Detecting small area similarities in the epidemiology of childhood acute lymphoblastic leukaemia and type 1 diabetes: a Bayesian approach

Richard G Feltbower1*, Samuel OM Manda2, Mark S Gilthorpe2, Roger C Parslow1 and Patricia A McKinney1.

1 Paediatric Epidemiology Group and 2Biostatistics Unit, Unit of Epidemiology and Health Services Research, University of Leeds, 30 Hyde Terrace, Leeds, LS2 9LN, UK.

*Tel +44(0)113 343 4841, fax +44(0)113 343 4877, Email r.g.feltbower@leeds.ac.uk.

Summary
Childhood acute lymphoblastic leukaemia (ALL) and type 1 diabetes have common epidemiological and etiological features including correlated international incidence and associations with infections. We tested whether their similar large-scale distributions are reflected in small geographical areas whilst examining the influence of sociodemographic characteristics. Details of 299 children (0-14 years) with ALL and 1551 with diabetes diagnosed between 1986-1998 were extracted from two registers in Yorkshire, UK. Standardized incidence ratios (SIRs) across 532 electoral wards (EW) were compared using Poisson regression, confirming significant associations between population mixing and the geographical heterogeneity of both conditions. Bayesian methods were applied looking for correlation between diseases by modeling a bivariate outcome based on their SIRs; spatial and heterogeneity components were included within a hierarchical random effects model. A small positive correlation between diseases of 0.13 (-0.40, 0.58) was observed and this reduced after controlling for population mixing (0.12), population density (0.06) and deprivation (0.05). The Bayesian approach showed a small but non-significant joint spatial correlation between diseases, only partially suggesting the risk of both was associated within some EWs. Population mixing remained significantly associated with both diseases using Bayesian methodology. Links between diabetes and ALL observed for large regions are considerably weaker for small areas. More powerful replications are needed for confirmation of these findings.

Introduction
Acute lymphoblastic leukemia (ALL) is the most common form of childhood leukemia comprising around 20% of all cancers in those diagnosed under the age of fifteen (UK Childhood Cancer Study Investigators, 2000). Childhood type 1 or insulin dependent diabetes arises as a result of an immune mediated destruction of the insulin producing beta cells of the pancreas (Atkinson and Maclaren, 1994) and in developed countries accounts for part of the increasing burden of chronic disease in children (Currie et al, 1997).

Exposure to infectious agents has been linked to the aetiology of both childhood ALL and type 1 diabetes over recent years (Greaves, 2002; Kyvik et al, 1995). One theory suggests that common infections in early life may reduce the risk of future development of disease by sufficient stimulation of the naïve immune system (Rook and Stanford, 1998). In other words, lack of exposure to such pathogens may trigger an abnormal immune response manifesting itself in disease onset. This idea has been encapsulated in the 'delayed infection' hypothesis for childhood ALL (Greaves, 1997) and the 'hygiene hypothesis' for both type 1 diabetes and allergy/asthma (Wills-Karp et al, 2001).

The joint environmental associations between these two conditions have been illustrated through the international correlation between their incidence rates (Feltbower et al, 2004). With access to two population-based registers of childhood diabetes (Feltbower et al, 2003) and cancer (McKinney et al, 1998) from the same geographical region of the UK, we were able to investigate the occurrence of ALL and diabetes across small areas. In view of the sparsely distributed case data, we adopted a Bayesian approach in examining smoothed standardised incidence ratios (SIR) and modelling their joint spatial association using a bivariate outcome in relation to socio-demographic factors, including a proxy measure for exposure to infections known as population mixing. The concept of population mixing has been applied to geographical and infectious disease epidemiology as a potential proxy measure for exposure to infections and its association quantified separately for both childhood leukaemia (Kinlen, 1988; Stiller and Boyle, 1996, Parslow et al, 2002; Dickinson et al, 2002, Law et al, 2003) and diabetes (Parslow et al, 2001).

We aimed to test (i) whether the occurrence of childhood ALL was spatially associated with diabetes in small areas and (ii) whether area-based socio-demographic risk factors affect the degree of correlation in incidence between these conditions.

