

# **Ethnicity and health in Scotland: can we fill the information gap?**

## **A demonstration project focusing on coronary heart disease and linkage of census and health records**

BHOPAL R<sup>1</sup>, FISCHBACHER CM<sup>1,2</sup>, STEINER M<sup>1</sup>, CHALMERS J<sup>1,2</sup>, POVEY C<sup>2</sup>, JAMIESON J<sup>2,3</sup>, KNOWLES D<sup>2</sup>

(<sup>1</sup>Public Health Sciences, University of Edinburgh, <sup>2</sup>Information and Statistics Division, National Services Scotland, <sup>3</sup>National Resource Centre for Ethnic Minorities, NHS Health Scotland)

*Address for correspondence: Prof R Bhopal, Public Health Sciences, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG.*

*Project done in collaboration with the General Register Office for Scotland*



## Contents page

<b>Acknowledgements</b> .....	<b>4</b>
<b>Abbreviations</b> .....	<b>5</b>
<b>A note on terminology relating to ethnicity.</b> .....	<b>6</b>
<b>Overall Summary</b> .....	<b>8</b>
<b>1 Introduction: ethnic variations in health, and the need for data in Scotland</b> .....	<b>10</b>
<b>1 Introduction: ethnic variations in health, and the need for data in Scotland</b> .....	<b>10</b>
1.1 Background .....	10
1.2 Why include ethnicity in routine health data collection? .....	11
1.3 Health inequalities: policy and legal framework for improving information.....	11
1.4 Purpose of this project and collaboration .....	11
1.5 Focus on coronary heart disease in this demonstration project.....	12
1.6 Aims and objectives .....	13
1.7 Outline of methods and data sources .....	13
1.8 Legal and ethical issues.....	14
1.9 Publicising the project .....	14
1.10 Context of this project in relation to national initiatives, and prospects for the future in the light of ethnic monitoring .....	15
<b>2 Current availability of information on ethnicity and health in Scotland</b> .....	<b>16</b>
2.1 Introduction .....	16
2.2 Census .....	16
2.3 Primary care.....	16
2.4 Scottish Morbidity Record and Scottish Cancer Registration systems .....	17
2.5 Child Health Information.....	18
2.6 Health Visitor/District Nurse databases (CIS) in Lothian.....	18
2.7 Patient Information Management System (PiMS) .....	19
2.8 Genitourinary medicine clinics .....	19
2.9 Drug misuse information .....	19
2.10 Diabetes core dataset standard .....	19
2.11 National work on data standards.....	19
2.12 Surveys .....	20
2.13 Conclusion .....	21
<b>3 The validity and value of name search algorithms for assessing ethnicity     applied to names from the Scottish population: analysis of SMR-01 (mortality)     and DARTS databases</b> .....	<b>24</b>
3.1 Background .....	25
3.2 Methods .....	25
3.3 Ethics .....	27
3.4 Results .....	27
3.5 Discussion.....	35
<b>4 Imputation and model estimates</b> .....	<b>38</b>
4.1 Background .....	38
4.2 Methods .....	39
4.3 Results .....	40
4.4 Discussion.....	43
<b>Overall, the results encourage us to develop further these modelling approaches     pending Scotland developing and undertaking a health survey of ethnic     minority groups.</b> .....	<b>43</b>
<b>5 All cause and cardiovascular mortality by country of birth</b> .....	<b>44</b>
5.1 Background .....	45
5.2 Methods .....	45
5.3 Ethics .....	46

5.4 Results .....	46
5.5 Conclusions.....	56
<b>6 Adding ethnicity codes to routine health data by record linkage to census records: surmounting the confidentiality challenge .....</b>	<b>63</b>
6.1 Introduction .....	64
6.2 Methods .....	64
6.3 Ethics .....	67
6.4 Results .....	69
6.5 Discussion.....	70
<b>7 Cohort analysis of CHD using the linked database.....</b>	<b>71</b>
7.1 Introduction .....	72
7.2 Methods .....	72
7.3 Results .....	73
7.4 Conclusions.....	83
<b>8 Conclusions and recommendations .....</b>	<b>88</b>
8.1 Main findings .....	88
8.2 Information gaps filled and left .....	89
8.3 Next steps, lessons learned and recommendations .....	90
8.4 Letting the information desert bloom.....	91
<b>References .....</b>	<b>93</b>

## **Acknowledgements**

We thank the Scottish Executive for funding this study and the two organisations that assisted the Executive in managing the grant and the work – the Public Health Institute for Scotland (now part of NHS Scotland), and Information Services (Scotland). Numerous people offered their expertise, guidance and support (too many to list comprehensively).

David Orr of GROS was intimately involved in the linkage study from an early stage and served in the Steering Committee until 2004 when his place was taken by Peter Scrimgeour and Ganka Mueller. Jo Fuchs, Ian Maté and Susan Wallace of GROS made many contributions to the linkage component, and Duncan Macniven of GROS gave helpful advice, particularly on dissemination of the findings. Professor Phil Hanlon acted enthusiastically and promptly to help make this study possible when it was first mooted in 2002. Dr Rafik Gardee ensured the study received the full backing of the National Resource Centre for Ethnic Minority Health. Mr Hector Mackenzie of Scottish Executive, Dr Mac Armstrong (CMO), and Chris Oswald (CRE) also played influential roles in maintaining the directions of the work. Prof Andrew Morris, Richie McAlpine and Philip Thompson provided access to and much assistance in analysis of the Dundee diabetes register (DARTS).

Dr Kiran Nanchahal, Isabel Silva and Punam Mangtani at the London School of Hygiene gave their help in relationship to the name search programme SANGRA. Kate McIntyre of the University of Glasgow and Niamh Murphy of the Western Infirmary Glasgow advised on the analysis of linked data. Dr Sarah Wild offered detailed comments on the penultimate draft.

Dr Paramjit Gill gave permission to reproduce the appendix in Chapter 2.

The name searchers who checked the name search codes were: Rekha Gupta, Amjad Chaudhry and Fulmaya Lama.

The project secretary was Jenny Hay and subsequently Tori Hastie.

Finally, we thank our employing organisations for their support of our work in this project.

### **Contributions and responsibility**

The authors of the report and individuals named in each chapter are responsible for the contents. Neither the funders, nor those named in the acknowledgements or associated with the individuals named, are responsible for any remaining errors in this work.

## Abbreviations

AMI	Acute myocardial infarction
CHD	Coronary heart disease
CIS	Continuous inpatient stays
COPPISH	Core patient profiles in Scottish Hospitals
DARTS	Diabetes audit and research Tayside Scotland
DM	Diabetes mellitus (type 2)
GPASS	General practice administration system for Scotland
GROS	General Register Office for Scotland
HbA1c	Glycosylated haemoglobin
ISD	Information Services, National Services Scotland
MI	Myocardial infarction
MREC	Multi-centre Research Ethics Committee
Non-SA	Non-South Asian
NP2	Nam Pehchan version 2
NRCEMH	National Resource Centre for Ethnic Minority Health
PAC	Privacy Advisory Committee
PEPI	Post-edit pre-imputation
PTI	Practice Team Information
SA	South Asian
SIMD	Scottish index of multiple deprivation
SMR	Standardised mortality/morbidity ratio
SMR01	Scottish morbidity record linked to mortality
Type 2 DM	Type 2 diabetes mellitus

## **A note on terminology relating to ethnicity.**

There is no consensus on appropriate terms for the scientific study of health by ethnicity, and published guidelines are yet to be widely adopted. We have followed general conventions used in the UK and, whenever appropriate, the terminology used by the original authors. For example, in the UK the term ethnic minority group usually refers to minority populations of non-European origin and characterised by their non-White status. (we use it this way here). The term South Asian refers to populations originating from the Indian Sub-continent, effectively, India, Pakistan, Bangladesh and Sri Lanka. White is the term currently used to describe people with European ancestral origins. By ethnicity we mean the group a person belongs to as a result of a mix of cultural factors including language, diet, religion, and ancestry.

Bhopal (2004) has published a glossary (Glossary of terms relating to ethnicity and race: for reflection and debate, J Epi Comm Health) from which we extract the key points for some relevant terms as follows:

### **African**

A person with African ancestral origins who self-identifies, or is identified, as African, but excluding those of other ancestry e.g. European and South Asian. This term is the currently preferred description for more specific categories, as in African American, for example. In practice, Northern Africans from Algeria, Morocco and such countries are excluded from this category. (See also Black.)

### **Afro-Caribbean/African Caribbean**

A person of African ancestral origins whose family settled in the Caribbean before emigrating and who self-identifies, or is identified, as Afro-Caribbean. (See also Black.)

### **Bangladeshi**

A person whose ancestry lies in the Indian subcontinent who self-identifies, or is identified, as Bangladeshi. (see also South Asian.) Between 1947 and 1971 the land known as Bangladesh was East Pakistan and before that India.

### **Chinese**

A person with ancestral origins in China, who self-identifies, or is identified, as Chinese.

### **Ethnic minority group**

Usually, but not always, this phrase is used to refer to a non-White population. Alternatively, it may be used to describe a specific identifiable group e.g. gypsy travellers, and less commonly, Irish in the UK. Some people consider the phrase inaccurate and prefer minority ethnic group, but the two phrases are used synonymously.

### **Ethnicity**

The social group a person belongs to, and either identifies with or is identified with by others, as a result of a mix of cultural and other factors including language, diet, religion, ancestry, and physical features traditionally associated with race. Increasingly, the concept is being used synonymously with race but the trend is pragmatic rather than scientific.

**Indian**

A person whose ancestry lies in the Indian sub-continent who identifies, or is identified, as Indian (see, South Asian). (Major changes to India's geographical boundaries took place in 1947 when Pakistan was created.)

**Irish**

A person whose ancestry lies in Ireland who self-identifies as Irish but generally restricted to the White population (see, White).

**Minority ethnic group**

See ethnic minority group. Increasingly used as the preferred phrase and replacing ethnic minority group.

**Non-Asian/Non-Chinese etc.**

This type of term is rarely defined but self-evidently implies those not belonging to the group under study. This degree of non-specificity is not usually recommended, but is used for the purposes of this report, which serves as a demonstration project.

**Pakistani**

A person whose ancestry lies in the Indian subcontinent who identifies, or is identified, as Pakistani (see South Asian). Some Pakistanis may have birth or ancestral roots in the current territory of India but identify with Pakistan, a country created in 1947.

**Reference/control/comparison**

This refers to the standard against which a population that is being studied can be compared to permit an analysis of similarities and differences. The concept is fundamental to epidemiology, and this terminology is preferable to non-specific ethnic or racial terms such as non-Asian, or general or even White population. In this report, however, we use the term non-South Asian.

**South Asian**

A person whose ancestry is in the countries of the Indian sub-continent, including India, Pakistan, Bangladesh and Sri Lanka. This label is usually assigned, for individuals rarely identify with it. (See also Indian, Indian Asian, Asian, Pakistani, Bangladeshi).

**White**

The term usually used to describe people with European ancestral origins who identify, or are identified, as White (sometimes called European, or in terms of racial classifications, the group known as Caucasian or Caucasoid). The word is capitalised to highlight its specific use. The term has served to distinguish these groups from those groups with skin of other colours (black, yellow etc), and hence derives from the concept of race but is used as an indicator of ethnicity. There are problems of poverty and excess disease in subgroups of the White population, which cannot be unearthed and tackled by using the label White.

## Overall Summary

**Introduction** The Race Relations (Amendment) Act 2000 and NHS policy require health services to show that they are promoting racial equality and reducing ethnic inequalities. Routine data sources in Scotland do not include the information needed to (a) measure health inequalities, (b) assess service use, and (c) demonstrate compliance with policy and legislation. The Scottish Executive funded a 30 month demonstration project (ending May 2005) aiming to extract data on the health status of ethnic minority groups from Scotland's major health databases. This report describes the conceptual, ethical, technical and practical challenges we faced, and gives an overview of the outputs. Our project focused on CHD as an example, with particular reference to South Asians.

**Objectives and Methods** We sought to (a) link the ethnic code in the Scottish Census to the Scottish NHS Community Health Index and subsequently to the Scottish Morbidity Record (SMR-01) database which has already been augmented with mortality data (Ch 6 and 7) (b) to test name search algorithms, validated elsewhere, in Scotland where 2 percent of the population are non-White Datasets used were mortality records, and the Dundee DARTS diabetes register (Ch 3) (c) Analyse country of birth as a proxy for ethnicity and as a variable of interest in its own right (Ch 5) (d) Extrapolate cardiovascular risk factor data from England to Scottish populations (Ch 4). In preparation we reviewed the actual, and potential, available data by ethnic group in Scotland (Ch 2).

**Results** (a) on the argument of overriding public interest, permission was obtained from MREC (Scotland), the Privacy Advisory Committee, and the General Register Office for Scotland to link census records to mortality/morbidity records. Overall, 94% of census records were matched to the community health index, with the figure for minority ethnic groups exceeding 85 %. The mortality and morbidity rates, and survival, for coronary heart disease by ethnic group showed important variations that require further research (Ch 6,7). The most important findings were that the incidence of acute myocardial infarction (AMI) in South Asians was, comparative to non-South Asians, about 60-70% higher. In contrast, survival following AMI was better, even taking into account age and sex differences.

(b) Computer based name search algorithms were found to be inaccurate in Scotland. However when the process was supplemented with visual inspection by knowledgeable observers, the algorithms performed sufficiently well to separate South Asian and non-South Asians and to perform some comparisons using the DARTS database. Despite equivalence in processes of care, the DARTS database showed South Asians had a higher HbA1c (a marker for high blood sugar) than non-South Asians. The difference was not present when the results were adjusted for age and sex differences between the populations. This observation needs further research. The age-adjusted prevalence rate in South Asians was 4-5 times higher than in non-South Asians, despite their lower BMI; an observation that corroborates international studies (Ch 3).

(c) Mortality by country of birth was calculated to show major variations in coronary heart disease. Scottish born residents of Scotland had much higher mortality rates than either the English and Welsh born people living in Scotland or the England and Wales population. The higher mortality rates from CHD in South Asian populations shown in England and Wales were demonstrable, but not statistically significantly so, in analyses with Scotland as the standard and using the age group 20-69. This difference and, the lower mortality from CHD in the Chinese and Hong Kong born, were clear cut when we used England and Wales as the standard. The key observation is that control of CHD in Scottish born, Irish born, and South Asian populations is a top priority (Ch 5).



(d) procedures for imputation of data by modelling were developed, giving some interim data on lifestyle that require further explanation and further development of methods (Ch 4).

(e) These observations, and the potential for further analysis, transform the meagre data currently available by ethnic group in Scotland (Ch 2 and 8).

**Conclusions** Information about the health of ethnic minorities can be extracted from routine databases in Scotland. Our project demonstrates that retrospective extraction of data by ethnic group from existing databases is possible at relatively low cost, and reasonable timescales, given co-operation between agencies and appropriate skills within the team. Methods are likely to be applicable beyond Scotland. The results on coronary heart disease demonstrate major variations that can be used to refine policies and services to improve the health of Scotland's ethnically diverse populations. The data also add to broader understanding of the causes and consequences of CHD. Similar data could be produced for a range of important disease. Pending the results of prospective ethnic coding in 5-10 years this approach fills a data vacuum.

### **Next steps and recommendations**

The data produced on CHD and its risk factors should be used to refine and adapt Scotland's policies on the prevention and control of CHD and its risk factors so they are relevant to our multi-ethnic society.

Subject to widespread consultation, and once the methods for census/SMR data linkage have been streamlined, ethical committee and PAC approval should be sought for a more comprehensive analysis of morbidity and mortality by ethnic group e.g. on cancers, respiratory diseases, infections etc.

The utility of the linkage methods should be evaluated in other UK nations and abroad.

The country of birth analysis should be extended to other major causes of death, with both Scotland and England and Wales as standard comparison populations to yield informative comparisons.

Computerised name search methods need to be further refined if they are to be used in a fully automated way. Investigators need to continue to include a visual inspection within the procedures

The imputation data underline the necessity for Scotland to undertake a survey of health and lifestyle in Scotland's minority ethnic populations

Pending the arrival of useful data from prospective ethnic coding – a process that could take 5 to 10 years – the approach described here will help fill an important gap. The project needs, therefore, to be continued for another 24 months after which it can be reviewed.

# 1 Introduction: ethnic variations in health, and the need for data in Scotland

BHOPAL R<sup>1</sup>, FISCHBACHER CM<sup>1,2</sup>, STEINER M<sup>1</sup>, CHALMERS J<sup>1,2</sup>, POVEY C<sup>2</sup>, JAMIESON J<sup>2,3</sup>, KNOWLES D<sup>2</sup>

(<sup>1</sup>Public Health Sciences, University of Edinburgh, <sup>2</sup>Information and Statistics Division, National Services Scotland, <sup>3</sup>National Resource Centre for Ethnic Minorities, NHS Health Scotland)

**Abstract:** The Race Relations (Amendment) Act 2000 and NHS policy require health services to show that they are promoting racial equality and reducing ethnic inequalities. Routine data sources in Scotland do not include the information needed to (a) measure health inequalities (b) assess service use (c) demonstrate compliance with policy and legislation. The Scottish Executive funded a 30 month demonstration project (ending May 2005) aiming to seed health databases with ethnic codes. This chapter describes the conceptual, ethical, technical and practical challenges faced, and summarises the anticipated outputs. Our project focused on CHD, particularly in South Asians as an example.

## 1.1 Background

People from minority ethnic groups (see p2 – terminology) show differences in health status and health care compared with the population as a whole for complex reasons, including their generally poorer socio-economic status, differences in environmental circumstances during the life-course, barriers to access to health services, genetic factors, and the effects of racism. For example, UK statistics show men born in the Indian sub-continent have a 38% higher death rate and women have a 43% higher death rate from heart disease than the general population. (Gill et al, 2005; Bhopal, 2000). The prevalence rate of diabetes is 3-4 times higher (NRCEMH, 2004). People from minority ethnic groups experience health and health care disadvantages and a key obstacle to improving matters in Scotland is the absence of any published data from routine monitoring of their health status and health service utilization.

Chapter 2 shows that in Scotland information on ethnic group is not consistently recorded in primary care and is not routinely held in computerised databases such as the General Practice Administrative System for Scotland (GPASS) or collected as part of the Practice Team Information (PTI) scheme in general practice. There is provision for the recording of ethnic group in NHS hospital discharge data, but this is usually omitted. Information collated by the General Register's Office for Scotland from death certificates includes country of birth but not ethnic group. For older people, country of birth has often been used as a proxy for ethnic group, but it is becoming less useful as the number of people from ethnic minority groups born in the UK increases. This said, even with such limitations birthplace based mortality data can shed important light upon ethnic variations in disease. In Scotland, however, such data have not been analysed. Diabetes and CHD registers do not hold accurate and comprehensive information on ethnic group but work is underway to achieve this.

The prospective development of ethnically coded information systems is a high priority for Scotland. However, the fruits of that work lie several years hence, and will not take advantage of information which is already available and which has the potential to yield insights into ethnic inequalities in health. This project aimed to meet needs pending the development of new ethnic monitoring systems.

## **1.2 Why include ethnicity in routine health data collection?**

Non-White ethnic groups accounted for 1.25% (63,000) of the Scottish population in 1991 and 2.01% (89,000) in 2001, an increase of 41%. Ethnic diversity in Scotland is likely to increase as recent immigrants are from more diverse backgrounds than previously. All of this means that the challenges of meeting the health care needs of Scotland's ethnic minorities are likely to grow, and that reliable data on ethnicity will be a crucial part of that task. Data collection is required on legal and equity grounds, and to address health needs (Scottish Executive Equality Unit, 2000).

Evidence from many sources outside Scotland indicates that there are likely to be wide inequalities in both risk factor levels and health outcomes by ethnic group. Gill *et al* studied the mortality rates between 1989 and 1992 of adults aged 25-74 living in England and Wales (2005). All cause mortality varied widely by country of birth. Standardised mortality ratios among women ranged from 70 (for those born in Bangladeshi) to 113 (Indian born) and among men from 79 (born in Hong Kong or China) to 114 (born in Bangladesh). Variations in specific causes of deaths were much wider, with nearly three-fold variations in CHD mortality (Gill et al, 2005). Such information is important to policy, planning and service delivery and raises questions important to public health and epidemiology research.

## **1.3 Health inequalities: policy and legal framework for improving information**

The 1999 White Paper "Towards a Healthier Scotland" identified the reduction of inequalities as a key priority and suggested that research effort "must focus on the causes of health inequalities and practical means to tackle them" (Scottish Executive, 1999). The Race Relations (Amendment) Act 2000 places a positive duty on public bodies to promote racial equality, to recognise institutional racism and to work to eradicate it. Scottish Executive Health Department Letter (HDL) 2002/51 requires Scottish NHS organisations to make a clear commitment to address the challenge of ethnic minority health, to develop a local demographic profile, to ensure equitable access to services by all ethnic groups, to include ethnicity in human resource strategies and to involve ethnic minority communities in service development (Scottish Executive, 2002). Without reliable information on ethnic group it will not be possible for NHS organisations to demonstrate that they have met legal and strategic obligations and may leave them vulnerable to legal challenge.

The Race Equality Advisory Forum report, "Making it Real; A Race Equality Strategy for Scotland" highlights the importance of better information in getting ethnic minority health into the mainstream of service provision, but also makes it clear that better information must become the base for prompt and efficient action (p10) (2001)

The almost total lack of routine data on health status and access to services by ethnic group has been recognised as a priority for action, partly because of policy and legal requirements. Without this information NHS Scotland is open to the accusation that it is not treating different ethnic groups equally.

In December 2002 an information gathering exercise was carried out by the Central Statistics Unit and the Equalities Research team to gain understanding of users needs in relation to ethnicity data. The Scottish Health Department response indicated that health status data were needed.

## **1.4 Purpose of this project and collaboration**

The purpose of the project was to demonstrate the feasibility of providing information about ethnic variations in diseases in Scotland using existing data sets, ultimately, to

underpin current policy initiatives to address ethnic health inequalities, inform more effective interventions at the community level and promote the equality agenda of the NHS and the Scottish Executive. The project focused on ethnic variation in CHD in Scotland, for reasons explained below. The project was a demonstration project intended to examine the feasibility and utility of the approaches outlined. Once established, these approaches might be used for work in other disease areas. The project is a collaborative partnership between the University of Edinburgh, the General Register Office for Scotland, Information Services (ISD) and the National Resource Centre for Ethnic Minority Health, part of NHS Health Scotland.

### **1.5 Focus on coronary heart disease in this demonstration project**

A focus on coronary heart disease (CHD) was justified for several reasons. CHD is a problem of major public importance for UK ethnic minority groups, particularly South Asians (Patel and Bhopal, 2004). The high level of CHD mortality in South Asians has already been mentioned. Using mortality data by country of birth from England and Wales the table below shows the extent of ethnic variations in this disease. In this table, the standard is 100, so a value exceeding this indicates a death rate above the population average. Scotland has an epidemic of CHD but the current Scottish CHD strategy does not include any references to ethnicity, at least partly because of lack of data (Scottish Executive, 2002). The investigators have current interests in the epidemiology of cardiovascular disease. Pragmatically, given the enormous potential scope of this project, it made sense to test the concepts and methods in one important disease area.

Among Pakistanis (the largest ethnic minority group in Scotland after the Irish) mortality among men was almost 50% higher than in the general population (table). As well as being relatively more common, in absolute terms CHD was also by far the most common single cause of death among Pakistani men. CHD mortality was similarly raised among Indian- and Bangladeshi-born men but substantially lower for Chinese men.

#### **Standardised mortality rates (95% confidence intervals) for ischaemic heart disease (ICD-9 410-414) by country of birth 1989-92, England and Wales (from Gill *et al*, 2005)**

	<b>India</b>	<b>Pakistan</b>	<b>Bangladesh</b>	<b>China *</b>	<b>Caribbean</b>	<b>Africa**</b>
<b>Men</b>	142 (137, 147)	148 (138, 158)	151 (136, 167)	44 (36, 54)	62 (58, 67)	58 (47, 70)
<b>Women</b>	158 (148, 168)	111 (93, 130)	91 (60, 133)	43 (30, 60)	86 (77, 96)	61 (37, 94)

\* including Hong Kong and Taiwan    \*\* West and South

There are also wide variations by country of birth and ethnic group in the prevalence of CHD risk factors. For example, while smoking is less common compared to the general population among Indian, Pakistani and Bangladeshi women, the proportion of Bangladeshi men who smoke (42%) is the highest of any ethnic group (Ehrens *et al*, 2001). While this demonstration project is focused on CHD, and particularly in South Asian populations, its methodological findings apply to all of Scotland's populations.

