Research Data Related to this Submission

Data set https://data.mendeley.com/datasets/j2wscc6tdd/draft?a=6a4b0cd9-f7f9-4426-8a88-3e1599d76f50

Data for: A Diabetes Needs Index for Small Areas in England

England MSOAs Data Diabetes Outcomes and Risk Factors

Highlights

- There are relatively few studies of the small area pattern of diabetes in the UK and of factors influencing it.
- Evidence is lacking on urban-rural contrasts and impacts of the social environment
- Develops a diabetes risk index for English MSOAs (with England-wide coverage), which can be applied to assess diabetes health care need
- The index shows the highest diabetes risk in more urban areas, and shows the clustering of higher diabetes in such areas
- Affluence offsets urbanity effect in some regions, while low area socio-economic status amplifies urbanity effect
- Impacts of social environment on diabetes remain after allowing for area SES and ethnicity

A Diabetes Needs Index for Small Areas in England

Corresponding Author:
Peter Congdon, School of Geography, Queen Mary University of London, Mile End Rd, London E1 4NS. Email: p.congdon@qmul.ac.uk. Phone: 020 7882 8200. Fax: 020 7882 7032

A Diabetes Risk Index for Small Areas in England

Abstract

UK and international studies point to significant area variation in diabetes risk, and summary indices of diabetic risk are potentially of value in effective targeting of health interventions and healthcare resources. This paper aims to develop a summary measure of the diabetic risk environment which can act as an index for targeting health care resources. The diabetes risk index is for 6791 English small areas (which provide entire coverage of England) and has advantages in incorporating evidence from both diabetes outcomes and area risk factors, and in including spatial correlation in its construction. The analysis underlying the risk index shows that area socio-economic status, social fragmentation and south Asian ethnic concentration are all positive risk factors for diabetes risk. However, urban-rural and regional differences in risk intersect with these socio-demographic influences.

Keywords. Diabetes; Risk; Urbanity; Mortality; Prevalence; Morbidity

1. Introduction

Increases in diabetes prevalence and related adverse outcomes are a concern for public health agencies. Estimated global diabetes prevalence (among 20-79 year olds) has risen from 151 million to 463 million in 2019 (International Diabetes Federation, 2020). UK figures show 3.8 million people with diagnosed diabetes (aged 17 and over) in 2018, 7.2% of the adult population (Diabetes UK, 2020), and an increase of 2.1 million since 1998.

UK and international studies confirm significant spatial inequalities in the patterning of diabetes (Cox et al, 2007; Toms et al, 2019). To tackle upwards diabetes trends, local and regional health agencies may therefore prioritise diabetes related interventions to small areas identified as having higher disease risk (Gabert et al, 2016; Matthews et al, 2013). However, there is a paucity of recent evidence on the pattern of diabetes outcomes in UK small areas, and on the factors affecting it, though the policy need for such information is now recognised (House of Commons Library, 2019).

In the face of research gaps and the need to use health care resources effectively, small area indices of diabetes risk or diabetes health care needs

have a major role. Area based health need indices to guide health interventions and resourcing are most commonly based on data reduction methods using socio-economic indicators (e.g. Earnest et al, 2015; Pampalon et al, 2009; Sundquist et al, 2003); that is, they do not include observed morbidity in their construction. A few studies include observed health outcomes such as prevalence, or health care activity (e.g. hospital admissions, prescribing rates, outpatient attendances). Thus Glover et al (2004) include a single health outcome (e.g. psychiatric hospital admissions) and regress it on socioeconomic risk indicators. Wang and Wall (2003), in the biostatistics literature, develop a cancer morbidity scale based on four types of cancer, though without incorporating socioeconomic indices.

Here we develop a health risk index for diabetes (which can serve as an index of need for diabetes health care) across 6791 English small areas, these areas covering all of England, so that the study is national in the sense that England is one of the UK nations. This index is distinct from generic deprivation indices (e.g. Allick et al, 2020) in that it is addressing the area specific risk of a particular disease. The index is partly based on diabetes outcomes (mortality and prevalence), but we also incorporate ecological (area-level) risk factors for diabetes, such as area socio-economic status, social fragmentation, obesity, and urbanity in the construction of the risk index.

The risk index has, as a major aim, to provide an index for allocating health resources, especially primary care and public health resources, as for existing indices of health need (e.g. Sundquist et al, 2003). A more general aim is to identify areas with a diabetogenic environment, where diabetes levels are high, and where contributory contextual factors are also elevated. It may also be used to identify aberrant health care activity (e.g. emergency diabetes hospitalizations in excess of what the need index suggests). The index can therefore serve more than one purpose: provide an overall morbidity measure, provide a measure of need for diabetes health care, or provide an index against which to assess diabetes activity levels.

The small areas used in the study were defined originally for UK Census purposes (Office of National Statistics, 2020), and are known as Middle Level Super Output Areas, or MSOAs for short. MSOAs are nested within larger agencies: 207 Clinical Commissioning Groups (CCGs) for health administration, 332 Local Authorities for local government, and nine regions (see Figure 1).

The MSOAs provide a workable spatial scale: they have an average (all ages, 2017 estimates) population of 8190, with 5th and 95th percentiles of 5730 and 11600. It is important that relatively low scales are used in ecological studies since “in choosing a suitable geographic area, those with the smallest possible population size are preferred as this is likely to mean better homogeneity among the population and reduce the risk of ecological fallacy” (Allick et al, 2020, p. 2). On the on the other hand, a lower spatial scale, such as the 32000 LSOAs in England, has the problem that health outcome data may become unduly sparse (as well as difficult to obtain), with a loss of information (e.g. diabetes deaths for LSOAs, even over a five year period, would imply many zero counts).

Spatial correlation in area health outcomes is frequently reported (Tosetti et al, 2018), due to clustering of high disease risk in neighbouring areas, and this clustering is apparent in diabetes outcomes (e.g. Green et al, 2003) A small area health disease risk or needs index should ideally incorporate such correlation, though this feature is not in fact included in existing need indices. We use a spatial factor method, with Bayesian estimation (Wang and Wall, 2003), allowing for spatial clustering in the risk index, and also for uncertainty in index scores (Marí-Dell'Olmo et al, 2011).

This approach is extended here to both include morbidity outcome indicators (such as mortality and prevalence), and potential risk factors (such as area socio-economic status). Both outcome indicators and risk factors are components of the overall risk score. In the multivariate literature, these are known as a reflective and formative indicators respectively (Coltman et al, 2008). Another terminology for this type of model is a “multiple indicators, and multiple causes” or MIMIC model (e.g. Proitsi et al, 2011; Wang and Wang, 2019), subject of course to caveats around causal interpretations being based on area data.

The technique used therefore has the benefit of providing a small area index of varying diabetes risk, and hence need for diabetes care, but also provides evidence on community factors affecting small area variations in diabetes morbidity. In particular, we report on regional differences in diabetes risk, on urban-rural risk variations, and risk contrasts according to socio-economic and behavioural factors.

2. Methods

2.1 Area Risk Factors

The spatial factor method is here adapted to include area-based risk factors. Research in the UK and elsewhere has shown strong associations between diabetes and area poverty and material deprivation (Connolly et al, 2000; Nishino et al, 2015; Walker et al, 2011; Hsu et al, 2012). These in turn may be linked to variations in individual risk behaviours, which vary according to area socioeconomic status. In the UK, Diabetes UK (2010) report that "[area] deprivation is strongly associated with higher levels of obesity, physical inactivity, unhealthy diet, smoking and poor blood pressure control, all of which are linked to the risk of developing Type 2 diabetes". UK studies have used summary deprivation measures to explain ecological variations in diabetes outcomes (e.g. Evans et al, 2000; Connolly et al, 2000; Walker et al, 2011; Fleetcroft et al, 2017) though these measures may incorporate illness measures in their construction, raising possible issues of endogeneity.

In the present study we use MSOA data based on the approximate social grade (ASG) socio-economic classification, produced by the UK Office for National Statistics. Links between diabetes risk and occupational structure are confirmed in other studies (Poulsen et al, 2014; Vazquez et al, 2019). In particular, we use an Index of Concentration at Extremes (Malla et al, 2020; Krieger et al, 2018). This is obtained as $[(DE)_i-(AB)_i]/T_i$, where T_i is the total population aged 16-64; $(DE)_i$ is the number in ASGs D and E, namely in semi-skilled and unskilled manual occupations, unemployed and lowest grade occupations; and $(AB)_i$ is the number in higher and intermediate managerial, administrative, and professional occupations. This index has a theoretical range between -1 and +1, and measures extreme concentrations of lower skill and low income groups. It gives a better fit than using the percent in ASGs D and E per se, or an income deprivation score (Smith et al, 2015). For brevity, we denote this measure as ICE-ASG.

