Does the "Scottish effect" apply to all ethnic groups? All-cancer, lung, colorectal, breast and prostate cancer in the Scottish Health and Ethnicity Linkage Cohort Study

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-001957
Article Type:	Research
Date Submitted by the Author:	13-Aug-2012
Complete List of Authors:	Bhopal, Raj; University of Edinburgh, Centre for Population Health Sciences Bansal, Narinder; University of Edinburgh, Centre for Population Health Sciences Steiner, Markus; University of Aberdeen, Brewster, David ; NHS Scotland National Service, Information Services Division,
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Oncology, Public health
Keywords:	EPIDEMIOLOGY, Breast tumours < ONCOLOGY, Gastrointestinal tumours < ONCOLOGY, Respiratory tract tumours < ONCOLOGY, Urological tumours < ONCOLOGY

Does the "Scottish effect" apply to all ethnic groups? All-cancer, lung, colorectal, breast and prostate cancer in the Scottish Health and Ethnicity Linkage Cohort Study

Short title: Cancer by ethnic group in Scotland

Bhopal R[1], Bansal N[1], Steiner M[1,2], Brewster DH[1, 3] on behalf of the Scottish Health and Ethnicity Linkage Study

[1]Edinburgh Ethnicity and Health Research Group, Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG

[2]Environmental & Occupational Medicine, Section of Population Health, University of Aberdeen, Polwarth Building 1st floor, Foresterhill, Aberdeen AB25 2ZD

[3]Information Services Division, NHS Scotland National Services, Gyle Square, 1 South Gyle Crescent, Edinburgh EH12 9EB

Correspondence: Professor Raj S. Bhopal, The Centre for Population Health Sciences, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, Scotland, UK. Tel: 0131 650 3216; Fax: (0131) 650 6909; Raj.Bhopal@ed.ac.uk.

Word count summary: 297

Abstract

Background and Objectives

While ethnic group variations in cancer exist, no multi-ethnic, population based, longitudinal studies are available in Europe. Our objectives were to examine ethnic variation in all-cancer, and lung, colorectal, breast and prostate cancers.

Design, Setting, Population, Measures and Analysis

This retrospective cohort study of 4.65 million people linked the 2001 Scottish Census (providing ethnic group) to cancer databases. With the White Scottish population as reference (value 100), directly age standardised rates and ratios (DASR and DASRR), and risk ratios, by sex and ethnic group with 95% confidence intervals (CI) were calculated for first cancers. In the results below the 95% CI around the DASRR excludes 100. Eight indicators of socio-economic position were assessed as potential confounders across all groups.

Results

For all-cancers the White Scottish population (100) had the highest DASRRs, Indians the lowest (men 45.9, women 41.2), and White British (men 87.6, women 87.3) and other groups were intermediate (e.g. Chinese men 57.6). For lung cancer the DASRRs for Pakistani men (45.0), and women (53.5), were low and for Any Mixed Background men high (174.5). For colorectal cancer the DASRRs were lowest in Pakistani (men 32.9 women 68.9), White British (men 82.4, women 83.7), Other White (men 77.2, women 74.9), and Chinese men (42.6). Breast cancer in women was low in Pakistanis (62.2), Chinese (63.0) and White Irish (84.0). Prostate cancer was lowest in Pakistanis (38.7),

Indian (62.6) and White Irish (85.4). No socio-economic indicator was a valid confounding variable across ethnic groups.

Conclusions

The 'Scottish effect' does not apply across ethnic groups for cancer. The findings have implications for clinical care, prevention and screening e.g. responding appropriately to the known low uptake among South Asian populations of bowel screening might benefit from modelling of cost-effectiveness of screening, given comparatively low cancer rates.

Article Summary

Article focus

The Scottish Health and Ethnicity Linkage Study examined whether all cancers, and lung, colorectal, breast and prostate cancer separately, in the period 2001-2008, varied by 2001 Scottish Census ethnic group categories.

Key messages

- The main public health lesson and challenge is for the majority population, for the "Scottish effect" in relation to cancer does not apply across Scotland's ethnic groups.
- This exemplifies how the study of ethnic variations provides a public health approach with potential to benefit the entire population.

Strengths and limitations of the study

- The strength of the study is the development of a retrospective cohort with high overall linkage rates in a national population; the exploration of the potential role of socio-economic variables and country of birth available in the Census; and the linkage of Census data to both cancer registry and community/hospital mortality data.
- The limitations include the small numbers of outcomes for some non-White populations, and the consequent aggregation of some ethnic groups; variation in linkage rates by ethnic group; inability to capture events that occur overseas outside the UK; and lack of individually linkable cancer risk factor data.

Introduction

Cancer is a dominant cause of death in industrialised countries, and particularly common in Scotland(1) Cancer incidence varies hugely across countries, between country of birth/ethnic groups and over time, thus clearly indicating that the causes of cancer are largely environmental. Examination of such variations, including by country, by country of birth and when possible ethnic group, has proven of value both in sparking causal research, and assessing disease burden, health care priorities and patients' needs.(2)

Given international variations, it is not surprising that major differences in cancer frequency are demonstrable by ethnic group.(2;3) Ethnic group studies on cancer have mostly utilised the proxy indicator country of birth, which is usually available in both population registries and censuses (supplying denominators) and sometimes in cancer and death registration systems. The limitations of this proxy have been discussed elsewhere,(4;5) including that, especially in European countries with colonial histories such as Scotland, many of the elderly were born abroad, and substantial proportions (often 50% or more) of resident ethnic minority populations are not born abroad. Name search methods are also popular(6-9) but have even more limitations e.g. they are not good for studying White minority groups and African and Caribbean origin Black populations.(4;5) A recent survey of European cancer registries concluded that while self-reported ethnicity was the exemplary variable, none of 79 registries analysed data this way, with Scotland being closest to achieving this goal.(4)

Within multi-ethnic countries proper ethnic group data are needed to maintain valid surveillance of cancer trends and inequalities, to set priorities, to ensure equitability of service delivery and to further develop hypotheses on causation.(10) The few studies that

use reported ethnic group in Europe may have a high proportion with missing ethnicity. The best such studies combine this with country of birth (3;8). Such is the scarcity of data that a 2007 paper reported on observer assigned ethnic group on 2713 people followed up for 19.9 years, yielding six cases in South Asian men, and 26 in African-Caribbean men.(11) Linkage of cancer registration and hospital episode statistics (providing ethnicity) in England is demonstrating the importance of this approach, despite some limitations e.g. missing data.(12-15) Most available studies in Europe analyse data at a point or period of time i.e. cross-sectional analyses using numerators and denominators from different sources, creating potential errors in calculations of rates.(3;16) The field is developing internationally with recent work using name search methods in Canada(6) and linkage methods in New Zealand(17), with interest in multination comparisons for specific ethnic groups.(18)

Ethnic variations in cancer, mostly using country of birth(3), have been noted with, for example, comparatively lower mortality for all combined and four major cancers in South Asian migrants in England and Wales(3;16) Studies based on country of birth(16) and ethnicity data in England and Wales(12) support the role of environmental factors in explaining this variation. There is evidence of change over time and across generations, though cancer inequalities persist, some narrowing, others widening.(19)

The Scottish Health and Ethnicity Linkage Study compared all cancers (without non-melanoma skin cancers), and lung, colorectal, breast and prostate cancer separately, in the period 2001-2008 by ethnic group categories as reported in the 2001 Scottish Census.(20) These cancers were chosen as the commonest cancers in Scotland, prioritised by national public health strategy.(1)

Scotland has a higher incidence of cancers compared to England and Wales and people born in Scotland living in England and Wales also have comparatively high rates.(16;19) The background information on Scotland's health services, cancer data systems, the ethnic mix of the population and previous research on Scottish populations by ethnic group has been summarised recently by Arnold and Brewster(4) (Ch 4.4 in ref no 4). Data on cancer by ethnic group in Scotland are old, limited in scope and from small-scale studies(2;9;21;22) focusing solely on Chinese, South Asians and Italians and published in the 1980's and early 1990's. These studies are summarised in box 1.

This paper reports new, more comprehensive data from Scotland using a national, retrospective cohort study. It also includes both an examination of the potential for adjusting for socio-economic confounding and studying the effects of country of birth in relation to ethnicity. Finally, using risk factor data from Health Survey for England and Scottish Health Survey we interpret our results indirectly (in the absence of linkable risk factor data).

Methods

The methods of our retrospective cohort study are published, and key details on linkage are also given in appendix 1.(20;23) We followed a strict protocol that preserved anonymity and maintained separation of personal data from the Census and NHS, and clinical data (see also ethics below). We used computerised matching of names, addresses and dates of birth to link the Census 2001 for Scotland , which provided ethnic group as reported by either individuals or the householder completing the form based on a question followed by a choice of 14 categories (appendix 1, table A1, which also provides linkage

by ethnic group), and other demographic and socio-economic variables, to the Scottish Community Health Index (CHI), which is a register of patients using the NHS. We then matched, using CHI number, to an already linked death in the community and hospital, and cancer registration records (SMR06) database.

Ethnic group is a legally required field that was well completed (95.8 %) and, after imputation (4.3%), available for 100% of those completing the census form (which is also a legal obligation). (For details see: http://www.gro-scotland.gov.uk/census/censushm/censcr02/data-quality/census-variables/results-and-conclusions/appendix-d-person-items-reports-and-tables-p10-to-p-17.html; accessed 26th of April 2012). About 95% of the people participating in the 2001 census (4.9 million) were linked as above to health records, i.e. 4.65 million, with 85% or more linked in every ethnic group(20) (see appendix 1). The total estimated Scottish population was 5.06 million so our cohort of 4.65 million includes about 92% of the 2001 population. While the identities of those not completing a census form are unknown it is estimated in census validity studies that a higher proportion of Non-White than White groups were non-completers – estimated at, for example, 10.2% of Pakistanis and 3.8% of White Scottish.

The ethnic group categories (and labels) follow those of the Scottish Census 2011, given in appendix 1.(20) Because of small numbers we grouped Bangladeshis with Other South Asians; and Caribbean, African and Black Scottish or Other Black, into one "African origin" group. Further grouping was sometimes necessary because of small numbers in analysis of specific cancers as described in the results. Mostly, following our analytical strategy, ethnic groups are sometimes omitted to avoid potential disclosure of identity.

About 90% of the cases were obtained from the cancer registry, 10% from mortality files. Cancers are registered at diagnosis, so mortality data add cases where the diagnosis was first made outside Scotland, which is especially important for mobile ethnic minority groups. A date of embarkation field is in the registry but we did not think this was reliable enough in relation to non-UK migration to use to adjust denominators. More than 90% of the Scottish Cancer Registry records for 2001-2008 were linked to our census extract file. We excluded non-melanoma skin cancer. The ICD codes used are in box 2. Other non-cancer health outcomes were excluded from the analysis file for reasons given in the ethics section below.

To minimise the numbers of age/sex cells with no cases, which creates instability in the analysis, we restricted analysis by age as follows: ≥ 20 years for all cancer; ≥ 30 years for lung cancer; ≥ 20 years for breast cancer; and ≥30 years for colorectal and ≥40 years for prostate cancer. This led to few omissions, ranging from 0.1% to 1.9 % depending on the specific diagnosis.

We analysed only first events i.e. newly diagnosed cancers occurring between 2001 and 2008. First event meant there was no record of the cancer diagnosis under study in the preceding 10 years in the mortality and cancer registration (SMR06) linked file. The cancer registry collects data from a range of sources including pathology laboratories, so our cases are likely to be new ones.

We calculated for first cancers for all and each cause, by sex: directly age standardised cumulative incidence rates (DASRs) per hundred thousand per year using 10-year age groups; DASR ratios (DASRRs); risk ratios (RRs) using Poisson regression with robust

variance adjusting for age and country of birth; and 95% confidence intervals (CI) around summary measures. To assess effects of out-migration we calculated RR using moving average for 3-year time periods 2001-2004, 2002-2005 etc. In appendix 2 we provide details of our approach to calculating rates and risk ratios, including details of the Poisson modelling. The standard reference population was the White Scottish population. For ease of interpretation we multiplied ratios by 100 to get whole numbers interpretable as percentages. We adjusted the RRs for country of birth being Scotland or outside Scotland. Relatively few cases in ethnic minority populations were born in Scotland e.g. for all cancers excepting non-melanoma, the proportion was 5.1% in Other White British, 11.2% in Indians, 18.5% in Pakistani, 8% of Chinese, and 36% of African origin groups. In the small Any Mixed Background group 64.7% were born in Scotland. For this reason i.e. statistical precision, analysis is not stratified by country of birth.

We examined, in each ethnic group, whether there was an association between eight indicators of socioeconomic position and all cancer rates (at all ages) and hence whether any were potentially valid confounding factors across all our ethnic groups. The indicators were: 1) the postcode (zipcode) based Scottish Index of Multiple Deprivation, 2) car ownership, 3) highest qualification of the individual, 4) highest qualification in the household, 5) National Statistics Socio-economic Classification at individual and 6) household levels, 7) household tenure and 8) economic activity in the previous week (of the Census completion date).

Data were analysed using SAS version 9 (SAS Institute Inc. Cary NC. USA) and Stata 11 (StataCorporation 2009; Statistical Software: Release 11.0; College Station, TX).

