The Ethnic Density Effect as a Contextual Influence in Ecological Disease Models: Establishing its Quantitative Expression

Peter Congdon, School of Geography, Queen Mary University of London, London E1 4NS, UK. p.congdon@qmul.ac.uk

Abstract. Research suggests higher neighbourhood ethnic minority density to be associated with lessened chances of ethnic group illness. We focus on the density effect on psychosis, arguing that (at higher ethnic concentrations) it acts as a contextual influence attenuating the compositional influence whereby minority ethnicity is associated with higher psychosis risk. In terms of ecological disease regression, the ethnic density effect will then be apparent in nonlinear impacts of minority concentration. Contextual effects may also be evident in spatially varying regression coefficient models for psychosis. Nonlinearity or heterogeneity may be associated with other contextual processes where geography modifies demography (e.g. deprivation amplification). We illustrate these issues with an analysis of psychosis prevalence in 4,835 London neighbourhoods. The data are collected in primary care (during 2019/20) using clinical diagnosis (e.g. based on referrals to specialists or psychosis hospitalisations), and currently under care: such care may extend retrospectively over several years. The data offer a complete population perspective in contrast to survey data which typically offer limited geographic perspectives. We consider impacts on psychosis prevalence of non-white ethnicity, as well as those of deprivation, social fragmentation and urbanicity. We find evidence suggesting nonlinear impacts of non-white ethnicity on psychosis (essentially flat risk above a threshold concentration), but find no evidence for deprivation amplification.

Highlights

Assesses role of ethnic density and deprivation amplification in varying neighbourhood prevalence of psychosis.

Finds evidence of threshold ethnic concentration beyond which there is no increase in psychosis risk, consistent with an ethnic density effect.

Finds no evidence supporting deprivation amplification in explaining varying neighbourhood psychosis rates

Uses ecological inference to assess dependence of non-white psychosis rates on minority concentration.

Keywords. Psychosis. Ethnic density effect. Deprivation amplification. Spatial regression. Ecological inference.

1. Introduction

Neighbourhood studies of disease variation in relation to area risk factors may identify contextual (place) effect or compositional effects (Leyland and Groenewegen, 2020). The latter simply reflect differences in population composition between areas without any mediation by place influences. For example, lower neighbourhood socioeconomic status is often associated with higher levels of serious mental illness, but this effect may be primarily because individuals with lower socioeconomic status are more at risk of serious mental illness (Read et al, 2013).

However, the spatial epidemiology literature has identified mechanisms which specifically imply place mediation. For example, several studies report an ethnic density effect on illness, especially certain types of mental illness (Becares et al, 2018). Other forms of protective density effect have been reported (Schofield et al, 2016). It has been argued that higher neighbourhood ethnic minority density may lead to lessened chances of illness due to the buffering effects against race discrimination in such areas. Becares et al (2009) refer to “social networks and supportive communities, possibly mitigating the detrimental impact of racism on the health of ethnic minority people”, and reinforced social capital in high density areas is also relevant (Bennett et al, 2020; Baker et al, 2021). With regard to psychosis, these influences suggest a flattening of psychosis risk at higher levels of ethnic minority concentration in regression studies (i.e. a form of nonlinearity), offsetting the compositional effect. If spatial heterogeneity in risk factor effects is introduced into regression approaches, the ethnic density effect may imply smaller coefficients on non-white ethnicity at higher minority densities.

Similarly, deprivation amplification is another contextual mechanism that acts to exaggerate the already negative health effects of poverty at the individual level (Mennis et al, 2012; Munford et al, 2022).

We focus in this paper on neighbourhood variations in psychosis, where ethnic concentration and area deprivation are among factors that have been found relevant in explaining area contrasts. Psychosis is distinctive in the number of research studies finding significant effects of place factors (e.g. Lin and Goodman, 1992; March et al, 2008). These include factors such as social fragmentation (Allardyce and Boydell, 2006) and urbanicity (Spauwen and Van Os, 2006).

We aim to establish the quantitative form of these neighbourhood risk factors in a study of psychosis variations across 4,835 micro-neighbourhoods in London, and with 111,729 total diagnosed cases of psychosis. We also seek to establish the extent to which there are ethnic density effects and deprivation amplification. As a preliminary to formal regression analysis, we seek, using ecological inference, to establish whether psychosis rates among non-whites show an ethnic density effect. We also use machine learning methods to make a preliminary examination of the form of the psychosis-ethnicity and psychosis-deprivation associations (e.g. their linearity or otherwise). We then consider, via spatial regression, aggregate (all population) rates of psychosis prevalence in relation to ethnic mix, deprivation, fragmentation and urbanicity. We seek to establish the extent to which there are nonlinear effects of non-white ethnicity on total psychosis rates – which would be a second form of evidence consistent with a density effect. By contrast, a linear relation, whereby the incremental effect of an increase in psychosis risk is the same for all ethnic minority concentrations, would not support a density effect (Roberts and Martin, 2006). We also examine evidence for deprivation amplification, and whether or not an incremental linear relation is appropriate for the psychosis-deprivation association.

2. The Quantitative Expression of Contextual Processes

A particular aim of the work here is to establish the quantitative form of the ethnic density effect, and of other place effects on psychosis such as those due to deprivation and social fragmentation.

Regarding the ethnic density effect in particular, this effect exists against the background of considerably higher rates of serious mental illness, such as psychosis, among non-white ethnic groups. For example, Qassem et al (2015) report an odds ratio of 2.9 for psychosis among black people as compared to whites. This excess risk would, if expressed as a compositional effect unmediated by area effects, simply show in regression as a positive monotonically increasing impact of neighbourhood non-white proportions on aggregate levels of psychosis.

Here we argue that ethnic density acts as a contextual influence to modify the compositional influence at higher minority concentrations.

In general, a purely compositional effect (of an adverse risk factor) will translate into monotonically increasing disease rates as the exposure level increases. Many epidemiological studies in fact show approximately linear risk-exposure relationships (e.g. Roberts and Martin, 2006), and a linearity assumption is typically the baseline in regression studies of health variation.

By contrast, we suggest that the ethnic density effect will be apparent in nonlinear impacts of minority concentration on total psychosis rates: initially prevalence or incidence rises at relatively low and medium ethnic concentrations, reflecting the compositional effect. However, at higher minority concentrations, the contextual density effect becomes apparent, such that the monotonic increase will attenuate or disappear, and the regression curve will flatten or possibly turn down. This will show in a nonlinear association between aggregate (all population) psychosis risk and non-white minority concentration.

A number of studies have focused on the ethnic density effect in terms of ethnic-specific psychosis rates (albeit, relevant UK studies generally refer to psychosis incidence). They report that higher ethnic concentrations to be associated with lower psychosis rates among the corresponding ethnic groups.However, interdependence between ethnic concentration and other area factors may mitigate against a simple linear negative association. Higher concentrations of ethnic minorities tend to be in more deprived areas; the London study below shows a 0.5 correlation between non-white ethnicity and income deprivation. Hence low ethnic concentrations are more likely in relatively affluent areas, and a simple linear negative effect would imply unduly high psychosis rates among (relatively affluent) ethnic minorities in such areas. Some nonlinearity may instead characterise associations between ethnic density and ethnic specific psychosis rates.

Contextual effects may also be evident in spatially varying regression coefficients (spatial heterogeneity) (Assuncao, 2003). In a regression with all population psychosis levels as the outcome, an ethnic density effect would imply lower coefficients on ethnic density (e.g. less positive effects on psychosis risk) in areas with higher ethnic minority concentrations.

