WORKING PAPER SERIES

http://www.casa.ucl.ac.uk/working_papers/paper86.pdf

 ISSN: 1467-1298

86
11 | 04

DEMOGRAPHIC AND DEPRIVATION RATIOS: EXAMPLES OF THEIR USE IN UNDERSTANDING UNDERLYING SPATIAL PATTERNS IN SOCIAL PHENOMENA

Richard Webber

www.casa.ucl.ac.uk

T +44 (0) 20 7679 1782 • T +44 (0) 20 7679 1813 • F +44 (0) 20 7813 2843 • E casa@ucl.ac.uk

Centre for Advanced Spatial Analysis • University College London • 1 - 19 Torrington Place • Gower St • London • WC1E 7HB

Demographic and Deprivation Ratios: examples of their use in understanding underlying spatial patterns in social phenomena.

Richard Webber:
Visiting Professor, Centre for Advanced Spatial Analysis, UCL

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

The intention of this paper is to explore the concept of standardized demographic or deprivation Ratios – what they are, why they might be useful, for what statistical distributions they can be built, how they can be constructed and which research activities and policy areas they might inform.

Such Ratios are designed to demonstrate the extent to which the local levels of various statistical measures are above or below the level that would be 'expected' on the basis of the demographic make up of local areas. They would answer questions such as 'Is unemployment in this town high for a place of this sort?'; 'Is the reason for the high level of vodka consumption in Scotland something to do with local history or local culture or can it be explained as a consequence of the demographics of the Scottish population?' or 'Is the level of burglary in Avon and Somerset above the level that it ought to be, bearing in mind the characteristics of its population?'

The analysis of standardised Ratios is also relevant to the study of regionalization. We are used to the administrative regions in terms of which government divide the country and publish statistics. The mapping of Standardised Ratios shows the extent to which these administrative boundaries correspond to the boundaries of 'natural' regions, these being defined as sets of adjacent areas sharing similar values on a broad range of Standardised Ratios.

Relating the actual levels of social statistics to some measure of what might be expected on the basis of the population is clearly relevant to the evaluation of local performance, whether in the private or the public sector and Ratios of this sort, for example Mortality Ratios, have been used for many years by health professional to benchmark local levels of mortality against the level which might be expected on the basis of the gender and age. However the mapping of the difference between actual and expected rates can often throw interesting light on cultural differences between regions and sub regions of the country which persist despite the homogenizing tendency of central government and national or even multinational retail multiples.

The paper illustrates the potential meaning and use of these Ratios by means of a set of twelve demographic and deprivation Ratios created from the 2001 census in the UK.

Levels of transformation that increase the ease with which local statistics can be used to interpret social processes

Demographic or deprivation Ratios can be viewed as representing one in a hierarchy of transformations that can be applied to databases of local statistics, each step in this hierarchy involving a transformation which improves the analyst's ability to recognize the impact of underlying social processes but demanding a corresponding increase in time spent by the user in understanding the nature of the transformation process..

With the advent of low cost mapping software and with improved physical access to an increasing range of geographically organized statistics, analysts and policy makers are becoming increasingly reliant on maps of social phenomena as inputs into resource allocation, investment decisions and local policy initiatives. Regional development agencies use maps of local unemployment as input to industrial development strategies; health authorities use maps based on medical diagnoses to target health prevention campaigns; police forces map the incidence of various crimes to identify and target areas for additional resources; retailers, property development and car dealers are just some of the commercial organization that rely on the mapping of demographic target groups to ensure their investments are located in areas of highest potential demand for their services.

Such maps as are made to support decisions in these application areas could be made simply by mapping the absolute counts of target groups, showing the numbers of people in specific localities that are unemployed, that have been admitted to hospital, that have been victims of crime, that are professionals or managers. Such maps, though useful in displaying local levels of demand for specific services, provide little understanding of local processes. To understand the reasons for local variations it is necessary to transform raw counts into percentage or per thousand rates by relating their size to the number of persons at risk.

In marketing and public health it is common practice to apply additional transformation to the original data, expressing local rates as a percentage of the national average rate. Such a relative rate, which is often expressed to a national average base of 100, is conventionally referred to by marketers as an ‘index’ value. Examples of a set of index values are to be seen in Table 1 which shows the index values for each Mosaic neighbourhood type on two variables, the proportion of households with access to two cars or vans and the proportion of persons aged 16- 74 who are unemployed. However index values can equally well be used to express the relative levels of such variables by local area as by type of neighbourhood.

Thus the person responsible for the Yorkshire territory of a national car dealership chain would expect to be given a map showing how local variations in rates of car ownership compare with the national average. Under this convention a car ownership rate equivalent to the national average would be indicated by a value of 100, a local rate twice the national average by a value of 200 and a local rate half the national average by a value of 50. Expressing statistical distributions in the form of ‘index’ values places local variations in a national context. Perhaps more importantly it makes it much easier to monitor trends in the pattern in the local concentration of a particular variable for different point in time. This is particularly, relevant in fields such as health, unemployment or house prices where national averages can change quite rapidly within comparatively short periods of time.

Comparing local index values for the same variable for different points of time makes it possible to establish whether the performance of a local area is improving or deteriorating

relative to the national level. Simple transformations of this sort are helpful to the extent that our policy interest is in looking as 'relative' rather than 'absolute' values and to provide information to those who want to 'benchmark' local patterns against some form of external or wider reference value.

The concept of a Standardised demographic Ratio takes this process further by attempting to measure the extent to which the level of some target group within a set of local areas is above or below what one might reasonably 'expect' in such areas bearing in mind the general characteristics of their local population.

The field in which Standardized Ratios are most widely used at the present time is the monitoring of mortality statistics. Given the variation in the likelihood of death between different age classes, it would be quite inappropriate to use crude death rates as a measure of the local effectiveness of different local NHS trusts. As a popular retirement centre it would be surprising if Worthing did not have a very much higher number of deaths per thousand inhabitants each year that Milton Keynes, a city very few of whose residents have yet reached pensionable age.

