COHORT PROFILE

Cohort profile: Scottish Health and Ethnicity Linkage Study of 4.65 million people exploring ethnic variations in disease in Scotland

Raj Bhopal,[1]* Colin Fischbacher,[1,2] Christopher Povey,[2] Jim Chalmers,[1,2] Ganka Mueller,[3] Markus Steiner,[1] Helen Brown,[1] David H Brewster[1,2] and Narinder Bansal[1]

[1]Public Health Sciences, Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK, [2]Information Services Division, NHS National Services Scotland, Edinburgh, UK and [3]General Register Office for Scotland, Edinburgh, UK

*Corresponding author. Edinburgh Ethnicity and Health Research Group, Public Health Sciences, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, UK. E-mail: raj.bhopal@ed.ac.uk

How did the study come about?

Many countries require health services to show that they are meeting the needs of ethnic minority populations.[1] This requires data on health status, healthcare uptake and outcomes and population denominators. Weaknesses in routine data collection often make such requirements difficult to meet. Routine data sources in Scotland, as in most countries, may not include a patient's ethnicity. In Scotland, the need for such data is driven by both policy and legislation responding to rapidly increasing ethnic diversity. Fair For All (2003), Scotland's policy, provides a strategic approach to improve the health of minority ethnic groups.[2] The UK Race Relations (Amendment) Act (2000) placed a duty on public bodies to promote racial equality.[3] These mandates are reflected in guidance on ethnic monitoring.[4–6] Appropriate service and research is undermined by the lack of data. Ethnic variations occur in all of Scotland's national health priority areas, including coronary heart disease/stroke,[7,8] cancer,[9–11] maternal and child health[12,13] and mental health.[14]

In view of the mismatch between need for and availability of data by ethnic group, Bhopal proposed a demonstration project to explore retrospective approaches.[15] The project tested proposals including name search methods, analyses by country of birth, modelling/extrapolation from other nations' datasets, and linkage methods. The demonstration project concluded that census health records linkage methods—in the context of this project first mooted by Povey—held most promise. To our knowledge, attempting matching of a national health dataset to a complete national census using demographic identifiers rather than national identity numbers had not been reported though health data linkage is well-established in the UK and internationally,[16] including exploring ethnicity and health.[17,18]

Who set it up, how, and why, and how was it funded

The study was set up in collaboration by the University of Edinburgh (R.B. and C.F.), the Information Services Division (ISD) of NHS National Services Scotland and the General Register Office for Scotland (GROS). ISD maintains hospital episode data (the Scottish Morbidity Record), which has already been linked to GROS mortality data. A permanently linked database joins together mortality to general hospital discharge records (SMR01), maternity and birth records (SMR02), psychiatric inpatient records (SMR04) and cancer registrations (SMR06). In addition we have accessed child immunization and health records (SIRS and CHSP-PS) and Mental Welfare Commission records, which are not in the ISD 'linked' dataset. We have linked the census, which contains ethnic codes, to all these data using methods reported in detail elsewhere.[15,19]

The Registrar General of Scotland judged that under present legislation, namely the Census Act 1920 and the Census Confidentiality Act 1991, the linkage described here was acceptable if the confidentiality

of individual personal information was maintained during linkage and analysis. We used both one-way encryption methods ('hashing') and organizational procedures already published[19]; e.g. GROS required that the linkage was performed on a stand-alone computer in a locked 'safe haven' within GROS premises. Modifications were made to the computer operating system to monitor activity and all peripheral devices were disabled. The room at GROS was accessed with the agreement of GROS, ISD and the PI (Bhopal). GROS maintains a register of visits. We obtained approval from the Scottish Multi-centre Research Ethics Committee (MREC) and the Privacy Advisory Committee (PAC) which advises ISD and GROS on the use of data.

This work was supported by the Commission for Racial Equality, NHS Health Scotland (through The National Resource Centre for Ethnic Minority Health), ISD, the Scottish Executive Health Department and the Chief Medical Officer. An ethical opinion, subsequently published, was given by a Professor of Medical Ethics.[20] For funding see acknowledgements.

