5.0,	80.5]
60-69	134,704	199.6	[188.8,	211.1]	306.8	[239.7,	392.7]	180.0	[114.8,	282.2]	122.6	[113.3,	132.5]	164.9	[119.5,	227.6]	66.0	[33.0,	132.0]
70+	199,054	303.1	[291.2,	315.6]	264.2	[196.6,	355.1]	115.6	[67.1,	199.0]	224.8	[215.4,	234.6]	247.0	[179.0,	341.0]	132.5	[81.2,	216.3]
Non- T2DM																			
CHD events		6,179			123			20			5,188			59			26		
30-39	386,710	7.8	[6.3,	9.6]	8.6	[2.8,	26.8]	17.2	[6.5,	45.8]	1.5	[1.0,	2.2]	2.5	[0.6,	10.1]	3.4	[0.8,	13.5]
40-49	376,761	40.4	[37.1,	44.1]	110.0	[74.7,	160.0]	17.4	[6.5,	46.2]	11.5	[10.0,	13.2]	11.5	[4.8,	27.6]	4.9	[1.2,	19.6]
50-59	323,448	89.4	[84.3,	94.9]	170.0	[120.0,	250.0]	25.6	[8.3,	79.4]	36.2	[33.3,	39.4]	36.4	[18.9,	70.0]	39.1	[17.6,	87.1]
60-69	267,650	160.0	[150.0,	170.0]	260.0	[180.0,	380.0]	30.8	[7.7,	120.0]	80.5	[75.7,	85.6]	130.0	[79.0,	200.0]	73.6	[33.1,	160.0]
70+	319,625	280.0	[270.0,	290.0]	420.0	[300.0,	580.0]	140.0	[66.1,	290.0]	200.0	[200.0,	210.0]	280.0	[190.0,	410.0]	160.0	[86.3,	300.0]

Rates are per 10,000 person-years [95% confidence interval]

10.6 Adjusted risk of fatal and non-fatal CHD by ethnic group

Table 10.6a presents the results of the Cox proportional hazards models for ethnic differences in fatal and non-fatal CHD risk stratified by diabetic status and gender. Compared with the full study sample, CHD risk differences in the ONS-linked subset were attenuated for the comparison between South Asians vs. White and increased for the comparison between Black African/Caribbean vs. White.

In both diabetic and non-diabetic males, adjusted risk of CHD was higher in the South Asian group than in the White group, and lower in the Black African/Caribbean group. Adjusted models for diabetic females showed no difference in CHD risk between South Asian and White groups and reduced risk in the Black African/Caribbean group. Adjusted models for non-diabetic females showed that the risk of CHD was significantly reduced for both South Asian and Black African/Caribbean groups compared to White.

10.6.1 Diabetic Cohort

In the fully adjusted models for diabetic males, risk of incident fatal and non-fatal CHD was increased by 25% for South Asians compared to White (HR 1.25, CI95% 1.04, 1.50) and reduced by 51% for Black African/Caribbean compared to White (HR 0.49, CI95% 0.34, 0.72). In the fully adjusted models for women, no excess risk was evident for South Asian women (HR 1.04, CI95% 0.82, 1.33) while risk was reduced by 45% for Black African/Caribbean women (HR 0.55, CI95% 0.36, 0.83).

10.6.2 Non-diabetic Cohort

In the fully adjusted models, the ethnic differences in the non-diabetic cohort were larger than in the diabetic cohort. Amongst non-diabetic males, the risk of incident fatal and non-fatal CHD was increased by 55% for South Asians compared with the White group (HR 1.55, CI95% 1.20, 2.00) and reduced by 72% for the Black African/Caribbean group compared with the White group (HR 0.28, CI95% 0.13, 0.63). No ethnic differences were apparent in the fully adjusted models for non-diabetic women, with all 95% confidence intervals crossing unity.

10.6.3 Sensitivity Analysis

Sensitivity analyses replicating all of the models described above using only the complete cases available for the fully adjusted models are presented in table 10.6b. In diabetics and non-diabetics of both genders, the point estimates are slightly larger, and confidence intervals are comparable to those derived in the main analysis, with no difference in the interpretation of the relative risk of incident fatal and non-fatal CHD by ethnic group.

Sensitivity analyses replicating the analysis of non-fatal CHD in patients without both T2DM and T1DM are presented in table 10.6c. The patterning of differences in CHD risk by ethnic group closely matches those of the main analysis.

Table 10.6a Incidence of fatal and non-fatal CHD amongst subjects with and without type 2 diabetes

		Males								Females							
		Crude		Age		Framingham		Full		Crude		Age		Framingham		Full	
T2DM	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
	South Asian	1.09	[0.95, 1.26]	1.34	[1.16, 1.55]	1.40	[1.19, 1.65]	1.25	[1.04, 1.50]	0.86	[0.72, 1.04]	1.27	[1.05, 1.53]	1.32	[1.06, 1.63]	1.04	[0.82, 1.33]
	Black	0.51	[0.38, 0.70]	0.56	[0.41, 0.76]	0.53	[0.37, 0.77]	0.49	[0.34, 0.72]	0.44	[0.31, 0.64]	0.55	[0.38, 0.79]	0.63	[0.43, 0.93]	0.55	[0.36, 0.83]
Subjects		44,413		44,413		42,313		39,637		42,365		42,365		40,480		37,460	
Non	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
T2DM	South Asian	1.22	[1.02, 1.46]	1.89	[1.58, 2.26]	1.73	[1.37, 2.20]	1.55	[1.20, 2.00]	0.60	[0.46, 0.77]	1.49	[1.15, 1.92]	1.58	[1.14, 2.18]	1.35	[0.95, 1.92]
	Black	0.28	[0.18, 0.43]	0.47	[0.30, 0.73]	0.40	[0.21, 0.76]	0.28	[0.13, 0.63]	0.35	[0.24, 0.51]	0.98	[0.66, 1.44]	0.89	[0.53, 1.51]	0.71	[0.41, 1.24]
Subjects		87,101		87,101		43,857		38,255		122,442		122,442		58,449		51,950	

Table 10.6b Incidence of fatal and non-fatal CHD amongst subjects with and without type 2 diabetes – complete case analysis

		Males								Females							
		Crude		Age		Framingham		Full		Crude		Age		Framingham		Full	
T2DM	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
	South Asian	1.10	[0.92, 1.31]	1.34	[1.12, 1.61]	1.38	[1.16, 1.66]	1.25	[1.04, 1.50]	0.79	[0.63, 1.00]	1.18	[0.93, 1.50]	1.27	[1.00, 1.62]	1.04	[0.82, 1.33]
	Black	0.51	[0.35, 0.74]	0.54	[0.37, 0.79]	0.55	[0.38, 0.80]	0.49	[0.34, 0.72]	0.53	[0.35, 0.79]	0.65	[0.43, 0.97]	0.67	[0.45, 1.02]	0.55	[0.36, 0.83]
Subjects		39,637		39,637		39,637		39,637		37,460		37,460		37,460		37,460	
Non	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
T2DM	South Asian	1.23	[0.95, 1.58]	1.67	[1.30, 2.16]	1.74	[1.35, 2.24]	1.55	[1.20, 2.00]	0.83	[0.58, 1.16]	1.51	[1.07, 2.13]	1.63	[1.15, 2.31]	1.35	[0.95, 1.92]
	Black	0.22	[0.10, 0.48]	0.30	[0.13, 0.66]	0.31	[0.14, 0.69]	0.28	[0.13, 0.63]	0.50	[0.29, 0.86]	0.89	[0.51, 1.53]	0.94	[0.54, 1.62]	0.71	[0.41, 1.24]
Subjects		38,255		38,255		38,255		38,255		51,950		51,950		51,950		51,950	

Table 10.6c Incidence of fatal and non-fatal CHD amongst subjects without type 2 diabetes – excluding patients with T1DM

		Males								Females							
		Crude		Age		Framingham		Full		Crude		Age		Framingham		Full	
Non	White (ref)	1	--	1	--	1	--	1	[1.00,1.00]	1	--	1	--	1	--	1	--
T2DM	South Asian	1.22	[1.02,1.46]	1.90	[1.59,2.28]	1.76	[1.39,2.23]	1.59	[1.24,2.05]	0.59	[0.46,0.77]	1.49	[1.15,1.93]	1.57	[1.13,2.18]	1.37	[0.96,1.95]
	Black	0.28	[0.18,0.43]	0.47	[0.30,0.73]	0.40	[0.21,0.78]	0.34	[0.16,0.72]	0.33	[0.23,0.50]	0.95	[0.64,1.41]	0.91	[0.53,1.53]	0.84	[0.49,1.42]
Subjects		86,616		86,616		43,415		38,083		122,060		122,060		58,115		51,971	

Table shows hazard ratios and 95% confidence intervals

10.7 Adjusted risk of fatal and non-fatal CHD by ethnic minority subgroup

Table 10.7a presents the results of the Cox proportional hazards models for ethnic differences in risk of fatal and non-fatal between the main South Asian and Black subgroups compared with the White group as a whole, stratified by diabetic status and gender. In the fully adjusted analyses for males, risk was significantly increased for all South Asian subgroups, and significantly reduced for Caribbean groups in relation to the White group. No differences between African and White males were evident in the adjusted analysis. Fully adjusted models for women showed no differences by ethnicity in the non-diabetic cohort, but a significant reduction in risk for diabetic women of Caribbean ethnicity.

10.7.1 Diabetic Cohort

The ethnic differences in risk were smaller for South Asian groups and larger for Black groups in the ONS-linked subset compared with the full study sample. In the fully adjusted models for diabetic males, risk of incident fatal and non-fatal CHD increased by 84% for Pakistani males compared with White males (HR 1.84, CI95% 1.23, 2.75) and reduced by 61% for Caribbean males compared with White males (HR 0.39, CI95% 0.23, 0.67). No differences in risk of fatal and non-fatal CHD were found between Indian, Bangladeshi and African males compared with White males in either the crude or adjusted analysis except for in the Framingham model, where risk was significantly increased in the Pakistani group.

Amongst diabetic females, all risk differences were smaller than for males. In the fully adjusted models for diabetic females, risk of incident fatal and non-fatal CHD was increased by 45% for Indian females compared with White females (HR 1.45, CI95% 1.11, 0.90) and reduced by 30% in Caribbean females compared with White females (HR 0.60, CI95% 0.37, .97). No differences in risk of fatal and non-fatal CHD were found between Pakistani, Bangladeshi and African females compared with White females in either the crude or fully adjusted analysis, though an increase in risk for Indian females was observed in the age- and Framingham-adjusted models.

10.7.2 Non-diabetic Cohort

The patterning of risk differences in the non-diabetic cohort differed from that of the diabetic cohort. Compared to diabetic males, no differences in risk between the South Asian subgroups and the White group were evident in the fully adjusted analysis, though risk was significantly increased for Indian and Pakistani groups in the Framingham adjusted model (Indian HR 1.56 CI95% 1.13,2.15 Pakistani HR 2.42 CI95% 1.30,4.51). The difference between Caribbean and White males was much larger in the non-diabetic cohort compared to the diabetic cohort, with a 90% reduction in risk amongst non-diabetics (HR 0.10 CI95% 0.01, 0.71).

Though the age adjusted models for female non-diabetics suggested an increase in risk for Pakistani women (HR 1.89 CI96 1.02, 3.52), no ethnic differences for any South Asian or Black subgroups were evident in the Framingham and fully adjusted models.

10.7.3 Sensitivity Analysis

Sensitivity analyses replicating all of the models described above using only the complete cases available for the fully adjusted models are presented in table 10.7b. In all subgroups the point estimates were increased slightly for all groups. No differences in the interpretation of the relative risk of incident non-fatal CHD by ethnic sub group were evident.

Table 10.7a Incidence of fatal and non-fatal CHD amongst subjects with and without type 2 diabetes – ethnic subgroup breakdown

		Male								Female							
		Crude		Age		Framingham		Full		Crude		Age		Framingham		Full	
T2DM	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
	Indian	1.01	[0.83, 1.23]	1.2	[0.99, 1.46]	1.31	[1.06, 1.62]	1.16	[0.92, 1.47]	0.99	[0.78, 1.25]	1.36	[1.06, 1.73]	1.45	[1.11, 1.90]	1.24	[0.92, 1.67]
	Pakistani	1.74	[1.30, 2.32]	2.28	[1.71, 3.04]	2.33	[1.68, 3.24]	1.84	[1.23, 2.75]	0.81	[0.53, 1.24]	1.34	[0.88, 2.05]	1.50	[0.94, 2.40]	1.03	[0.61, 1.76]
	Bangladeshi	1.12	[0.62, 2.03]	1.38	[0.77, 2.50]	1.43	[0.71, 2.86]	1.37	[0.65, 2.89]	0.52	[0.22, 1.26]	1.01	[0.42, 2.44]	0.56	[0.14, 2.26]	0.5	[0.12, 2.00]
	African	0.64	[0.38, 1.07]	0.78	[0.46, 1.31]	0.78	[0.43, 1.41]	0.71	[0.37, 1.37]	0.33	[0.15, 0.75]	0.49	[0.22, 1.09]	0.64	[0.29, 1.44]	0.62	[0.26, 1.48]
	Caribbean	0.46	[0.30, 0.71]	0.47	[0.30, 0.71]	0.43	[0.26, 0.71]	0.39	[0.23, 0.67]	0.49	[0.32, 0.77]	0.56	[0.36, 0.87]	0.6	[0.37, 0.97]	0.57	[0.35, 0.93]
Subjects		44413		44413		42313		39637		42365		42365		40480		37460	
Non	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
T2DM	Indian	1.24	[0.97, 1.58]	1.68	[1.32, 2.14]	1.56	[1.13, 2.15]	1.4	[1.00, 1.97]	0.63	[0.45, 0.89]	1.36	[0.97, 1.90]	1.4	[0.91, 2.15]	1.22	[0.76, 1.95]
	Pakistani	1.29	[0.86, 1.95]	2.83	[1.88, 4.26]	2.42	[1.30, 4.51]	1.9	[0.98, 3.68]	0.61	[0.33, 1.13]	1.89	[1.02, 3.52]	1.99	[0.95, 4.18]	1.73	[0.82, 3.65]
	Bangladeshi	0.91	[0.38, 2.18]	1.71	[0.71, 4.11]	1.66	[0.53, 5.15]	1.16	[0.29, 4.64]	0.23	[0.03, 1.60]	0.79	[0.11, 5.61]	1.38*	[0.19, 9.77]	--*	--
	African	0.26	[0.13, 0.49]	0.5	[0.26, 0.96]	0.53	[0.22, 1.27]	0.36	[0.12, 1.12]	0.28	[0.15, 0.53]	1	[0.54, 1.86]	0.98	[0.40, 2.35]	0.69	[0.26, 1.83]
	Caribbean	0.26	[0.13, 0.56]	0.35	[0.17, 0.73]	0.19	[0.05, 0.75]	0.1	[0.01, 0.71]	0.49	[0.30, 0.82]	1.01	[0.61, 1.67]	0.85	[0.43, 1.71]	0.73	[0.36, 1.46]
Subjects		87101		87101		43857		38255		122442		122442		58449		51950	

*Confidence intervals not estimated due to small sample size and zero counts in at least one subgroup

Table 10.7b Incidence of fatal and non-fatal CHD amongst subjects with and without type 2 diabetes – ethnic subgroup breakdown, complete case analysis

		Male								Female							
		Crude		Age		Framingham		Full		Crude		Age		Framingham		Full	
T2DM	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
	Indian	1.06	[0.84, 1.34]	1.25	[0.99, 1.58]	1.06	[0.84, 1.34]	1.25	[0.99, 1.58]	0.97	[0.72, 1.29]	1.37	[1.02, 1.83]	1.47	[1.09, 1.98]	1.24	[0.92, 1.67]
	Pakistani	1.57	[1.06, 2.33]	2.07	[1.40, 3.07]	1.57	[1.06, 2.33]	2.07	[1.40, 3.07]	0.74	[0.44, 1.26]	1.25	[0.74, 2.12]	1.37	[0.81, 2.32]	1.03	[0.61, 1.76]
	Bangladeshi	1.19	[0.57, 2.49]	1.51	[0.72, 3.18]	1.19	[0.57, 2.49]	1.51	[0.72, 3.18]	0.32	[0.08, 1.26]	0.58	[0.15, 2.34]	0.65	[0.16, 2.60]	0.5	[0.12, 2.00]
	African	0.59	[0.31, 1.14]	0.73	[0.38, 1.41]	0.59	[0.31, 1.14]	0.73	[0.38, 1.41]	0.44	[0.18, 1.05]	0.64	[0.27, 1.54]	0.68	[0.28, 1.63]	0.62	[0.26, 1.48]
	Caribbean	0.46	[0.27, 0.77]	0.44	[0.26, 0.75]	0.46	[0.27, 0.77]	0.44	[0.26, 0.75]	0.62	[0.39, 1.00]	0.7	[0.43, 1.13]	0.73	[0.45, 1.17]	0.57	[0.35, 0.93]
Subjects		39637		39637		39637		39637		37460		37460		37460		37460	
Non	White (ref)	1	--	1	--	1	--	1	--	1	--	1	--	1	--	1	--
T2DM	Indian	1.17	[0.83, 1.64]	1.48	[1.06, 2.08]	1.57	[1.12, 2.20]	1.4	[1.00, 1.97]	0.81	[0.51, 1.29]	1.33	[0.84, 2.12]	1.43	[0.90, 2.28]	1.22	[0.76, 1.95]
	Pakistani	1.38	[0.72, 2.66]	2.71	[1.41, 5.23]	2.52	[1.31, 4.85]	1.9	[0.98, 3.68]	0.96	[0.46, 2.03]	2.09	[0.99, 4.39]	2.29	[1.09, 4.81]	1.73	[0.82, 3.65]
	Bangladeshi	0.91	[0.23, 3.65]	1.43	[0.36, 5.72]	1.3	[0.32, 5.19]	1.16	[0.29, 4.64]	--*	--	--*	--	--*	--	--*	--
	African	0.23	[0.08, 0.73]	0.36	[0.12, 1.13]	0.38	[0.12, 1.17]	0.36	[0.12, 1.12]	0.36	[0.13, 0.95]	0.82	[0.31, 2.19]	0.89	[0.33, 2.37]	0.69	[0.26, 1.83]
	Caribbean	0.09	[0.01, 0.67]	0.11	[0.01, 0.76]	0.11	[0.02, 0.78]	0.1	[0.01, 0.71]	0.68	[0.34, 1.36]	0.93	[0.47, 1.87]	0.97	[0.49, 1.95]	0.73	[0.36, 1.46]
Subjects		38255		38255				38255		51950		51950		51950		51950	

*Confidence intervals not estimated due to small sample size and zero counts in at least one subgroup

10.8 Comparison of results from the main and ONS-linked studies

Forest plots illustrating the crude and adjusted hazard ratios for South Asian and Black African/Caribbean groups compared with the White group for both of the CPRD and ONS-linked study populations are displayed in figure 10.4.

Overall, ethnic differences in CHD risk were comparable for the study of non-fatal CHD alone and for fatal and non-fatal CHD combined.