In the results text below we provide both absolute (DASRs) and ratio (DASRRs and RRs) measures and describe findings where the 95% CI does not include 100, the value for the reference White Scottish population.

Ethics and disclosure

The work was approved by the Multicentre Research Ethics Committee for Scotland and the Privacy Advisory Committee of NHS National Services Scotland. The ethical and other permissions and related issues have been reported in detail,(20;23) including an independent assessment by an ethicist.(24) To comply with the Data Protection Act and safe-setting rules the data set only contained cancer outcomes. Other outcomes were excluded to minimise risks of inadvertent disclosure of identity. The analysis was conducted on a standalone computer in a locked room in the General Register Office for Scotland (GROS), now known as National Records Scotland, by named researchers (NB, MS,GB – see contributors), following a strict disclosure protocol. Outputs leaving the safe setting (including this paper) were screened by a GROS disclosure committee.

Results

All cancers without non-melanoma skin cancer

Table 1 and figure 1 show that in men and women, with the exception of men in the Any Mixed Background group (where the 95% CI included the reference value), the White Scottish population had the highest rates and ratios of cancer (DASRR of 100 by definition), above even other White groups. The rates (and DASRRs) were particularly low in Indian (45.9 in men, 41.2 in women), Pakistani (49.3 in men, 65.0 in women) and Chinese (57.6 in men) populations. Including country of birth as a co-variate, as shown by comparing the age-adjusted and age and country of birth adjusted risk ratios (table 1),

only slightly altered these patterns, though in this analysis 95% CIs were more likely to include the reference value. Generally, this adjustment closed the gap slightly between the reference and each comparison population.

As shown in appendix 3, table A2, excepting for the African origin group, and Other South Asian women, RRs were similar in the time period 2001-4, 2002-5, 2003-6, 2004-7, and 2005-8, indicating that, with the few exceptions above, unmeasured, differential emigration was not underlying these ethnic variations.

Lung cancer

Table 2 and figure 2 shows that with the exception of the White Irish (similar), and any mixed background men (higher), all other ethnic groups had lower lung cancer standardised rates (and ratios) than the White Scottish population. The low DASRR for Pakistani men (45.0) and Chinese men (63.1) and a high DASRR for Any Mixed Background men (174.5) were notable. The DASRs show that, in every group excepting Chinese, men had much higher rates of lung cancer than women.

Including country of birth as a covariate raised the risk ratios in every ethnic group, indicating Scottish born people in these ethnic groups are at higher risk of lung cancer than those born abroad.

Colorectal Cancer

Table 3 and figure 3 shows large differences by ethnic group, with the highest DASRs for colorectal cancer in White Scottish and Irish men. Pakistani men (DASRR = 32.9) and women (68.9) and Chinese men (42.6) had very low ratios with Other White British (82.4

in men, 83.7 in women) and Other White (77.2 in men, 74.9 in women) groups being intermediate. (Data for Indians are omitted for risks of disclosure reasons, but the results have been examined and the pattern is similar to that in Pakistanis.)

Including country of birth as a covariate made little difference to the patterns observed e.g. the risk ratio in Pakistani men changed from 45.6 to 46.4.

Breast cancer in women

Table 4 and figure 4 show large ethnic variations (but, for once, no advantage to the Other White British population). White Irish populations (84.0) had lower DASRRs than the White Scottish population but DASRRs were especially low for Pakistani (62.2) and Chinese (63.0) populations. For Indian (86.5) and other South Asian (88.2) groups the rate ratios were closer to the reference value and the 95% CI included this. Adjustment for country of birth hardly altered the results.

Prostate cancer

Table 5 and figure 5 show large ethnic differences in prostate cancer, with DASRRs as low as 38.7 in the Pakistanis group, and considerably lower than in Indians (62.6). The Other White British group (111.8) had a higher DASRR for prostate cancer than the White Scottish reference, while the White Irish (85.4) had a lower one. The African origin population had a high DASRR (138.1) but the 95% CI's included 100. (Moving average analysis showed little variation across time periods, but the data were not released because of risks of disclosure.) Adjustment for country of birth attenuated the risk difference in Other White British, but across the other ethnic groups the risk ratios were lowered suggesting that being born in Scotland was protective.

Socio-economic factors

Appendix 4 (tables A 3 and A 4) shows the relationship between eight socio-economic variables and all cancers (all ages) by ethnic group. There was inconsistency in the relationships with no variable being consistently associated in the same direction with cancer in each ethnic group. These variables, therefore, did not meet the requirement of a confounding variable for our purposes.

Discussion

(a) Principal findings

To our knowledge this is the first reported European census-to-cancer data linkage exploring ethnic variations, though similar work has been done without the ethnicity angle in Iceland (25). Developing the method is, therefore, a key result. While disaggregating White sub-populations has been recommended, (26) examples are rare (27), even though country of birth work in England shows substantially higher all-cause, cardiovascular and cancer mortality rates in Ireland and Scotland born residents.(16) Even recent incidence studies have omitted this opportunity.(12;13) The observation that the White Scottish population, excepting breast cancer and prostate cancer in the Other White British, generally have higher rates than other ethnic groups in the same environment, further emphasises the challenge in Scotland.(1) Differences in cancer rates between many non-White and White populations have been demonstrated previously,(3) including in Scotland(2;3;9;21). Our advances here are to provide (retrospective) cohort data; to use the recommended measure of reported and not observer assigned ethnic group; to provide data by a broad range of ethnic groups including White subgroups; to examine the associations with socio-economic factors to assess validity of potential confounding

factors; to include country of birth in analyses; and provide updated data on a national scale.

The results have clinical and public health repercussions. For example, there is concern about low uptake of cancer screening services by South Asians.(28;29) Breast cancer screening services need to achieve greater ethnic equity(30), especially as breast cancer mortality seems to be converging towards the historically high rates in the UK(19) and ethnic minority women seem to be presenting with a comparatively high proportion of late stage disease. (31) However, before implementing new interventions to raise the rate of colorectal cancer screening, given the low relative rates of this cancer and that rapid convergence is not evident e.g. in Pakistani born people(9) – we might wish to review the cost-effectiveness of screening in such ethnic groups first.

(b) Strengths and limitations of the study

Retrospective cohort studies have the advantage of being low cost and fast in delivering results and, unlike case-control studies, provide incidence rates.(32) The strength of the study is the development of new methods creating a retrospective cohort; high overall linkage rates (95%); a large national population (4.65 million people); the availability of reported ethnic data on a wide range of ethnic groups; a check on whether differential emigration by ethnic groups might be creating spurious differences by analysis over time using moving averages; the exploration of the potential role of socio-economic variables and country of birth available in the Census; and the linkage of Census data to both cancer registry and community/hospital mortality data, so differences in rates do not simply reflect varying entry by ethnic group to the health system.

Audits show high completeness and quality of the SMR06 file for cancer diagnoses though such statistics by ethnic group are not available.(33;34) All deaths are certified by a doctor in Scotland and all hospitals are required to submit cancer registration data.

The validity of available indicators of socio-economic position, particularly area based ones derived from postcode and census data, is not established in multi-ethnic studies, yet they are usually used in cancer research.(13;35) Harding's study of mortality including cancer is a rare example of using other indicators.(36) We tested eight indicators and found none were consistently associated in the same direction with the outcome (cancer) and hence none were valid confounding variables suitable for across-ethnic group comparisons. The recommendation that studies of ethnic and racial variations adjust for socio-economic variables is sound but is not readily achievable as using invalid variables will generate spurious results.

Convergence of rates across generations is the predicted pattern.(2) A recent review indicated convergence was cancer site specific and occurring slower than expected in Europe.(3) We explored this using the country of birth variable in the Census and found this pattern was only evident for lung cancer. We acknowledge this may change as more cases occur in Scotland-born ethnic minority populations. In future as those born in Scotland increase in age, examining cancer by ethnic group stratified by country of birth will be important. These data break new ground in Europe, both in terms of findings and in linkage methods.(3)

The limitations of the study include the small numbers of outcomes for some non-White populations, and the consequent aggregation of some ethnic groups, though the numbers

are large compared with a recent paper (11). The result is imprecision of estimates and insufficient numbers to examine survival as others have done. (37) We had some variation in linkage rates by ethnic group (ranging from 85.1% in Other South Asian to 95.3% in White Scottish) but the potential bias is unknown. We think such bias would be small as the variation in linkage is most probably due to random causes e.g. variations in the spelling of unfamiliar names or mis-recording of date of birth in NHS databases. Similarly, there may have been differences in response rates by ethnic group in the census but the potential bias cannot be assessed for lack of data on non-responders. Inability to capture events that occur overseas outside the UK is a problem that is not easily resolved. Deaths of UK residents are reported back via several channels, including embassies and consulates, and the primary care registration systems. Such reports, however, may not give accurate cause of death. 'Salmon bias', whereby sick people return to countries of origin to die or for treatment, is potentially important but we think it unlikely in Scotland, and not a central issue for this analysis. Firstly, in contrast to cancers, we find high rates of cardiovascular disorders, including chronic ones such as heart failure, in South Asian populations. (38) A 'salmon bias' is not likely to be specific to cancer but to life threatening chronic illness. NHS Scotland provides excellent services free at the point of use so cancer patients are likely to stay not emigrate. Finally, 90% of our events are incident cases, not mortality, and the bias applies to mortality data. Denominator bias would arise from differential migration by ethnic group. If this occurred then rate ratios would alter over time. Appendix 3 table A2 shows this did not happen for most ethnic groups for all cancers.

The greatest limitation of retrospective cohort studies is inability to specify which confounding variables and risk factors are to be studied, and also to control the quality and

completeness of outcome data.(32) In our case the census gave access to a wide range of relevant exposure and potential confounding variables. The outcome data are of high quality and completeness in Scotland. The lack of cancer risk factor data in our retrospective cohort is a limitation, as in many studies of this design. We have no specific risk factor data to explore hypotheses though we are commencing a pilot project reporting in 2013 on linking risk factor data held in primary care to our data but even if successful we do not envisage having such data till about 2015. In the meantime, we have used data from national health surveys(39;40) to help interpret the cancer patterns (table 6) as discussed below.

c) Findings in relation to the Literature

i. The Scottish Context

Scotland has high cancer rates, probably reflecting historically high exposure to causal factors such as smoking, and a diet high in processed foods and low in fruit and vegetables.(1;41) These factors combine with comparatively poor socio-economic status, in ways that are not properly understood. It is of both scientific and public health significance that people of other ethnic groups in Scotland do not share White Scottish residents' propensity to cancer. This applies to both White and non-White subgroups alike, though particularly the latter. Other White British in Scotland, predominantly English, have lower rates of a range of problems (including all cancer, but not breast or prostate cancer in these results). Similar results were found for those born in England and Wales and living in Scotland, and those born in Scotland and living in England and Wales e.g. lower cancer mortality(16) and all cause mortality and cardiovascular,(42;43) and alcohol related mortality(44;45) in the England and Wales born. These differences are

probably linked to the higher socio-economic status and lower exposure to causal factors of these Other White British (predominantly English) populations compared to the White Scottish group. This is a less likely explanation for White Scottish people having (mostly) higher cancer rates than White Irish, and White Other groups. Examination of White subgroups in epidemiology is uncommon. Given the potential interest demonstrated here more work is warranted especially in the acquisition of risk factor data that are integral to the cohort analysis.

The main non-White populations of Scotland are Pakistani, Indian and Chinese. They are well established, with about half of the population born in the UK.(46;47) The main Indian, Pakistani and Chinese population migrations to Scotland occurred in the mid-1950s through to the 1970s. People from these ethnic groups born abroad have lived on average in Scotland for several decades although exact data are not available. In 2001 about half of these three ethnic populations lived in the West of Scotland in Greater Glasgow and Lanarkshire health board areas (http://www.scotpho.org.uk/downloads/ethnic_pop_by_hb.xls, accessed 26th of April 2012) comprising some of the most socio-economically deprived areas in Western Europe, known for their high mortality rates for chronic diseases, including cancer.(48)

ii. Risk factors and socioeconomic status

The socio-economic status of Indian, Pakistani and Chinese populations in Scotland is hard to assess, as on some indicators they are better e.g. housing tenure, on others they are worse e.g. employment status.(49) Overall, Indians, Pakistanis and White Scottish populations seem to be similar and Chinese slightly

worse off. South Asian populations have higher cardiovascular disease (CVD) rates(23) and higher rates of diabetes than the White Scottish population(50) and given that CVD and cancer share risk factors, and diabetes may raise cancer risk, there is no *prior* reason to expect cancer rates to be low in these populations in Scotland, especially in those born, or long-settled, in Scotland. Notwithstanding previous work elsewhere,(3) and Scotland (9) it is a surprise, therefore, to find that all cancers and some common cancers are still, decades after Matheson et al (9) and Merchant et al (22) reported, substantially less common in non-White populations, especially in South Asians. Unlike much previous research using country of birth and deaths data, where wariness about data artefacts, particularly numerator and denominator mismatch bias,(3;16) cautions against accepting large variations as correct,(5) this linked cohort analysis indicates that differences are possibly even larger then reported hitherto using proxy measures of ethnicity. (3;16;19) Reduction in the strength of the association is a typical outcome of non-differential (non-systematic) mismeasurement error so the increased variations are in line with epidemiological principles.