Some studies, though none UK based, have considered impacts of area social cohesion on diabetes outcomes, including diabetes self care (which influences mortality) (Gariepy et al, 2013; Walker et al, 2016). Gebreab et al (2017) suggest that adverse neighbourhood social environments may "contribute to T2DM through stress, transmission of negative health behaviours, and lack of social support". Here we assess effects on diabetes

risk of an inverse measure of social cohesion, namely a social fragmentation index (SFI) (Pabayo et al, 2014; Congdon, 1996). Pampalon et al (2009) refer to social fragmentation as a form of social deprivation, which they find relevant, along with material deprivation, to explaining small area variation in premature (all cause) mortality. The SFI index is used as an inverse measure of community cohesion in recent government initiatives to measure wellbeing (Brown et al, 2017).

The SFI index is derived from a principal component analysis of four Census 2011 variables for the 6791 MSOAs (percentages of one person households, of adults over 15 not married, of private sector renting, and of migration within the previous year): scores on the first component account for 73% of the variation in the four indicators. Scores on the SFI are higher in transient areas with high population turnover, high numbers of non-family and one person households, and high private sector renting. In the UK, private sector renting is typically short-stay insecure accommodation and surveys show private renters less likely to trust neighbours (Swales and Tipping, 2017). Low scores on the SFI occur in family-oriented areas, with relatively low residential turnover.

The ethnic structure of area populations is also relevant to variations in diabetes risk. Raised diabetes risk among south Asian groups has been attributed to increased insulin resistance, even adjusting for adiposity (McKeigue et al, 1991; Barnett et al, 2006). Nishino et al (2015) show area differences in emergency diabetes admissions to be linked to south Asian ethnicity, and non-UK studies also report higher diabetes prevalence among South Asians (Middelkoop et al, 1999). Here we measure the potential impact of area ethnic mix by the proportion of the MSOA population with south Asian ethnicity (from the 2011 UK Census).

A further source of geographic variability in diabetes outcomes is urban status. UK and international studies have shown lower T2D rates in areas with more greenspace (Bodicoat et al, 2014; Astell-Burt et al, 2014), while Cox et al (2007) report higher diabetes incidence in urban areas in Scotland. Links between air pollution (typically elevated in urban areas) and diabetes have also been proposed (Rajagopalan and Brook, 2012). As a measure of urban-rural status, we use a ridit score (Ernstsen et al, 2012) based on the UK 2011 Census rural-urban classification (Bibby and Brindley, 2013). This is an ordered 8-fold classification of MSOAs from most to least urban, with the extremes being “urban major conurbation” and “rural village and

dispersed in a sparse setting". The score is defined to be highest for urban areas, lowest for rural areas.

Further measures of the environment are provided by levels of physical inactivity and of overweight/obesity for adults aged 18 and over (these are percentage data for 2015/16 and at local authority level, available at https://fingertips.phe.org.uk/profile/physical-activity and https://fingertips.phe.org.uk/search/overweight). There is considerable research showing links between inactivity, excess weight and diabetes risk at both individual and ecological levels (Scarborough et al, 2011; Allender et al, 2007; Geiss et al, 2017). These factors have been adduced to explain increasing T2D levels among younger adults in the UK (e.g. Wilmot et al, 2010).

In summary, there are six area-risk factors included in this analysis, namely a measure of area socio-economic status; social fragmentation; the proportion of MSOA populations of south Asian ethnicity; an index of urbanity; a measure of adult physical inactivity; and a measure of adult overweight.

2.2 Sources of Data on Diabetes Outcomes

The two forms of diabetes outcome data are for deaths and diagnosed prevalence. Diabetes deaths for MSOAs in England during 2013-17 cover ICD-10 codes E10-E14, and hence encompass both type 1 and type 2 diabetes. There are 25868 diabetes deaths during 2013-17. The geographic scope provided by the 6791 MSOAs is the entirety of England, one of the constituent countries of the United Kingdom. The data were obtained by request from the Office of National Statistics and are available at: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/adhocs/009524deathsfromdiabetesbysexandmiddlelayersuperoutputareamsoaengland2013to2017

Diabetes prevalence data for these MSOAS (with both diabetes types included) are for ages 17 and over and are for 2017/18, and obtained from the House of Commons Dashboard for chronic diseases included in the Quality Outcomes Framework (House of Commons, 2019). The Quality and Outcomes Framework (QOF) is an annual reward and incentive programme for all GP surgeries in England, which includes maintenance of registers of

diagnosed chronic diseases such as diabetes. The prevalence data can be obtained at: https://commonslibrary.parliament.uk/social-policy/health/constituency-data-how-healthy-is-your-area/.

Total diabetes cases for each MSOA are obtained by multiplying 2017 MSOA adult populations (ages 17 and over) by the crude diabetes prevalence rates for each MSOA (i.e. unadjusted for age structure), which are provided by the Dashboard. We then adjust for differing age structures between MSOAs by including an offset for expected deaths and prevalent cases in the Poisson statistical model (see next section and Appendix 1).

Caveats regarding the health outcome data may be mentioned. The first is that both types of data (deaths and prevalence) include both diabetes type 1 and 2. Type 1 diabetes (T1D, with disease code ICD E10) usually has child or adolescent onset, and accounts for around 10% of prevalent cases among adults (Diabetes UK, 2010). However, the QOF data is not available by diabetes type, and deaths data by type would be sparse (overly thinly scattered) at MSOA level for T1D. Data from NOMIS (https://www.nomisweb.co.uk) show that 6.2% of diabetes deaths in England over 2013-17 were from T1D, an average of 0.24 per MSOA.

There are relatively few studies explaining differences in T1D risk specifically, though regarding area socio-economic status, a meta-analysis by Scott et al (2017) found that “low socioeconomic status is associated with higher levels of mortality and morbidity for adults with type 1 diabetes”, while Collier et al (2015) and Govan et al (2012) report that complication rates among diabetes type 1 patients (and hence mortality risk) increase with area deprivation. Overweight may be relevant to development of T1D, as it is for T2D (Islam et al, 2014).

Other caveats may be mentioned: the diabetes prevalence data omits undiagnosed cases, estimated at just under 1 million in the UK (Diabetes UK, 2020). Also diabetes deaths may be undercounted: Stokes and Preston (2017), in a US study, report that the proportion of deaths with diabetes assigned as the underlying cause of death (3.3–3.7%) understates the contribution of diabetes to US mortality.

2.3 Statistical Methods: Roles of Different Types of Indicator in the Model

A Bayesian estimation strategy is adopted, with estimation using Markov chain Monte Carlo (MCMC) techniques in the WINBUGS package (Lunn et al, 2009). Estimates are based on the last 15000 iterations of two chain runs of 20000 iterations with convergence assessed using Brooks-Gelman-Rubin diagnostics (Brooks and Gelman, 1998). We adopt regression techniques appropriate to count data (death and prevalence totals), allowing also for spatial dependence in the latent factor (see Appendix 1). Uncertainty about the factor scores (e.g. in terms of 95% credible intervals, or posterior density plots) is available as part of the estimation procedure, by analysing MCMC samples of the individual factor scores (Marí-Dell'Olmo et al, 2011).

As mentioned above, the factor analysis model includes both morbidity indicators as responses (denoted Y_1 and Y_2), and risk variables (X_1 to X_6) as predictors. In the terminology of factor analysis, the two response variables (mortality and prevalence) are reflective indicators, while the area risk factors (area SES, social fragmentation, obesity, urbanity, etc.) are formative indicators (Coltman et al, 2008; Diamantopoulos et al, 2008). The usual type of factor analysis, available in statistical packages, such as SPSS and Stata, involves reflective indicators only. Principal component analysis to develop socio-economic needs scores (e.g. Pampalon et al, 2009) is similar in intent to using factor analysis with reflective indicators only.

As noted by Coltman (2008): “most scholars assume that this relationship between construct and indicator is reflective...with reflective (or effect) measurement models, causality flows from the latent construct to the indicator[s]”. However, not all observed indicators can be so regarded. As an example, consider a latent construct “life stress”; relevant reflective indicators might be poor sleep, or use of tranquilizer medication, while formative indicators might be marital problems, or job insecurity.

In a model with reflective indicators only, the scores on the latent construct are determined by these indicators only (these indicators are the “components” of the construct); Wang and Wall (2003) provide a model with reflective indicators only. In a model with formative indicators only, these indicators determine the values of latent construct. In a model with both reflective and formative indicators, all the indicators (here, these are morbidity outcomes and area risk factors respectively) influence scores on the latent construct.

In the present application we can envisage the latent construct, a summary measure of diabetes risk, as an intermediate variable between the causes of varying risk (formative variables, here area or contextual risk factors which influence diabetes risk) and morbidity indicators (reflective measures). The model can be represented (Fleischer and Roux (2008) as a directed acyclic graph (DAG); see Figure 2, with observed data in squares, and latent data in circles.