The risk of incident CHD was increased in the South Asian population in both study cohorts in males and females with and without type 2 diabetes. Amongst patients with diabetes, the magnitude of the effect size was larger in the full CPRD cohort than in the ONS-linked cohort. Amongst patients without diabetes, the effect size was comparable between the CPRD and ONS-linked cohorts.

For Black African/Caribbean patients with type 2 diabetes, the risk of CHD was decreased uniformly for males and females. Amongst those without type 2 diabetes, there was strong evidence of a large risk decrease for males, but no evidence of a risk difference for females. The reduction in CHD risk for males without type 2 diabetes was larger than for males with type 2 diabetes in both the main CPRD and ONS-linked cohorts.

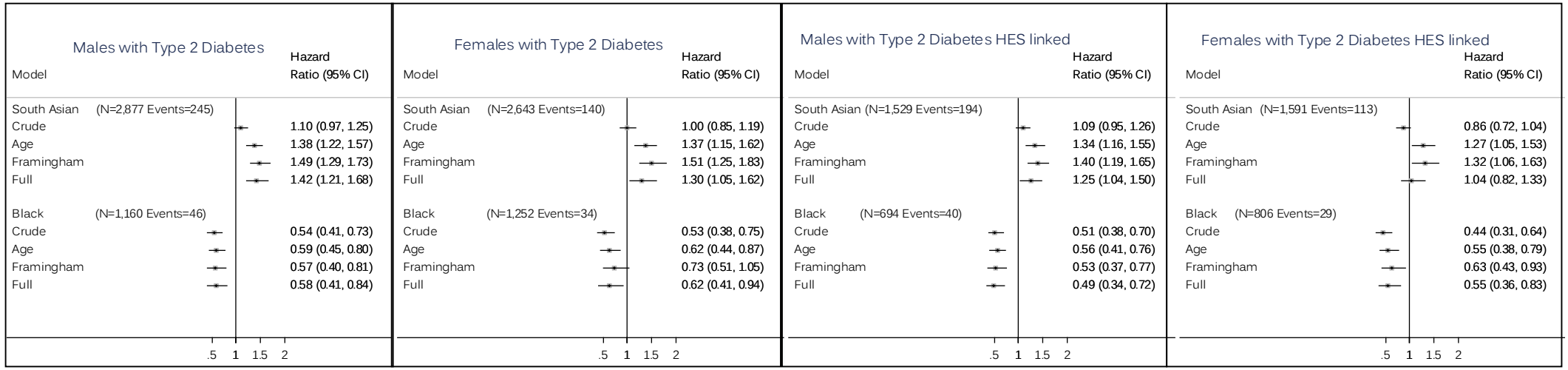


Figure 10.5 Crude and multivariable adjusted hazard ratios for all study populations stratified by gender and diabetic status

Chapter 11
Discussion of Ethnic differences in Diabetes Mellitus and Coronary Heart Disease

11.1 Summary

This study utilized a cohort of almost 900,000 patients registered with the CPRD to identify ethnic differences in risk of non-fatal coronary heart disease amongst adults aged 30 and over with and without type 2 diabetes. This is the first UK- based study to make use of routinely recorded ethnicity data to explore the relationship between ethnicity, diabetes and CHD in a large population-based database known to be representative of the UK population as a whole. Linkage between the CPRD and ONS mortality data further enabled further exploration of ethnic differences in fatal and non-fatal CHD combined in a subset of almost 300,000 patients.

Summarised below are the findings of the study in relation to the objectives described in chapter 8, to other published studies and the strengths and limitations of the study.

11.2 Main findings

11.2.1 Differences in CHD risk according to diabetic status

The first hypothesis of higher CHD risk amongst patients with diabetes compared to those without was supported by the findings of the study. Crude results showed that the presence of diabetes increased the risk of CHD by almost 40% in both the main and ONS-linked cohorts, reducing to 20% after accounting for age, gender and deprivation. The excess risk associated with diabetes was markedly higher for ethnic minority groups in the main study cohort, with an adjusted increase of 60% and 75% in South Asian and Black groups respectively, compared with 28% in the White groups. In the ONS subgroup, the excess risk of fatal and non-fatal CHD combined due to diabetes was comparable in White and South Asian groups, at around 20%, but notably higher for Black African/Caribbean patients, at 46%.

11.2.2 Ethnic differences in CHD amongst patients with type 2 diabetes

The second hypothesis of the study was borne out, with higher incidence in South Asian and lower incidence in Black African/Caribbean groups, with differences more pronounced for males than females. These findings are consistent with previous literature on the risk of CHD according to ethnic group and diabetic status.(277,306)

In both the main and ONS-linked cohorts, the risk of either non-fatal CHD alone or in combination with fatal CHD was not significantly different between South Asian and White groups in the crude model. Adjustment for age increased the strength of the association between South Asian ethnicity and CHD risk significantly. In both diabetic cohorts, the largest difference between South Asian and White groups was found after adjusting for the variables found in the Framingham risk score (blood pressure, smoking status and total cholesterol) with further adjustment for clinical covariates, medications and demographic factors attenuating the excess risk in South Asians, to the point of parity for diabetic females in the ONS subgroup.

The risk of CHD for Black African/Caribbean groups compared with the White group was consistently reduced in both the crude and adjusted models in both the main and ONS-linked cohorts by approximately 50%, with increasing levels of adjustment having little or no effect on the observed relationship.

The estimates found in both our study and the CALIBER study may be lower than those reported elsewhere, where diabetes mellitus is reported to double the risk of CHD. This may be due to our use of patients aged 30 and over, rather than a more restricted age range, such as 45 and older, as was used in the Atherosclerosis Risk in Communities (ARIC) study, in the US, which reported a doubling of incident CHD and CHD mortality amongst individuals with diabetes during the period 1987–2009.(307)

Furthermore, our study uses an unselected general practice population rather than a population of hospital or clinic attendees. Lower estimates may also be due to poorer capture of the outcome compared with intervention studies or trials, where all relevant data are completely recorded for all participants. However, the concordance between these two observational studies using the same database over a comparable time period is reassuring.

11.2.3 Ethnic differences in CHD amongst patients without type 2 diabetes

The third hypothesis that ethnic differences in CHD risk would be more pronounced in the diabetic population compared with the non-diabetic population was not supported by the results of this study. Ethnic differences in CHD risk amongst those without diabetes were consistently more pronounced than for those with diabetes in both the main and ONS-linked study populations. In the main cohort, CHD risk was increased by 50% in the diabetic group and approximately 65% in the non-diabetic group, while in the ONS-linked subgroup, risk was increased by up to 73% in non-diabetic males compared with 40% in diabetic males (Framingham models).

Since the primary mechanism by which ethnicity increases CHD is via diabetes and insulin resistance, it is possible that the independent effect of ethnicity on CHD risk was attenuated in the diabetic population. Furthermore, though risk factors such as blood pressure, cholesterol level and BMI were worse in diabetic cohort at baseline, our use of time-updated covariates may mean that individuals saw improvements in their risk factors between the time of diabetes onset and first incident CHD event, due to behavioural change or pharmacological intervention. The relationship between the coronary heart disease risk and pharmacological treatment is complex and cyclical. Medication use is both an indicator of severity of disease and also improvement in cardiovascular risk profile. The time-updated values for clinical covariates may result from increasing severity, which necessitates medication use, which then improves the status of the clinical measures, and accounts for a reduction in both absolute risk and relative differences by ethnicity.

This mechanism may be supported by our data, as Cox regression models adjusting for medication use revealed a smaller risk difference between South Asian and White groups than the models adjusting only for the variables in the Framingham risk score, suggesting that inclusion of these measures attenuates the independent effect of ethnicity on CHD risk.

Type 1 diabetes, undiagnosed diabetes, or pre-diabetic states such as insulin resistance or metabolic syndrome may have also have been present in the non-diabetic sample, thus attenuating differences between the diabetic and non-diabetic cohorts.

11.2.4 Differences in CHD incidence according to ethnic minority subgroup

The fourth hypothesis, positing that CHD risk will be raised for all South Asian subgroups, and be highest for the Pakistani group, in comparison with the White group was supported by the results of this study in the male population only. Amongst females, risk was highest in the Indian population compared with the White population in both diabetic and non-diabetic cohorts. One exception was in the ONS-linked cohort, where the risk amongst non-diabetic females was highest in the Pakistani group.

11.3 Comparisons with previous research

The findings of this study echo those of others based in the UK. Ethnic differences in cardiovascular outcomes of CHD were explored in the UK Prospective Diabetes Study (UKPDS).(308) Similar to our study, the UKPDS results demonstrated a protective effect of Black African/Caribbean ethnicity on the incidence of myocardial infarction and death compared with the White population. In contrast, no differences with respect to cardiovascular risk between the White and South Asian populations were evident.

Results from the Southall and Brent Revisited (SABRE) study examining CHD risk in a multi-ethnic London-based cohort lend further support to the findings reported here. The study authors reported an adjusted hazard ratio of 1.45 (CI95% 1.28, 1.64) for South Asian compared

with White participants and 0.74 (CI95% 0.60, 0.92) for Black African/Caribbean compared with White participants.(282)

A recent study in the CALIBER (CArdiovascular disease research using LInked Bespoke studies and Electronic health Records) database, a resource which contains CPRD patients with data linked to HES and the Myocardial Ischaemia National Audit Project (MINAP), found an increase in risk for 9 out of 12 CVD subtypes amongst patients with T2DM after adjusting for age, sex and cardiovascular risk factors.(309) Though ethnicity data was available in the dataset, it was not considered as an explanatory variable or as a covariate for adjustment. The relative risk increases attributable to T2DM observed in this study were comparable to those found in the study presented above in both magnitude and direction.

The differences between ethnic minority subgroups reported in this study are supported in the literature; previous studies have demonstrated that risk in Pakistani and Bangladeshi is higher than in Indians.(6,310–312) Furthermore, the decline in CHD mortality since 1971 in the UK has been slower for South Asian men and women than for the rest of the UK population.(313) In both study cohorts there was weak or no evidence of a difference between Black African/Caribbean subgroups and the White population with respect to CHD risk, regardless of diabetes status.

Much of the variation in health status by ethnic group can be attributed to differences in socio-economic status. For conditions such as diabetes and CHD though, as discussed in the background of this study, the findings of my study suggest an independent effect of ethnicity above and beyond that of social deprivation, with significant differences in CHD risk apparent after adjusting for social deprivation, both at the practice level in the main study and at patient level in the ONS-linked study.

Finally, increased risk in South Asians may be related to factors around access to and the use of health services, which may be worse for older patients or new migrants, who may find it

difficult to communicate in English. Though provision of translated materials and interpreting services is a central aspect of equitable care in the NHS, the social distance created in interpreted consultations has been shown to hinder the sharing of pertinent information, and to sometimes create misunderstandings between the patient and practitioner.(77,314–316)

11.4 Strengths

The study made use of a population-based sample using patients drawn from primary care practices across the UK. In order to be eligible for inclusion into the study, all patients had to have been registered for a minimum of 12 months at their practice, allowing sufficient time for disease outcomes and risk factors to be recorded by the GP.

The study included non-fatal CHD, thus allowing for the inclusion of patients who do not reach hospital. Including both fatal and non-fatal CHD allowed the effect of ethnicity in the general population to be established, which is difficult to do in smaller or regional studies, which are less generalisable to the UK as a whole. ONS mortality data are near complete, due to mandatory death registration in the UK.

11.5 Limitations

11.5.1 Missing data

Missingness of covariate data differed according to study population and diabetic status. The completeness of ethnicity and covariate recording was higher in the diabetic population than the non-diabetic population. This contributed to problems with small sample size in adjusted models in the non-diabetic and ethnic minority populations due to lack of covariate availability. Greater amounts of missing data in the non-diabetic cohort compared with the diabetic cohort meant that individuals included in the adjusted analysis may not have been representative of the wider CPRD or UK population. Sensitivity analysis explored this and found the results between the full and complete-case cohorts to be comparable. Greater levels of missing data in the population without diagnosed diabetes mellitus is to be expected as these individuals are likely to be considered low risk.

Missingness of clinical measures such as BMI, Hba1c, cholesterol level, smoking status and alcohol consumption is a limitation of routine electronic patient data not collected for the express purpose of research. This limited the power to determine an independent association between these covariates and the outcome. Furthermore, this may have introduced selection bias in the multivariate analysis. Covariate completeness is likely to be higher for patients who are perceived to be at higher risk by their general practitioners, for those have pre-existing conditions and for those who consult more often. This may include differential measurement of risk factors by ethnicity due to known differences in risk.

Because the mechanisms of missingness cannot be determined and the assumption of that it is random phenomenon may not be valid, multiple imputation was not pursued. A further exploration of the missingness mechanism with more data may aid our interpretation of these results and provide better evidence for the ways in which ethnicity, diabetes and CHD are causally related.

The fact that the fully adjusted models for CHD risk showed a smaller effect size than the models adjusted for the Framingham models is likely due to the fact that the additional covariates tended to reduce the number of complete cases available for analysis.

11.5.2 Timing and ascertainment of clinical events

The first diagnosis in the primary care record may not necessarily correlate to onset of disease. Patients may not consult with their primary care physician in the first instance, and may instead present in hospital. Some patients may not present in primary care at all, leading to an underestimation of the outcome – though this is unlikely to account for the observed differences by ethnicity.

Secondly, the non-diabetic population may be artificially healthy, as, in order to be included in the analysis, they must have been disease free from 1990 onwards. Our non-diabetic population does not include individuals who were disease-free in that period who went on to

develop type 2 diabetes during the follow-up period. CHD may be under-captured in primary care, as acute myocardial infarction tends to present initially in hospital.

In addition, the amount of follow-up time available in the CPRD may have been insufficient to adequately typify the relationship between diabetes and incident CHD. The median follow-up time for all patients contributing to the study was five years. Since the South Asian population was, on average, 10 years younger at T2DM onset, it is possible that they may not yet have experienced the CHD outcome of interest. This group may have subclinical CHD which had not been detected, or at least recorded, by the general practitioner. A study with longer follow-up time may yield a better measure of ethnic differences in this population.

Time updated measures of BMI, BP, cholesterol and HbA1c have cyclical relationship with medications, showing improvements in response to treatment. As a result, the risk of the outcome may not have been clearly related to these confounders, which may have been better controlled in the time leading up to the CHD event.

The data available may have led to an underestimate of the time lag between T2DM and CHD. Type 2 diabetes has been found to be identified in primary care on average 7 years after its true onset; therefore, the effect of diabetes on CHD may have been underestimated because we have not allowed sufficient time with the disease to pass.

Next steps would be to make more use of the linked HES data. For example, additional incident non-fatal events could be identified from the database. Furthermore, ethnic differences in fatal CHD alone could be typified to better understand the contribution of death to the estimates of ethnic differences in incidence.

11.5.3 Prescriptions

Cardiovascular prescriptions based on those issued in primary care. Issues of prescriptions do not tell us whether the medication was actually used as indicated, this drug exposure may

have been misclassified. This study is also unable to account for over-the-counter prescriptions of aspirin, leading to a potential underestimate of the exposure to this particular drug. We took only the first ever prescription to denote an ever/never exposure. This is a simple and pragmatic approach as we were not hypothesising about the relationship between drug exposure and the outcome, though we did want to account for the fact that exposures will influence the pathway between ethnicity and outcome.

11.5.4 Co-morbidities

The analysis did not adjust for co-morbidities such as chronic kidney disease or cerebrovascular disease (principally stroke or transient ischaemic attacks), which contribute to the 2-fold increase in CVD risk for diabetics observed globally.(275,306,317) There is a complex relationship between all cardio-metabolic diseases and hypothesising about where they lie on the causal pathway can be problematic. Further research into this topic would benefit from including co-morbidities as confounders in the analysis model.

Using body mass index instead of central adiposity or waist circumference, which are not routinely available in electronic health records, may not have adequately captured the excess risk conferred by central adiposity known to manifest more frequently in the South Asian population.

Chapter 12
Discussion

12.1 Introduction

This chapter integrates the key discussion points of the three main studies comprising this thesis. Following a summary of the research project as a whole and a review of the current used of routinely recorded ethnicity data for UK-wide epidemiological research, the completeness and usability of ethnicity data in routine electronic health databases are considered. Following a presentation of the key findings of each of the studies, the chapter discusses the strengths and limitations of the work overall, and the implications for future policy and research.

Three distinct but related studies were undertaken over the course of the research presented in this thesis to answer the question of whether routine electronic health records can be used to conduct epidemiological research that would substantively increase our knowledge of the relationship between ethnicity and health. Ethnic differences in diabetes and coronary heart disease were chosen as example conditions because of their increasing prevalence in the UK and their strong relationship with modifiable risk factors, which can be targeted for primary and secondary prevention.

Prior to the commencement of this research project, no reliable data were available on:

1. the quality and completeness of ethnicity data in UK-based primary and secondary care electronic health databases;
2. contemporary measures of ethnic differences in the prevalence and incidence of type 1 and type 2 diabetes in the CPRD;
3. ethnic differences in the initial presentation of coronary heart disease for patients with and without type 2 diabetes in the CPRD.

12.2 Summary of research undertaken

I. A systematic literature review to identify the extent to which patient-level ethnicity routinely available in electronic health databases has been used for UK-wide epidemiological research.

II. A cross-sectional investigation of the completeness and quality of ethnicity data available in the CPRD and HES and the comparability of ethnic population estimates in the CPRD to those of the UK population according to the 2011 Census.

III. A retrospective study using prospectively collected data of ethnic differences in the incidence and prevalence of type 1 and type 2 diabetes mellitus using definitions of diabetes derived using algorithms designed to improve the correct classification and identification of diabetes in electronic health records.

IV. A prospective study of ethnic and gender differences in the incidence of fatal and non-fatal CHD amongst CPRD patients with and without type 2 diabetes mellitus.

12.3 Review of current uses of routinely recorded ethnicity data for UK wide epidemiological research

12.3.1 Key findings

- From a total of 1,232 observational studies using UK-wide electronic health databases identified, 15 (1.3%) utilised the patient-level ethnicity data available. Ten studies utilised HES, three utilised the QRESEARCH database and two utilised the THIN database. None of the included studies drew on the CPRD.
- Completeness of ethnicity data in the studies was relatively high and allowed important differences by ethnic group to be distinguished.
- The fact that all 15 studies were published after the financial incentivisation of ethnicity recording in primary care in 2004 may reflect increased confidence in data quality and completeness brought about by this top-down initiative.

12.3.2 Findings in the context of previous research

- The literature review established that routinely collected ethnicity data in UK based healthcare databases are under-utilised in observational epidemiological studies. To date, these data have not been widely used to investigate possible inequalities between ethnic groups in healthcare usage and disease outcomes across the whole UK population. Principal reasons for this may include perceptions of lack of completeness and poor quality of these data.(318)

12.4 Completeness and usability of ethnicity data in routine electronic health databases

12.4.1 Key findings

- The completeness and consistency of routinely recorded ethnicity data in UK primary and secondary care has improved considerably over time and, with certain caveats, can be usefully incorporated into health research.
- In primary care, recording of valid ethnicity for new patients registering with general practices across the UK improved dramatically following incentivisation under the Quality and Outcomes Framework. At the time of the study in 2012, valid ethnicity was being recorded for 86% of newly registered patients in primary care, 77% of HES inpatients, and 50% of both HES A&E patients and outpatients.
- Over 80% of patients with multiple ethnicities in both CPRD and HES had ethnicity which was truly identical, or fell into the same five high-level categories. One exception was for the HES outpatient population, where only 10% of patients with multiple ethnicities had codes which were truly identical or categorically matched, limiting the usefulness of the ethnicity data available in this dataset.
- Linking the CPRD and HES inpatient data improved completeness of ethnicity recording to 88.9% overall and to 97.1% for those registered from 2006 onwards, with high levels of agreement between linked sources for patients of White and South Asian ethnicity.

