This is an author produced version of a paper published in *British Journal of Cancer*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/42873/

**Published paper**

http://dx.doi.org/10.1038/sj.bjc.6605903
Cancer incidence among the south Asian and non-south Asian population under 30 years of age in Yorkshire, UK


1Paediatric Epidemiology Group, Division of Epidemiology, University of Leeds, LS2 9NL, UK.

2Paediatric Oncology and Haematology, St James’s University Hospital, Beckett Street, Leeds, LS9 7TF, UK.

3Institute of Oncology, Bexley Wing, St James’s University Hospital, Beckett Street, Leeds, LS9 7TF, UK.

4School of Geography, University of Leeds, Leeds, LS2 9JT, UK.

*Correspondence: Dr Richard G. Feltbower; Email: r.g.feltbower@leeds.ac.uk

Running title: Incidence of cancer in 0-29 year olds by ethnic group.
Abstract

BACKGROUND:
Few studies have examined epidemiological differences between ethnic groups for children and young adults with cancer.

METHODS:
Subjects aged 0-29 diagnosed between 1990 and 2005 in the former Yorkshire Regional Health Authority were included in the analysis. Ethnicity (south Asian or not) was assigned using name analysis programs and Hospital Episode Statistics data. Differences in incidence (per 1,000,000 person-years) rates and trends were analysed using joinpoint and Poisson regression analysis.

RESULTS:
Overall cancer incidence was similar for south Asians (12.1, 95% CI 10.7-13.5; n=275) and non-south Asians (12.6, 95% CI 12.2-13.1; n=3259). Annual incidence rates increased significantly by 1.9% per year on average (95% CI 1.2-2.6%), especially for south Asians (7.0%; 95% CI 4.2-9.9%).

CONCLUSION:
If present trends continue, the higher rate of increase seen among the south Asian population in Yorkshire will result in 3-times higher incidence than non-south Asians by 2020.

Keywords: Childhood; Adolescent cancer; Ethnicity; Incidence
Introduction

Within Europe and the UK, several cancer epidemiology studies have shown differences in incidence rates between ethnic groups across all ages (Jack et al, 2009; Lane et al, 2007; NCIN, 2009).

Differences in incidence by ethnic group have also been observed amongst children (0-14 year olds) in the UK. Excesses of all cancers combined, leukaemia, acute myeloid leukaemia (AML), lymphoma, germ cell and hepatic tumours has been observed amongst south Asians compared to non-south Asians over the last 20 years (Cummins et al, 2001; McKinney et al, 2003; Powell et al, 1994; Winter et al, 1999). However, south Asian children have displayed lower incidence rates of Wilms’ tumours and rhabdomyosarcoma (Powell et al, 1994; Stiller et al, 1991). Furthermore, the well-described incidence peak for leukaemia amongst 2-6 year olds has been observed later in life for south Asians (5-9 year olds) than for non-south Asians (0-4 year olds) (McKinney et al, 2003; McNally & Eden, 2004). In the US, incidence of leukaemia was higher among Hispanics and lower amongst Blacks compared to White children and adolescents (Bleyer, 2002; Matasar et al, 2006).

While an increasing number of papers have recently reported on the descriptive epidemiology of teenage and young adult cancer in the UK, none have focused on differences by ethnic group (Birch et al, 2003; Birch et al, 2008). Moreover, it is unclear whether incidence differences exist exclusively in the older adult population or whether they begin to emerge in adolescence or young adulthood.

The aim of the work was to investigate for the first time in the UK incidence trends by ethnic group (south Asian or not) across the childhood (0-14 years) and young adult (15-29 years) age ranges exploiting the availability of population-based data on paediatric and
young adult cancer, whilst utilising information on ethnic group derived and validated using a combination of two name analysis programs and linked hospital episode statistics (HES) data.

MATERIALS AND METHODS

Case data

The data used for this study were extracted from the Yorkshire Specialist Register of Cancer in Children and Young People (YSRCCYP), a population-based dataset of children (ages 0-14) and young adults (ages 15-29) diagnosed with cancer whilst resident in the former Yorkshire Regional Health Authority since 1974 (Feltbower et al, 2009). Patients were eligible if diagnosed between 1990 and 2005 aged between 0-29 years. Skin carcinomas and melanomas were excluded. Case details were ascertained from hospital clinics and neuropathology departments across the region, and additional ascertainment checks undertaken for completeness were carried out with the National Registry of Childhood Tumours (www.ccrg.ox.ac.uk) and the Northern and Yorkshire Cancer Registry and Information Service (www.nycris.org.uk). 85% of all diagnoses recorded on the YSRCCYP have been histologically verified (Feltbower et al, 2009).