Figure 1, republished from open access ref.,[19] illustrates in concept how record linkage was based on the use of subsets of information from three datasets: healthcare records, which include personal identifiers and clinical information; the Community Health Index (CHI) dataset which contains personal identifiers and the CHI number; and the census file which contains personal identifiers and details of individuals' ethnicity. The CHI dataset lists everyone registered with a General Practitioner or eligible for NHS screening services and forms a unique identifier for NHS use. More than 99% of the Scottish population is listed on the CHI.

Date of birth, surname (using soundex codes to allow for variations in spelling), forename, address and full postcode, which were available in both datasets albeit not always recorded identically, were used to link the census number to the CHI. At this stage, other data fields in the two datasets were disconnected from identifying variables. CHI and the census unique number were encrypted prior to linkage. A one-way cryptographic ('hashing') algorithm (currently impossible to reverse) was used to encrypt the CHI number. The census number was encrypted using an algorithm developed by GROS. For the records deemed to be matches, 73.6% were exact matches. For the remainder, a probability matching process was performed. Here, the rate of false positives is critical. Methods have been developed to identify how false positives occur and what kind of strategies a human checker employs to decide whether a pair match is 'good'.[16] These decision strategies were built into a 'partitioning' computer algorithm. These 'partitions' then allow the allocation of effort to the most profitable 'partitions' which yield the lowest false-positive rates and highest true-positive rates.

Once the linkage was completed, personal identifying variables (such as names, address, postcode and dates of birth) were removed, leaving a file with an encrypted CHI number and its corresponding encrypted census number (look-up file). A census extract containing ethnic code (and limited other data including age, sex and indicators of socio-economic status (see Table 1)) and an encrypted census number was joined to the above look-up file using the encrypted census number. The encrypted census numbers were then discarded leaving the ethnicity code, some other variables from the census and the encrypted CHI number. The relevant parts of the ISD linked database were linked via the encrypted CHI numbers. The encrypted CHI was replaced with an unrelated serial number (to keep together the multiple records on the same people, known as hospital admission spells), resulting in depersonalized clinical health records carrying census ethnicity codes. Using methods previously described, we estimated an upper limit to the false-positive linkage rate of 0.08%.[19]

What does it cover?

For Phase 1 (demonstration project), we studied myocardial infarction, since it is one of the most important conditions in Scotland in terms of both mortality
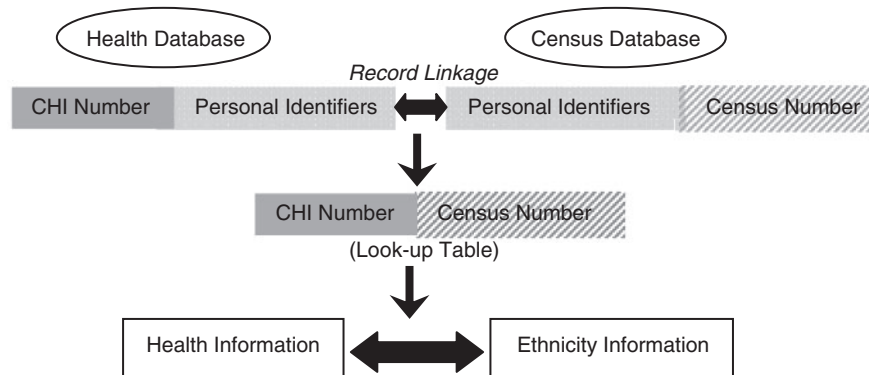