Using the Health Surveys for England(39;40) and the Scottish Health Survey table 6 summarises the best available data on some major cancer risk factors, as identified by Cancer Research UK (http://info.cancerresearchuk.org/healthyliving/; accessed 26th of April 2012). The Scottish population data were collected separately using very similar methods to those in the Health Survey for England, excepting the red meat question. Except for physical activity, which may be a reporting artefact, the White Scottish population has the highest, or among the highest prevalence, of all nine risk factors, with the non-White populations,

especially women, having the lowest prevalences. These patterns are in alignment with the results on all cancer (table 1 and figure 1). The Scottish Health and Lifestyle Surveys have very small numbers of people from these populations so Scottish data have not been published by ethnic group (51). While little is known about the risk factor profile of ethnic minority groups in Scotland, some data are available for Glasgow, the home to a high proportion of Scotland's Non-White population, where questionnaire based health and lifestyle surveys have been done.(52) These Glasgow data lend support to findings from the Health Survey for England in table 6 (40) e.g. smoking is uncommon in South Asian women and in Indian men but common in Pakistani men; drinking alcohol is uncommon in South Asian women and Pakistani men (mostly Muslim) but not so in Indian men; and the diet is a mix of traditional and Scottish foods with high fat content,(53) on a par with local populations. While substantial numbers of Indians are vegetarians, or occasional eaters of meat, Pakistani populations are not, with red meat (particularly lamb) being a key dietary component. (54;55)

iii. Implications for research, public health and clinical practice

More fundamental research is required to explain ethnic variations. This requires basic science cancer researchers to join forces with epidemiologists, so hypotheses can be both generated and tested in multidisciplinary research groups. In practical terms we propose that a research unit for the focused study of ethnic variations in cancer be set up. In such a research environment, for example, hypotheses for the differences in colorectal cancer risk could be systematically tested, rather than the current ad-hoc approach, where interesting observations are made but not studied in depth, a problem exemplified in the UK since at least 1984.(56) A full

discussion of biomedical hypotheses is beyond the scope of this paper but we consider in a little detail colorectal cancer, and very briefly the other three specific cancers, in relation to risk factors to illustrate the potential.

The well-known "deficit" of colorectal cancer in South Asian populations has led to interest in dietary components, especially spices such as curcumin (a component of turmeric) and capsaicin,(57) fibre and other complex carbohydrates influencing bile acid metabolism and bowel flora, as protective agents.(18;58-60) This line of reasoning assumes a protective agent in South Asian populations. An alternative, perhaps more promising line, is to assume less exposure to carcinogenic agents in the South Asian lifestyle. Meat, particularly red meat, is a postulated source of such carcinogens,(61;62) yet Pakistani populations are keen red meat consumers (see table 6). It may be that processing agents for meat are more important than the meat itself as indicated, especially, in the earlier (62) of recent systematic reviews(63) and also recently suggested for cardiovascular risk.(64) It is possible that the Pakistani diet contains less processed meat. Health Survey for England data, unfortunately, combine all red meats (table 6). Unpublished data on the diet of infants and very young children in Bradford indicates this is correct- processed meats were a common reported component in White English Bradford infants, but not in Pakistanis (data examined by RSB as co-investigator of the Born In Bradford study, communication of findings with permission from John Wright, PI of Born in Bradford Project). South Asians are also less likely to smoke heavily and smoking has been associated with both colonic and rectal cancer in the Whitehall 1 cohort study (65). In terms of well established associations for colorectal cancer (table 6) the picture is less clear – South Asians report eating

more fruit and vegetables and have a lower BMI, which are protective, but higher waist/hip ratio and central obesity, and lower physical activity which are risks. These small and inconsistent variations do not reconcile with the major differences in disease outcomes. These kind of hypotheses, which may explain the sustained low rates of colorectal cancer in some ethnic minority populations, need detailed study.

The challenges for public health include maintaining the low rates of cancer in non-White population while reducing them in White populations. This is one example where the general goal of narrowing inequalities needs careful specification of the change needed.(66) In all likelihood, given the anticipated tendency to convergence of disease risks in migrant populations,(3) cancer rates will rise in non-White ethnic minority groups in Scotland, so reducing inequalities but worsening public health. The greater challenge is to reduce inequalities by finding strategies that encourage convergence of the majority White populations to the low rates in the non-White groups. Already, however, it may be too late for breast cancer (12) but fortunately not for many other cancers.(3;19)

Given the likely lower level of reproductive risk factors (early menarche, late first child, small family, no breast feeding) and of smoking we expected substantially less breast cancer in non-White women, especially South Asians, but the rates were similar in Indians and other South Asians although still substantially lower in Pakistanis. The pattern for breast cancer in Pakistanis accords with historically relatively early marriage, children, and breast feeding – all more common than in the White Scottish. Unpublished recent data from Scotland indicate only small

ethnic differences in age at first birth, but substantially more breast feeding in all non-White groups, especially South Asians (personal observation as PI using Scottish Health and Ethnicity Linkage Study maternity data, paper in preparation). Table 6 shows non-White women were far less likely to report taking hormone replacement therapy than White groups. While non-White groups had less overweight/obesity than White Scottish women, with little difference in waist/hip ratio, they are generally more adipose, a phenomenon known to be present at a very young age, as reflected in skinfold thickness and direct measure of fat in children.(67) We found no evidence for an excess of breast cancer in African origin populations, as reported with contested data in England, (68-70) although our numbers are very small. The comparatively low uptake of breast cancer screening in ethnic minority populations requires urgent action ,(29) including in Scotland where we have corroborated the findings in England (30). Breast cancer screening leads to earlier diagnosis and reduced case fatality so increased participation may lead to convergence of incidence rates but better outcome.

Screening rates for colorectal cancer are also low in South Asians,(29) although Scottish data are not yet available by ethnic group. Since screening leads to both reduced incidence (removal of polyps and premalignant lesions) and early diagnosis we would expect even lower rates of colorectal cancer if South Asians participated equally in this service. The effectiveness / cost-effectiveness data on which colo-rectal screening is based, although solid,(71) are probably not applicable to populations with both low rates of colorectal cancer and low participation, such as Pakistanis. Modelling of cost-effectiveness may help to

decide how to proceed, especially on the urgency of implementing new interventions to raise colorectal screening rates in South Asians.(29)

Given the primary cause of lung cancer is tobacco, and tobacco smoking is relatively uncommon in South Asians, particularly women and Indians, though less so for Pakistani and Bangladeshi men (40;72) the low rates are in line with cigarette smoking patterns (table 6). We note the high prevalence of lung cancer in mixed population men but such data need corroboration. Pakistani men have the same high prevalence of current smoking as White Scottish men but the amount smoked is lower, which together with the fact that lifetime exposure to tobacco matters, probably explains their lower risk of lung cancer. The traditional taboos against smoking in South Asian women are holding, contrary to earlier expectations and predictions.(40) The same is not true of men, particularly Bangladeshi and Pakistani men,(40) and there is evidence in Glasgow that the prevalence of smoking in school leaving age South Asian boys is similar to that in White boys (73). Until recently, in England and Wales, smoking cessation services were not accessed well by ethnic minority populations, though this has changed recently. The situation in Scotland is unknown. Lung cancer in South Asian men is likely to converge towards the White Scottish population rate as implied by our analysis with country of birth as a covariate, as it has done in England and Wales.

Prostate cancer is known to vary greatly by ethnic group, with high rates in African origin (Black) populations and low rates in South Asian groups.(13;74) We corroborated these patterns, though the risk estimates for this ethnic group are

imprecise. In a recent review, the age standardised incidence rate in the Black population in the UK PROCESS study was estimated at 166 per 100,000, three times higher than in the White Population, 258 per 100,000 in the US and 304 per 100,000 in Jamaica. Our estimate of 326.6 per hundred thousand in African origin populations fits with these data. Additionally, our data suggest differences between Indians and Pakistanis and low rates in Chinese. We also noted low rates in White Irish, and high rates in Other White British. The causes of these variations are unknown though the patterns have potential to generate testable hypotheses. However, ethnic group variations in testing for prostate specific antigen, and subsequent biopsy, are likely to be a major determinant of variations in the incidence of diagnosed cancer as implied by recent studies in Scotland and Ireland. (75;76) In contrast to our findings, prostate cancer rates are comparatively high in the Republic of Ireland. This may reflect higher rates of PSA testing and greater use of biopsies there. Biological understanding of such ethnic variations is limited, with current attention focused on genetics, hormones, and fat, dietary factors including fatty acids, and vitamin D.(74) These ethnic variations provide a good model for disentangling causal hypotheses e.g. our findings do not support a major causal role for vitamin D, as the lowest rates are in populations with the lowest vitamin D levels i.e. South Asians.(77) One hypothesis currently of interest is dietary factors, such as lycopene in tomatoes being protective. Valid dietary data across ethnic groups are few(55) but tomatoes are integral to the preparation of many common meals in the South Asian cuisine. Table 6 indicates a higher level of fruit and vegetable consumption in all the non-White groups – which fits with the low risk of prostatic cancer in South Asian and Chinese men but not with the higher risk (though 95% CI includes 100) in the African origin groups.

Since Scotland has high rates of cancer we would expect that non-White Scottish ethnic groups born in Scotland would have higher rates than their parents/grandparents born abroad. Generally, adding country of birth led to modest narrowing of the risk difference, but in the age group developing cancers, relatively few non-White minority patients were born in Scotland. Unsurprisingly, the adjustment had most impact for lung cancer, as the major risk factor of smoking is socially patterned. It appears that the protection enjoyed by minority groups may be sustained for some cancers across generations, and convergence may be slower than expected as indicated from studies in Europe(3). Definitive analyses will need to wait until the Scottish born ethnic minority populations have moved into the age groups where cancers are common.

d) Conclusions

Powerful calls have been made for the collection of data by ethnic group and not by other proxies. (2;3;78)The Scottish Health and Ethnicity Linkage Study has shown how to obtain national cancer statistics by ethnic group. The same methods could be applied wherever a population census or database records ethnic group, as in England and Wales, where the large numbers will permit a finer disaggregation of ethnic groups with the potential of incorporating important covariates such as religion, country of birth and social circumstances. The advantages over solely relying on NHS databases (12;13) are a more reliable denominator and linked numerator data, longitudinal analysis of outcomes and access to relevant economic and social variables not available in NHS databases. The findings on all cancers, and specific cancers (particularly colorectal, prostate and breast), raise important questions on causation, and on public health and clinical policies. Risk

factor data are required to help explain such variations better. Ideally, these would be collected within prospective cohort studies. We also need to find ways of linking risk factor data from other sources such as primary care. In the meantime we need better and ongoing multi-ethnic cross sectional health surveys across the UK to augment the 1999 and 2004 Health Surveys for England.(39;40) The study contradicts the usual viewpoint that the health status of ethnic minorities is poor-at least for all-cancers and common cancers. The main public health lesson and challenge is for the majority population, for the "Scottish effect" in relation to cancer does not apply across Scotland's ethnic groups. Can the White Scottish population change to enjoy the low rates of cancer seen in other ethnic groups in the country? Also, can the Non-White groups avoid the high risks of cancer in Scotland across the generations? This exemplifies how the study of ethnic variations provides a public health approach with potential to benefit the entire population.

Acknowledgements, Funding and Independence

We thank the Chief Scientist's Office for a grant (CZH/4/432), NHS Health Scotland for a supplementary grant and the Equality and Diversity Information Programme of the Information Services Division (ISD) of NHS National Services Scotland for support. ISD and the General Register Office for Scotland both made 'in-house' contributions to the work. Joan Jamieson (ISD) was a co-investigator in the early stages and general adviser. Helen Brown was senior research fellow in the earlier stages of the study. Anne Houghton and Arti Nair gave secretarial help to prepare the paper and to general administration. Sarah Wild and Colin Simpson provided helpful comments on an earlier manuscript. BMJ editors and referees (Ruth Jack and Philippe Autier) provided invaluable feedback to improve this paper.

The researchers acted independently of the funding body and the study sponsor (the University of Edinburgh) at all stages of the work.

Authors' Contributions

Bhopal was the PI and lead writer, Brewster was a co-investigator and Chair of Cancer sub-group of SHELS, Bansal was the research fellow and co-ordinator of the study, and Steiner was researcher and primary analyst. All authors helped plan the study, evolve analysis plans, interpret data and critically revise successive drafts of the manuscript.

Other Contributors from the Scottish Health and Ethnicity Linkage Study investigators

Colin Fischbacher was co-principal applicant and Chairman of the Steering Committee. Chris Povey was a co-investigator and had the idea of linking the census data to the data

held by ISD and performed most of the linkage work including developing linkage methods. Jim Chalmers was a co-investigator and had the original idea for the use of one-way encryption. Ganka Mueller was a collaborator and was key in linking Census data to health data. Ms Genevieve Brin led the analysis of socioeconomic position. Kirsty MacLachlan advised throughout. These important contributions did not meet ICMJE authorship requirements.

The authorship, the authorship byline, and note of contributions has been agreed by all the investigators named.

Competing Interests: We have no competing interests.

Data sharing

The data are only available in a data safe haven with restricted access at National Records Scotland, and governed by strict ethical and other restrictions on access. Individual consent for linking these records was not sought. The analysts did not have access to identifiable data.