Tumour diagnoses were categorised into histological groups according to the International Classification of Childhood Cancer (ICCC) (Kramárová et al, 1996). For the purposes of certain analyses, malignancies were also grouped into the following four categories to retain statistical power: leukaemia, lymphoma, central nervous system (CNS) tumours and other solid tumours, corresponding to ICCC codes I, II, III and IV-XII.

The former Yorkshire region contained a population of approximately 1.4 million 0-29 year olds at the time of the 2001 census, 8% of whom described are of south Asian ethnic origin. This was an increase of 1.5% in the south Asian population since 1991. In 2001,
the study area contained 11% of all 0-29 year old south Asians in England, compared to only 7.4% of all 0-29 year old non-south Asians.

**Assignment of ethnic group**

Historically in the UK cancer registry data on ethnic group has not been routinely, or accurately, collected and thus alternative methods have been proposed in order to allocate ethnicity. Ethnicity was assigned as either south Asian (i.e. of Pakistani, Indian, or Bangladeshi origin) or non-south Asian (all other ethnicities) initially using the south Asian names analysis program Nam Pehchan and the South Asian Name and Group Recognition Algorithm (SANGRA), with visual inspection by local experts in case of any discrepancy. Both programs have a high level of accuracy and Nam Pehchan is particularly suited to the south Asian population of Yorkshire (Cummins et al, 1999; Nanchahal K et al, 2001).

In addition to both name analysis programs, independent cross-checking was completed using ethnicity data available from linked in-patient HES data. HES records were matched via NHS number, date of birth, postcode and sex. 62% (N=2195) of register patients were successfully linked to at least one HES record, no ethnic group was recorded in HES for 24% (N=517) of the linked patients, thus 1678 (47.5%) of patients’ ethnic classification derived from Nam Pehchan and/or SANGRA included in the analysis were validated using HES. For every hospital episode, ethnicity details are collected separately which can lead to an individual being allocated multiple ethnicity codes. The most frequently recorded ethnicity code was used to overcome this issue if present. Ethnicity codes within HES are based on the 16 categories as used in the 2001 census. These groups were aggregated to a binary classification of south Asian and non-south Asian. In cases where the most common code was listed as ‘unknown’, the most commonly occurring known group was used instead.
Where clear differences existed between the original 16 category HES code and the results of the name analysis (e.g. White British vs. south Asian), HES superseded the name analysis classification and the data were changed accordingly. However, in cases where the HES ethnicity code was ambiguous, such as ‘any other Asian’ or ‘any other ethnic group’ compared to a name classification of south Asian, the latter ethnicity assignment was retained. There was disagreement between assigned ethnicity based on HES data and name analysis in just 2.4% (n=40) of cases, of which 23 were re-allocated.

**Statistical Analysis**

Incidence rates and trends were examined overall and by major histological subtype. Internally age- and sex-standardised incidence rates (per 1,000,000 person years) and 95% confidence intervals by ethnic group were calculated using mid-year population estimates of the former Yorkshire Regional Health Authority based on the 1991 and 2001 UK census available by single year of age, sex and ethnic group. Population estimates used between 1990 and 2001 incorporated non-responses which were not included in the original UK census estimates; these figures also take into account changes in census geographies over time (Norman et al, 2008; Sabater & Simpson, 2009). Mid-year population estimates between 2002 and 2005 were experimental estimates by ethnic group provided by the Office for National Statistics (ONS, 2001 Census).

Incidence rates were calculated by 5-year age band (0-4, 5-9, 10-14, 15-19, 20-24, 25-29 years), diagnostic group and gender. Incidence trends were assessed by deriving smoothed 3-year moving averages for the study period in order to highlight temporal patterns within the data. Initially, joinpoint regression was used to compare differences in the incidence of cancer between south Asian and non-south Asian children and young people whilst adjusting for sex, age and year of diagnosis. Poisson regression models
were used when joinpoint analysis provided no evidence against a constant average annual percentage change (AAPC) in incidence rates. An interaction term between year and ethnic group was included in the best fitting model to assess whether the effect of ethnic group changed significantly over time. Incidence trends by ethnicity were forecasted to 2020 by extrapolating predicted values from the final model. Statistical analysis were implemented mostly in Stata 11, joinpoint regression was performed using the Joinpoint Regression Program version 3.4.3.