Figure 1 Overview of record linkage process

Table 1 Baseline 2001 variables extracted from census: Phase 2

Ethnic group
Religion, current
Religion of upbringing
Country of birth
Age
Sex
Long-term illness
Self-assessed health
Marital status
Labour force status
Socio-economic status
Highest qualification
Scottish Index of Multiple Deprivation decile
Car ownership
Housing tenure
Household size
Number of rooms
Urban/rural indicator
Health board (Glasgow, Lothian, Tayside, Other)
Mobile (temporary) accommodation
Self-contained accommodation
Central heating
Moved within last year
Activity last week

and morbidity, is common in ethnic minority groups, and likely to be recorded in hospital morbidity records.[19,21] For Phase 2, we are studying cardiovascular disease, maternal and child health, cancer and mental health—see Table 2. We are currently planning Phase 3, studying a wider range of conditions and linkage to primary care records. In theory, we can study long-term outcomes for any health problem coded in NHS records (and in the records of any other organizations that are able and willing to provide data for linkage). In practice, our work is heavily governed and controlled by data privacy and other ethical considerations, and practical matters of resource, expertise and interest.

Who is in the sample?

The objective was to link all 5 million or so people resident in Scotland in April 2001, the time of the census, but we set an arbitrary standard of ⩾80% in each ethnic group, on the basis of a judgement that potential biases would be too great for linkages less than that. About 94% overall were linked in the Phase 1 demonstration project using the methods described below, and 95% in Phase 2, with the minimum standard of 80% linkage met for every ethnic group (see Table 3), thus creating a retrospective cohort study of ~4.65 million people. Taking into account the census enumeration rate (96%), this represents ~91% of the population resident in Scotland in April 2001.

How often have they been followed up and what has been measured? How many events related to the health outcomes are there?

The cohort focuses on ethnicity (as reported on the census form completed by the head of household) as the primary exposure variable. Other census information is included, and is listed in Table 1. Religion (both current and that of upbringing) and country of birth is added to the study database and used to provide a more comprehensive assessment of the relationship between an individual's health and their ethnic group. Socio-economic status is an important co-variable that could potentially explain some aspects of ethnic variation. Accordingly, a careful analysis of the influence of socio-economic status is being built into the analysis. Census variables relating to employment and educational attainment, both of the individual and the head of household, together with area-based socio-economic indicators, are used to build up a comprehensive picture of socio-economic status. The Scottish Index of Multiple Deprivation (SIMD), based on area of residence, and included in many Scottish health databases, is used. This will enable the complex relationships between health, ethnicity and socio-economic status to be elucidated. In addition, data on self-reported long-standing illness are available from the census. We also have travel times from home to the nearest hospital.

The outcomes currently under study are given in Table 2. The number of events varies, obviously, by ethnic group, disease outcome, and number of years of follow-up. Scotland's non-White ethnic minority population is small at ~2%. Table 3 gives the number of people linked by ethnic group. For White minority populations, e.g. other British and Irish, the population size and hence number of events is large. In Phase 2, we are analysing 7 years of follow-up for most events. In Phase 3, currently planned, we anticipate 9 years of follow-up. For most outcomes, given the degree of ethnic inequality, our analyses are statistically robust. For many endpoints, ethnic inequalities are large, with 2-fold-plus risks being common. Differences of interest (e.g. a 50% difference) can be detected for most end points. For example, asthma is a relatively important but rare cause of hospitalization and Indians comprised 0.3% of the Scottish population in 2001 (compared with Pakistanis, who were 1%). There are more than 3500 hospital discharges and deaths for asthma per year. With 8 years of outcomes we accrue ~28 000 admissions. Assuming 2% of them are in non-White populations, and that is conservative given previous studies showing a relative excess of hospitalization in minority populations,[22] that makes 560. If 0.3% of the total hospitalization/deaths are in Indians, that is ~84 admissions. With