This Figure first shows the two-way feedback between areas on the latent construct F, via the model assumed for spatial dependence between them (see Appendix 1). Thus the factor score for the i^{th} MSOA is based on average factor scores in the locality of the i^{th} area (the set of areas surrounding that area and adjacent to it), denoted by $j \neq i$.

Figure 2 also shows λ coefficients linking the latent construct F to relative risks for the two morbidity outcomes, R_1 and R_2, and coefficients β linking the area risk variables X_1 to X_6 to the latent construct. (The u_1 and u_2 terms represent Poisson heterogeneity; see Appendix 1). Relative risks are unknown measures of disease risk with national average 1 (Richardson et al, 2004).

Assume F is constrained (as in this study) to be a positive measure of risk: such that higher F scores imply higher R_1 and R_2, and hence higher Y_1 and Y_2. Then all the observed variables (whether reflective or formative) impact on the values of F, and hence are "components" of it. If the risk variables are also positive measures of risk, and the β coefficients are positive, then higher X values necessarily translate into higher F scores (and thereby imply higher Y values). Similarly high Y values feed back (since this raises the model likelihood) to raise the F scores via the λ coefficients, assuming these coefficients are positive.

We compare two basic modelling approaches. The full model contains both reflective and formative indicators, as in Figure 2. Another analysis involves a model with the two reflective indicators only, namely diabetes mortality and prevalence, dispensing with area risk factors; Wang and Wall (2003) provide a model with reflective indicators only. These approaches are compared in terms of fit using the Deviance Information Criterion, or DIC measure, of Spiegelhalter et al (2002).

3. Results

3.1 Regression Parameters

Table 1 shows the regression coefficients (the β coefficients in Figure 2) for the impacts of the six area variables on the diabetes risk score. The Table shows posterior means and standard deviations of the coefficients together with 95% credible intervals (analogous to 95% confidence intervals). It can be seen that all the area variables are all positive risk factors in the sense that higher values on them are associated with higher mortality and prevalence. The impacts are all significant in that their 95% credible intervals are positive. The most important risk factors are the area occupational status measure, proportions South Asian, and social fragmentation. Less important are urbanity, physical inactivity levels, and overweight levels, but they are still significant influences. So the results in Table 1 show that all the risk variables used in the procedure play a significant role: none are redundant.

3.2 Contrasts in Risk Scores and Modelled Outcomes

To show broad geographical contrasts in average risk and morbidity outcomes, Table 2 presents average risk scores and average mortality and prevalence rates (MSOA modelled relative risks times the national mortality and prevalence rates) disaggregated by standard region (Figure 1) and urban-rural category. The latter is collapsed to 5 categories[1].

Average risk scores, mortality and prevalence are highest in the 338 MSOAs classified as “urban major conurbation” in the West Midlands, and lowest in the most rural category (village and dispersed) in the South East region. Contrasts in modelled mortality and prevalence approach or exceed two-fold (86% for prevalence, 104% for mortality) when comparing the most contrasting region-urban categorisations.

As one way of validating the model, we show how differences in risk scores translate into gradients in both morbidity and in area risk factors (area occupational status, social fragmentation, etc.). So we seek to show that method does what it purports to do: identify extreme environments in terms of diabetic risk, namely small areas where morbidity is high, and risk factors are elevated, and small areas where morbidity is low, and risk factors are also low.

MSOAs are accordingly grouped into decile categories on the risk score (Table 3). MSOAs with lowest risk (decile 1) have low mortality and prevalence, low scores on the ICE-ASG index, fragmentation and percent south Asian ethnicity, and low urbanity scores. By contrast, MSOAs with highest risk (decile 10) have considerably elevated mortality and prevalence, high scores on the ICE-ASG and fragmentation, notably high percentages south Asian, and are urban on average (high urbanity scores). Relative mortality and prevalence ratios approach 2.5-fold when comparing the highest and lowest risk deciles.

Despite the urban bias for high diabetes risk, there are considerable risk variations within major urban centres To illustrate this, Figure 3 shows wide variability in the scores in Birmingham MSOAs (Birmingham is a major conurbation in the West Midlands), but also clustering of high scores, an issue pursued below.

3.3 Model Comparison

Subsidiary analysis involves a comparison with a model with reflective indicators only (diabetes mortality and prevalence) without area risk factors. The DIC of this model is 95106, as compared to 94153 for the model with both reflective and formative indicators. Hence there is a clear gain in using contextual risk predictors of diabetes morbidity. According to the criteria discussed by Spiegelhalter et al (2002, page 613) the reflective-formative model has “considerably more support”. Posterior predictive checks on the best fitting model, using the Poisson deviance, are also satisfactory (Gelman et al, 1996).

3.4 Validation Against Other Diabetes Outcome Indicators

Another form of model validation involves evaluation of the predictive utility of the diabetes risk measure for a diabetes outcome not included in the model. However, this may be problematic to some degree as there is no “gold standard” measure of diabetes risk, and available measures may reflect care variations as well as variations in true diabetes need (Booth and Hux, 2003).

MSOA data on diabetes incidence (the other main epidemiological measure apart from mortality and prevalence) is not available. Here we consider hospital admissions for diabetic complications for 207 Clinical

Commissioning Groups (CCGs) in England in 2016/17, and CCG data on diabetes prescribing per head of adult population (net ingredient cost) for 2017/18.

Clinical variations in these outcomes may reflect partly levels of diabetes but also quality of care variations. For example, National Health Service (2011) refers to “unwarranted variation in diabetes care” and admissions for diabetes complications may be preventable to some extent (Booth and Hux, 2003). Similarly House of Commons (2013) mention that “there will always be some variation in performance because some populations have more people at risk of diabetes than others but, [it is] recognised that variation is mostly driven by differences in how primary care trusts deliver diabetes care and in clinical practice between healthcare professionals”.

We undertake linear regressions (207 observations) of the two CCG outcomes on average CCG risk factor scores, and on region (as a variable with 9 categories), to account for geographic performance variation beyond risk. For complication rates (which are indirectly standardised ratios, with average 100), the R-squared is 59%. There is a significant positive slope on the risk score, 5.54 (s.e. 1.01), which implies a range in predicted complication rates from 87.5 to 118.2. For prescribing, the R-squared is 58% with a significant positive slope on the risk score, 11.7 (s.e. 2.1), implying a predicted range in NIC per head from 176.5 to 241.4. So the risk score has utility in predicting two important activity indicators for diabetes.

3.5 High Risk Clusters

As mentioned above, the risk scores may be applied in detecting high and low risk spatial clusters, with high risk clusters (adjacent MSOAs all with high diabetes risk scores), which is of potential importance regarding potential intervention. A LISA cluster map is available in the GEODA package (Anselin et al, 2006), classifying each MSOA by comparing risk scores in that MSOA and average risk scores in contiguous MSOAs; see Figure 4, based on Moran LISAs (Martins-Melo et al, 2012). There are 1011 high-high MSOAs and 1478 low-low MSOAs in Figure 4.

Remaining categories in Figure 4 can be interpreted as spatial outliers, namely high-low (high risk MSOAs surrounded by low risk), and low-high (low risk areas surrounded by high). These categories only account for 25 MSOAs combined. A residual class entitled “not significant” in Figure 4,

includes MSOAs in which contiguous MSOAs have intermediate (non-significant) risk scores.

High risk clusters are generally located in the most urbanised settings (MSOAs in urban major conurbations), namely 775 of 1011. Whereas 15% of all MSOAs are classed as high-high, proportions are much higher in metropolitan local authorities such as Birmingham (64% of its 132 MSOAs being high-high), Sandwell (with all 38 MSOAs being high-high) and Newham (all 37 of its MSOAs being high-high). By contrast, low-low clusters tend to be located in mainly less urbanised lower density settings, with MSOAs of larger geographic extent, so tending to dominate the map in Figure 4. Of the 1478 MSOAs in low-low clusters, 38% are in the two most rural settings ("Rural Town", "Rural Village and Dispersed"), whereas these MSOAs only account for 18% of all 6791 English MSOAs.

However, there are some low-low clusters in more urbanised settings, such as London, reflecting impacts of intra-urban variations in deprivation (see Figure 5). The cluster of high risk in north east London reflects locations of both large south Asian ethnic populations, and high deprivation (as in London boroughs such as Tower Hamlets and Newham). By contrast, affluent central boroughs (Kensington & Chelsea), and outer suburban boroughs (Richmond, Kingston, Bromley), have concentrations of low-low clustering.

4.Discussion

There is a lack of UK based studies investigating the spatial patterning of diabetes, the extent of spatial clustering in this patterning, and the association of geographic diabetes variations with area sociodemographic characteristics. Research on small area diabetes contrasts in the UK is generally regional in focus. In particular, there are no UK based studies with a national focus (England being a constituent nation of the UK) aiming to produce a small area index of diabetes risk or health care need, as does the present study.