Materials and Methods
We extracted data on children diagnosed with ALL and type 1 diabetes between 1986-1998 from two population-based disease registers covering the former Yorkshire Regional Health Authority in the north of the UK. We limited the case series to a period centred at the time of the 1991 national UK census to ensure the
inclusion of relevant socio-demographic denominator data. Patients’ addresses and postcodes at the time of diagnosis were validated and linked to one of 532 Electoral Wards in existence in Yorkshire at the time of the 1991 Census. These small geographical areas have a median population count of 4,200 (range 500 to 25 000). Population estimates from the 1991 Census were used to calculate age-sex standardised incidence rates.

For each disease, a Poisson regression model was fitted to the observed numbers of counts in each Ward. We initially examined the distribution of cases for both diseases across Wards by calculating and mapping spatially (locally) smoothed estimates of the SIRs. Secondly, we compared separately the risk for both diseases from three socio-demographic factors previously linked to disease onset. These included: (i) Population mixing, measured using the Shannon index (Stiller and Boyle, 1996), reflecting the diversity of origins of incomers into each Ward and calculated for the childhood (0-14 years) population; (ii) Person-based childhood population density and (iii) Deprivation, measured using the Townsend Score, standardised to all Wards in Yorkshire.

Thirdly, we modelled the two disease counts jointly, examining the effects from each covariate using two Bayesian forms of spatial smoothing. This was done using spatially unstructured and structured random effects which control for the unobserved spatial covariates. A convolution model developed by Besag and colleagues (1991) was used to model each pair of ALL and diabetes counts for the 532 Wards. The effect on the degree of correlation between both diseases was examined before and after allowing for each socio-demographic factor previously linked to the spatial distribution of disease incidence. In our Bayesian analysis of the model, all parameters were given standard and relatively non-informative prior distributions. Posterior estimation of all the model parameters was done using the Gibbs sampling algorithm implemented in the software package WinBUGS.

Results

We identified 299 children aged 0-14 years diagnosed with ALL and 1551 diagnosed with type 1 diabetes. Lower rates of ALL and diabetes were seen in the more urban county of West Yorkshire than other parts of the Region, whilst higher rates of both conditions were observed in the more rural county of North Yorkshire. The south-eastern part of the Region below the Humber estuary showed large fluctuations in SIRs for both disease groups (figures 1a-1b).

Figure 1a Spatially smoothed standardized incidence ratios for childhood type 1 diabetes diagnosed between 1986-1998 across electoral wards in Yorkshire, UK

Classical approach:

Higher rates of diabetes and ALL were present in areas of low population mixing, and this effect remained after adjusting for the other two covariates. Significantly lower rates of ALL were observed in areas with very high population mixing, although no difference in incidence was seen for diabetes. An inverse association was present for population density for each condition with lower rates associated with higher levels of density. However, once the effects from population mixing and deprivation were taken into consideration, the
association with population density disappeared for diabetes and was reversed for ALL. There was some evidence of a negative association between deprivation and diabetes, with lower rates observed in more deprived areas. There was no systematic relationship between deprivation and ALL.

Bayesian approach:
We modeled the effects of both disease counts together as a bivariate outcome. Assuming dependent random effects between diseases with no adjustment for covariates, we found a small degree of positive correlation between diseases of 0.129 (-0.399, 0.579). Compared to the classical univariate model, the parameter estimates largely remained the same after allowing for dependent random effects and the contribution of each socio-demographic covariate on its own. After separately accounting for population mixing, population density, and deprivation, the partial correlation between diseases fell from 0.129 to 0.117 (-0.429, 0.564), 0.062 (-0.454, 0.561) and 0.045 (-0.521, 0.509), respectively. This corresponded to an approximate reduction of 10%, 50% and 65%, respectively. After adjusting for all three covariates simultaneously (table 1), the correlation almost disappeared entirely, falling to 0.016 (-0.499, 0.522). The parameter estimates were similar to the adjusted IRRs from the classical approach.