To convert crude death rates per thousand inhabitants into a statistic which is comparable across NHS trusts serving areas of differing age distributions, the crude death rates for Worthing, Milton Keynes or wherever are compared with the death rates that would have occurred had each inhabitant died with a probability identical to the overall national average for a person of that age and gender. Once this adjustment is made, by standardising the local rate to take into account the age and gender profile of the trust, residents of Worthing can be shown to experience relatively lower likelihood of death at any point in time that the average for Britain as a whole.

Whilst the risk of death varies very significantly by age and gender, it is their occupation group which best predicts the risk of a person aged 16 – 65 being unemployed. Statistics from the 2001 census show that people whose current or previous occupation was a professional or manager were far less likely to be enumerated as unemployed than people whose current or previous occupation was unskilled. It is for this reason that a map showing variations in the local rate of unemployment in a British conurbation looks very similar to a map showing the distribution of people in unskilled occupations. Such a map shows where pockets of unemployed people are concentrated. But this map would not necessarily reveal the local areas where it is difficult to find a job or where a local employer may have recently made workers redundant.

Arguably such a map would be a very valuable input into determining where to site training and re-training facilities. However it would not necessarily highlight localities to which inward investment should be targeted on account of local weakness in the health of the economy.

It is to show which areas of the country are suffering most seriously from weak demand for labour that a 'Standardised' unemployment Ratio would be most useful, a measure which compared actual local levels of unemployment with the levels which would be

'expected' on the basis of their occupational composition. The effect of a standardized map such as this would be to reveal the economic weakness of towns such as Great Yarmouth and Torbay, whose high relative levels of unemployment are concealed by occupational profiles which are much higher than those of traditional unemployment black spots such as Liverpool and Middlesbrough.

Alternative approaches to the calculation of standardized Ratios: Mortality Rates.

Government uses a variety of statistical approaches to help it compare the performance of local service providers in ways which take into account the demographics of the areas that they serve.

We have already described the approached used by public health professionals, Standardised Mortality Rates, to benchmark local mortality rates against an appropriate norm based on the age and gender mix of the local area. This approach is often described as the measurement of the 'compositional effects' underlying variations in a particular distribution.

This method works particularly well in applications, such as mortality, where there are one or two very powerful discriminators, in this case age and gender, which are overwhelmingly more predictive of the observed outcome, death, than any other. A very large proportion of the original variation in crude death rates is accounted for by controlling for age and gender.

However it could be supposed that the 'compositional effects' captured by Standardised Mortality Rates could be made more accurate by adding further variables. One might ask, for example, why not include the effects of variations in occupational groupings or even smoking on the argument that variations in both social class and smoking would narrow the gap between actual and 'expected' mortality levels.

Technically it would not be difficult to extend the calculation of the SMR to include social class. After all information exists on the mortality rates of different social classes within age and gender categories. The number of residents in each NHS Trust in each permutation of social class, age and gender can be established from the 2001 census. By contrast even if we were to be able obtain statistics on the level of smoking by NHS trust area, which might be possible using the results of lifestyle surveys, it would not be possible to incorporate the compositional effect smoking because we do not measure mortality rates of smokers disaggregated by age and gender.

However we should bear in mind that the purpose of the standardized process is not necessarily to 'explain away' the maximum possible proportion of the original variance of a distribution. One of the purposes of a SMR is precisely to highlight inequalities in mortality rates between health trusts resulting from differences in levels of deprivation. Reducing inequalities of health between occupational groupings may therefore be one of

the most effective means of reducing variations in the Ratio. By contrast a policy of reducing inequalities may not be appropriate to the other demographic groups used in the standardization process. For example so far no government has sought to reduce mortality inequalities between age groups or between genders. These differences are perceived as exogenous factors and assumed to be unavoidable.

It is nevertheless worth noting that it is not the purpose of the SMR method to create indicators of the effectiveness of local NHS trusts. That Salford has one of the highest SMRs in England implies that it has a strong justification for discrimination in terms of resource allocation, not that the local NHS management is ineffective. In circumstances where a benchmark system seeks to evaluate the performance of local service management, for instance in the case of the Police Service with Her Majesty's Inspectorate of Constabulary (HMIC) comparison of Most-Similar Forces (MSF), then it becomes important to introduce a much larger number of variables into the evaluation system.

Comparator Analysis

Because of the rather different objectives of the Police Performance Assessment Framework, it uses a very different methodology to that used in calculating SMRs. In this exercise each individual police authority in England and Wales has been compared with each other police authority in terms of its position on a wide range of geographical, demographic and socio economic indicators deemed to be predictive of the local need for policing resources. By means of this process a table has been built which contains a statistical measure of the similarity of every permutation of two policy authority areas based on their similarity in respect of demographic and other relevant characteristics.

From this table it is possible to establish for each individual police authority (other than the City of London Police Force whose territory is considered too different from that of any other Police Force) a set of between four and eight other police authorities which are broadly comparable in terms of their geographic, demographic and socio economic make up. These closest 'neighbours' are then considered 'comparators'. Given that their composition is broadly similar to that of the target police authority on a range of indicators that can be shown to proxy different policing environments, then it is assumed that their crime levels, their clear up rates, their response times should also be broadly similar. This approach we call the use of 'comparators'. It is very similar to the appraisal methods used by surveyors and valuers to form appraise fair prices or rents for commercial or residential properties.

The advantage of the comparator approach over the compositional approach is that it does not require information at a fine level of resolution to be available. To benchmark police forces against each other nor is there is any requirement to establish anything about the relationship between crime, age, gender, occupational group or indeed any other key discriminator[1]. The information required for a comparator evaluation need be

no more than a common set of geographic statistics for the areas being matched, in this case police authorities.

The disadvantage of the comparator approach is that it can only be applied at the level of geography for which the analysis is undertaken. For example whilst a set of comparator forces can be produced for the Devon and Cornwall Police Authority, it is not possible from this exercise to subsequently benchmark the performance of an administrative subdivision of the Authority, such as Plymouth Basic Command Unit, or indeed to set targets for individual wards of Plymouth or indeed for individual police beats. Indeed, the Police Performance Assessment Framework involves a very different and disparate comparator approach to group similar Basic Command Units and Crime and Disorder Reduction Partnerships, but no such analyses have been conducted at finer spatial granularities.