Similarly, nonlinearity or heterogeneity may occur under other contextual processes, such as other forms of group density effect (Schofield et al, 2016), or deprivation amplification. Under deprivation amplification, one would expect higher levels of mental illness among those in poverty to be further enhanced by higher levels of deprivation in an area. This implies an enhancement of risk at higher levels of deprivation in regression studies (an upward concave association). For a spatially varying coefficients approach, deprivation amplification would imply higher coefficients (more enhanced risk) in more deprived areas.

3. Methods

3.1 Study Design

We illustrate these issues with an analysis of psychosis prevalence in 4,835 lower super-output areas (LSOAs) in London, considering impacts of deprivation, social fragmentation and urbanicity, as well as of non-white ethnicity. These are Census micro-neighbourhoods with an average population of 1,800. The benefits of a small neighbourhood scale for analysis of health variations have been discussed in several studies (e.g. Thunhurst, 2009), and for psychosis in particular by Schofield et al (2011).

Psychosis is for the year 2019/20, and defined according to the Quality Outcomes Framework or QOF (Department of Health, 2022), a scheme for monitoring chronic disease in UK primary care. Psychosis is defined as diagnosed "schizophrenia, bipolar affective disorder and other psychoses". Numbers of cases by small area are provided by the UK House of Commons Library (2022); totals are also available by general practitioner (GP) practice. There are a total of 111,747 diagnosed psychosis cases, or around 1.3% of the total population of London. The data offer a complete population perspective in contrast to survey data (such as the Adult Psychiatric Morbidity Survey in the UK) which typically offer limited geographic perspectives.

Typically, a diagnosis of psychosis would not be made in primary care but would be based on referrals to specialists or on psychosis hospitalisations. However, patient care in the UK is commissioned by primary care general practitioners, and ongoing care would be primary care based or commissioned: such care may extend retrospectively over several years. According to a review by Public Health England (PHE, 2016) on psychosis data: "the [QOF] register is a cumulative count of all identified cases, so as the register builds it will come to show a primary care-based lifetime prevalence".

We consider four predictors of area psychosis relativities: a social fragmentation score (Allardyce et al, 2005); a measure of urbanicity; a measure of income deprivation from the 2019 Index of Multiple Deprivation (MHCLG, 2019); and a measure of non-white ethnicity, namely the percent of LSOA populations (2021 Census) of all non-white ethnicity, including mixed ethnicity.

Urbanicity is measured by combining information (via principal component analysis) on (a) population density (2021 Census), (b) the percentage of dwellings that are flats, from the 2021 Census (cf. Kovess-Masfety et al, 2005), and (c) a green space measure developed by the Consumer Data Research Centre (CDRC) in collaboration with Sentinel (https://sentinel.esa.int/web/sentinel/sentinel-data-access). The first two of these indicators are direct (positive) measures of urbanicity, the third is a negative indicator.

The fragmentation score, obtained by principal component analysis (PCA), is a measure of population transience, and of household structure not oriented to familism and settled marital status (Timms, 1971). The score is based on five 2021 Census indicators: % of population with different address one year previously; all one person households (as % of household heads of all ages); one person households under 65 (as % of household heads aged under 65); private rented household tenures (not including socially rented housing); and married couple households. There is a correlation of 0.89 between the two one person household indicators which the PCA will take account of; if two variables are highly correlated, then they will load highly on the same principal component. A population churn indicator available from the CDRC is also used as an additional measure of transience in deriving the PCA score. Evidence from Nazroo and King (2002) points to higher psychosis among single, widowed and divorced people, and among renters. The focus on private renting reflects its role in the UK as a relatively short-stay, less secure form of accommodation, with much higher turnover than other housing sectors (in other countries private renting may not have these characteristics) (e.g. Bone, 2014; Bone and O'Reilly,2010)

A central feature of London's population is its diversity, with non-white ethnic groups (including mixed ethnicity) accounting for 3.94 million of the total population of 8.6 million (2021 Census figures). Non-white ethnicity in the analysis here includes all ethnicities not classed as white; in terms of actual 2021

Census designations (https://www.nomisweb.co.uk/), these comprise Asian, Black, Mixed or Multiple, or Other. Census ethnicity is based solely on self-identification. Non-white ethnicity in London is associated with, and overlaps with, area deprivation: 18% of the population in the 10% least deprived LSOA neighbourhoods are non-white, compared to 55% in the most deprived neighbourhoods. So possible interactions (or offsetting effects) between poverty and ethnicity may be important as an influence on psychosis rates (Becares et al, 2018).

3.2 Data Analysis: Ecological Inference

We consider two preliminary forms of analysis which complement and inform interpretation of the subsequent spatial regression analysis.

The first preliminary analysis involves the ecological inference (EI) technique (King, 1997). The EI approach has been most commonly applied in political science applications, for example where (all race) voting registration rates are available for districts, as are data on percentages black and white, and the goal is to estimate registration rates by race. Here we use the same principle, implemented via the R package ei, to estimate psychosis rates among non-white people in London LSOAs, using known data on total psychosis rates (cases per total population, and on Census 2021 ethnic data in LSOAs (proportions of population non-white and white).

A primary goal is to assess the extent to which non-white psychosis shows a density effect with regard to non-white ethnic concentration. If confirmed, this would provide one form of evidence supporting (or not) a density effect, which can be compared with the evidence from spatial regression on a density effect.

The EI method is based on the identity (for neighbourhood i)

$$\pi_i=\pi_{bi}(x_i)+\pi_{wi}(1-x_i),$$

where x_i is the proportion non-white in neighbourhood i, π_i is the overall psychosis rate, π_{bi} is the psychosis rate for non-whites, and π_{wi} is the psychosis rate for whites. π_i and x_i are known, whereas π_{bi} and π_{wi} are unknown.

The EI method is also used to provide a preliminary assessment of the extent of any deprivation amplification effect on psychosis, with the deprived category in each neighbourhood being based on Census 2021 information on individual deprivation dimensions. The deprived or poverty group in each LSOA is based on the 2021 UK Census Table TS011 (Households by Deprivation Dimensions), and those with two or more dimensions of the four possible (low education, lack of employment, poor health, or poor housing) are taken as the deprived group in each LSOA. These account for 18% of households across England (and 20% in London). The extent to which EI analysis does support deprivation amplification is one form of evidence which can be checked against what spatial regression shows.

3.3 Data Analysis: Machine Learning

The second preliminary analysis is a machine learning (ML) approach to investigate the shape of the relationships between psychosis (the total psychosis rate) and the four neighbourhood characteristics (non-white ethnicity, deprivation, social fragmentation, urbanicity). ML is also used to assess the most important interactions between the predictors. ML methods have established utility to assess nonlinearity, and the most important interactions, without detailed preliminary parametric specification as in formal regression (Ryo and Rillig, 2017; Hu et al, 2009; Lampa et al, 2014).

For a response variable, we follow the blended approach of Hu et al (2009), using empirical Bayesian smoothed prevalence ratios (SPR) for psychosis (with average ratio of 1, and analogous to a standard mortality ratio), which take account of the spatial structure of the observed responses. The smoothing is carried out via the program GeoDa using observed and expected psychosis totals in each LSOA. Spatial interaction is defined by binary adjacency (weights of 1 for adjacent LSOAs, 0 otherwise), while expected cases are derived using psychosis by age schedules from Nazroo and King (2002).

The SPRs are then used in conjunction with machine learning, specifically random forest regression as implemented in the R program randomForestSRC. The random forest algorithm combines ensemble learning methods with a decision tree approach to generate multiple randomly drawn decision trees from a dataset, averaging results to provide predictions without overfitting (Ishwaran, 2007). Each decision tree is based on a random sample of observations from the dataset and a random sample of the features (predictors).