RESULTS
A total of 3534 children and young adults were registered with a tumour between 1990 and 2005 whilst living in the former Yorkshire region, of which 275 (7.8%) were south Asian.

Incidence
The number of cases, comparative frequencies and age-sex standardised incidence rates by diagnostic group, age group and gender for south and non-south Asians are given in Table 1. The distribution of cases was very similar across ethnic groups, except for a higher frequency of 25-29 year old non-south Asian cases compared to south Asians. There was a notable excess of male cases compared to females which was similar across ethnic groups. The incidence rate (per 1,000,000 person-years) for all cancers in 0-29 year olds between 1990 and 2005 was 12.6 (95% CI 12.2-13.0) and was similar overall between ethnic groups. Rates were non-significantly higher for south Asians with leukaemia and lymphoma but were lower for central nervous system (CNS) and other solid tumours. Slightly higher south Asian rates for all cancers combined were observed for those aged 5-9, 10-14 and 15-19, however, this was not significant. There was no significant difference in incidence when comparing male and female south Asians and non-south Asians for all tumour groups combined.
Age-specific incidence rates by ethnic group for each diagnostic group are given in Figure A of the online supplementary material. A small deficit amongst 25-29 year old south Asians compared to non-south Asians was seen consistently across all diagnostic groups, and was most evident amongst lymphomas and CNS tumours.

Poisson regression models showed that there was a significant difference between the incidence rates of south and non-south Asians by diagnostic group (likelihood-ratio test for ethnic group and diagnostic group interaction; p=0.01), adjusting for age group, sex and year of diagnosis. No significant difference in incidence between age group and ethnicity was apparent when testing for an interaction between the two variables.

**Temporal Trends in Incidence**

Figure 1 displays incidence trends between 1990 and 2005 for all cancers combined and each diagnostic group using smoothed 3-year moving averages (MA). Joinpoint regression analysis revealed that the best fitting model for all cancers combined and for each diagnostic group was one with no joinpoints, therefore Poisson regression was used to calculate the average annual percentage change (AAPC) in each case. Overall, there was a significant AAPC of 1.9% (95% CI 1.2-2.6%; p-value < 0.001). A significant change in incidence was also seen for leukaemia (AAPC=1.8%; 95% CI 0.2-3.4%; p-value = 0.027) and other solid tumours (AAPC=2.9%; 95% CI 1.7-4.1% p-value < 0.001). The latter change in incidence was accounted for by rises among bone tumours (4.3%; n=151), germ cell neoplasms (4.4%; n=614) and carcinomas (5%; n=94). No significant change in incidence was apparent for lymphoma (1.1%) or CNS (1.0%) tumours.
Likelihood-ratio tests showed significant differences between the temporal change in incidence by ethnic group (an interaction between year and ethnicity) for all cancers combined (p=0.0002). Joinpoint regression analysis again showed that the best fitting model in each case for both south and non-south Asians had no joinpoints, and thus the AAPC was calculated using Poisson regression. The rate of increase among south Asians was much higher (AAPC=7%; 95% CI 4.2-9.9%) compared to non-south Asians (AAPC 1.5%; 95% CI 0.8-2.3%); this finding was even more pronounced for those south Asians aged 15-29 (AAPC=11%; 95% CI 4.0-18.0%). Similarly, there was a significantly higher rate of increase for south Asians with lymphoma (AAPC 12.0%, 95% CI 5%-18%) compared to non-south Asians (AAPC 0.1%; 95% CI -1.5% to 1.7%), although the number of south Asians with lymphoma between 1990 and 2000 was extremely small (n=28). This meant that rates of all cancers combined among south Asians matched those of the indigenous population by 2005 despite exhibiting much lower rates before the mid-1990s (Figure 1). This difference in incidence trends by ethnic group was not mirrored among those with leukaemia, CNS tumours or other solid tumours.