- The ethnic breakdown of the CPRD, which has already been shown to be representative of the UK population in terms of age and gender, was found to be comparable to that of the combined censuses for England, Wales, Scotland and Northern Ireland.

12.4.2 Findings in the context of previous research

- Previous methods of assigning patient ethnicity such as using name-recognition software or extrapolating from area-based measures of ethnic prevalence are of questionable validity, particularly for individuals of mixed ethnicity and for descendants of migrants.
- The overall completeness of ethnicity recording in the whole of the CPRD and HES was found to be comparable to that reported in other recent studies using these resources.(122,125)
- The findings around agreement between linked datasets mirror those of a recent NHS-based study which reported high accuracy of coding for patients of White British ethnicity, but far weaker agreement for all other ethnic groups amongst patients on a cancer registry.(319)
- Since the publication of my work in 2013, a further study comparing the CPRD population to the 2011 census has been conducted. As with the majority of research conducted using the CPRD to date, this contemporaneous project did not examine ethnicity, but did further confirm that the CPRD database is representative of the UK population with respect to age, gender and region.(320)
- My study has not explored ethnicity beyond what has been recorded in routine databases. Due to the observational nature of this study, we are not in a position to explore the reasons as to why ethnicity is not recorded, or recorded inconsistently, or indeed incorrectly, over time and between data sources.
- As service-level factors are not recorded in routine health databases, this study was unable to establish whether the likelihood of having complete and consistent ethnicity data was related to service provider factors such as staff availability, workload and

time pressures at the surgery/ward, or time available in the consultation in which to ask about ethnicity.

12.5 Ethnic differences in the incidence and prevalence of diabetes mellitus in the CPRD

12.5.1 Key findings

- Based on populations identified using the algorithms, there were a total of 334,584 prevalent cases of diabetes, giving an overall prevalence of 0.26% for type 1 diabetes and 2.87% for type 2 diabetes over the study period of January 1990 to August 2013. While the prevalence of type 1 diabetes increased minimally over time, the prevalence of type 2 diabetes rose 6-fold over the study period, increasing from 0.6% in 1990 to 3.7% in 2012.
- The prevalence of type 1 diabetes in the White population was over double that of the non-White ethnic groups, with incidence significantly raised rates in the White group compared with the Black and South Asian groups. Age standardisation did not alter the crude prevalence in any notable way.
- The prevalence of type 2 diabetes was highest for South Asian groups (4.7%) followed by White (3.9%) and Black African/Caribbean groups (3.2%). The age-standardised prevalence was doubled in the South Asian and Black African/Caribbean populations, to 10.7% and 7.4%, respectively.
- The incidence of type 2 diabetes was raised 2-fold in the Black African/Caribbean group and 3-fold in the South Asian group relative to the White population.
- The prevalence and incidence of both type 1 and type 2 diabetes was higher in males than in females. This pattern was displayed consistently across all calendar years, ethnic groups and age groups.
- T2DM onset occurred a decade earlier in the South Asian and Black African/Caribbean groups compared with the White group. Mean age at onset declined over the study

period, with the largest decline observed in the Black African/Caribbean population and smallest decline in the South Asian population.

- The study revealed significant heterogeneity in diabetes risk according to ethnic minority subgroup: type 2 diabetes risk was increased by 2.6-fold in Indians, 4-fold in Pakistanis and 6-fold in Bangladeshis relative to the White population. Risk in both African and Caribbean subgroups was doubled relative to the White population.
- These findings support use of ethnicity over race as an epidemiological variable, as it is able to identify differences between populations which may arise from cultural and socio-demographic differences rather than genetic differences alone.

12.5.2 Findings in the context of previous research

- This study represents the first implementation of a series of algorithms in a large database representative of the UK population as a whole, and also the first assessment of ethnic differences in diabetes burden for the entire cohort of patients contributing to the CPRD.
- This study improved upon the original implementation of the algorithms in the Welsh SAIL database by utilising the patient-level ethnicity data available in the CPRD to improve the identification of patients with type 2 diabetes.(321)
- The study confirmed the hypothesis that onset of T2DM occurs at younger ages in South Asian populations compared with the White population. The 10-year difference in median age at onset was comparable to that found in non-database studies of diabetes in the UK.(12,195,322)
- The spike in incidence of both type 1 and type 2 diabetes in the year 2006 may be related to the Quality and Outcomes Framework, which changed from rewarding the coding of diabetes as a whole to rewarding the coding of specific diabetes type.(323) The fact that incidence rates returned to 2005 levels the year after suggests that this finding is artefactual, reflecting a shift in case ascertainment rather than a true jump in incidence.

- Globally, the UK has the fifth highest incidence of T1DM amongst children aged under 14 years. A report by the International Diabetes Federation found that in 2013, the incidence of T1DM in this age group was 2.5 per 10,000 person-years.(229) This is comparable with the findings of this study, where the incidence of T1DM was 2.06 per 10,000 person-years in the group aged 0–9 years and 2.96 per 10,000 person-years in the group aged 10–19 (averaging 2.51 across the two age groups).
- Diabetes is primarily managed in primary care, so the CPRD should capture all cases. The fact that estimates of prevalence and incidence are comparable to those found in other UK-based studies indicates that the CPRD is a valuable resource for conducting research in this area and can be used more widely for health policy and establishing the evidence base.

12.6 Ethnic and gender differences in the risk of incident CHD amongst patients with and without type 2 diabetes mellitus

12.6.1 Key findings

- The presence of diabetes increased the risk of CHD by 40% in both the main and ONS-linked cohorts, although this was reduced to 20% after accounting for age, gender and deprivation. The excess risk associated with diabetes was markedly higher for ethnic minority groups compared with the White population in the main study cohort.
- In the ONS subgroup, the excess risk of fatal and non-fatal CHD combined due to diabetes was significantly higher in the Black African/Caribbean group compared with the White and South Asian groups.
- Though tests for effect modification by gender, age and deprivation were found to be statistically significant, no significant interaction between diabetic status and ethnic group was found, suggesting that the effect of ethnicity on CHD risk does not differ for those with and without diabetes.
- The risk of incident CHD was significantly raised in the South Asian population relative to the White population for both males and females. The magnitude of the risk difference

between ethnic groups was larger for those without type 2 diabetes than for those with diabetes.

- The protective effect of Black African/Caribbean ethnicity on CHD risk was stronger for males than for females relative to the White population.

12.6.2 Findings in the context of previous research

- This is the first UK-based study to make use of routinely recorded ethnicity data to explore the relationship between ethnicity, diabetes and CHD in a large population-based database known to be representative of the UK population as a whole.
- The findings of this study echo those of others based in the UK. Findings from both the UKPDS and the SABRE studies have demonstrated a protective effect of Black African/Caribbean ethnicity on the incidence of myocardial infarction and death compared with the White population.
- The findings suggest that the CPRD is well suited for observational research into ethnic inequalities and that important ethnic differences can be well characterised using this resource.

12.7 Strengths and limitations of electronic health data

A discussion of the methodological strengths and weaknesses of each of the three main studies has been reported in chapters 5, 7 and 11. This section focuses on the strengths and limitations of the databases used to address the study questions, which influence all aspects of the research reported in this thesis.

12.7.1 Strengths

a) Sample size

The size of the CPRD patient population resulted in a large number of incident and prevalent events of diabetes and coronary heart disease being identified during the study period. This allowed for sufficient power to detect relationships between populations stratified by both gender and ethnic group, which is often unfeasible in smaller studies where population sizes do not allow for such granular comparisons. Often other studies resort to collapsing together

disparate ethnic groups to make general comparisons, for example between "White" and "Non-White" populations.(121) This study identified over 340,000 patients with diabetes mellitus in the whole of the CPRD and over 36,000 patients with incident CHD in a subsample of the database. National databases such as the CPRD provide enormous gains in terms of efficiency and cost over studies which involve more costly and time-consuming methods of participant recruitment and data collection.

b) Prospectively collected data

The data in the CPRD are prospectively collected and, as a result, the data are not subject to recall bias (the presence of a disease outcome affects the reporting of exposure status) or observer bias (the knowledge of the patient's disease status influences ascertainment or recording of exposure). Furthermore, the CPRD is updated regularly, allowing for the interrogation of the complete database at any time to investigate a wide range of epidemiological research questions using timely and relevant data. The CPRD holds all coded data entered onto the patient's medical record and thus can provide a comprehensive overview of the patient's care history throughout their period of registration with the general practitioner.

c) Representativeness of the UK population

Though general practices opt in to contributing data to the CPRD, the patient population has been found to be broadly representative of the UK population as a whole with respect to age, gender and region – and, now, ethnicity.(74,143,320) In 2015, the CPRD database will expand to include general practices using clinical systems other than Vision, further improving the representation of the UK population in this resource.

d) Linkage to the hospital episode statistics and ONS data

Linkage of datasets allowed an analysis of concordance of ethnicity between primary and secondary care sources, and provided a measure of assurance regarding the reliability of ethnic coding and completeness from using combined sources. Linkage to the HES and ONS

data further allowed the identification of deaths related to coronary heart disease and information on patient-level deprivation for the final study.

12.7.2 Potential limitations

a) Capture of clinical measures

The primary purpose of the clinical data held in the CPRD is for patient care, rather than research. By its nature it only includes information gathered at consultation and is thus routinely collected rather than researcher-led. As a result, the completeness and accuracy of data are subject to temporal changes in coding practices, health priorities and population need.

Anything not reported to the general practitioner is necessarily not recorded. The absence of a code does not necessarily mean that an individual is free from that condition, but must also be interpreted as being unknown. Similarly, covariate data are missing if unmeasured. This is a greater problem for studies focusing on these measures as a main exposure or outcome. For the purposes of this study, clinical measures such as blood pressure, BMI and cholesterol were considered in order to reduce confounding on the pathway between ethnicity and the outcome.

In addition to incomplete data, a further potential problem with routinely electronic health records is incorrect coding stemming from errors in the way data is entered. A wide range of studies have found the validity of diagnoses and process of care measures in CPRD to be high.(140,142,324,325) Combined with the fact that the CPRD data are subject to ongoing internal quality checks and that concerns with data quality are fed back to the general practices, researchers can be reassured that errors which do occur in the database are kept to a minimum.

b) Capture of prescriptions

All prescriptions issued in primary care are automatically added to the electronic patient record. Though electronic health records contain a complete history of prescriptions issued by the GP there is no way to determine whether the drugs prescribed are taken appropriately. Furthermore, there is no way to capture over-the-counter prescriptions for medications such as aspirin or non-steroidal anti-inflammatory drugs (NSAIDs), or prescriptions issued in secondary care. This can lead to ascertainment bias due to the potential for differential capture of prescriptions depending on the propensity to have them issued in primary care.

Furthermore, though some medications are available over the counter, patients eligible for free prescriptions such as those under 16 or over 60, with certain medical conditions including diabetes, or on a low income, may be more likely to obtain their necessary medications via prescription rather than over the counter, further complicating the issue of ascertainment bias. Though medications issued in secondary care can be examined for patients with linkage to HES, there is no way to link medications purchased over the counter to the primary care record.

Although the effect of medications was not a primary question under study for the thesis, the potential for residual confounding due to the incomplete ascertainment of medications, particularly aspirin, may have impacted the observed relationship between ethnicity, diabetes and coronary heart disease. Determining the causal relationships between medication use, disease process measures and outcomes is complex; as it was not the main purpose of this study, a simple measure of ever/never exposure to medications was used.

c) Capture of biomarker and biological data

Neither of the electronic health databases used in the research captured information on biomarkers or biological measures salient to our understanding of the underlying mechanisms relating ethnicity to diabetes and heart disease. It is known that ethnic differences in the pathogenesis of these conditions relate to differences in insulin resistance, endothelial function, adiposity and the atherosclerotic process. Despite the data available in the CPRD

having limited scope for investigating the pathogenesis of cardio-metabolic disease, the data do allow for the characterisation of observed ethnic differences in disease risk and outcomes, which is vital for developing health policies and care guidelines which are appropriate to the needs of the UK population. Planned linkages between the CPRD and resources such as the UK Biobank and laboratory results data may make studies into these areas feasible in the future.

d) Capture of socio-demographic data

Researchers only have access to the health and demographic data available in GP records and linked datasets. Other factors salient to ethnic differences such as education, income, household size, home ownership, living conditions, migration history are all relevant but unavailable in routine electronic health records. In my study, I was able to use the index of multiple deprivation in my analysis, and found that, after adjustment for socio-economic status, significant ethnic differences remained apparent. Data on country of birth, language and religion are available in the CPRD – though not for all individuals; making more use of these data in the future may prove valuable in better typifying population groups, and first- and subsequent-generation migrants.

12.8 Ethnicity in research

When considering how ethnicity can be best used in epidemiological and social research, two main questions arise. Firstly, how well does the 'ethnicity' variable reflect the sociological concept? Secondly, how useful is this variable in identifying populations between whom important differences in health exist? This thesis has focused on the latter by exploring how routinely recorded ethnicity as captured in primary and secondary care can be maximised for research into healthcare and outcomes for the entire UK population.

Ethnic monitoring was introduced into the NHS for the purposes of describing variations between groups and uncovering potential patterns of disadvantage and racism which may otherwise remain insidious.(94) The categories are intended to delineate populations that share some combination of traits such as language, culture, migration history, biological background, religion and socio-political influences. These factors shape the use of healthcare services and impact upon the natural history, prevalence, and management of disease.(326) The resulting single variable available for researchers is intended to be a distillation of each of these factors, which have been operationalised into a pragmatic and usable version of the concept of ethnicity. Currently, the UK is the only European country which collects ethnicity data in the census. This express intention differentiates the UK from countries such as France, where ethnicity data are considered sensitive and monitoring is prohibited on the grounds of discrimination.(46)

A key challenge when using ethnicity in research is the tension between ethnicity as a sociological concept and ethnicity as an epidemiological variable. A principle upheld across the UK National Health Service is that ethnicity should be self-reported. A question for debate is whether the variable of ethnicity has construct validity, that is, how well the variable captures the multidimensional concept it is intended to represent.(136,327) Measures of ethnicity can be conceptualised as falling along a continuum from the least stable (self-reported) to most stable (country of birth/ethnic origin), with ethnicity defined using broad categories falling in between. Self-reported ethnicity is arguably the best measure of the sociological concept of

ethnicity, in that it captures the individual's own identity, encompassing both ancestry and social group membership.

By its nature, this allows individuals to change their ethnic identification over time should they choose to. This poses a challenge when utilising the variable of ethnicity in epidemiological research as it is impossible to ascertain from a coded anonymised database whether changes in ethnicity coding over time are errors, or actual changes, and how these changes relate to health profile. The utility of repeat coding of ethnicity over multiple consultations has been debated; a UK-based qualitative study of South Asian patients reported that patients felt that being asked about their ethnicity repeatedly was not necessary. Instead, they felt that wider questions about religion and diet would be more useful, as these dimensions are also salient to understanding lifestyle influences on health risk.(112) For the purposes of this thesis I have considered discrepant ethnicities to be erroneous, and have strived to develop a classification system which assigns each individual a single ethnic category which is fixed over time.

The observational data available to researchers using the CPRD and HES contains no information on how ethnicity is captured; while we can assume that the information was self-reported by the individual via questionnaire or a face-to-face consultation, there is no way to verify this. It is possible that in some conditions, ethnicity information may be recorded without directly consulting the patient. Such scenarios include occasions where the clinician has previously asked the individual and is updating the record, or where the individual is unable to communicate the information directly.

A second tension in the use of ethnicity for research is that between using ethnicity to describe disease patterns and using ethnicity to attribute disease causation. There is concern that though ethnicity is a useful lens through which to identify and describe differences in health and healthcare, it may not be as well suited to identifying causal mechanisms through which disease outcomes occur, particularly when relying on standard categories such as those of the census.(328) The intended use and meaning of ethnicity must be explicitly stated by the

authors, as it will depend on the context of the research.(329) Use of pragmatic categorisation which recognises that heterogeneity within groups may exist and that the category may not apply equally to all members is necessary in order for research into inequalities to progress. Furthermore, it is important for researchers to recognise that meaningful categorisation must be responsive to the social and political context, with relevance changing over time and between regions.(330)

Finally, previous studies have ascribed patient ethnicity indirectly by using name-recognition software or by using country of birth as a proxy.(102,287,331–337) Each of these methods is of questionable validity, particularly for individuals of mixed ethnicity and for descendants of migrants, and is likely to become less meaningful over time. Though these methods have been useful for certain situations in the past, they are increasingly less useful now, especially in countries such as the UK, where large proportions of current ethnic minority groups are UK born.

Taken in isolation, the variable of ethnicity does not necessarily capture all of the information necessary to understand variations between population groups. In subsequent research, making use of ethnicity data combined with information on religion, language, country of birth, diet and deprivation will be necessary to gain as complete a picture as possible. Data relating to social class, education, housing and employment are not available in routine health databases. Access to these data could be made possible via linkage between the CPRD and the Census for England and Wales, as in Scotland, where linkage is routinely performed for the purposes of health research.(338)

12.9 Implications for health policy and clinical practice

Increasing levels of ethnicity recording in routine databases will prove vital for estimating the healthcare needs of the UK population as it ages. The incentivisation for the recording of ethnicity for all newly registered patients was discontinued in 2011 as it was expected that practices would continue to record this information routinely in order to best assess the needs of their patient population.(339) However, it will be important to examine whether rates of ethnicity recording in primary care continue to stay high or decline in the absence of financial reward.

Ethnicity is an important factor in planning and costing future health need in the UK. As the population makeup changes and life expectancy increases, rates of chronic disease will increase significantly in ethnic minority populations, and at a faster rate than in the white population. A recent study simulating the future burden of chronic disease across England has demonstrated that the greatest health need is projected to be in areas with large South Asian populations such as Leicester and east London, where the prevalence of both diabetes and heart disease are estimated to be the highest.(340)

Though ethnicity has now been incorporated into care guidelines, such as those for hypertension and diabetes, there is still much scope to extend this to a wider range of conditions known to be ethnically patterned.(299,341) NICE acknowledges in its existing guidance that further research into ethnic differences is necessary in order for it to better tailor its recommendations to the UK population.(342,343) Research using national databases such as the CPRD, THIN and QRESEARCH will be able to provide much of the evidence base for treatment and care guidelines.

Ethnicity is already routinely incorporated into primary care risk scores such as those developed by QRESEARCH for cardiovascular disease, diabetes and kidney disease. (114–116) These data will provide a cost-effective resource for investigating ethnic differences in comparison with interventional research. One example is the use of these databases for

pragmatic clinical trials, which offer cost and efficiency benefits relating to the recruitment of patients into trials and collection of outcomes data in comparison with traditional randomised controlled trials. There are also scientific benefits in that it is possible to conduct comparisons with patients not recruited as their data are still available in the database.(344)

12.10 Future Research

12.10.1 Increasing the use of ethnicity in research using electronic health records

Continuing to streamline data entry into medical records will ensure that diseases can be identified using standard code sets. At the practice, where data are first entered, it is important to ensure that input of codes and corresponding dates is as accurate as possible. Standardised data entry templates have been shown to greatly improve the accuracy of disease coding as well as the completeness of relevant clinical information such as routine measures and test results.(345) Increasing the use of such templates across a range of clinical domains may serve to improve data quality throughout the medical record, and subsequently improve their suitability for research purposes.