Another potential gain in ML over standard regression methods in detecting interactions (McClelland and Judd, 1993), which can be assessed without extensive fitting of alternative models. We use the iml and randomForest libraries (Boehmke and Greenwell, 2019), and measure interactions using the two-way H-statistic (Inglis et al, 2022). This represents interaction strength, and is between 0 (no interaction) and 1, if all variation in the predicted outcome depends on a given interaction.

3.4 Data Analysis: Spatial Regression

We then consider a formal spatial regression whereby the relative risk of diagnosed psychosis is related to the four neighbourhood characteristics. A log-link Poisson regression appropriate to the response variable (psychosis cases) is adopted.

The predictors are standardized, and so estimated means of the regression coefficients can be seen as measuring the relative importance of predictors. A different perspective on relative importance is provided by the Pratt index (Aschard, 2016), obtained by multiplying standardized regression coefficients by the correlation between the predictor and the response (estimated psychosis relative risk).

A spatially correlated random effect is used to represent the overall impact of unobserved covariates, as is standard in Bayesian disease mapping (Byers and Besag, 2000). Initial analysis, not presented here in detail, suggests the presence of “spatial confounding” whereby correlation between the four observed area characteristics and the random effect affects regression results. Therefore, the method of Reich et al (2006) is used in the spatial regression, as this adjusts for confounding. We use the R program RASCO as developed by Azevedo et al (2020) in this analysis.

As posited above, an ethnic density effect would be evident if the aggregate psychosis rate demonstrated non-monotonicity and non-linearity in relation to proportions non-white in neighbourhoods. This would imply rising prevalence or at relatively low and medium ethnic concentrations, consistent with a compositional effect, but at higher minority concentrations, the contextual density effect will mean the monotonic increase is attenuated or eliminated, and the regression curve will flatten or possibly turn down.

As also mentioned, it is also relevant to assess whether, and how, contextual effects (e.g. ethnic density) are expressed when spatial heterogeneity in regression is allowed. We use the R program INLA (Gómez-Rubio, 2020) to implement a spatially varying coefficients approach, again with Poisson regression. Methods which admit spatial heterogeneity while also adjusting for confounding have yet to appear in the literature, so interpretations of this analysis are subject to that caveat.

4. Results

4.1 Ecological Inference

We find that the overall psychosis rate among London's non-white population (the unknown π_{bi} in section 3.2) is estimated at 0.0199 (i.e around 19.9 per 1000), and that 78,237 or around 70% of London's diagnosed cases of psychosis are among non-whites. So, non-whites are a clear majority of the total burden of psychosis in London (Rössler et al, 2005), and by extension the related health care costs (Qassem et al, 2016). The psychosis rate among whites is estimated at 0.0072 (or 7.2 per 1000), with a non-white to white relative risk of 2.76. The rate for whites is consistent with findings of Qassem et al (2016) and Nazroo and King (2002), while the relative risk is comparable to the ethnic group relative risk estimates in Fearon et al (2006), although these refer to psychosis incidence.

The neighbourhood rates for non-white psychosis have a correlation of -0.45 with percentages non-white across LSOAs, with a 95% confidence interval (-0.47,-0.42), providing strong evidence for a negative association, in line with an ethnic density effect. The shape of the relationship is however nonlinear (see Figure 1a), with the peak rate of non-white psychosis coinciding approximately with the average concentration of non-whites, around 39%. At higher rates of ethnic density, the risk of psychosis among non-whites tails off. The same nonlinearity characterises a plot of the relative psychosis risk (the ratio of non-white to white psychosis rates) against the LSOA percent non-white (Figure 1b).

The ecological inference approach can also be applied to assess deprivation amplification. Under such amplification, one would expect that the psychosis rate among those in poverty would be elevated in highly deprived neighbourhoods as compared to less deprived areas.

We find from the R program ei that deprived people (those with two or more dimensions of deprivation according to the 2021 UK Census) have a psychosis rate of 37.9 per 1000, whereas the non-deprived have a rate of 6.9 per 1000. Deprivation amplification would be supported by a positive correlation between neighbourhood psychosis rates in the poverty group and the neighbourhood index of income deprivation. In fact, we find a negative correlation of -0.12, so not supporting the presence of deprivation amplification in psychosis variations across London.

4.2 Machine Learning: Nonlinearity, Marginal Dependency Plots, and Interactions

The preceding analysis has established that the majority of diagnosed psychosis cases in London are among non-whites. It follows that impacts of neighbourhood non-white ethnicity on total (all population) psychosis cases are likely to reflect ethnic density effects. We therefore seek to establish the form of the relationship between non-white concentration and the all population psychosis rate.

A preliminary impression of the form of this impact can be obtained, using smoothed prevalence ratios (Hu et al, 2009) as a measure of psychosis risk. A bivariate plot (Figure 2) of the ratios against % non-white shows a tailing off in increased psychosis risk at higher non-white concentrations.

To assess the association via ML analysis, we use the smoothed prevalence ratios as the response in a random forest regression on the four predictors: deprivation, percent non-white, social fragmentation and urbanicity. Of interest are the marginal dependence plots which represent the overall trend in predictor/response relations (Ehrlinger, 2016). Figure 3 shows the marginal plots for the four predictors in terms of their impact on prevalence ratios. The horizontal axes express the predictors in standardized form. The plots for deprivation and urbanicity are approximately linear, though the urbanization curve tails off slightly at higher values. Hence the deprivation plot does not provide evidence for deprivation amplification, which would be suggested if the gradient of the plot steepened at high deprivation levels.

By contrast, the plot for non-white ethnicity shows an attenuated, essentially flat, effect on psychosis risk for percentages non-white above average. This type of nonlinearity is in accord with the pattern posited above, with the protective ethnic density effect coming into play at higher levels of non-white ethnicity. The fragmentation effect also tails off at higher fragmentation values, possibly also some sort of density effect – for example, van Os et al (2000) find a protective density effect (regarding schizophrenia) for people with single marital status.

Regarding potential interactions, we find the highest H-statistics to be between deprivation and % non-white (0.41), between deprivation and fragmentation (0.35), and between deprivation and urbanicity (0.32).

4.3 Results: Spatial Regressions

We now apply spatial regression, including a comparison between a linear effects model with one nonlinear in the effects of % non-white and fragmentation. The dependent variable is the relative risk of psychosis in neighbourhoods, with the average risk being 1. A subsidiary analysis allows for spatial heterogeneity, through spatially varying coefficients. The aim in both non-linear and varying coefficients analyses is to assess evidence of contextual effects, and their quantitative form, especially the ethnic density effect.

In terms of the patterning in psychosis we seek to explain, Figure 4 shows smoothed prevalence ratios (as an estimator of relative risk) across London LSOAs, with higher psychosis levels in much of inner London, though also characterising some inner suburbs.

4.3.1 Linear vs Non-linear Spatial Regressions

In the first analysis, we use the method of Reich et al (2006) to investigate the relationship between psychosis and the four predictors. We first apply a spatial regression with linear effects of the four predictors, plus the three interaction terms noted above. We aim to assess whether, and how far, model fit is improved by allowing nonlinear effects in two of the four covariates.

Specifically, based on the exploratory machine learning analysis summarised above, we use nonlinear functions (linear splines) to represent the impacts on psychosis risk of % non-white and fragmentation. More complex nonlinear functions (e.g. cubic splines) were found to be imprecisely estimated. For the non-white spline, knots are at values -1,-0.5,0, and 0.5 in the standardized scale; while for fragmentation they are at -1,0, and 1.

Table 1 summarizes the parameters of the two spatial regression models, with fit represented by the DIC criterion (Spiegelhalter et al, 2002) – lower DIC values mean better fit. We assess significance by whether the 95% interval excludes zero.