Temporal incidence rates comparing 0-14 year olds and 15-29 year olds by diagnostic group showed a significant increase in incidence rates for 15-29 year olds between 1990 and 2005 for all cancers combined (AAPC=2.9% 95% CI 2.0-3.9%), leukaemia (AAPC=4.2%, 95% CI 1.3-7.1%) and other solid tumours (AAPC=4.4%, 95% CI 3.0-5.9%). No significant change in incidence was apparent for 0-14 year olds (0.01% for all cancers combined).

Figure 2 shows the age- and sex-standardised incidence rates for south and non-south Asians for all cancers combined alongside predicted incidence rates from 2005 to 2020. If the AAPC continues to increase at a constant rate, the incidence of cancer
amongst south Asians is set to increase almost 3-fold from 2005 onwards to be approximately 56 per million compared to 17 per million for non-south Asians by 2020.

**DISCUSSION**

We report details of the first population-based epidemiological study looking at incidence rates and trends across the childhood, teenage and young adult age range using a validated ethnic group classification. Previous epidemiological studies focusing on ethnic minority populations have relied on a single source of data on ethnic group (Cummins *et al.*, 2001; McKinney *et al.*, 2003). However, our analysis incorporated information from the patient’s name alongside a self-reported ethnic classification derived from routine hospital admissions. Although we were able to validate only half of the cohort using this data source, mis-classification rates were extremely low suggesting that the primary methodology based on name recognition software algorithms was an accurate and reliable approach. An estimated 60% of the south Asian population within the study region (mainly West Yorkshire) originates from Mirpur in rural Pakistan; making it one of the few regions in the UK which allows for a detailed analysis of a relatively homogeneous south Asian population (The Change Institute & Department for Communities and Local Government, March 2009; Turton & Gonzáles, 1999). We were also able to exploit mid-year population estimates broken down by age, sex and crucially ethnic group. These were derived from recent research to enhance previous census and mid-year estimates with adjustments for differential non-response by age-sex and ethnic group and upon the most recent ONS estimates. We are therefore confident in the accuracy of our population denominator and ethnic group assignment.

We found a significantly higher rate of increase in incidence among the south Asian population of Yorkshire in comparison to non-south Asians, which resulted in similar
overall incidence rates between ethnic groups across the study period (1990-2005). This difference in incidence trends was apparent for all cancers combined, but most pronounced for lymphoma, and seen in particular among 15-29 year olds.

Our findings suggest there may be an aetiological component, possibly environmental in nature, which could be specific to the context of the south Asian population in Yorkshire. Other epidemiological studies focusing on chronic disease epidemiology in Yorkshire have suggested there may be a dietary (Edwards et al, 2006) or infectious (Feltbower et al, 2005; Parslow et al, 2001) component to aetiology and future work should look in more detail especially within the lymphoma group to identify plausible causal hypotheses. The temporal variation in incidence trends by ethnic group could in part be confounded by socioeconomic status (Varghese et al, 1996) or genetic differences between Mirpuri south Asians and other ethnic groups (De Vos et al, 2006). We plan to examine ethnicity with respect to socioeconomic status in a further detailed ecological analysis.

The differential increase in incidence among the south Asian population also has implications for healthcare resource planners in the region. Our incidence projections suggest the burden of disease for south Asians is likely to result in almost 3-times higher rates by 2020 assuming that the linear increase since 1990 continues and the age structure of the Yorkshire population remains the same over time. The highest proportion of south Asians reside within the highly urbanised district of Bradford in West Yorkshire, comprising 48% of the total south Asian population of Yorkshire, and it is here that the increase in the burden of disease on the NHS will be most strongly felt.
We saw a slightly higher incidence of leukaemia and lymphoma among south Asians, but a deficit for CNS tumours and other solid tumours. This supports findings from previous UK childhood investigations (Cummins et al, 2001; McKinney et al, 2003; Powell et al, 1994; Stiller et al, 1991; Winter et al, 1999). Studies which looked at the older childhood and young adult age range revealed similar temporal increases in incidence which was more pronounced among the other solid tumour groups, particularly melanoma and carcinoma (Birch et al, 2002; Kroll et al, 2006; Reedijk et al, 2005). These changes in incidence could in part be attributed to improvements in diagnostic techniques and better case ascertainment more recently, although the latter is unlikely due to annual cross-checks with other regional and national cancer registries and neuropathology databases since the early 1990s (see methods), whilst histological verification of tumours has exceeded 85%.