At the researcher end, where data are extracted, standardising the definitions of disease using uniform code lists and algorithms, as demonstrated in this thesis, will allow for research to be replicable, comparable and validated. For example, the CALIBER programme based at University College London has set out to establish a common data model with reproducible variables and analytic protocols available to anyone interested using UK-based primary and secondary care electronic health records.(346)

12.10.2 Validating ethnicity coded on patient records

Research validating the coding of ethnicity for patients in the CPRD will give insight into the extent to which ethnicity is accurately coded on medical records and may also allow an opportunity to investigate reasons why ethnicity may be unrecorded. These data are not missing at random and the absence of data can be informative as this group may differ in important ways such as health-seeking behaviour and beliefs, access to healthcare and socio-economic status. Furthermore, such qualitative research will also provide valuable insight into

what individuals perceive ethnicity to mean, and how closely the standard categories align with individual self-identification. This will prove particularly important for individuals with mixed ethnic backgrounds and for UK-born descendants of migrants

12.10.3 Identifying health need in the future

The advantage of routine electronic health databases is that they are regularly updated and can be used to provide timely information on the demographic makeup of the general population and on areas of growing healthcare need. Examining ethnic differences regularly will be of great importance as the makeup of the UK population changes. In the future the ethnically 'mixed' population will become much larger and their health experiences will differ both from those of previous generations and from those of their contemporaries from different ethnic groups. Not only will healthcare needs evolve over time, but our use of ethnicity as a concept in health research will need to evolve in order to continue to be a meaningful marker of population groups whose health needs and experiences may differ.

Continued and critical appraisal of the concept of ethnicity is necessary. The standard categories of the Census will continually need revision in order to best capture the ethnic diversity of the UK population. The inclusion of the group "Gypsy or Irish Traveller" on the 2011 census marks a recognition by the UK government that the concept of ethnicity can be used to identify population groups with differing needs and health experiences based defined by socio-economic and cultural factors rather than solely genetic factors. In the future, the concept of ethnicity as a social stratification variable may become less salient for identifying groups with different health need and outcomes; factors such as educational attainment, employment status and deprivation may better explain disparities in health outcomes in the UK population as the currently observed ethnic differences potentially diminish over subsequent generations.

12.10.4 Ethnicity and cardio-metabolic disease

As with the ethnicity, the validation of adjudicated diabetes status will be an essential next step in determining the merit of the algorithms in improving the identification and classification of diabetes in electronic health records. An extension to this would be to use linked data from hospitals or disease registries to provide further information to support or refute diabetes classification. One such resource is the UK Biobank, which may be better positioned to characterise the ethnic differences in cardio-metabolic risk in the UK population. In addition to a planned future linkage to the CPRD, Biobank already includes data gathered by interview pertaining to waist circumference, fat mass, diet, education and employment status.

As discussed in chapter 7, correctly identifying and reducing the burden of undiagnosed diabetes in the UK is a key public health priority. Routine electronic health records will be of increasing value in generating timely estimates of the prevalence of undiagnosed diabetes across the UK as a whole. In this thesis, algorithms identifying patients with diagnostic codes for diabetes were used. This work is being taken forward by the Biobank diabetes adjudication team, who are in the process of developing new algorithms to identify diabetes in patients without diagnostic Read codes, but with supporting information such as prescriptions, raised blood glucose and process-of-care measures. The testing of these algorithms in population databases such as the CPRD and validation of the resulting diagnoses will help improve the way in which screening and diagnosis of diabetes is primary care are undertaken.

The use of risk stratification and clinical prediction tools is widespread in UK primary care. While some risk scores recommended for use by the NHS already incorporate ethnicity into the risk prediction models, several in current use do not.(347–351) These tools rely almost exclusively on data coded onto the electronic patient record. Improving the quality and completeness of ethnicity and clinical management data in routine electronic health records will improve the identification of populations with the highest health need. There are concerns that existing risk scores may not be effective at screening for undiagnosed conditions, and thus improving these scores, will be of great value.(352)

12.11 Overall Conclusions

This thesis has provided a demonstration of the value of electronic health records for examining ethnic differences in health for UK populations. The significance of the results presented in the thesis is three-fold. Firstly, the thesis has reported novel results which provide valuable information about the usability and generalisability of ethnicity data available in UK-based electronic health records. Secondly, the research has replicated findings from other studies around the prevalence and incidence of diabetes and patterning of ethnic differences in heart disease outcomes, providing reassurance about the quality of these data and their suitability for investigating ethnic differences in major disease outcomes. Thirdly, the thesis has extended our knowledge of the ethnic patterning of disease by further investigating trends in ethnic minority subgroups.

Together, the findings reported in this thesis provide a unique insight into the ways in which routinely recorded ethnicity data can be maximised for the purposes of epidemiological research into health inequalities across the UK. Each of the studies undertaken represents a novel use of the CPRD and its linked datasets. The results suggest these data need to be used more widely in order to examine a wide range of questions relating to ethnic differences in the utilisation of health services, management of existing health conditions and predicting the future burden of disease in the UK population.

References

1. Bhopal RS. Ethnicity, Race, and Health in Multicultural Societies: Foundations for better epidemiology, public health, and health care. Eur J Public Heal [Internet]. 2007/06/05 ed. New York: Oxford University Press; 2007;17(3):238–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17545231

2. Mathur R, Hull SA, Badrick E, Robson J. Cardiovascular multimorbidity: the effect of ethnicity on prevalence and risk factor management. Br J Gen Pract [Internet]. 2011/05/31 ed. 2011;61(586):e262–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21619750

3. M.S. C, M.G. P. Tackling health inequalities by means of an evidence-based health promotion programme [Internet]. International Journal of Pharmacy Practice. M.S. Culshaw, University of Huddersfield, United Kingdom. E-mail: m.culshaw@hud.ac.uk: Pharmaceutical Press; 2012. p. 20–1. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed10&NEWS=N&AN=70760058

4. Improving the health of black and minority ethnic groups: a guide for primary care organisations. Kings Fund [Internet]. 2000;71. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=sopp&AN=230604-75668

5. Smith NR. Intergenerational Continuities in Ethnic Inequalities in Health in the UK. 2010;

6. Ethnic variations in the UK incidence of and survival from myocardial infarction: Cohort study linking census, mortality and morbidity data. Eur J Cardiovasc Prev Rehabil [Internet]. Conference:EuroPRevent 2011 Geneva Switzerland. Conference St. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed10&AN=70472417

7. Streetly A. Health of gypsies/travellers. Lancet. 1987/09/12 ed. 1987;2(8559):630.

8. Salway S, Barley R, Allmark P, Gerrish K, Higginbottom G, Ellison GTH. Ethnic diversity and inequality: ethical and scienfici rigour in social research. [Internet]. Joseph Rowntree Foundation, editor. York; 2011. Available from: http://www.jrf.org.uk/sites/files/jrf/ethnicity-social-policy-research-full.pdf

9. E. B, S. H, R. M, S. S, K. B, S. B, et al. Health equity audits in general practice: a strategy to reduce health inequalities. Prim Health Care Res Dev [Internet]. E. Badrick; 2014 Jan

[cited 2014 Dec 29];15(1):80–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23375244

10. Diabetes UK. the Cost of Diabetes Report. 2014.

11. Chiu M, Austin PC, Manuel DG, Shah BR, Tu J V. Deriving ethnic-specific BMI cutoff points for assessing diabetes risk. Diabetes Care [Internet]. 2011/06/18 ed. 2011;34(8):1741–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21680722

12. Gholap N, Davies M, Patel K, Sattar N, Khunti K. Type 2 diabetes and cardiovascular disease in South Asians. Prim Care Diabetes [Internet]. 2010/09/28 ed. 2011;5(1):45–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20869934

13. Bellary S. Enhanced diabetes care to people of South Asian ethnicity (UKADS): A thesis submitted to The University of Birmingham for the degree of Doctor of Medicine. 2010.

14. Kurian AK, Cardarelli KM. Racial and ethnic differences in cardiovascular disease risk factors: a systematic review. Ethn Dis [Internet]. 2007/02/06 ed. 2007;17(1):143–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17274224

15. Freedman BJ. For debate...Caucasian. Br Med J (Clin Res Ed). 1984/03/03 ed. 1984;288(6418):696–8.

16. D.R. W, Williams DR. The concept of race in Health Services Research: 1966 to 1990. Health Serv Res [Internet]. 1994;29(3):267–74. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed3&AN=1994271693

17. Bhopal R. Is research into ethnicity and health racist, unsound, or important science? BMJ [Internet]. 1997/06/14 ed. 1997;314(7096):1751–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9202509

18. Fullilove MT. Comment: abandoning “race” as a variable in public health research--an idea whose time has come. Am J Public Heal [Internet]. 1998/09/16 ed. 1998;88(9):1297–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9736864

19. United Nations Economic and Security Council. Statement by Experts on Problems of Race. Am Anthropol. 1951;53(1):142–5.

20. Bhopal R. Race and ethnicity: responsible use from epidemiological and public health perspectives. J Law, Med Ethics [Internet]. 2006;34(3):479,500–7. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=17144172

21. Cooper RS, Warren RC, Hahn RA, Bristow L, Yu ES. A case study in the use of race and ethnicity in public health surveillance. Public Heal Rep [Internet]. 1994/01/01 ed. 1994;109(1):46–52. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed3&AN=1994066511

22. Carter B, Fenton S. Not thinking ethnicity: A critique of the ethnicity paradigm in an over-ethnicised sociology. J Theory Soc Behav. 2010;.40(1):pp.

23. Department of Health. A practical guide to ethnic monitoring in the NHS and social care [Internet]. 2005. Available from: http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/publicationsandstatistics/publications/publicationspolicyandguidance/browsable/DH_4116927

24. Kohut K, D'Mello L, Bancroft EK, Thomas S, Young M-A, Myhill K, et al. Implications for cancer genetics practice of pro-actively assessing family history in a General Practice cohort in North West London. Fam Cancer [Internet]. 2012 Mar [cited 2014 Oct 17];11(1):107–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22016143

25. Ethnicity questions and antenatal screening for sickle cell/thalassaemia [EQUANS] in England: A randomised controlled trial of two questionnaires. Ethn Heal [Internet]. 2006;11(2):169–89. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed7&AN=2006174445

26. Nazroo J. The health of Britain's ethnic minorities: findings from a national survey [Internet]. PSI; 1997 [cited 2014 Aug 27]. Available from: http://discovery.ucl.ac.uk/179498/

27. Nathanson C, Hopper K. The Marmot Review - social revolution by stealth. Soc Sci Med [Internet]. 2010/08/10 ed. 2010;71(7):1237–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20692085

28. Noble D, Smith D, Mathur R, Robson J, Greenhalgh T. Feasibility study of geospatial mapping of chronic disease risk to inform public health commissioning. BMJ Open [Internet]. British Medical Journal Publishing Group; 2012 Jan [cited 2014 Dec 29];2(1):e000711. Available from: http://bmjopen.bmj.com/content/2/1/e000711.abstract

29. Chaikiat A, Li X, Bennet L, Sundquist K. Neighborhood deprivation and inequities in coronary heart disease among patients with diabetes mellitus: a multilevel study of 334,000 patients. Health Place [Internet]. England: Center for Primary Health Care Research, Lund University, Malmo, Sweden.; 2012;18(4):877–82. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=22498356

30. Garner S, Bhattacharyya G. Poverty , ethnicity and place. 2011.

31. Department of Health. Tackling Health Inequalities: A program for Action [Internet]. Publications D of H, editor. London; 2003. Available from: www.doh.gov.uk/healthinequalities/programmeforaction

32. Bhopal RS. The quest for culturally sensitive health-care systems in Scotland: insights for a multi-ethnic Europe. J Public Health (Oxf) [Internet]. 2012 Mar [cited 2014 Oct 14];34(1):5–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22267290

33. Aung NC, Rechel B, Odermatt P. Access to and utilisation of GP services among Burmese migrants in London: a cross-sectional descriptive study. BMC Health Serv Res [Internet]. BioMed Central Ltd; 2010 Jan [cited 2014 Oct 14];10(1):285. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2970605&tool=pmcentrez&rendertype=abstract

34. Topal K, Eser E, Sanberk I, Bayliss E, Saatci E. Challenges in access to health services and its impact on quality of life: a randomised population-based survey within Turkish speaking immigrants in London. Health Qual Life Outcomes [Internet]. BioMed Central Ltd; 2012 Jan [cited 2014 Oct 17];10(1):11. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3275503&tool=pmcentrez&rendertype=abstract

35. Misra A, Ganda OP. Migration and its impact on adiposity and type 2 diabetes. Nutrition [Internet]. 2007/08/07 ed. 2007;23(9):696–708. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17679049

36. Bentham G. Migration and morbidity: implications for geographical studies of disease. Soc Sci Med [Internet]. 1988 Jan [cited 2014 Oct 17];26(1):49–54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3353753

37. Smith NR, Kelly YJ, Nazroo JY. The effects of acculturation on obesity rates in ethnic minorities in England: evidence from the Health Survey for England. Eur J Public Health [Internet]. 2012 Aug [cited 2014 Sep 19];22(4):508–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21697245

38. Smith NR, Kelly YJ, Nazroo JY. Intergenerational continuities of ethnic inequalities in general health in England. J Epidemiol Community Health [Internet]. 2009;63(3):253–8. Available from: http://jech.bmj.com/content/63/3/253.abstract

39. A review of ethnicity, health and nutrition-related diseases in relation to migration in the United Kingdom. Public Health Nutr [Internet]. 2001;4(2 B):647–57. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed5&AN=2001207488

40. Hurcombe R, Bayley M, Goodman A. Ethnicity and alcohol: a review of the UK literature [Internet]. 2010. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=sopp&AN=192386-116429

41. Chris W Smith PSD. Immigration Patterns of Non-UK Born Populations in England and Wales in 2011 [Internet]. [cited 2014 Dec 29]. Available from: http://www.ons.gov.uk/ons/rel/census/2011-census-analysis/immigration-patterns-and-characteristics-of-non-uk-born-population-groups-in-england-and-wales/story-on-immigration-patterns-of-non-uk-born-populations-in-england-and-wales-in-2011.html?format=print

42. Survey H, Health Survey for England. The health of minority ethnic groups. 2004.

43. Tucker CM, Herman KC, Pedersen TR, Higley B, Montrichard M, Ivery P. Cultural sensitivity in physician-patient relationships: perspectives of an ethnically diverse sample of low-income primary care patients.[Erratum appears in Med Care. 2003 Dec;41(12):1330]. Med Care [Internet]. 41(7):859–70. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=12835610

44. Simon BP. " Ethnic" statistics and data protection in the Council of Europe countries. October. 2007;

45. Psoinos M, Hatzidimitriadou E, Butler C, Barn R. Ethnic monitoring in Healthcare Services in the UK as a mechanism to address Health Disparities: A Narrative Review [Internet]. London: Swan IPI; 2011. Available from: http://eprints.kingston.ac.uk/18866/

46. Simon P. "Ethnic" statistics and data protection in the Council of Europe countries. Strasbourg: Council of Europe; 2007.

47. UK Government. Race Relations (Amendment) Act 2000 [Internet]. Crown Corporation; 2000. Available from: http://www.legislation.gov.uk/ukpga/2000/34

48. Baxter C, Baxter D. Racial inequalities in health: a challenge to the British National Health Service. Int J Heal Serv. 1988;18:563–71.

49. Silvera M. Proposal for guidelines on monitoring the ethnic origin of service users. London: Parkside Health Authority; 1991.

50. National Association of Health Authorities in England and Wales. Actions not words. Birmingham: National Association of Health Authorities; 1988.

51. Parliament of the United Kingdom, editor. Census Order 1920 [Internet]. England: the King's Printer of Acts of Parliament; 1920. Available from: http://www.legislation.gov.uk/ukpga/1920/41/pdfs/ukpga_19200041_en.pdf

52. London Health Observatory. Ethnicity Coding [Internet]. 2012. Available from: http://www.lho.org.uk/LHO_Topics/National_Lead_Areas/EHIP/EthnicityCoding.aspx#History

53. Office for National Statistics. Labour Force Survey User Guide [Internet]. Volume 1 - LFS Background and Methodology 2011. 2011. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/labour-market-statistics/index.html

54. Ballard R. Britain's Visible Minorities: A demographic overview [Internet]. Stalybridge: Centre for Applied South Asian Studies; Available from: http://www.casas.org.uk/papers/pdfpapers/demography.pdf

55. H. B. Ethnic and racial questions in the Census: The home affairs committee report. J Ethn Migr Stud. 1983;11(1-2):83–91.