In Table 1(a) for the linear model, the impact of %non-white (NW) is likely to be understated as the overall NW slope includes the flattened risk at high NW values (see Figures 2 and 3). The non-linear model, as in Table 1(b), has better fit (a lower DIC), and better represents the steep upward slope in psychosis risk at lower values of ethnic concentration.

Table 1 Regression Summary, Spatial Regression					
(a) Linear Model					
	Mean	Median	2.5%	97.5%	Pratt Index
(Intercept)	-0.051	-0.051	-0.057	-0.043	
Fragmentation	0.083	0.083	0.073	0.091	0.031
Deprivation	0.144	0.144	0.136	0.150	0.109
% Non-white	0.066	0.066	0.058	0.073	0.034
Urbanicity	-0.013	-0.013	-0.023	-0.003	-0.007
Interactions					
Non-white and Deprivation	0.027	0.027	0.018	0.035	0.006
Fragmentation and Deprivation	0.016	0.016	0.008	0.025	0.000
Urbanicity and Deprivation	0.007	0.006	-0.001	0.014	0.002
Fit					
DIC	26627.5				
(b) Non-Linear Model					
	Mean	Median	2.5%	97.5%	Pratt Index
(Intercept)	0.360	0.363	0.201	0.492	
Fragmentation (Linear term only)	0.124	0.124	0.067	0.175	0.041
Deprivation	0.132	0.132	0.123	0.140	0.104
% Non-white (Linear term only)	0.354	0.355	0.187	0.517	0.227
Urbanicity	-0.008	-0.008	-0.018	0.001	-0.004
Interactions					
Non-white and Deprivation	0.035	0.036	0.027	0.045	0.006
Fragmentation and Deprivation	0.011	0.011	0.002	0.018	0.000
Urbanicity and Deprivation	0.034	0.034	0.025	0.042	0.011
Fit					
DIC	26597.1				

From Table 1(b), it is apparent that the most important influences on neighbourhood psychosis risk, based on mean standardized regression coefficients, are % non-white, deprivation and fragmentation. Interaction effects are small, albeit more significant than in the linear model. The more important interactions are between deprivation and urbanicity, and between deprivation and % non-white.

The Pratt index emphasizes more the importance of deprivation as an influence on neighbourhood psychosis, though in the non-linear model also stresses the role of non-white ethnicity.

The regressions here therefore suggest that high psychosis in inner London (see Figure 4) are especially associated with higher deprivation levels in some parts of inner London, and also with the residential patterning of non-white groups. The insignificant effect of urbanicity (and its slightly negative Pratt index) may reflect multicollinearity (Thomas et al, 2007).

Figure 5a shows the predicted effect on relative psychosis risk of percentages non-white from the spline representation in the non-linear model. As in Figures 2 and 3 above, the attenuation of impact at higher ethnic concentrations is readily apparent, consistent with an ethnic density effect. Figure 5a shows a threshold ethnic concentration (around half the population) beyond which there is no increase in psychosis risk. By contrast, Figure 5b shows only a slight attenuation in the impact of fragmentation at higher scores.

4.3.2 Varying Spatial Coefficients

The main output of interest from the varying coefficient analysis is the extent to which varying coefficients are systematically associated with the original predictor values. Figure 6 accordingly shows varying coefficients relating relative psychosis risk to % Non-White, plotted against the actual values for this predictor.

Figure 6 shows the highest positive coefficients are at low values of non-white concentration: the average coefficient is 0.035, but coefficients two or three times the average are apparent at the lowest non-white concentrations. The average coefficient for the lowest quartile of non-white concentrations is 0.05. The correlation between the varying slopes and the percents non-white is -0.19, with 95% interval (-0.22, -0.16).

By contrast, there is no association between varying deprivation coefficients and the deprivation score – whereas a positive association would be expected under deprivation amplification. There is very little variation in deprivation slopes around the average of 0.143: the 5^{th} and 95^{th} percentiles in the distribution of varying deprivation slopes are 0.141 and 0.146.

5. Discussion

The above analysis has shown consistent evidence, across different methods, of an ethnic density effect for all non-white groups combined. The analysis supports existing research, such as Bosqui et al (2014), who report "an overall ethnic density dose effect for ethnic minorities".

The results of spatial regression support the postulated nonlinear association in line with a density effect: initially all population psychosis prevalence rises sharply at relatively low levels of ethnic concentration, reflecting the compositional effect. However, at higher minority concentrations, the regression curve linking the overall psychosis rate to non-white ethnic concentration flattens, with no further increase in risk.

The ethnic density effect is apparent in a different way in a model using spatially varying coefficients. The highest impacts of non-white ethnic concentration on all population psychosis rates are at the lowest levels of concentration, with coefficients two or three times the average.

Such findings may have implications for health resourcing (Qassem et al, 2015). Thus, a simplistic linear formula linking resources to ethnic minority concentration might underestimate need in (say) deprived areas with relatively low ethnic minority concentration - where need would be elevated both by deprivation and relatively high psychosis among the ethnic minority. So nonlinearity in impacts of neighbourhood risk factors is relevant to need for funding and to resourcing that matches need.

In this paper we have focused mainly on the quantitative expression of two types of contextual effect: the ethnic density effect and deprivation amplification, finding in fact no evidence of the latter in the case of London's psychosis prevalence. These two processes are akin in their operation to differential registration by race across voting districts, especially in the US, which Wakefield (2004) summarises as a process whereby "registration of an individual will be associated with both their own race and the race of those around them (that is, we believe that registration will be a function of both demography and geography)". This is an alternative way of saying there is interplay between compositional and contextual effects.

There is related research on the quantitative expression of contextual effects for direct environmental risks (where there is no interaction between demography and geography), such as pollution or green space levels. For example, Richardson et al (2018) assess nonlinearity in impacts on psychosis of direct environment measures using categorisation. Assessing nonlinearity may be relatively complex, and different methods may have some impact on findings. Interestingly, the threshold established above (section 4) beyond which higher ethnic minority concentration implies no increase in psychosis risk (for example, see Figures 3 and 5) contrasts with a “threshold below” in the case of impacts of pollution on mortality (e.g. Roberts and Martin, 2006).

A small neighbourhood focus on psychosis prevalence, as undertaken here, has value in complementing the relatively limited evidence available from national community surveys. Studies such as Nazroo and King (2002) and Qassem et al (2015) provide prevalence estimates based on small numbers of identified cases, particularly in separate ethnic groups. Hence there is utility in estimates derived from small area administrative data, especially for the purposes of geographically specific resource and needs assessment for mental health care.

Limitations on the above analysis may be mentioned. Firstly, the ideal study framework to assess interactions between individual and environment is a multilevel one, though relevant studies are generally limited in geographic scope (e.g. of patient populations in single UK local authorities). The QOF data considered here could also possibly be analyzed at general practice level rather than by LSOA, though no patient details are available apart from diagnosis. A multilevel analysis might be possible for primary care register data in countries (such as Denmark) where patient details are released for research. The available data from the QOF are also not cross-hatched (i.e. the patients LSOA and general practice are not both known). All that is available is either psychosis totals by general practice or psychosis totals by LSOA. General practices are in a sense spatial entities (e.g. one could summarize their location by their postcode), so spatial clustering in psychosis could be considered, but analysis by practice rather than LSOA has disadvantages: for example, Census data is available for LSOAs but not practices.

An ecological study with a region-wide (or possibly national) perspective, as here, has the benefit of comprehensive coverage of different contextual settings. The research here shows the potential utility of clinical register data to enable neighbourhood level analysis, and so provide a geographical perspective typically not possible with national survey data.

A second possible limitation is that psychosis diagnosed in primary care may omit some part of the total community prevalence. Diagnosis rates may differ between ethnic groups and hence between neighbourhoods (Schwartz and Blankenship, 2014). Third, there is heterogeneity in the total psychosis category considered here, and any findings may differ between specific disorders such as bipolar and schizophrenia.