## **1.6 Aims and objectives**

### **1.6.1 Overall aim**

Using CHD as a specific example, to explore the potential for retrospective analysis of routine data sources on health in Scotland to provide information about:

- the current health status of ethnic minorities in Scotland and trends in their health
- ethnic inequalities in health and trends in inequalities

### **1.6.2 Specific objectives at the inception of the project**

1. Document the current and potential availability of information about the health of ethnic minorities from routine data sources in Scotland, focussing on CHD.
2. Explore ways of extracting such information retrospectively using names, country of birth and possibly language, religion or other clues to ethnicity. In particular,
  - (a) Obtain and test ethnic minority name search algorithms used and validated elsewhere in the UK
  - (b) Explore the possibility of validating name search algorithms using census data in Scotland
  - (c) Use name search algorithms to allocate ethnic group to routine health data sets
3. Analyse such ethnically coded routine health data sets to describe (for the specific example of CHD):
  - (d) Current utilisation of health services by ethnic group
  - (e) Ethnic patterns of health outcomes
  - (f) Inequalities in health outcomes
4. Demonstrate the ways in which such data could be presented and used to help improve the health of Scotland's ethnic minority groups

### **1.6.3 Additional objective tackled during the project that became the central focus**

We tested the feasibility and value of linking the ethnic code in the census data to health information held in the SMR database at ISD. This idea was foreseen but not included as a specific objective in the beginning because we did not envisage that legal and ethical problems could be solved. When we foresaw solutions to those problems we made this a central objective.

## **1.7 Outline of methods and data sources**

### **1.7.1 SMR-01 database**

We focused on the SMR-01 databases because they hold important data that is linked to mortality outcomes and because of the availability of names and other information such as addresses. The SMR mortality linked database currently holds information on around 20 million health service contacts with 5 million individuals. It includes information about all discharges from hospitals from acute specialties in Scotland since 1981 and includes up to 6 diagnosis codes. Mortality data from GROS has been linked to this database.

### **1.7.2 Name searching**

The name search method is a way of assigning ethnic group to people with distinct names but it is unsuitable for ethnic groups who do not have sufficiently distinctive names (such as Irish or Afro-Caribbean groups) and therefore fails to address the needs of all Scottish ethnic groups in an equitable way. It ignores the current recommendation that ethnic group should be self-assigned. While recognising that this approach on its own does not provide an adequate solution, we explored its value in both the SMR database and the DARTS (diabetes) database in Dundee.

### **1.7.3 Imputing (modelling) cardiovascular risk factors**

Existing data about levels of risk factors among ethnic minorities in other parts of the UK may be relevant to assessing levels of need (Ehrens et al, 2002). Data are available from the 1999 Health Survey of England. We used these data, together with population data on ethnic minorities in Scotland, to impute the burden of selected cardiovascular risk factors in ethnic minority groups in Scotland.

### **1.7.4 Addition of an ethnic group indicator to an anonymised extract of health service records using linked census data**

As part of the 2001 Census respondents were asked to define their own ethnicity. Through the Community Health Index (CHI) number this information was linked to hospital discharge data and mortality data held by (ISD). Personal identifiers were removed from the linked dataset. The advantages of this approach were that it:

- provided information for all ethnic groups and not only those with distinctive names
- was based on self defined ethnicity, the currently recommended approach
- capitalised on the experience of data linkage in Scotland.

This work on data linkage was, in our judgment, allowed under the Data Protection Act, was required under the Race Relations (amendment) Act 2000 and was within the duties of the Registrar General. The work followed the principles of the Data Protection Act and two National Statistics (NS) Protocols on Data Access and Confidentiality, and Data Matching. This work was supported by the Commission for Racial Equality, NHS Health Scotland (through NRCEMH), ISD, the Scottish Executive Health Department and the Chief Medical Officer. The methods ensured that no identified individual response to the Census was linked to health records and that no identified health records were linked to Census information.

## ***1.8 Legal and ethical issues***

As appropriate to each element of the research we also obtained approval from the Scottish Multi-centre Research Ethics Committee. Legal advice (obtained by GROS) indicated that the proposed use of census data was compliant with the Census Act 1920 and the Data Protection Act 1998. We also obtain approval from the Privacy Advisory Committee (PAC) which advises ISD and GRO(S) on the use of data, and from Malcolm McWhirter, the lead Scottish Director of Public Health for data confidentiality issues, and local research ethics committees for components of the project when necessary.

## ***1.9 Publicising the project***

We disseminated information about the project to the Commission for Racial Equality, through the PHIS and NHS Health Scotland websites, the National Heart Forum newsletter, electronic mailing lists and the BE4 black and minority ethnic group website. Through these and other channels, including presentations at national and international

conferences, we made efforts to encourage people to question and discuss the proposal and the early findings.

### ***1.10 Context of this project in relation to national initiatives, and prospects for the future in the light of ethnic monitoring***

The long term solution to addressing the needs for information on the health of minority ethnic groups is the prospective collection of information about ethnic group including language, religion, dietary requirements, relevant cultural practices and country of ancestral origin

Experience over many years from England and Wales and recent pilot studies indicate that the work involved in the development of such an ethnic monitoring system (including raising public awareness, training staff to collect data etc) takes many years.

When this work was instigated, implementing prospective information systems was at the very early stage of discussion. The development of such systems is now a high priority for the Scottish Executive, ISD and the National Resource Centre on Ethnic Minority Health at NHS Health Scotland. With full commitment, reliable data should be available in about five years. This project fills the gap in the interim. If the project is judged successful in adding ethnic information to routine data sources in relation to coronary heart disease it would be appropriate to extend it to examine other health issues.

## 2 Current availability of information on ethnicity and health in Scotland

FISCHBACHER CM,<sup>1,2</sup> JAMIESON J,<sup>2,3</sup> BHOPAL R,<sup>1</sup> POVEY C,<sup>2</sup> KNOWLES D,<sup>2</sup> CHALMERS J,<sup>1,2</sup>  
(<sup>1</sup>Public Health Sciences, University of Edinburgh, <sup>2</sup>Information and Statistics Division, National Services Scotland, <sup>3</sup>National Resource Centre for Ethnic Minorities, NHS Health Scotland.,

**Abstract** Information on ethnicity and health is sparse in Scotland. The 1991 and 2001 censuses provide ethnic group, and other relevant variables including country of birth. There are no usable data from primary care, the Scottish Morbidity Register, child health information systems or from large-scale social surveys. Data from genitourinary services and drug misuse services are collected but have not been analysed. There are attempts to remedy this problem, including the inclusion of ethnicity and religion in the planned e-Health National Clinical Dataset. The need for this current project is, therefore, explicit.

### 2.1 Introduction

This chapter reviews the availability of information on ethnicity and health from routine sources in Scotland. We have defined information on ethnicity as information that would allow routinely collected data to be disaggregated and presented by separate ethnic group. Other information is also collected on characteristics relevant to ethnicity, such as religion and language. We have defined health broadly. By routine we mean sources that are collected as part of routine patient care rather than as part of specific research or audit studies. This chapter does not cover information from research studies of the health of minority ethnic groups in Scotland, which has been reviewed elsewhere by Netto and colleagues (2001). Sources of information from regular surveys are summarised. Information sources have been divided into general sources and those collected by the health service, and the latter have been divided into those collected in primary care settings, in secondary care and in specialist services.

In August 2000 the Scottish Executive published "Equality in Scotland: Guide to Data Sources". An updated version is available at: <http://www.scotland.gov.uk/stats/egds/egds-00m.asp>. This is a wide ranging review.

### 2.2 Census

The 1991 and 2001 Censuses in Scotland and England collected information on country of birth and ethnic group (Registrar General for Scotland, 2003; National Statistics, 2003). The 2001 census collected data on religion of upbringing and current religion. The ethnic group categories used in the census forms 1991 and 2001 are compared by Gill et al (2005) and the table is reproduced in the appendix at the end of this chapter. Standard Census reports provide population estimates by ethnic group and cross tabulations of ethnic group by country of birth, age and sex. The census includes questions about self assessed health and the presence of limiting long term illness (LLTI). Results are published in five ethnic groups (White, Indian, Pakistani and other South Asian, Chinese and Other) and measures of health are available by age, sex and these five ethnic groupings.

### 2.3 Primary care

Most information on primary care in Scotland is collected for administrative purposes, for prescribing and for direct patient management, and little clinical information is available centrally. Around 900 (85%) of Scotland's 1050 GP practices use General Practice Administration System for Scotland (GPASS) software for administrative purposes, for



prescribing and for recording clinical care. Other systems used in Scotland include EMIS, Vision, Torex and Exeter systems. Information on ethnicity is not routinely collected by GPs or other primary care staff. In addition, there are issues around coding and storing ethnicity data in GP computer systems. There are local projects underway to help rectify this for the future.

There is no agreed terminology or coding scheme for ethnicity in primary care. The Read coding system is widely used to code clinical diagnosis and other characteristics of patients and clinical consultations. The Read system includes a large number of codes that cover race or ethnic group. These include a variety of approaches, such as classification on the basis of appearance: "O/E Europeanoid" (2261.); place of birth: "Born in Pakistan" (13eW.); "race": "RACE: Pakistani" (134M.); family origin: "Indian origin" (1347.) They include ill-defined terms such as "Mongoloid origin". There is no common agreement about which codes to use. Read codes for the 2001 census categories were added only recently. Even if information on ethnicity was routinely collected in primary care, it would not be useful unless either standard categories were used or categories were used that had an agreed mapping to a set of standard categories. There are plans to introduce the SNOMED CT system in place of the Read coding system. SNOMED CT includes codes for ethnic group but these do not exactly match those used in the 2001 Census in Scotland. None of the commonly used software packages used in primary care have a routine means of recording and storing ethnic group, unlike other personal details such as age, sex or marital status.

The PTI scheme collects data from a sample of 54 Scottish practices and is the main source of information in Scotland on the pattern of consultations in primary care. Information on ethnic group is not included in the scheme.

## ***2.4 Scottish Morbidity Record and Scottish Cancer Registration systems***

A wide range of information on hospital episodes (inpatients, day cases and outpatients) is recorded and reported through the Scottish Morbidity Record (SMR-01) scheme. Between 1996 and 2004 a two character code for ethnic group was an optional item. The guidance stated "Although not mandatory, it is strongly recommended that these items be completed whenever the information is available." The codes introduced to the SMR-01 scheme in 1996 were as follows:

1. White
2. Black Caribbean
3. Black African
4. Black Other
5. Indian
6. Pakistani
7. Bangladeshi
8. Chinese
9. Other Ethnic Group
10. Not Known
11. Refused

This classification differs from that used in the 2001 Census in Scotland mainly in that it does not include "mixed" or "other Asian" categories. ISD in April 2004 asked data

providers to stop collecting data using this classification and to use codes based on the 2001 Census. However, the present data collection system and validation rules are not able to store these census codes centrally.

Existing data on ethnicity in the SMR-01 system is severely incomplete, being missing in around 94% of records. An analysis of hospital discharges in the SMR-01 scheme during 2002 shows that 10 NHS Boards had ethnicity recorded in less than 1% of returns, 3 had less than 50%, one had 52% and one 60%. In addition it is not clear how the information is collected, that is, whether it is based on assessment of appearance by administrative staff or whether patients (or relatives) are asked to state the ethnic group.

Ethnic group is recorded in 18% of Scottish cancer registrations. Information for registrations are obtained from clinical notes which often do not record ethnic group (D Brewster, personal communication).

## **2.5 Child Health Information**

The Child Health Surveillance Programme (CHSP) supports the Developmental Screening and Immunisation Programme for children in Scotland. Currently 10 NHS Boards participate in the scheme. Information is collected by Health Visitors at the first visit to a new baby. Ethnic group is categorised (by parents) as follows:

- Caucasian
- Indian
- Pakistani/Bangladeshi
- West Indian
- Chinese
- Other – please specify

For births in 2001, 11 health boards recorded information on ethnic group, with completeness of recording varying from 1-74% (median 22%). The programme has identified several obstacles, including that the ethnic codes are not compliant with the Scottish Census 2001

Further information is available at:

[\(http://www.isdscotland.org/isd/info3.jsp?pContentID=1022&p\\_applic=CCC&p\\_service=Content.show&\)](http://www.isdscotland.org/isd/info3.jsp?pContentID=1022&p_applic=CCC&p_service=Content.show&)

## **2.6 Health Visitor/District Nurse databases (CIS) in Lothian**

This system was designed to provide management information and to support health visitors and district nurses in the management of individual clients. Ethnic group is recorded and the field is said to be 100% complete. There are several recognised problems in relation to this system:

- Although the field for ethnic group is mandatory, in practice almost all entries are “white” or “not known”
- Staff have not received training in the approach to encourage self ethnicity classification
- There is no mandatory reporting of contacts by ethnic group
- Census 1991 codes are being used

## **2.7 Patient Information Management System (PiMS)**

PiMS is used in primary care and mental health divisions in six NHS Boards in Scotland. Although information is not collected currently, this system has the potential to collect ethnic group, religion and language and to meet, for example, the Census 2001 classifications as the database can be altered to suit Scottish data codes and standards. As with other systems, there is a need for staff training.

## **2.8 Genitourinary medicine clinics**

Reporting from genitourinary medicine clinics uses an anonymised individual patient return for new cases of sexually acquired infection. Twenty-two of twenty-four clinics collect information on ethnic status. In the year ending 31st March 2002 there were 39,220 attendances and ethnic status was recorded in 98.6% of all returns (source ISD (D)5). The proportion of attendances with an ethnicity other than White was 1.94%, but this varied from 0 to 3.3% for clinics with more than 100 attendances annually. There is anecdotal evidence that for some clinics "White" may be recorded as the default value. To date, no requests for analyses of these ethnic data have been received. The ethnic classification currently remains as Census 1991. A new web-based system was introduced in 2004 and uses Census 2001 categories as from 2005. Further information is available at: <http://www.show.scot.nhs.uk/support/sti/>

## **2.9 Drug misuse information**

National statistics provided by ISD from the Scottish Drug Misuse Database on "new" individuals attending drug treatment services include a breakdown by ethnic group. During the three year period from 2001/02 to 2003/04 information on ethnicity was reported for 89% of new clients, using the categories in the 2001 census. Only 0.76% of clients were from ethnic minority groups. Further information is available at: <http://www.drugmisuse.isdscotland.org/publications/04dmss/04sdmd.htm>.

## **2.10 Diabetes core dataset standard**

The diabetes core dataset includes a field for self-assigned ethnic group, using categories based on the 2001 Scottish Census. The 2003 Scottish Diabetes Survey reported that only 7 of 15 Scottish NHS Boards were able to provide information on the ethnic group of patients with diabetes. Overall 37% of patients had ethnic group recorded, and for the boards that provided information on ethnic group, the completeness of recording ranged from 1.1% to 100%.

## **2.11 National work on data standards**

The e-Health National Clinical Datasets Development Programme (NCDDP) aims to develop a set of agreed definitions for data items collected by the NHS in Scotland. As part of this programme Generic Core Data Standards have been developed jointly with Social Care Data Standards in Scotland. These were approved for use in December 2004 and will be published in the Health and Social Care Data Dictionary (<http://www.datadictionary.scot.nhs.uk/>). The Generic Core Data Standards include self-defined ethnic group categories based on the Scottish 2001 Census, with the addition of the categories "Not disclosed/ patient refused" and "Not known".

## 2.12 Surveys

The following surveys collect information from respondents on their ethnic group. However survey results are often not disaggregated by ethnic group. Some of these surveys are UK wide surveys that include only small numbers of participants from Scotland.

**Scottish Household Survey** (Scottish Executive). This is a continuous survey covering around 3,900 households each quarter with a focus on social justice, transport and housing issues. Information on smoking, limiting long-standing illness, health problems and disability is collected. Ethnic group is classified using the 2001 Scottish Census categories.

**British Household Panel Survey.** This is an UK-wide annual survey which includes around 2000 households in Scotland and collects information on ethnic group using a classification based on the 2001 Census. Individual and household details include: neighbourhood and individual demographics, current employment, labour and non-labour income, health and caring, employment history, values and opinions. Issues asked every 2 waves of survey include: attitudes towards cost/payments for health care, distribution of wealth - social justice, government's roles and responsibilities, environment, management of household expenditures. There are also different topical issues in each wave of the survey.

**Children Looked After Survey** (Scottish Executive). This is an annual survey that collects information on the number of children looked after by statute group, accommodation type, etc. It also includes numbers of children starting and ceasing to be looked after and in short-term placements. Information on ethnicity is collected using 2001 Scottish census categories.

**Employment Service (now Jobcentre Plus, part of the Department of Work and Pensions)** Data on ethnicity and disability are collected but totals involved are said to be too small to be worth reporting routinely. They are however available on request.

**Family Resources Survey – (Department for Work and Pensions).** This is an annual UK wide survey which includes approximately 2,200 households (4,900 individuals) in Scotland. Information is collected on receipt of Social Security benefits, housing costs, assets and savings, type of family, mean income, source of income. Information on ethnic group is available for all 3,747 adults in the sample but not for the children.

**Households Applying as Homeless** - Scottish Executive. This quarterly survey collects details of households applying to local authorities under the Homeless Persons Legislation and includes information on ethnic group using the 2001 Census classification.

**Households Below Average Income** - Department for Work and Pensions. This is a UK-wide survey derived from the Family Resources Survey, and previously (pre-1993/94), derived from the Family Expenditure Survey. It includes data from the Inland Revenue's Survey of Personal Income (SPI). It collects information on household income by characteristics of the household and includes approximately 2,200 households in Scotland. Information on ethnic group is derived from the Family Resource Survey.

The following surveys include participants in Scotland and gather information on ethnic group, but have sample sizes that are too small to allow any useful analyses

**General Household Survey - Office for National Statistics.** A UK-wide survey that includes around 800 households in Scotland.

**Omnibus Survey - Office for National Statistics.** This survey is carried out 8 times yearly and includes 150 people per month in Scotland.

**Scottish House Condition Survey (Communities Scotland).** This survey provides information by age, including information on household composition, employment details,

income and housing costs, satisfaction with housing and neighbourhood environment, heating use, and, in 2002, the health of respondents and children. The survey includes 18,158 households for social data and 15,051 households for both social and physical data. A breakdown by ethnicity is not available due to small sample size in non-White groups.

**Scottish Social Attitudes Survey** – (National Centre for Social Research). This is an annual survey of around 1600 adults in Scotland covering social attitudes.

### ***2.13 Conclusion***

Routine health data in Scotland generally provide very patchy and severely incomplete information about ethnicity. Several issues are common to all the sources reviewed: incomplete data, inconsistency about the categories used for ethnic group, a lack of training for staff on collecting data on ethnicity, and a lack of provision within IT systems for information on ethnicity to be recorded and exchanged. The need for new, innovative and intensive efforts to improve the situation is now evident.

## Appendix

We are grateful to Dr Paramjit Gill for permission to reproduce the text below and the table from Gill et al (Website – see text).

### Census 2001

As shown below, the Census 2001 question was significantly different as it asked questions on people of Irish descent and mixed parentage. Also, for the first time, the Northern Ireland 2001 Census included an ethnic group question thereby providing a comprehensive picture of the UK population ethnic group.

However, both the Scotland and Northern Ireland Census Offices adopted a modified version of the ethnic group question to that used in England and Wales. In both England and Wales and Scotland, a “two-tier” question was used: people were first invited to choose whether they were “white”, “mixed”, “Asian”, “Black” or “Other” and then directed to choose a more specific category within these broad groups. In Northern Ireland, a single-tier question was used. Census data is available by country on:

<http://www.statistics.gov.uk/statbase/explorer.asp?CTG=3&SL=&D=4712&DCT=32&DT=32#4712>  
(for England & Wales);

<http://www.scrol.gov.uk/scrol/common/home.jsp> (for Scotland); and

[http://www.nisra.gov.uk/Census/Census2001Output/standard\\_tables1.html](http://www.nisra.gov.uk/Census/Census2001Output/standard_tables1.html) (for Northern Ireland).

**Table 42 (in original publication): Census ethnic group classification in 1991 and 2001.**

<b>1991 Great Britain Equivalent</b>	<b>England and Wales (2001)</b>	<b>Scotland (2001)</b>	<b>Northern Ireland (2001)</b>
White	White: British  White: Irish White: Other White	White Scottish  Other White British White Irish Other White	White   Irish Traveller
Black-Other	Mixed: White and Black Caribbean Mixed: White and Black African	Any Mixed Background	Mixed
Other-Other	Mixed: White and Asian Mixed: Other Mixed		
Indian	Asian or Asian British: Indian	Asian, Asian Scottish or Asian British: Indian	Indian
Pakistani	Asian or Asian British: Pakistani	Asian, Asian Scottish or Asian British: Pakistani	Pakistani
Bangladeshi	Asian or Asian British: Bangladeshi	Asian, Asian Scottish or Asian British: Bangladeshi	Bangladeshi
Other-Asian	Asian or Asian British: Other Asian	Asian, Asian Scottish or Asian British: Any other Asian background	Other Asian
Caribbean	Black or Black British: Caribbean	Black, Black Scottish or Black British: Caribbean	Black Caribbean
African	Black or Black British: African	Black, Black Scottish or Black British: African	Black African
Other	Black or Black British: Other Black	Black, Black Scottish or Black British: Other Black	Other Black
Chinese	Chinese or other ethnic group :Chinese	Asian, Asian Scottish or Asian British: Chinese	Chinese
Other-Other	Chinese or other ethnic group: Other Ethnic Group	Other ethnic Background	Other ethnic group

### 3 The validity and value of name search algorithms for assessing ethnicity applied to names from the Scottish population: analysis of SMR-01 (mortality) and DARTS databases

STEINER M, FISCHBACHER CM, BHOPAL R, POVEY C, CHALMERS J, JAMIESON J, KNOWLES D

The DARTS database analysis was done in association with Professor Andrew Morris, Ritchie McAlpine and Philip Thomson at the University of Dundee.

#### Abstract

**Introduction:** One approach to adding ethnicity to databases that hold names is the application of name search algorithms to infer this information.

**Objectives and Methods:** We aimed to assess the validity of the name search algorithm Nam Pehchan 2 (NP2) and supplemented it with visual inspection by expert observers. We used (a) SMR-01 death records for the years 1997 to 2002 (b) the DARTS database. After we applied NP2 codes to our data, expert observers performed visual inspection of all NP2 identified South Asian (SA) names plus a random sample of non-South Asian names. We calculated interrater agreement and kappa for the independent expert observers, and the positive predictive value of NP2 with expert opinions as gold standard. We then analysed the DARTS database to assess the utility of the method.

**Results:** The SMR-01 data included 95,339 names. NP2 assessed 96.9% of these as non-SA origin. Name inspection was performed on 2,915 names identified by NP2 as SA and a random selection of 1,000 names identified as non-SA. Each observer independently categorised approximately 1,200 names as South Asian. The agreement between the three observers was 93.3% (kappa 0.9 [CI: 0.88-0.91]). Of 2,915 names identified by NP2, only 1,091 were confirmed by all three observers as SA. The positive predictive value for NP2 was therefore only 37.4%. All non-South Asians names were confirmed as non-South Asian by the observers.

In the DARTS database of 448 names identified by NP2 as South Asian, 274 were agreed by all three expert observers as South Asian (PPV = 61.2%). Of these 274, 176 were suitable for analysis. Prevalence of type 2 diabetes was about 4 times higher in South Asians than non-South Asians in Tayside. Indicators of processes of care were similar in the two ethnic groups. HbA1c levels and the prevalence of retinopathy were comparatively greater in South Asian, though they were younger. After age and sex adjustment these difference were not seen.

**Discussion:** The application of automated name search algorithms to a major population based Scottish disease register does not give satisfactory results. SA status needs to be confirmed by additional expert observers. Name search algorithms are useful as a screening tool but when the proportion of South Asians in the population is low, visual inspection by expert observers is essential for accurate classification. Our analysis of DARTS data shows that these methods are valuable in the absence of alternative methods – producing important information at relatively low costs. Despite satisfactory processes of care, the outcomes for South Asians were not as good as for the population as a whole, though the difference may be a result of differences in the age and sex structures of the two populations. This needs further research. Overall, the equitability of the care given is extremely encouraging.



### **3.1 Background**

One option for obtaining information on the health of ethnic minority groups is to assign by name search an ethnic code to already available routine health data. Software packages can use names to assign ethnic group on a probability basis and can be applied to assign ethnicity in large databases (Nicholl et al, 1986; Coldman et al, 1988). Classification of ethnic group by inspecting names has been used to identify South Asian, Chinese and Hispanic populations and was successfully used 20 years ago to obtain information about ethnic differences in cancer incidence in Scotland (Matheson et al 1990). Computerised name search algorithms have been available for some time and RSB has previously described the use of a name search programme to allocate Chinese and South Asian ethnic group. More recently at least two other algorithms have been developed-Nam Pehchan 2 and SANGRA (Nanchahal et al, 2001; Cummins et al, 1999).