In summary, despite the similarity in incidence rates between south Asian and non-south Asians in Yorkshire, our data revealed a sharper increase for the south Asian population which if present trends persist will result in a 3-fold higher incidence by 2020 compared to non-south Asians.

ACKNOWLEDGEMENTS

We thank the Candlelighters Trust for funding the Yorkshire Specialist Register of Cancer in Young People. We acknowledge the help of Isabel dos Santos Silva from the London School of Hygiene and Tropical Medicine for allowing the use of the SANGRA software package. We are enormously indebted to the technical expertise of James Thomas, Eva Morris and Rebecca Birch based at the Northern and Yorkshire Cancer Registry and Information Service for HES data linkage and for helpful discussions over ethnic group validation; we also thank the Thames Cancer Registry for collating the HES data on behalf of the UK Association of Cancer Registries. Finally,
we are grateful to Paula Feltbower for meticulous data collection and the co-operation of all oncologists, pathologists, GPs and medical records staff in Yorkshire.


ONS (2001 Census) Standard Area Statistics (England and Wales) Experimental Statistics Tables EE1-EE4. ESRC/JISC Census Programme, Census Dissemination Unit. Mimas (University of Manchester)


The Change Institute, Department for Communities and Local Government (March 2009) *The Pakistani Muslim Community in England, Understanding Muslim Ethnic Communities*; Crown Copyright.


Table 1
Number of cases and age-sex standardised incidence rates for south Asians and non-south Asians by diagnostic group, age group and gender (1990-2005)

<table>
<thead>
<tr>
<th>Diagnostic Group</th>
<th>Number of Cases</th>
<th>Incidence Rates</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>South Asian</td>
<td>Non-south Asian</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>Cases (column %)</td>
<td>Cases (column %)</td>
<td>Rate (95% CI)</td>
</tr>
<tr>
<td>Leukaemia</td>
<td>66 (24%)</td>
<td>649 (20%)</td>
<td>715</td>
</tr>
<tr>
<td>Lymphoma</td>
<td>71 (26%)</td>
<td>689 (21%)</td>
<td>760</td>
</tr>
<tr>
<td>CNS Tumours</td>
<td>37 (13%)</td>
<td>562 (17%)</td>
<td>599</td>
</tr>
<tr>
<td>Other Solid Tumours</td>
<td>101 (37%)</td>
<td>1359 (42%)</td>
<td>1,460</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Age at diagnosis (years)</th>
<th>Number of Cases</th>
<th>Incidence Rates</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>South Asian</td>
<td>Non-south Asian</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>Cases (column %)</td>
<td>Cases (column %)</td>
<td>Rate (95% CI)</td>
</tr>
<tr>
<td>0-4</td>
<td>66 (24%)</td>
<td>650 (20%)</td>
<td>716</td>
</tr>
<tr>
<td>5-9</td>
<td>40 (15%)</td>
<td>364 (11%)</td>
<td>404</td>
</tr>
<tr>
<td>10-14</td>
<td>37 (13%)</td>
<td>375 (12%)</td>
<td>412</td>
</tr>
<tr>
<td>15-19</td>
<td>42 (15%)</td>
<td>463 (14%)</td>
<td>505</td>
</tr>
<tr>
<td>20-24</td>
<td>41 (15%)</td>
<td>597 (18%)</td>
<td>638</td>
</tr>
<tr>
<td>25-29</td>
<td>49 (18%)</td>
<td>810 (25%)</td>
<td>859</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Number of Cases</th>
<th>Incidence Rates</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>South Asian</td>
<td>Non-south Asian</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>Cases (column %)</td>
<td>Cases (column %)</td>
<td>Rate (95% CI)</td>
</tr>
<tr>
<td>Male</td>
<td>159 (58%)</td>
<td>1995 (61%)</td>
<td>2,154</td>
</tr>
<tr>
<td>Female</td>
<td>116 (42%)</td>
<td>1264 (39%)</td>
<td>1,380</td>
</tr>
<tr>
<td>Total</td>
<td>275</td>
<td>3,259</td>
<td>3,534</td>
</tr>
</tbody>
</table>