Declaration of competing interest.

None

Data availability.

The data used in this study may be obtained from https://github.com/houseofcommonslibrary/local-health-data-from-QOF/blob/main/README.md, the UK Census site (https://www.nomisweb.co.uk/), and the CDRC site (https://www.cdrc.ac.uk/).

References

Allardyce J, Gilmour H, Atkinson J, Rapson T, Bishop, J, McCreadie, R (2005) Social fragmentation, deprivation and urbanicity: relation to first-admission rates for psychoses. Brit J Psychiatry, 187(5), 401-406

Allardyce, J., Boydell, J. (2006). Environment and schizophrenia: review: the wider social environment and schizophrenia. Schizophrenia Bulletin, 32(4), 592-598.

Aschard, H. (2016) A perspective on interaction effects in genetic association studies. Genetic Epidemiology, 40(8), 678-688.

Assuncao, R (2003) Space varying coefficient models for small area data. Environmetrics, 14(5), 453-473

Azevedo, D, Bandyopadhyay, D., Prates, M, Abdel-Salam, A, Garcia, D (2020) Assessing spatial confounding in cancer disease mapping using R. Cancer Reports, 3(4), e1263.

Baker S, Jackson M, Jongsma H, Saville C (2021). The ethnic density effect in psychosis: a systematic review and multilevel meta-analysis. The British Journal of Psychiatry, 219(6):632-43.

Bécares, L., Nazroo, J., Stafford, M. (2009). The buffering effects of ethnic density on experienced racism and health. Health & place, 15(3), 700-708.

Bécares L, Dewey M, Das-Munshi J (2018) Ethnic density effects for adult mental health: systematic review and meta-analysis of international studies. Psychological medicine. 2018 Sep;48(12):2054-72.

Bennett N, Salway S., Piekut A (2020) What pathways have been theorised and tested between ethnic density and mental ill-health?: A theory-based systematic review Journal of Epidemiology and Community Health 74 Suppl 1 A17 -A17

Boehmke, B., Greenwell, B. (2019) Hands-on Machine Learning with R. Chapman and Hall/CRC.

Bone, J. (2014) Neoliberal nomads: housing insecurity and the revival of private renting in the UK. Sociological Research Online, 19(4), 1-14.

Bone, J, O'Reilly, K. (2010) No place called home: The causes and social consequences of the UK housing 'bubble'. The British Journal of Sociology, 61(2), 231-255.

Bosqui T, Hoy K, Shannon C (2014) A systematic review and meta-analysis of the ethnic density effect in psychotic disorders. Social Psychiatry and Psychiatric Epidemiology. 49:519-29.

Byers, S., Besag, J. (2000) Inference on a collapsed margin in disease mapping. Statistics in medicine, 19(17-18), 2243-2249.

Ehrlinger, J. (2016) ggRandomForests: random forests for regression. ArXiv preprint arXiv:1501.07196.

Fearon P, Kirkbride J, Morgan C (2006) Incidence of schizophrenia and other psychoses in ethnic minority groups: results from the MRC AESOP study. Psychological Medicine, 36:1541–50.

Gómez-Rubio, Virgilio (2020). Bayesian Inference with INLA. Chapman & Hall/CRC Press. Boca Raton, FL.

Hu, W., Mengersen, K., Tong, S. (2009) Spatial analysis of notified cryptosporidiosis infections in Brisbane, Australia. Annals of Epidemiology, 19(12), 900-907.

Inglis, A., Parnell, A., Hurley, C (2022). Visualizing variable importance and variable interaction effects in machine learning models. Journal of Computational and Graphical Statistics, 1-13.

Ishwaran, H. (2007). Variable importance in binary regression trees and forests. Electronic Journal of Statistics, 1, 519-537.

King, G (1997) A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data. Princeton University Press.

Kovess-Masfety, V., Lecoutour, X., Delavelle, S. (2005) Mood disorders and urban/rural settings: comparisons between two French regions. Social Psychiatry and Psychiatric Epidemiology, 40, 613-618.

Lampa, E., Lind, L., Lind, P, Bornefalk-Hermansson, A. (2014). The identification of complex interactions in epidemiology and toxicology: a simulation study of boosted regression trees. Environmental Health, 13(1), 1-17

Leyland, A, Groenewegen, P (2020) Context, composition and how their influences vary. In: Multilevel Modelling for Public Health and Health Services Research. Springer.

Lin, S, Goodman, A (1992) Geographic variations in the prevalence of schizophrenia. Archives of General Psychiatry, 49(11), 906-907.

March, D., Hatch, S, Morgan, C., Kirkbride, J, Bresnahan, M., Fearon, P., Susser, E. (2008) Psychosis and place. Epidemiologic Reviews, 30(1), 84-100.

McClelland, G, Judd, C (1993). Statistical difficulties of detecting interactions and moderator effects. Psychological Bulletin, 114(2), 376-390.

Munford, L., Khavandi, S., Bambra, C. (2022). COVID-19 and deprivation amplification: an ecological study of geographical inequalities in mortality in England. Health and Place, 78, 102933.

Nazroo J, King M (2002) Psychosis—symptoms and estimated rates. In: Sproston K, Nazroo J (eds) Ethnic minority psychiatric illness rates in the community (EMPIRIC)—quantitative report. TSO, London, pp 47–63

Public Health England (2016) Psychosis Data Report. PHE, London,UK

Qassem, T., Bebbington, P, Spiers, N, McManus, S, Jenkins, R., Dein, S. (2015) Prevalence of psychosis in black ethnic minorities in Britain: analysis based on three national surveys. Social Psychiatry and Psychiatric Epidemiology, 50(7), 1057-1064.

Read, J., Johnstone, L., Taitimu, M. (2013) Psychosis, poverty and ethnicity. In Models of Madness (pp. 217-235). Routledge.

Reich, B, Hodges, J, Zadnik, V. (2006) Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. Biometrics, 62(4),1197–1206.

Richardson, L., Hameed, Y., Perez, J., Jones, P, Kirkbride, J (2018) Association of environment with the risk of developing psychotic disorders in rural populations: findings from the social epidemiology of psychoses in East Anglia study. JAMA Psychiatry, 75(1), 75-83.

Roberts, S., Martin, M (2006) The question of nonlinearity in the dose-response relation between particulate matter air pollution and mortality: can Akaike's Information Criterion be trusted to take the right turn? American Journal of Epidemiology, 164(12), 1242-1250.

Rössler, W., Salize, H. J., Van Os, J., Riecher-Rössler, A. (2005). Size of burden of schizophrenia and psychotic disorders. European Neuropsychopharmacology, 15(4), 399-409.

Ryo, M, Rillig, M. C. (2017). Statistically reinforced machine learning for nonlinear patterns and variable interactions. Ecosphere, 8(11), e01976.

Schofield, P., Ashworth, M., Jones, R. (2011) Ethnic isolation and psychosis: re-examining the ethnic density effect. Psychological Medicine, 41(6), 1263-1269.

Schofield, P., Das-Munshi, J., Bécares, L., Morgan, C., Bhavsar, V., Hotopf, M., Hatch, S (2016). Minority status and mental distress: a comparison of group density effects. Psychological Medicine, 46(14), 3051-3059.

Schwartz, R, Blankenship, D (2014) Racial disparities in psychotic disorder diagnosis: a review of empirical literature. World Journal of Psychiatry, 4(4), 133-140.

Spauwen, J., Van Os, J. (2006). The psychosis proneness: psychosis persistence model as an explanation for the association between urbanicity and psychosis. Epidemiology and Psychiatric Sciences, 15(4), 252-257.