There are some disadvantages of this approach including:

- it performs poorly in women whose married name does not reflect their ethnic group.
- it is unsuitable for ethnic groups who do not have distinctive names (such as Irish or Afro-Caribbean groups) and are, therefore, assigned to the White group;
- it therefore fails to address the needs of all Scottish ethnic groups in an equitable manner and may not meet the requirements of the Race Relations Amendment Act to “promote good relations between persons of different racial groups”; and
- it ignores the current recommendation that ethnic group should be self assigned.

Nonetheless, there are times when the method works and there is no superior alternative. The aim of this work is to assess the validity of the software solutions when applied to names in the Scottish population. A subsidiary aim was to assess the value of the analysis arising from the name search. Once the census to SMR-01 linkage was achieved (see chapter 7) there was no purpose in analysing the SMR data so only the DARTS database analysis is presented.

The Diabetes Managed Clinical Network (MCN) of the Tayside region has established an efficient clinical governance programme, which spans primary and secondary care, to monitor progress and inform service development required to deliver a clinically effective diabetes service (2002). The indicators monitored and fed into the Diabetes Audit and Research in Tayside Scotland (DARTS) data base are on diabetes type 1 and type 2, and include HbA1c measures, blood pressure measures, cholesterol testing, eye screening and treatment (diet, tablets, insulin). The DARTS data base is a valuable source of information about patients with diabetes and their treatment, outcome (e.g. myocardial infarction, stroke etc.) and service quality. This is the first time that an analysis has been done by ethnic group. We focused on type 2 diabetes as this is the type that is seen in marked excess in South Asians. On first principles, in the light of potential linguistic, cultural and access barriers, we anticipated that the quality of care would be comparatively poor in South Asians. We expected South Asians to have a much higher prevalence of type 2 diabetes.

### **3.2 Methods**

#### **3.2.1 Name searching (SMR and DARTS databases)**

Name search was applied to mortality data extracts from the Scottish Morbidity Record (SMR01) and (DARTS). The data extract from SMR01 contained the names from death records between 1997 and 2002. The data extract from DARTS consisted of patients with diagnoses of diabetes mellitus type 2 who entered the register till 04/11/2003.

Both data sources were analysed using Nam Pehchan 2 (NP2). Names identified by the software as South Asian were presented to expert observers with experience of South Asian cultures and communities. The list of names came from databases of confidential and personal medical information. To decrease the risk of disclosure of confidential information, South Asian names were taken from the publicly available electoral register and mixed randomly with the names of patients. The expert observers were not provided with any information about these people other than their name. They were asked to keep these names confidential and a confidentiality agreement was signed.

We calculated positive predicted values and agreement between raters. Sensitivity and specificity could not be calculated for NP2 as only NP2 positive results were examined by the expert observers. We calculated kappa as a measure for inter-rater agreement between expert observers for South Asian affiliation and between the expert observers and NP2 for language and religion. We followed Altman's (2000) assessment of the quality of agreement as given below:

Kappa	Quality of agreement
< 0.20	Poor
0.21 – 0.40	Fair
0.41 – 0.60	Moderate
0.61 – 0.80	Good
0.80 – 1.00	Very good

Our focus in this report is on NP2's capacity to identify South Asian status. The programme also assigns language and religion. We found that our expert observers' own assessments of language were in disagreement. There were also substantial disagreements among the observers on religion. This line of enquiry was, therefore, not pursued. There is an alternative computer programme called SANGRA. We noted that NP2 and SANGRA differed greatly in the numbers of names identified as South Asian. Further observations may be reported in due course in collaboration with the team that developed SANGRA (see acknowledgements).

### 3.2.2 Analysis of DARTS database

An extract from the DARTS database was taken on the 04/11/2003 of patients with diabetes type 2 only. Nam Pehchan software was applied to all names in the database and experts assessed the output and produced a list of South Asian names (see above).

The status of the persons was checked to see if they were still alive and/or living in the Tayside area on the date the extract was taken. The figures were calculated in accordance with the demographic and clinical guidelines set out in the Scottish Diabetes Survey 2002. The group was assessed till the end of 2004 for diabetes monitoring, process of care indicators, and diabetes outcomes. We calculated unadjusted distributions and figures for patients categorised by ethnic status (South Asian versus non-South Asian) by age categories and sex. Prevalence was calculated for people alive and living in Tayside on the 04/11/2003 using the census population 2001 for Tayside categorised by age, sex, and ethnicity as the denominator.

We also calculated age adjusted figures for outcomes using all available patients (non-South Asians plus South Asians) in the data extract as the reference population. For categorical variables (e.g. prevalence, eye screening, treatment, etc.) we used indirect and direct age-standardisation, while for continuous variables (e.g. blood pressure, HbA1c, cholesterol, etc.) we used linear regression methods.

### 3.3 Ethics

The work on the SMR-01 database was approved by the MREC Scotland and the Privacy Advisory Committee of ISD, and that on the DARTS database by the Tayside Local Research Ethics Committee.

### 3.4 Results

#### 3.4.1 Name search-SMR01 death records

SMR01 records contained 95339 names for the years 1997 to 2002. 2915 (3%) of these names were identified by NP2 as South Asian based on the full names. 1000 random non-South Asian names identified from the electoral register were added before these were presented to the experts.

The three experts assigned 1210, 1229 and 1230 names, respectively, as South Asians. Approximately 31% of the names identified by NP2 as South Asian were confirmed as South Asians. All three experts rated all 1000 non-South Asian names as non-South Asian origin (table 3.3).

**Table 3.1 Comparison of Nam Pehchan 2 and expert observers' assignment of South Asian ethnicity.**

NP2 ethnicity	NP2 number	Expert 1 agrees	Expert 2 agrees	Expert 3 agrees	All three experts agree
South Asian	2915	1210	1229	1230	1091
Non-south Asian	1000	1000	1000	1000	1000

The agreement between the experts showed a very good kappa of 0.90 (0.89-0.91) overall, and 0.88 (0.87-0.90) for South Asian names only. When NP2 was added as an independent rater kappa dropped to 0.35. The positive predicted value i.e. that names identified by NP2 will be South Asian on expert review, was only 37.4% (1091/2915).

#### 3.4.2 Name search-DARTS register

The DARTS register held data from 17442 patients with diabetes mellitus 1 and 2 up to the 04/11/2003. Of these patients 12958 were diagnosed with diabetes type 2. After correcting for those dying or leaving Tayside 10,509 patients were left for analysis. 676 persons died or left Tayside after the snapshot date and were excluded from analysis on follow up examinations.

The Nam Pehcham (NP2) software identified 448 names as South Asian. Another 448 names from the electoral register were added before these were presented to the expert observers for checking. The observers confirmed 274 names of the 448 initially identified as South Asian by NP2 were of South Asian origin, as shown in table 3.2. The positive predictive value for NP2 was 61.2% (274/448) which was similar to results reported elsewhere. For electoral register names the result was 71% (318 of 448 names).

**Table 3.2 Agreement on ethnic coding of names among expert observers**

Type of names	All experts agree			Experts initially disagree	After discussion agree	
	Total	South Asian	Non-South Asian		South Asian	Non-South Asian
All names	896	514	183	199	78	121
DARTS names	448	247	102	99	27	72
Electoral Register names	448	267	81	100	51	49

The agreement measured as kappa for the three expert observers on two outcomes (SA/Non-SA) for South Asian names is presented in table 3.3:

**Table 3.3: Inter-rater agreement for three expert observers for South Asian names**

	Total	kappa	95% CI for kappa
All names	896	0.66	0.63-0.71
DARTS names	448	0.68	0.63-0.73
Electoral Register names	448	0.64	0.53-0.70

This kappa shows a good inter-rater agreement according to Altman et. al (2000).

### 3.4.3 Diabetes and diabetes care in South Asians and non-South Asians in the DARTS database

The database extract was on 12958 persons in Tayside with type 2 diabetes (of our total of 17,442 placed on the database up to 2003).

The age structure of this population is shown in table 3.4. South Asians were much younger than non-South Asians.

**Table 3.4 Age by ethnic group in the DARTS extract**

Age groups	Non-South Asians	South Asians
< 50	8.8% (1116)	32.7% (81)
50 to < 65	26.9% (3414)	33.1% (82)
65+	64.4% (8179)	34.3% (85)
Age missing	0.0% (1)	

The following tables and figures show the results after removing persons who were not entitled to be included. Table 3.5 shows that relatively few people were removed in this process, and that 158 South Asians were left of the 274 identified by the name search process (table 3.2 -247 plus 27 shown in the DARTS names row, total 274)

**Table 3.5 No. of people in database extract and still alive and in Tayside**

	Data extract		Still in Tayside		Change
	(n)	%	(n)	%	
Non SA	10333	98.3%	9675	98.4%	-6.4%
SA	176	1.7%	158	1.6%	-10.2%

Table 3.6 shows that in non-South Asians the age-specific prevalence of type 2 diabetes was higher in every age group in men than women. Except in one age (75-84) age specific prevalences were similar in South Asian men and women. In the younger age groups (<55 years) in particular, the prevalence of diabetes was much higher in South Asians than in non-South Asians. The small difference in total prevalence reflects the relatively young age distribution of South Asians.

**Table 3.6 Prevalence of type 2 diabetes based on DARTS extract sample and Tayside census.**

Age range	Non -South Asian			South Asian			Prevalence ratio (SA/non-SA)
	Census n	type 2 DM n	%	Census	type 2 DM	%	
<b>Males</b>							
0 - 14	33734	0	0.0%	466	0	0.0%	-
15 - 24	24479	4	0.0%	507	0	0.0%	-
25 - 34	22807	40	0.2%	304	7	2.3%	11.5
35 - 44	27527	229	0.8%	223	13	5.8%	7.3
45 - 54	26175	709	2.7%	198	32	16.2%	6.0
55 - 64	21462	1381	6.4%	115	29	25.2%	3.9
65 - 74	17179	1805	10.5%	59	20	33.9%	3.2
75 - 84	9288	1142	12.3%	28	6	21.4%	1.7
≥ 85	2217	189	8.5%	2	0	0.0%	n/a
Total	184868	5499	3.0%	1902	107	5.6%	
Age adjusted rate			3.0%	10.5%			
95% CI			2.9-3.0%	3.0-12.4%			
<b>Females</b>							
0 - 14	32186	0	0.0%	407	0	0.0%	-
15 - 24	23152	7	0.0%	430	0	0.0%	-
25 - 34	24552	35	0.1%	262	1	0.4%	3.0
35 - 44	29376	168	0.6%	235	11	4.7%	7.8
45 - 54	27040	499	1.8%	147	24	16.3%	9.0
55 - 64	22811	993	4.4%	60	16	26.7%	6.1
65 - 74	20792	1588	7.6%	49	15	30.6%	4.0
75 - 84	14652	1203	8.2%	19	2	10.5%	1.3
≥ 85	6070	341	5.6%	2	0	0.0%	-
Total	200631	4834	2.4%	1611	69	4.3%	
Age adjusted rate			2.4%	9.8%			
95% CI			(2.3-2.5%)	(7.6-12.1%)			

Table 3.7 shows modest and clinically unimportant differences in the approach to medication in the two groups.

**Table 3.7: Comparison of treatment**

	Non-SA		SA	
	n	%	n	%
Insulin	1645	15.9%	25	14.2%
Tablets	5945	57.5%	111	63.1%
Combination*	180	1.7%	3	1.7%
Diet alone	2514	24.3%	37	21.0%
Unknown	49	0.5%	0	0.0%
Total	10333		176	

\*Combination = Insulin plus tablets

South Asians were more likely to have a structured review as shown in table 3.8.

**Table 3.8: Structured clinical review appointment in prior 15 months (period as used in the Scottish Diabetes Survey).**

Review	Non-SA		SA	
	n	%	n	%
Yes	7984	82.5%	145	91.8%
No	1691	17.5%	13	8.2%
Total*	9675		158	

\*(in order to limit the effects of mortality and migration, these figures are restricted to those patients who were still alive and registered on 31/11/2004)

After adjustment for age, and examining data for men and women separately, there was a modest but not statistically significant difference, in favour of South Asians.

Table 3.9 Part A shows that measurement of HbA1c was equally common in both groups. Part B shows that in the South Asians glycaemic control, as reflected by HbA1C, was comparatively poor with about 20% (compared with 13%) having a value of >9, reflecting poor control. Taking into account age and sex differences in a multiple linear regression model there was an association between South Asian ethnicity and a higher HbA1c though this was not statically significant at the 5% level (adjusted B coefficient = 0.1988, p=0.095, 95% confidence interval of -0.03 – 0.43). In a logistic regression where South Asians were compared with non-South Asians the odds ratio was raised for an HbA1c of  $\geq 7.5\%$  (OR =1.53, 95% CI= 1.1-2.1) but not for an HbA1c of  $\geq 9\%$  (OR=1.17, 95% CI = 0.78 – 1.76). These findings require corroboration in larger studies.

**Table 3.9 A: Comparison of HbA1c – testing**

In prior 15 months	Non-SA		SA	
	n	%	n	%
Yes	9299	96.1%	152	96.2%
No	376	3.9%	6	3.8%
Total	9675		158	

**Table 3.9 B: Comparison of HbA1c – results**

HbA1c Range*	Non-SA		SA	
	n	%	n	%
<7.5	5495	59.1%	64	42.1%
7.5-9	2559	27.5%	57	37.5%
>9	1245	13.4%	31	20.4%
Total**	9299		152	

*\*(HbA1c range as defined in the Scottish Diabetes Survey)*

*\*\* (Limited to those tested in the prior 15 months and based on the most recent value)*

As shown in table 3.10 A and B both the likelihood of a cholesterol measurement and of being within the advised range was similar in both groups. Age standardisation did not alter this interpretation. In a logistic regression model comparing South Asian with non-South Asians the odds ratio associated with a cholesterol of > 5mmol/l was 0.72 (95% CI = 0.50 – 1.05).

**Table 3.10A: Comparison of cholesterol – testing**

In prior 15 months	Non-SA		SA	
	n	%	n	%
Yes	9032	93.4%	151	95.6%
No	643	6.6%	7	4.4%
Total	9675		158	

**Table 3.10B: Comparison of cholesterol – results**

Cholesterol Range*	Non-SA		SA	
	n	%	n	%
≤5	6587	72.9%	110	72.8%
>5	2445	27.1%	41	27.2%
Total**	9032		151	

*\*(Cholesterol range as defined in the Scottish Diabetes Survey)*

*\*\* (Limited to those tested in the prior 15 months and based on the most recent value)*

Table 3.11 A and B shows that blood pressure management and outcome was slightly better in South Asians than non-South Asians, the latter possibly partly reflecting their younger age structure. In a logistic regression model comparing South Asians the odds ratio associated with a systolic blood pressure >140mmHg was 0.74 (95% CI =0.52 – 1.04).



**Table 3.11A: Comparison of blood pressure - testing**

In prior 15 months	Non-SA		SA	
	n	%	n	%
Yes	8759	90.5%	146	92.4%
No	916	9.5%	12	7.6%
Total	9675		158	

**Table 3.11B: Comparison of blood pressure - results**

Systolic BP Range	Non-SA		SA	
	n	%	n	%
≤140 mmHg	4894	55.9%	96	65.8%
>140 mmHg	3865	44.1%	50	34.2%
Total**	8759		146	

\*(Blood pressure range as defined in the Scottish Diabetes Survey)

\*\* (Limited to those tested in the prior 15 months and based on the most recent value)

Despite equivalence in retinal screening, slightly more South Asians had retinopathy (27.4% compared with 22.4%), despite being younger (table 3.12).

**Table 3.12A: Retinal Screening – testing**

In prior 15 months	Non-SA		SA	
	n	%	n	%
Yes	8186	84.6%	135	85.4%
No	1489	15.4%	23	14.6%
Total	9675		158	

**Table 3.12B Retinal Screening – results**

Retinal Status	Non-SA		SA	
	n	%	n	%
Normal	5711	69.8%	86	63.7%
Retinopathy	1832	22.4%	37	27.4%
Screened - result unknown	643	7.9%	12	8.9%
Total**	8186		135	

\*\* (Limited to those tested in the prior 15 months and based on the most recent value)

Table 3.13 A and B shows that weight management was similar in both ethnic groups, and that comparatively fewer South Asians were obese (BMI >30). Multiple regression analysis showed that after controlling for age and sex, South Asians had a mean BMI that was 3.3 units lower than in non-South Asians (95% CI = 2.39 to 4.2 units lower). In a multiple logistic regression model comparing South Asians and non-South Asians the OR associated with a BMI of ≥25, defined as overweight, was 0.39 (95% CI = 0.27 – 0.55) and for ≥30 was 0.38 (95% CI = 0.26-0.55). As obesity, which is closely reflected in BMI, is a dominant causal factor for diabetes these results emphasise that the South Asians' comparatively high prevalence of diabetes occurs despite comparatively low BMI.

**Table 3.13A Comparison of body mass index – testing**

In prior 15 months	Non-SA		SA	
	n	%	n	%
Yes	8347	86.3%	144	91.1%
No	1328	13.7%	14	8.9%
Total	9675		158	

**Table 3.13B Comparison of body mass index – results**

BMI Range	Non-SA		SA	
	n	%	n	%
<18	44	0.5%	0	0.0%
18-25	1853	22.2%	48	33.3%
26-30	3045	36.5%	53	36.8%
31-35	2106	25.2%	33	22.9%
36-40	843	10.1%	6	4.2%
>=41	456	5.5%	4	2.8%
Total**	8347		144	

*\*(BMI ranges as defined in the Scottish Diabetes Survey)*

*\*\* (Limited to those tested in the prior 15 months and based on the most recent value)*

Table 3.14 shows that the prevalence of ever having one of four cardiovascular outcomes was lower in South Asians – probably reflecting their younger age structure. We examined the age standardised morbidity ratio by sex. The numbers were too small for precision.

**Table 3.14 Comparison of cardiovascular outcomes**

History of:	Non-SA		SA	
	N=9675	%	N=158	%
Myocardial Infarct	1314	13.6%	14	8.9%
CVA (Stroke)	755	7.8%	8	5.1%
Coronary revascularisation	588	6.1%	6	3.8%
Hypertension	5987	61.9%	61	38.6%

*\*(in order to limit the effects of mortality and migration, these figures are restricted to those patients who were still alive and registered on 31/11/2004)*

No amputations were reported for the South Asian population for this time period in Tayside.

## **3.5 Discussion**

### **3.5.1 Name search**

The proportion of South Asians in Scotland is approximately 2% and it is obvious that the same figure probably applies to South Asian names. When the proportion of people in the population with the factor of interest, in this case South Asian name, is low the positive predictive value (PPV) of a screening test is low (Bhopal, 2002). This explains why NP2 performs less well in Scotland than in England.

The results for both data bases suggest that the name search algorithm on its own applied to names in Scottish databases does not provide satisfactory results. It is recommended that people knowledgeable on South Asian names confirm the South Asian status conferred by the computer programme.

We conclude that this approach on its own does not provide an adequate solution even for South Asian names, and is likely to be worse for other ethnic groups such as the Chinese.

Current name search algorithms can be useful as a screening tool that needs to be accompanied by visual inspection by expert observers.

### **3.5.2 Value of name search based analysis of DARTS database**

The analysis of DARTS data has provided important and novel insights and shown the strengths of the name search based approach as an exploratory method. An estimate of age specific prevalence of type 2 DM was possible using census data. The anticipated comparatively high prevalence of diabetes in South Asians was confirmed. The extremely high rates in the younger South Asians may, in part, be an artefact of an underestimated denominator at Census 2001, but more likely they reflect an earlier onset of disease in South Asians.

Where this work breaks new ground is in showing that on important indicators of quality of care South Asians were not disadvantaged. This contradicted our prior expectation. Nonetheless, on two vital matters there was some discrepancy – South Asians were worse for glycaemic control as shown by HbA1c and retinopathy. We would expect both problems to be greater in older people. As South Asians are, on average, comparatively young we anticipated that after age adjustment the differences would be the case. The opposite was true – the difference diminished. These observations deserve more attention. Future research needs to consider the effect of duration of diabetes, treatment and other factors that affect HbA1c. Analyses of BMI and waist sizes by ethnic group, in future, could use the lower cut-points specific to South Asian populations. These name search based data add substantially to our knowledge of diabetes in South Asian in Scotland and point to the need for more work.

In conclusion, where other methods of identifying minority ethnic groups are not available, name searching can be a useful tool that gives useful data, albeit in Scotland it needs augmenting with visual inspection by expert observers.

## Appendix:

### Information for the observers assessing the names

A list of names was shown on a screen in a spreadsheet format as below

validation index no	forename	surname	South Asian	sure	language	religion
287	Ammar	Al-Rikabi				
466	Tasnim	Rafiq				
563	Surriya	Hussain				
654	Martin	Safdar				
668	Mohammed	Nazim Sattar				
676	Brenda	Morrison				

The first column is an identifier number for the purposes of this study. It does not relate to any number used in the health service. The second and third columns are the forename and surname to be inspected. The instructions and codes are given below.

**South Asian column.** South Asian ethnic origin is defined as people whose ancestors originally lived in the countries that are now India, Pakistan, Bangladesh and Sri Lanka.

**Sure column.** This gauges the confidence with which the assignment is made by the observers.

**Language column.** Based on the name, the most likely South Asian language a person with this name might speak should be entered. Many people with these names will have English as their first language but the language that traditionally might have been associated with such a name is asked for. Expert agreement was low so these data are not presented in this report.

**Religion column.** Based on the name, the most likely religious group a person with this name would belong to should be entered. Expert agreement was good. These analyses are not reported here.

### List of codes for data entry

#### South Asian

- 0 name not South Asian
- 1 name South Asian

#### Sure (of SA status)

- 0 not sure
- 1 sure

#### Language

- 0 not South Asian
- 1 Bengali
- 2 Gujerati
- 3 Hindi
- 4 Punjabi
- 5 Pashtu
- 6 Singhalese
- 7 Sylheti
- 8 Tamil
- 9 Urdu
- 10 Other
- 11 Muslim ( up to3 languages)
- 12 Hindu, Sikh, Buddhist (up to 3 languages)
- 13 Common (more than one language possible)
- 99 Don't know

#### Religion

- 0 not South Asian
- 1 Common (more than one religion possible)
- 2 Hindu
- 3 Muslim
- 4 Sikh
- 5 Buddhist
- 6 Other
- 99 Don't know

## 4 Imputation and model estimates

STEINER M<sup>1</sup>, FISCHBACHER CM<sup>2</sup>, BHOPAL R<sup>1</sup>, POVEY C<sup>2</sup>, CHALMERS J<sup>2</sup>, JAMIESON J<sup>3</sup>, KNOWLES D<sup>2</sup>

(<sup>1</sup>Public Health Sciences, University of Edinburgh, <sup>2</sup>Information and Statistics Division, National Services Scotland, <sup>3</sup>National Resource Centre for Ethnic Minorities, NHS Health Scotland.,

### Abstract

**Introduction** Unlike England and Wales there are no data on the health related lifestyles of Scotland's ethnic minority populations. If the minority ethnic populations in Scotland and England are similar in their lifestyles then, following adjustment for age and social class differences, the English data could be used to estimate the situation in Scotland.

**Objective** To develop and test models to impute lifestyle data for Scotland using information from England.

**Methods** Smoking cigarettes and blood pressure were taken as examples of a self-reported binary variable and a continuous variable. Census 2001 Scotland gave population data by age, sex and ethnic group categories similar to those in the Health Survey for England 1999. Social class adjustment was not done for lack of time.

The methods of standardisation and linear and logistic regression modelling were used.

**Results** The standardisation method produced age adjusted point estimates of prevalence of smoking that were, with the exception of Caribbean men (38.2% in Scotland, 34.8% in England), within 2% of the value in England. The standardisation method did not yield confidence intervals. Regression methods did not yield valid data.

For every group except the Chinese and Irish women, and Caribbean men, the standardised blood pressures were within 2 mmHg of the HSE 99 values. These differences were also seen using the regression method, but in addition there were substantial differences for Indian, Pakistani, Bangladeshi and Chinese men. So, standardisation and regression gave importantly different results. The regression method yielded confidence intervals.

**Conclusion** The concept of in estimating lifestyle data for Scottish ethnic minority groups by using appropriate modelled English data proved technically demanding. As the cost and timescales of a multiethnic Scottish health survey are challenging this interim solution is worth pursuing. The next stage is to include adjustment for social class.