Box 1 Brief overview of Scottish studies on ethnic variations in cancer

- Muir reported that Harkness (1993, unpublished) examined nasopharyngeal cancers in Scotland, identifying Chinese people by name recorded on the cancer register. The age standardised rate was 0.3 per 100,000 in the entire Scottish population, and 13.7 in people with Chinese names(2).
- Black found substantial differences between Italian born residents and the Scottish population in laryngeal and stomach cancer (higher in men) and lung cancer (lower in men and women)(21).
- Merchant et al identified Indian and Pakistani men by name in the cancer registry and compared cancer rates to those of the Bombay cancer registry and the whole Scottish population(22). Oral cancer in Scottish Indians/Pakistanis was intermediate between the Bombay and Scotland rates. Similar observations were made for lung cancer in men and breast cancer in women.
- Matheson et al found cancers between 1961-1981 in South Asian adults by name search in the West of Scotland, reporting comparatively low rates of colorectal, breast and bronchial cancer, but high rates of cervical cancer.(9)

Box 2: ICD codes used in the study

Up to 31 Dec 1996 ICD9 codes were used by the Cancer Registry (needed for 10 year look-back):

Lung cancer	ICD9 162
Breast cancer	ICD9 174
Prostate cancer	ICD9 185
Colorectal cancer	ICD9 153 - 154
All cancers	ICD9 140 - 208
All cancer without non-melanona skin cancers	ICD9 140 - 172 & 174 - 208;

From 1st Jan 1997 in Cancer Registry and from 1st Jan 2000 in mortality data ICD10 codes were used:

Lung cancer	ICD10 C33 - C34
Breast cancer	ICD10 C50
Prostate cancer	ICD10 C61
Colorectal cancer	ICD10 C18 - C21
All cancers	ICD10 C00 - C97*
All cancer without non-melanona skin cancers	ICD10 C00 - C43 & C45 - C97

*C97 is multiple cancer sites – used in mortality data only

Table 1. All first cancer excluding non-melanoma skin cancer occurring between May 2001 and April 2008: directly age standardised annual rates per 100000 population per year by ethnic group and sex, and related rate ratios, and age and country of birth adjusted risk ratios (Poisson regression), with corresponding 95% CIs

Ethnic group	Events (n)	Population	Directly standardised rate		Age standardised rate ratio (as %)		Age adjusted risk ratio		Age and country of birth adjusted risk ratio	
			Rate	95% CI	Rate ratio	95% CI	Risk ratio	95% CI	Risk ratio	95% CI
MEN										
White Scottish	71094	1433584	**708.5**	*703.5, 713.4*	**100.0**	. .	**100.0**	. .	**100.0**	. .
Other White British	5848	136352	**620.4**	*605.3, 635.4*	**87.6**	*85.3, 89.8*	**88.5**	*85.0, 92.2*	**93.2**	*87.9, 98.9*
White Irish	940	18389	**671.3**	*630.4, 712.2*	**94.8**	*88.9, 100.6*	**95.6**	*91.5, 100.0*	**99.6**	*93.9, 105.6*
Other White	766	23517	**598.9**	*557.2 640.7*	**84.5**	*78.6, 90.5*	**85.1**	*80.5, 89.9*	**89.0**	*83.1, 95.4*
Any Mixed Background	65	2243	**794.8**	*617.2, 972.4*	**112.2**	*87.1, 137.3*	**106.0**	*82.0, 137.1*	**108.4**	*83.6, 140.7*
Indian	57	4522	**325.1**	*233.7, 416.4*	**45.9**	*33.0, 58.8*	**43.8**	*33.0, 58.2*	**46.1**	*35.1, 60.5*
Pakistani	103	7557	**349.0**	*266.2, 431.8*	**49.3**	*37.6, 61.0*	**53.1**	*43.7, 64.4*	**55.8**	*46.2, 67.5*
Other South Asian	36	2287	**492.1**	*313.7, 670.5*	**69.5**	*44.3, 94.6*	**63.1**	*51.1, 77.9*	**65.9**	*52.3, 82.9*
African origin	39	2308	**491.8**	*328.1, 655.6*	**69.4**	*46.3, 92.5*	**75.2**	*60.3, 93.8*	**78.3**	*63.5, 96.7*
Chinese	73	4343	**408.2**	*310.1, 506.3*	**57.6**	*43.8, 71.5*	**64.3**	*58.1, 71.1*	**67.6**	*59.4, 76.9*
Other Ethnic Group	21	2510	**333.3**	*152.4, 514.3*	**47.0**	*21.5, 72.6*	**46.1**	*30.8, 69.2*	**48.4**	*33.0, 71.0*
WOMEN										
White Scottish	76485	1643684	**664.8**	*660.2, 669.3*	**100.0**	.	**100.0**	. .	**100.0**	. .
Other White British	5855	151335	**580.0**	*565.7, 594.4*	**87.3**	*85.0, 89.5*	**88.5**	*85.8, 91.3*	**94.8**	*90.7, 99.1*
White Irish	986	21354	**590.4**	*554.4, 626.5*	**88.8**	*83.4, 94.3*	**90.4**	*86.4, 94.5*	**95.7**	*91.1, 100.6*
Other White	816	29392	**531.8**	*495.4, 568.1*	**80.0**	*74.5, 85.5*	**81.5**	*78.1, 85.1*	**86.6**	*80.2, 93.5*
Any Mixed Background	65	2826	**575.1**	*434.7, 715.5*	**86.5**	*65.4, 107.6*	**82.6**	*63.0, 108.3*	**85.3**	*66.3, 109.7*
Indian	48	4054	**274.2**	*183.1, 365.3*	**41.2**	*27.5, 54.9*	**45.3**	*36.5, 56.2*	**48.4**	*38.7, 60.6*
Pakistani	103	7430	**432.4**	*299.5, 565.3*	**65.0**	*45.0, 85.0*	**61.0**	*54.5, 68.2*	**65.1**	*57.7, 73.4*
Other South Asian	42	1889	**583.5**	*399.8, 767.1*	**87.8**	*60.1, 115.4*	**84.3**	*69.2, 102.8*	**88.7**	*71.6, 109.9*

African origin	41	2132	**516.9**	*344.3, 689.5*	**77.8**	*51.8, 103.7*	**76.9**	*58.7, 100.7*	**80.9**	*58.4, 112.0*
Chinese	102	4681	**545.5**	*421.3, 669.8*	**82.1**	*63.4, 100.8*	**84.3**	*66.9, 106.2*	**90.2**	*72.6, 112.3*
Other Ethnic Group	50	3141	**513.2**	*320.3, 706.2*	**77.2**	*48.2, 106.2*	**75.0**	*60.5, 93.1*	**80.0**	*63.1, 101.5*

Table 2. Lung cancer: directly age standardised annual rates per 100000 population per year by ethnic group and sex, and related rate ratios, and age and country of birth adjusted risk ratios (Poisson regression), with corresponding 95% CIs

Ethnic group	Events (n)	Population	Directly standardised rate		Age standardised rate ratio (as %)		Age adjusted risk ratio		Age and country of birth adjusted risk ratio	
			Rate	95% CI	Rate ratio	95% CI	Risk ratio	95% CI	Risk ratio	95% CI
MEN										
White Scottish	15155	1212648	**178.5**	*175.7, 181.3*	**100.0**	. .	**100.0**	. .	**100.0**	. .
Other White British	983	116075	**124.5**	*116.7, 132.2*	**69.7**	*65.2, 74.2*	**70.4**	*63.4, 78.2*	**84.3**	*74.5, 95.4*
White Irish	211	15453	**174.6**	*151.3, 197.9*	**97.8**	*84.6, 110.9*	**99.7**	*90.9, 109.3*	**114.8**	*99.9, 131.9*
Other White	151	17335	**141.5**	*118.5, 164.6*	**79.3**	*66.3, 92.2*	**80.7**	*74.2, 87.7*	**94.2**	*82.4, 107.6*
Any mixed background	21	1400	**311.6**	*184.1, 439.0*	**174.5**	*103.1, 245.9*	**172.3**	*100.5, 295.5*	**184.9**	*114.3, 299.3*
Pakistani	18	5353	**80.3**	*37.4, 123.2*	**45.0**	*21.0, 69.0*	**48.1**	*33.7, 68.7*	**57.7**	*37.6, 88.6*
Chinese	15	3004	**112.6**	*53.7, 171.6*	**63.1**	*30.1, 96.1*	**68.3**	*49.7, 93.9*	**81.5**	*55.3, 120.3*
WOMEN										
White Scottish	12996	1408621	**131.8**	*129.6, 134.0*	**100.0**	. .	**100.0**	. .	**100.0**	. .
Other White British	626	127254	**74.7**	*68.8, 80.5*	**56.7**	*52.1, 61.2*	**57.3**	*52.3, 62.7*	**79.9**	*72.0, 88.7*
White Irish	177	17924	**119.2**	*101.7, 136.7*	**90.4**	*77.0, 103.8*	**91.7**	*75.9, 110.7*	**120.8**	*94.5, 154.5*
Other White	99	21210	**81.0**	*64.8, 97.1*	**61.4**	*49.1, 73.7*	**62.4**	*50.1, 77.6*	**83.5**	*66.5, 105.0*
Any Mixed Background	10	1849	**115.5**	*43.5, 187.5*	**87.6**	*33.0, 142.3*	**86.3**	*62.7, 118.8*	**97.6**	*62.9, 151.4*
Pakistani	8	4963	**70.6**	*0.0, 149.6*	**53.5**	*0.0, 113.5*	**37.7**	*21.3, 66.8*	**52.7**	*29.5, 94.2*
Chinese	15	3250	**127.3**	*50.7, 203.9*	**96.6**	*38.4, 154.7*	**93.5**	*64.6, 135.3*	**130.7**	*95.7, 178.4*

* Indian, other South Asian, African origin and Other Ethnic Groups numbers were small and judged to be potentially disclosive

Table 3. Colorectal cancer: directly age standardised annual rates per 100000 population per year by ethnic group and sex, and related rate ratios, and age and country of birth adjusted risk ratios (Poisson regression), with corresponding 95% CIs

Ethnic group	Events (n)	Population	Directly standardised rate		Age standardised rate ratio (as %)		Age adjusted risk ratio		Age and country of birth adjusted risk ratio	
			Rate	95% CI	Rate ratio	95% CI	Risk ratio	95% CI	Risk ratio	95% CI
MEN										
White Scottish	11155	1212648	**131.4**	*129.0, 133.8*	**100.0**	. .	**100.0**	. .	**100.0**	. .
Other White British	863	116075	**108.3**	*101.1, 115.5*	**82.4**	*76.7, 88.1*	**83.3**	*80.8, 85.9*	**84.8**	*77.5, 92.9*
White Irish	163	15453	**136.1**	*115.4, 156.9*	**103.6**	*87.7, 119.5*	**105.0**	*97.3, 113.4*	**106.5**	*96.7, 117.3*
Other White	108	17335	**101.4**	*81.9, 121.0*	**77.2**	*62.2, 92.2*	**77.2**	*65.7, 90.7*	**78.4**	*66.4, 92.7*
Pakistani	13	5353	**43.2**	*16.8, 69.6*	**32.9**	*12.8, 53.0*	**45.6**	*28.9, 72.0*	**46.4**	*29.1, 73.9*
Chinese	10	3004	**56.0**	*17.9, 94.1*	**42.6**	*13.6, 71.6*	**59.7**	*24.6, 144.9*	**60.7**	*24.8, 148.9*
WOMEN										
White Scottish	9885	1408621	**100.3**	*98.3, 102.2*	**100.0**	. .	**100.0**	. .	**100.0**	. .
Other White British	704	127254	**83.9**	*77.7, 90.1*	**83.7**	*77.3, 90.1*	**84.4**	*78.8, 90.4*	**79.1**	*72.1, 86.8*
White Irish	155	17924	**105.5**	*88.9, 122.2*	**105.3**	*88.5, 122.0*	**105.9**	*91.0, 123.4*	**100.2**	*86.6, 116.0*
Other White	93	21210	**75.1**	*59.6, 90.6*	**74.9**	*59.4, 90.4*	**77.2**	*62.8, 95.0*	**72.9**	*61.0, 87.1*
Pakistani	6	4963	**69.0**	*0.0, 150.8*	**68.9**	*0.0, 150.4*	**37.4**	*20.3, 68.7*	**35.0**	*18.6, 65.8*
Chinese	17	3250	**111.9**	*47.5, 176.3*	**111.6**	*47.3, 175.9*	**137.4**	*85.0, 222.1*	**128.8**	*80.4, 206.4*

* Indian, other South Asian, African origin and Other Ethnic Groups numbers were small and judged to be potentially disclosive

Table 4. Breast cancer in women: directly age standardised annual rates per 100000 population per year by ethnic group, and related rate ratios, and age and country of birth adjusted risk ratios (Poisson regression), with corresponding 95% CIs