56. Coleman D, Salt J. Ethnicity in the 1991 census: volume 1: demographic characteristics of ethnic minority populations [Internet]. HMSO, editor. London; 1996. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=sopp&AN=239325-312

57. Bulmer M. A controversial census topic: race and ethnicity in the British census. J Off Stat [Internet]. 1986/01/01 ed. 1986;2(4):471–80. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12341365

58. Northern Ireland Statistics & Research Agency. Comparability with the 1991 Census [Internet]. Available from: http://www.nisra.gov.uk/archive/census/Metadata/Chapter 4 - Comparability with the 1991 Census.pdf

59. Heath I. The role of ethnic monitoring in general practice. Br J Gen Pract [Internet]. 41(349):310–1. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=1777272

60. Simpson L, Akinwale B. Stability and change in ethnic groups in England and Wales. Popul Trends, no [Internet]. 2005;121. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=sopp&AN=46261-051018205

61. Goldring S, Newman J. The ONS Longitudinal Study--a prestigious past and a bright future. Popul Trends [Internet]. 2010/04/10 ed. 2010;(139):4–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20379275

62. Centre for Longitudinal Study Information and User Support. 2001 Census Appendix 13: Ethnic Group [Internet]. Available from: http://www.celsius.lshtm.ac.uk/dataDict/appendices/01app13.html

63. Office for National Statistics. Final recommended questions for the 2011 Census in England and Wales: National Identity [Internet]. 2009. Available from: http://www.ons.gov.uk/ons/search/index.html?pageSize=50&newquery=census+2011+final+recommended+questions

64. Aspinall PJ. Describing the "white" ethnic group and its composition in medical research. Soc Sci Med [Internet]. 1998;47(11):1797–808. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=9877349

65. U.S Census Bureau. Directive No. 15: Race and Ethnic standards for federal statistics and administrative reporting [Internet]. Bureau USC, editor. 1977. Available from: http://wonder.cdc.gov/wonder/help/populations/bridged-race/Directive15.html

66. Bhopal RS. Heterogeneity among Indians, Pakistanis, and Bangladeshis is key to racial inequities. BMJ [Internet]. 2002/10/19 ed. 2002;325(7369):903. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12386051

67. Senior PA, Bhopal R. Ethnicity as a variable in epidemiological research. BMJ [Internet]. 1994;309(6950):327–30. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=8086873

68. Bhopal R. Glossary of terms relating to ethnicity and race: for reflection and debate. J Epidemiol Community Heal [Internet]. 2004/05/15 ed. 2004;58(6):441–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15143107

69. Mason D. A rose by any other name...? Categorisation, identity and social science. J Ethn Migr Stud [Internet]. Routledge; 1990 Jun 30 [cited 2014 Oct 15];17(1):123–33. Available from: http://dx.doi.org/10.1080/1369183X.1990.9976227

70. Chaturvedi N, McKeigue PM. Methods for epidemiological surveys of ethnic minority groups. J Epidemiol Community Health [Internet]. 1994/04/01 ed. 1994;48(2):107–11. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8189161

71. Proposals for the Responsible Use of Racial and Ethnic Categories in Biomedical Research: Where Do We Go From Here? Proceedings of a conference, April 2005. J Law

Med Ethics [Internet]. 2006;34(3):483–619. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed7&AN=17203547

72. Bhopal R, Donaldson L. White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health. Am J Public Heal [Internet]. 1998/09/16 ed. 1998;88(9):1303–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9736867

73. Lockwood MJ. General Practice Research Database provides detailed anonymised data. BMJ [Internet]. 1996/09/21 ed. 1996;313(7059):757. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8819474

74. Walley T, Mantgani A. The UK General Practice Research Database. Lancet [Internet]. 1999/04/23 ed. 1997;350(9084):1097–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10213569

75. Lawson DH, Sherman V, Hollowell J. The General Practice Research Database. Scientific and Ethical Advisory Group. QJM [Internet]. 1998/08/26 ed. 1998;91(6):445–52. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9709463

76. Crisis UK. Critical Condition: Vulnerable single homeless people and access to GPs [Internet]. London; 2002. Available from: http://www.crisis.org.uk/data/files/publications/GP_policybrief[1].pdf

77. Accessing primary health care: A meta-ethnography of the experiences of British South Asian patients with diabetes, coronary heart disease or a mental health problem. Chronic Illn [Internet]. 2012;8(2):135–55. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed10&AN=2012278679

78. Klinker S, Fitzpatrick S. Reaching out: a study of black and minority ethnic single homelessness and access to primary health care [Internet]. 2000. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=sopp&AN=260685-0000072453+

79. Carey IM, Cook DG, De Wilde S, Bremner SA, Richards N, Caine S, et al. Developing a large electronic primary care database (Doctors' Independent Network) for research. Int J Med Inform [Internet]. 2004;73(5):443–53. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=15171985

80. Van Staa TP, Goldacre B, Gulliford M, Cassell J, Pirmohamed M, Taweel A, et al. Pragmatic randomised trials using routine electronic health records: putting them to

the test. BMJ [Internet]. 2012/02/09 ed. 2012;344:e55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22315246

81. Gallagher AM, Smeeth L, Seabroke S, Leufkens HGM, van Staa TP. Risk of Death and Cardiovascular Outcomes with Thiazolidinediones: A Study with the General Practice Research Database and Secondary Care Data. PLoS One [Internet]. 2011/12/14 ed. Public Library of Science; 2011;6(12):e28157. Available from: http://dx.doi.org/10.1371/journal.pone.0028157

82. Department of Health. Launch of the Clinical Practice Research Datalink [Internet]. 2011. Available from: http://www.dh.gov.uk/health/2011/10/cprd-launch/

83. Marston L, Carpenter JR, Walters KR, Morris RW, Nazareth I, White IR, et al. The validity of routinely recorded smoking status in UK primary care : a cross-sectional study. 2014;1–7.

84. Herrett E, Shah AD, Boggon R, Denaxas S, Smeeth L, van Staa T, et al. Completeness and diagnostic validity of recording acute myocardial infarction events in primary care, hospital care, disease registry, and national mortality records: cohort study. BMJ [Internet]. 2013;346(7909):f2350. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medc&AN=23692896

85. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. Br J Clin Pharmacol [Internet]. 2010/01/19 ed. 2010;69(1):4–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20078607

86. Garcia Rodriguez L. Case validation in research using large databases. Ann Rheum Dis [Internet]. 2013;72:31–2. Available from: <Go to ISI>://WOS:000331587901135

87. Health and Social Care Information Centre. GP Systems of Choice — Health and Social Care Information Centre [Internet]. [cited 2014 Dec 29]. Available from: http://systems.hscic.gov.uk/gpsoc

88. Chisholm J. The Read clinical classification. BMJ [Internet]. 1990/04/28 ed. 1990;300(6732):1092. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2344534

89. Le H V, Truong CT, Webb D, Mynepalli L, Hodgson E, Bennett D, et al. A clinical classification dictionary for read codes in the general practice research database. Pharmacoepidemiol Drug Saf [Internet]. 2012;Conference:28th International Conference on Pharmacoepidemiol. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed10&AN=70935822

90. Hughes G. Hospital Episode Statistics: are they anything to write home about? Emerg Med J [Internet]. 2009/05/26 ed. 2009;26(6):392. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19465603

91. Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. BMJ [Internet]. 2003;326(7398):1070. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=12750210

92. Keable-Elliott D, Campion AM. Ethnic monitoring in general practice. Br J Gen Pract [Internet]. 47(418):329–30. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=9219419

93. Collecting race, ethnicity, and language data to identify and reduce health disparities: Perceptions of health plan enrollees. Med Care Res Rev [Internet]. 2011;68(3):367–81. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed10&AN=2011259917

94. Aspinall PJ, Anionwu E. The Role of Ethnic Monitoring in Mainstreaming Race Equality and the Modernization of the NHS: A Neglected Agenda? Crit Public Health [Internet]. 2002;12(January 2015):37–41. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed5&AN=2002196678

95. Kumarapeli P, Stepaniuk R, de Lusignan S, Williams R, Rowlands G. Ethnicity recording in general practice computer systems. J Public Health (Bangkok) [Internet]. 28(3):283–7. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=16840765

96. Aspinall PJ. Department of Health's requirement for mandatory collection of data on ethnic group of inpatients. BMJ [Internet]. 1995;311(7011):1006–9. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=7580590

97. Health and Social Care Information Centre. How good is HES ethnic coding and where do the problems lie [Internet]. 2011. Available from: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=171

98. Clinical Informatics Research Group. Ethnicity Coding [Internet]. 2012. Available from: http://www.clininf.eu/ethnicity.html

99. Home Office Identity & Passport Service. Official information on births, marriages and deaths [Internet]. 2012. Available from: http://www.gro.gov.uk/gro/content/certificates/most_customers_want_to_know.asp#CertificateInformation

100. Christie B. Scotland introduces record of ethnicity on death certificates. BMJ. 2012;344.

101. Jones M, Kai J. Capturing ethnicity data in primary care: challenges and feasibility in a diverse metropolitan population. Divers Heal Soc Care [Internet]. 2007;4(3):211–20. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=sopp&AN=123989-

102. Nitsch D, Kadalayil L, Mangtani P, Steenkamp R, Ansell D, Tomson C, et al. Validation and utility of a computerized South Asian names and group recognition algorithm in ascertaining South Asian ethnicity in the national renal registry. QJM [Internet]. 2009/10/16 ed. 2009;102(12):865–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19828643

103. Hull SA, Rivas C, Bobby J, Boomla K, Robson J. Hospital data may be more accurate than census data in estimating the ethnic composition of general practice populations. Inform Prim Care [Internet]. 17(2):67–78. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=19807948

104. Wong EC, Palaniappan LP, Lauderdale DS. Using name lists to infer Asian racial/ethnic subgroups in the healthcare setting. Med Care [Internet]. 48(6):540–6. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=20421828

105. General Practitioners Committee. Ethnicity and first language recording- GPC guidance [Internet]. British Medical Association, editor. 2011. Available from: http://www.bma.org.uk/images/gpcethnicityfirstlanguagemay2011_v2_tcm41-206033.pdf

106. Millett C. Incentives to improve ethnicity coding in primary care. Equality and inequality in Health . London: The Kings Fund; 2010.

107. NHS Employers. Changes to QOF 2011/12 [Internet]. 2011 [cited 2015 Jun 12]. Available from: http://www.nhsemployers.org/your-workforce/primary-care-contacts/general-medical-services/quality-and-outcomes-framework/changes-to-qof-201112

108. Department of Health. Working for patients: Working paper 11. Framework for information systems: the next steps. DoH, editor. London; 1990.

109. Aspinall PJ. The mandatory collection of data on ethnic group of inpatients: experience of NHS trusts in England in the first reporting years. Public Heal [Internet]. Nature Publishing Group; 2000;114(4):254–9. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=10962586

110. Health and Social Care Information Centre. HES User Guide [Internet]. 2010. Available from: http://www.hesonline.nhs.uk/Ease/servlet/ContentServer?siteID=1937&categoryID=87

111. Group NIM. Group - 1994 - Data Set Change Notice Ethnic Group Data Collection.pdf. 1994.

112. Iqbal G, Johnson MRD, Szczepura A, Wilson S, Gumber A, Dunn JA. UK ethnicity data collection for healthcare statistics : the South Asian perspective UK ethnicity data collection for healthcare statistics : the South Asian perspective. 2012;243(March).

113. Pringle M, Rothera I. Practicality of recording patient ethnicity in general practice: Descriptive intervention study. BMJ Br Med J (International Ed [Internet]. 1996;312(7038):1080–2. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=9605141064&site=ehost-live

114. Hippisley-Cox J, Coupland C. Predicting the risk of chronic Kidney Disease in men and women in England and Wales: prospective derivation and external validation of the QKidney Scores. BMC Fam Pract [Internet]. 2010 Jan [cited 2014 Oct 18];11(1):49. Available from: http://www.biomedcentral.com/1471-2296/11/49

115. 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 [Internet]. 2010/12/15 ed. J. Hippisley-Cox, Division of Primary Care, University Park, Nottingham NG2 7RD, United Kingdom. E-mail: juliahippisleycox@gmail.com: BMJ Publishing Group (Tavistock Square, London WC1H 9JR, United Kingdom); 2011;342(7788):93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21148212

116. Collins GS, Altman DG. External validation of QDSCORE((R)) for predicting the 10-year risk of developing Type 2 diabetes. Diabet Med [Internet]. 2011/04/13 ed. 2011;28(5):599–607. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21480970

117. Collins GS, Altman DG. An independent and external validation of QRISK2 cardiovascular disease risk score: a prospective open cohort study. BMJ [Internet]. 2010 Jan [cited 2014 Oct 17];340:c2442. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2869403&tool=pmcentrez&rendertype=abstract

118. Billings J, Dixon J, Mijanovich T, Wennberg D. Case finding for patients at risk of readmission to hospital: development of algorithm to identify high risk patients. BMJ [Internet]. 2006 Aug 12 [cited 2014 Aug 21];333(7563):327. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1539047&tool=pmcentrez&rendertype=abstract

119. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: Prospective derivation and validation of QRISK2 [Internet]. BMJ. J. Hippisley-Cox, Division of Primary Care, Tower Building, University Park, Nottingham NG2 7RD. E-mail: Julia.hippisley-cox@nottingham.ac.uk: BMJ Publishing Group; 2008. p. 1475–82. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed8&NEWS=N&AN=2008334775

120. Shah A, Diggens N, Stiller C, Murphy D, Passmore J, Murphy MFG. Place of death and hospital care for children who died of cancer in England, 1999-2006. Eur J Cancer [Internet]. 2011 Sep [cited 2014 Oct 18];47(14):2175–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21531128

121. Hacker J, Stanistreet D. Equity in waiting times for two surgical specialties: a case study at a hospital in the North West of England. J Public Health (Oxf) [Internet]. 2004 Mar [cited 2014 Oct 18];26(1):56–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15044576

122. Mindell J, Klodawski E, Fitzpatrick J. Using routine data to measure ethnic differentials in access to coronary revascularization. J Public Health (Oxf) [Internet]. 2008 Mar [cited 2014 Oct 18];30(1):45–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18042654

123. Bragg F, Cromwell DA, Edozien LC, Gurol-Urganci I, Mahmood TA, Templeton A, et al. Variation in rates of caesarean section among English NHS trusts after accounting for maternal and clinical risk: cross sectional study. BMJ [Internet]. 2010 Jan [cited 2014 Oct 18];341:c5065. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2950923&tool=pmcentrez&rendertype=abstract

124. Cooper SC, Day R, Brooks C, Livings C, Thomson CS, Trudgill NJ. The influence of deprivation and ethnicity on the incidence of esophageal cancer in England. Cancer Causes Control [Internet]. 2009 Oct [cited 2014 Oct 18];20(8):1459–67. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19533393

125. Jack RH, Davies EA, Møller H. Testis and prostate cancer incidence in ethnic groups in South East England. Int J Androl [Internet]. 2007 Aug [cited 2014 Oct 18];30(4):215–20; discussion 220–1. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17573844

126. Downing A, West RM, Gilthorpe MS, Lawrence G, Forman D. Using routinely collected health data to investigate the association between ethnicity and breast cancer incidence and survival: what is the impact of missing data and multiple ethnicities? Ethn Health [Internet]. 2011 Jun [cited 2014 Oct 18];16(3):201–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21462016

127. Mann AG, Trotter CL, Balogun MA, Ramsay ME. Hepatitis C in ethnic minority populations in England. J Viral Hepat [Internet]. 2008 Jun [cited 2014 Oct 18];15(6):421–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18208498

128. Bottle A, Aylin P, Majeed A. Identifying patients at high risk of emergency hospital admissions: a logistic regression analysis. J R Soc Med [Internet]. 2006 Aug [cited 2014 Oct 18];99(8):406–14. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1533524&tool=pmcentrez&rendertype=abstract

129. Collins GS, Altman DG. An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study. BMJ [Internet]. 2009/07/09 ed. 2009;339:b2584. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19584409

130. Neumann CS, Cooper BT. Ethnic differences in gastro-oesophageal reflux disease. Eur J Gastroenterol Hepatol [Internet]. 11(7):735–9. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=10445793

131. Mathur R, Noble D, Smith D, Greenhalgh T, Robson J. Quantifying the risk of type 2 diabetes in East London using the QDScore: A cross-sectional analysis. Br J Gen Pract [Internet]. 2012 Oct [cited 2014 Dec 29];62(October):663–70. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3459773&tool=pmcentrez&rendertype=abstract

132. P. S, O. S, M. A, Schofield P, Saka O, Ashworth M, et al. Ethnic differences in blood pressure monitoring and control in South East London. Br J Gen Pract [Internet]. 2011/03/29 ed. P. Schofield, King's College London, Department of Primary Care and Public Health Sciences, Capital House, 42Weston Street, London, SE1 3QD, United Kingdom. E-mail: peter.1.schofield@kcl.ac.uk: Royal College of General Practitioners (14 Princes Gate, London SW7 1PU, United Kingdom); 2011;61(585):e190–6. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed10&NEWS=N&AN=2011289473

133. Fischbacher CM, Bhopal R, Steiner M, Morris AD, Chalmers J, C.M. F, et al. Is there equity of service delivery and intermediate outcomes in South Asians with type 2 diabetes? Analysis of DARTS database and summary of UK publications. J Public Health (Oxf)

[Internet]. England: Information Services Division, NHS National Services Scotland, Edinburgh EH12 9EB, UK. colin.fischbacher@isd.csa.scot.nhs.uk; 2009;31(2):239–49. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed9&NEWS=N&AN=19196794

134. Brown R, Shantsila E, Varma C, Lip G. Ethnic and gender differences in symptom-to-door times in patients presenting with st elevation myocardial infarction [Internet]. Heart. R. Brown, University of Birmingham, Centre for Cardiovascular Sciences, United Kingdom: BMJ Publishing Group; 2014. Available from: http://japha.org/article.aspx?articleid=1838416

135. Kong M, Meakin L, Donley P, Gregory R, Scudamore I. Evidence of an "epidemic" of gestational diabetes (19952008): Implications for service delivery [Internet]. Diabetic Medicine. M. Kong, Department of Diabetes, University Hospitals of Leicester NHS Trust, Leicester, United Kingdom: Blackwell Publishing Ltd; 2010. p. 168. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed9&NEWS=N&AN=70087859

136. Ahdieh L, Hahn RA. Use of the terms "race", "ethnicity", and "national origins": a review of articles in the American Journal of Public Health, 1980-1989. Ethn Heal. 1996/03/01 ed. 1996;1(1):95–8.

137. Bhopal R, Rankin J. Concepts and terminology in ethnicity, race and health: be aware of the ongoing debate. Br Dent J [Internet]. 1999 May 22 [cited 2014 Oct 18];186(10):483–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10379073

138. Oakley A. Ethnicity and research evaluating interventions: issues of science and ethics. In: Nazroo JY, editor. Health and Social Research in Multiethnic Populations. Abingdon: Routledge; 2006. p. 142–64.

139. Mathur R, Grundy E, Smeeth L. Availability and use of UK based ethnicity data for health research [Internet]. National Centre for Research Methods Working Paper Series. London: n/a; 2013 Mar [cited 2014 Oct 18]. Report No.: 01/13. Available from: http://eprints.ncrm.ac.uk/3040/1/Mathur-_Availability_and_use_of_UK_based_ethnicity_data_for_health_res_1.pdf

140. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. Br J Gen Pract [Internet]. 2010 Mar [cited 2014 Sep 24];60(572):e128–36. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2828861&tool=pmcentrez&rendertype=abstract

141. Thomas SL, Edwards CJ, Smeeth L, Cooper C, Hall AJ. How accurate are diagnoses for rheumatoid arthritis and juvenile idiopathic arthritis in the general practice research

database? Arthritis Rheum [Internet]. 2008/09/02 ed. 2008;59(9):1314–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18759262

142. Jick SS, Kaye JA, Vasilakis-Scaramozza C, Garcia Rodriguez LA, Ruigomez A, Meier CR, et al. Validity of the general practice research database. Pharmacotherapy [Internet]. 2003/05/14 ed. 2003;23(5):686–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12741446

143. Hollowell J. The General Practice Research Database: quality of morbidity data. Popul Trends [Internet]. 1997/04/01 ed. 1997;(87):36–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9134574

144. Millett C, Netuveli G, Saxena S, Majeed A. Impact of pay for performance on ethnic disparities in intermediate outcomes for diabetes: a longitudinal study. Diabetes Care [Internet]. 2009;32(3):404–9. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=19073759

145. Hull S a., Mathur R, Badrick E, Robson J, Boomla K. Recording ethnicity in primary care: Assessing the methods and impact. Br J Gen Pract [Internet]. 2011;61(May):e290–4. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=21619753

146. Nolan CJ, Damm P, Prentki M. Type 2 diabetes across generations: from pathophysiology to prevention and management. Lancet [Internet]. Elsevier Ltd; 2011 Jul 9 [cited 2014 Jul 11];378(9786):169–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21705072

147. Stumvoll M, Goldstein BJ, Haeften TW Van. Type 2 diabetes : principles of pathogenesis and therapy. 2010;365.

148. Lebovitz HE. Type 2 diabetes: the evolution of a disease. Br J Diabetes Vasc Dis [Internet]. 2013 Jan 7 [cited 2014 Sep 22];12(6):290–8. Available from: http://dvd.sagepub.com/cgi/doi/10.1177/1474651412470289

149. Kuzuya T, Nakagawa S, Satoh J, Kanazawa Y, Iwamoto Y, Kobayashi M, et al. Report of the Committee on the classification and diagnostic criteria of diabetes mellitus. Diabetes Res Clin Pract [Internet]. 2002 Jan [cited 2014 Jul 14];55(1):65–85. Available from: http://www.sciencedirect.com/science/article/pii/S0168822701003655

150. Taylor SI. Deconstructing type 2 diabetes. Cell. 1999;97:9–12.

151. Lebovitz HE. Science, clinical outcomes and the popularization of diabetes surgery. Curr Opin Endocrinol Diabetes Obes [Internet]. 2012 Oct [cited 2014 Nov 14];19(5):359–66. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22914565

152. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2013/12/21 ed. 2014;37 Suppl 1:S81–90.