### 4.1 Background

To identify actions to prevent and control diseases it is necessary to have knowledge about the frequency and distribution of risk factors. However, currently there are no such data in Scotland. A potential way out of this dilemma is to search for data sources that might have such information available in other populations with a similar population structure. Then we could extrapolate lifestyle and risk factor profiles from ethnic minority groups in other populations (e.g. England) to the ethnic groups in the Scottish population. The assumption behind this approach is that the distribution of these factors is about the same in the ethnic groups in both populations except that demographic and social structures differ. A potential data source for this approach is the Health Survey for England 1999 (HSE 99) with a boost sample for ethnic minority groups (Erens et al, 2002). In this data set the estimates for the ethnic minority groups are based on a sufficient sample size.

## 4.2 Methods

The data sources used for this approach were the Health Survey for England (HSE) 1999 (Erens et al, 2001) and the Census data from Scotland for 2001. The HSE 99 concentrated on health in ethnic minorities and therefore the ethnic groups were boosted to get valid estimates. Because of the disproportionate sampling of the population, sample weights were assigned to the observations and our analysis of these survey data had to respect these weights. The age and sex distribution for the different ethnic groups in Scotland were taken from the Census data 2001. This analysis was based on self assigned ethnicity – this question was in both the HSE 99 and the Census 2001 and was used in a similar way. This allows the direct comparison of both groups.

The following ethnic groups were used: Caribbean, Indian, Pakistani, Bangladeshi, Chinese, Irish, and all others (called Core99, which was the general population excluding the other listed ethnic groups).

First, the age and sex distributions for the ethnic groups in the 2001 Scottish Census were compared with the distribution in the HSE 1999. This was done visually showing the age and sex distribution for the several ethnic groups in histograms and a chi-squared test to compare the two frequency distributions. The sample size in each group was checked to establish whether a further breakdown for social class was possible. Assuming that the population distributions were comparable estimates calculated from the HSE 99 could be assigned to the Census 2001 data. The potential variables we were looking for at this stage were smoking and blood pressure. We have taken into account the specific sampling method in this survey and used appropriate survey methods to allow for the stratified nature of the sample.

We used two approaches to calculate estimates for the Scottish population. The first approach is similar to standardisation, easy to understand and to apply. Proportions and means calculated by age, sex, and ethnic categories from the HSE 1999 data are simply applied to the same categories in the Scottish census and then aggregated according to the population structure in the latter. The second approach used linear and logistic regression models on the HSE 99 data and then adjusting the regression parameters for the age and sex groups for each ethnic group.

For prevalence data on smoking we simply applied the proportions found in the HSE data to the Scottish census data stratified by ethnic group, sex and age category. This approach did not easily allow the calculation of confidence intervals. The second method using regression models and adjusting the model parameters to the Scottish census population 2001 did not produce results and was omitted.

For continuous variables we checked the distribution and transformed the data if necessary. Systolic blood pressure values were log transformed prior to calculating means due to non normal distribution. Therefore we present geometric means after back transformation.

First for each combination of age category, sex and ethnic group we calculated the mean of the variable of interest. Then we applied these mean values to the Scottish census data multiplying the mean value times the count of people. The data were cumulated for each ethnic and sex category and divided again by the count of persons in this group. The disadvantage of this approach is the lack of confidence intervals so we can not tell anything about the precision of the values. Second, we used linear regression methods and adjusted the parameters according to the age and sex distribution in the Scottish census. This allowed us to calculate confidence intervals for the estimated systolic blood pressure in each ethnic group. Confidence limits were omitted for the general population due to insufficient numbers of measurements in the HSE 99 data.

The underlying assumption for all approaches here is that people in Scotland and England are comparable regarding the variable of interest within each age, sex and ethnic category. However, this method only adjusts for the age and sex structure in each ethnic group separately and allows us to adjust for age differences between the ethnic groups. Differences in the age structure by sex within each ethnic group could not be taken account of.

### 4.3 Results

Table 4.1 shows the distribution of ethnic groups in both data sets, HSE 1999 and Scottish census. Survey weights were applied to the HSE 1999 data to correct for the over sampling in the ethnic groups. The presented numbers reflect the true distribution in the population. The distribution of ethnic groups were significantly different between the census data 2001 and the Health Survey for England data. England has a higher percentage of South Asians with approximately 1.3, 1 and 0.4% for Indian, Pakistani and Bangladeshi in each group versus 0.3, 0.6 and 0.04% in the Scottish census data.

**Table 4.1: Distribution of ethnic groups in the Scottish census 2001 and the HSE data 1999**

Ethnic group		Data source	
		Scottish Census	HSE
Caribbean	N	1,778	1,294
	%	0.04	1.01
Indian	N	15,037	1,694
	%	0.3	1.33
Pakistani	N	31,793	1,216
	%	0.63	0.95
Bangladeshi	N	1,981	483
	%	0.04	0.38
Chinese	N	16,310	292
	%	0.32	0.23
Irish	N	49,428	5,281
	%	0.98	4.14

Comparing the age distribution for each ethnic group, separately for men and women, only Bangladeshi women and Pakistani men had a similar age structure in census and HSE data. The chi square test was significant for all other ethnic and sex groups (Data not shown).

Table 4.2 shows that within the Scottish population the age structure between men and women differed significantly in most ethnic minority groups with Indian, Pakistani and Bangladeshi women younger than men.



**Table 4.2: Age distribution by sex for each ethnic group in the Scottish census for people age 16 and older. Figures are percentages.**

Age-group	Caribbean		Indian		Pakistani		Bangladeshi		Chinese		Irish		General popul.	
	men	women	Men	women	men	women	men	women	men	women	men	women	men	women
Total N	736	728	6,187	5,377	10,799	10,502	792	580	6,334	6,452	21,970	24,680	1,883,890	2,103,418
16-24	17.1	18.1	24.0	24.1	26.6	28.4	25.1	28.1	29.0	26.8	12.8	12.2	14.6	12.9
25-34	23.8	25.7	28.0	28.0	26.7	28.2	27.3	32.2	23.3	23.5	17.7	15.0	17.4	16.6
35-44	28.5	23.2	18.5	19.7	20.8	20.8	22.9	18.8	21.2	22.4	18.1	14.7	19.6	18.7
45-54	13.5	16.1	13.8	13.9	12.1	11.9	12.6	11.4	14.0	15.3	16.9	15.2	17.7	16.2
55-64	7.9	9.1	8.9	7.9	8.1	6.0	7.2	5.3	6.9	5.5	14.9	14.8	13.8	13.3
65-74	4.4	3.7	5.2	4.2	4.5	3.8	4.2	2.1	4.3	4.3	11.6	14.2	10.4	11.5
75-84	4.5	3.0	1.4	1.9	1.0	0.7	0.3	1.0	1.1	1.9	6.8	10.5	5.3	7.9
85+	0.4	1.1	0.3	0.4	0.2	0.3	0.5	1.0	0.3	0.5	1.4	3.5	1.2	3.1

Further, variables to adjust for social category were not comparable between HSE 1999 and Census 2001. Social class for example was measured in HSE 1999 as Register general social class categories I to V while in the Scottish census NS-SEC (National Statistics Socio-Economic Classification) was used.

### 4.3.1 Smoking

Table 4.3 shows the prevalence for smoking for the HSE 1999 data and age adjusted results for the Scottish population. Caribbean men had a higher prevalence in smoking in Scotland compared to data from England. The other ethnic groups showed a very similar smoking prevalence in Scotland compared to England. These differences are based on the different age structure of the ethnic groups between England and Scotland.

**Table 4.3: Percentage of smokers by ethnic category for HSE 1999 data and adjusted for the age distribution of the Scottish census population**

<i><b>Ethnicity</b></i>	<b>HSE 1999</b>			<b>Scotland</b>
	<b>Prevalence</b>	<b>Lower CI</b>	<b>Upper CI</b>	<b>Prevalence (modelled)</b>
<b>Women</b>				
Caribbean	25.1%	21.4%	28.8%	26.9%
Indian	5.8%	3.8%	7.8%	6.5%
Pakistani	4.7%	2.8%	6.5%	4.5%
Bangladeshi	1.1%	0.1%	2.2%	1.2%
Chinese	9.1%	5.6%	12.5%	10.9%
Irish	32.8%	28.5%	37.0%	31.2%
<b>Men</b>				
Caribbean	34.8%	30.2%	39.5%	38.2%
Indian	22.7%	18.9%	26.5%	23.3%
Pakistani	26.2%	22.6%	29.9%	26.2%
Bangladeshi	44.1%	39.7%	48.4%	44.0%
Chinese	17.2%	11.4%	23.0%	18.1%
Irish	39.0%	34.5%	43.6%	38.0%

### 4.3.2 Systolic blood pressure

Table 4.4 shows the proportion of valid measurements in the HSE 1999 for each ethnic group. Blood pressure measurements were taken in 50 to 62% of most ethnic minority groups, but in Bangladeshis only 39% were measured.

**Table 4.4: Proportions of valid blood pressure readings by ethnic group for persons 16 years and older in the HSE 1999 data set**

Ethnic group	Adjusted N	Adjusted % of valid measurements
Caribbean	870	52.6
Indian	1212	61.0
Pakistani	728	50.1
Bangladeshi	275	38.9
Chinese	208	58.6
Irish	3534	61.8

Table 4.5 shows the geometric means using both the standardisation method and adjustment using linear regression. For women we found only small differences between the standardisation method and adjusted regression. The differences between these both estimation methods were more obvious for men. Here we found lower estimated values with the regression method compared to the standardisation method. For adjusted regression we were able to calculate confidence intervals.

**Table 4.5: Geometric means for blood pressure from HSE 1999 and adjusted to the age distribution of the Scottish census population using standardisation and regression methods.**

Ethnicity	HSE 1999			Standardi- sation g. mean	Scottish estimates		
	g. mean	95% CI Lower CI Upper CI			g. mean	Adjusted regression 95% CI Lower CI Upper CI	
<b>Women</b>							
Caribbean	127.97	126.02	129.95	126.43	126.48	124.91	128.08
Indian	124.68	122.58	126.82	122.82	123.87	122.19	125.58
Pakistani	121.77	119.88	123.69	122.81	121.86	120.09	123.66
Bangladeshi	119.26	116.67	121.91	119.32	118.25	116.34	120.19
Chinese	123.89	120.66	127.21	121.51	121.41	119.56	123.28
Irish	128.94	126.93	130.99	132.33	130.47	128.87	132.08
<b>Men</b>							
Caribbean	135.25	133.17	137.36	132.90	130.28	128.50	132.09
Indian	132.93	130.82	135.08	131.42	127.59	125.80	129.40
Pakistani	128.76	126.81	130.73	128.63	125.52	123.63	127.44
Bangladeshi	125.43	123.11	127.78	123.96	121.80	119.80	123.84
Chinese	129.69	127.00	132.45	128.17	125.05	123.01	127.13
Irish	134.75	132.64	136.89	135.15	134.39	132.58	136.22

*g.mean: geometric mean*

*CI: confidence interval*

*HSE 1999 geometric means and confidence interval calculated with svymean (Stata8.2); Scottish Census data calculated using linear regression (svyreg).*

## 4.4 Discussion

We were able to demonstrate that it is possible to impute reasonable estimates for prevalence of health related risks. Also it is possible to estimate the distribution of continuous variables and present these as an average (mean) in a distinct ethnic group and give confidence limits. This said, this work was more difficult than we expected.

Both presented approaches have their limitations. The standardisation method is easy to implement and the principle is easy to understand. It can be applied to all sorts of variables, binary (e.g. smoking yes/no), categorical (e.g. body mass index < 25, 25-30, >30) or continuous measurements (e.g. blood pressure). But we can only present the point estimate and do not get confidence limits in the same easy way. Using adjusted regression methods we get confidence intervals from our analysis. As shown for blood pressure the mean values differed between the standardisation method and regression methods mainly for males. This is due to the fact that for this regression we assumed that men and women show a similar age distribution within each ethnic group. As we have shown in table 4.2 this assumption does not hold and we would need to include more variables (interaction terms) in our model to adjust for this fact.

A third approach to model the distribution of risk factors in Scottish minority groups is available called probabilistic sampling. This method is used mainly in the topic of exposure assessment where no data are available for the population under risk. A population equal to the population of interest (e.g. ethnic minority groups in Scotland) is simulated together with the risk factors. Instead of applying only a point estimator the frequency distribution or a point estimate and its variance (and other potential distribution parameters) are determined. Monte Carlo methods are used to assign a random value according to its probability to the census data. This artificial data set represents similar distributions for each single variable of interest for each group as found in the HSE 1999 data set. The variance is kept in the data set for analysis and conventional statistical methods can be used. However, this approach is very time and resource intensive and was beyond the scope of this project as the resources were limited.

Despite the limitations of the methods used and explained above the estimated results for the Scottish population allow insights into what we might expect in Scotland in different ethnic groups and where further investigation might be necessary.

The current results show no big difference between Scottish and English population for smoking and blood pressure in different ethnic groups. It is worth adjusting the models additionally for social status. As mentioned above different scales were used in HSE 1999 and Scottish census. National statistics made a linkage table available which allows to make these scales to a certain amount comparable ([http://www.statistics.gov.uk/methods\\_quality/ns\\_sec/continuity.asp](http://www.statistics.gov.uk/methods_quality/ns_sec/continuity.asp)). Further research is necessary to evaluate this effect as social structures in ethnic groups are probably different for Scotland and England.

**Overall, the results encourage us to develop further these modelling approaches pending Scotland developing and undertaking a health survey of ethnic minority groups.**

## 5 All cause and cardiovascular mortality by country of birth

FISCHBACHER CM<sup>2</sup>, STEINER M<sup>1</sup>, BHOPAL R<sup>1</sup>, CHALMERS J<sup>2</sup>, POVEY C<sup>2</sup>, JAMIESON J<sup>3</sup>, KNOWLES D<sup>2</sup>.

In association with GROS.

(<sup>1</sup>Public Health Sciences, University of Edinburgh, <sup>2</sup>Information and Statistics Division, National Services Scotland, <sup>3</sup>National Resource Centre for Ethnic Minorities, NHS Health Scotland.)

### Abstract

**Objective.** Country of birth is commonly used as a marker of ethnic group. For recent migrants the individual's country of birth may be an accurate proxy e.g. most people of Chinese birth would also be ethnically Chinese. Major differences by country of birth in all cause, CHD and stroke mortality have been demonstrated in a number of countries, including England and Wales. One of the most important findings from such research is the high mortality from CHD and stroke in people born in the Indian Subcontinent. There are no published data on such variations in Scotland.

**Design.** We calculated standardised mortality ratios (SMRs) by country of birth for Scottish residents aged 25 years and over, and separately in those 25-69, for a 6.25 year period between January 1997 and March 2003 using routine mortality statistics and 2001 census denominators for Scotland, adjusted appropriately for the time period in question. For the standardisation we used both mortality rates and population structures from England and Wales, and from Scotland.

**Results.** For some countries e.g. India, substantial numbers of people born there were of White ethnic origin, this applying particularly to people over 60 years of age. Comparisons with England and Wales as the standard showed high all-cause, CHD and stroke mortality in most country of birth groups living in Scotland. With Scotland as the standard, most country of birth groups were either similar to, or had lower SMRs, than the Scottish born. In men born in India or Pakistan CHD SMRs were not significantly raised compared to men born in Scotland. However, compared to men born and resident in England and Wales, SMRs were significantly raised in men born in India and in Scotland. Finally, in the age group 25-69 the excess of CHD anticipated in Indian subcontinent born populations was clear though not statistically significant.

**Conclusions.** These are the first data on country of birth by mortality in Scotland. As elsewhere, major variations have been demonstrated. Comparisons within the Scottish population might be interpreted as reassuring for South Asians born outside Scotland, since they do not show the excesses in CHD mortality by country of birth reported in England and Wales. However, the use of England and Wales as a comparison group shows a substantial excess of CHD risk among South Asians in Scotland, comparable to that reported in England and Wales. This analysis highlights the importance of the choice of reference, and age groups in assessing relative mortality risks. In relation to control of coronary heart disease either in South Asians in Scotland, or all those born in Scotland, there is no room for complacency. On a broader note the analyses show the potential in building up a picture of the health of Scotland's minority populations who are recent migrants. Future work can use methods to examine variations and trends over the period 1971-2004. The massive challenge of securing for Scotland the much lower mortality rates seen in England is made explicit.

## **5.1 Background**

Country of birth is recorded on death certificates and census returns. It provides a fair guide to ethnicity for older people from several minority populations and for those groups who migrated to Scotland more recently. However, 40-50% of people from minority ethnic groups were born in the UK (with higher proportions among younger people) and country of birth information will become decreasingly useful. This said, most deaths occur in the older age groups, so country of birth remains of value for studies of mortality. Country of birth is potentially misleading for people who were born abroad but would classify themselves as of "White" ethnicity. Country of birth provides only a partial solution to the need for data by ethnic group. Nonetheless, in the absence of valid ethnic group data, country of birth analysis provides important insights into health and disease patterns. These analyses are also of value in their own right i.e. not merely as a proxy for ethnic group.

Death rates for all cause mortality, mortality from CHD and from stroke are presented here based on the country of birth. We chose CHD and stroke for reasons explained in chapter 1, but in brief because of the importance of these problems, the known variations in England and Wales and the specialist knowledge of the study team. Our analyses follow a long tradition including work done in England and Wales around the censuses of 1971 (Marmot et al, 1984), 1981 (Balarajan et al, 1990) and 1991 (Wild and McKeigue, 1997; Gill et al 2005). Patterns found in Scotland can be assessed against those already reported.

## **5.2 Methods**

Electronic death registrations were obtained from GROS for people resident in Scotland whose deaths occurred between 1st January 1997 and 31st March 2003. The records included codes for country of birth, which were grouped into ten categories (Scotland, India, Pakistan, Bangladesh, other UK (excluding Ireland), Northern Ireland and the Republic of Ireland, China, Hong Kong, and the rest of the world (excluding areas already mentioned). Other UK is effectively England and Wales.

To estimate the denominator population, commissioned tables were obtained from GROS giving the Scottish population from the 1991 and 2001 Census by country of birth categories shown above and aged 0-24 and then in 5 year age bands up to the age of 90, with one category for those aged over 90. A linear extrapolation between the census years 1991 and 2001 was used to calculate the denominator for the years in-between censuses and after 2001. This was necessary as the South Asian population in Scotland showed sizeable changes over time. Using the census data from only 2001 as denominator showed increasingly implausible results for the years further away from the census year.

Deaths from all causes and where the primary underlying cause was CHD (ICD9 410-414, ICD10 I20-I25) or stroke (ICD9 430-438, ICD10 I60-I67) were studied.

We calculated standardised mortality ratios (SMRs) with 95% confidence intervals for Scottish residents 25 years and over for a 6.25 year period using routine mortality statistics and adjusted census denominators (as above). Standardised mortality ratios were calculated using a comparison against the England and Wales population, which permits us to compare patterns obtained there. The SMRs (indirectly adjusted death rates) for each country of birth group can only be compared to the SMR of the population of England and Wales and not with each other. In addition we calculated age-adjusted rates by direct standardisation but the numbers in some country of birth groups were too small to get valid rates. However, these rates allow direct comparisons between the

different countries of birth groups. The results of the directly age-standardised death rates are presented in the appendix.

A similar analysis was done with the population of Scotland as the standard. Finally, the analysis was repeated and restricted to the age group 25-69. This was done in response to the finding that in Scotland, unlike England and Wales, many of the deaths in the Indian born were in White people probably born in India during the era of empire to 1947. This latter group would be in the oldest age group.

### **5.3 Ethics**

As only anonymised data were used ethical approval was not required.

### **5.4 Results**

#### **5.4.1 Comparison using England and Wales as standard**

Of 362,029 deaths in Scotland in those aged 25 years and above, there were 892, 281 and 14 deaths among those born in India, Pakistan and Bangladesh, respectively. We would have expected the number of deaths in Pakistanis to exceed those in Indians, as the population size of the former is much greater than the latter. These numbers indicated that many of the deaths in Indian born were not in people of Indian ethnic origin. Rather, those deaths were most likely to be in White people born in India when it was part of the British Empire. This point is likely to apply to Northern Ireland born and possible Hong Kong born and, of course, other countries that were part of the British Empire. This would not apply to Pakistan or Bangladesh because these countries were post-Empire creations. Most White people born in India would have been born before 1947, making them at least 50 years of age and most would be much older than that. We re-ran the analysis for the age group 25-69 to see the effects of reducing the number of White people in overseas countries of birth groups e.g. Indian.

As expected, cardiovascular diseases were dominant i.e. 22% of all deaths were primarily caused by CHD and 11.7% by stroke.

Table 5.1 shows that compared with the England and Wales population, all cause SMRs [95% confidence interval] were significantly higher among women born in Scotland, N and Ro Ireland and India; and among men born in Scotland and N and Ro Ireland. Men and women living in Scotland and born in other UK showed a significantly reduced mortality for all causes of deaths compared to those living and born in England and Wales. This is probably a healthy migrant effect. Most of these people are from England and Wales but their mortality is lower than in England and Wales. The numbers of deaths in those born in Bangladesh are too low to interpret.

**Table 5.1 Observed and expected deaths among Scottish residents (25 years and older) from all causes for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, with standardised mortality ratios (95% confidence interval) using death rates from England and Wales as reference**

Country of birth	Observed	Expected	SMR	95% confidence interval	
				lower limit	upper limit
<b>Women</b>					
<i>Engl/Wales(ref)</i>	1764645	1764642.5	100	-	-
Scotland	171488	147781.1	116.0***	115	117
UK (other) +	12827	13452.8	95.3***	94	97
N. Ireland	1336	1236.8	108.0**	102	114
R/Ireland	1716	1453.2	118.1***	113	124
India	419	365.3	114.7**	104	126
Pakistan	110	126.2	87.2	72	105
Bangladesh	5	9.6	51.9	17	121
China	71	81.8	86.8	68	110
Hong Kong	93	89.3	104.2	84	128
Rest of the world	2760	2830.7	97.5	94	101
<b>Men</b>					
<i>Engl/Wales(ref)</i>	1589979	1589975.5	100	-	-
Scotland	152456	125068.9	121.9***	121	123
UK (other)	11889	12923.2	92.0***	90	94
N. Ireland	1204	988.5	121.8***	115	129
R/Ireland	1507	1165.6	129.3***	123	136
India	473	432.2	109.5	100	120
Pakistan	171	207.6	82.4*	70	96
Bangladesh	9	15.4	58.6	27	111
China	67	75.9	88.3	68	112
Hong Kong	103	114.7	89.8	73	109
Rest of the world	3325	3192.8	104.1*	101	108

(\*) Two sided  $p < .05$

(\*\*) Two sided  $p < .01$

(\*\*\*) Two sided  $p < .001$

+ Effectively UK (other) means people born in England and Wales living in Scotland.

As shown in table 5.2, mortality from coronary heart disease (CHD) among men and women born in India, Scotland, Northern and Republic of Ireland and rest of the world (men) was significantly higher than the England/Wales population. Men living in Scotland, and born in other UK and men born in Hong Kong had a significantly reduced mortality rate for CHD compared to England and Wales.

**Table 5.2 Observed and expected deaths among Scottish residents (25 years and older) from coronary heart disease for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, with standardised mortality ratios (95% confidence interval) using death rates from England and Wales as reference**

Country of birth	Observed	Expected	SMR	95% confidence interval	
				lower limit	upper limit
<b>Women</b>					
<i>Engl/Wales(ref)</i>	315887	315888	100	100	100
Scotland	33785	26442	127.8***	126	129
UK (other)	2448	2389	102.5	98	107
N. Ireland	275	223	123.5***	109	139
R/Ireland	369	269	137.4***	124	152
India	91	65	139.4**	112	171
Pakistan	26	19	140.3	92	206
Bangladesh	2	1	134	16	484
China	13	15	89.2	47	153
Hong Kong	18	14	124.1	74	196
Rest of the world	536	503	106.5	98	116
<b>Men</b>					
<i>Engl/Wales(ref)</i>	385800	385797	100	100	100
Scotland	37445	30550	122.6***	121	124
UK (other)	3011	3128	96.3*	93	100
N. Ireland	328	239	137.0***	123	153
R/Ireland	371	288	128.8***	116	143
India	136	108	126.2**	106	149
Pakistan	64	49	129.8	100	166
Bangladesh	2	4	56.5	7	204
China	15	18	81.6	46	135
Hong Kong	17	27	62.5*	36	100
Rest of the world	870	759	114.6***	107	122

(\*) Two sided  $p < .05$

(\*\*) Two sided  $p < .01$

(\*\*\*) Two sided  $p < .001$

Table 5.3 shows that in comparison with England and Wales in most country of birth groups there was a comparatively high SMR for stroke among men and women (Scotland, Ireland (Northern and Republic), India and rest of the world). The small numbers of stroke deaths in Pakistanis and Bangladeshis prevents firm conclusions from being drawn. For UK (other) the mortality was higher in women.