Ethnic group	Events (n)	Population	Directly standardised rate		Age standardised rate ratio (as %)		Age adjusted risk ratio		Age and country of birth adjusted risk ratio	
			Rate	95% CI	Rate ratio	95% CI	Risk ratio	95% CI	Risk ratio	95% CI
WOMEN										
White Scottish	22361	1643684	**194.3**	*191.8, 196.9*	**100.0**	. .	**100.0**	. .	100.0	. .
Other White British	2049	151335	**199.0**	*190.4, 207.6*	**102.4**	*97.8, 107.0*	**103.1**	*97.5, 108.9*	103.0	*94.0, 112.8*
White Irish	258	21354	**163.2**	*143.2, 183.2*	**84.0**	*73.6, 94.3*	**84.7**	*78.1, 91.9*	84.7	*73.6, 97.5*
Other White	277	29392	**174.4**	*153.4, 195.5*	**89.8**	*78.9, 100.6*	**90.9**	*84.2, 98.2*	90.8	*81.8, 100.9*
Any mixed background	25	2826	**208.5**	*123.1, 293.9*	**107.3**	*63.3, 151.3*	**101.4**	*68.4, 150.3*	101.3	*68.9, 148.9*
Indian	29	4054	**168.2**	*95.6, 240.7*	**86.5**	*49.2, 123.9*	**79.3**	*53.4, 117.7*	79.2	*52.4, 119.9*
Pakistani	37	7430	**120.8**	*70.7, 170.9*	**62.2**	*36.4, 88.0*	**61.5**	*51.7, 73.1*	61.4	*50.8, 74.4*
Other South Asian	13	1889	**171.4**	*70.6, 272.1*	**88.2**	*36.3, 140.0*	**78.4**	*59.0, 104.3*	78.4	*58.1, 105.8*
African origin	16	2132	**133.8**	*57.9, 209.6*	**68.8**	*29.8, 107.9*	**87.1**	*63.9, 118.7*	87.0	*56.1, 135.1*
Chinese	31	4681	**122.4**	*71.4, 173.4*	**63.0**	*36.7, 89.2*	**73.6**	*49.5, 109.5*	73.6	*49.0, 110.6*
Other Ethnic Group	26	3141	**232.3**	*102.8, 361.9*	**119.5**	*52.9, 186.2*	**103.7**	*84.2, 127.8*	103.7	*79.7, 134.9*

Table 5. Prostate cancer: directly age standardised annual rates per 100000 population per year by ethnic group, and related rate ratios, and age and country of birth adjusted risk ratios (Poisson regression), with corresponding 95% CIs

Ethnic group	Events (n)	Population	Directly standardised rate		Age standardised rate ratio (as %)		Age adjusted risk ratio		Age and country of birth adjusted risk ratio	
			Rate	95% CI	Rate ratio	95% CI	Risk ratio	95% CI	Risk ratio	95% CI
MEN										
White Scottish	15213	918632	**236.6**	*232.9, 240.3*	**100.0**	. .	**100.0**	. .	**100.0**	.
Other White British	1575	88919	**264.5**	*251.7, 277.4*	**111.8**	*106.1, 117.5*	**112.4**	*109.2, 115.7*	**100.0**	*89.9, 111.3*
White Irish	183	11977	**202.1**	*173.1, 231.2*	**85.4**	*73.1, 97.8*	**85.7**	*78.6, 93.5*	**77.9**	*66.0, 92.1*
Other White	171	11817	**218.4**	*184.9, 251.9*	**92.3**	*78.1, 106.5*	**91.8**	*84.6, 99.7*	**82.9**	*72.1, 95.2*
Any Mixed Background	12	832	**240.2**	*106.7, 373.7*	**101.5**	*45.1, 158.0*	**99.9**	*68.5, 145.5*	**95.1**	*64.0, 141.3*
Indian	16	2021	**148.1**	*70.4, 225.8*	**62.6**	*29.7, 95.5*	**63.2**	*45.4, 88.0*	**56.3**	*40.4, 78.4*
Pakistani	17	3256	**91.6**	*44.2, 139.0*	**38.7**	*18.7, 58.8*	**46.4**	*33.4, 64.6*	**41.2**	*29.3, 58.2*
African origin	13	917	**326.6**	*147.6, 505.7*	**138.1**	*62.4, 213.8*	**142.8**	*95.9, 212.7*	**130.5**	*90.4, 188.6*
Chinese	18	1956	**176.1**	*92.6, 259.5*	**74.4**	*39.1, 109.7*	**84.6**	*59.4, 120.4*	**75.4**	*52.9, 107.5*

Other South Asian and Other Ethnic group numbers were small and judged potentially disclosive

Table 6: Pattern* of smoking, alcohol, physical activity, fruit/vegetables and meat eating, hormone replacement therapy, and obesity/central obesity by 6 ethnic group from Health Surveys for England (1999 and 2004) and Scottish Health Survey (2003 and 1999)

Ethnic Group Men (16 yrs or more)	N for current smoking (varies for each variable)	% currently smoking cigarettes (16 yrs+)	% not current alcohol drinker	% meeting physical activity guidelines[+] 16+ years in HSE 16-74 yrs in Scotland	% consuming 5 or more portions of fruit/ vegetables/ day	% eats red meat 2+ times/wk[Δ] (HSE 1-6 times wk)	% eats meat products 2+ times wk Δ	% overweight (BMI >25) or obese (BMI >30)	% raised waist/hip ratio (> 0.95)	% ever used hormone replacement therapy (HSE 1999, 16 yrs +)
Scottish population (predominantly White Scottish)	3582	29	8	44	20	6	38	65	29	-
General population (predominantly White English)	2855	24	8	37	23	-	-	67	33	-
White Irish	496	30	10	39	26	79		67	36	-
Indian	547	20	33	30	37	45		53	38	-
Pakistani	423	29	89	28	33	64		55	36	-
Chinese	345	21	19	30	36	80		37	17	-
Black - African	379	21	32	35	31	-		62	16	-
Caribbean	403	25	15	37	32	68		67	25	-

Ethnic Group Women (16 yrs or more)	N for current smoking (varies for each variable)	% currently smoking cigarettes (16 yrs+)	% not current alcohol drinker	% meeting physical activity guidelines[+] 16+ Years in HSE 16-74 yrs in Scotland	% consuming 5 or more portions of fruit/ vegetables/ day	% eats red meat 2+ times/wk[Δ] (HSE 1-6 times wk)	% eats meat products 2+ times wk [Δ]	% overweight (BMI >25) or obese (BMI >30)	% raised waist/hip ratio (> 0.95)	% ever used hormone replacement therapy (HSE 1999, 16 yrs +)
Scottish population (predominantly White Scottish)	4514	28	13	33	22	56	21	60	37	17**
General population (predominantly White English)	3805	23	14	25	27	-	-	57	30	18
White Irish	653	26	11	29	32	67	-	58	37	19
Indian	547	5	59	23	36	34	-	55	30	7
Pakistani	423	5	95	14	32	62	-	62	39	5
Chinese	345	8	33	17	42	72	-	25	32	8
Black - African	379	10	45	29	32	-	-	70	32	-
Caribbean	403	24	21	31	31	61	-	65	37	8

* Comparative data for predominantly White (general) population are from the 2003 HSE
+ 30 minutes or more moderate to vigorous activity on 5 days/week or more
Δ These data are from HSE 1999 as they were not published in HSE 2004. HSE equivalent question is 1-6 times/wk
** 1998 Scottish Health Survey 25-74 yrs, not 16 yrs+ as in HSE

Reference List

(1) Scottish Executive. White Paper: Towards a Healthier Scotland.Edinburgh: The Stationery Office; 1999.

(2) Muir CS. Epidemiology of cancer in ethnic groups. Br J Cancer 1996 Sep;29:S12-S16.

(3) Arnold M, Razum O, Coebergh JW. Cancer risk diversity in non-western migrants to Europe: An overview of the literature. Eur J Cancer 2010 Sep;46(14):2647-59.

(4) Razum O, Spallek J, Reeske A, et al. Migration-sensitive Cancer Registration in Europe. Challenges and Potentials. Frankfurt: Peter Lang International Academic Publishers; 2011.

(5) Gill PS, Kai J, Bhopal R.S, et al. Health Care Needs Assessment: Black and Minority Ethnic Groups. In: Stevens A, et al, editors. Health Care Needs Assessment. The epidemiologically based needs assessment reviews. Third Series ed. Abingdon: Radcliffe Medical Press Ltd; 2007. p. 227-389.

(6) Hislop TG, Bajdik CD, Saroa SR, et al. Cancer incidence in Indians from three areas: Delhi and Mumbai, India, and British Columbia, Canada. J Immigr Minor Health 2007 Jul;9(3):221-7.

(7) Winter H, Cheng K, Cummins C, et al. Cancer incidence in the south Asian population of England (1990-1992). British Journal of Cancer 1999;79(3/4):645-54.

(8) Mangtani P, Maringe C, Rachet B, et al. Cancer mortality in ethnic South Asian migrants in England and Wales (1993-2003): patterns in the overall population and in first and subsequent generations. Br J Cancer 2010 Apr 27;102(9):1438-43.

(9) Matheson LM, Donnigan MG, Hole D, et al. Incidence of colorectal, breast and lung cancer in a Scottish Asian population. Health Bulletin (Scotland) 1985;43(5):245-9.

(10) Cancer Research UK. Equality in Cancer Prevention. London: Cancer Research UK; 2004. Report No.: 1.

(11) Lane DA, Lip GY, Beevers DG. Ethnic differences in cancer incidence and mortality: the Birmingham Factory Screening Project. QJM 2007 Jul;100(7):423-31.

(12) Jack RH, Davies EA, Moller H. Breast cancer incidence, stage, treatment and survival in ethnic groups in South East England. Br J Cancer 2009 Feb 10;100(3):545-50.

(13) Jack RH, Davies EA, Moller H. Prostate cancer incidence, stage at diagnosis, treatment and survival in ethnic groups in South-East England. BJU Int 2010 May;105(9):1226-30.

(14) National Cancer Intelligence Network. Cancer Incidence and Survival by Major Ethnic Group, England, 2002 - 2006. London: Cancer Research UK; 2009.

(15) Downing A, West RM, Gilthorpe MS, et al. Using routinely collected health data to investigate the association between ethnicity and breast cancer incidence and survival: what is the impact of missing data and multiple ethnicities? Ethn Health 2011 Jun;16(3):201-12.

(16) Wild SH, Fischbacher CM, Brock A, et al. 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 Apr 10;94:1079-85.

(17) Jeffreys M, Stevanovic V, Tobias M, et al. Ethnic inequalties in cancer survival in New Zealand: Linkage study. Research and Practice 2005;95:834-7.

(18) Rastogi T, Devesa S, Mangtani P, et al. Cancer incidence rates among South Asians in four geographic regions: India, Singapore, UK and US. Int J Epidemiol 2008 Feb;37(1):147-60.

(19) Harding S, Rosato M, Teyhan A. Trends in cancer mortality among migrants in England and Wales, 1979-2003. European Journal of Cancer 2009 Aug;45(12):2168-79.

(20) Bhopal R, Fischbacher C, Povey C, et al. Cohort profile: Scottish Health and Ethnicity Linkage Study of 4.65 million people exploring ethnic variations in disease in Scotland. Int J Epidemiol 2010 Jul 24.

(21) Black RJ. Cancer in Italian migrant populations. Scotland. IARC Sci Publ 1993;(123):186-92.

(22) Merchant NE, Ferguson MM, Ali A, et al. Oral carcinoma in the Indian and Pakistani community in Scotland. J Oral Med 1986 Jan;41(1):62-5.

(23) 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 Jul 5;7(1):142.

(24) Boyd KM. Ethnicity and the ethics of data linkage. BMC Public Health 2007 Nov 8;7(1):318.

(25) Vidarsdottir H, Gunnarsdottir HK, Olafsdottir EJ, et al. Cancer risk by education in Iceland; a census-based cohort study. Acta Oncol 2008 Jan 1;47(3):385-90.

(26) Bhopal R, Donaldson L. White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health. Am J Public Health 1998 Sep;88(9):1303-7.

(27) Williams R, Ecob R. Regional mortality and the Irish in Britain: findings from the ONS Longitudinal Study. Sociology of Health & Illness 1999;21(3):344-67.

(28) Price CL, Szczepura AK, Gumber AK, et al. Comparison of breast and bowel cancer screening uptake patterns in a common cohort of South Asian women in England. BMC Health Serv Res 2010;10:103.

(29) Szczepura AK, Price CL, Gumber AK. Breast and Bowel Cancer Screening Uptake Patterns over 15 Years for UK South Asian Ethnic Minority Populations, Corrected for Differences in Socio-demographic Characteristics. BMC Public Health 2008 Oct 2;8(1):346.

(30) Bansal N, Bhopal RS, Steiner MF, et al. Major ethnic group differences in breast cancer screening uptake in Scotland are not extinguished by adjustment for indices of geographical residence, area deprivation, long-term illness and education. Br J Cancer 2012 Mar 13;106(8):1361-6.

(31) Cuthbertson SA, Goyder EC, Poole J. Inequalities in breast cancer stage at diagnosis in the Trent region, and implications for the NHS Breast Screening Programme. J Public Health (Oxf) 2009 Sep;31(3):398-405.

(32) Szklo M, Nieto FJ. Epidemiology: Beyond the basics. 2 ed. Maryland: Aspen Publishers Inc; 2007.

(33) Brewster DH, Stockton D, Harvey J, et al. Reliability of cancer registration data in Scotland, 1997. Eur J Cancer 2002 Feb;38(3):414-7.

(34) Brewster DH, Stockton DL. Ascertainment of breast cancer by the Scottish Cancer Registry: an assessment based on comparison with five independent breast cancer trials databases. Breast 2008 Feb;17(1):104-6.