The reliability of the comparator method is sensitive to the number of geographical pieces used in the analysis. Whilst the demographics of Greater London are considered too different from that of any other force for the Metropolitan Police to have comparators, it could be argued that the most rural force in England and Wales, Dyfed – Powys, is hardly a realistic comparator for any other force. Traditionally it has enjoyed lower crime rates and higher clear up rates than its comparators. This is more likely be a consequence of it having a much lower proportion of its population resident in urban centres of any size than of the effectiveness of its Chief Constable.

Multivariate regression

A third approach to benchmarking is the use of various forms of regression. Essentially the approach views the target variable – crime rates, unemployment rates, sickness levels, the attainment levels of pupils in particular schools – as though it were a dependent variable whose levels are dependent upon the observed levels of a number of other variables. Statistical techniques are then used to identify a set of supposedly independent variables from among those available which in combination provide the most reliable 'expected' value of the dependent variable.

Though this methodology works well in a number of statistical applications, it has a number of limitations in this particular one.

One of the limitations of the technique in this particular application is that the requirement for independence between the so called 'independent' variables is seldom met. Most demographic variables that can be accessed at a local authority level, whether from the census or other sources, are highly correlated with each other. The reason why this is problematic is that the decision made by the algorithm to select one variable over another for use in the predictive model can often have a very significant impact on the 'residuals' (the difference between the actual and the predicted values). For example a model to predict local crime levels would include some measure of deprivation. If the algorithm found that sickness was the best proxy for deprivation, then South Wales,

which has a much higher level of sickness than it does of deprivation, would be identified by the model as likely to have a high level of crime.

If on the other hand the algorithm found that the percentage of households living in social or council housing was the best predictor of crime levels, then South Wales, which for historic reasons has very low levels of public housing, would be predicted not to have a particularly high level of crime. Whereas in the first model South Wales's crime would probably have been demonstrated to have been lower than 'expected', the second model would suggest South Wales had a higher than 'expected' level of crime. Quite apart from these different results challenging the validity of the estimate, they open the process of setting targets to political manipulation since different local government areas will have interests in different variable being used.

Given that many of the input variables have distributions at the local level which are significantly affected by regional or historical patterns, the residuals are as likely to be a reflection on local deviations in the expected levels of the 'independent' variables on which the model relies as they are a reflection of exceptional local deviations in the variable one is trying to standardize. This rather defeats the purpose of the exercise.

Another difficulty in the regression approach is that the models are sensitive to the level of geographic resolution at which they are calibrated. Results therefore can only reliably be applied at the level in the geographic hierarchy at which they are built. This effect is illustrated by the example of a model built to predict the Labour share of the vote at ward level using demographic data at the ward level, the purpose of the model being to evaluate to the performance of the party organization in each ward. It may well be the case that Labour systematically achieved more votes than the model predicted in those wards which were surrounded by wards in which Labour could expect to achieve a high share of the vote. Likewise it is quite likely that it would under perform in wards surrounded by others in which Labour could not expect to perform well. If this pattern were to be typical, then a regression model built at the constituency level would have a very different structure to one built at as ward level and the constituencies in which Labour appeared to over perform based on the ward model would be very different from those in which it appeared to over perform based on the constituency model.

Geodemographic composition

The use of geodemographic neighbourhood to create benchmark targets against which local performance can be evaluated incorporates some elements from each of the three approaches considered above, the SMR, the comparator approach and the multivariate regression approach.

The central idea underlying geodemographics is that in modern industrial societies the majority of residential areas conform with a reasonable degree of approximation to a limited number of types of neighbourhood; that most urban areas contain residential

neighbourhoods which are broadly similar to those that one would find within other urban areas; and that these distinct types of neighbourhood do tend to cluster geographically so that some areas of the country are very much richer in certain types than they are in others. If therefore one can define these types of neighbourhood, and if one can reliably assign each unit postcode to the type to which it most closely approximates, then it becomes possible, by appending type of neighbourhood to records of survey respondents, customer files, patient records, crime incidents or of pupil attainments, to identify the variations in average performance of different types of neighbourhood which can then be used to measure 'expected' performance at all higher levels of geographical resolution.

Whereas most classification systems, such as age, gender, occupation, are based on the use of a single dimension, neighbourhood classifications are by contrast multi-variate. Indeed the more different criteria that are available for measuring the similarity of a postcode to different types of neighbourhood the more effective the classification is likely to be in predicting variations in behaviour not included as classification criteria. The classification used in this analysis, Mosaic, makes use of 400 different criteria for judging the similarities between postcodes and the 61 different types of neighbourhoods which it defines. These 400 criteria are taken from the 27 topic headings listed in Table 2.

The composition approach used to create SMRs takes into account age and gender. An equivalent approach can use (Mosaic) neighbourhood type rather than age and gender to create ratios for a wide range of other target groups. The first step in this process involves the calculation of the relative incidence of the target group nationally in each of the 61 Mosaic types. This relationship can be established for a large number of demographic variables using the census itself. Alternatively it can be established for health, education and crime variables by appending Mosaic codes to datasets compiled from administrative records such as the Hospital Episode Statistics (HES) or Pupil Level Annual School Census (PLASC). The relationship can also be established by appending Mosaic to national sample surveys such as the British Crime Survey (BCS). The relative incidence is typically expressed in the form of what is known as a geodemographic profile, as illustrated for the variables 'households with two cars or vans' and 'persons aged 16-74 unemployed' in Table 1.

Once a profile has been created the second step involves the accumulation for each unit of any given geography, for example Primary Healthcare Trusts, Police Authorities, Car Dealership Areas, the percentage number of households (or adults, or persons) resident in each neighbourhood type. The third step involves the multiplication of the proportion of residents in each neighbourhood type in each unit of geography by the relative incidence of the target variables within that type of neighbourhood at a national level. Summing these values it is now possible to calculate a likely level of incidence of the target group within any local area. This method rests on the assumption that residents in each type of neighbourhood have a similar likelihood of belonging to the target group to that of residents in neighbourhoods of matching demographics elsewhere in the country.