**Table 5.3 Observed and expected deaths among Scottish residents (25 years and older) from stroke for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, with standardised mortality ratios (95% confidence interval) using death rates from England and Wales as reference**

Country of birth	Observed	Expected	SMR	95% confidence interval	
				lower limit	upper limit
<b>Women</b>					
<i>Engl/Wales(ref)</i>	225313	225311	100	-	-
Scotland	24001	18596	129.1***	127	131
UK (other)	1862	1685	110.5***	106	116
N. Ireland	207	158	130.8***	114	150
R/Ireland	261	185	141.2***	125	159
India	63	44	143.9**	111	184
Pakistan	13	12	109.4	58	187
Bangladesh	1	1	89.5	2	499
China	12	10	118	61	206
Hong Kong	13	9	137.9	73	236
Rest of the world	405	337	120.0***	109	132
<b>Men</b>					
<i>Engl/Wales(ref)</i>	136064	136063	100	-	-
Scotland	13934	10447	133.4***	131	136
UK (other)	1061	1083	97.9	92	104
N. Ireland	123	84	146.7***	122	175
R/Ireland	143	103	138.5***	117	163
India	53	34	154.0**	115	201
Pakistan	19	14	140.4	85	219
Bangladesh	2	1	221.9	27	802
China	5	6	79.8	26	186
Hong Kong	9	8	115	53	218
Rest of the world	324	276	117.6**	105	131

(\*) Two sided  $p < .05$

(\*) Two sided  $p < .05$

(\*\*\*) Two sided  $p < .001$

#### 5.4.2 Comparisons using Scotland as the standard

Table 5.4 uses Scotland born as the standard. The pattern is that nearly all country of birth groups have similar or lower all-cause mortality than Scotland though sometimes the differences are not statistically significant, the exception being Republic of Ireland – men born there had higher mortality than the standard. For example, the results show that compared with those born in Scotland, all cause SMRs were significantly lower among men born in India, Pakistan and Bangladesh, and among women born in Pakistan, China

and Northern Ireland. Men and women born in other UK had a significantly reduced mortality for all causes of deaths compared with those born in Scotland.

**5.4 Observed and expected deaths among Scottish residents (25 years and older) from all causes for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, with standardised mortality ratios (95% confidence interval) using death rates among those born in Scotland as reference**

Country of birth	Observed	Expected	SMR	95% confidence interval	
<b>Women</b>					
Scotland	171488	171488.0	100	100	100
UK (other)	12827	15634.1	82.0***	81	83
N. Ireland	1336	1429.5	93.5*	89	99
R/ Ireland	1716	1687.9	101.7	97	107
India	419	430.9	97.2	88	107
Pakistan	110	155.6	70.7***	58	85
Bangladesh	5	11.3	44.1	14	103
China	71	95.3	74.5*	58	94
Hong Kong	93	108.0	86.1	69	105
Rest of the world	2760	3332.1	82.8***	80	86
<b>Men</b>					
Scotland	152456	152456.0	100	99	101
UK (other)	11889	15765.8	75.4***	74	77
N. Ireland	1204	1200.6	100.3	95	106
R/ Ireland	1507	1393.4	108.2**	103	114
India	473	535.3	88.4**	81	97
Pakistan	171	271.6	62.9***	54	73
Bangladesh	9	20.7	43.6**	20	83
China	67	93.0	72.0**	56	91
Hong Kong	103	148.5	69.3***	57	84
Rest of the world	3325	3848.0	86.4***	83	89

(\*) Two sided  $p < .05$

(\*\*) Two sided  $p < .01$

(\*\*\*) Two sided  $p < .001$

At all ages, our results as shown in table 5.5, did not show the pattern of raised CHD mortality previously reported in England and Wales for people born in India, Pakistan and Bangladesh. Rather SMRs were very similar – unlike all cause mortality, there was no advantage to most of the immigrant populations. Again, those born in other UK had a significantly reduced mortality rate for CHD compared to those born in Scotland. The comparatively low CHD mortality in Hong Kong (men only) populations reported in England was seen in Scotland.

**Table 5.5 Observed and expected deaths among Scottish residents (25 years and older) from coronary heart disease for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, with standardised mortality ratios (95% confidence interval) using death rates among those born in Scotland as reference**

<b>Country of birth</b>	<b>Observed</b>	<b>Expected</b>	<b>SMR</b>	<b>95% confidence interval</b>	
<b>Women</b>					
<i>Scotland</i>	33785	33785.0	100	99	101
UK (other)	2448	3054.5	80.1***	77	83
N. Ireland	275	283.0	97.2	86	109
R/ Ireland	369	342.5	107.7	97	119
India	91	84.7	107.4	87	132
Pakistan	26	25.6	101.5	66	149
Bangladesh	2	1.9	103.8	13	375
China	13	18.6	69.8	37	119
Hong Kong	18	19.4	92.8	55	147
Rest of the world	536	650.9	82.4***	76	90
<b>Men</b>					
<i>Scotland</i>	37445	37445.0	100	99	101
UK (other)	3011	3834.0	78.5***	76	81
N. Ireland	328	292.2	112.3*	100	125
R/ Ireland	371	348.1	106.6	96	118
India	136	133.9	101.6	85	120
Pakistan	64	63.6	100.6	77	128
Bangladesh	2	4.6	43	5	155
China	15	22.5	66.8	37	110
Hong Kong	17	34.9	48.7**	28	78
Rest of the world	870	914.1	95.2	89	102

(\*) Two sided  $p < .05$

(\*\*) Two sided  $p < .01$

(\*\*\*) Two sided  $p < .001$

Mortality from stroke among Scottish residents born in India was, compared with Scotland born, slightly but non-significantly raised among men and women in contrast to the larger excesses previously reported for men and women resident in England and Wales and born in India. Small numbers of stroke deaths in Pakistanis and Bangladeshis did not allow conclusions.

Scottish residents born in other UK and rest of the world (men) had a comparatively low SMR.

**Table 5.6 Observed and expected deaths among Scottish residents (25 years and older) from stroke for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, with standardised mortality ratios (95% confidence interval) using deaths among those born in Scotland as reference**

Country of Birth	Observed	Expected	SMR	95% confidence interval	
<b>Women</b>					
<i>Scotland</i>	24001	24001.0	100	99	101
UK (other)	1862	2173.6	85.7***	82	90
N. Ireland	207	204.0	101.5	88	116
R/ Ireland	261	237.8	109.8	97	124
India	63	56.7	111.1	85	142
Pakistan	13	15.7	83	44	142
Bangladesh	1	1.4	69.6	2	388
China	12	13.1	91.8	47	160
Hong Kong	13	12.3	106	56	181
Rest of the world	405	434.2	93.3	84	103
<b>Men</b>					
<i>Scotland</i>	13934	13934.0	100	98	102
UK (other)	1061	1443.1	73.5***	69	78
N. Ireland	123	111.6	110.2	92	131
R/ Ireland	143	137.4	104.1	88	123
India	53	46.2	114.8	86	150
Pakistan	19	18.3	103.6	62	162
Bangladesh	2	1.2	162.9	20	588
China	5	8.4	59.7	19	139
Hong Kong	9	10.6	85.3	39	162
Rest of the world	324	366.4	88.4*	79	99

(\*) Two sided  $p < .05$

(\*\*\*) Two sided  $p < .001$

### 5.4.3 Comparisons within the age group 25-69 using both England and Wales, and Scotland, as the standard

Tables 5.7-5.10 restrict analysis to the 25-69 year age group, and to all causes and CHD (too few strokes for analysis). In this analysis, as stated before, there would be relatively few White people in country of birth groups such as India or Hong Kong. Obviously the numbers of deaths are lower. Table 5.7 shows the same kind of pattern described under 5.1.

**Table 5.7 Observed and expected deaths among Scottish residents (25 – 69 years) from all causes for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, with standardised mortality ratios (95% confidence interval) using death rates from England and Wales as reference**

County of Birth	Observed	Expected	SMR	95% confidence interval	
<b>Women</b>					
<i>Engl/Wales(ref)</i>	304571	304570	100	100	100
Scotland	36585	27748	131.8***	130	133
UK (other)	2415	2627	91.9***	88	96
N. Ireland	216	204	106.1	92	121
R/Ireland	253	235	107.6	95	122
India	107	86	124.9*	102	151
Pakistan	78	68	114.9	91	143
Bangladesh	3	3	94.4	19	276
China	12	17	72.2	37	126
Hong Kong	25	33	76.5	50	113
Rest of the world	553	596	92.8	85	101
<b>Men</b>					
<i>Engl/Wales(ref)</i>	480417	480415	100	100	100
Scotland	57352	41344	138.7***	138	140
UK (other)	3888	4293	90.6***	88	93
N. Ireland	365	308	118.7**	107	131
R/Ireland	426	291	146.4***	133	161
India	173	173	99.8	85	116
Pakistan	121	132	91.4	76	109
Bangladesh	6	12	50.8	19	111
China	26	26	99.9	65	146
Hong Kong	62	67	92.4	71	118
Rest of the world	884	824	107.3*	100	115

(\*) Two sided  $p < .05$

(\*\*) Two sided  $p < .01$

(\*\*\*) Two sided  $p < .001$

Table 5.8 shows that the pattern of CHD mortality by country of birth demonstrated in England and Wales, not shown clearly in table 5.2, is indeed present in Scotland. The highest rates of CHD mortality are in the Indian subcontinent populations (women and men from Pakistan), and though not precisely estimated the rates were low in Chinese

and Hong Kong born. The excess mortality in Scottish born in comparison to England and Wales is noteworthy and is accentuated for CHD.

**Table 5.8 Observed and expected deaths among Scottish residents (25 – 69 years) from coronary heart disease for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, with standardised mortality ratios (95% confidence interval) using death rates from England and Wales as reference**

Country of birth	Observed	Expected	SMR	95% confidence interval	
<b>Women</b>					
<i>Engl/Wales(ref)</i>	38017	38018	100	99	101
Scotland	5466	3526	155.0***	151	159
UK (other)	282	317	89	79	100
N. Ireland	36	25	143.2*	100	198
R/Ireland	54	33	165.6***	124	216
India	23	11	209.8**	133	315
Pakistan	19	7	258.4***	156	404
Bangladesh	2	0	741.8	90	2680
China	1	2	50	1	279
Hong Kong	3	3	88	18	257
Rest of the world	66	68	96.8	75	123
<b>Men</b>					
<i>Engl/Wales(ref)</i>	118464	118462	100	99	101
Scotland	13944	10255	136.0***	134	138
UK (other)	967	1045	92.6*	87	99
N. Ireland	111	75	148.4***	122	179
R/Ireland	100	76	131.7**	107	160
India	61	44	137.8*	105	177
Pakistan	49	31	159.9**	118	211
Bangladesh	1	3	37.7	1	210
China	3	6	48.5	10	142
Hong Kong	8	16	51.3	22	101
Rest of the world	209	182	114.7	100	131

(\*) Two sided  $p < .05$

(\*\*) Two sided  $p < .01$

(\*\*\*) Two sided  $p < .001$

Table 5.9 is standardised to the Scottish population, and it shows all cause mortality is favourable (and statistically so in most comparisons) in comparison to Scotland born in every group except Republic of Ireland men.

**Table 5.9 Observed and expected deaths among Scottish residents (25 – 69 years) from all causes for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, with standardised mortality ratios (95% confidence interval) using death rates from Scottish born in Scotland as reference**

	Observed	Expected	SMR	95% confidence interval	
<b>Women</b>					
Scotland (ref)	36585	36585	100	99	101
Engl/Wales	304571	401483	75.9***	76	76
UK (other)	2415	3462	69.8***	67	73
N. Ireland	216	268	80.5**	70	92
R/Ireland	253	311	81.5***	72	92
India	107	113	94.5	77	114
Pakistan	78	89	87.5	69	109
Bangladesh	3	4	72.1	15	211
China	12	22	55.0*	28	96
Hong Kong	25	43	58.3**	38	86
Rest of the world	553	782	70.7***	65	77
<b>Men</b>					
Scotland (ref)	57352	57352	100	99	101
Engl/Wales	480417	667337	72.0***	72	72
UK (other)	3888	5991	64.9***	63	67
N. Ireland	365	429	85.2**	77	94
R/Ireland	426	400	106.4	97	117
India	173	240	72.2***	62	84
Pakistan	121	185	65.4***	54	78
Bangladesh	6	17	36.3**	13	79
China	26	36	71.9	47	105
Hong Kong	62	94	66.0***	51	85
Rest of the world	884	1157	76.4***	71	82

(\*) Two sided  $p < .05$

(\*\*) Two sided  $p < .01$

(\*\*\*) Two sided  $p < .001$

Table 5.10 shows that in comparison to Scotland born, the highest SMRs for CHD were in the Indian subcontinent women though none are statistically significant and the lowest SMRs were in Chinese/Hong Kong born (statistically significant in Hong Kong men).

**Table 5.10 Observed and expected deaths among Scottish residents (25 – 69 years) from coronary heart disease for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, with standardised mortality ratios (95% confidence interval) using death rates from Scottish born in Scotland as reference**

Country of birth	Observed	Expected	SMR	95% confidence interval	
<b>Women</b>					
Scotland (ref)	5466	5466	100	97	103
Engl/Wales	38017	59015	64.4***	64	65
UK (other)	282	494	57.1***	51	64
N. Ireland	36	39	92.2	65	128
R/Ireland	54	50	107.6	81	140
India	23	17	135	86	203
Pakistan	19	12	164.3	99	257
Bangladesh	2	0	458.2	55	1655
China	1	3	32.3	1	180
Hong Kong	3	5	55.6	11	163
Rest of the world	66	106	62.1***	48	79
<b>Men</b>					
Scotland (ref)	13944	13944	100	98	102
Engl/Wales	118464	161191	73.5***	73	74
UK (other)	967	1427	67.7***	64	72
N. Ireland	111	102	108.9	90	131
R/Ireland	100	103	97.2	79	118
India	61	60	101.5	78	130
Pakistan	49	42	117.2	87	155
Bangladesh	1	4	27.7	1	154
China	3	8	35.8	7	104
Hong Kong	8	21	37.5**	16	74
Rest of the world	209	248	84.2*	73	96

(\*) Two sided  $p < .05$

(\*\*) Two sided  $p < .01$

(\*\*\*) Two sided  $p < .001$

## 5.5 Conclusions

### 5.5.1 Methods

The interpretation of country of birth analyses done using census denominators and death certificate data is complex, and has been discussed by Gill et al (2005). In brief, the count of death (numerator) may be wrong because (a) some people may enter the country and die here but are not normally resident (b) some people who are normally resident may prefer to leave the country to die elsewhere. Even if their deaths are notified here, which is possible for administrative and other reasons, the cause of death may not be known. These effects are seldom studied and hitherto have not been examined in Scotland. The limitations of the cause of death information are well known. Diagnosing the cause of



death without an autopsy is an imprecise part of the art of medicine. Whether there are variations by country of birth in the accuracy of the cause of death is not known, although on first principles one might anticipate that there might be, for example, the medical history may be less accurate in those people born overseas who do not speak English as a first language.

The numerator is turned into a disease rate using a census denominator. The census is only held once every 10 years, and is most imprecise for the young adults and inner city populations (though the published data are adjusted for undercounts). There are, therefore, likely to be variations in the precision of denominators by country of birth group. In this study we have used 1991 and 2001 censuses to estimate the correct population sizes by simple linear extrapolation. While this is a conceptually straightforward action, we are not aware of others who have taken advantage of such a safeguard.

Sometimes country of birth is not being studied solely as a marker of environmental exposures in childhood but as marker of ethnicity. In our analysis we noted that the Pakistani ethnic group population of Scotland was nearly twice the size of the Indian ethnic group and of similar age structure as Indians yet there were about three times as many deaths in Indians as in Pakistanis. Scotland had a major role in the British rule of India. People born in the Indian subcontinent of White ethnic groups would have been there before and until about 1947 (when Pakistan was created as a separate nation and India gained independence). In the India-born group, but not in Pakistan or Bangladeshi born, a high proportion of people more than 56 years of age would be of non-Indian ethnic groups. In corresponding analyses in England this would not matter so much because of the comparatively large ethnic Indian population in relation to Indian born White people. In Scotland this matters more. While 98.9% of White Scottish people were born in Scotland or England, this still means over 1% (about 50,000 people) were born elsewhere. Some of them, especially the older ones contributing most deaths – were born in British colonies. Indeed, an analysis of the census data confirmed this, with most people born in India who were > 65 years of age in 2001 being White. Only in the middle range of age-groups- roughly, 40-65- is country of birth a reasonable marker of ethnicity.

Large numbers of deaths were in the rest of the world group – this group is worthy of deeper exploration in future analyses particularly focussing on African origin populations.

Given appropriate interpretation the data are of potential value not least because they are available for a long period, as country of birth data have been collected at each census. This value needs to be gauged in relation to similar analysis from the linkage data in the next chapter.

### **5.5.2 Conclusions: results**

The first and most important result is that analysis by country of birth has been achieved so that its value can be assessed. The number of deaths in Scottish residents born outside Scotland is substantial – more than 10% of the total. Most of these were in people born in other parts of the UK or the Republic of Ireland. There are also substantial numbers of deaths in the 'rest of the world' group. In this demonstration project focused on CHD, we have not unpicked this large group but it may be valuable, both for minority ethnic populations (e.g. of African ancestry) and for White minorities e.g. Eastern Europeans. For specific causes the number of deaths in the minority ethnic groups focused on here – South Asians and Chinese – was relatively small but the findings can be interpreted well because of similar data from England and Wales. Minority populations are ageing and over the next 5-15 years the number of deaths will climb rapidly.

Despite all the caveats the findings are intriguing and ought to have an impact on public health policy. For example, for all causes we see Scottish born residents fare comparatively badly (table 5.1, 5.4, 5.7, 5.9) and are on a par with or marginally better

than Scottish residents born in the Republic of Ireland. That living in Scotland, in itself, is not necessarily bad for mortality is suggested by the lower death rates in a range of populations including people born in other UK countries i.e. England and Wales. There is the possibility of a healthy migrant effect i.e. the healthiest people migrate (whether into or out of Scotland). This possibility needs to be borne in mind. People born outside of Scotland (the so-called salmon bias) may also return to their country of birth to die and hence artificially lowering the death rate. Overall, this analysis is good news for Scottish residents born outside Scotland and gives cause for concern about those born and being raised in Scotland.

The tables on CHD (5.2, 5.5, 5.8) also show how the Scottish born are highly prone to this problem. However, we see, especially in the younger group (table 5.8) major variations by country of birth – with excesses in Indian subcontinent groups and deficits in far Eastern ones – that echo those demonstrated in England and Wales and other countries. For stroke (tables 5.3 and 5.6) the number of cases are small but it is likely substantial variations will arise in time. The comparatively low rates in those born in other UK are a pointer to achieving lower mortality from stroke. This work suggests that South Asians in Scotland already are, or are likely to become, highly prone to death from cardiovascular disease, but are less susceptible to other causes of death, as found in England and Wales.

### **5.5.3 Conclusions – overall**

These results provide data, using the proxy of country of birth, on ethnic variations in mortality in Scotland. As in other countries, major variations have been demonstrated. Comparisons within the Scottish population might be interpreted as reassuring for those born outside Scotland, since they do not show the wide inequalities in CHD mortality by country of birth reported in England and Wales. However, the use of England and Wales as a comparison group shows a substantial excess of CHD risk among South Asians in Scotland, comparable to that reported in England and Wales. The high risks are clearer in the age group 25-69. This analysis highlights the importance of the choice of reference and age groups in assessing relative mortality. In relation to control of mortality both overall, and for cardiovascular disease, either in South Asians in Scotland, or those born in Scotland, there is no room for complacency. The methods developed here can be applied to a broader range of causes, and to a broader range of country of birth groups. Furthermore, using our approach of estimating between census denominators, we could analyse data for a much longer period of time, say 1969-2004 i.e. a 25 year period. The country of birth approach is likely to be complementary to the linkage methods described in chapters 6 and 7.

**Appendix**  
Table A5. 1

Descriptive statistics: counts of deaths for all causes, CHD, and stroke

By country of birth

Country	all deaths	CHD	stroke	% of death by CHD	% of death by stroke
Scotland	323944	71230	37935	22.0	11.7
UK (other)	24716	5459	2923	22.1	11.8
N. Ireland	2540	603	330	23.7	13.0
R/ Ireland	3223	740	404	23.0	12.5
India	892	227	116	25.4	13.0
Pakistan	281	90	32	32.0	11.4
Bangladesh	14	4	3	28.6	21.4
China	138	28	17	20.3	12.3
Hong Kong	196	35	22	17.9	11.2
Rest of the world	6085	1406	729	23.1	12.0
<i>Total</i>	<i>362029</i>	<i>79822</i>	<i>42511</i>	<i>22.0</i>	<i>11.7</i>

(a) By country of birth and sex

	all deaths	CHD	stroke	% CHD	% stroke
Women					
Scotland	171488	33785	24001	19.7	14.0
UK (other)	12827	2448	1862	19.1	14.5
N. Ireland	1336	275	207	20.6	15.5
R/ Ireland	1716	369	261	21.5	15.2
India	419	91	63	21.7	15.0
Pakistan	110	26	13	23.6	11.8
Bangladesh	5	2	1	40.0	20.0
China	71	13	12	18.3	16.9
Hong Kong	93	18	13	19.4	14.0
Rest of the world	2760	536	405	19.4	14.7
Men					
Scotland	152456	37445	13934	24.6	9.1
UK (other)	11889	3011	1061	25.3	8.9
N. Ireland	1204	328	123	27.2	10.2
R/ Ireland	1507	371	143	24.6	9.5
India	473	136	53	28.8	11.2
Pakistan	171	64	19	37.4	11.1
Bangladesh	9	2	2	22.2	22.2
China	67	15	5	22.4	7.5
Hong Kong	103	17	9	16.5	8.7
Rest of the world	3325	870	324	26.2	9.7

The next six tables give the directly standardised rates per 1000.