153. Schindler C. Influences of diabetes on the development and progression of cardiovascular disease. Br J Diabetes Vasc Dis [Internet]. 2010 Apr 12 [cited 2014 Sep 22];10(2):54–6. Available from: http://dvd.sagepub.com/cgi/doi/10.1177/1474651410367260

154. A. O, I. D, Y. K, H. C, A. K, Onat A, et al. Type-2 diabetes and coronary heart disease: Common physiopathology, viewed from autoimmunity. Expert Rev Cardiovasc Ther [Internet]. A. Onat, Department of Cardiology, Cerrahpasa Medical Faculty, Istanbul University, Istanbul, Turkey. E-mail: alt_onat@yahoo.com.tr: Expert Reviews Ltd.; 2014 Jun [cited 2014 Aug 27];12(6):667–79. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24846677

155. Cowie CC, Port FK, Wolfe RA, Savage PJ, Moll PP, Hawthorne VM. Disparities in incidence of diabetic end-stage renal disease according to race and type of diabetes. N Engl J Med [Internet]. 321(16):1074–9. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med3&AN=2797067

156. Raymond NT, Paul O'Hare J, Bellary S, Kumar S, Jones A, Barnett A. Comparative risk of microalbuminuria and proteinuria in UK residents of south Asian and white European ethnic background with type 2 diabetes: A report from UKADS. Curr Med Res Opin [Internet]. 2011;27(SUPPL. 3):47–55. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed10&AN=2011645453

157. Fox CS, Sullivan L, D'Agostino RBS, Wilson PWF, Study FH. The significant effect of diabetes duration on coronary heart disease mortality: the Framingham Heart Study. Diabetes Care [Internet]. United States: National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts 01702, USA. caroline@fram.nhlbi.nih.gov; 2004;27(3):704–8. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=14988289

158. Soderberg S, Cameron A, Magliano D, Alberti G, Tuomilehto J, Kowlessur S, et al. Mortality, all-cause and CVD, over 15 years in multi-ethnic Mauritius: The impact of diabetes and obesity. Scand Cardiovasc J [Internet]. 2010;European Society of Cardiology. Available from:

http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed10&AN=70764386

159. Zabetian A, Sanchez I., Narayan KM., Hwang C., Ali M. Global rural diabetes prevalence: a systematic review and meta-analysis covering 1990-2012. Diabetes Res Clin Pract [Internet]. Ireland: Elsevier Ireland Ltd; 2014;104(2):206–13. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=24507869

160. Beagley J, Guariguata L, Weil C, Motala A. Global estimates of undiagnosed diabetes in adults [Internet]. Diabetes Research and Clinical Practice. J. Beagley, International Diabetes Federation, Brussels, Belgium. E-mail: jessica.beagley@oxon.org: Elsevier Ireland Ltd; 2014. p. 150–60. Available from: www.elsevier.com/locate/diabres

161. Leahy J. Pathogenesis of type 2 diabetes mellitus. Arch Med Res [Internet]. J.L. Leahy, University of Vermont, College of Medicine, Given Building, C-331, 89 Beaumont Ave., Burlington, VT 05405, United States. E-mail: jleahy@uvm.edu: Elsevier Inc.; 2005;36(3):197–209. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=2005244331

162. Naser K., Gruber A, Thomson G. The emerging pandemic of obesity and diabetes: Are we doing enough to prevent a disaster? [Internet]. International Journal of Clinical Practice. Blackwell Publishing Ltd; 2006. p. 1093–7. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=2006390512

163. Zabetian A, Keli HM, Echouffo-Tcheugui JB, Narayan KMV, Ali MK. Diabetes in the Middle East and North Africa. Diabetes Res Clin Pract [Internet]. Ireland: Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, United States. azabeti@emory.edu.; 2013;101(2):106–22. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=23642969

164. Naqshbandi M, Harris S., Esler J., Antwi-Nsiah F. Global complication rates of type 2 diabetes in Indigenous peoples: A comprehensive review [Internet]. Diabetes Research and Clinical Practice. Elsevier Ireland Ltd; 2008. p. 1–17. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed8&NEWS=N&AN=2008459691

165. Chiha M, Njeim M, Chedrawy EG, Maguy Chiha, Mario Njeim EGC. Diabetes and coronary heart disease: a risk factor for the global epidemic. Int J Hypertens [Internet]. Hindawi Publishing Corporation; 2012 Jan [cited 2014 Aug 27];2012:697240. Available from: /pmc/articles/PMC3483823/?report=abstract

166. Liu L, Yin X, Morrissey S. Global variability in diabetes mellitus and its association with body weight and primary healthcare support in 49 low- and middle-income developing countries. Diabet Med [Internet]. England: Blackwell Publishing Ltd (9600 Garsington Road, Oxford OX4 2XG, United Kingdom); 2012;29(8):995–1002. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed10&NEWS=N&AN=2012424848

167. Guariguata L, Whiting D, Weil C, Unwin N, L. G, D. W, et al. The International Diabetes Federation diabetes atlas methodology for estimating global and national prevalence of diabetes in adults. Diabetes Res Clin Pract [Internet]. Ireland: Elsevier Ireland Ltd (P.O. Box 85, Limerick, Ireland); 2011;94(3):322–32. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed10&NEWS=N&AN=2011675236

168. Sicree R, Shaw J. Type 2 diabetes: An epidemic or not, and why it is happening. Diabetes Metab Syndr Clin Res Rev [Internet]. 2007 Jun [cited 2014 Nov 13];1(2):75–81. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1871402106000130

169. WHO | Pacific islanders pay heavy price for abandoning traditional diet. World Health Organization; [cited 2015 Jun 12]; Available from: http://www.who.int/bulletin/volumes/88/7/10-010710/en/

170. Win Tin ST, Lee CMY, Colagiuri R. A profile of diabetes in Pacific Island Countries and Territories. Diabetes Res Clin Pract [Internet]. 2015 Feb [cited 2015 Mar 11];107(2):233–46. Available from: http://www.sciencedirect.com/science/article/pii/S0168822714004707

171. Karter AJ, Schillinger D, Adams AS, Moffet HH, Liu J, Adler NE, et al. Elevated rates of diabetes in Pacific Islanders and Asian subgroups: The Diabetes Study of Northern California (DISTANCE). Diabetes Care [Internet]. 2013 Mar [cited 2015 Jun 12];36(3):574–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3579366&tool=pmcentrez&rendertype=abstract

172. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res Clin Pract [Internet]. Elsevier Ireland Ltd; 2014 Feb [cited 2014 Jul 13];103(2):137–49. Available from: http://www.sciencedirect.com/science/article/pii/S0168822713003859

173. Ramachandran A, Wan Ma RC, Snehalatha C. Diabetes in Asia. Lancet [Internet]. Elsevier Ltd; 2010;375(9712):408–18. Available from: http://dx.doi.org/10.1016/S0140-6736(09)60937-5

174. Scheen AJ, Van Gaal LF, Gaal LF Van. Combating the dual burden: therapeutic targeting of common pathways in obesity and type 2 diabetes. lancet Diabetes Endocrinol

[Internet]. 2014 Feb 19 [cited 2014 Oct 28];2:1–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24731666

175. Scheen AJ, Gaal LF Van. Combatting the dual burden: therapeutic targeting of common pathways in obesity and type 2 diabetes. 2:1–40.

176. Savage DB, Petersen KF, Shulman GI. Disordered lipid metabolism and the pathogenesis of insulin resistance. Physiol Rev [Internet]. 2007 Apr [cited 2014 Oct 29];87(2):507–20. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2995548&tool=pmcentrez&rendertype=abstract

177. Thomas G, Khunti K, Curcin V, Molokhia M, Millett C, Majeed A, et al. Obesity paradox in people newly diagnosed with type 2 diabetes with and without prior cardiovascular disease. Diabetes Obes Metab [Internet]. 2014 Apr [cited 2014 Dec 19];16(4):317–25. Available from: <Go to ISI>://MEDLINE:24118783

178. Oliveros E, Somers VK, Sochor O, Goel K, Lopez-Jimenez F. The concept of normal weight obesity. Prog Cardiovasc Dis [Internet]. Elsevier Inc.; 2014 [cited 2014 Jul 17];56(4):426–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24438734

179. Gluckman PD, Hanson MA, Cooper C, Thornburg KL. Effect of In Utero and Early-Life Conditions on Adult Health and Disease. N Engl J Med [Internet]. 2014 [cited 2014 Nov 15];359:61–73. Available from: http://www.nejm.org/doi/full/10.1056/NEJMra0708473

180. Scheuner MT, Setodji CM, Pankow JS, Blumenthal RS, Keeler E. Relation of familial patterns of coronary heart disease, stroke, and diabetes to subclinical atherosclerosis: the multi-ethnic study of atherosclerosis. Genet Med [Internet]. United States: RAND Corporation, Santa Monica, California 90407-2138, USA. scheuner@rand.org; 2008;10(12):879–87. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=19092440

181. Sarlund H, Laakso M, Voutilainen E, Penttila I, Pyorala K. Familial aggregation of non-insulin dependent diabetes and coronary heart disease are accompanied by different effects on serum lipids, lipoproteins and apolipoproteins. Atherosclerosis [Internet]. NETHERLANDS: Department of Medicine, Kuopio University Central Hospital, Finland.; 1991;86(1):17–29. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med3&NEWS=N&AN=2064632

182. Morris AP, Voight BF, Teslovich TM, Ferreira T, Segrè A V, Steinthorsdottir V, et al. Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. Nat Genet [Internet]. 2012 Sep [cited 2014 Nov 14];44(9):981–90. Available from:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3442244&tool=pmcentrez&rendertype=abstract

183. Bonnefond A, Froguel P, Vaxillaire M. The emerging genetics of type 2 diabetes. Trends Mol Med [Internet]. 2010 Sep [cited 2014 Nov 14];16(9):407–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20728409

184. Forouhi NG, Wareham NJ. The EPIC-InterAct Study: A Study of the Interplay between Genetic and Lifestyle Behavioral Factors on the Risk of Type 2 Diabetes in European Populations. Curr Nutr Rep [Internet]. 2014 Jan [cited 2014 Nov 12];3(4):355–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25383255

185. Vassy JL, Meigs JB. Is genetic testing useful to predict type 2 diabetes? Best Pract Res Clin Endocrinol Metab [Internet]. Elsevier; 2012 Apr 4 [cited 2014 Nov 14];26(2):189–201. Available from: http://www.bprcem.com/article/S1521690X11001229/fulltext

186. Giugliano D, Maiorino MI, Bellastella G, Chiodini P, Esposito K, D. G, et al. Treatment regimens with insulin analogues and haemoglobin A1c target of <7% in type 2 diabetes: a systematic review. Diabetes Res Clin Pract [Internet]. Ireland: Elsevier Ireland Ltd (P.O. Box 85, Limerick, Ireland); 2011;92(1):1–10. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed10&NEWS=N&AN=2011209676

187. S. K. Cardiometabolic risk factors and their treatment in patients with type 2 diabetes [Internet]. Expert Review of Endocrinology and Metabolism. S. Kumar, University of Warwick, Clinical Sciences Research Institute, Warwick Medical School, Coventry CV4 7AL, United Kingdom. E-mail: sudhesh.kumar@warwick.ac.uk: Future Drugs Ltd; 2007. p. 331–9. Available from: http://www.future-drugs.com/doi/pdf/10.1586/17446651.2.3.331

188. S. P. Glycaemic management of type 2 diabetes [Internet]. Clinical Medicine, Journal of the Royal College of Physicians of London. S. Page, Dept. Diabet., Endocrinol./Clin. N., University Hospital NHS Trust, Queen's Medical Centre, Derby Road, Nottingham NG7 2UH, United Kingdom. E-mail: simon.page@mail.qmcuh-tr.trent.nhs.uk: Royal College of Physicians; 2004. p. 302–6. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed6&NEWS=N&AN=2004358903

189. Inzucchi SE, Bergenstal RM, Buse JB, Diamant M, Ferrannini E, Nauck M, et al. Management of hyperglycemia in type 2 diabetes: a patient-centered approach: position statement of the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). Diabetes Care [Internet]. 2012 Jun [cited 2014 Jul 10];35(6):1364–79. Available from:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3357214&tool=pmcentrez&rendertype=abstract

190. Gillett M, Royle P, Snaith A, Scotland G, Poobalan A, Imamura M, et al. Non-pharmacological interventions to reduce the risk of diabetes in people with impaired glucose regulation: A systematic review and economic evaluation. Health Technol Assess [Internet]. N. Waugh, Department of Public Health, University of Aberdeen, Aberdeen, United Kingdom: National Co-ordinating Centre for HTA (Bouldrewood, Mail Point 728, Highfield, Southhampton, United Kingdom); 2012;16(33):1–235. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=22935084

191. Penn L, White M, Lindstrom J, den Boer AT, Blaak E, Eriksson JG, et al. Importance of Weight Loss Maintenance and Risk Prediction in the Prevention of Type 2 Diabetes: Analysis of European Diabetes Prevention Study RCT. PLoS One [Internet]. United States: Institute of Health and Society, Newcastle University, Newcastle upon Tyne, United Kingdom. linda.penn@ncl.ac.uk; 2013;8(2):e57143. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=23451166

192. Bakker LE, Sleddering MA, Schoones JW, Meinders AE, Jazet IM. Pathogenesis of type 2 diabetes in South Asians. Eur J Endocrinol [Internet]. 2013/08/14 ed. 2013;169(5):R99–114. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=23939919

193. Gujral UP, Pradeepa R, Weber MB, Narayan KM, Mohan V. Type 2 diabetes in South Asians: similarities and differences with white Caucasian and other populations. Ann N Y Acad Sci [Internet]. 2013/01/16 ed. 2013;1281:51–63. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=23317344

194. Tillin T, Hughes AD, Godsland IF, Whincup P, Forouhi NG, Welsh P, et al. Insulin resistance and truncal obesity as important determinants of the greater incidence of diabetes in Indian Asians and African Caribbeans compared with Europeans: the Southall And Brent REvisited (SABRE) cohort. Diabetes Care [Internet]. 2013 Feb [cited 2014 Nov 13];36(2):383–93. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3554271&tool=pmcentrez&rendertype=abstract

195. Hartland AJ, DeVille-Almond J, Orton M, Hartland V. Age of diagnosis of type 2 diabetes is more than 9 years younger in Asian-British compared with the rest of the UK population [Internet]. Diabetes. A.J. Hartland: American Diabetes Association Inc.; 2013.

p. A365. Available from: http://diabetes.diabetesjournals.org/content/62/Supplement_1/A365.full.pdf+html

196. Barnett a H, Dixon a N, Bellary S, Hanif MW, O'hare JP, Raymond NT, et al. Type 2 diabetes and cardiovascular risk in the UK south Asian community. Diabetologia [Internet]. 2006 Oct [cited 2014 Nov 15];49(10):2234–46. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16847701

197. Tillin. T, Hughes AD, Whincup P, Forouhi NG, Godsland IF, Sattar N, et al. Excess diabetes in ethnic minorities-incidence and predictors. Diabetologia [Internet]. 2011;Conference:47th Annual Meeting of the European Association fo. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed10&AN=70562372

198. Choi SE, Liu M, Palaniappan LP, Wang EJ, Wong ND. Gender and ethnic differences in the prevalence of type 2 diabetes among Asian subgroups in California. J Diabetes Complications [Internet]. 2013;27(5):429–35. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed11&AN=2013570399

199. Laws A, Jeppesen JL, Maheux PC, Schaaf P, Chen YD, Reaven GM. Resistance to insulin-stimulated glucose uptake and dyslipidemia in Asian Indians. Arterioscler Thromb Vasc Biol [Internet]. 1994 Jun 1 [cited 2014 Nov 14];14(6):917–22. Available from: http://atvb.ahajournals.org/content/14/6/917.short

200. P.M. M. Metabolic consequences of obesity and body fat pattern: lessons from migrant studies [Internet]. Ciba Foundation symposium. 1996. p. 54–193. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed4&NEWS=N&AN=9017274

201. Joshi SR. Metabolic syndrome--emerging clusters of the Indian phenotype. J Assoc Physicians India [Internet]. 2003 May [cited 2014 Nov 15];51:445–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12974423

202. Ramachandran A, Snehalatha C, Yamuna A, Murugesan N, Narayan KMV. Insulin resistance and clustering of cardiometabolic risk factors in urban teenagers in southern India. Diabetes Care [Internet]. 2007 Jul 1 [cited 2014 Nov 14];30(7):1828–33. Available from: http://care.diabetesjournals.org/content/30/7/1828.long

203. Misra A, Vikram NK, Sharma R, Basit A. High prevalence of obesity and associated risk factors in urban children in India and Pakistan highlights immediate need to initiate primary prevention program for diabetes and coronary heart disease in schools. Diabetes Res Clin Pract [Internet]. Ireland; 2006;71(1):101–2. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med5&NEWS=N&AN=16112243

204. McKeigue PM, Marmot MG, Adelstein AM, Hunt SP, Shipley MJ, Butler SM, et al. Diet and risk factors for coronary heart disease in Asians in northwest London. Lancet [Internet]. 1985 Nov 16 [cited 2014 Dec 22];2(8464):1086–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2865567

205. Banerji MA, Faridi N, Atluri R, Chaiken RL, Lebovitz HE. Body Composition, Visceral Fat, Leptin, and Insulin Resistance in Asian Indian Men1 [Internet]. Endocrine Society; 2013 [cited 2014 Dec 22]. Available from: http://press.endocrine.org/doi/abs/10.1210/jcem.84.1.5371?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub=pubmed&

206. World Health Organization, Organization WH. Public health Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Public Health [Internet]. 2004 Jan 10 [cited 2014 Jul 20];363(9403):157–63. Available from: http://www.thelancet.com/journals/a/article/PIIS0140-6736(03)15268-3/fulltext

207. Bhopal RS, Rafnsson SB. Could mitochondrial efficiency explain the susceptibility to adiposity, metabolic syndrome, diabetes and cardiovascular diseases in South Asian populations? Int J Epidemiol [Internet]. 2009 Aug [cited 2014 Nov 17];38(May):1072–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19423659

208. Hales CN, Barker DJ. Type 2 (non-insulin-dependent) diabetes mellitus: the thrifty phenotype hypothesis. Diabetologia [Internet]. 1992 Jul [cited 2014 Nov 17];35(7):595–601. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1644236