The index here has the benefit of (a) incorporating evidence of risk both from diabetes outcomes and area risk factors, such as area socio-economic status, and of (b) including spatial correlation in its construction. It also has the benefit of extensive geographic coverage, namely all English MSOAs. This type of framework, with both reflective disease measures, and formative

socio-economic contextual variables, could be adapted to using dynamic (change) measures of area ill-health and context (Norman, 2010).

Among limitations of the present study are its ecological nature, so that inferences about causal impacts on area diabetes risk cannot be made, and impacts of area characteristics may reflect compositional as well as contextual effects. However, there is increasing evidence that neighbourhood factors per se may contribute to higher diabetes outcomes, for example, through limited access to healthy foods and physical activity, via housing market factors affecting ethnic concentration (Phillips, 1998), or via neighbourhood social environments (Gebreab et al, 2017; Smalls et al, 2015). The present study confirms positive impacts of physical inactivity and overweight on diabetes risk, and also finds a positive impact of an index of social fragmentation on diabetes risk, a link not considered in existing UK studies of diabetes geographic variation.

The present study also confirms existing UK-based findings that area ethnic mix significantly affects area diabetes variations (Nishino et al, 2015). However, the study provides new evidence on the role of urban-rural status in diabetes spatial patterns in England. Urbanity effects remain after allowing for area socio-economic status, ethnicity, and social fragmentation. Further research is indicated on what urbanity impacts on diabetes might stem from, for example, built environment and air quality factors (Stewart et al, 2011; Rajagopalan and Brook, 2012).

The present study highlights concentrations of high diabetes risk, and spatial clustering of high risk, in more urbanised settings, with the urbanity effect strongest in particular English regions (e.g. the West Midlands). However, other dimensions of risk such as that of area socio-economic status, intersect the urban effect, as illustrated by clusters of diabetes risk in London. Detection of high risk and of high risk clusters are of particular importance for effective targeting of health resources and public health interventions to tackle diabetes and other chronic conditions. The present study describes a method applicable across chronic conditions for developing small area risk indices to guide such interventions.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Declarations of interest: none

Notes

1. These 5 categories collapse over the three separate “sparse settings” categories (Bibby and Brindley, 2013), which together account for only 78 of 6791 England MSOAs. The collapsed categories are 1 Major Conurbation, 2 Minor Conurbation, 3 City and Town, 4 Rural Town and Fringe, 5 Rural Village and Dispersed.

Appendix 1 Formal Description of Methods

For analysis of the two reflective outcomes (diabetes deaths and prevalence), we use total events as the basis of the regression modeling. We assume Poisson distributed outcomes, $Y_{ij} \sim \text{Poisson}(E_{ij}R_{ij})$ for areas i=1,..,N, and outcomes j=1,..,J, where J=2, and j=1 for deaths, j=2 for prevalence. There are N=6791 areas. The E_{ij} denote expected event totals, obtained by applying national age specific outcome rates to MSOA populations. The R_{ij} denote relative risks of mortality and prevalence, with national averages of 1 (Richardson et al, 2004).

These two outcomes are taken as reflective indicators of a single latent risk factor F, so that log-link regressions for the relative risks have the form

$$\log(R_{ij})=\alpha_j+\lambda_j F_i+u_{ij},$$

where α_j are intercepts, and λ_j are loadings of $\log(R_{ij})$ on the score for common latent risk score F_i.The u_{ij} are unstructured residuals (i.e. with no spatial structure), $u_{ij} \sim N(0,\omega_j^2)$, which may be needed to account for spatially unstructured Poisson over-dispersion. By contrast, we assume the F_i are spatially structured, as set out below.

As mentioned in Section 2.3, we allow area risk factors X_{i1} to X_{i6} (the formative indicators) to impact on the diabetes risk factor, with $\beta=(\beta_1,...,\beta_6)$ the corresponding regression parameters. The risk factors are ICE-ASG, social fragmentation, south Asian ethnicity, urbanity, physical inactivity, and obesity. All risk factors are expressed in (0,1) form (0 for the minimum risk factor value, 1 for the maximum) so that their relative importance can be assessed by simply comparing coefficient sizes. Let $\eta_i= \beta_1X_{i1}+ \beta_2X_{i2}+... +\beta_6X_{i6}$ denote their combined impact on F_i, the factor score for the i^{th} MSOA (see Figure 2). The β_j are assigned N(0,100) priors (prior densities).

To represent spatial dependence we assume factor scores in area i depend on nearby areas, according to the scheme of Besag et al (1991). First denote $e_i=F_i-\eta_i$, and let w_{ij} denote spatial adjacency indicators ($w_{ij}=1$ if MSOAs i and j are adjacent, $w_{ij}=0$ otherwise), and let $m_i=\sum_{i\neq j}w_{ij}$ denote the number of MSOAs adjacent to the i^{th} MSOA. Then, as in Stern and Cressie (2000, eqn 5), the conditional means for F_i are

$$E(F_i|F_{-i})= \eta_i+\sum_{i\neq j}w_{ij}e_j/m_i,$$

with conditional variances τ^2/m_i, where F_{-i} denotes all MSOA risk scores apart from the score for MSOA i.

As in other latent factor applications, identifying restrictions are needed, either on the variance of the latent factor, or on one of the loadings λ_j. Here the variance parameter τ^2 is set to one (i.e. factor standardization), and so both loadings are unknowns (Skrondal and Rabe-Hesketh, 2007). To ensure unique labelling of the common factor F_i as a positive measure of risk (i.e. ensure that higher scores F_i denote higher diabetes risk), the loadings λ_j are assigned gamma $\Gamma(1,0.1)$ priors. The intercept parameters α_j are assigned N(0,100) priors, and variances ω_j^2 are assigned $\Gamma(1,0.1)$ priors.

The BUGS code (Lunn et al, 2009) for the model (essential elements only) is as follows:

```
model {
for (i in 1:N) {# mortality and prevalence counts, Y1 and Y2, in N=6791 areas
Y1[i] ~ dpois(mu[i,1])
Y2[i] ~ dpois(mu[i,2])
# expected events
E[i,1] <- E1[i]; E[i,2] <- E2[i]
# Relative Risk Models for J=2 outcomes
for (j in 1:2) {mu[i,j] <- E[i,j]*R[i,j]
log(R[i,j]) <- alpha[j]+lambda[j]*F[i]+u[i,j]
# iid effects for Poisson overdispersion
u[i,j] ~ dnorm(0,inv.omega2[j])}}
# Priors
for (j in 1:2) {lambda[j] ~ dgamma(1,0.1)}
# variance=100, so precision=0.01
for (j in 1:2) {alpha[j] ~ dnorm(0,0.01)}
for (j in 1:6)  {beta[j] ~ dnorm(0,0.01)}
for (j in 1:2) {inv.omega2[j]~ dgamma(1,0.1)}
```

```
# Risk Index Model
tau.F <- 1
# error vector over areas adjacent to area i (in vector adj), NN=total adjacencies (sum of num[i])
for (i in 1:NN) { We[i] <- e[adj[i]] }
for (i in 1:N) {F[i]  ~ dnorm(F.nei[i],tauF[i])
# prediction of risk
eta[i] <- beta[1]*ice_asg[i]+beta[2]*sas[i]+beta[3]*urb[i]
+beta[4]*frg[i]+beta[5]*inact[i]+beta[6]*ovrwgt[i]
e[i] <- F[i]-eta[i]
# num[i] is number of neighbours of area i
tauF[i]   <-  tau.F*num[i]
# vector cum.num[], of length NN+1, contains cumulative num[i], with elements 0,num1,num1+num2, etc.
F.nei[i] <- eta[i]+sum(We[cum.num[i]+1:cum.num[i+1] ])/num[i]}}}
```

To obtain posterior summaries of the needs scores (and their uncertainty) one monitors the vector F[1:6791].

The vectors adj and num can be obtained in R using functions from spdep and maptools. One first reads in a shapefile using the maptools command:

```
shape = readShapePoly(("file.shp")
```

then commands spatial=poly2nb(shape) and WB=nb2WB(spatial) create a list with adj and num vectors WB$adj and WB$num.

References

Allender, S., Foster, C., Scarborough, P.,Rayner, M., 2007. The burden of physical activity-related ill health in the UK. Journal of Epidemiology and Community Health, 61(4), pp.344-348.

Allik, M., Leyland, A., Ichihara, M, Dundas, R., 2020. Creating small-area deprivation indices: a guide for stages and options. J Epidemiol Community Health, 74(1), pp.20-25

Anselin L, Syabri I, Kho Y, 2006. GeoDa: an introduction to spatial data analysis. Geographical Analysis, 38(1),5–22.

Astell-Burt, T., Feng, Kolt, G, 2014. Is neighborhood green space associated with a lower risk of type 2 diabetes? Evidence from 267,072 Australians. Diabetes care, 37(1), pp.197-201.