**Table A5.2 Crude and directly age adjusted mortality per 1000 among Scottish residents 25 years of age and older and from all causes for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, using the population structure from England and Wales(1) and Scottish born in Scotland(2) as reference group**

Country	Crude rate	Adjusted for English population (1)	Adjusted for Scottish population (2)	95% - Confidence Interval for adjusting by Scottish population (2)	
				Lower	Upper
<b>Female</b>					
Scotland	17.05	14.81	14.35	14.18	14.51
Other UK	12.32	12.08	11.63	11.14	12.13
N. Ireland	16.48	13.66	13.17	11.44	14.91
R/Ireland	24.14	14.81	14.24	12.59	15.88
India	15.02	14.45	13.92	10.75	17.1
Pakistan	3.54	8.08	8.12	3.64	12.59
Bangladesh	2.3	9.28	9.46	-18.66	37.58
China	9.05	10.95	10.5	4.55	16.44
Hong Kong	5.75	19.44	18.32	9.6	27.04
Rest of the world	9.04	12.39	11.91	10.83	12.99
<b>Male</b>					
Scotland	17.4	21.98	21.44	21.17	21.7
Other UK	12.03	17.2	16.66	15.92	17.39
N. Ireland	16.87	22.88	22.21	19.23	25.2
R/Ireland	29.66	23.93	23.32	20.45	26.2
India	15.41	20.59	19.95	15.38	24.51
Pakistan	5.06	12.09	12	5.69	18.3
Bangladesh	2.78	23.16	21.84	-33.19	76.87
China	10.55	15.97	15.57	5.95	25.19
Hong Kong	5.97	15.43	15.2	6.34	24.05
Rest of the world	12.56	19.34	18.81	17.22	20.41

**Table A5.3 Crude and directly age adjusted mortality per 1000 among Scottish residents 25 years of age and older and from CHD for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, using the population structure from England and Wales(1) and Scottish born in Scotland(2) as reference group**

Country	Crude rate	Adjusted for English population (1)	Adjusted for Scottish population (2)	95% - Confidence Interval for adjusting by Scottish population (2)	
				Lower	Upper
<b>Female</b>					
Scotland	3.36	2.9	2.8	2.73	2.88
Other UK	2.35	2.31	2.22	2	2.44
N. Ireland	3.39	2.8	2.69	1.89	3.5
R/Ireland	5.19	3.11	2.99	2.23	3.76
India	3.26	3.06	2.97	1.45	4.49
Pakistan	0.84	2.08	2.06	-0.34	4.47
Bangladesh	0.92	6.2	6.67	-20.44	33.77
China	1.66	2.1	1.96	-0.69	4.61
Hong Kong	1.11	4.03	3.8	-1.29	8.9
Rest world	1.76	2.41	2.31	1.82	2.8
<b>Male</b>					
Scotland	4.27	5.3	5.2	5.06	5.33
Other UK	3.05	4.27	4.17	3.79	4.55
N. Ireland	4.6	6.08	5.94	4.32	7.57
R/Ireland	7.3	5.7	5.57	4.13	7.02
India	4.43	5.08	5.02	2.78	7.26
Pakistan	1.89	4.45	4.41	0.61	8.21
Bangladesh	0.62	2.36	2.46	-7.18	12.09
China	2.36	4.38	4.14	-1.61	9.9
Hong Kong	0.99	3.57	3.4	-1.5	8.31
Rest of the world	3.29	5.1	4.99	4.13	5.84

**Table A5.4 Crude and directly age adjusted mortality per 1000 among Scottish residents 25 years of age and older and from CVD/stroke for 6.25 years (Jan 1997 – Mar 2003), by country of birth and sex, using the population structure from England and Wales(1) and Scottish born in Scotland(2) as reference group**

Country	Crude rate	Adjusted for English population (1)	Adjusted for Scottish population (2)	95% - Confidence Interval for adjusting by Scottish population (2)	
				Lower	Upper
<b>Female</b>					
Scotland	2.39	2.01	1.91	1.85	1.97
Other UK	1.79	1.72	1.62	1.44	1.81
N. Ireland	2.55	2.02	1.91	1.25	2.56
R/Ireland	3.67	2.2	2.07	1.44	2.7
India	2.26	2.28	2.16	0.84	3.48
Pakistan	0.42	1.27	1.25	-0.69	3.19
Bangladesh	0.46	1.44	1.3	-4.55	7.16
China	1.53	1.76	1.66	-0.69	4.01
Hong Kong	0.8	2.63	2.49	-1.47	6.46
Rest of the world	1.33	1.89	1.78	1.35	2.21
<b>Male</b>					
Scotland	1.59	2.24	2.14	2.05	2.23
Other UK	1.07	1.7	1.62	1.37	1.87
N. Ireland	1.72	2.65	2.49	1.38	3.61
R/Ireland	2.81	2.27	2.18	1.26	3.1
India	1.73	2.66	2.53	0.69	4.37
Pakistan	0.56	2.21	2.12	-1.19	5.43
Bangladesh	0.62	16.62	15.28	-37.1	67.66
China	0.79	1.41	1.34	-1.67	4.34
Hong Kong	0.52	1.75	1.73	-1.45	4.9
Rest of the world	1.22	2.04	1.95	1.39	2.5

## 6 Adding ethnicity codes to routine health data by record linkage to census records: surmounting the confidentiality challenge

POVEY C<sup>2</sup>, FISCHBACHER CM<sup>2</sup>, CHALMERS J<sup>2</sup>, KNOWLES D<sup>2</sup>, STEINER M<sup>1</sup>, JAMIESON J<sup>3</sup>, SCRIMGEOUR P<sup>4</sup>, BHOPAL R<sup>1</sup>.

(<sup>1</sup>Public Health Sciences, University of Edinburgh, <sup>2</sup>Information and Statistics Division, National Services Scotland, <sup>3</sup>National Resource Centre for Ethnic Minorities, NHS Health Scotland, <sup>4</sup>GROS.)

In association with D Orr, G. Mueller, J Fuchs, I Mate and S Wallace from GROS.

### Abstract

**Introduction.** The Race Relations (Amendment) Act 2000 and NHS policy require health services to publish data showing they are promoting racial equality and reducing ethnic inequalities. However, routine data sources in Scotland rarely include a patient's ethnicity. One approach is to use record linkage techniques. Linkage using personal identifiable information has to be dealt with using a strict protocol to prevent even the remotest possibility of leakage of personal information. This chapter describes our solution.

**Methods.** The health data are maintained by Information Services of NHS National Services Scotland (ISD), and the census data are maintained by the General Register Office for Scotland (GROS). A copy of the Community Health Index (CHI) dataset (the Scottish equivalent to the English NHS number) was used. This contains the demographic data and CHI number for every individual registered with a General Practitioner or in receipt of screening services. A one-way encryption algorithm was used to convert the CHI number into a code in a way that ensured it would be virtually impossible to reverse the process. A data extract with relevant personal identifiers and a census form number was taken from GROS's census file, the census number was encrypted and the file was linked to the encrypted CHI dataset with the same personal identifiers using probability matching. The personal identifiers were then stripped off. A further extract containing ethnic code and an encrypted census number was joined to the first file using the encrypted census number, which was then discarded leaving the ethnicity code and the encrypted CHI number. The relevant hospital and death datasets were then linked via the encrypted CHI number, after which the personal identifiers and other information not needed for analysis were discarded.

**Results.** Overall, 94% of the 4.9 million census records were matched to a CHI record with an estimated false positive rate of less than 0.1 %. The majority of matching failures are thought to be caused by poor transcription of names from the census form, and variations between the census data and NHS records for items such as date of birth and the detailed spelling of names.

**Conclusion.** In our judgement the technique described meets ethical, professional and legal concerns about the linkage of census and health data and is transferable outside Scotland. The linkage makes it possible to do a cohort analysis of morbidity and mortality on individuals included in the census.

## **6.1 Introduction**

The limitations and strengths of name search, imputation and country of birth analysis were considered in earlier chapters. Notwithstanding the value of the data captured by such methods, they are not based on individual data including self-reported ethnic group. The reasoning behind the decision to use record linkage has been discussed in chapter one. Essentially, it provides a practical way to add self-assigned ethnicity to individual medical and death records, although strictly speaking it captures the ethnicity assigned by the person who fills in the census form, usually the head of household.

There was an understandable concern about linking health service data to census data. This approach was only considered acceptable on the condition that at no stage in the process would an identified individual's response to the Census ever be linked to their health records and that no identified health records would ever be linked to Census information. Linkage using personal identifiable information was dealt with using a strict protocol to prevent even the remotest possibility of leakage of personal information. Prior to the matching starting we considered that the analytical work would be satisfactory if 90% or more of records overall could be matched, and 80% or more for the minority ethnic groups. These conditions were challenging and required an innovative approach.

## **6.2 Methods**

### **6.2.1 Preparatory work - provision of secure facilities, computer and software**

The General Register Office for Scotland (GROS) provided a locked room within their offices at Ladywell House to which access could be controlled jointly by ISD and GROS. This was to provide a secure area for the processing of the data which ensured that none of the census data were removed from GROS's offices, and additionally that none of the ISD data would be available to GROS staff. The processing was performed on a stand-alone, password-protected, personal computer.

### **6.2.2 Existing record linkage**

Record linkage has been performed within ISD for many years using well established probability matching techniques to link the data on different datasets which relate to the same individual (Kendrick and Clarke, 1993). On the basis of forename and surname, previous names, postcode and date of birth, computer algorithms are used to assess the probability that two records relate to the same person. As long as these "linking" fields are present in two or more datasets, it is possible to link all the information relating to an individual on one dataset to information relating to the same (or probably the same) individual on the other datasets. The datasets then form one linked dataset which can be analysed.

ISD's main dataset consists of all episodes of hospital discharge from acute (non-maternity and non-psychiatric) hospitals (SMR01) and death data provided by GROS. Thus, information from all episodes of acute care for an individual can be joined together right up to the time of their death, and information derived from the death registration process can be included. This dataset is referred to as the "SMR01-death linked dataset" or SMR01 for short.



### **6.2.3 Linking the SMR01-death linked dataset to the census-derived ethnicity code**

For each individual on ISD's existing linked dataset, we wished to add a code for ethnicity derived from the census. The work was based on the use of the Community Health Index (CHI) number as a unique identifier. Each person in Scotland has a unique CHI number which is either generated shortly after birth when information about the baby is sent to the local NHS Health Board, or on first registration with a GP. Prior to 1997, a person's CHI number could change if they moved to a different Health Board area of Scotland, but since 1997, the CHI number has been fixed regardless of where they move within Scotland.

The complete linkage exercise was performed as summarised in the figure (page 67) and in the following independent steps.

#### **Step 1. Production of encrypted codes derived from the census ID and CHI numbers using a one-way hash algorithm.**

We converted the census ID and CHI numbers in every file into an encrypted code, thus providing additional assurance that these numbers could not be disclosed deliberately or accidentally to any outside agency or individual. This was done using a publicly available one-way encryption hash algorithm of the type acknowledged by cryptographers as secure -SHA256. One-way encryption means that a particular CHI number can be converted to a new identifier but it is virtually impossible to reverse the process and derive the CHI number from the encrypted revision.

The encryption of the Census ID was done via the substitution of a unique serial number. The start position for serial number generation was chosen at random.

#### **Step 2. Creating CHI to Census identifier look up**

##### **Census extract**

A data extract was taken from the census file where each record contained the following fields:

1. Anonymised census person identifier (one way encrypted)
2. Names
3. Gender code
4. Date of birth
5. Postcode of enumeration
6. Address (including postcode)

This extract was from the "post-edit pre-imputation" (PEPI) data set. This dataset contains data which have been cleaned and edited but does not include One Number Census (ONC) changes, such as imputed records for households and individuals estimated to have been missed by the census.

##### **CHI (2001) extract**

The CHI data extract for May 2001 was used as it was considered to be the most compatible with the census performed on 29 April 2001.

A file with the following fields was produced from the CHI dataset:

1. Encrypted CHI number
2. All surnames
3. Forenames

4. Date of birth
5. Postcodes (up to two - current and previous)
6. Addresses (up to two – current and previous)

This file contained approximately 5.3 million records, which is the population of Scotland currently registered with a GP. This is slightly greater than the Scottish population because the CHI data set includes people who have left the UK without notification. This extract contained approximately 2% previously undetected duplicated records. When this file was linked to the census extract 85% of the census records matched.

### **CHI (2002) extract**

Another CHI data extract was made from the May 2002 data set to help capture the 15% of census records unmatched from the first linkage.

A file with the following fields was produced from the CHI dataset:

1. Encrypted CHI number (current and historical)
2. CHI status (current, historical, corrected, deleted etc)
3. All surnames (past and present)
4. Forenames
5. Date of birth
6. Date of death
7. Postcodes
8. Addresses
9. NHS number (for people born in Scotland this is their birth registration number)

This file contained approximately 9.5 million records. This is considerably greater than the Scottish population of approximately 5 million. This difference is because the full CHI dataset includes records of people who have died since CHI was started in the 1970s, those who have migrated out of Scotland, those who moved into another CHI administration area within Scotland, duplicate records, and faulty (but fixed) records.

The second extract should capture late GP registrations and any other unusual processing. Additional fields were included because the manual checking was much more rigorous in the second match.

This linkage yielded a further 9% matches, i.e. 94% overall.

A lookup file was created from this linkage exercise. Each lookup entry consisted of the encrypted CHI number and the encrypted census identification number. This file provided the 'bridge' for applying census data to health records.

### **Step 3. Creation of the SMR dataset**

ISD created an extract file from the existing SMR01-death linked dataset. The criterion for selection was any mention of coronary heart disease during any inpatient stay during 1999-2003, in any of the six possible diagnostic positions. To this file was added information on hospitalisations of these individuals back to 1983. This extracted file contained all hospital admissions and/or deaths. These extracted records were then linked to the CHI dataset to acquire the CHI (encrypted) number. This linkage was done to add missing CHI numbers and check the linkage methodology, particularly calculating a false positive rate by comparing CHI numbers added clerically with CHI numbers added by the linkage process.

#### **Step 4. Merge of SMR data with census ethnicity data**

Using the look up file, a census extract containing ethnicity code and encrypted CHI number was created. The SMR01-death linked file with attached encrypted CHI number was taken to GROS, and matched with this census extract. Once the matching process was complete, the encrypted CHI number was discarded to leave an anonymised dataset based on the SMR01-death linked dataset with the addition of an ethnicity code.

This process ensures that unauthorised people did not see named personal data.

#### **Step 5. Minimisation of data set for analysis**

The relevant data protection legislation suggests that such data should be “acceptably anonymised” prior to analysis. Therefore, prior to analysis dates of birth were converted to ages, residential information was coded at aggregate geographical level (see below) and deprivation scores or categories assigned.

The variables remaining on the analysis file were these:

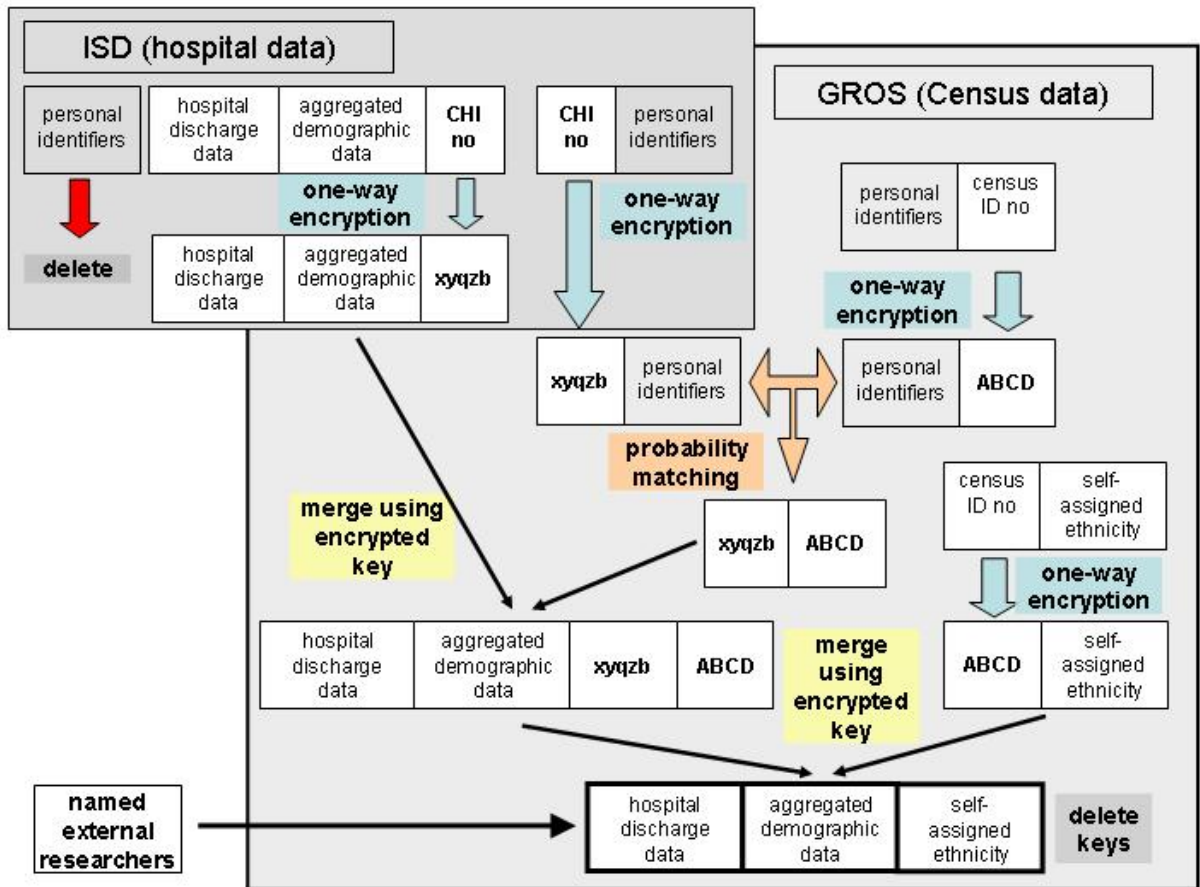
1. Person ID (consecutive)
2. CIS (Hospital stay no., consecutive)
3. Month/Year of admission for all hospital admissions and deaths
4. Survival in days from first MI admission to death
5. LOS (length of stay) in days
6. ICD diagnostic codes (6)
7. Age in single years
8. Sex
9. Deprivation categories (SIMD, Carstairs 5/7)
10. High level geographies (urban/rural, Glasgow/non-Glasgow, health board)
11. Ethnic code
12. COB (country of birth)

Access to the analysis file was restricted to 2 named researchers and was carried out in the secure setting created by GROS for this project.

### **6.3 Ethics**

The work was approved by MREC (Scotland), PAC (ISD) and GROS.

Figure 6.1 Diagrammatic conceptualisation of the linkage procedure



## 6.4 Results

Table 6.1 demonstrates the effectiveness of the matching process (94% overall). The table percentages are all proportions of the total number of census records. Exact means that there was fairly standard exact (deterministic) matching. Probability means that there was probability matching performed on residual unmatched census records (Newcombe, 1988).

**Table 6.1. Effectiveness of matching process**

	<b>Exact</b>	<b>Probability</b>	<b>Combined exact &amp; probability</b>
CHI (2001) – Census	69%	16%	85%
CHI (2002) – Census	2%	7%	9%
Combined 2001 & 2002 - Census	71%	23%	94%

This group of linkages used a methodology that severely restricts false positive matches while still yielding a high overall match rate. The methodology works by grouping the matched pair records by ‘quality’ of match, by using automatic procedures, which mimic what a manual (human) checker would do. Some groups are very robust, ranging down to some that are more questionable. Each group was checked manually using an established method (Newcombe, 1988). The performance of the matching process, for each ethnic group, is shown in Table 6.2. The match exceeded 80% in every ethnic group i.e. meeting our pre-specified standard of 80%.

**Table 6.2. Numbers by ethnic group in census, and matching performance**

<b>Ethnic group</b>	<b>Unmatched</b>	<b>Matched</b>	<b>Total</b>	<b>% matched</b>
White Scottish	201306	3890972	4092278	95.1%
Other White British	24976	321278	346254	92.8%
White Irish	4332	41078	45410	90.5%
Other White	9943	61279	71222	86.0%
Indian	1651	11690	13341	87.6%
Pakistani	3580	23660	27240	86.9%
Bangladeshi	261	1463	1724	84.9%
Other South Asian	980	4526	5506	82.2%
Chinese	2459	12161	14620	83.2%
Caribbean	196	1400	1596	87.7%
African	723	3598	4321	83.3%
Black Scottish or Other Black	139	878	1017	86.3%
Any Mixed Background	1139	10429	11568	90.2%
Other Ethnic group	1446	7115	8561	83.1%
Non Resident Students	2887	33272	36159	92.0%
No Response	37768	180872	218640	82.7%
Scotland	293786	4605671	4899457	94.0%

The figures relate to the PEPI Census data and are therefore not identical to the final adjusted figures published by GROS.

The results of the SMR01-death/CHI linkage give an insight into the false positive rate for the linkage methodology. 630121 SMR01-death records had a non-blank entry in CHI number field, i.e. the record came to ISD with a CHI number. 629493 yielded the same CHI number on matching to CHI dataset using the linkage methodology. 628 records had differing CHI numbers, and are therefore potential false linkages from this linkage. 125 of the 628 had CHI entries on SMR01-death which do not appear in the CHI data set. 503 (628-125 entries) false positives might have arisen in the linkage process.

Applying the worst assumption that all the 503 CHIs derived from the linkage were incorrect and the CHIs as applied by the hospitals were correct, this implies the worst possible false positive linkage rate of 0.08 %.

## **6.5 Discussion**

We succeeded in our goals, achieving our pre-set standards for linkage. The proportion of records which were ascribed an ethnicity code was smaller than that normally achieved for record linkage of files within ISD. There are two reasons for this. Firstly, the bulk of linkages done within ISD do not require great attention to be paid to the rate of false positive matching. It is evident that the ethnic population is so small compared with the whole Scottish population that the false positive match rate is critical. Consequently the matching and checking methodology appropriate for administrative matching was used instead of the usual standard for epidemiological research. This reduces the expected apparent match rate from 98% to 94%.

Secondly, the method used for collecting and capturing Census information was challenging. Census forms are completed by members of the public and processed electronically using a combination of optical recognition and keying from form images. The success of these processes depends on a reasonable degree of legibility. In addition, it is possible that items such as date of birth may be wrongly recorded by a third party onto NHS records and never checked by the individual, but perhaps might be more accurate on the census form. Similarly, the detailed spelling of a name, particularly if it is associated with a minority ethnic group, is likely to be prone to error when transcribed by a third party onto NHS records. Such errors would lead to a different accuracy of linkage depending on ethnic group. Nevertheless, it is believed that the overall degree of matching accuracy and completeness is adequate for the analyses described in the next chapter. Moreover, the methods created here provide an innovative, cost-effective and ethical way to extract information by ethnic group from Scotland's health databases. In due course the methods might be streamlined and improved.

The epidemiological and public health value of this linkage was assessed for CHD as described in the next chapter.

## 7 Cohort analysis of CHD using the linked database

FISCHBACHER CM<sup>2</sup>, STEINER M<sup>1</sup>, BHOPAL R<sup>1</sup>, POVEY C<sup>2</sup>, CHALMERS C<sup>2</sup>, JAMIESON J<sup>3</sup>, KNOWLES D<sup>2</sup>

(<sup>1</sup>Public Health Sciences, University of Edinburgh, <sup>2</sup>Information and Statistics Division, National Services Scotland, <sup>3</sup>National Resource Centre for Ethnic Minorities, NHS Health Scotland.)

In association with Kate McIntyre of the University of Glasgow and Niamh Murphy of the Western Infirmary, Glasgow

### Abstract

**Introduction and objectives.** There is a shortage of data on disease incidence and outcomes following disease by ethnic group. The value of the linked data set was tested by analysing incidence and survival from acute myocardial infarction in South Asians. The prior expectation was that incidence would be higher and survival lower in the South Asian group.

**Methods.** The study cohort was the population at census 2001 who were matched to the CHI i.e. 94% of the census population. SMR01 data for people who had had a diagnoses of CHD since 1981 were extracted. The SMR01 data set was reorganised by continuous inpatient stays (CIS), and length of stay calculated from date of admission to date of discharge or death. Follow up mortality data were available to December 2003. For each matched person with CHD we obtained all SMR01 data back to 1981. We analysed all AMI events (attack rates), new AMI events (incidence rates), and survival following AMI. Survival was adjusted for sex, age, prior history of diabetes, and deprivation index in a logistic regression model, and separately in a Cox regression model. Comparisons were between South Asians and non-South Asians (for brevity this is not specified below).

**Results.** The dataset consisted of 2.6 million hospital spells in the linked census population who had a CHD diagnosis since 1981. The spells were part of 2.08 million continuous inpatient stays (CIS) in 248,290 people, of whom 58596 died. There were 1170 South Asians who had died (10,085 CIS). These people contributed 455 AMI in the principal diagnosis position and another 22 in other positions. Reflecting their younger population structure, deaths in South Asians were at a comparatively younger age and had more men. The incidence rate and attack rate for AMI was higher at virtually every age group in South Asians. Age adjusted rates were about 65% higher in South Asians. Survival following AMI was much better in South Asians. After adjusting for age, sex and deprivation in a logistic regression model the OR for South Asians was 0.55, (95% CI = 0.36 – 0.84) and with age, sex and diabetes controlled the OR was 0.6 (95% CI = 0.44 – 0.82) in a Cox regression model.

**Conclusions.** The multilayered and complex SMR01 database can be analysed to provide useful and novel information. The numbers of cases, though small, are sufficient to provide information relevant to health needs assessment, especially in conjunction with other data. The incidence of AMI was shown to be comparatively high in South Asians, while survival after MI was relatively good. Both are unique observations in the Scottish context and add in important ways to the national and international data on the epidemic of CHD in South Asians.

## **7.1 Introduction**

In UK ethnicity and health research on mortality and morbidity investigators have, mainly, used mortality by country of birth based on numerators and denominators from different data systems; cross-sectional surveys; and analysis of large data sets using name search. There is a dearth of cohort type analysis based on individuals. The main source of such data is the ONS 1% sample of the census longitudinal study. This is England and Wales data where the principal linkage is to mortality data (Harding et al, 1997). Analysis by ethnic group is being done, and some preliminary data were reported by Gill et al (2005) and Harding et al (1997).

The approach taken here, developed as described in chapter 6, is an individual linked analysis i.e. cohort analysis linking to morbidity and mortality, with mortality being used to calculate survival. The cohort is the census population that was matched i.e. 94% of approximately 5 million people. The advantages are many, including of course that ethnicity is self-defined, or strictly, provided by the person completing the census form, usually the head of the household. The disadvantages include the technical difficulties and the fact that death outcomes before the census 2001 cannot be included (people who died were not in the Census).

In this analysis we test the potential of the linked data set. The focus is on CHD for reasons given earlier. There were, at the time of writing, no satisfactory UK data on CHD incidence or survival from acute myocardial infarction (AMI) by ethnic group. There was a prior expectation, inferred from mortality data and small scale local studies in England and Wales, and internationally, that incidence of CHD in South Asians would be higher and that survival after MI would be poorer.