The theoretical merits of this approach are as follows. As compared with compositional models which rely on one or two variables only, it incorporates virtually all demographic dimensions likely to have a material impact on the distribution of the target group; however, compared with regression, no single individual variable has a sufficiently large impact on the result of the estimation process as to render it particularly sensitive to systematic local or regional distortions in the expected value of these variables; benchmarks can be created for all units of any set of geographical zones and the result of the estimation process is no longer dependent on the level of geography of the units used to create the estimate. The method can be applied to any data items that can be obtained either from the census or from a national administrative dataset. However the data on which the benchmarks are calibrated do not necessarily need to be available for every region to which the benchmark is applied. For example even though PLASC is collected for English Local Education Authorities only, benchmarks can be generated for Local Authorities or indeed for individual wards within Wales and Scotland also.

Hitherto composition models based on Mosaic have been applied using the results of national market research surveys such as the British Market Research Bureau's Target Group Index. For example the leading door to door distribution companies now offer clients targeting facilities whereby they will drop leaflets or samples exclusively in UK postcode sectors which, on the basis of the mix of their residents by Mosaic neighbourhood type, have the highest likely consumption of a brand or product category included in Target Group Index corresponding to the target group of interest to their clients. Such product categories might include likelihood of purchasing a conservatory, being a heavy user of take away meals or being a regular user of instant coffee. Similar methods are used to evaluate the extent to which residents within the catchment areas of different retail outlets are likely to cite price, quality or convenience as the reason for deciding to shop where they do.

Although the method has been extensively used, no attempts have been made to validate the accuracy of the approach. However with the reconstruction of Mosaic following the publication of the results of the 2001 census the opportunity exists to use census variables themselves to demonstrate the methodology and to test the plausibility of the results.

Standardised demographic and deprivation Ratios from the 2001 census.

To demonstrate the composition effect method of calculating standardized demographic and deprivation Ratios using a multi-variate postcode level geodemographic classification, an exercise was undertaken on a set of twelve statistics taken from the results of the UK's 2001 census. The test that was undertaken uses the 9577 postcode sectors (eg N6 4) into which the Post Office's postcode system divides the United Kingdom.

The variables used in the test are listed in Table 3.

To undertake the test, values were extracted from the census for each of the twelve variables for each of the 9577 postcode sectors. These data were extracted in the form of index values. In other words the values were represented as a rate expressed as a percentage of the UK average rate, such that a level equal to the UK average would be given a value of 100.

The second step involved the extraction of a matching set series of twelve 'profiles', each profile containing a set of 61 index values, these indicating the extent to which the national average level of that type on each of the 12 variables differed from the national average for that variable. This information was also expressed in index form, ie with a value of 100 indicating a level of a variable in a Mosaic neighbourhood type identical to the national average level for that variable.

Each of the twelve profiles was then matched against the distribution of resident population by Mosaic type in each of the 9577 postcode sectors. By multiplying the proportion of the population in each Mosaic type with the corresponding index value for each of the twelve variables it was possible to identify the 'expected' level of that variable in each of the 9577 sectors. The index values for these estimates were then placed alongside the actual index values to as to establish for each sector whether it had an actual value higher or lower than its 'expected' score. By the size of this difference as a percentage of the national average level of that variable we were then in a position to establish the Ratios for each postcode sector on each of the test variables[2].

Before the results were analysed a decision was made to remove from the database all non residential postcode sectors (typically identifying large users of mail) together with all others containing fewer than 300 persons. Due to the absence of census data in the Belfast postal area for some of the chosen variables it was decided to remove Northern Ireland postcodes from the database. A decision was also made to remove those postcode sectors where more than 5% of the population lived in institutions. As a result of these modifications the overall number of sectors used in the study was reduced from 9577 to 8482.

To make maps of the Standardised Ratios for the twelve variables, the distribution of Ratio scores for the 8482 records were classed into eight bands containing equal numbers of postcode sectors. The maps were created using an orange to green palette such that sectors with highest Standardised Ratios are represented in darkest shades of orange, sectors with lowest Standardised Ratios being represented in darkest shades of green.

Deprivation Ratios

The exercise involved the construction of six standardized deprivation Ratios. The variables used to create these Ratios are listed in Table 4 together with their index values for each of the 61 Mosaic neighbourhood types. Four of these involve disadvantaged groups, persons aged 16 – 74 who were unemployed at the time of the 2001 census,

persons with no qualifications, households living at more than one person per room and households consisting of a lone parent with dependent children.

Two others indicate affluent groups, persons in ‘good’ health and households with the use of two cars or vans, low levels of which could also be viewed as an indicator of deprivation.

The national map of standardized unemployment Ratios highlights a number of economically disadvantaged areas which are less visible on maps showing crude unemployment rates, notably the Lincolnshire Wolds and Gainsborough, Great Yarmouth and Lowestoft, the Isle of Wight, West Cornwall, North Devon, much of North West and South West Scotland. By contrast areas of low unemployment that stand out are North Yorkshire and the North Pennines, the inner areas of Leeds and West Yorkshire, North East Lancashire, the Potteries, the upper Severn Valley (Welshpool / Newtown), much of Dorset and the Cotswolds.

The map of London shows that many of its lowest standardized unemployment rates occurring in Tower Hamlets and Newham.

In the North East standardized unemployment rates are very low around the Team Valley trading estate, around Sedgefield and Peterlee but much higher in and around South Shields and Sunderland.

Ratios for persons with no qualifications show very interesting patterns within urban regions. Thus whilst West London has fewer people with no qualifications than would be expected, East London and the Thames Estuary has many more. Likewise South and West Manchester has fewer people with no qualifications than would be expected whilst North and East Manchester has many more. Tyneside has significantly lower Ratios than Durham. In general both the West and East Midlands seem to have unexpectedly large numbers with no qualifications, in particularly the area around the Wash, in the Nottinghamshire / Derbyshire coalfield and in the Potteries. The Welsh valleys and much of rural South Wales also have above average Ratios.

By contrast one finds low Standardised Ratios in many areas of high landscape and amenity value, for instance in the Lake District, the Yorkshire Dales and in particular in the Cotswolds. The scenically attractive Somerset / Dorset border has much low standardized scores than the less interesting Somerset plain.