Barnett, A., Dixon, A., Bellary, S., Hanif, M., O'Hare, J., Raymond, N., Kumar, S., 2006. Type 2 diabetes and cardiovascular risk in the UK south Asian community. Diabetologia, 49(10), pp.2234-2246.
Besag, J, York, J, Mollie, A. ,1991. Bayesian image restoration, with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics,43(1), 1-21.
Bibby, P, Brindley, P, 2013. The 2011 Rural-Urban Classification for Small Area Geographies: A User Guide; UK Government Statistical Service: London, UK
Bodicoat, D, O'Donovan, G., Dalton, A, Gray, L., Yates, T., Edwardson, C., Hill, S., Webb, D, Khunti, K., Davies, M, Jones, A, 2014. The association between neighbourhood greenspace and type 2 diabetes in a large cross-sectional study. BMJ open, 4(12), p.e006076.
Booth, G, Hux, J, 2003. Relationship between avoidable hospitalizations for diabetes mellitus and income level. Archives of Internal Medicine, 163(1), pp.101-106.
Brooks S, Gelman A ,1998. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7(4), 434-455.
Brown, H., Abdallah, S, Townsley, R., 2017. Understanding local needs for wellbeing data: measures and indicators. What Works Well Centre for Well-Being. https://whatworkswellbeing.org/resources/understanding-local-needs-for-wellbeing-data/
Collier, A., Ghosh, S., Hair, M, Waugh, N., 2015. Impact of socioeconomic status and gender on glycaemic control, cardiovascular risk factors and diabetes complications in type 1 and 2 diabetes: a population based analysis from a Scottish region. Diabetes & metabolism, 41(2), pp.145-151.
Coltman, T., Devinney, T, Midgley, D. Venaik, S., 2008. Formative versus reflective measurement models: Two applications of formative measurement. Journal of Business Research, 61(12), pp.1250-1262.
Congdon, P ,1996. Suicide and parasuicide in London: a small-area study. Urban Studies 33(1), 137-158
Connolly V, Unwin N, Sherriff P, Bilous R, Kelly, W, 2000. Diabetes prevalence and socioeconomic status: a population based study showing increased prevalence of type 2 diabetes mellitus in deprived areas. J of Epid & Community Health, 54(3), 173-177
Cox, M., Boyle, P, Davey, P, Feng, Z, Morris, A, 2007. Locality deprivation and Type 2 diabetes incidence: a local test of relative inequalities. Social Science and Medicine, 65(9), pp.1953-1964.

Diabetes UK, 2010. Diabetes in the UK 2009: Key statistics on diabetes. Diabetes UK, London
Diabetes UK, 2020, Diabetes Prevalence 2018. https://www.diabetes.org.uk/professionals/position-statements-reports/statistics/diabetes-prevalence-2018. Accessed 11-03-2020
Diamantopoulos, A., Riefler, P, Roth, K, 2008. Advancing formative measurement models. Journal of Business Research, 61(12), pp.1203-1218.
Earnest A, Ong M, Shahidah N, Chan A, Wa, W, Thumboo J., 2015. Derivation of indices of socioeconomic status for health services research in Asia. Preventive Medicine Reports, 2, 326-332.
Ernstsen, L.; Strand, B.; Nilsen, S; Espnes, G.; Krokstad, S. Trends in absolute and relative educational inequalities in four modifiable ischaemic heart disease risk factors. BMC Public Health 2012, 12(1), 266.
Evans, J, Newton, R, Ruta, D, MacDonald, T, Morris, A, 2000. Socio-economic status, obesity and prevalence of Type 1 and Type 2 diabetes mellitus. Diabetic Medicine, 17(6), pp.478-480
Fleetcroft, R., Asaria, M., Ali, S, Cookson, R., 2017. Outcomes and inequalities in diabetes from 2004/2005 to 2011/2012: English longitudinal study. Br J Gen Pract, 67(654), pp.e1-e9.
Fleischer, N, Roux, A. 2008. Using directed acyclic graphs to guide analyses of neighbourhood health effects: an introduction. Journal of Epidemiology & Community Health, 62(9), pp.842-846.
Gabert, R, Thomson, B, Gakidou, E, Roth, G. ,2016. Identifying high-risk neighborhoods using electronic medical records: a population-based approach for targeting diabetes prevention and treatment interventions. PLoS One, 11(7) e0159227.
Gariepy, G., Smith, K.J. and Schmitz, N., 2013. Diabetes distress and neighborhood characteristics in people with type 2 diabetes. Journal of Psychosomatic Research, 75(2), pp.147-152.
Gebreab, S, Hickson, D, Sims, M., Wyatt, S, Davis, S, Correa, A.,Diez-Roux, A, 2017. Neighborhood social and physical environments and type 2 diabetes mellitus in African Americans: The Jackson Heart Study. Health & Place, 43, 128-137.
Geiss, L, Kirtland, K., Lin, J., Shrestha, S., Thompson, T., Albright, A, Gregg, E., 2017. Changes in diagnosed diabetes, obesity, and physical inactivity prevalence in US counties, 2004-2012. PLoS One, 12(3).
Gelman, A., Meng X-L, Stern H, 1996. Posterior predictive assessment of model fitness via realized discrepancies. Statistica Sinica, 6(4), 733–807.

Glover G, Arts G, Wooff D, 2004. A needs index for mental health care in England based on updatable data. Social Psychiatry and Psychiatric Epidemiology, 39(9), 730-8.
Govan, L., Maietti, E., Torsney, B., Wu, O., Briggs, A., Colhoun, H, Fischbacher, C, Leese, G, McKnight, J, Morris, A, Sattar, N., 2012. The effect of deprivation and HbA 1c on admission to hospital for diabetic ketoacidosis in type 1 diabetes. Diabetologia, 55(9), pp.2356-2360.
Green, C., Hoppa, R., Young, T, Blanchard, J, 2003. Geographic analysis of diabetes prevalence in an urban area. Social Science & Medicine, 57(3), pp.551-560.
House of Commons, 2013. The management of adult diabetes services in the NHS. Committee of Public Accounts, London: House of Commons
House of Commons Library, 2019. Constituency Data: How Healthy is your Area? https://commonslibrary.parliament.uk/social-policy/health/diseases/constituency-data-how-healthy-is-your-area/. Accessed 11-06-2019.
Hsu, C, Lee, C, Wahlqvist, M, Huang, H., Chang, H, Chen, L., Shih, S., Shin, S., Tsai, W, Chen, T, Huang, C, 2012. Poverty increases type 2 diabetes incidence and inequality of care despite universal health coverage. Diabetes Care, 35(11), pp.2286-2292.
International Diabetes Federation, 2020. IDF Diabetes Atlas, 9th edn. Brussels, Belgium: International Diabetes Federation. http://www.diabetesatlas.org Accessed 18-03-2020
Islam, S, Srinivasan, S, Craig, M, 2014. Environmental determinants of type 1 diabetes: a role for overweight and insulin resistance. Journal of Paediatrics and Child Health, 50(11), pp.874-879.
Krieger, N., Kim, R., Feldman, J, Waterman, P, 2018. Using the Index of Concentration at the Extremes at multiple geographical levels to monitor health inequities in an era of growing spatial social polarization: Massachusetts, USA (2010–14). International Journal of Epidemiology, 47(3), pp.788-819.
Lunn, D., Spiegelhalter, D., Thomas, A, Best, N., 2009. The BUGS project: evolution, critique and future directions. Statistics in Medicine 28 (25), 3049--3082.
Matthews, L, Kirk, A, MacMillan, F, Mutrie, N.,2013. Can physical activity interventions for adults with type 2 diabetes be translated into practice settings? A systematic review using the RE-AIM framework. Translational Behavioral Medicine, 4 (1) 60-78