209. Wang Y, Rudd A., Wolfe CD. Ethnic differences in survival after stroke: Follow-up study of the south London stroke register [Internet]. Cerebrovascular Diseases. Y. Wang: S. Karger AG; 2012. p. 60. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed12&NEWS=N&AN=71636816

210. Ethnic aspects of emotional distress in patients with diabetes - the Amsterdam Health Monitor Study. Diabet Med [Internet]. 2013;30(1):e25–31. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed11&AN=2012733050

211. Abbotts J, Harding S, Cruickshank K. Cardiovascular risk profiles in UK-born Caribbeans and Irish living in England and Wales [Internet]. Atherosclerosis. Elsevier Ireland Ltd; 2004. p. 295–303. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed6&NEWS=N&AN=2004307910

212. George G. Understanding of Coronary Heart Disease in South Asian Migrant Men in the UK. 2010.

213. Rafnsson SB, Bhopal RS. Large-scale epidemiological data on cardiovascular diseases and diabetes in migrant and ethnic minority groups in Europe [Internet]. European Journal of Public Health. 2009. p. 484–91. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed9&NEWS=N&AN=2009517553

214. Smith C. International Migrants in England and Wales 2011 [Internet]. [cited 2015 Jan 7]. Available from: http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-for-local-authorities-in-england-and-wales/rpt-international-migrants.html#tab-Country-of-birth

215. Diabetes UK. Statistics: Diabetes Prevalence [Internet]. 2013. Available from: http://www.diabetes.org.uk/About_us/What-we-say/Statistics/Diabetes-prevalence-2012/

216. Health and Social Care Information Centre. National Diabetes Audit Report 1 : Care Processes and Treatment Targets. 2012;

217. ADDITION Leicester: Prevalence of impaired glucose regulation (IGR) and screen detected Type 2 diabetes (T2D) in a mixed ethnic UK population. Diabet Med [Internet]. Conference:Diabetes UK's Annual Professional Conference 2009 . Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed10&AN=70342102

218. Davis TMEME, T.M.E. D. Ethnic diversity in Type 2 diabetes. Diabet Med [Internet]. T. M. E. Davis, University of Western Australia, School of Medicine and Pharmacology, Fremantle Hospital, P.O. Box 480, Fremantle, WA 6959, Australia. E-mail: tdavis@cyllene.uwa.edu.au: Blackwell Publishing Ltd; 2008;25(SUPPL. 2):52–6. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed8&NEWS=N&AN=2008348339

219. National Paediatric Diabetes Audit Project Board. National Paediatric Diabetes Audit [Internet]. 2013. Available from: http://www.hqip.org.uk/assets/NCAPOP-Library/NCAPOP-2013-14/NPDA-report-2011-12-low-res-FINAL.pdf

220. Stone MA, Camosso-Stefinovic J, Wilkinson J, de Lusignan S, Hattersley AT, Khunti K. Incorrect and incomplete coding and classification of diabetes: A systematic review. Diabet Med [Internet]. 2010/06/12 ed. 2010;27(5):491–7. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20536944

221. Rollason W, Khunti K, de Lusignan S, Lusignan S De, Res M, Citp F. Variation in the recording of diabetes diagnostic data in primary care computer systems: implications for the quality of care. Inform Prim Care [Internet]. 2009/10/08 ed. 2009;17(2):113–9.

Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19807953

222. Hippisley-Cox J, O'Hanlon S. Identifying patients with diabetes in the QOF - two steps forward one step back. BMJ [Internet]. 2006 [cited 2014 Dec 31];333:672. Available from: http://www.bmj.com/rapid-response/2011/10/31/identifying-patients-diabetes-qof-two-steps-forward-one-step-back

223. UK Biobank. UK Biobank : Protocol for a large-scale prospective epidemiological resource UK Biobank Coordinating Centre Stockport. 2007.

224. Collins R. What makes UK Biobank special? Lancet [Internet]. Elsevier Ltd; 2012 Mar 31 [cited 2014 Aug 13];379(9822):1173–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22463865

225. Sudlow C. UK Biobank: Follow up and adjudication of health outcomes [Internet]. 2014 [cited 2014 Sep 11]. Available from: http://www.ukbiobank.ac.uk/wp-content/uploads/2014/06/1410-Cathie-Sudlow-follow-upoutcomes_Frontiers.pdf

226. UK Biobank. Data-Field 2443: Diabetes diagnosed by Doctor [Internet]. 2013 [cited 2014 Sep 11]. Available from: http://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=2443

227. Office for National Statistics. Revised European Standard Population 2013 [Internet]. 2013. Available from: http://www.ons.gov.uk/ons/guide-method/user-guidance/health-and-life-events/revised-european-standard-population-2013--2013-esp-/index.html#

228. S. S, M.J. D, S. M, S. de L, K. K. Prevalence and characteristics in coding, classification and diagnosis of diabetes in primary care [Internet]. Postgraduate Medical Journal. S. Seidu, Leicester Diabetes Centre, Leicester General Hospital, Gwendolen Road, Leicester LE5 4WP, United Kingdom. E-mail: scelee@doctors.org.uk: BMJ Publishing Group (Tavistock Square, London WC1H 9JR, United Kingdom); 2014. p. 13–7. Available from: http://pmj.bmj.com/content/90/1059/13.full.pdf+html

229. Iacobucci G, G. I. UK has fifth highest rate of type 1 diabetes in children, new figures show. BMJ [Internet]. G. Iacobucci; 2013;346:f22. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed11&NEWS=N&AN=23288146

230. Agyemang C, Kunst AE, Bhopal R, Anujuo K, Zaninotto P, Nazroo J, et al. Diabetes prevalence in populations of South Asian Indian and African origins: a comparison of England and the Netherlands. Epidemiology [Internet]. United States: Department of Public Health, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands. c.o.agyemang@amc.uva.nl; 2011 Jul [cited 2015 Jan 13];22(4):563–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21610499

231. Khunti K, Morris DH, Weston CL, Gray LJ, Webb DR, Davies MJ. Joint prevalence of diabetes, impaired glucose regulation, cardiovascular disease risk and chronic kidney disease in South Asians and White Europeans. PLoS One [Internet]. 2013 Jan [cited 2014 Jul 14];8(1):e55580. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3559442&tool=pmcentrez&rendertype=abstract

232. Holden SHE, Barnett AH, Peters JR, Jenkins-Jones S, Poole CD, Morgan CL, et al. The incidence of type 2 diabetes in the United Kingdom from 1991 to 2010. Diabetes Obes Metab [Internet]. C.J. Currie, School of Medicine, Cardiff University, Pharma Research Centre, Cardiff MediCentre, Cardiff CF14 4UJ, United Kingdom. E-mail: currie@cardiff.ac.uk: Blackwell Publishing Ltd (9600 Garsington Road, Oxford OX4 2XG, United Kingdom); 2013;15(9):844–52. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed11&NEWS=N&AN=2013478272

233. Dorsett R. Ethnic minorities in the inner city [Internet]. 1998. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=sopp&AN=256778-0000074131+

234. Greenhalgh PM. Diabetes in British south Asians: nature, nurture, and culture. Diabet Med [Internet]. ENGLAND: Joint Department of Primary Care and Population Sciences, University College London Medical School, UK.; 1997;14(1):10–8. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=9017348

235. The impact of ethnicity on survival in adults with diabetes: A prospective populationbased cohort study. Diabet Med [Internet]. 2013;Conference:Diabetes UK Professional Conference 2013 Mancheste. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed11&AN=71019153

236. Bhopal R, Unwin N, White M, Yallop J, Walker L, Alberti KGMM, et al. Heterogeneity of coronary heart disease risk factors in Indian, Pakistani, Bangladeshi, and European origin populations: cross sectional study. BMJ [Internet]. ENGLAND: Department of Epidemiology and Public Health, Medical School, University of Newcastle, Newcastle upon Tyne NE2 4HH.; 1999;319(7204):215–20. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=10417082

237. Bhopal R, Unwin N, White M, Yallop J, Walker L, Alberti KG, et al. Heterogeneity of coronary heart disease risk factors in Indian, Pakistani, Bangladeshi, and European origin populations: cross sectional study. BMJ. 1999;319:215–20.

238. Harris TJ, Cook DG, Wicks PD, Cappuccio FP. Impact of the new American Diabetes Association and World Health Organisation diagnostic criteria for diabetes on subjects from three ethnic groups living in the UK. Nutr Metab Cardiovasc Dis [Internet]. T.J. Harris, Department of General Practice and Primary Care, St. George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK.: Department of General Practice and Primary Care, St. George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, UK.; 2000;10(6):305–9. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed5&NEWS=N&AN=11302004

239. Decode Study Group. Will new diagnostic criteria for diabetes mellitus change European epidemiological data. 1998;

240. The expert committee on the diagnosis and classification of diabetes Mellitus. Follow-up Report on the Diagnosis of Diabetes Mellitus. Diabetes Care. 2003. p. 3160–7.

241. Bannister CA, Holden SE, Jenkins-Jones S, Morgan CL, Halcox JP, Schernthaner G, et al. Can people with type 2 diabetes live longer than those without? A comparison of mortality in people initiated with metformin or sulphonylurea monotherapy and matched, non-diabetic controls. Diabetes Obes Metab [Internet]. 2014 Nov [cited 2014 Dec 30];16(11):1165–73. Available from: http://www.readcube.com/articles/10.1111/dom.12354?r3_referer=wol&tracking_action=preview_click&show_checkout=1

242. Holman R. Metformin as first choice in oral diabetes treatment: the UKPDS experience. Journ Annu Diabetol Hotel Dieu [Internet]. 2008/07/10 ed. 2007;13–20. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=18613325

243. National Institute for Health and Care Excellence. The management of type 2 diabetes. 2014.

244. Hooper J, Chohan P, Caley M. Case detection of disease by NHS Health Checks in Warwickshire, England and comparison with predicted performance. Public Health [Internet]. The Royal Society for Public Health; 2014 May [cited 2014 Nov 28];128(5):475–7. Available from: http://dx.doi.org/10.1016/j.puhe.2014.01.013

245. NHS Employers. 2014/15 General medical services (GMS) contract quality and outcomes framework (QOF). 2014.

246. Choudhury S, Hussain S, Yao G, Hill J, Malik W, Taheri S. The impact of a diabetes local enhanced service on quality outcome framework diabetes outcomes. PLoS One [Internet]. 2013 Jan [cited 2014 Dec 19];8(12):8–14. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3866172&tool=pmcentrez&rendertype=abstract

247. Kontopantelis E, Reeves D, Valderas JM, Campbell S, Doran T. Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study. BMJ Qual Saf [Internet]. 2013;22(1):53–64. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medc&AN=22918988

248. Harris MI, Klein R, Welborn TA, Knuiman MW. Onset of NIDDM occurs at least 4-7 yr before clinical diagnosis. Diabetes Care [Internet]. 1992 Jul [cited 2014 Nov 28];15(7):815–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1516497

249. The NHS Information Centre. National Diabetes Audit 2012-2013 Report 1: Care Processes and Treatment Targets. 2013;

250. The NHS Information Centre. National Diabetes Audit Executive Summary 2009-2010. 2010.

251. Pierce MB, Zaninotto P, Steel N, Mindell J. Undiagnosed diabetes-data from the English longitudinal study of ageing. Diabet Med [Internet]. England: Department of Epidemiology and Public Health, University College London, London, UK. mary.pierce@nshd.mrc.ac.uk; 2009 Jul [cited 2014 Dec 15];26(7):679–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19573116

252. Holt T a., Stables D, Hippisley-Cox J, O'Hanlon S, Majeed A. Identifying undiagnosed diabetes: Cross-sectional survey of 3.6 million patients' electronic records. Br J Gen Pract [Internet]. 2008 Mar [cited 2014 Dec 15];58(March):192–6. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2249795&tool=pmcentrez&rendertype=abstract

253. NHLBI NIH. What Is Atherosclerosis? [Internet]. [cited 2014 Oct 10]. Available from: http://www.nhlbi.nih.gov/health/health-topics/topics/atherosclerosis/

254. Reddy KS. Cardiovascular diseases in the developing countries: dimensions, determinants, dynamics and directions for public health action. Public Health Nutr [Internet]. 2006 Dec 22 [cited 2014 Oct 11];5(1a):231–7. Available from: http://www.journals.cambridge.org/abstract_S1368980002000320

255. Enas EA. Lipoprotein(a) as a determinant of coronary heart disease in young women: a stronger risk factor than diabetes?. Circulation [Internet]. UNITED STATES; 1998;97(3):293–5. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=9462536

256. Nag T, Ghosh A. Cardiovascular disease risk factors in Asian Indian population: A systematic review. J Cardiovasc Dis Res [Internet]. Elsevier Ltd; 2013 Dec [cited 2014

Aug 27];4(4):222–8. Available from: http://www.sciencedirect.com/science/article/pii/S0975358314000059

257. World Health Organization. Global atlas on cardiovascular disease prevention and control [Internet]. World Health Organization; 2011 [cited 2014 Sep 18]. Available from: http://www.who.int/cardiovascular_diseases/publications/atlas_cvd/en/

258. Smith SC, Collins A, Ferrari R, Holmes DR, Logstrup S, McGhie DV, et al. Our time: a call to save preventable death from cardiovascular disease (heart disease and stroke). J Am Coll Cardiol [Internet]. Elsevier Inc.; 2012 Dec 4 [cited 2014 Oct 10];60(22):2343–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22995536

259. Nichols M, Townsend N, Scarborough P, Rayner M. Cardiovascular disease in Europe 2014: epidemiological update. Eur Heart J [Internet]. 2014 Aug 19 [cited 2014 Aug 21]; Available from: http://www.ncbi.nlm.nih.gov/pubmed/25139896

260. Proudfit WL. Origin of concept of ischaemic heart disease. Heart [Internet]. 1983 Sep 1;50(3):209–12. Available from: http://heart.bmj.com/cgi/doi/10.1136/hrt.50.3.209

261. Grimes DS. An epidemic of coronary heart disease. QJM [Internet]. 2012 Jun [cited 2014 Oct 10];105(6):509–18. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22262810

262. Stehbens WE. The Hypothetical Epidemic of Coronary Heart Disease and Atherosclerosis. 1995;449–54.

263. Oppenheimer GM. Profiling risk: the emergence of coronary heart disease epidemiology in the United States (1947-70). Int J Epidemiol [Internet]. 2006 Jun [cited 2014 Oct 11];35(3):720–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16481365

264. WHO MONICA Project. WHO MONICA Project: Objectives and Design. Int J Epidemiol [Internet]. 1989 Jan 1;18(Supplement 1):S29–37. Available from: http://www.ije.oxfordjournals.org/cgi/doi/10.1093/ije/18.Supplement_1.S29

265. Jones DS, Greene J a. The decline and rise of coronary heart disease: understanding public health catastrophism. Am J Public Health [Internet]. 2013 Jul [cited 2014 Sep 28];103(7):1207–18. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23678895

266. Appleby J. Does poor health justify NHS reform ? BMJ. 2011;2–6.

267. Tan S, Scott W, Panoulas V, Sehmi J. Review article Coronary heart disease in Indian Asians. Glob Cardiol Sci Pract. 2014;

268. Ramaraj R, Chellappa P. Cardiovascular risk in South Asians. Postgrad Med J. 2008;84:518–23.

269. McKeigue PM, Ferrie JE, Pierpoint T, Marmot MG. Association of early-onset coronary heart disease in South Asian men with glucose intolerance and hyperinsulinemia. Circulation. 1993;87:152–61.

270. Khafaji HAH, Suwaidi JM Al. Atypical presentation of acute and chronic coronary artery disease in diabetics. World J Cardiol [Internet]. 2014 Aug 26 [cited 2014 Oct 9];6(8):802–13. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4163709&tool=pmcentrez&rendertype=abstract

271. Haffner SM, Lehto S, Rönnemaa T, Pyörälä K, Laakso M, From M, et al. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. N Engl J Med [Internet]. UNITED STATES: Department of Medicine, University of Texas Health Science Center at San Antonio, 78284-7873, USA.; 1998 Jul 23 [cited 2014 Aug 7];339(4):229–34. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-0032560807&partnerID=tZOtx3y1

272. Grundy SM, Cleeman JI, Merz CNB, Brewer HB, Clark LT, Hunninghake DB, et al. Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. Circulation [Internet]. 2004 Jul 13 [cited 2014 Jul 17];110(2):227–39. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15249516

273. Yusuf PS, Hawken S, Ôunpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): Case-control study. Lancet. 2004;364:937–52.

274. Anand SS, Yusuf S, Vuksan V, Devanesen S, Teo KK, Montague P a, et al. Differences in risk factors, atherosclerosis and cardiovascular disease between ethnic groups in Canada: the study of health assessment and risk in ethnic groups (SHARE). Indian Heart J [Internet]. 2000;52:S35–43. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11339439

275. Gupta R, Joshi P, Mohan V, Reddy KS, Yusuf S. Epidemiology and causation of coronary heart disease and stroke in India. Heart [Internet]. 2008 Jan [cited 2014 Oct 12];94(1):16–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18083949

276. Zaman MJS, Philipson P, Chen R, Farag A, Shipley M, Marmot MG, et al. South Asians and coronary disease: Is there discordance between effects on incidence and prognosis? Heart [Internet]. M.J. S. Zaman, Department of Epidemiology and Public Health, University College London, 1-19 Torrington Place, London WC1E 6BT, United Kingdom. E-mail: justin.zaman@nhs.net: BMJ Publishing Group (Tavistock Square, London WC1H 9JR, United Kingdom); 2013;99(10):729–36. Available from:

http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=23406688

277. Yusuf S, Reddy S, Ounpuu S, Anand S. Global burden of cardiovascular diseases: Part II: variations in cardiovascular disease by specific ethnic groups and geographic regions and prevention strategies. Circulation. 2001;104:2855–64.

278. Tan S-T, Scott W, Panoulas V, Sehmi J, Zhang W, Scott J, et al. Coronary heart disease in Indian Asians. Glob Cardiol Sci Pract [Internet]. S.-T. Tan, National Heart and Lung Institute, Imperial College London, Hammersmith Hospital, Du Cane Road, London W11 0HS, United Kingdom. E-mail: sian-tsung.tan@imperial.ac.uk: Bloomsbury Qatar Foundation Journals; 2014 Jan [cited 2014 Aug 27];2014(1):13–23. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4104373&tool=pmcentrez&rendertype=abstract

279. Tillin T, Hughes AD, Godsland IF, Whincup P, Forouhi NG, Welsh P, et al. Insulin Resistance and Truncal Obesity as Important Determinants of the Greater Incidence of Diabetes in Indian Asians and African Caribbeans Compared With Europeans: The Southall and Brent Revisited (SABRE) cohort. Diabetes Care [Internet]. 2012/09/12 ed. 2012; Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=22966089

280. Bansal N, Fischbacher CM, Bhopal RS, Brown H, Steiner MF, Capewell S. Myocardial infarction incidence and survival by ethnic group: Scottish Health and Ethnicity Linkage retrospective cohort study. BMJ Open [Internet]. 2013 Jan [cited 2014 Jul 26];3(9):e003415. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed11&AN=2013640647

281. Wild SH, Fischbacher C, Brock A, Griffiths C, Bhopal R. Mortality from all causes and circulatory disease by country of birth in England and Wales 2001-2003. J Public Health (Bangkok). 2007;29:191–8.