Spiegelhalter, D, Best, N, Carlin, B, van der Linde, A (2002) Bayesian measures of model complexity and fit. J Royal Statistical Society, Ser B. 64(4): 583–639

Thomas, D, Zhu, P., Decady, Y (2007) Point estimates and confidence intervals for variable importance in multiple linear regression. Journal of Educational and Behavioral Statistics, 32(1), 61-91.

Thunhurst C (2009) Measuring the health of urban populations: what progress have we made? Public Health. 123(1):e40-4

Timms G (1971) The Urban Mosaic: towards a Theory of Residential Differentiation. Cambridge University Press.

van Os J, Driessen G, Gunther N, Delespaul P (2000) Neighbourhood variation in incidence of schizophrenia: evidence for person-environment interaction. British Journal of Psychiatry 176, 243–248.

Wakefield, J. (2004) Prior and likelihood choices in the analysis of ecological data, Chapter 1 in "Ecological Inference: New Methodological Strategies", eds G King, O Rosen, M Tanner (pp. 13-50)

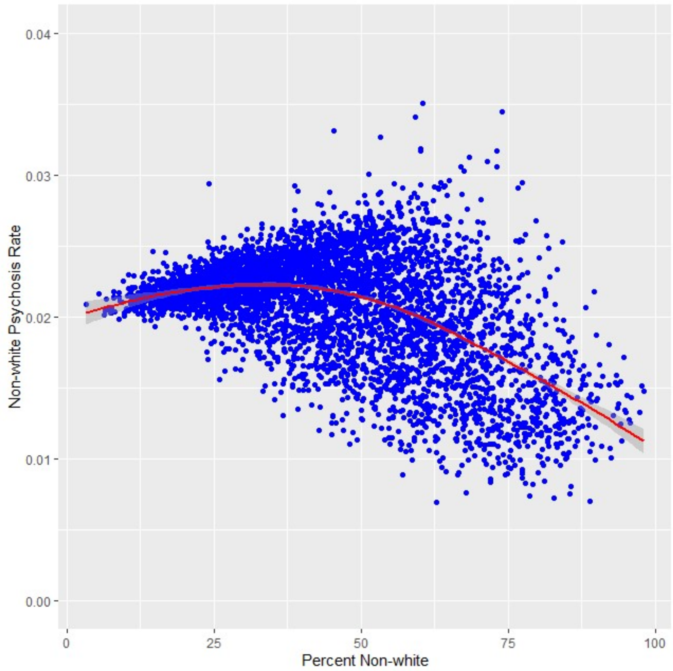

Figure 1a Percent Non-white and Non-white Psychosis Rate, London LSOAs

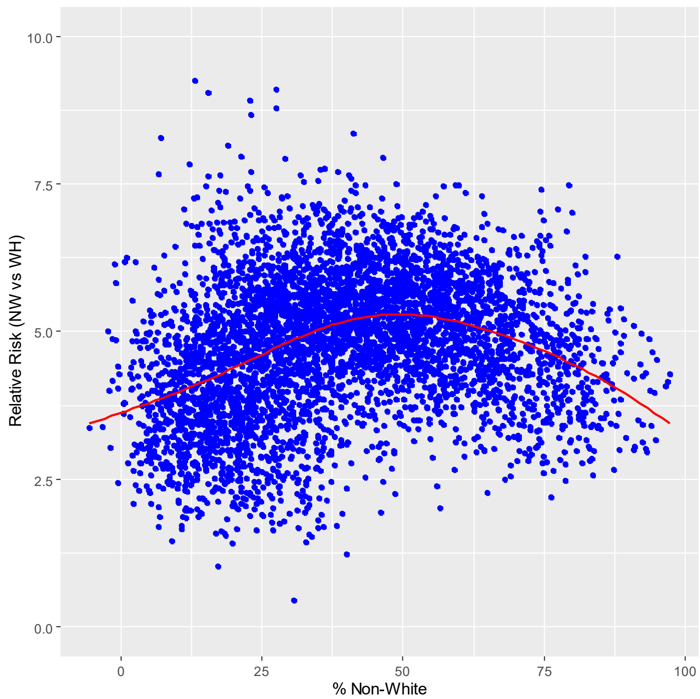

Figure 1b Percent Non-white and Ratio of Non-white to White Psychosis Rate, London LSOAs

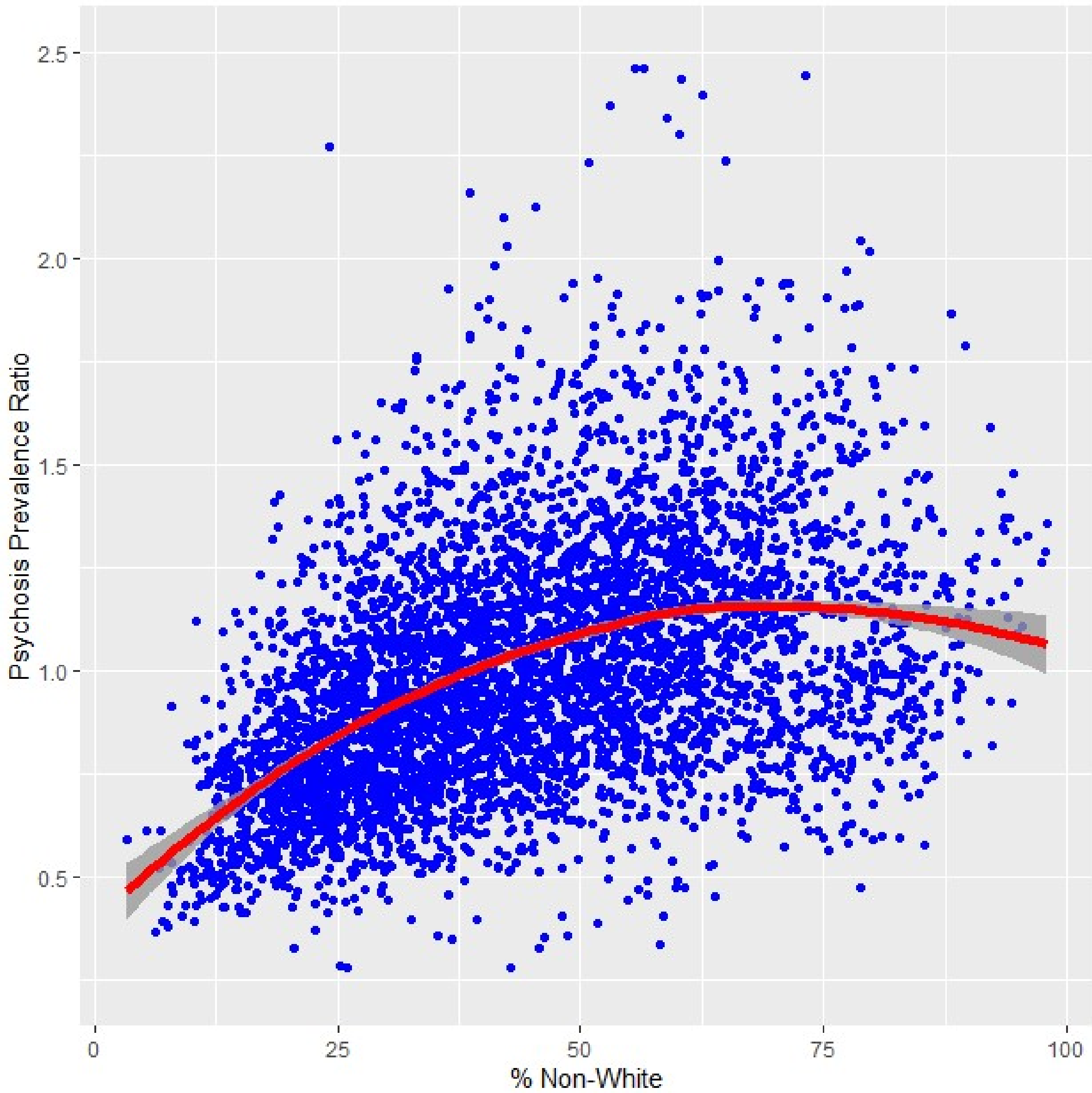

Figure 2 Psychosis Prevalence Ratio (Total Population Prevalence) and % Non-White, London LSOAs