McKeigue P, Shah B, Marmot M, 1991. Relation of central obesity and insulin resistance with high diabetes prevalence and cardiovascular risk in South Asians. Lancet 337, 382–386.
Malla, G., Long, D, Black, N, Zhu, S., Uddin, J., Cherrington, A., Dutton, G., Safford, M, Cummings, D., Judd, S, Levitan, E, 2020. Abstract P447: Index of Concentration at the Extremes and Diabetes Prevalence: The Reasons for Geographic and Racial Differences in Stroke (REGARDS) Study. Circulation, 141(Suppl_1), pp.AP447-AP447.
Marí-Dell'Olmo, M., Martínez-Beneito, M, Borrell, C., Zurriaga, O., Nolasco, A, Domínguez-Berjón, M, 2011. Bayesian factor analysis to calculate a deprivation index and its uncertainty. Epidemiology, pp.356-364.
Martins-Melo F, Ramos A, Alencar C, Lange W, Heukelbach, J. ,2012. Mortality of Chagas' disease in Brazil: spatial patterns and definition of high-risk areas. Tropical Medicine & International Health, 17(9), 1066-1075
Matthews, L, Kirk, A, MacMillan, F (2013). Can physical activity interventions for adults with type 2 diabetes be translated into practice settings? A systematic review using the RE-AIM framework. Translational Behav Med, 4, 60-78
Middelkoop, B., Kesarlal-Sadhoeram, S, Ramsaransing, G, Struben, H, 1999. Diabetes mellitus among South Asian inhabitants of The Hague: high prevalence and an age-specific socioeconomic gradient. International Journal of Epidemiology, 28(6), pp.1119-1123.
National Health Service (NHS) (2011) NHS Atlas of Variation in Healthcare for People with Diabetes. NHS, London.
Nishino, Y., Gilmour, S., Shibuya, K, 2015. Inequality in diabetes-related hospital admissions in England by socioeconomic deprivation and ethnicity: facility-based cross-sectional analysis. PloS one, 10 (2), e0116689.
Norman, P., 2010. Identifying change over time in small area socio-economic deprivation. Applied Spatial Analysis and Policy, 3(2-3), pp.107-138.
Office of National Statistics (2020) Census Geography: An Overview of the Various Geographies used in the Production of Statistics Collected via the UK Census (accessed 10-03-2020).
https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography
Pabayo, R., Molnar, B, Cradock, A, Kawachi, I. ,2014. The relationship between neighborhood socioeconomic characteristics and physical inactivity among adolescents living in Boston, Massachusetts. American Journal of Public Health, 104 (11), e142-e149.

Pampalon, R., Hamel, D., Gamache, P, Raymond, G., 2009. A deprivation index for health planning in Canada. Chronic Dis Can, 29(4), pp.178-91.
Phillips, D., 1998. Black minority ethnic concentration, segregation and dispersal in Britain. Urban Studies, 35(10), 1681-1702.
Poulsen K, Cleal B, Willaing I, 2014. Diabetes and work: 12-year national follow-up study of the association of diabetes incidence with socioeconomic group, age, gender and country of origin. Scand J Public Health, 42:728–33
Proitsi, P, Hamilton, G., Tsolaki, M., Lupton, M., Daniilidou, M., Hollingworth, P, Todd, S. ,2011.. A multiple indicators multiple causes ,MIMIC model of behavioural and psychological symptoms in dementia ,BPSD. Neurobiology of Aging, 32 (3), 434-442
Rajagopalan, S, Brook, R ,2012. Air pollution and type 2 diabetes: mechanistic insights. Diabetes, 61 (12) 3037-3045.
Scarborough, P., Bhatnagar, P., Wickramasinghe, K, Allender, S., Foster, C. Rayner, M., 2011. The economic burden of ill health due to diet, physical inactivity, smoking, alcohol and obesity in the UK: an update to 2006–07 NHS costs. Journal of public health, 33(4), pp.527-535
Richardson, S., Thomson, A., Best, N, Elliott, P., 2004. Interpreting posterior relative risk estimates in disease-mapping studies. Environmental Health Perspectives, 112(9), pp.1016-1025.
Scott, A., Chambers, D., Goyder, E, O'Cathain, A., 2017. Socioeconomic inequalities in mortality, morbidity and diabetes management for adults with type 1 diabetes: A systematic review. PloS one, 12(5).
Skrondal A, Rabe-Hesketh S, 2007. Latent variable modelling: a survey. Scandinavian Journal of Statistics. 34 (4),712-45
Smalls, B, Gregory, C, Zoller, J, Egede, L, 2015. Assessing the relationship between neighborhood factors and diabetes related health outcomes and self-care behaviors. BMC Health Services Res, 15(1) 445.
Smith, T, Noble, M, Noble, S, Wright, G, McLennan, D, Plunkett, E, 2015. The English Indices of Deprivation 2015; Technical Report. London: Department for Communities and Local Government.
Spiegelhalter, D, Best, N, Carlin, B, van der Linde, A, 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society, Series B. 64(4), 583–639
Stern, H, Cressie, N. ,2000. Posterior predictive model checks for disease mapping models. Statistics in Medicine, 19 (17-18), 2377-2397.
Stewart, J, Battersby, S, Lopez-De Fede, A., Remington, K, Hardin, J., Mayfield-Smith K, 2011.. Diabetes and the socioeconomic and built environment: geovisualization of disease prevalence and potential

contextual associations using ring maps. International journal of health geographics, 10 (1), 18.
Stokes, A, Preston, S, 2017. Deaths attributable to diabetes in the United States: comparison of data sources and estimation approaches. PloS one, 12(1).
Sundquist, K., Malmström, M., Johansson, S, Sundquist, J. (2003). Care Need Index, a useful tool for the distribution of primary health care resources. Journal of Epidemiology & Community Health, 57(5), 347-352.
Swales K, Tipping S (2018) Fragmented Communities? The role of cohesion, community involvement and social mixing. NatCen Social Research, London
Toms, R., Bonney, A, Mayne, D, Feng, X, Walsan, R., 2019. Geographic and area-level socioeconomic variation in cardiometabolic risk factor distribution: a systematic review of the literature. International Journal of Health Geographics, 18(1), 1.
Tosetti E, Santos R, Moscone F, Arbia G, 2018. The Spatial Dimension of Health Systems, in Oxford Research Encyclopedia of Economics and Finance. OUP, New York.
Vazquez, L, Calvo-Bonacho, E., Reviriego, J., García-Margallo, T., Caveda, E, Goday, A., 2019. Incidence of Diabetes in the Working Population in Spain: Results from the ICARIA Cohort. Diabetes Therapy, 10(1), pp.57-69.
Walker, J, Livingstone, S, Colhoun, H, Lindsay, R, McKnight, J., Morris, A., Petrie, J., Philip, S., Sattar, N., Wild, S, 2011. Effect of socioeconomic status on mortality among people with type 2 diabetes: a study from the Scottish Diabetes Research Network Epidemiology Group. Diabetes Care, 34(5), pp.1127-1132.
Walker, R., Williams, J, Egede, L, 2016. Influence of race, ethnicity and social determinants of health on diabetes outcomes. The American Journal of the Medical Sciences, 351(4), pp.366-373.
Wang, F., Wall, M., 2003. Generalized common spatial factor model. Biostatistics, 4(4), pp.569-582.
Wang, J., Wang, X., 2019. Structural Equation Modeling: Applications using Mplus, 2nd Edition. Wiley.
Wilmot, E., Davies, M, Yates, T., Benhalima, K., Lawrence, I, Khunti, K., 2010. Type 2 diabetes in younger adults: the emerging UK epidemic. Postgraduate Medical Journal, 86(1022), pp.711-718.