It would not be difficult to support the proposition that areas with low scores on no qualifications highlight those parts of the country that have a popular image that attracts footloose well educated people such as journalists or consultants working from home.

The standardized overcrowding Ratio highlights the very much higher household densities that are characteristic of Scotland than of England and Wales. Much of London, though interesting not the London Borough of Lambeth, also has higher than expected proportions of households living at densities of over one person per room. Low

Standardised Ratios by characterize much of South Wales and of the Nottingham / Derbyshire and South Yorkshire coalfield. Low Standardised Ratios are also typical of that part of East Anglia that lies east of the Fens.

Ratios for lone parents with dependent children tend to be higher in the western half of Britain than in the east. Particularly high rates occur in rural Wales, in Greater Manchester, Merseyside, Fylde and the Lancaster area. Most of Outer London also has higher than expected proportions of lone parent families. By contrast low rates are characteristic of most of the West Yorkshire conurbation, and of the Yorkshire, Nottinghamshire and Derbyshire coalfields, Tyneside, Norfolk and Cambridgeshire. Figure 1 illustrates a clear cultural boundary between Birmingham and the Black Country with Birmingham sectors having high Standardised Ratios, the Black Country sectors very low Standardised Ratios.

Good Health Ratios are high throughout the South East (other than in London and along the coast) and the South Midlands. They are also high in the East Midlands south of the Trent, in North Wales (other than in the seaside towns) and in most of the uplands North of England. Low levels of good health, by contrast are focused in four regions, in an area crossing industrial South Wales and extending into Herefordshire and the Forest Dean, in the Yorkshire / Nottinghamshire coalfield (see Figure 2), in most of the North East, including the Durham Pennines and the Cleveland coast, and in the vicinity of The Wash. Poor health Ratios characterise most of rural Britain as well as rural Cornwall and North and West Devon. By contrast central and inner London and most of outer London is characterized by high standardized good health Ratios.

Ratios for two car ownership tend to divide the country along urban / rural lines, with lower than expected Ratios being a particular feature of suburban London, Merseyside and Greater Manchester, Tyneside and Teesside. York and Blackpool also have unusually low levels of two car ownership, as does much of rural Wales and rural Lincolnshire.

By contrast higher than expected levels of two car ownership characterize the more affluent, recently built commuter areas in the South East, Midlands and Cheshire.

Demographic Ratios

The standardized demographic Ratios that have been created apply to two types of house, terraced and detached, to being a Christian, to two age group, aged 20 – 24 and aged 65+, and to belonging to the ‘A/B’ social grade.

The Index values of the 61 Mosaic neighbourhood types on each of these variables are shown in Table 5.

Higher than expected proportions of terraced houses are particularly a feature of the Pennines, both the historic textile communities of Lancashire and the West Riding but

also the rural North Pennines and Lake District. The village of Tebay in Cumbria, Figure 5, is a typical example of a rural community which has many more households living in terraced housing than would be the case in East Anglia for example. Terraces are also relatively common in much of the South Midlands and South West of England. By contrast this housing type is relatively uncommon in East Anglia, even after taking into account the rural character of the region, particularly compared with rural areas in North Wales. The Black Country and the East Midlands coalfield are areas in which terraced housing is relatively uncommon.

The pattern in South Wales highlights a very nice distinction between the earlier, Cardiff facing mining communities and the more recent industrial communities based on anthracite mining and tinplate manufacture, which face Swansea.

The Standardised Detached Ratio is the mirror of the standardized Terraced Ratio, with detached houses being especially common in areas surrounding the Wash, in Lincolnshire, Leicestershire, Nottinghamshire and South Yorkshire. The Welsh Marches and much of North East Wales also record high Ratios. By contrast Figure 3 shows particularly low Standardised Ratios occur throughout the rural Pennines and the Peak District as well as in the historic Pennine textile towns. Lower than expected proportions of detached houses are also a feature of most of the South East England and of the Midlands as far North as Leicester and Birmingham.

The population describing itself as Christian at the time of the census is significantly higher than would be expected in the North West of England and along the eastern side of the country from Peterborough as far North as Newcastle. Figure 6 shows an example of the type of street in inner Liverpool where Christians are much more numerous than they are in equivalent inner areas in other parts of the country, even after controlling as the method does for differences in ethnic origin. For a variable with as high an average rate as the percentage Christian, an index value of 116 represents a very significant difference from the expected score and may perhaps be a consequence of the traditional sectarian identification of Orange and Catholic communities on Merseyside. Outside Birmingham it is also strong in the West Midlands. By contrast people living in Wales are less likely to declare themselves to be Christian than would be expected, especially those living in the older South Wales valley communities. South East London has average expected proportions of Christians by contrast to the North and West of the city whose low Ratios are the result of being home to large Jewish as well as affluent Asian populations.

Outside London lower than expected levels of Christians are found in the large cities of the East Midlands, Leicester and Nottingham, and in Bradford, Leeds and Sheffield. Low Ratios are also recorded in cities without significant immigrant populations such as Brighton and Hove, Norwich, Oxford and Cambridge. Rural East Anglia east a line between Cromer and Ely and the rural South West have lower than expected levels of Christians, particularly when considering the density of their medieval churches.

The distribution of the standardized Ratio for the age group 20 - 24 is the most random of the twelve, with there being no clear regional pattern at all. North Wales is the only part

of the country to have consistently larger proportions of people 20 - 24 than might be expected. Within cities such as Bristol, Leeds and Oxford, those sectors closest to their universities have high standardized scores but other sectors not far away have lower than expected proportions of this age group.

By contrast the standardized Ratio of the age group 65+ clearly isolates environmentally attractive areas of the country that older people would like to retire to, the rural triangle contained by Exmouth, Yeovil and Weymouth, the Welsh Marches, East Lincolnshire, the Sussex coast and the Norfolk and Suffolk coasts. Pensioners, by contrast, are less common than expected in most areas of recent population growth, particularly the M4 / M3 silicon corridors, the M11 and the areas around Bedford, Milton Keynes and Aylesbury.