Table 2 Primary endpoints and datasets for Phase 2 project

Priority area	Health outcome/disease/condition	Service use	Outcome	Datasets
Cardiovascular	Angina Chest pain MI Heart failure Stroke All CHD All CVD	Angiography Angioplasty CABG Length of hospital stay	Survival following AMI 30-day survival following AMI	SMR01, GROS mortality records
Cancer	Breast cancer Colorectal cancer Lung cancer Prostate cancer	Uptake of breast cancer screening Stage at presentation Treatment Place of death	Survival following first registration of: Breast cancer Colorectal cancer Lung cancer Prostate cancer (stage and grade to be included in analyses where possible)	SMR01, SMR06, GROS mortality records, Scottish Breast Screening Programme
Maternal and child health	Age of mother at birth Smoking at first visit Premature births Abortion Miscarriage Perinatal mortality Hospital admissions Childhood obesity	Mode of delivery Analgesia at delivery Immunization Child health visits completed	Breastfeeding Birth-weight Height/weight of infant at health visits	SMR01, SMR02 Child health surveillance data
Mental health	All psychiatric hospital admissions including Intentional self-harm Suicide	Length of hospital stay Admission type detention Discharge-type Readmissions	Survival to second episode of self-harm Survival following attempted suicide Time between admission(s)	SMR01, SMR04 MWC detention data GROS mortality records

AMI: acute myocardial infarction, CHD: coronary heart disease, CVD: cerebrovascular disease, CABG: coronary artery bypass graft, SMR: Scottish morbidity record, MWC: Mental Welfare Commission.

Table 3 Linkage rates by ethnic group

		Number	Percentage
1	White Scottish	4 290 153	95.3
2	Other White British	357 788	93.6
3	White Irish	47 173	92.2
4	Other White	74 655	87.9
5	Any mixed background	12 117	91.7
6	Indian	13 717	89.9
7	Pakistani	28 538	89.8
8	Bangladeshi	1783	88.0
9	Other South Asian	5810	85.1
10	Caribbean	1659	89.5
11	African	4514	86.5
12	Black Scottish or Other Black	1057	89.1
13	Chinese	15 115	87.4
14	Other ethnic group	8945	86.2

84 expected events, we can detect a difference of 44% in the rate, with ~80% power and 0.05 statistical significance. In Phase 2, we have already demonstrated that robust statistical analyses is possible within this dataset for many of the cardiovascular and maternal/child health outcomes in the table (analyses on mental health and cancer are underway but we cannot comment at this point).

What is attrition like?

The 5% of the Scottish population that is not linked is being characterized to assess potential biases by comparing with the 95% that was linked. It is not possible to compare with the additional 4% of the Scottish population that was not enumerated at the 2001 census. People who migrate, or are hospitalized or die, outside Scotland are poorly characterized in our databases and, to date, we have not developed a sound strategy for estimating the potential biases. We particularly welcome collaborations on this challenge.

What has it found? Key findings and publications

Overall, 94% in Phase 1 and 95% in Phase 2 of the 4.9 million census records were matched to a CHI record with an estimated false-positive rate of <0.1%. Phase 1 data showed that between April 2001 and December 2003, there were 126 first episodes of acute myocardial infarction (AMI) among South Asians and 30 978 among non-South Asians. The incidence rate ratio was 1.45 [95% confidence interval (CI) 1.17–1.78] for South Asian compared to non-South Asian men and 1.80 (95% CI 1.31–2.48)

for South Asian women. After adjustment for age, sex and any previous admission for diabetes the hazard ratio for death following AMI was 0.59 (95% CI 0.43–0.81), reflecting better survival among South Asians.[19] Phase 2 analyses of cardiovascular diseases are underway and will be submitted for publication shortly. They seem to both corroborate and extend Phase 1 findings e.g. confirming increased incidence rather than increased case fatality in explaining the high CHD mortality rate in the South Asian population and showing heterogeneity between Indians and Pakistanis. For example, directly standardized rate ratios for first MI (in the period May 2001 to April 2008) in males were highest for Pakistanis, followed by Other South Asians and Indians. For females, rates were higher for Other South Asians, followed by Pakistani and Indian. Pakistani males had significantly higher rate ratios of first MI than their White Scottish counterparts. After adjustment for diabetes and the Scottish index of social and economic deprivation (SIMD), Pakistani males and females had significantly better survival 28 days after first MI than the White Scottish population. Analyses relating to Phase 2 goals on cardiovascular disease, cancer, maternal and child health and mental health will be reported in 2010.