Figure 1b Spatially smoothed standardized incidence ratios for childhood acute lymphoblastic leukemia diagnosed between 1986-1998 across electoral wards in Yorkshire, UK

Conclusion
Despite the reported similarities in their epidemiology, this had only been addressed formally in a study comparing the international correlation in their respective incidence rates (Feltbower et al, 2004), which found a strong positive association of 0.53 (0.36, 0.72). We performed an analogous study investigating whether the risk for both diseases was similar across small geographical areas in the north of the UK, showing a small and weaker joint spatial correlation of 0.13 (-0.40, 0.58). This finding was also illustrated by mapping the spatial distribution of each disease: generally rates were lower in the more populated county of West Yorkshire than other areas and higher in the less populated county of North Yorkshire.

The correlation almost disappeared entirely once we allowed for the effects of population density and deprivation; no effect was observed on the size of the correlation after adjusting for population mixing. The absence of any influence from population mixing on the joint geographical distributions of both diseases is a new observation. This is not necessarily inconsistent with previously reported studies. One possible explanation for this is that population mixing may have had a strong but erratic effect across Wards yielding an association with disease occurrence in the classical approach, whereas deprivation/population density had a modest but consistent effect across small areas, which disappeared when considering the joint spatial correlation. There was considerable heterogeneity in incidence across Wards accounting for three-quarters of the observed variation in disease outcome for each condition and it may be surprising that the data only revealed a small, non-significant degree of spatial correlation.
Table 1  Fixed and random effects estimates (median and 95% credible intervals) from a Bayesian bivariate model with dependent errors and all three covariates considered simultaneously

<table>
<thead>
<tr>
<th>Random effects</th>
<th>Diabetes</th>
<th>Leukemia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-spatial (heterogeneity)</td>
<td>0.067 (0.042-0.106)</td>
<td>0.130 (0.066-0.256)</td>
</tr>
<tr>
<td>Spatial*</td>
<td>0.121 (0.062-0.239)</td>
<td>0.237 (0.095-0.628)</td>
</tr>
<tr>
<td>Spatial correlation</td>
<td>0.016 (-0.499, 0.522)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Population density</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>medium</td>
<td>1.021 (0.830-1.259)</td>
<td>1.152 (0.731-1.885)</td>
</tr>
<tr>
<td>High</td>
<td>0.957 (0.746-1.123)</td>
<td>1.250 (0.737-2.186)</td>
</tr>
<tr>
<td>Population mixing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.282 (0.904-1.780)</td>
<td>1.242 (0.523-2.614)</td>
</tr>
<tr>
<td>10&lt;sup&gt;th&lt;/sup&gt;-90&lt;sup&gt;th&lt;/sup&gt;</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>&gt;90&lt;sup&gt;th&lt;/sup&gt;</td>
<td>0.947 (0.800-1.117)</td>
<td>0.646 (0.448-0.916)</td>
</tr>
<tr>
<td>Deprivation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (lowest)</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>2</td>
<td>1.084 (0.863-1.362)</td>
<td>1.462 (0.868-2.520)</td>
</tr>
<tr>
<td>3</td>
<td>1.091 (0.876-1.364)</td>
<td>1.373 (0.829-2.342)</td>
</tr>
<tr>
<td>4</td>
<td>0.931 (0.740-1.178)</td>
<td>1.096 (0.652-1.885)</td>
</tr>
<tr>
<td>5 (highest)</td>
<td>0.832 (0.650-1.070)</td>
<td>0.871 (0.506-1.543)</td>
</tr>
</tbody>
</table>

Discussion

Although childhood ALL and diabetes display strongly correlated incidence rates at the national level, this association is weakened considerably when investigating the distribution across small geographical areas. Deprivation and population density appear to explain more of the spatial correlation between the diseases than population mixing. The parallels in the descriptive epidemiology of type 1 diabetes and childhood ALL and the role of deprivation/population density/population mixing suggests that a joint investigation of common causal pathways and underlying genetic susceptibility in individuals would be informative. Firstly, however, our findings need to be replicated in other populations and we are planning to conduct more extensive geographical analyses for other countries.

References

Greaves M. Childhood leukemia. BMJ 2002;324:283-7
Greaves M. Etiology of acute leukemia. Lancet 1997;349:344-349
Rook GAW, Stanford JL. Give us this day our daily germs. Immunol today 1998;19:113-116