## **7.2 Methods**

We linked SMR01 data from 1981 to December 2003 for AMI to census ethnic code as described in chapter 6. The first step was cleaning the data and preparing it for analysis. The data set is built up from spells of hospital care. One hospital admission can generate more than one spell e.g. if a person is admitted with a heart attack and develops a peptic ulcer requiring transfer to a gastro-enterologist, that means two spells. Spells were checked for consistent age and sex for each patient. Missing age was replaced by the age from a previous spell if it was consistent. Inconsistent age was replaced by the most frequent age for that patient.

The spells were summarised as continuous inpatient stays (CIS). Admission date was taken from the earliest recorded date, while the discharge date was the last recorded date found for spells within one CIS. If hospital based death occurred the finishing date of a CIS has been calculated as the date the death occurred.

Death records were coded to indicate whether it happened in hospital (death date found to be equal or less than the discharge date of the last hospital spell of a person) or in the community.

Only persons on the SMR01/CHD list who were alive at the Census 2001 (April 2001) could be linked to the census and provided with an ethnic group code. The time period available for follow up was, therefore, May 2001 till December 2003. Data prior to May 2001 were used for retrospective checking of previous admissions for diseases of interest and for co-morbidities.

The disease investigated was acute myocardial infarction (AMI). This was defined as diagnosis code 410 in ICD9 (prior to 1995) or I21/I22 in ICD10 for any hospital admission or death. Diagnoses could be in any position (of six provided). Procedures such as angioplasty were not available on the data file, but could be added for future analysis.



A new case was defined as someone who had not had an admission to hospital for AMI in the previous 10 years. Survival analysis was for the period up to 2.7 years after the AMI admission. For survival analysis we looked for any history of diabetes prior to the diagnosis of a first AMI. This included diabetes in any diagnosis position, not only the principal one.

The denominator for the analysis was the number of people included in the 2001 Census who were matched to the CHI index. There is, therefore, an exact correspondence between numerators and denominators.

We calculated age and sex specific hospital admission rates for all AMI admissions, for first occurrence of an AMI (this corresponds closely to incidence), and for subsequent all cause mortality. Incidence and mortality were directly and indirectly standardised using the whole matched population.

We conducted logistic regression for 30 days survival and survival analysis for survival after first AMI in the time period between 01 May 2001 and 31 December 2003, taking age (in categories), sex, ethnicity, prior history of diabetes and deprivation index into account as explanatory variables.

## 7.3 Results

### 7.3.1 Spells, admissions, deaths and rates

These results were based on people with hospital admissions and deaths with diagnostic codes for coronary heart disease (CHD) in any of the six possible positions on the SMR form. Table 7.1 shows 2,616,439 spells, including death records, from 248,290 persons. These spells were summarised as continuous inpatient stays (CIS, hospital admissions) including death records.

**Table 7.1: Distribution of spells**

Codes	Kind of spells	Observations
	Total	2616437
D2	Death records	58596
H1	Pre-Coppish hospital records	1112721
H2	Coppish hospital records	1445122

Table 7.2 shows the count of spells over the years. Deaths prior to April 2001 are likely to be errors since the dataset includes only people who respond to the 2001 Census. These errors may arise from erroneous completion of census forms e.g. a dead person is said to be living, or from the data linkage methods. The numbers are small and do not have any appreciable effect on the results. Deaths for 2004 were not available at the time of linkage.

**Table 7.2: Distribution of spells over the years**

Year	D2 (Deaths)	H1/H2 (Hospital)	Total
1981	0	31,028	31,028
1982	0	29,902	29,902
1983	0	35,720	35,720
1984	0	39,211	39,211
1985	0	42,138	42,138
1986	0	45,237	45,237
1987	0	49,587	49,587
1988	0	54,637	54,637
1989	0	59,703	59,703
1990	0	65,190	65,190
1991	0	73,215	73,215
1992	0	83,529	83,529
1993	0	96,620	96,620
1994	0	110,861	110,861
1995	0	123,704	123,704
1996	0	137,138	137,138
1997	8	151,853	151,861
1998	10	170,323	170,333
1999	19	188,193	188,212
2000	13	210,496	210,509
2001	13,786	256,293	270,079
2002	22,079	254,226	276,305
2003	22,681	230,918	253,599
2004	0	18,121	18,121
<i>Total</i>	<i>58,596</i>	<i>2,557,843</i>	<i>2,616,439</i>

Table 7.3 shows that the numbers and percentages of South Asians was close to that expected on the basis of the population size and structure i.e. just under 1% of the Scottish population was South Asian but we would expect fewer admissions and deaths as they are comparatively young. The number of South Asian people (1170) and CIS's (10085) is substantial. The deaths over 2 years (104) compares with about 300 in those born in India, Pakistan and Bangladeshi over 6.25 years period in the age group 25-69 (table 5.5). These comparisons give confidence that the data set is valid.

**Table 7.3: Distribution of SMR01 data by ethnic group (row percentages, unless stated otherwise)**

	<b>Non South Asians</b>	<b>South Asians</b>	<b>Total</b>
Spells total	2604035 (99.5%)	12400 (0.47%)	2616435 (100%)
Hospital spells only	2545547 (99.4%)	12296 (0.59%)	2557843 (100%)
CIS	2078257 (99.5%)	10085 (0.48%)	2088342 (100%)
Persons	247120 (99.5%)	1170 (0.47%)	248290 (100%)
Deaths total	58488 (99.8%)	104 (0.17%)	58592
Deaths in hospital (column %)	28579 (48.9%)	56 (53.9%)	28635 (47.4%)
Persons with death record only, no hospital data	1936 (99.7%)	6 (0.31%)	1942 (100%)

Table 7.4 shows the relative contribution of different CHD diagnosis (AMI, Angina and other) to the CIS's. Nearly all AMI's were in the principal diagnostic position indicating that this was the primary reason for admission. (Note that the totals in this table exceed the CIS in table 7.3 – the reason being that each person may have more than 1 CHD diagnosis.)

**Table 7.4: Distribution of continuous inpatient hospital stays**

<b>CIS</b>	<b>Non South Asians N=2078257</b>	<b>South Asians N=10085</b>	<b>Total 2088342</b>
AMI principal position	105067 (5.1%)	525 (5.2%)	105592 (5.1%)
AMI all positions	114575 (5.5%)	572 (5.7%)	115147 (5.5%)
Angina principal position	131115 (6.3%)	748 (7.4%)	131863 (6.3%)
Angina all positions	220096 (10.6%)	1238 (12.3%)	221334 (10.6%)
Other CHD principal position	155499 (7.5%)	1197 (11.9%)	156696 (7.5%)
Other CHD all positions	397257 (19.1%)	2421 (24.0%)	399716 (19.1%)
All CHD principal position	380095 (18.3%)	2383 (23.6%)	382478 (18.3%)
All CHD all positions	620522 (29.9%)	3520 (34.9%)	624042 (29.9%)

Table 7.5 shows the number of people by diagnosis. The key point is that there are substantial numbers of cases for analysis.

**Table 7.5: Distribution of persons by ethnic group and CHD diagnosis.**

Persons	Non South Asians	South Asians	Total
	N=247120	N=1170	248290
AMI principal position	96615 (39.1%)	445 (38.0%)	97060 (39.1%)
AMI all positions	101893 (41.2%)	467 (39.9%)	102360 (41.2%)
Angina principal position	76538 (31.0%)	463 (39.6%)	77001 (31.0%)
Angina all positions	110642 (44.8%)	633 (54.1%)	111275 (44.8%)
Other CHD principal position	91864 (37.2%)	626 (53.5%)	92490 (37.3%)
Other CHD all positions	165178 (66.8%)	874 (74.7%)	166052 (66.9%)

Table 7.6(a) shows there were more men than women with CHD diagnoses and 7.6 (b) shows South Asians died, on average, younger than non-South Asians, reflecting their younger age.

**Table 7.6a: Sex distribution and age at death in South Asians and Non South Asians**

	Non South Asians	South Asians
N	247120	1170
Female sex	44.7%	30.8%

**Table 7.6b: Average age at death in South Asians and Non South Asians**

	Non South Asians	South Asians
N	58492	104
Mean age at death (years)	77.5	70.6
Median age at death	79	72.5

Table 7.7 shows that while the sex ratio of deaths in the non-South Asian population was close to 1, in South Asians it was between 2 to 3 to 1. This difference is probably a result of the different age structures.

**Table 7.7: Deaths by year, sex and ethnic group**

Year	Non South Asians		South Asians	
	F	M	F	M
2001	6815	6950	5	16
2002	10948	11080	13	34
2003	11215	11429	10	25

About one third (33.8%) of all deaths in the CHD database had AMI in any position for the death certificate or the accompanying hospital stay. Table 7.8 shows that the proportions dying in and out of hospital were similar by ethnic group. Again, the slightly higher figure for South Asians dying in hospital probably reflects the younger population.

**Table 7.8: Deaths by place of death for AMI by ethnic group**

	Non South Asian	South Asian	Total
Death with AMI	19747 (100%)	54 (100%)	19801 (100%)
Community	12017 (60.9%)	30 (55.6%)	12047 (60.8%)
Hospital	7730 (39.2%)	24 (44.4%)	7754 (39.2%)

These deaths, occurring between May 2002 and December 2003 permit us to calculate mortality rates in a way that is complementary, yet superior, to the country of birth analysis in chapter 5. The database held 527,085 CIS with start date between 01 May 2001 and 31 Dec 2003 for people of 25 years of age and older. Out of these 41,619 (7.9%) were due to a condition with AMI. Of these admissions 171 (0.41%) were among South Asians; 154 were in the principal position and 126 were first AMI's. These cases permit us to calculate incidence/attack rates for AMI. For age standardisation we used the matched PEPI population of 25 years and older (table 7.12).

Table 7.9 and 7.10 show the hospital admission rate for AMI in principal diagnosis position per 1000 people per year, for first and all AMI's respectively. The rate is higher in men than women and rises with age (for both groups). Table 7.9(a) shows that in South Asian women the incidence rate is higher than in non-South Asian women in every age group except 25-34 years. The same applies to men with the exception of the very youngest and oldest age groups where the number of cases was too low for interpretation. The age standardised rates in men and women were more than 50% higher in South Asians than non-South Asians (table 7.9(b))

**Table 7.9(a): Acute first MI admissions (5/2001 – 12/2003) in principal position by ethnic group.**

	First AMI		Incidence / 1000 / year	
	Non South Asians	South Asians	Non South Asians	South Asians
Female				
25 to 34	29	0	0.03	0.00
35 to 44	199	2	0.20	0.28
45 to 54	585	6	0.68	1.36
55 to 64	1408	8	2.00	3.49
65 to 74	3043	10	5.05	8.39
75 to 84	4605	8	11.33	24.19
85 and over	3406	4	21.92	42.86
Total	13276	38	2.91	1.53
Male				
25 to 34	86	0	0.12	0.00
35 to 44	609	11	0.69	1.58
45 to 54	1948	12	2.36	2.56
55 to 64	3525	20	5.43	6.36
65 to 74	5091	27	10.28	15.29
75 to 84	4726	16	18.77	42.55
85 and over	1710	2	31.05	26.79
Total	17696	88	4.53	3.40

**Table 7.9 (b): Directly standardised first AMI rates (incidence) using the matched population of 25 years and older from the Scottish census population 2001 as reference**

Sex/Ethnicity	Person years	Crude rate / 1000	Adjusted rate / 1000	95% confidence interval for age adjusted rate
Female				
Non SA	4557730	2.91	2.56	2.51 – 2.60
SA	24762	1.53	4.86	3.05 – 6.67
Male				
Non SA	3905224	4.53	5.00	4.93 – 5.08
SA	25885	3.40	7.71	5.68 – 9.75

Table 7.10 shows that the pattern for first and subsequent MIs is very similar to that described for table 7.9. The total crude (actual) rates are higher in non-South Asians because this population is older.

**Table 7.10: All acute MI hospital admissions (first and subsequent (5/2001 – 12/2003) in principal position**

	AMI admissions 5/2001 – 12/2003		AMI admission rate / 1000 / year	
	Non South Asians	South Asians	Non South Asians	South Asians
Female				
25 to 34	34	0	0.04	0.00
35 to 44	220	2	0.23	0.28
45 to 54	655	6	0.77	1.36
55 to 64	1601	8	2.27	3.49
65 to 74	3537	12	5.87	10.07
75 to 84	5468	8	13.45	24.19
85 and over	4019	4	25.87	42.86
Total	15537	40	3.41	1.62
Male				
25 to 34	93	0	0.13	0.00
35 to 44	672	14	0.76	2.01
45 to 54	2272	17	2.76	3.63
55 to 64	4166	24	6.41	7.63
65 to 74	6152	36	12.42	20.39
75 to 84	5790	21	23.00	55.85
85 and over	2046	2	37.15	26.79
Total	21192	114	5.43	4.40

Table 7.11 shows that after age standardisation South Asian men and women had substantially raised rates of myocardial infarction. The confidence intervals for the adjusted rates for South Asians and non-South Asians do not overlap. South Asian women had a 69% excess in the adjusted rate compared to non-South Asians. Note, their age adjusted rate is about 50% of that for South Asian men i.e. the pattern was similar to that in non-South Asians. South Asian men had a rate 64% higher than in non-South Asians. These excesses are almost identical to those reported for mortality data by country of birth in England and Wales, but are not seen in relation to country of birth using the Scotland country of birth analysis in Chapter 5. This paradox suggests that contrary to our prior predictions, survival after MI in South Asians might be better.

**Table 7.11: Crude and age adjusted rates of myocardial infarction by sex and ethnic group using the matched population of 25 years and older from the Scottish census population 2001 as reference.**

<b>Sex/Ethnicity</b>	<b>Person years</b>	<b>Crude rate / 1000</b>	<b>Adjusted rate / 1000</b>	<b>95% confidence interval for adjusted rate</b>
Female				
Non-SA	4557730	3.42	2.99	2.94 – 3.03
SA	24762	1.62	5.07	3.23 – 6.91
Male				
Non-SA	3905224	5.44	6.01	5.92 – 6.08
SA	25885	4.40	9.88	7.63 – 12.14

Table 7.12 gives the population data underlying the above calculations

**Table 7.12: Census population and matched population for South Asians and Non South Asians- aged 25 years and older**

	<b>Non South Asian</b>		<b>South Asian</b>	
	<b>Census population 2001</b>	<b>Matched population</b>	<b>Census population 2001</b>	<b>Matched population</b>
Female				
25 to 34	355958	320018	4653	3519
35 to 44	398718	366337	3352	2647
45 to 54	345475	322103	2058	1653
55 to 64	283160	264030	1086	860
65 to 74	245338	225973	633	447
75 to 84	167914	152421	186	124
85 and over	65412	58266	56	35
Total	1861975	1709148	12024	9285
Male				
25 to 34	333956	278236	4830	3328
35 to 44	375219	332767	3575	2612
45 to 54	338779	308996	2264	1756
55 to 64	264009	243602	1477	1179
65 to 74	199225	185773	837	662
75 to 84	102214	94429	198	141
85 and over	22849	20655	38	28
Total	1636251	1464458	13219	9706

### 7.3.2 Survival analysis and logistic regression

The Kaplan-Meier survival plot (figure 7.1) for survival after first AMI shows better survival after first AMI for South Asians than the non-South Asian population. For South Asians no



obvious differences are visible between male and females. In the non-South Asian population, males show a better survival than females. However, all these differences might be due to age differences between the groups and are evaluated further below.

**Figure 7.1 Kaplan-Meier survival plot for survival in days after AMI**

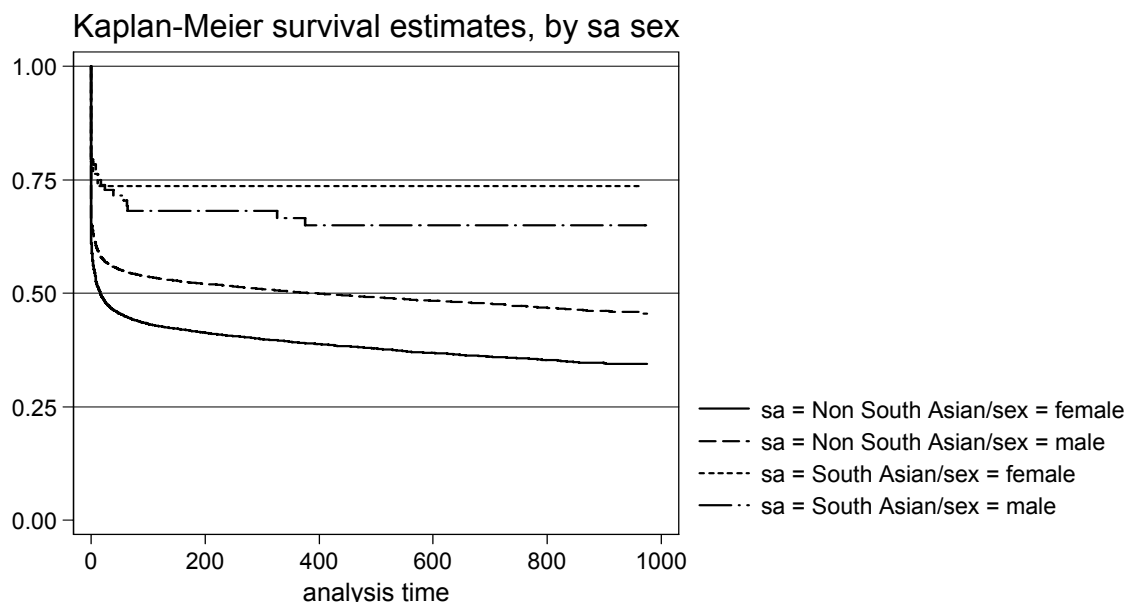


Table 7.13 and 7.14 show age (and sex) specific survival analysis for South Asians. We did not pursue additional age adjustment for single age within age groups because of the small numbers in South Asians and because no additional improvement was found when such age adjustment was applied. In most age (and sex) groups South Asians showed a better survival after AMI than the non-South Asian population. This is shown by a hazard ratio less than 1. The numbers of cases by age group was too small for a statistically robust analysis but the pattern was consistent and contrary to previous predictions.

**Table 7.13 Age specific hazard ratios for survival over 2.6 years after AMI in South Asians with first diagnosis of AMI in principal position**

Age group	Deaths (All)	Deaths South Asian	Hazard ratio (HR) for South Asian HR	95%-Confidence Incidence
35-44	142	2	0.87	.21 - 3.5
45-54	483	4	1.19	.44 - 3.18
55-64	1641	6	0.61	.27 - 1.35
65-74	4170	13	0.62	.36 - 1.06
75-84	6459	11	0.56	.31 - 1.01
85+	4203	4	0.87	.33 - 2.33

Table 7.14, with AMI in any position, confirms the pattern in table 7.13.

**Table 7.14 Age specific hazard ratios for survival over 2.6 years after AMI in South Asians with first diagnosis of AMI in any position**

Age group	Deaths (All)	Deaths South Asian	Hazard ratio (HR) for South Asian HR	95%-Confidence Incidence
35-44	162	2	0.73	.18 - 2.93
45-54	513	4	1.23	.46 - 3.29
55-64	1784	7	0.64	.30 - 1.34
65-74	4586	16	0.7	.43 - 1.14
75-84	7093	12	0.57	.32 - 1.01
85+	4598	4	0.92	.35 - 2.46

The data were combined and subsequently analysed in multiple logistic regression and Cox regression models to help verify these findings and to understand the factors that might lead to better survival in South Asians.

Without adjusting the results for sex, age and deprivation, South Asians showed a significantly better survival after first AMI. Table 7.15 shows the odds ratio for this analysis was substantially below one and statistically significantly so.

**Table 7.15 Odds ratio for 30 day survival following AMI in South Asians compared to non-South Asians**

	OR	95% Confidence Interval		p-value
		Lower	Upper	
South Asian	.43	.29	.64	< 0.001

Table 7.16 shows the odds ratio changed little after adjustment for important confounding factors.

**Table 7.16 Logistic regression analysis of 30-day survival in South Asians compared to non-South Asians, adjusting for SIMD\*, sex, and age.**

	OR	95% Confidence Interval		p-value
		Lower	Upper	
South Asian	.55	.36	.84	0.005
SIMD 2	1.16	1.06	1.26	0.001
SIMD 3	1.20	1.11	1.30	< 0.001
SIMD 4	1.18	1.09	1.28	< 0.001
SIMD 5	1.26	1.17	1.36	< 0.001
Male	.87	.82	.91	< 0.001
Age 0-34	.15	.09	.24	< 0.001
Age 35-45	.15	.12	.18	< 0.001
Age 45-55	.18	.16	.20	< 0.001
Age 55-65	.34	.32	.36	< 0.001

\* SIMD: Scottish index of multiple deprivation

Cox regression for survival after AMI without taking into account any age or sex differences or other covariates showed similar results as in the logistic regression for 30-

day survival – South Asians presenting a hazard ratio of 0.48 (95% CI = 0.35 – 0.66) therefore having a significant better survival after AMI than non-South Asians (table 7.17).

**Table 7.17 Hazard ratio for survival after 2.6 years following AMI-comparison of South Asians and non-South Asians.**

	Hazard ratio	95% Confidence Interval		p-value
		Lower	Upper	
South Asian	0.48	0.35	0.66	< 0.001

Even after adjustment for sex, age and an index of multiple deprivation the odds ratio for survival was statistically significantly lower at 0.55.

Adjusting the Cox regression survival analysis for simultaneously age, sex and previous diagnosis of diabetes the difference in the hazard ratio for survival in South Asians changed only slightly. The different age structure in South Asians – being a younger population than the non-South Asians-explained only a small part of the differences in survival. A diagnosis for diabetes did not show a significant effect on survival (table 7.18).

**Table 7.18 Hazard ratio for survival after 2.6 years following AMI – comparison of South Asians and non-South Asians after adjustment for age, sex and diabetes**

	Hazard ratio	95% Confidence Interval		p-value
		Lower	Upper	
South Asian	0.60	0.44	0.82	0.001
Diabetes	1.01	.97	1.06	0.551
0-34	.18	.11	.29	< 0.001
35-44	.21	.18	.25	< 0.001
45-54	.23	.21	.25	< 0.001
55-64	.43	.41	.45	< 0.001
Male sex	.91	.88	.94	< 0.001

We examined survival following MI at 30, 90, 180, 270, 360 and 720 days. The results were consistent with the picture shown above. Analysis by sex, in Glasgow and outside Glasgow, and in more and less economically deprived areas of Scotland all showed the same picture. The analysis is shown in the appendix tables.

## 7.4 Conclusions

### 7.4.1 Methods

Our linkage is a 'data rich' linkage i.e. it was done using names, dates of birth and addresses. In this respect this may be a unique approach to achieving data on ethnic variations. Blakely and Salmond reported a probability match in New Zealand (2002). Theirs was a 'data impoverished' match without actual names and addresses but using various codes e.g. country of origin, area of residence etc. Jeffreys et al reported a linkage study of cancer, focusing on New Zealand Maoris, but theirs was a linkage of

hospital and mortality databases and not population based. Our method gives high confidence that with few errors we have one-to-one matching.

Cohort analyses can be done on data sets that link individuals to future outcomes on a one-to-one basis, on the assumption that the probability linkage is accurately linking individuals on the census to individual outcomes on the SMR01 database, via CHI, on a one-to-one basis, our output is a cohort analysis, with the census providing baseline data. We have voluntarily stripped off the individual identifiers, and we have taken out of the baseline (census) database those who are not matched. As such, the denominators are the numbers of census records that are matched i.e. 85-95% percent of about 5 million people, the match varying slightly by ethnic group and by age group. At this point we have no information on whether those who are not matched differ from those who are but this could be explored in future work.

The numerators are the morbidity and mortality counts on the SMR. All people on our linked database who do not have a link to the SMR record for acute myocardial infarction can be assumed, reasonably for this severe outcome, to have had no outcome. For the period of time following the census, the outcomes can be analysed using a person-time denominator or the denominator of people at risk. This analysis makes an assumption, which does not seem unreasonable, that those resident in Scotland at the census, in the main, continue to be resident in the country. Morbidity data prior to the census held on the SMR01 database back to 1981 can be used to assess whether the events following the census are new ones or recurrent ones. As such, we get an estimate of the incidence of acute myocardial infarction based on new cases and an estimate of both new and recurrent cases (attack rate).