What are the main strengths and weaknesses?

We had to balance individuals' right to data privacy (individual consent was impossible) and the potential benefits to society of producing information derived from potentially sensitive data.[20] We and a professional ethicist[20] judge that our methods strike a balance. The proportion of records successfully linked was only slightly lower than that normally achieved within ISD in internal linkages (typically ~98%), even although we linked data held by two separate agencies. As the non-White ethnic minority population (~2%) is comparatively small, the false positive rate is critical. A rigorous matching methodology appropriate for administrative matching was, therefore, used.

Census forms are completed by the public and processed electronically using optical recognition and keying from images. The success of these processes depends on legibility. The detailed spelling of a name, particularly if it is associated with a minority ethnic group, is more likely to be prone to error when transcribed by a third party onto NHS records, than in the census where the census informant writes it. Such errors could lead to varying accuracy of linkage by ethnic group. Nevertheless, we met our prior stated standard of 80% for every ethnic group, lower in non-White than White populations. People who are not matched are not included in our analyses. At present, we do not know about the characteristics of those matched compared

with those not matched, but work on this is underway.

The major strength of this analysis is the population coverage with the inclusion of community and hospital deaths as the SMR database contains generally validated and complete data, e.g. over 99% of hospital admissions for AMI in Scotland (A. Redpath, ISD, personal communication, 2006). Possible, as opposed to probable or definite, AMIs are coded as chest pain and are excluded from AMI figures, so the AMI figures could be underestimates. The quality of data on AMI in the database has been validated as reliable.[19] Information about the quality of SMR01, SMR04 and SMR06 is available.

Validation has not, however, been done by ethnic group. For example, the diagnosis of AMI was based on criteria used by clinicians to make the diagnosis, which would be inconsistent over time. This should not, however, affect ethnic groups differently. If misclassification of diagnosis is non-differential then the differences would probably be underestimated, but if they were differential, differences would be exaggerated. We do not have data to assess these options, and the anonymized methods precluded validity studies. We do know that South Asian populations in the UK, and clinicians looking after them, are sensitized to the high rate of CHD, and early admission and treatment,[23,24] concordant with excellent survival,[19] is likely.

Linkage rates were slightly lower for South Asians than non-South Asians. If non-linkage occurred at random, this would reduce the power of the study, but not bias the results. Factors reducing the power of the study are particularly important for small ethnic minority populations and rare outcomes. If those not matched were at different risk of hospitalization or death from those who were matched, which seems quite likely, this would bias the results. We intend to explore this in future research.

Potential sources of bias in the analyses have not yet been assessed systematically. A separate data linkage evaluation exercise is planned in collaboration with NHS Central Register to provide alternative linkage outcomes for a small sample of records that can help assess the accuracy of the main data linkage project. A comparison of census characteristics of matched and non-matched individuals is also planned. Quantifying the potential impact of deaths abroad, to account for the possibility that minority ethnic groups may be more likely to move abroad when they are elderly or ill with chronic conditions (the so-called salmon bias) is proving problematic. This bias would improve apparent survival. Such an effect is unlikely to be large, as reflected in mortality by country of birth analyses in England and Wales showing that mortality is not uniformly low in most ethnic minority groups, but we will scrutinize our results to detect such bias (e.g. as reflected in relatively good long-term compared to short-term survival).

The statistical power is governed by the size of the population and event rate which determine number of cases, and the size of the differences between comparison groups. At the 2001 Census ~2% of the Scottish population was not White (~100 000 people) and most of them were South Asian (Pakistani and Indian) or Chinese. These groups will be our main foci for analysis, although where there is sufficient statistical power for other ethnic groups we will examine them, e.g. of African ethnicity. We are also paying attention to White minorities in Scotland, e.g. the other UK (mainly English and Irish) populations. In our Phase 2 analysis and interpretation, we are examining heterogeneity of broadly defined ethnic groups, e.g. South Asian. For these non-White ethnic groups, we are finding major differences in event rates in relation to comparison populations. One limitation is that we are unable to study Eastern Europeans and any other ethnic groups which were not separately identified in 2001. (This kind of work may be a focus for future work based on census 2011.)