Social Grade AB Ratios are characteristic of areas that have become more attractive than they previously were to high income groups, very often places with a shortage of older high status neighbourhoods. Key concentrations occur in the Western half of London, around Reading, around Warwick and Rugby, in the Cheshire commuter belt and in most of the stone built Pennine textile towns. Much of central Scotland has high scores on this Ratio. By contrast the social AB group is lower than one would expect given the mix of neighbourhoods in and around Bournemouth, in Cornwall and North West Devon, though much of Wales and in most of Southern Scotland. Very low Ratios occur in the area around the Wash, in Lincolnshire and much of East Yorkshire. As is shown in Figure 4 very low Ratios also occur among the inter war suburbs of East London, along the Lea Valley and along both side of the Lower Thames.

Properties of Ratios

These brief summaries of the geographical distribution of ratios may or may not be interesting in their own right. But they illustrate a number of interesting difference ways in which Ratios can vary over space.

There are a number of instances where the Ratios provide clear evidence of local cultural divides. For instance the boundary between the city of Birmingham and the Black Country is very clearly delineated by the map of ratios for Lone Parents. Likewise the Standard Ratio for Terraced Houses provides a very clear delineation between the older South Wales mining communities built around the mining of steam coal and the later anthracite mining communities. The distinctiveness of the Fens and the surrounding areas of Norfolk and Lincolnshire is clearly visible from many of the maps. Postcode sectors in this part of the country tend to have Standardised Ratios which are very similar to each other and which are typically very high or very low.

Whilst the cultural boundary between Scotland and England is sharply delineated in many of the statistics, the same can not be said about Wales. In respect of housing, North Wales, Cheshire and Shropshire share a common set of local characteristics. In respect of

sickness, industrial South Wales shares many similarities with Herefordshire and the Forest of Dean.

Some patterns, such as in respect of Terraced Houses, have a distinct regional pattern, with very large numbers of contiguous postcode sectors deviating from the national average in the same direction. There are other patterns, such as that of the Christian Religion, which, though regional in character, could more appropriately be described as sub regional. For some of these patterns, such as lone parents, the sub regional element of the pattern reveals a shared economic history, for example a past reliance on coal mining.

In other instances the pattern of Ratio scores reflects the geography of a characteristic which is inadequately captured by Mosaic notwithstanding its multi-dimensional nature. For example the distribution of 2 car households is related to the extent to which residents in different rural areas have access to office jobs in nearby employment centres and the distribution of the social grade 'A/B' to the attractiveness of the local landscape.

There are also instances of where the pattern reflects the distribution of a very specific target group which is inadequately captured in the Mosaic classification. For example whilst Mosaic included a number of ethnic neighbourhood types, it is evident that the spread of Asian people across all types of neighbourhood in Leicester and Bradford, not just those dominated by Asians, contributes to the lower than expected proportions of people describing themselves as Christian. A similar pattern applies to the distribution of persons aged 65+. It is evident that in areas that are attractive to the elderly one finds not just a high proportion of the population living in retirement enclaves but pensioners being disproportionately more numerous in other types of neighbourhood.

Notwithstanding all these different spatial patterns it is interesting to note that there is one of the twelve Ratios, persons aged 20 -24, which appears to have no significant regional or sub regional pattern to it.

An evaluation of Standardised Ratios as a modelling tool

In order to evaluate the accuracy of the geodemographic composition method, we undertook a simple analysis of the extent to which the actual level of the twelve variables could be predicted from the Mosaic profile of the area using the geodemographic composition method. Table 6 shows that on average the correlation between the actual levels of the twelve variables and their Standardised Ratio was $R = +0.894$[3].

The geodemographic composition method worked particularly well for predicting levels of two car ownership, the proportion of social group AB, the proportion of houses that are detached and the proportion of the population without education qualifications. By contrast the method was least effective in predicting the proportions of Christians and the distribution of terraced houses.

The actual average size of the correlation is probably larger than many people might have expected and suggests that the methodology is probably a safe method for the regional estimation of those products whose distribution is only available from sample surveys and whose distribution is thought to be dependent on demographics rather than to regionality. To a very small degree there is a slight circularity in this test with census variables since these variables have of course themselves been used as contributory criteria for the measurement of similarity in the clustering process.

However it should be borne in mind that each individual variable is just one of a set of 400 variables used to build the classification. Whilst it is possible that the score of a postcode on such a variable could have influenced the Mosaic classification that it was given, and hence affected the estimated value in the direction of the actual, the relative influence of any individual one out of the 400 variables used will seldom be sufficient on its own to alter the cluster code allocated to any but a very small number of postcodes[4].

Typically one would suppose that the distribution of consumptions not covered by the census would correlate at a level similar to r=0.891. Nonetheless it is important to recognize that not all behaviours are equally well predicted by this method and that it would be prudent only to apply the method to behaviours whose regional distribution is not partly the result of local historical or geological factors. For example it would be a relatively straightforward exercise to calculate the Standardised Ratio for a behaviour included in a national research survey at the level of the standard region and to then compare this Standardised Ratio with the observed value of that behaviour by standard region to identify whether there are significant regional effects (as with porridge for example) which are over and beyond those that can be attributed to demographic differences.

The maps of the Ratios show very clear variations in the case of certain variables, such as terraced houses, much more fragmented distributions in others, such as aged 20 – 24. It would be possible and indeed a very interesting exercise to calculate the level of autocorrelation of each of these patterns, thereby measuring the extent to which it is local regional factors that explain that part of the variability in the original distribution which is not explained by the demographic make up of the postcode sector. A key point to bear in mind is that when the geodemographic composition method is used to predict the level of variables whose errors are randomly (rather than systematically or regionally) distributed, errors will tend to cancel themselves out once postcode sector estimates are aggregated to larger geographic units, such as the catchment areas of shopping centres or areas under the control of particular police forces. By contrast where Ratios are highly regionalized, local errors in estimation are likely to be compounded rather than eliminated by aggregation of smaller units into larger ones.