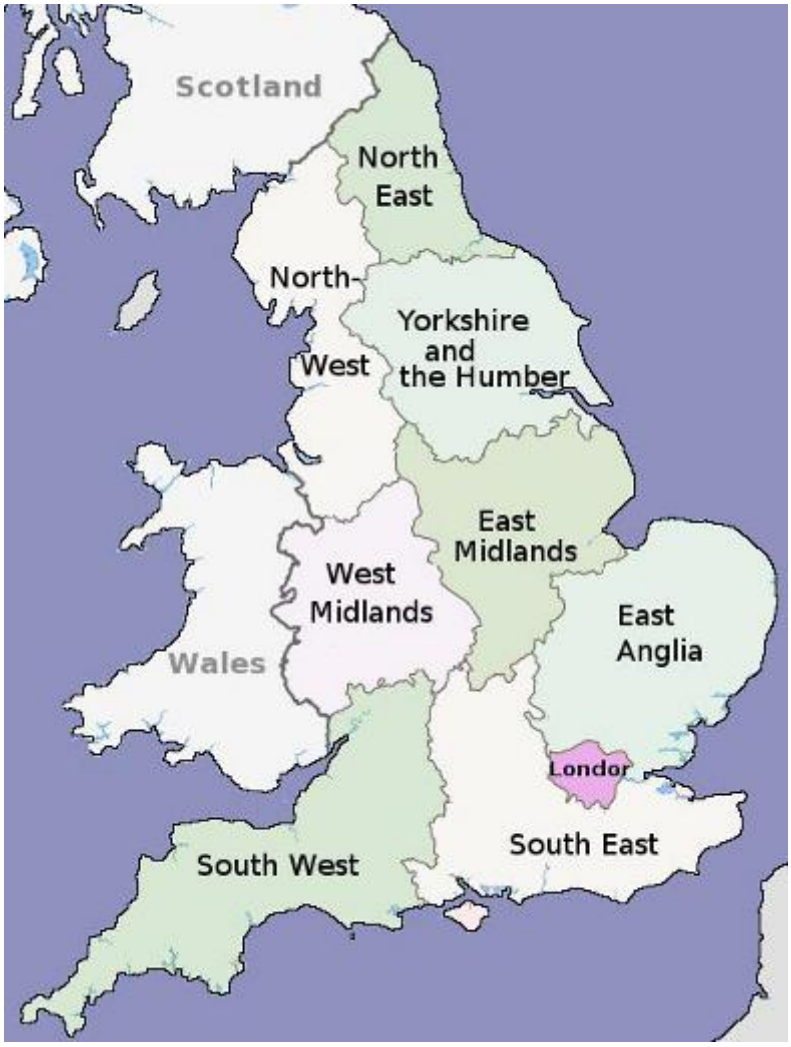

Figure 1 Regions of England

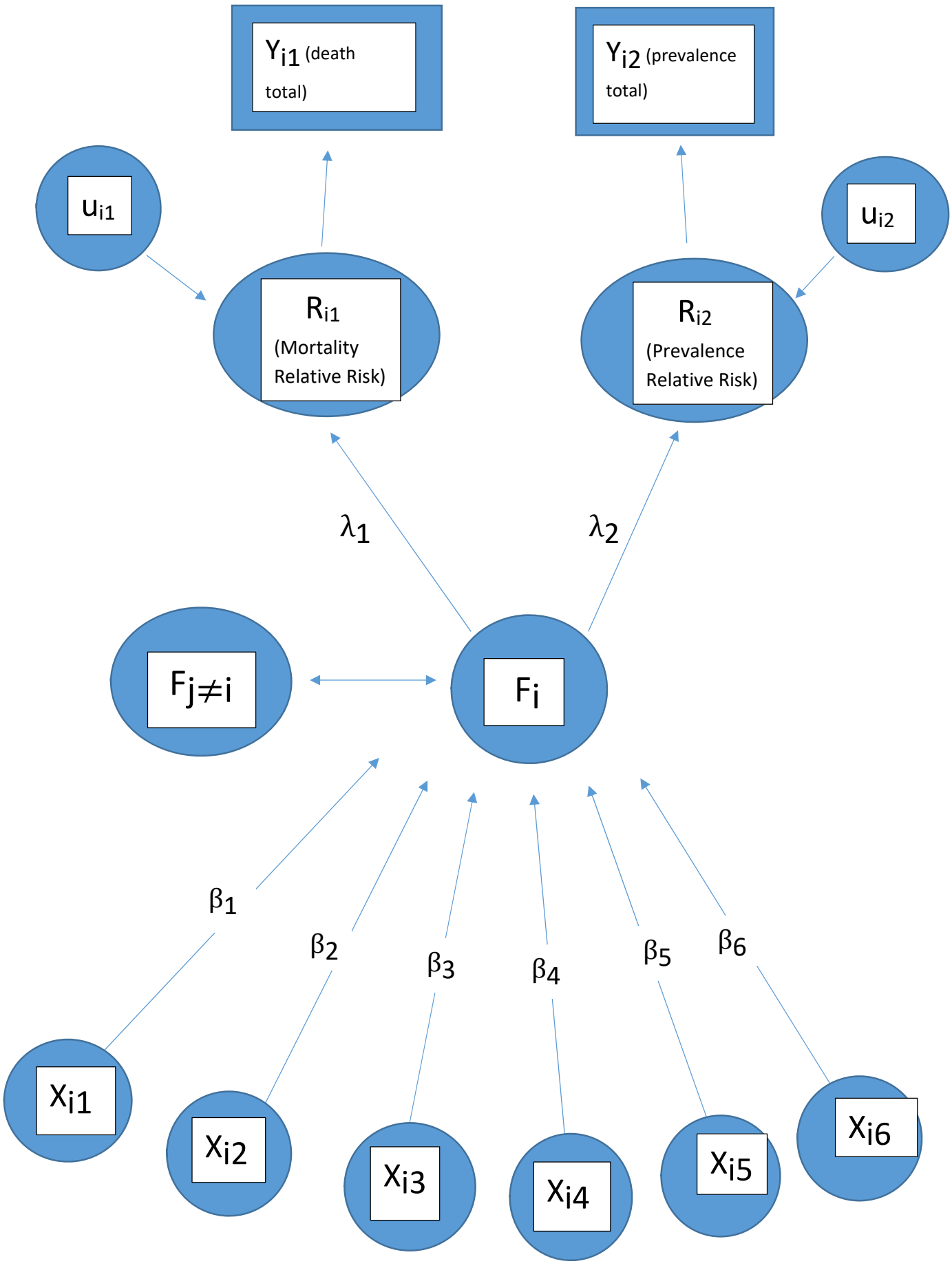

Figure 2 Directed Acyclic Graph for Diabetes Risk Model

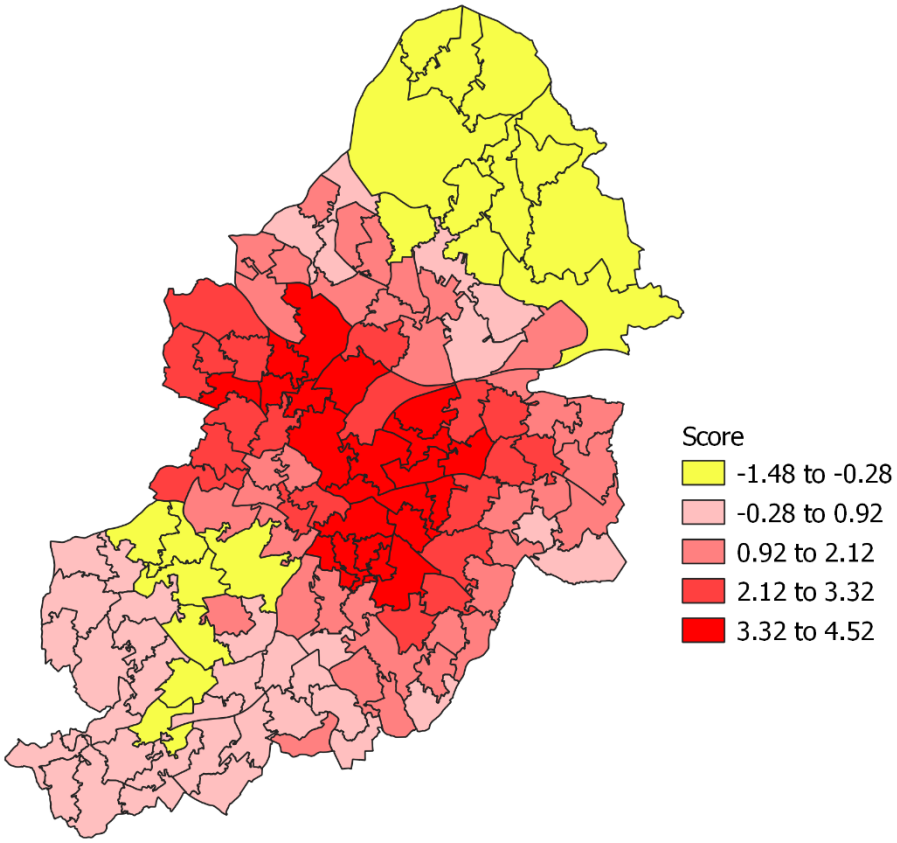

Figure 3 Factor Scores in Birmingham MSOAs

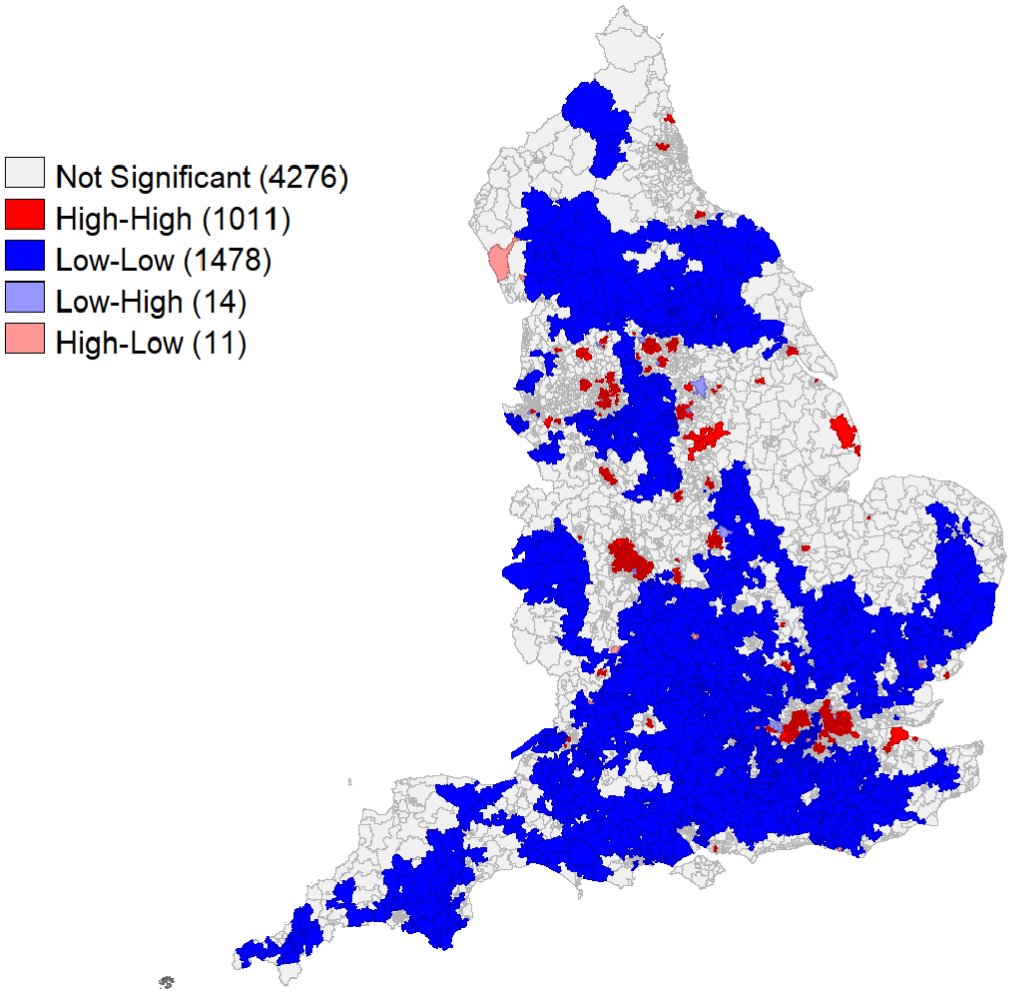

Figure 4 LISA Cluster Categories for England MSOAs

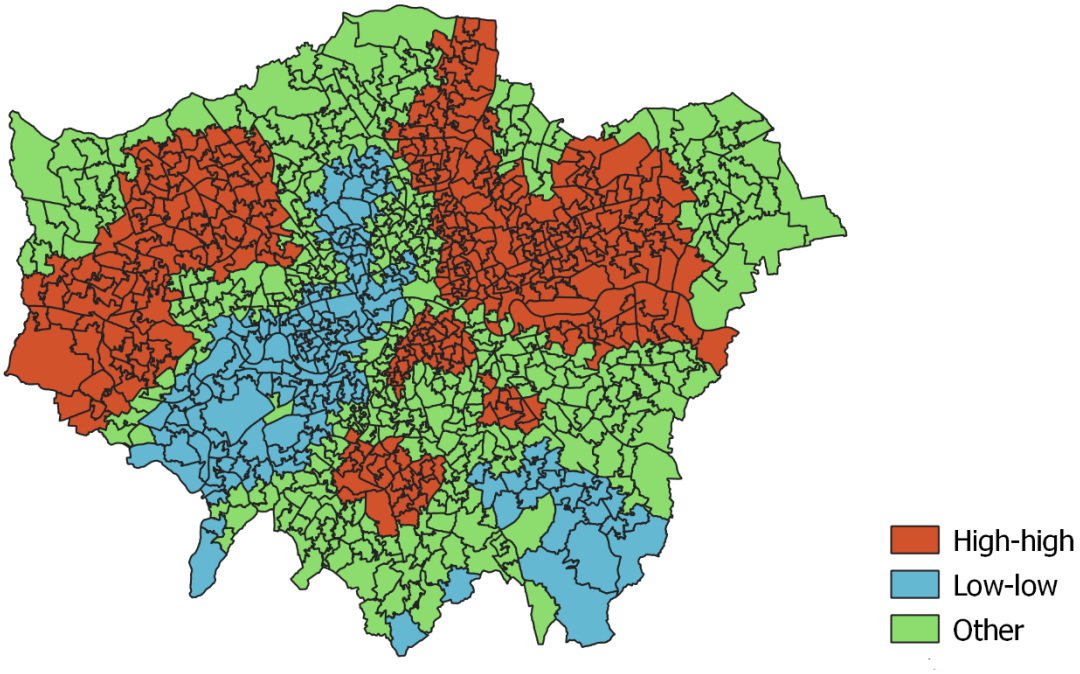

Figure 5. Need Cluster Categories based on LISA Indices (Greater London)

Table 1 Regression Parameter Summary

Regression Effects	Mean	St devn	2.5%	97.5%
ICE-ASG, β_1	2.55	0.066	2.42	2.67
Fragmentation, β_2	2.50	0.093	2.32	2.69
South Asian, β_3	3.06	0.110	2.84	3.27
Urbanity, β_4	0.44	0.059	0.33	0.56
Physical inactivity, β_5	0.34	0.093	0.16	0.53
Overweight/Obese, β_6	0.43	0.123	0.18	0.67
Other Parameters	**Mean**	**St devn**	**2.5%**	**97.5%**
α_1	-0.022	0.014	-0.021	0.002
α_2	0.009	0.012	0.013	0.027
λ_1	0.197	0.006	0.197	0.209
λ_2	0.195	0.002	0.195	0.199

Table 2a Average Need Scores by Region and Urban-Rural Category

	Urban Major Conurb-ation	Urban Minor Conurb-ation	Urban (City and Town)	Rural Town and Fringe	Rural Village and Dispersed	All Categories
London	0.53	-	-1.44	-0.02	-	0.53
South East	-0.73	-	-0.47	-1.28	-1.69	-0.70
South West	-	-	-0.33	-0.78	-1.13	-0.54
East of England	-0.11	-	-0.25	-0.82	-0.94	-0.43
East Midlands	-0.70	0.23	0.50	-0.45	-0.67	0.17
West Midlands	1.35	-	0.33	-0.59	-0.98	0.63
Yorkshire/Humberside	0.70	0.43	0.03	-0.63	-1.08	0.20
North West	0.43	-	0.11	-0.58	-1.04	0.20
North East	0.17	-	0.27	0.19	-0.72	0.18
All Regions	0.54	0.35	-0.10	-0.72	-1.11	0.00

Table 2b Modelled Mortality Rates (per 100,000) by Region and Urban-Rural Category

	Urban Major Conurb-ation	Urban Minor Conurb-ation	Urban (City and Town)	Rural Town and Fringe	Rural Village and Dispersed	All Categories
London	47.87	-	28.87	35.52	-	47.82