Power calculations (80% power at the 5% level of significance) have shown that differences of a size that are common and of interest (e.g. a 50% increase) can be detected for most end-points, for the three largest non-White minority populations, namely Pakistanis, Indians and Chinese. Our calculations assume 5 years of follow-up data but for most outcomes we have 7 years, and this will increase in future linkages.

Can I get hold of the data? Where can I find out more?

This project has created a retrospective cohort of ~4.6 million people living in Scotland at the time of the 2001 Census. Phase 2 work is focusing on cardiovascular disease, cancer, mental health and maternal and child health. Subject to ethical approval, future work will have potential to examine healthcare procedures and adjust for other variables that are available in the census. Research using the datasets requires approval from GROS, ISD, a multicentre research ethics committee, Privacy Advisory Committee and the Directors of Public Health. Approvals have been given *only* to examine ethnic variations in health and healthcare.

The researchers appreciate the benefits of collaboration and sharing of data, and are in principle open to proposals for relevant collaborative analysis of datasets. There are two products of this data linkage project that may be of value to other researchers:

- The CHI-census linkage file, containing the encrypted census identifier numbers, and encrypted CHI numbers (dataset #1); and
- The matched, anonymized dataset containing census variables and health records (dataset #2).

Use of the data files requires the use of specialist software and the expertise of research team members, not least in obtaining necessary approvals. Collaboration will be pursued if the prime objectives of the project are not impaired or delayed in the four areas under study in Phase 2. Collaborators would be primarily responsible for funding the additional research, including the cost of extra work for the core research team. We are committed to seek ways of increasing access to the database.

Conclusion

Our approach, potentially, has international applicability. It demonstrates how the glaring absence of cohort studies reporting by ethnic group in Europe can be overcome.[25] There is considerable potential in linking databases that have previously been considered too sensitive for record linkage or where linkage is restricted by data protection legislation. The methods described here and in more detail in our report and publication[19,21] have the potential to fill the information gap on data by ethnic group. This gap will persist until we have high-quality prospective ethnic group coding systems in healthcare databases and the inclusion of a valid ethnic code on birth and death registration—both formidable long-term challenges, hitherto unachieved in either Europe or North America.[1] It is noteworthy, however, that prospective ethnic coding will not provide the detailed socio-demographic data available in the census, and so will not replace our method.

Funding

Acknowledgements

References

1 Bhopal RS. *Ethnicity, Race, and Health in Multicultural Societies; Foundations for Better Epidemiology, Public Health, and Health Care*. Oxford: Oxford University Press, 2007.

2 Scottish Executive. *Fair for All: Working Together Towards Culturally-Competent Services*, NHS HDL(2002)51. Edinburgh: Scottish Executive, 2002.

3 The Home Office. *Race Relations (Amendment) Act 2000: New Laws for a Successful Multi-Racial Britain*. London: Home Office Communication Directorate, 2001.

4 *The Duty to Promote Race Equality: A Guide for Public Authorities in Scotland*. UK: Commission for Racial Equality (CRE), 2002.

5 Department of Health, Health and Social Care Information Centre, NHS employers. *Practical Guide to Ethnic Monitoring in the NHS and Social Care*. UK: London, Department of Health (DOH), 2005. http://www.hcna.b-ham.ac.uk/documents/04_HCNA3_D4.pdf (1 July 2010, date last accessed).

6 NHS Health Scotland and NHS National Services Scotland. *Ethnic Monitoring Toolkit & Communication Guideline*. Edinburgh: NHS Health Scotland, 2005. http://www.isdscotland.org/isd/5875.html (1 July 2010, date last accessed).