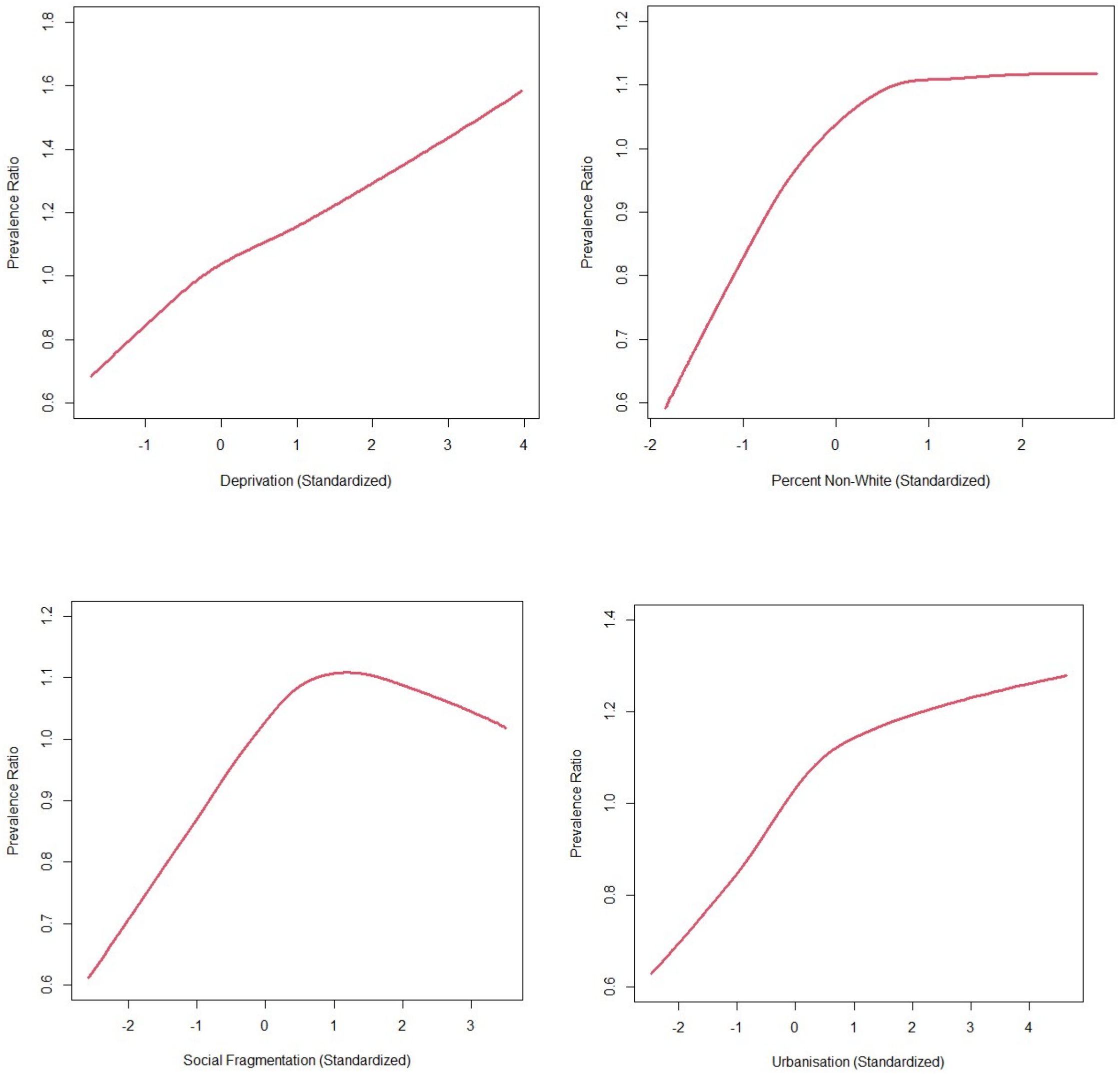

Figure 3. Marginal Dependence, Population Psychosis Prevalence and Neighbourhood Characteristics