One interesting line of enquiry would be to attempt to define natural or cultural regions using measurements of the similarities of the Ratio values for contiguous postcode sectors or other geographies. As a rule this method of defining regions is likely to be more reliable than that based on the measurement of actual values. This is because it will remove differences within a region which are the result of highly localized variations in

social status. For example the values for Beverley in East Yorkshire on demographic variables are likely to very different from those of its less affluent near neighbour Hull. In terms of actual values the two towns are dissimilar. However it would not be surprising Beverley's values on the twelve Standardised Ratios calculated in this test were in fact quite similar to those of Hull's, in which case it would be a demographic barrier rather than a regional or cultural one which separated the two towns.

There is no reason why, taking these twelve Ratios, we should not create a measure which expresses how similar or different adjacent postcode sectors may be and therefore builds up a representation which highlights higher than average local differences in these values. Such a representation could be a very useful device for delineating cultural fault lines, an activity which is not irrelevant to debates on how existing administrative areas should be divided or grouped together into larger entities. Measuring the similarities of the Ratios in Cumbria to those of the rest of North West and again to those of the North East of England could make a useful contribution to the debate over which of these two regions it would be more appropriate for Cumbria to become a part of.

If the Standardised Ratios are effective at eliminating those spatial variations in a key distribution which are the consequence of the demographic make up of an area, then it may be instructive to measure the correlations between different Standardised Ratios. For example the correlation between the actual proportions of terraced houses and detached houses at the postcode sector level is R = -0.5592. Clearly where one variable is high the other is low. We might expect that a significant proportion of the variation in the percentage of houses that were terraced and detached would be explained by variations in demographics, such as the status of the neighbourhoods and their degree of rurality. The fact that the correlation coefficient created by comparing the distribution of the Standardised Ratios is negative at a level of R = -0.329 suggests quite a high level of similarity in the regional component which underlies both these distributions. Clearly there is an element in each of the two distributions which is related to some further underlying distribution, which is not captured by the demographic factors used in building Mosaic. The ability of Ratios to delineate the boundaries of what were once coal-mining communities is a good example of how such an 'external' factor could be identified which explains some of residual left over from direct correlation of different variables. Such a factor ought logically to be considered as an additional factor in any formula used to set local performance benchmarks.

The pattern of association between terraced and detached houses is very different from the pattern which occurs when we correlate the level of unemployment with the proportion of households with two cars or vans. The overall correlation of these two distributions, r =0.745, is even stronger than the correlation between terraced and detached houses, r = -0.559. However when we remove the effect of demographic composition and correlate the Standardised Unemployment Ratio with the Standardised two car ownership Ratio, the correlation drops to R = -0.148. That an area has a higher than expected unemployment rate provides no predictive insight into telling whether it has higher or lower levels of two car ownership than we would expect. By contrast there is a rather stronger relationship between the Standardised Ratio for good health and the

Standardised Ratio for two car ownership, R = +0.282. These two factors either seem to have a slight influence on each other or otherwise are commonly influenced by a factor not included in Mosaic, which we might hypothesise would be a measure such as the age of the housing or the level of population density in the immediate vicinity. In summary therefore the use of Standardised Ratios may be useful for identifying that part of the linkage between distributions that is caused by variations in the geodemographic composition of areas and that which is exogenous to a purely demographic model.

Figure 1: Standardised Ratios: Lone Parent Households, Birmingham and Black Country

Figure 2: Standardised Ratios: Good Health, North Midlands and North of England

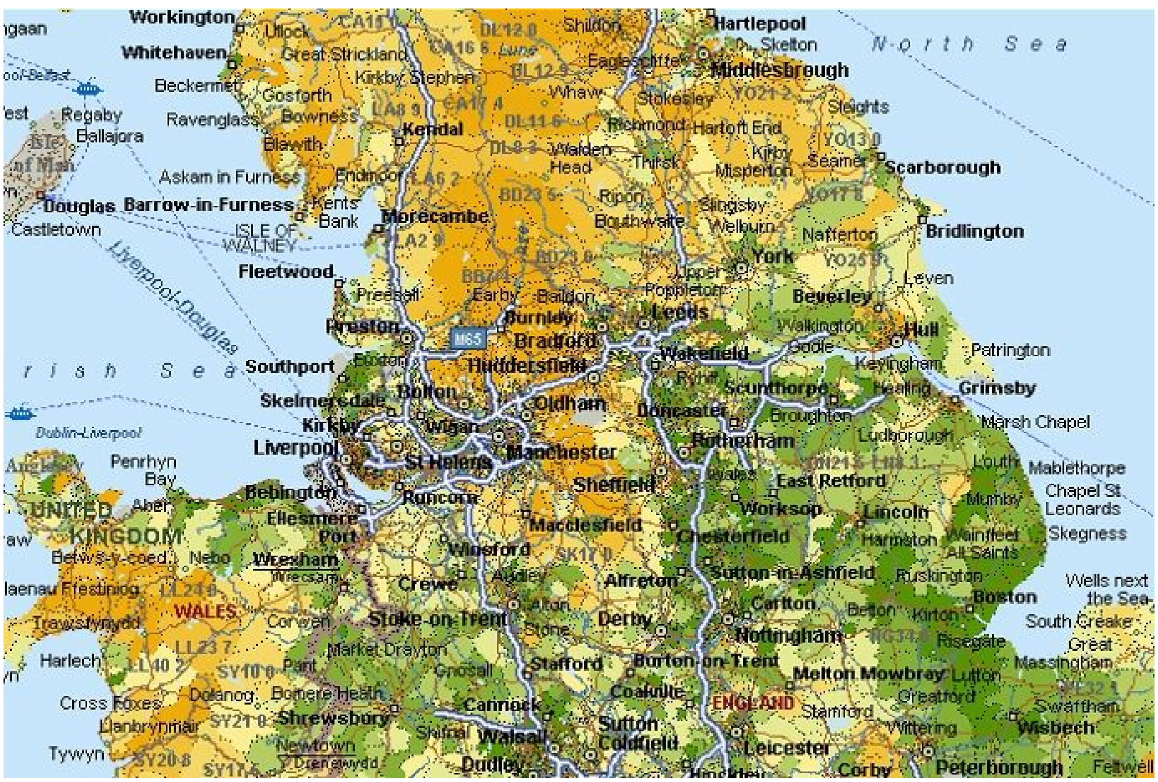

Figure 3: Standardised Ratios: Terraced Houses, North Midlands and North of England

Figure 4: Standardised Ratios: Social Grade AB, London and Thames Estuary


Figure 5: Tebay, Cumbria (CA10 3): The Standardised Ratio for Terraced Houses is 136.