7 Gill PS, Kai J, Bhopal RS, Wild S. Health care needs assessment: black and minority ethnic groups. Chapter 4. In: Rafferty J (ed.). *Health Care assessment. The Epidemiologically Based Needs Assessment Reviews*. Third Series. Abingdon: Radcliffe Medical Press Ltd, 2006; pp. 227–339; http://www.hcna.bham.ac.uk/documents/04_HCNA3_D4.pdf.

8 Bhopal R. What is the risk of coronary heart disease in South Asians? A review of UK research. *J Public Health Med* 2000;**22:**375–85.

9 Bahl V. Cancer and ethnic minorities – the Department of Health's perspective. *Br J Cancer Suppl* 1996;**29:**S2–10.

10 Bhopal RS, Rankin J. Cancer in minority ethnic populations: priorities from epidemiological data. *Br J Cancer Suppl* 1996;**29:**S22–32.

11 Wild SH, Fischbacher CM, Brock A, Griffiths C, Bhopal R. Mortality from all cancers and lung, colorectal, breast and

prostate cancer by country of birth in England and Wales, 2001–2003. *Br J Cancer* 2006;**95:**424.

[12] Collingwood BA. Investigating variations in infant mortality in England and Wales by mother's country of birth, 1983–2001. *Paediatr Perinat Epidemiol* 2006;**20:** 127–39.

[13] Gardosi J, Chang A, Kalyan B, Sahota D, Symonds EM. Customised antenatal growth charts. *Lancet* 1992;**339:** 283–87.

[14] Sashidharan SC. *Inside Outside: Improving Mental Health Services for Black and Minority Ethnic Communities in England*. England: Department of Health, 2003.

[15] Bhopal R, Fischbacher CM, Steiner M *et al*. *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*. Edinburgh, University of Ediburgh. http://www.chs.med.ed.ac.uk/phs/research/ Retrocoding%20final%20report.pdf (1 July 2010, date last accessed).

[16] Newcombe H. *Handbook of Record Linkage: Methods for Health and Statistical Studies, Administration and Business*. Oxford: Oxford University Press, 1988.

[17] Jeffreys M, Stevanovic V, Tobias M *et al*. Ethnic inequalities in cancer survival in New Zealand: linkage study. *Res Practice* 2005;**95:**834–37.

[18] Harding S, Balarajan R. Mortality of third generation Irish people living in England and Wales: longitudinal study. *Br Med J* 2001;**322:**466–67.

[19] Fischbacher CM, Bhopal R, Povey C *et al*. Record linked retrospective cohort study of 4.6 million people exploring ethnic variations in disease: myocardial infarction in South Asians. *BMC Public Health* 2007;**7:**142.

[20] Boyd KM. Ethnicity and the ethics of data linkage. *BMC Public Health* 2007;**7:**318.

[21] Bhopal R, Fischbacher CM, Steiner M *et al*. Can we retrospectively analyse Scotland's health databases by ethnic group? A demonstration project focusing on linkage of census and health records. Conference Assistant. *Social Dimensions of Health Institute, International Conference, 18–20 September 2007, Universities of St. Andrews and Dundee*. Social Dimensions of Health Institute, 2007.

[22] Netuveli G, Hurwitz B, Levy M *et al*. Ethnic variations in UK asthma frequency, morbidity, and health-service use: a systematic review and meta-analysis. *Lancet* 2005;**365:** 312–17.

[23] Chaturvedi N, Rai H, Ben-Shlomo Y. Lay diagnosis and health-care-seeking behaviour for chest pain in south Asians and Europeans. *Lancet* 1997;**350:**1578–83.

[24] Ben-Shlomo Y, Naqvi H, Baker I. Ethnic differences in healthcare-seeking behaviour and management for acute chest pain: secondary analysis of the MINAP dataset 2002-2003. *Heart* 2008;**94:**354–59.

[25] Ranganathan M, Bhopal R. Exclusion and inclusion of nonwhite ethnic minority groups in 72 North American and European cardiovascular cohort studies. *PLoS Med* 2006;**3:**e44.