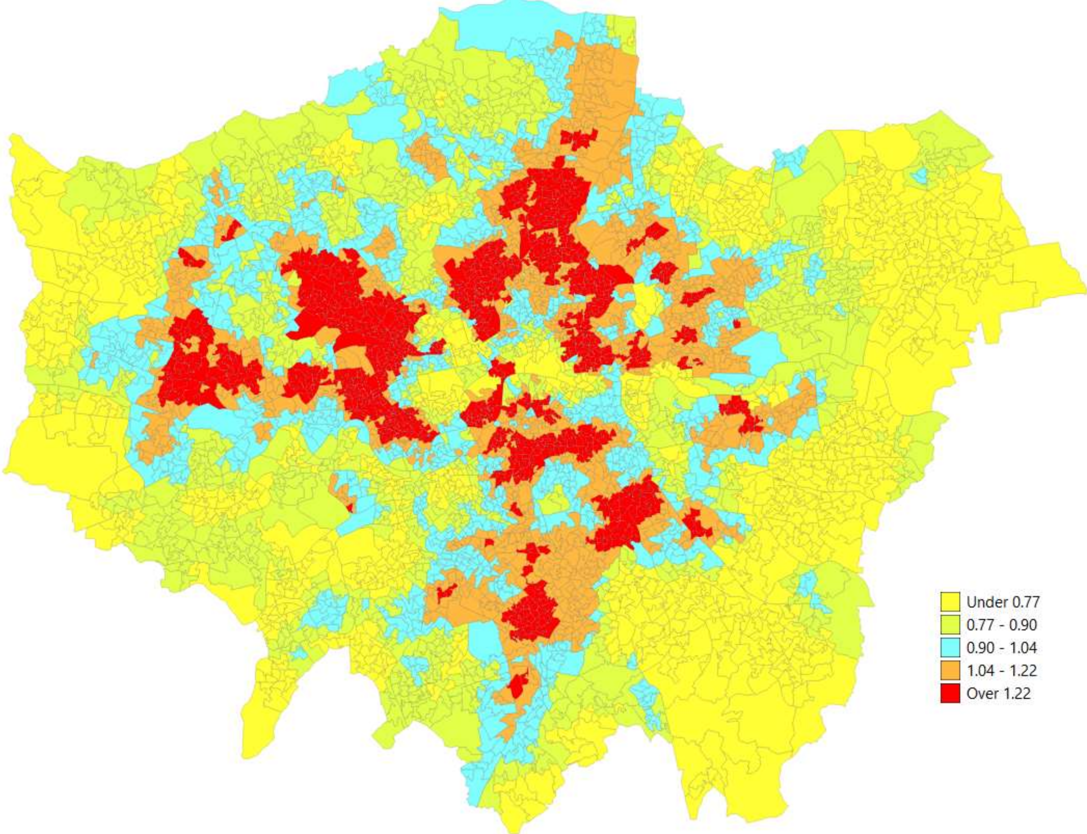

Figure 4 Prevalence Ratios for All Population Psychosis, London LSOAs

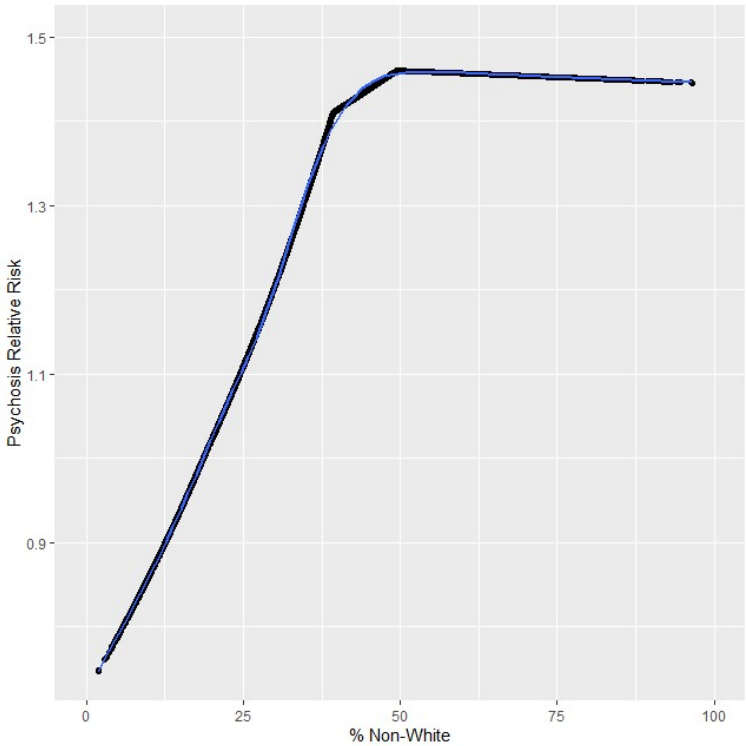

Figure 5a Spatial Regression, Impact of % Non-White on Population Psychosis Relative Risk

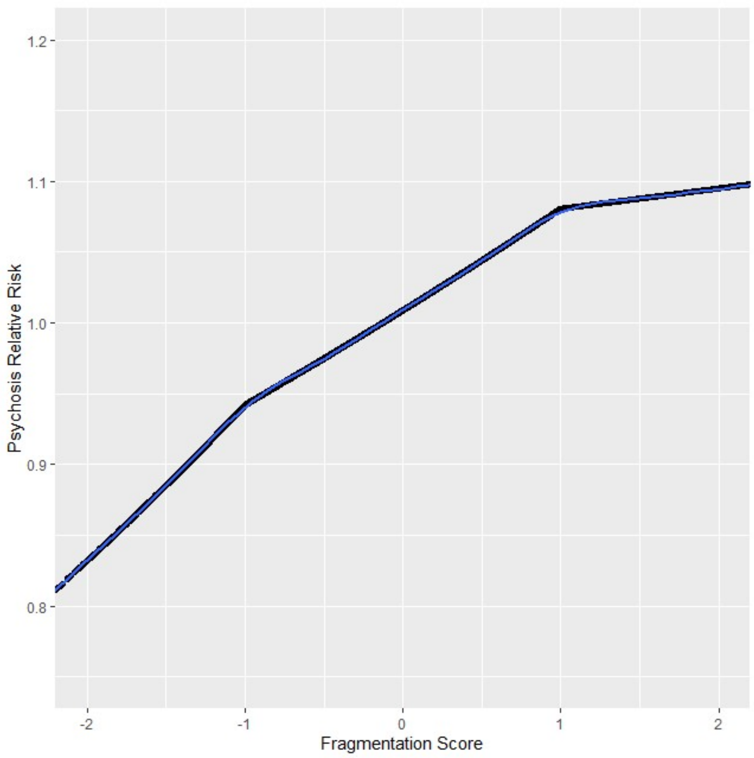

Figure 5b. Spatial Regression, Impact of Fragmentation on Population Psychosis Relative Risk

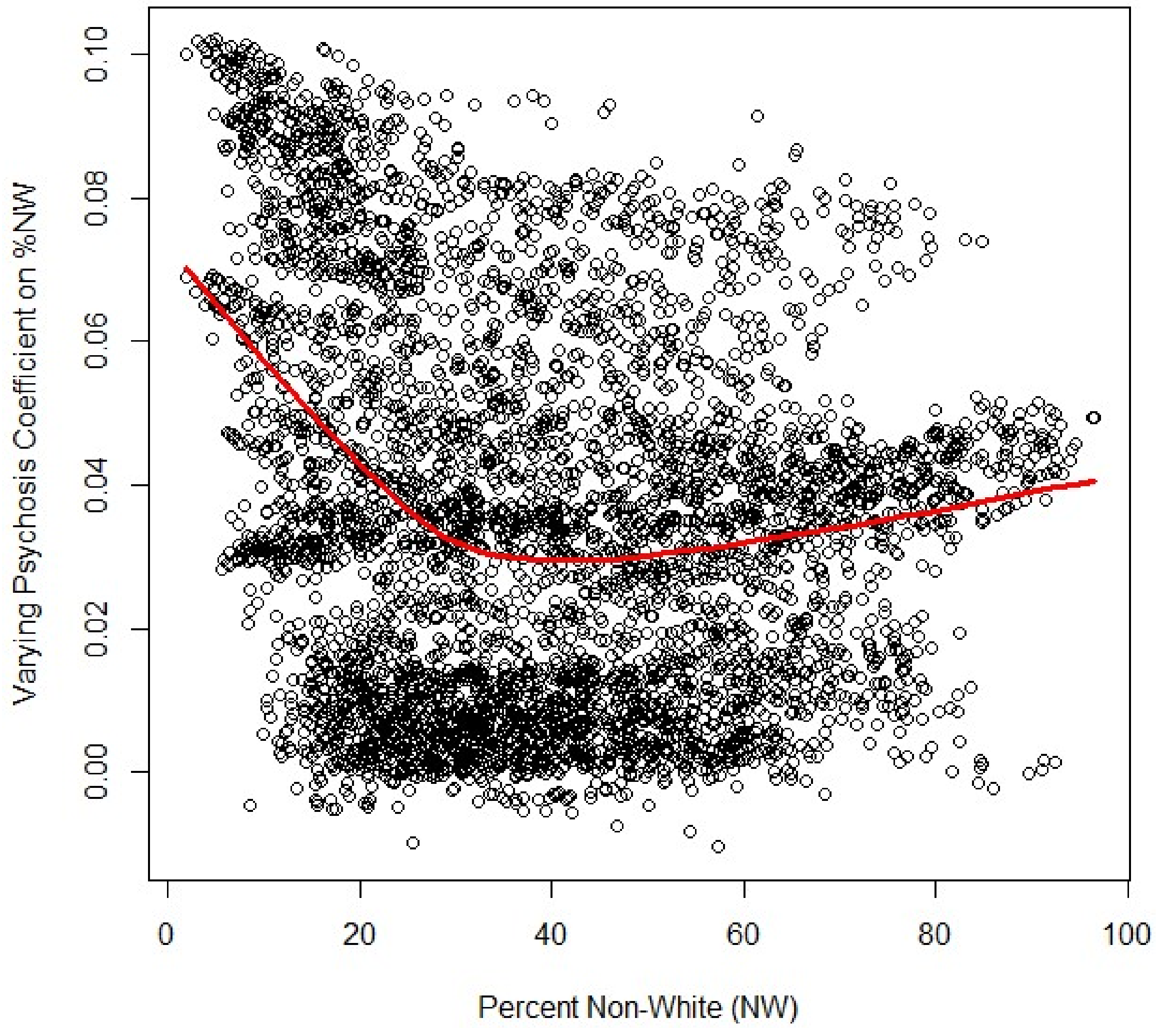

Figure 6. Varying Coefficient Analysis: Varying Impacts of % Non-White on Population Psychosis Risk.