South East	34.99	-	39.04	33.15	29.99	37.18
South West	-	-	41.17	36.80	34.25	39.26
East of England	41.21	-	42.22	36.68	35.47	40.25
East Midlands	34.70	49.14	48.36	39.15	37.96	45.90
West Midlands	58.79	-	46.03	38.77	36.24	50.62
Yorkshire/Humberside	50.02	44.49	45.31	38.03	34.08	45.24
North West	44.83	-	42.90	36.68	34.21	43.25
North East	41.63	-	44.58	45.90	34.94	43.11
All Regions	47.80	46.39	42.62	37.32	34.33	43.40

Table 2c Modelled Prevalence (Percentage) by Region and Urban-Rural Category

	Urban Major Conurb-ation	Urban Minor Conurb-ation	Urban (City and Town)	Rural Town and Fringe	Rural Village and Dispersed	All Categories
London	7.70	-	4.94	6.45	-	7.70
South East	5.89	-	6.19	5.19	4.79	5.92
South West	-	-	6.30	5.74	5.36	6.05
East of England	6.54	-	6.44	5.69	5.59	6.21
East Midlands	5.85	7.01	7.48	6.15	5.90	7.01
West Midlands	8.90	-	7.17	5.94	5.51	7.75
Yorkshire/Humberside	7.90	7.33	6.77	5.93	5.43	7.10
North West	7.36	-	6.89	5.98	5.48	7.04
North East	6.91	-	7.04	6.92	5.88	6.93
All Regions	7.65	7.20	6.63	5.82	5.40	6.83

Table 3 Outcome and Risk Factor Profiles according to Need Score Category

MSOAs by Need Score Decile

Need Score Decile	Modelled Mortality Rate (per 100,000)	Modelled Prevalence Rate (Percent)	Index of Concentrat-ion at Extremes (ASG's DE vs ASG's AB)	Social Fragment-ation Score (0,1 scale)	% South Asian	Urbanity (0,1 scale)	Physically Inactive Adults (%)	Over-weight Adults (%)
1 (Lowest Need)	28.1	4.4	-0.29	0.22	1.7	0.38	19.0	56.8
2	32.6	5.1	-0.18	0.20	1.6	0.36	20.1	59.9
3	35.1	5.6	-0.11	0.21	1.7	0.39	20.9	62.0
4	37.6	6.0	-0.06	0.22	2.0	0.46	21.8	62.6
5	40.1	6.3	0.00	0.23	2.3	0.50	22.1	63.0
6	42.2	6.7	0.06	0.24	2.5	0.54	23.2	64.1
7	45.2	7.2	0.12	0.27	2.8	0.55	23.3	64.3
8	49.3	7.6	0.19	0.30	3.6	0.58	23.8	64.8
9	54.7	8.4	0.23	0.33	6.1	0.63	24.0	63.6
10 (Highest Need)	68.8	10.9	0.24	0.32	26.8	0.74	25.5	61.3