We currently do not know how long the people at Census 2001 were resident in Scotland prior to the census. Obviously, if they were linked to the SMR01 database they probably were resident for some time. If they were not linked, then we cannot be sure. In future projects there may be other ways of answering this question. Firstly, despite the lack of names on the electronic files from the 1991 census it may be possible to probability match that (though we acknowledge it may prove too difficult). This information could be used to construct denominators for earlier years. Secondly, we might be able to use older CHIs, to see if people were listed. If successful this approach could be used for a retrospective cohort with much larger numbers. Even for this analysis with prospective data for a little more than two years there were sufficient events to permit an analysis. One could envisage analysis with 5-7 year of outcome data in future. In developing our methods we have aggregated the South Asian populations. For most analyses the numbers would be too small for disaggregating until more time elapses and we have longer follow up. For some of the commoner events analysis could be done for other ethnic groups. CHD related events are not the only common events for hospital admission so analysis of this kind could be done on respiratory, gastrointestinal and neoplastic disorders. For disorders that are severe, and usually lead to hospitalisation or death, an estimate of disease incidence can be obtained, as here for myocardial infarction.

This is a complex data set, with multilayered information (spells, CIS, new and old episodes, and ultimately individuals with sometimes long, linked records reaching from 1981 to death). Analysing it is demanding. While we were able to learn from the work of others, our approaches and methods were developed afresh and were in several respects better for the purpose at hand. For example, while Macintyre (2001) and colleagues have focused on new AMIs we considered new and old. This has benefits in the context of ethnicity and incidence as attack rates data are sparse and needed, and as numbers of cases are low we need methods that maximise the number of cases.

## 7.4.2 Findings

The most important findings were that there were substantial numbers of completed inpatient stays in sufficient numbers of people to permit an interesting and novel analysis. The analysis provides, for the first time in a UK population, an indication of attack rates (all AMIs) and incidence (new cases), which were both higher, and survival which was better. These survival findings were only little influenced by the age and sex structure, and after adjusting for these the differences persisted. Deprivation did not show any relevant effect on survival in our data, contradicting results presented elsewhere where a lower deprivation category had a negative effect on survival after AMI (Macintyre et al, 2001). The results in relation to survival contradicted our prior expectation, and the relatively little information on survival following AMI in South Asians. The findings were, however, consistent – whether we used age-specific survival rates or regression with logistic or Cox modelling. Our intensive efforts to find errors or confounding factors failed to uncover any that countered the findings.

There was an important excess of AMI in the Scottish South Asian population, a finding that was hinted at in the country of birth analysis shown in chapter 5, particularly in the age group 25-69 (differences not statistically significant there). This implies that the high rates of CHD mortality derive from high incidence and not high case-fatality and that is, to our knowledge a unique demonstration. Indeed, the picture seems to be that South Asians have much higher morbidity rates for AMI, but because of the lower case-fatality, these difference are not seen clearly in mortality data (country of birth based analysis). These findings need corroboration in other data sets.

## Appendices

**Table A7.1: Description of survival following an acute AMI after 30, 90, 180, 270, 360, and 720 days by ethnic group**

		Non South Asian		South Asian		Total	
		N	%	N	%	N	%
30 days	alive (%)	15868	52.48	88	72.13	15956	52.56
	dead (%)	14369	47.52	34	27.87	14403	47.44
	Total (N)	30237		122		30359	
90 days	alive (%)	14056	48.65	83	70.34	14139	48.74
	dead (%)	14838	51.35	35	29.66	14873	51.26
	Total (N)	28894		118		29012	
180 days	alive(%)	12345	45.29	79	69.3	12424	45.39
	dead(%)	14910	54.71	35	30.7	14945	54.61
	Total (N)	27255		114		27369	
270 days	alive (%)	10713	41.79	70	66.67	10783	41.9
	dead (%)	14920	58.21	35	33.33	14955	58.1
	Total (N)	25633		105		25738	
360 days	alive(%)	9118	37.93	57	61.96	9175	38.02
	dead(%)	14923	62.07	35	38.04	14958	61.98
	Total (N)	24041		92		24133	
720 days	alive(%)	3416	18.62	30	46.15	3446	18.72
	dead(%)	14926	81.38	35	53.85	14961	81.28
	Total(N)	18342		65		18407	

**Table A7.2: Description of survival after 30, 90, 180, 270, 360, and 720 days and by sex.**

		Non SA		SA	
		female	male	female	male
30 days	alive(%)	47.26	56.4	70.59	72.73
	dead(%)	52.74	43.6	29.41	27.27
	Total(N)	12,974	17,257	34	88
90 days	alive(%)	42.7	53.13	69.7	70.59
	dead(%)	57.3	46.87	30.3	29.41
	Total(N)	12,415	16,473	33	85
180 days	alive(%)	39.19	49.94	67.74	69.88
	dead(%)	60.81	50.06	32.26	30.12
	Total(N)	11,773	15,476	31	83
270 days	alive(%)	35.96	46.31	65.52	67.11
	dead(%)	64.04	53.69	34.48	32.89
	Total(N)	11,186	14,441	29	76
360 days	alive(%)	32.26	42.38	61.54	62.12
	dead(%)	67.74	57.62	38.46	37.88
	Total(N)	10,579	13,456	26	66
720 days	alive(%)	14.86	21.82	50	44.44
	dead(%)	85.14	78.18	50	55.56
	Total(N)	8,418	9,921	20	45

**Table A7.3: Description of survival after 30, 90, 180, 270, 360, and 720 days by Glasgow and not Glasgow.**

		Non Glasgow		Glasgow	
		Non-SA	SA	Non-SA	SA
30 days	alive(%)	52.8	71.67	50.92	72.58
	Dead(%)	47.2	28.33	49.08	27.42
	Total	25,041	60	5,196	62
90 days	Alive(%)	49.05	68.42	46.66	72.13
	Dead(%)	50.95	31.58	53.34	27.87
	Total (N)	23,963	57	4,931	61
180 days	Alive(%)	45.75	67.27	43.09	71.19
	Dead(%)	54.25	32.73	56.91	28.81
	Total (N)	22,614	55	4,641	59
270 days	Alive(%)	42.33	64	39.18	69.09
	Dead(%)	57.67	36	60.82	30.91
	Total (N)	21,289	50	4,344	55
360 days	Alive(%)	38.46	56.1	35.32	66.67
	Dead(%)	61.54	43.9	64.68	33.33
	Total (N)	19,956	41	4,085	51
720 days	Alive(%)	19.11	41.94	16.31	50
	Dead(%)	80.89	58.06	83.69	50
	Total (N)	15,184	31	3,158	34

**Table A7.4: Description of survival after 30, 90, 180, 270, 360, and 720 days by deprivation score**

		Deprivation score < 21.62		Deprivation score > 21.62		Deprivation score missing	
		Non SA	SA	Non-SA	SA	Non-SA	SA
30 days	alive(%)	53.51	68.57	52.96	78	12.48	50
	dead(%)	46.49	31.43	47.04	22	87.52	50
	Total(N)	15,876	70	13,784	50	577	100
90 days	alive(%)	49.62	65.67	49.15	77.55	11.52	50
	dead(%)	50.38	34.33	50.85	22.45	88.48	50
	Total(N)	15,160	67	13,161	49	573	2
180 days	alive(%)	46.24	64.62	45.83	77.08	9.14	0
	dead(%)	53.76	35.38	54.17	22.92	90.86	100
	Total(N)	14,287	65	12,410	48	558	1
270 days	alive(%)	42.73	61.67	42.32	75	7.99	0
	dead(%)	57.27	38.33	57.68	25	92.01	100
	Total(N)	13,420	60	11,662	44	551	1
360 days	alive(%)	38.77	56.6	38.51	71.05	6.8	0
	dead(%)	61.23	43.4	61.49	28.95	93.2	100
	Total(N)	12,556	53	10,941	38	544	1
720 days	alive(%)	19.24	42.5	18.96	54.17	1.93	0
	dead(%)	80.76	57.5	81.04	45.83	98.07	100
	Total(N)	9,522	40	8,303	24	517	1

## 8 Conclusions and recommendations

BHOPAL R<sup>1</sup>, FISCHBACHER CM<sup>2</sup>, STEINER M<sup>1</sup>, CHALMERS C<sup>1,2</sup>, POVEY C<sup>2</sup>, JAMIESON J<sup>2,3</sup>, KNOWLES D<sup>2</sup>

*(<sup>1</sup>Public Health Sciences, University of Edinburgh, <sup>2</sup>Information and Statistics Division, National Services Scotland, <sup>3</sup>National Resource Centre for Ethnic Minorities, NHS Health Scotland.,*

### 8.1 Main findings

Routine health data in Scotland generally provide very patchy and severely incomplete information about ethnicity. Several issues are common to all the sources reviewed: incomplete data, inconsistency about the categories used for ethnic group, a lack of training for staff on collecting data on ethnicity, and a lack of provision within information systems for ethnicity to be recorded (chapter two).

Until now information about the health of ethnic minorities has not been extracted from databases in Scotland. Our project demonstrates that retrospectively this is possible at relatively low cost, and reasonable timescales, given co-operation between agencies and appropriate skills within the research team. Our methods, including those for data linkage, are likely to be applicable beyond Scotland.

The results from our analysis show that the application of South Asian name search algorithms by computer alone to major population based Scottish disease registers does not give satisfactory results. South Asian status needs to be confirmed by expert observers. Name search algorithms are useful as a screening tool but when the proportion of South Asians in the population is low, visual inspection by expert observers is essential for accurate classification. This said, when circumstances require it, and alternatives are not available, name search methods can yield valuable insights that can be the basis of both service planning and further research – as demonstrated in the analysis of the DARTS database (chapter three). For South Asians the name search of the DARTS database showed that the prevalence rate of type 2 diabetes mellitus was 3-4 times higher than in the non-South Asian population – as predicted from experience elsewhere, but for the first time using reliable Scottish data. There were reassuring findings that indicators of quality of care were similar in South Asian and non-South Asian groups. The HbA1c levels and retinopathy were slightly higher in the South Asian populations, indicating that glycaemic control is poorer.

Country of birth is a variable of interest in epidemiology. In certain circumstances it is a reasonable proxy for ethnic group e.g. most people living in Scotland who were born in China are ethnic Chinese. By contrast a fair number of White Scottish people were born in India. Census 2001 data show that among those aged 60 years or more, 1015 White Scottish and 510 other White British people were born in India. They outnumber the 1013 Scottish resident ethnic Indians over 60 years born there. (In those under 60 years there were 991 Scottish White people born in India compared to 5217 ethnic Indians.) This point has been of lesser importance in England and Wales because the Indian ethnic group is much larger there. Given a degree of caution in analysis and interpretation the findings from country of birth data are extremely interesting (chapter five).

We showed major variations by country of birth in all cause and CHD deaths. Country of birth analyses showed substantial variations in mortality from CHD. The deficit in Chinese and African or African Caribbean born follows the pattern demonstrated in England. The SMR for Indian Subcontinent born (South Asian) population was not raised in the initial country of birth analysis using the Scotland population as standard. Further analysis



using the England and Wales population as the standard showed that these populations have very high rates but then so do the Scots-born. Comparisons with the Scottish population might be interpreted as reassuring, since they do not show the excesses in CHD mortality in South Asian born populations as reported in England and Wales. However, the high risks are being disguised by using a reference group that itself is at high risk i.e. Scottish born, and in the all-age analysis diluting the differences by including a large number of ethnically White people. This analysis highlights the importance of the choice of reference group and age group in assessing relative mortality risks by country of birth. In relation to control of coronary heart disease either in South Asians in Scotland, or those mainly White people born in Scotland, there is no room for complacency.

Our major innovation in this project was the linkage of ethnic codes from the census to the SMR01 database (chapters six and seven). The technique described meets ethical, professional and legal concerns about the linkage of census and health data and is potentially transferable outside Scotland. This analysis is in individual based data – both demographic and health outcomes – so the design is that of a cohort study, without the limitations imposed by country of birth analysis. There was a clear excess in incidence of acute myocardial infarction remarkably similar to the excess of mortality in Indian Subcontinent born in England and Wales. Survival data have been sparse. South Asians had a better survival following myocardial infarction for reasons that remain mysterious, and merit further investigation. The data on incidence and survival fill an important gap in the research literature and are likely to be of international interest.

We need lifestyle and health data to interpret disease variations and to inform interventions to reduce inequalities. There are no cardiovascular risk factor data in Scotland to help explain these ethnic variations. Our limited use of imputed data (chapter four) gave some pointers towards acquiring quickly modelled estimates of risk factors for Scotland's ethnic minority populations. The work proved more technically demanding than we anticipated but it needs to be extended, pending the development and implementation of a Scottish multiethnic lifestyle and health survey.

## **8.2 Information gaps filled and left**

As a consequence of the project a steering group has formed and gained essential, and perhaps unique, experience that will be vitally important in developing this field of work in Scotland. Extremely good working relations between the University of Edinburgh, GROS, NRCEMH and ISD have resulted. A gap in infrastructure has been filled, albeit temporarily, that complements the work underway on prospective ethnic coding of information systems. This gain is not to be underestimated and needs to be nurtured.

The second gap filled is on the understanding of health status in some of Scotland's minority ethnic groups. This project was deliberately focused on CHD, the dominant cause of death in Scotland. Our primary goal was to test whether our methods were suited to the task. Our initial expectation was that name search would be our best approach, and we did not anticipate the linkage methods could be anonymised sufficiently to win approval. These expectations were overturned. We have shed light, from a number of perspectives, on ethnic variations in all cause mortality, CHD mortality and morbidity and to a lesser extent on diabetes prevalence and quality of care.

Our observations will help policy makers, researchers, planners, clinical networks, health promoters and service providers. The need, previously identified largely on data from England, for vigorous CHD and diabetes prevention and control services, such as Khush Dil, a heart disease and diabetes prevention and control project in Edinburgh that serves the South Asian populations, is now clear cut.

Similar data are required for cancers, stroke, and other major causes of death, disability and ill-health including mental health and maternal and child problems. We also need to

attend to the needs of a broader range of ethnic minority groups. The methods and infrastructure developed here can help to fill these and many other gaps.

### **8.3 Next steps, lessons learned and recommendations**

This demonstration project has tested and enhanced several existing methods (name searching, country of birth, imputation, record linkage) for achieving understanding of ethnic variations in the health of Scotland's ethnic minority populations. The strengths and weakness of these methods are now better known than hitherto.

Many lessons have been learned, and others will result from the reflection and feedback that will result from this report and the process of publication of papers. Some of these lessons are outlined below.

Given an agreement on a need, resources, teamwork, high level support, skill and will, what at first seems nigh impossible (extracting ethnicity data from Scotland's health databases) can be achieved. In particular, this applies to the linkage of census and SMR01 data, but also to other components of the work.

In surmounting the numerous obstacles we encountered, the value of both cooperation between agencies, and of moral support from their leaders, was crucial. The substantial amount of time spent in consultation and in communication was repaid.

As with most novel projects the time and resources required were greater than anticipated. Projects like this require – and in our case had – considerable freedom in management of work and timescales. Future projects of this kind should note the substantial amounts of time required for consultation and for gaining necessary permissions.

This demonstration project focused on one health problem (CHD) and on one major and heterogeneous ethnic group (South Asians), though not exclusively so. On many occasions we were tempted to broaden the scope of the work, not least because it could be considered that a narrow focus is unfair e.g. to those interested in other conditions and to ethnic groups that have not been studied in detail. We were wise not to do this, as it would have run a high risk of failure to deliver. These gaps can be filled now that the methods are developed and tested.

Scotland's ethnic minority populations are relatively small. The question of whether there would be sufficient deaths/cases to permit statistically valid analyses was hanging over this project. The results show that while the numbers are indeed relatively small, the findings are of interest and of value. Scotland's ethnic minority populations are growing and ageing so the numbers of deaths/cases will be growing rapidly. Scotland has been, in the main, reliant on data from England and Wales for inferring the pattern of mortality and morbidity by ethnic group. Our results from the imputation project and other findings on mortality give some confidence that this approach has some value in relation to lifestyle and health data.

Our recommendations are:

**Recommendation 1:** Name search methods developed in England will need considerable extra refinement if they are to be used in automatic mode i.e. without visual inspection by experts. Such work is not a priority in relation to the need to develop prospective information systems, ultimately obviating the need for name searching. Investigators requiring name searching methods should be aware of their limitations and accept that the specificity is low in a Scottish setting, thus requiring visual inspection of ethnic code outputs.

**Recommendation 2:** Country of birth is not only a reasonable proxy for ethnicity for most of Scotland's ethnic minority population in the middle years of life but is of interest in itself.

Country of birth can continue to be of use in analysis of mortality data. Analysis of such data on a broader range of causes of death is recommended to

- add to our understanding of ethnicity and health for policy and service purposes
- contribute ideas for epidemiological and clinical research
- compare and contrast findings with a large and growing international research literature

In future analysis the country of birth groupings we used need to be reviewed.

**Recommendation 3:** Imputation has shown its value as a stop-gap measure. We recommend, however, that Scotland fills the data vacuum by conducting a Scottish multiethnic lifestyle and health survey. Lifestyle and health data need to be local to hold the credibility needed to spur action. The work done here indicates that there are likely to be important and substantial ethnic variations in key risk factors in Scotland.

**Recommendation 4:** This project has innovated in linking ethnic codes on the census to mortality and morbidity databases. To our knowledge this is a unique achievement in the UK and possibly internationally. In some European countries population registers can be similarly linked but those are not censuses. Our recommendations are that the work is continued to complete the analysis for other disease categories and to develop more advanced analysis, including exploration of health service interventions e.g. procedures and operations. Further, the methods are open to refinement. The demonstration project was, rightly, done to exacting security standards. If the method is to be adopted more widely and in a service context a more streamlined approach will be feasible. This is probably justifiable as the demonstration project has shown the outputs are valuable. The next step is to open a dialogue between ISD, GROS, Privacy Advisory Committee and the Research Ethics Committee for Scotland on how the process might be made more efficient. Whether streamlined or not, the work needs to be completed. It is likely to be at least 5-years before the prospective ethnic coding system for Scotland is able to produce data of the kind this approach gives. It is quite possible, but hopefully it will not be required, that the method will need to be used again even after the 2011 census.

**Recommendation 5:** The probability linkage method is likely to be exportable, not only to the other UK countries, but internationally wherever there is a census recording ethnic codes and an electronic health database with administrative details and, preferably, a unique number such as the CHI. GROS and ISD should be resourced to open a dialogue with their counterparts in other UK countries. The needs for data are so great that multi-ethnic societies throughout Europe need access to new methods urgently.

**Recommendation 6:** The master file linking the encrypted user code to the encrypted CHI number is held safely. Before it can be used again, we need PAC and MREC approval. Gaining approval, streamlining the procedures, consulting with key data users, undertaking the additional analyses and writing them up in a way that will help policy makers, service commissioners and service providers, and achieve journal publication requires that the project is extended for at least 2 years. The extra work should extend to diseases other than CHD and to a broader range of ethnic minority groups. The linkage analyses should also extend beyond ethnic group to religion and census based on country of birth. The Steering Group that successfully managed the current project should be invited to continue the work.

## ***8.4 Letting the information desert bloom***

The costs of retrospective extraction of ethnicity and health data from Scotland's information systems in the ways demonstrated here are small in comparison to achieving it prospectively. Nonetheless, legal, ethical and policy related drivers require prospective coding. This project, and its proposed extension, are necessary measures, but hopefully

interim ones. In describing this project's original ambitions the Principal Investigator (R.S.B) has often said "Scotland's health information systems, as far as ethnicity and health concerned, are a desert. Our aim is to water a part of this desert and see whether it is fertile". It turns out that it is. The water can now be sprinkled more widely to let this information desert bloom.

## References

- Agyemang C, Bhopal R. Is the blood pressure of South Asian adults in the UK higher or lower than that in European white adults? A review of cross-sectional data. *Journal of Human Hypertension*. 2002; 16: 739-751
- Altman DG, *Practical Statistics for Medical Research*, Chapman&Hall/CRC,1990 (reprint 2000)
- Balarajan R, Bulusu L. Mortality among immigrants in England and Wales, 1979 - 83. In: Britton M, editor. *Mortality and Geography*. London: HMSO, 1990.
- Becker T, Wiggins C, Key C, Samet J. Ischaemic heart disease mortality in Hispanics, American Indians, and non-Hispanic whites in New Mexico, 1958-1982. *Circulation* 1988;78:302-9
- Bhopal R S. *Concepts of Epidemiology*. Oxford, Oxford University Press, 2002, pp317
- Bhopal R. What is the risk of coronary heart disease in South Asians? A review of UK research. *Journal of Public Health Medicine* 2000;22:375-385.
- Blakely T, Salmond C. Probabilistic record linkage and a method to calculate the positive predicitive value. *International Epidemiological Association* 2002; 31:1246-1252.
- Coldman A, Braun T, Gallagher R. The classification of ethnic status using name information. *J Epidemiol Community Health* 1988;42:390-95.
- Cummins C, Winter H, Cheng KK, Maric R, Silcocks P, Varghese C. An assessment of the Nam Pehchan computer program for the identification of names of South Asian ethnic origin. *Journal of Public Health Medicine* 1999; 21:401-406.
- Diabetes Services Managed Clinical Network TDAG Annual Report October 2002. Tayside, 2002.
- Erens B, Primatesta P, Prior G (eds). *Health Survey for England: The Health of Minority Ethnic Groups '99 Vols. 1 & 2*, ix-633. The Stationery Office. London, 2001.
- Gill PS, Kai J, Bhopal RS, Wild S. Health Care Needs Assessment: Black and Minority Ethnic Groups. In: Raftery J (ed). *Health Care Needs Assessment. The epidemiologically based needs assessment reviews. Third Series*. Abingdon: Radcliffe Medical Press Ltd. (in press, expected 2005). At: <http://hcna.radcliffe-oxford.com/bemgframe.htm>
- Harding S. Mortality of migrants from the Indian Subcontienent to England and Wales: Effect of Duration of Residence. *Epidemiology* 2003; 14:287-292.
- Jeffreys M, Stevanovic V, Tobias M, Lewis C, Ellison-Loschmann L, Pearce N et al. Ethnic inequalities in cancer survival in New Zealand: Linkage study. *Am J Public Health* 2005; 95:834-837.
- Macintyre K, Stewart S, Chalmers J, Pell J, Finlayson A, Boyd J et al. Relation between socioeconomic deprivation and death from a first myocardial infarction in Scotland: population based analysis. *BMJ* 2001; 322:1152-1153.
- Marmot MG, Adelstein AM, Bulusu L. Immigrant mortality in England and Wales 1970-78: studies of medical and population subjects, no 47. OPCS:London: HMSO 1984
- Matheson L, Dunnigan M, Hole D, Gillis C. Incidence of colo-rectal, breast and lung cancer in a Scottish Asian population. *Health Bulletin* 1990;43:245-49

- Nanchahal K, Mangtani P, Alston M, dos Santos Silva I. Development and validation of a computerized South Asian names and group recognition algorithm (SANGRA) for use in British health-related studies. *Journal of Public Health Medicine* 2001; 23:278-285.
- National statistics. Regional distribution: by minority ethnic population: Census update. Accessed 10th November 2003  
<http://www.statistics.gov.uk/statbase/ssdataset.asp?vlnk=6589&More=Y>
- Netto G, Arshad R, de Lima P, Diniz FA, MacEwen M, Patel V, Syed R; Audit of Research on Minority Ethnic Issues in Scotland from a 'Race' Perspective; Scottish Executive Central Research Unit; 2001.
- Newcombe H. Handbook of record linkage: methods for health and statistical studies, administration and business. Oxford: Oxford University Press, 1988.
- Nicoll A, Bassett K, Ulijaszek S. What's in a name? Accuracy of using surnames and forenames in ascribing Asian ethnic identity in English populations. *J Epidemiol Community Health* 1986;40:364-68
- NRCEMH (National Resource Centre for Minority Ethnic Health). Diabetes in Ethnic Minority Groups in Scotland, 2004.
- Patel KC, Bhopal RS. (Editors) The epidemic of coronary heart disease in South Asian populations: causes and consequences London: South Asian Health Foundation; 2004, pp145.
- Race Equality Advisory Forum Report, Making it Real – A Race Equality Strategy for Scotland, Edinburgh, Scottish Executive, 2001.
- Race Relations (Amendment) Act 2000. New laws for a successful multi-racial Britain. London: Home Office, 2000.
- Registrar General for Scotland: Scotland's Census 2001, Reference Volume. General Register Office for Scotland, Edinburgh, 25 March 2003
- Scottish Executive Equality Unit, Towards an Equality Strategy, 2000
- Scottish Executive Health Department, Fair for All: Working Together Towards Culturally-Competent Services, NHS HDL (2002) 51, Edinburgh, Scottish Executive
- Scottish Executive, Coronary Heart Disease and Stroke Strategy for Scotland, Reference Group. Scottish Executive, Edinburgh, August 2002
- Scottish Executive White Paper, Towards a Healthier Scotland, Cm 4269, The Stationery Office, 1999
- Wild S, McKeigue P. Cross sectional analysis of mortality by country of birth in England and Wales, 1970-92. *BMJ* 1997;314:705-10.