282. Tillin T, Hughes AD, Mayet J, Whincup P, Sattar N, Forouhi NG, et al. The relationship between metabolic risk factors and incident cardiovascular disease in Europeans, South Asians, and African Caribbeans: SABRE (Southall and Brent Revisited) -- a prospective population-based study. J Am Coll Cardiol [Internet]. United States: International Centre for Circulatory Health, National Heart and Lung Institute, Imperial College London, London, United Kingdom. t.tillin@imperial.ac.uk; 2013 Apr 30 [cited 2014 Aug 6];61(17):1777–86. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=23500273

283. Reddy KS. Cardiovascular Disease in Non-Western Countries. N Engl J Med. 2004;350(24):2438–40.

284. Ghosh a. Comparison of anthropometric, metabolic and dietary fatty acids profiles in lean and obese dyslipidaemic Asian Indian male subjects. Eur J Clin Nutr [Internet]. 2007 Mar [cited 2014 Oct 12];61(3):412–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17006446

285. Harding S, Rosato M, Teyhan A. Trends for coronary heart disease and stroke mortality among migrants in England and Wales, 1979-2003: slow declines notable for some groups. Heart. 2008;94:463–70.

286. Williams ED, Stamatakis E, Chandola T, Hamer M. Physical activity behaviour and coronary heart disease mortality among South Asian people in the UK: an observational longitudinal study. Heart. 2011;97:655–9.

287. Bhopal R, Bansal N, Fischbacher C, Brown H, Capewell S. Ethnic variations in chest pain and angina in men and women: Scottish Ethnicity and Health Linkage Study of 4.65 million people. Eur J Prev Cardiol [Internet]. 2012;19:1250–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21975324

288. Zaninotto P, Mindell J, Hirani V. Prevalence of cardiovascular risk factors among ethnic groups: Results from the Health Surveys for England. Atherosclerosis. 2007;195.

289. World Health Organization. The global burden of disease. 2004.

290. Anand SS, Islam S, Rosengren A, Franzosi MG, Steyn K, Yusufali AH, et al. Risk factors for myocardial infarction in women and men: insights from the INTERHEART study. Eur Heart J [Internet]. 2008 Apr [cited 2014 Sep 24];29(7):932–40. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18334475

291. Tan YY, Gast G-CM, van der Schouw YT. Gender differences in risk factors for coronary heart disease. Maturitas [Internet]. 2010 Feb [cited 2014 Sep 17];65(2):149–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19897327

292. Jousilahti P, Vartiainen E, Tuomilehto J, Puska P. Sex, Age, Cardiovascular Risk Factors, and Coronary Heart Disease : A Prospective Follow-Up Study of 14 786 Middle-Aged Men and Women in Finland. Circulation [Internet]. 1999 Mar 9 [cited 2014 Oct 11];99(9):1165–72. Available from: http://circ.ahajournals.org/cgi/doi/10.1161/01.CIR.99.9.1165

293. Lee WL, Cheung AM, Cape D, Zinman B. Impact of Diabetes on Coronary Artery Disease in Women and Men: A meta-analysis of prospective studies. 2000;(January).

294. Barrett-Connor E. Why Women Have Less Heart Disease Than Men and How Diabetes Modifies Women's Usual Cardiac Protection: A 40-Year Rancho Bernardo Cohort Study. Glob Heart [Internet]. World Heart Federation (Geneva); 2013 Jun 1 [cited 2014 Sep 2];8(2):95–104. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3810980&tool=pmcentrez&rendertype=abstract

295. Lewis JD, Bilker WB, Weinstein RB, Strom BL. The relationship between time since registration and measured incidence rates in the General Practice Research Database. Pharmacoepidemiol Drug Saf [Internet]. 2005/05/18 ed. 2005;14(7):443–51. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15898131

296. Mattace-Raso FUS, van der Cammen TJM, Hofman A, van Popele NM, Bos ML, Schalekamp MADH, et al. Arterial stiffness and risk of coronary heart disease and stroke: the Rotterdam Study. Circulation [Internet]. 2006 Feb 7 [cited 2014 Jul 11];113(5):657–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16461838

297. World Health Organization. WHO :: Global Database on Body Mass Index [Internet]. [cited 2014 Dec 15]. Available from: http://apps.who.int/bmi/index.jsp?introPage=intro_3.html

298. National Institute for Health and Care Excellence. Type 2 diabetes | Guidance and guidelines | NICE. NICE; [cited 2014 Dec 15]; Available from: http://www.nice.org.uk/guidance/cg87

299. National Institute for Health and Care Excellence. Clinical management of primary hypertension in adults. 2011.

300. CALIBER. CALIBER data portal: Health Behaviour [Internet]. 2014 [cited 2015 Jan 7]. Available from: https://www.caliberresearch.org/portal/chapter/2#Alcohol

301. StataCorp. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP; 2013.

302. Wannamethee SG, Shaper AG, Lennon L, Morris RW. Metabolic syndrome vs Framingham Risk Score for prediction of coronary heart disease, stroke, and type 2 diabetes mellitus. Arch Intern Med [Internet]. United States: Department of Primary Care and Population Sciences, Royal Free and University College Medical School, Rowland Hill Street, London NW3 2PF, England. goya@pcps.ucl.ac.uk; 2005;165(22):2644–50. Available from: http://archinte.jamanetwork.com/article.aspx?articleid=766882

303. Cameron AC, Miller DL. A Practitioner ' s Guide to Cluster-Robust Inference. 2013;

304. Rubin DB. Inference and missing data. Biometrika [Internet]. 1976 Dec 1 [cited 2014 Nov 17];63(3):581–92. Available from: http://biomet.oxfordjournals.org/content/63/3/581.short?rss=1&ssource=mfc

305. Roth MA, Aitsi-Selmi A, Wardle H, Mindell J. Under-reporting of tobacco use among Bangladeshi women in England. J Public Health (Oxf) [Internet]. 2009 Sep [cited 2014 Dec 18];31(3):326–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19561042

306. Saito I. Epidemiological Evidence of Type 2 Diabetes Mellitus, Metabolic Syndrome, and Cardiovascular Disease in Japan. Circ J [Internet]. 2012 [cited 2014 Oct 9];76(5):1066–73. Available from: http://joi.jlc.jst.go.jp/JST.JSTAGE/circj/CJ-11-1519?from=CrossRef

307. Carson AP, Tanner RM, Yun H, Glasser SP, Woolley JM, Thacker EL, et al. Declines in coronary heart disease incidence and mortality among middle-aged adults with and without diabetes. Ann Epidemiol [Internet]. 2014 Aug;24(8):581–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24970491

308. Davis TME, Coleman RL, Holman RR. Ethnicity and long-term vascular outcomes in Type 2 diabetes: A prospective observational study (UKPDS 83) [Internet]. Diabetic Medicine. T.M.E. Davis, School of Medicine and Pharmacology, University of Western Australia, Fremantle Hospital, Fremantle, WA, Australia. E-mail: tim.davis@uwa.edu.au: Blackwell Publishing Ltd (9600 Garsington Road, Oxford OX4 2XG, United Kingdom); 2014. p. 200–7. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed11&NEWS=N&AN=2014048534

309. Shah AD, Langenberg C, Rapsomaniki E, Denaxas S, Pujades-rodriguez M, Gale CP, et al. Type 2 diabetes and incidence of cardiovascular diseases : a cohort study in 1 · 9 million people. 2014;8587(14):1–9.

310. Bhopal R, Hayes L, White M, Unwin N, Harland J, Ayis S, et al. Ethnic and socio-economic inequalities in coronary heart disease, diabetes and risk factors in Europeans and South Asians. J Public Health Med [Internet]. England: Oxford University Press; 2002;24(2):95–105. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=med4&NEWS=N&AN=12141592

311. Brindle P, May M, Gill P, Cappuccio F, D'Agostino Sr. R, Fischbacher C, et al. Primary prevention of cardiovascular disease: a web-based risk score for seven British black and minority ethnic groups. Heart [Internet]. 2006/06/10 ed. 2006;92(11):1595–602. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16762981

312. Bhopal R, Fischbacher C, Vartiainen E, Unwin N, White M, Alberti G. Predicted and observed cardiovascular disease in South Asians: Application of FINRISK, Framingham and SCORE models to Newcastle Heart Project data [Internet]. Journal of Public Health. R. Bhopal, Section of Public Health Sciences, Medical School, Teviot Place, Edinburgh EH8 9AG, United Kingdom. E-mail: Raj.Bhopal@ed.ac.uk: Oxford University Press; 2005. p. 93–100. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=2005155080

313. Wild S, McKeigue P. Cross sectional analysis of mortality by country of birth in England and Wales, 1970-92. BMJ [Internet]. 1997 Mar 8 [cited 2014 Dec 22];314(7082):705–10. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2126166&tool=pmcentrez&rendertype=abstract

314. Pearson K, Haigh D, Williams J. To improve the quality of clinical care and self-care support, to South Asians with diabetes and at risk of diabetes, living in a deprived area of Huddersfield North within Kirklees, West Yorkshire [Internet]. Diabetic Medicine. K. Pearson, Public Health, NHS Kirklees, Huddersfield, United Kingdom: Blackwell Publishing Ltd; 2010. p. 138. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed9&NEWS=N&AN=70087772

315. Alam R, Speed S, Beaver K. A scoping review on the experiences and preferences in accessing diabetes-related healthcare information and services by British Bangladeshis. Heal Soc Care Community [Internet]. England: Blackwell Publishing Ltd (9600 Garsington Road, Oxford OX4 2XG, United Kingdom); 2012;20(2):155–71. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed10&NEWS=N&AN=21883609

316. Seale C, Rivas C, Kelly M. The challenge of communication in interpreted consultations in diabetes care: a mixed methods study. Br J Gen Pract [Internet]. England: Department of Sociology and Communications, Brunel University, Uxbridge, UK. clive.seale@brunel.ac.uk; 2013;63(607):e125–33. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=23561691

317. Varughese GII, Tomson J, Lip GYH, G.I. V, J. T, G.Y. L. Type 2 diabetes mellitus: A cardiovascular perspective. Int J Clin Pract [Internet]. G.Y.H. Lip, Haemostasis Thrombosis and Vascular Biology Unit, University Department of Medicine, City Hospital, Birmingham B18 7QH, United Kingdom. E-mail: g.y.h.lip@bham.ac.uk: Blackwell Publishing Ltd; 2005;59(7):798–816. Available from:

http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=2006289845

318. Somers EC, Thomas SL, Smeeth L, Schoonen WM, Hall AJ. Incidence of systemic lupus erythematosus in the United Kingdom, 1990-1999. Arthritis Rheum [Internet]. 2007/05/02 ed. 2007;57(4):612–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17471530

319. Saunders CL, Abel G a, El Turabi a, Ahmed F, Lyratzopoulos G. Accuracy of routinely recorded ethnic group information compared with self-reported ethnicity: evidence from the English Cancer Patient Experience survey. BMJ Open [Internet]. 2013 Jan [cited 2014 Dec 31];3(6):e002882 – . Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3696860&tool=pmcentrez&rendertype=abstract

320. Campbell J, Dedman DJ, Eaton SC, Gallagher AM, Williams TJ. Is the CPRD GOLD Population Comparable to the UK Population? Pharmacoepidemiol Drug Saf [Internet]. 2013;22:280–1. Available from: <Go to ISI>://WOS:000334876101135

321. De Lusignan S, Khunti K, Belsey J, Hattersley A, van Vlymen J, Gallagher H, 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 [Internet]. 2010/06/16 ed. 2010;27(2):203–9. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=20546265

322. Winkley K, Thomas SM, Sivaprasad S, Chamley M, Stahl D, Ismail K, et al. The clinical characteristics at diagnosis of type 2 diabetes in a multi-ethnic population: the South London Diabetes cohort (SOUL-D). Diabetologia [Internet]. 2013/03/16 ed. 2013;56(6):1272–81. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=23494447

323. NHS Employers. Revisions to the GMS contract. 2006.

324. Herrett EL, Thomas SL, Smeeth L. Validity of diagnoses in the general practice research database. Br J Gen Pract [Internet]. 2011/07/05 ed. 2011;61(588):438–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21722461

325. Booth HP, Prevost AT, Gulliford MC. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. Pharmacoepidemiol Drug Saf [Internet]. 2013;22(12):1357–61. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed11&AN=2013754442

326. Ellison GTH, Smart A, Tutton R, Outram SM, Ashcroft R, Martin P. Racial categories in medicine: a failure of evidence-based practice? PLoS Med / Public Libr Sci [Internet]. 2007;4(9):e287. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=17896860

327. Sheldon TA, Parker H, T.A. S, H. P. Race and ethnicity in health research. J Public Health Med [Internet]. 1992/06/01 ed. 1992;14(2):104–10. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed2&AN=1992197604

328. Braun L, Fausto-Sterling A, Fullwiley D, Hammonds EM, Nelson A, Quivers W, et al. Racial categories in medical practice: how useful are they? PLoS Med / Public Libr Sci [Internet]. 2007;4(9):e271. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=17896853

329. Crews DE, Bindon JR. Ethnicity as a taxonomic tool in biomedical and biosocial research. Ethn Dis [Internet]. 1991/01/01 ed. 1991 Jan 1 [cited 2015 Jan 13];1(1):42–9. Available from: http://www.researchgate.net/publication/21312504_Ethnicity_as_a_taxonomic_tool_in_biomedical_and_biosocial_research

330. Katikireddi SV, Valles S a. Coupled Ethical-Epistemic Analysis of Public Health Research and Practice: Categorizing Variables to Improve Population Health and Equity. Am J Public Health [Internet]. 2015 Jan [cited 2015 Jan 13];105(1):e36–42. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4265925&tool=pmcentrez&rendertype=abstract

331. Macfarlane GJ, Lunt M, Palmer B, Afzal C, Silman AJ, Esmail A. Determining aspects of ethnicity amongst persons of South Asian origin: The use of a surname-classification programme (Nam Pehchan). Public Health [Internet]. 2007 Mar 1 [cited 2015 Jan 13];121(3):231–6. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed8&AN=2007082043

332. Nijjar APK, Wang H, Quan H, Khan NA. Ethnic and sex differences in the incidence of hospitalized acute myocardial infarction: British Columbia, Canada 1995-2002. BMC Cardiovasc Disord [Internet]. 2010;10:38. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=20723259

333. Maizlish NA, Herrera L. A record linkage protocol for a diabetes registry at ethnically diverse community health centers. J Am Med Informatics Assoc [Internet]. 12(3):331–7.

Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med4&AN=15684130

334. Harron KL, McKinney PA, Feltbower RG, Bodansky HJ, Norman PD, Campbell FM, et al. Incidence rate trends in childhood type 1 diabetes in Yorkshire, UK 1978-2007: effects of deprivation and age at diagnosis in the South Asian and non-South Asian populations. Diabet Med [Internet]. England: Blackwell Publishing Ltd (9600 Garsington Road, Oxford OX4 2XG, United Kingdom); 2011;28(12):1508–13. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=medl&NEWS=N&AN=21838766

335. Harron KL, McKinney PA, Feltbower RG, Stephenson C, Bodansky HJ, Norman P, et al. Ethnic differences in incidence rates of childhood type 1 diabetes in Yorkshire 1978-2007. Diabetologia [Internet]. 2010;Conference:46th Annual Meeting of the European Association fo. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed9&AN=70262381

336. Prevalence and characteristics of multiple sclerosis in a multi-ethnic population in Bradford, United Kingdom. Mult Scler [Internet]. Conference:25th Congress of the European Committee for Treatm. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=emed9&AN=70100347

337. Chen W, Petitti DB, Enger S. Limitations and potential uses of census-based data on ethnicity in a diverse community. Ann Epidemiol [Internet]. 2004/06/05 ed. 2004;14(5):339–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15177273

338. Bhopal R, Sheikh A. The Scottish Health and Ethnicity Linkage Study shows the value of national datasets and that data can be securely linked. BMJ [Internet]. 2014 Jan [cited 2015 Jan 7];348(April):g2389. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24699120

339. NHS Employers. Contract changes 2011/12 [Internet]. [cited 2014 Dec 18]. Available from: http://www.nhsemployers.org/your-workforce/primary-care-contacts/general-medical-services/gms-contract-changes/contract-changes-201112

340. Clark SD, Birkin M, Heppenstall A. Sub regional estimates of morbidities in the English elderly population. Health Place [Internet]. Elsevier; 2014 May [cited 2014 Dec 18];27:176–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24631924

341. National Institute for Health and Care Excellence. Preventing type 2 diabetes : risk identification and interventions for individuals at high risk. 2012;(July).

342. National Institute for Health and Care Excellence. Chronic obstructive pulmonary disease. 2010;(April 2007).

343. National Institute for Health and Care Excellence. Prevention of cardiovascular disease. 2010;(June).

344. Van Staa T-P, Dyson L, McCann G, Padmanabhan S, Belatri R, Goldacre B, et al. The opportunities and challenges of pragmatic point-of-care randomised trials using routinely collected electronic records: evaluations of two exemplar trials. Health Technol Assess [Internet]. 2014 Jul [cited 2014 Dec 17];18(43):1–146. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25011568

345. Swinglehurst D, Greenhalgh T, Roberts C. Computer templates in chronic disease management: ethnographic case study in general practice. BMJ Open [Internet]. 2012 Jan [cited 2014 Dec 18];2(6):1–16. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3533031&tool=pmcentrez&rendertype=abstract

346. Denaxas SC, George J, Herrett E, Shah AD, Kalra D, Hingorani AD, et al. Data resource profile: cardiovascular disease research using linked bespoke studies and electronic health records (CALIBER). Int J Epidemiol [Internet]. 2012;41(6):1625–38. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=medl&AN=23220717

347. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. BMJ [Internet]. 2007/07/07 ed. 2007;335(7611):136. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17615182

348. Savill P. QDScore: a risk prediction tool for diabetes. Practitioner [Internet]. 2009/05/26 ed. 2009;253(1717):5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19462740

349. Lip GYH. Can we predict stroke in atrial fibrillation? Clin Cardiol [Internet]. 2012 Jan [cited 2014 Dec 19];35:21–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22246948

350. Pisters R, Lane D a, Nieuwlaat R, de Vos CB, Crijns HJGM, Lip GYH. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. Chest [Internet]. 2010 Nov [cited 2014 Oct 17];138(5):1093–100. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20299623

351. Billings J, Blunt I, Steventon A, Georghiou T, Lewis G, Bardsley M. Development of a predictive model to identify inpatients at risk of re-admission within 30 days of discharge (PARR-30). BMJ Open [Internet]. 2012;2(4). Available from: <Go to ISI>://MEDLINE:22885591

352. Witte DR, Shipley MJ, Marmot MG, Brunner EJ, D.R. W, M.J. S, et al. Performance of existing risk scores in screening for undiagnosed diabetes: An external validation study. Diabet Med [Internet]. D. R. Witte, Steno Diabetes Center, Niels Steensens Vej 1, 2820 Gentofte, Denmark. E-mail: DRW@steno.dk: Blackwell Publishing Ltd (9600 Garsington Road, Oxford OX4 2XG, United Kingdom); 2010 Jan [cited 2014 Dec 15];27(1):46–53. Available from: http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed9&NEWS=N&AN=2010035808