(35) Ali R, Barnes I, Kan SW, et al. Cancer incidence in British Indians and British whites in Leicester, 2001-2006. Br J Cancer 2010 Jun 29;103(1):143-8.

(36) Harding S. Mortality of migrants from the Caribbean to England and Wales: effect of duration of residence. Internation journal of epidemiology 2003;33:382-6.

(37) Tehranifar P, Neugut AI, Phelan JC, et al. Medical Advances and Racial/Ethnic Disparities in Cancer Survival. Cancer Epidemiol Biomarkers Prev 2009 Oct 1;18(10):2701-8.

(38) Bhopal RS, Bansal N, Fischbacher CM, et al. Ethnic variations in heart failure: Scottish Health and Ethnicity Linkage Study (SHELS). Heart 2012 Jan 27.

(39) The Department of Health. Health Survey for England 1999: the health of minority ethnic groups. Vol 1 and Vol 2 ed. London: The Stationary Office; 1999.

(40) Becker E, Boreham R, Chaudhury M, et al. Health Survey for England 2004 - The health of minority ethnic groups. London: National Centre for Social Research; 2006.

(41) Lawder R, Harding O, Stockton D, et al. Is the Scottish population living dangerously? Prevalence of multiple risk factors: the Scottish Health Survey 2003. BMC Public Health 2010;10:330.

(42) Wild SH, Fischbacher C, Brock A, et al. Mortality from all causes and circulatory disease by country of birth in England and Wales 2001-2003. J Public Health (Oxf) 2007 Jun;29(2):191-8.

(43) Fischbacher CM, Steiner M, Bhopal R, et al. Variations in all cause and cardiovascular mortality by country of birth in Scotland 1997-2003. Scottish Medical Journal 2007;52(4):5-10.

(44) Bhala N, Fischbacher C, Bhopal R. Mortality for Alcohol-related Harm by Country of Birth in Scotland, 2000-2004: Potential Lessons for Prevention. Alcohol and Alcoholism 2010;45(6):552-6.

(45) Bhala N, Wild S, Brock A, et al. Mortality from liver cirrhosis and hepatocellular cancer by country of birth in England and Wales. Journal of Epidemiology & Community Health 2007;61(Suppl 1):A37-A38.

(46) 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. http://www chs med ed ac uk/phs/research/Retrocoding%20final%20report pdf 2006 August 7

(47) Scottish Ethnicity and Health Research Strategy Working Group. Health in our Multi-ethnic Scotland: Future Research Priorities. 1 ed. Edinburgh: NHS Health Scotland; 2009.

(48) Walsh D, Bendel N, Jones R, et al. It's not 'just deprivation': why do equally deprived UK cities experience different health outcomes? Public Health 2010 Sep;124(9):487-95.

(49) Scottish Executive. Analysis Of Ethnicity In The 2001 Census.Edinburgh: Scottish Executive; 2007. p. 1-55.

(50) Fischbacher CM, Bhopal R, Steiner M, et al. Is there equity of service delivery and intermediate outcomes in South Asians with type 2 diabetes? Analysis of DARTS database and summary of UK publications. J Public Health 2009 Jun 1;31(2):239-49.

(51) Scottish Executive. The Scottish health survey 2003. Edinburgh: Scottish Executive; 2005.

(52) Heim D, MacAskill S. Black and Minority Ethnic Health in Greater Glasgow: A Comparative Report on the Health and Well-Being of African & Caribbean, Chinese, Indian and Pakistani People and the General Population. Glasgow: NHS Greater Glasgow; 2006.

(53) Anderson AS, Bush H, Lean M, et al. Evolution of atherogenic diets in South Asian and Italian women after migration to a higher risk region. Journal of Human Nutrition and Dietetics 2005;18(1):33-43.

(54) Vyas A, Greenhalgh A, Cade J, et alet al. Nutrient intakes of an adult Pakistani, European and African-Caribbean community in inner city Britain. J Hum Nutr Diet 2003 Oct;16(5):327-37.

(55) Leung G, Stanner S. Diets of minority ethnic groups in the UK: influence on chronic disease risk and implications for prevention. Nutrition Bulletin 2011;36(2):161-98.

(56) Marmot MG, Adelstein AM, Bulusu L. Immigrant mortality in England and Wales 1970 -78. Causes of death by country of birth. London: HMSO; 1984.

(57) Aggarwal BB, Kunnumakkara AB, Harikumar KB, et al. Potential of Spice-Derived Phytochemicals for Cancer Prevention. Planta Med 2008;74(EFirst):1560,1569.

(58) McKeigue PM, Adelstein AM, Marmot MG, et al. Diet and fecal steroid profile in a South Asian population with a low colon-cancer rate. Am J Clin Nutr 1989;50(1):151-4.

(59) Reddy S, Sanders TA, Owen RW, et al. Faecal pH, bile acid and sterol concentrations in premenopausal Indian and white vegetarians compared with white omnivores. Br J Nutr 1998 Jun;79(6):495-500.

(60) Haines A, Hill MJ, Thompson MH, et al. A prospective study of faecal bile acids and colorectal cancer. Eur J Cancer Prev 2000 Oct;9(5):317-23.

(61) Larsson SC, Wolk A. Meat consumption and risk of colorectal cancer: A meta-analysis of prospective studies. Int J Cancer 2006;119(11):2657-64.

(62) Sandhu MS, White IR, McPherson K. Systematic Review of the Prospective Cohort Studies on Meat Consumption and Colorectal Cancer Risk. Cancer Epidemiology Biomarkers & Prevention 2001 May 1;10(5):439-46.

(63) Larsson SC, Wolk A. Meat consumption and risk of colorectal cancer: A meta-analysis of prospective studies. Int J Cancer 2006;119(11):2657-64.

(64) Micha R, Wallace SK, Mozaffarian D. Red and Processed Meat Consumption and Risk of Incident Coronary Heart Disease, Stroke, and Diabetes Mellitus: A Systematic Review and Meta-Analysis. Circulation 2010 Jun 1;121(21):2271-83.

(65) Morrison DS, Batty GD, Kivimaki M, et al. Risk factors for colonic and rectal cancer mortality: evidence from 40 years follow-up in the Whitehall I study. J Epidemiol Community Health 2011 Nov 1;65(11):1053-8.

(66) Commission of the European Communities. Solidarity in Health: Reducing Health Inequalities in the EU.Brussels: Commission of the European Communities; 2009. p. 1-11.

(67) Nightingale CM, Rudnicka AR, Owen CG, et al. Patterns of body size and adiposity among UK children of South Asian, black African–Caribbean and white European origin: Child Heart And health Study in England (CHASE Study). International Journal of Epidemiology 2010.

(68) Bowen RL, Duffy SW, Ryan DA, et al. Early onset of breast cancer in a group of British black women. Br J Cancer 2008 Jan 29;98(2):277-81.

(69) Cichowska A, Fischbacher CM, Brock A, et al. Early onset of breast cancer in British Black women. Br J Cancer 2008 Jun 17;98(12):2011.

(70) Ingleby JD. Early onset of breast cancer in Black British women: how reliable are the findings? Br J Cancer 2008 Sep 16;99(6):986-7.

(71) Lansdorp-Vogelaar I, Knudsen AB, Brenner H. Cost-effectiveness of Colorectal Cancer Screening. Epidemiol Rev 2011 Jul 1;33(1):88-100.

(72) Jack RH, Davies EA, Moller H. Lung cancer incidence and survival in different ethnic groups in South East England. Br J Cancer 2011 Sep 27;105(7):1049-53.

(73) Kohli HS. A comparison of smoking and drinking among Asian and white schoolchildren in Glasgow. Public Health 1989 Nov;103(6):433-9.

(74) Kheirandish P, Chinegwundoh F. Ethnic differences in prostate cancer. Br J Cancer 2011 Aug 9;105(4):481-5.

(75) Brewster DH, Fraser LA, Harris V, et al. Rising incidence of prostate cancer in Scotland: increased risk or increased detection? BJU Int 2000 Mar;85(4):463-72.

(76) Carsin AE, Drummond FJ, Black A, et al. Impact of PSA testing and prostatic biopsy on cancer incidence and mortality: comparative study between the Republic of Ireland and Northern Ireland. Cancer Causes Control 2010 Sep;21(9):1523-31.

(77) Smith M. Seasonal, ethnic and gender variations in serum vitamin D3 levels in the local population of Peterborough. Bioscience Horizons 2010 May 3;10.1093/biohorizons/hzq016.

(78) Palmieri C, Coombes RC, Kim SB, et al. Ethnicity and breast cancer research. Lancet 2008 Jul 19;372(9634):188-9.

Appendix 1: Details on linkage methods (text lightly edited from open access publication(23))

Appendix figure 1, republished from our open access publication (23) illustrates in concept how record linkage was based on information from three datasets: health care records, which include personal identifiers and clinical information; the Community Health Index (CHI) which contains personal identifiers and the CHI number; and the census file which contains personal identifiers and details of individuals' ethnicity. The 14 ethnic groups are given in appendix table A1. The CHI dataset lists in Scotland 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 estimated to be 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 data sets, 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". 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 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, the encrypted CHI number and a newly generated index number unrelated to other numbers for the exclusive use of this project. 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), resulting in depersonalised 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%(23).

Appendix table A1. Linkage Rates by Ethnic group

		Number	Percentage
1	White Scottish	4290153	95.3
2	Other White British	357788	93.6
3	White Irish	47173	92.2
4	Other White	74655	87.9
5	Any Mixed Background	12117	91.7
6	Indian	13717	89.9
7	Pakistani	28538	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	15115	87.4
14	Other Ethnic Group	8945	86.2

Appendix 2: Methods for calculating rates, ratio and risk ratios

To calculate directly age standardised rates (DASR's) we used the cohort denominator at April 2001, and for the numerator the first event cancers for 7-years thereafter. We divided the result by 7 to get an annual rate. We had no information on emigration to recalculate denominators over time. Non-cancer outcomes were not available because of concerns over disclosure (see ethics and disclosure). We did not adjust the denominator to remove 50% of the people who developed cancer because the outcome is rare. For example, for all cancer in the White Scottish population the adjusted denominator would be 1433584-(0.5 x 71094) = 1398037, which is 97.5%. (It is standard practice to remove half of the numerator from the denominator when readjusting denominators in these circumstances (32). The recalculated directly standardised rate is 726.5 compared with our reported figure of 708.5, a 2.5% difference, for our commonest outcome. The difference would be much smaller for the specific cancers.. The effect on rate ratio and risk ratios would be very small, and less than this. Our approach has the merit of simplicity and is standard in descriptive epidemiology for rare outcomes(32) and has been adopted across SHELS analyses. The approach here – modelling cumulative incidence (risks) rather than person-time incidence- is appropriate when the numbers no longer at risk at the end of the observation period is not high (as here), when the period of observation is not highly variable (as here) and when the main comparisons are with a general population (as here). Szklo and Nieto's established textbook notes that the cumulative incidence approach we have used leads to a lower absolute value for the incidence than with a person-time rate but when events are rare (as here) the discrepancy is small(32).

We constructed Poisson models with age only and then included variables where we had a specific hypothesis; so there was no unspecified exploration (fishing), and no modelling with forward or backward selection to include as many significant cofactors as possible.

With robust variance we mean the empirical (robust) estimator of the covariance matrix. It has the property of being a consistent estimator of the covariance matrix, even if the working correlation matrix is misspecified. Some relevant papers are:

Zeger, S. L., Liang, K. Y., and Albert, P. S. (1988), "Models for Longitudinal Data: A Generalized Estimating Equation Approach," Biometrics, 44, 1049-1060.

Royall, R. M. (1986), "Model Robust Inference Using Maximum Likelihood Estimators," International Statistical Review, 54, 221-226.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," Econometrica, 50, 1-25

We used SAS for our statistical analysis and the user documentation advises that if you include the statement "REPEATED SUBJECT = ... /TYPE = unstr;" that empirical (or robust) estimators are produced, even if you have only one observation per subject. The subject identifier needs also be put in the CLASS statement. We can supply the full computer code to interested readers.

Appendix 3. Table A2. Number of cases and age-adjusted risk ratios (RR) for 5 overlapping time periods for all cancers (except non-melanoma skin cancer) by ethnic group

	2001-2004		2002-2005		2003-2006		2004-2007		2005-2008	
	N	RR	N	RR	N	RR	N	RR	N	RR
Men										
White Scottish	29719	100	29784	100	29358	100	29510	100	29392	100
Other White British	2413	87.6	2432	88.1	2425	89	2427	88.4	2432	88.8
White Irish	383	92.1	404	97.6	385	95	403	99.6	395	98.7
Other White	344	89.8	327	86.3	318	86.1	309	84	288	79.2
Any Mixed Background	24	95	26	102.1	29	114.6	27	105.1	27	104.4
Indian	18	34.7	22	41.3	27	50.1	30	53.9	28	49.1
Pakistani	40	52.2	40	50.5	36	44.7	39	46.7	47	54.5
Other South Asian	13	58.3	15	64.9	17	71.9	16	64.9	16	62.5
African Origin	19	92	19	90	18	84.1	13	58.8	12	52.7
Chinese	27	60.3	27	58.3	30	63.7	31	63.4	36	71.3
Other Ethnic Group	7	40.7	*	*	6	31.9	7	35.1	14	66.3
Women										
White Scottish	31535	100	31650	100	31854	100	32080	100	31926	100
Other White British	2453	90.5	2527	92.6	2456	89.2	2465	88.6	2395	86.1
White Irish	403	88.7	399	87.9	411	90.5	413	90.9	426	94.8
Other White	320	79.1	319	77.6	330	79	335	78.6	362	84.3
Any Mixed Background	28	89.2	32	99.7	23	70.1	29	85.9	24	70
Indian	14	34.3	18	42.3	22	49.7	21	45.3	24	50.2
Pakistani	38	59.4	36	53.5	46	65.1	50	66.9	48	61.7
Other South Asian	21	107	20	99.1	16	77.1	13	60.5	15	68.2
African origin	22	106.3	22	102.4	16	71.9	17	73.1	11	45.9
Chinese	40	85.6	45	92.5	48	95	48	90.7	43	78.8
Other Ethnic Group	18	71.4	18	67.9	22	79.1	24	81.3	25	81.2

*Potentially disclosive so suppressed.

For text interpreting these results see: Results. All cancers without non-melanoma skin cancer and discussion (strengths and weaknesses)

Table A3:Age-adjusted relative increase (%) in risk of cancer [and 95% confidence interval] for each category increase in the variable for census derived socio-economic variables

a- Men

Ethnic group	SIMD	Highest qualification (individual)	Highest qualification (household)	NS-SeC (individual)	NS-SeC (household)	Car Ownership	Household Tenure	Activity last wee
	quantitative	*quantitative*	*quantitative*	*quantitative*	*quantitative*	*1+ vs. 0*	*Owned vs. Rented*	*Working vs. Inac*
White Scottish	5.5 [4;7]*	9.9 [8;11.7]	9.9 [7.5;12.2]	7.9 [5.9;9.8]	7.8 [5.5;10.1]	15.5 [8.7;21.9]	11.8 [5.3;17.9]	25.9 [2
Other White British	4.5 [2.8;6.3]*	5.5 [3.4;7.7]	7.2 [5.3;9]	5.4 [3;7.8]*	6.8 [3.5;10.2]*	8 [2.8;13.1]	6.5 [1.9;11]	16.2 [1
White Irish	4.4 [0.5;8]*	5 [0;9.9]	6.5 [3.4;9.4]	9.8 [1.8;18.4]*	5.5 [-1.3;12.9]	12.7 [4.4;20.4]	10.1 [3.5;16.1]	24.8 [14.
Other White	4.4 [0.4;8.2]*	7.5 [-0.7;15.4]	6.7 [-5.5;17.5]	14.6 [7.5;22.1]	7.3 [-4.3;20.4]	4.2 [-11.1;17.4]	-7 [-10.9;-3.3]	20.8 [
Any mixed background	12 [-3.8;25.4]	-15 [-29.4;-2.2]	-9.2 [-42.5;16.3]	6.7 [-26.3;54.5]	35.8 [3.5;78.1]	28.9 [17.1;39]	8.5 [-29.1;35.3]	23.7 [-54.6
Indian	2.5 [-8.8;12.7]	6 [-20.2;26.7]	19.6 [0;35.3]	2.7 [-24.5;39.8]	0.9 [-24;34]	-100.9 [-328.9;5.9]	33.4 [9.3;51.1]	56.6 [36.9
Pakistani	4.7 [-8.8;16.6]	8.5 [-14.1;26.7]	4.5 [-11.4;18.3]	22.4 [-14.8;75.8]	4.4 [-13.8;26.5]	7.4 [-66.9;48.6]	-17.8 [-43.1;3]	26.9 [10.5
Other South Asian	0 [-13.5;12.1]	8.5 [-13.4;26.2]	24.3 [6.9;38.4]	29.4 [-7;80]	11.3 [-19.8;54.4]	19.6 [-26.1;48.7]	-5 [-70.2;35.1]	47.2 [-37.8
Black	11.5 [-13.5;31]	-2 [-57.3;33.7]	-4.3 [-48.6;21.4]	2 [-18.6;28]	-10.1 [-35.3;24.9]	14.7 [-45.6;50]	6.3 [-39.3;37]	25 [-19.4
Chinese	-2.2 [-14.2;8.5]	-17.8 [-56.6;11.5]	2.2 [-19.1;19.7]	-7.3 [-38.1;38.9]	-5.4 [-24.9;19]	35.2 [22.2;46]	34.4 [16.1;48.7]	22.9 [-30
All Other Ethnic group	12 [-21.5;36.3]	-3.3 [-37.6;22.5]	-12.8 [-130.6;44.8]	-48.2 [-77.5;19.2]	-39.9 [-68.9;16.3]	-5 [-171.1;59.3]	-105.8 [-531.6;33]	-46.9 [-292.6

* trend of increase across categories shows a significant departure from linearity
† figures are for each quintile increase in SIMD
§ figures are for each category increase in highest qualification - ie from none to low and low to high

b- Women

Ethnic group	SIMD	Highest qualification (individual)	Highest qualification (household)	NS-SeC (individual)	NS-SeC (household)	Car Ownership	Household Tenure	Activity last we
	quantitative	*quantitative*	*quantitative*	*quantitative*	*quantitative*	*1+ vs. 0*	*Owned vs. Rented*	*Working vs. In*
White Scottish	4.7 [3.9;5.4]	6.3 [4.5;8]*	6.7 [5.2;8.2]*	3 [0.5;5.9]*	3 [0.9;5.4]	12.9 [9.5;16]	8.9 [3.4;14.1]	14.6 [10.
Other White British	3.5 [1.2;5.8]*	4.5 [2.5;6.7]	7.2 [5;9.3]	0.2 [-2;2.8]	4 [0.7;7.5]*	15.5 [9;21.5]	0.9 [-8.8;9.7]	7 [2.
White Irish	0.4 [-3;3.8]	-1 [-10.9;8]	-0.5 [-10.2;8.4]	-14.6 [-24.1;-4]*	-4.5 [-11.6;3]	7.3 [0.5;13.5]	8.4 [-1.2;17.1]	-0.4 [-17.
Other White	4 [0.7;7.4]*	7.4 [4;10.7]	8.4 [0;16.2]	0.9 [-7.4;10]	10.7 [2.5;19.5]	10 [-1.4;20.1]	-22.5 [-40.3;-7]	4.5 [-19.
Any mixed background	10.8 [-2.9;22.6]	6.8 [-20;27.6]	6.2 [-21;27.2]	23.5 [-11.4;72.3]	11.7 [-18.2;52.6]	16.6 [-17.4;40.7]	19.6 [-20.6;46.4]	-7.5 [-37.
Indian	-11.6 [-39.6;10.9]	-20 [-33.4;-7.9]	-18.2 [-42.2;1.7]	-2.5 [-17.6;15]	7.9 [-22.1;49.5]	-13.5 [-96.3;34.4]	-20.4 [-94.9;25.6]	-44.1 [-9
Pakistani	-3.3 [-15.9;7.9]	6.5 [-41.3;38.3]	6.8 [-10.2;21.1]	-8.7 [-24.5;10.4]	0 [-15.8;18.7]	23.5 [3.5;39.4]	-10.1 [-97.8;38.7]	-21.4 [-15
Other South Asian	-14.1 [-34.5;3.2]	-9 [-44.9;18.1]	-5.5 [-43.2;22.1]	-5.5 [-32.8;32.8]	-21.8 [-45.1;11.4]	-65.6 [-189.9;5.4]	-44.8 [-107.9;-0.7]	24.6 [-126.
Black	5 [-25.6;28.1]	-14 [-44.5;10]	0.2 [-38.1;27.9]	4.7 [-26.9;49.9]	-13.3 [-34.4;14.6]	-39.5 [-166.4;26.9]	-21.3 [-105.6;28.4]	-21.2 [-89.
Chinese	8.2 [-2.5;17.8]	-1.8 [-30.5;20.6]	-16.2 [-34.6;-0.2]	-12.8 [-27.5;4.7]	-3.9 [-19.3;14.5]	-17.4 [-51.9;9.3]	1.5 [-39;30.2]	-38.2 [-68.5
All Other Ethnic group	7 [-4.5;17.4]	0.2 [-39.4;28.7]	-73.2 [-147.5;-21.2]	2.9 [-22.4;36.5]	-0.2 [-21.2;26.3]	3.3 [-62.6;42.5]	-40.9 [-96.5;-1]	34.8 [7.

¶ figures are for each category change in NS-SEC grouping, from N (never worked) to M (managerial and professional groups)
** figures indicate difference in incidence between those who do not own cars and those who do
†† figures indicate difference in incidence between those who rent and those who own their house
§§ figures indicate difference in cancer between those who were inactive and those working last week

**The analyses were on all cancers at all ages, so the no. of cases differs slightly from table 1. Data Disclosure Committee ruled that publication of numerators here was not permissible as it would be potentially disclosive.

.

Appendix 4: Assessing the potential to adjust for putative confounding variables

The data in table A3(a) for men and A2(b) for women show that none of the 8 variables was consistently associated with cancer i.e. in the same direction of association. Mostly the variables were associated as expected (though not always with linear effects) in the White groups but less so in the non-White ethnic groups. For example, in men SIMD (Scottish Index of Multiple Deprivation) the association varied widely across ethnic groups, from a decrease in cancer with increase in deprivation (-2.2%) to an increase in most groups e.g. 5.5% in White Scottish. In addition, SIMD did not show a linear increase in cancer with each category change in score (indicated by asterisk).

Appendix Table A4. Summary of performance of the 8 measures of socio-economic position as potential confounding factors

Table A4 shows that for no variable was the direction of association the same in all ethnic groups. SIMD was closest (10/11 times in men, 8/11 in women). However, our prior agreed definition for a valid confounding variable for the purposes of our analysis was that the direction of association should be the same in all ethnic groups. The alternative would have been to exclude some populations from adjustment for confounders. However, there are two good reasons for not doing this 1) it would be against the general approach of examining across groups and would go counter to our prior analysis strategy 2) the scientific literature generally shows that area based measures are not consistent confounders across ethnic groups. We concluded, therefore, that adjusting using these variables would be open to criticism.

Ethnic group		SIMD	Highest qualification (individual)	Highest qualification (household)	NS-SeC (individual)	NS-SeC (household)	Car Ownership	Household Tenure	Activity last week
Range of relative increase (%) across ethnic groups	Men	-2.2, 12	-17.8, 9.9	-12.8, 19.6	-8.2, 29.4	-39.9, 35.8	-100.9, 35.2	-100.5, 34.4	-38.2, 34.8
	Women	-11.6, 10.8	-20, 7.4	-73.2, 8.4	-14.6, 23.5	21.8, 11.7	-65.6, 23.5	-44.8, 19.6	-46.9, 56.6
No. of time association was in expected direction	Men	10	8	8	9	8	9	7	5
	Women	8	6	6	5	6	7	5	6
	Overall	18	14	14	14	14	16	12	11
No. of times association was non-linear	Men	4	0	0	2	1	NA	NA	NA
	Women	2	1	1	2	1	NA	NA	NA
	Overall	6	1	1	4	2	NA	NA	NA

*Out of a possible 11 for each sex, and 22 times in total.

A STROBE Statement—checklist of items that should be included in reports of observational studies

Authors' response is given in italics.

	Item No	Recommendation
Title and abstract	1	Indicate the study's design with a commonly used term in the title or the abstract
		(*b*) Provide in the abstract an informative and balanced summary of what was **done** and what was found *Both done*
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses *Both done*
Methods		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection *both done*
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Done*
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *not appropriate*
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. *Effect modifiers not appropriate, all other variables defined* Give diagnostic criteria, if applicable *ICD codes given*
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group *Insofar as necessary, done*
Bias	9	Describe any efforts to address potential sources of bias *Done*
Study size	10	Explain how the study size was arrived at *explained-maximum number of cases used*
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why

		Done wherever applicable
Statistical methods	12	Describe all statistical methods, including those used to control for confounding *done*
		(*b*) Describe any methods used to examine subgroups and interactions *Done wherever applicable, but interactions not examined for lack of study power*
		(*c*) Explain how missing data were addressed *Potential biases discussed*
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Not applicable for this retrospective cohort study but issues discussed*
		(*e*) Describe any sensitivity analyses *Not applicable*

Continued on next page

Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed *Reported*
		(b) Give reasons for non-participation at each stage *reported*
		(c) Consider use of a flow diagram *given simplicity of the design, not necessary; diagrams on linkage methods already published and given in appendix 1*
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders *Insofar as necessary, this has been given; but such data were extensively published in census publications*
		(b) Indicate number of participants with missing data for each variable of interest *Done*
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount) *Analysis does not use person time denominators so this is not necessary*
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time *Done*
Main results	16	*(a)* Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included *done*
		(*b*) Report category boundaries when continuous variables were categorized *not applicable*
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period *Done*
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses *not applicable*
Discussion		
Key results	18	Summarise key results with reference to study objectives *done*
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias *Done*

Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence *Done*
Generalisability	21	Discuss the generalisability (external validity) of the study results *done through discussion of the international literature*
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based *Done*

Figure 1. Any cancer: age standardised rate ratio by ethnic group

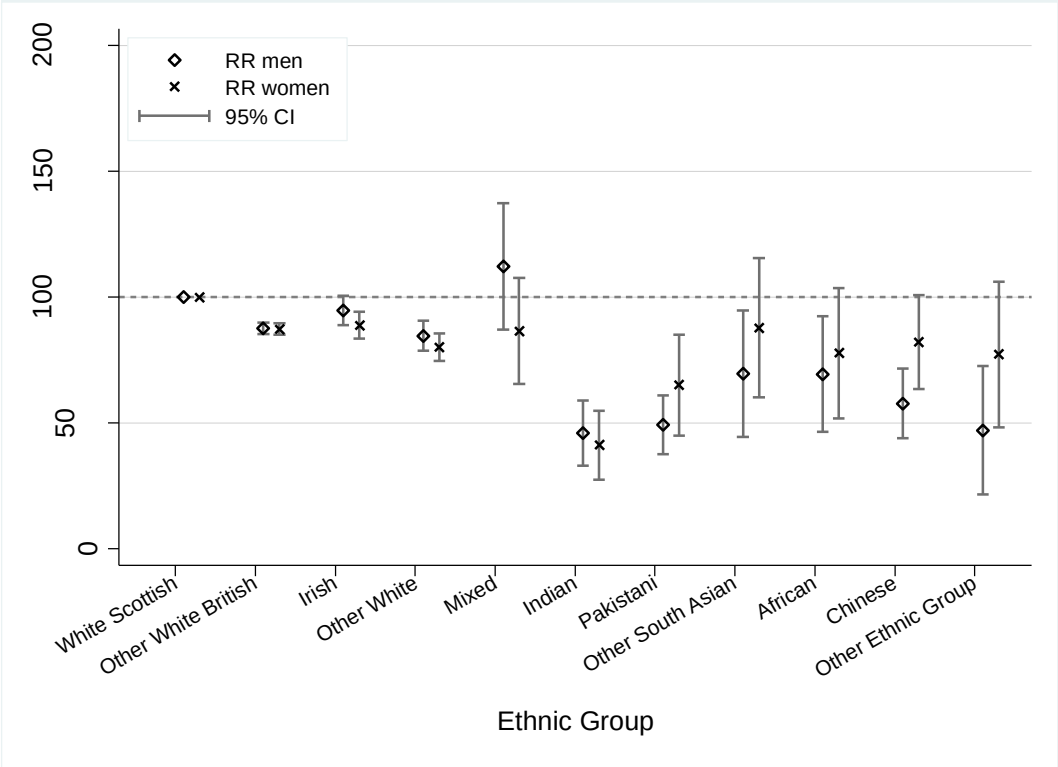


Figure 2. Lung cancer age standardised rate ratio by ethnic group

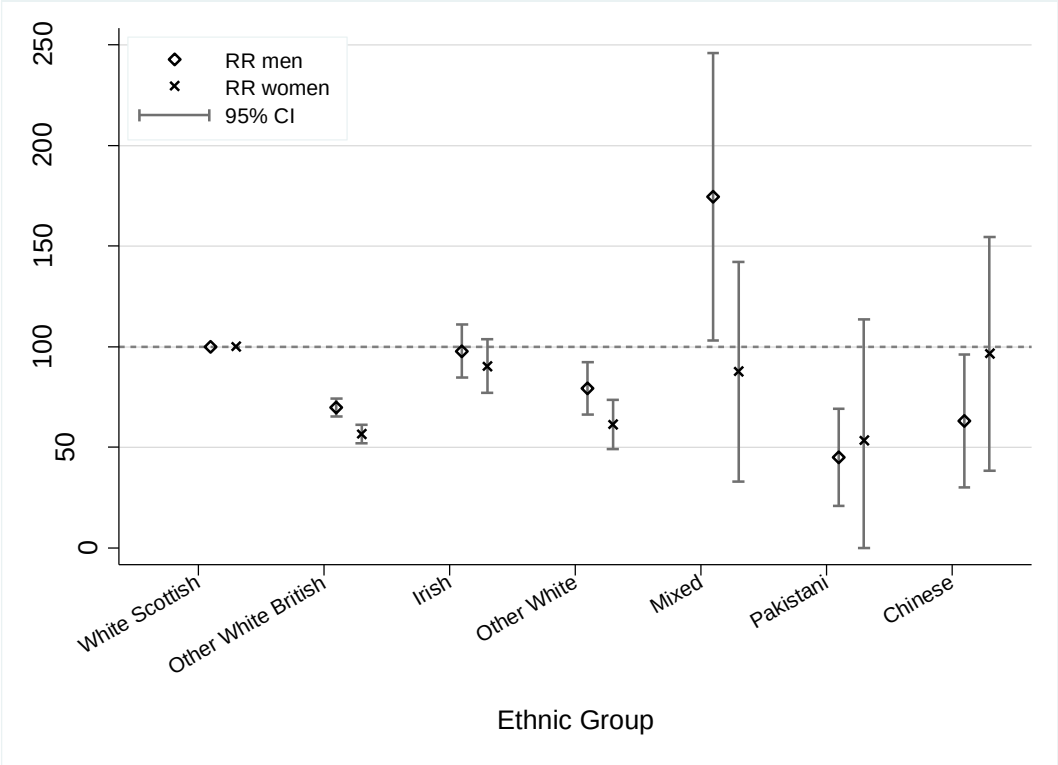


Figure 3. Colorectal cancer age standardised rate ratio by ethnic group

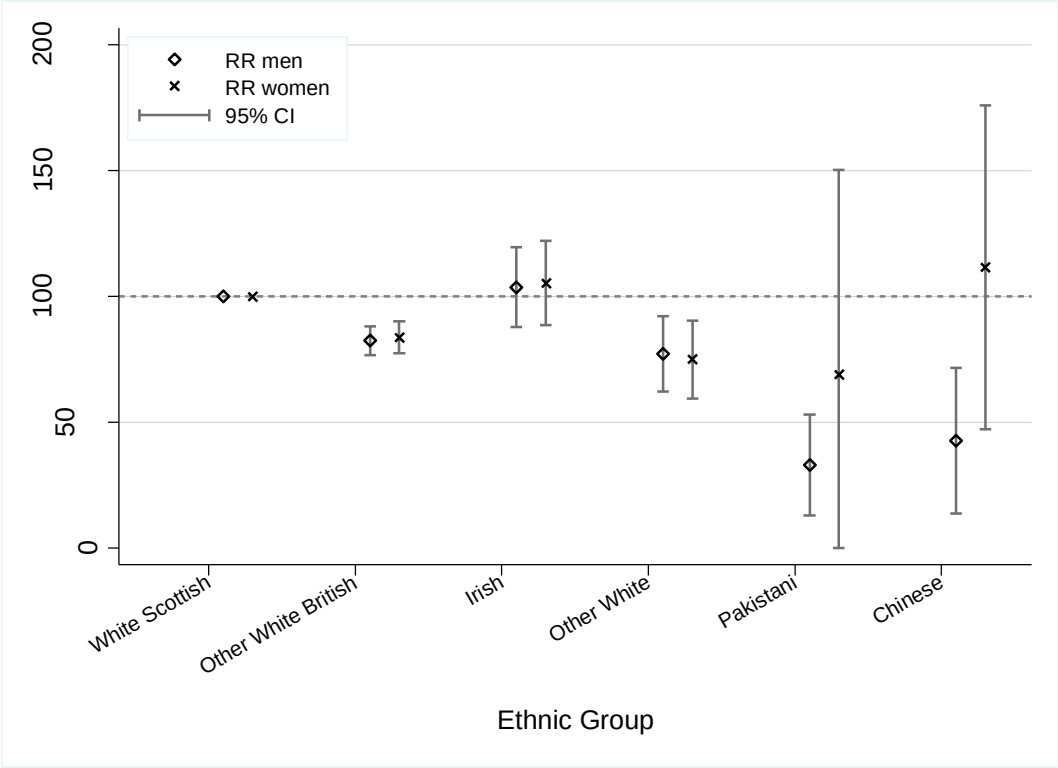


Figure 4. Breast cancer age standardised rate ratio by ethnic group

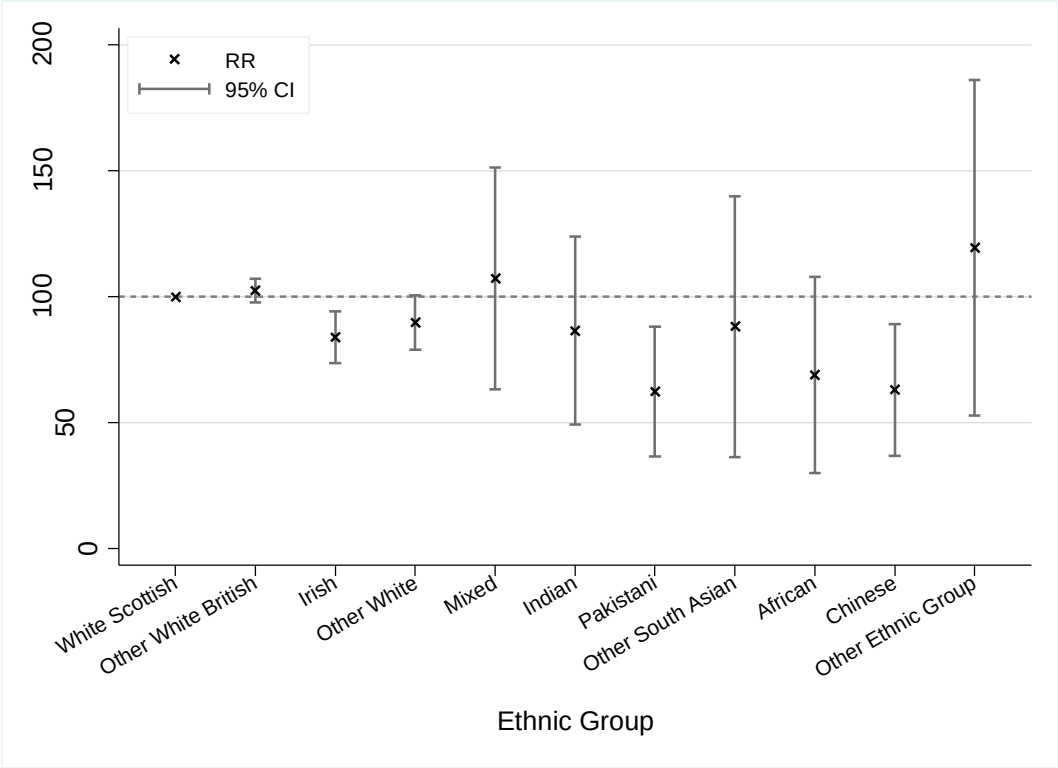
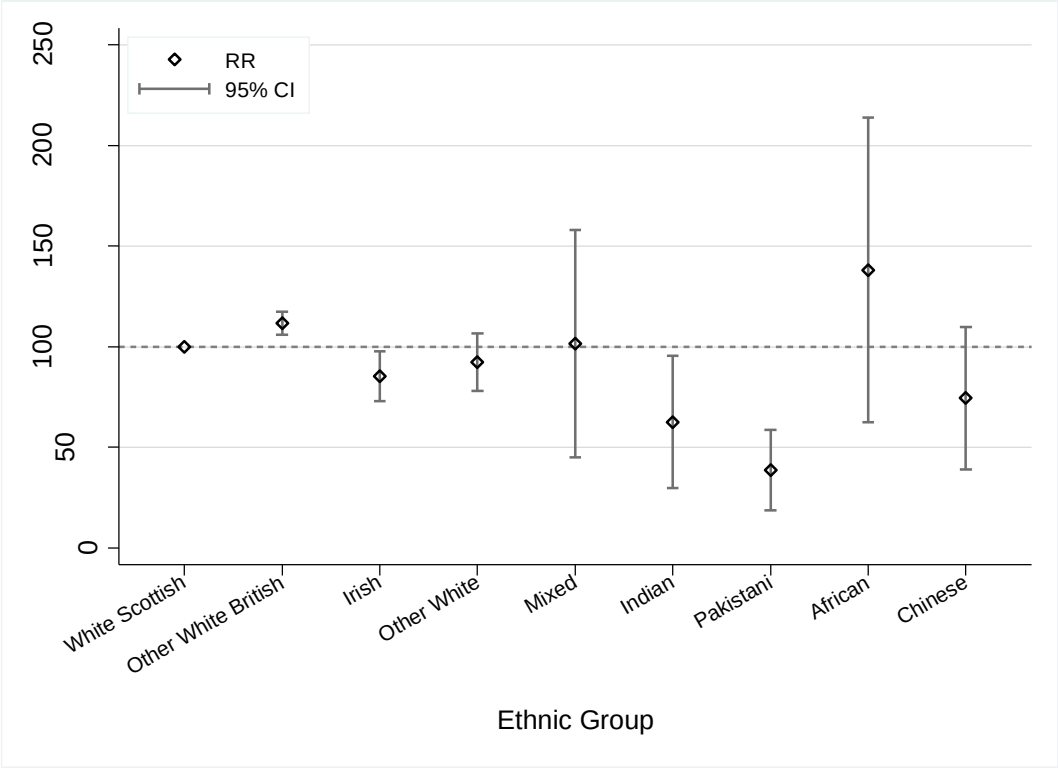


Figure 5. Prostate cancer age standardised rate ratio by ethnic group



Appendix Figure 1 Overview of Record Linkage Process