Figure 6: Liverpool 8 (L8 4) : The Standardised Ratio for Christians is 116

		Households with two cars or vans	Economic status : Unemployed
	GB average rate (as % of base count)	**22.842**	**3.562**
Mosaic neighbourhood type		**Index**	**Index**
Code	**Descriptor**	(GB=100)	(GB=100)
1	Global Connections	50	95
2	Cultural Leadership	123	65
3	Corporate Chieftains	207	43
4	Golden Empty Nesters	193	42
5	Provincial Privilege	154	49
6	High Technologists	207	46
7	Semi-Rural Seclusion	190	45
8	Just Moving In	104	102
9	Fledgling Nurseries	179	52
10	Upscale New Owners	233	45
11	Families Making Good	161	54
12	Middle Rung Families	147	64
13	Burdened Optimists	108	83
14	In Military Quarters	112	69
15	Close to Retirement	162	50
16	Conservative Values	124	51
17	Small Time Business	154	55
18	Sprawling Subtopia	129	60
19	Original Suburbs	136	60
20	Asian Enterprise	102	114
21	Respectable Rows	97	75
22	Affluent Blue Collar	113	71
23	Industrial Grit	93	90
24	Coronation Street	49	150
25	Town Centre Refuge	49	166
26	South Asian Industry	42	202
27	Settled Minorities	54	151
28	Counter Cultural Mix	30	162
29	City Adventurers	43	105
30	New Urban Colonists	87	79
31	Caring Professionals	75	92
32	Dinky Developments	80	80
33	Town Gown Transition	57	88
34	University Challenge	62	44
35	Bedsit Beneficiaries	25	178
36	Metro Multiculture	23	223
37	Upper Floor Families	32	225
38	Tower Block Living	10	323
39	Dignified Dependency	15	252
40	Sharing a Staircase	17	225
41	Families on Benefits	39	234
42	Low Horizons	35	214
43	Ex-industrial Legacy	37	159
44	Rustbelt Resilience	63	134
45	Older Right to Buy	82	103
46	White Van Culture	68	97
47	New Town Materialism	74	138
48	Old People in Flats	16	115
49	Low Income Elderly	46	97
50	Cared for Pensioners	58	106
51	Sepia Memories	56	65
52	Childfree Serenity	77	75
53	High Spending Elders	134	46
54	Bungalow Retirement	83	54
55	Small Town Seniors	97	76
56	Tourist Attendants	82	88
57	Summer Playgrounds	129	80
58	Greenbelt Guardians	187	49
59	Parochial Villagers	152	69
60	Pastoral Symphony	172	61
61	Upland Hill Farmers	172	62

Table 1: Index values for 61 Mosaic neighbourhoods on two census variables

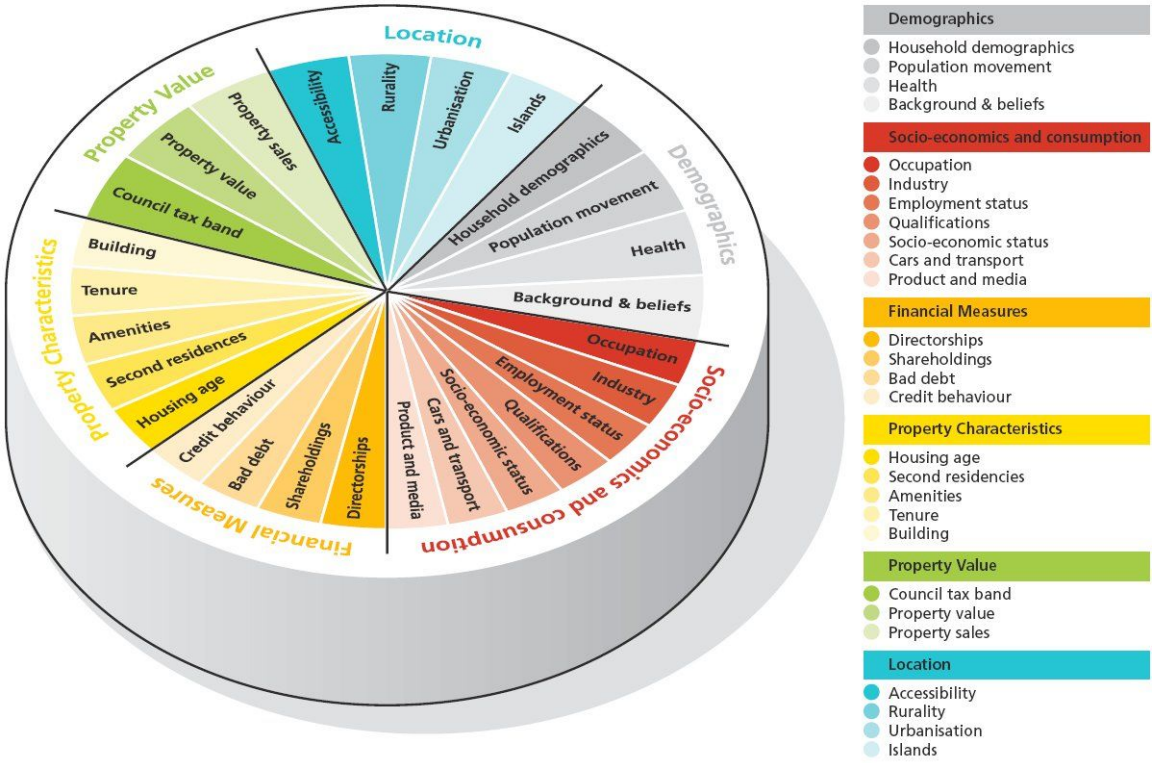

Table 2: Mosaic UK data sources

Variable	Denominator	GB average rate (%)
Deprivation Indicators		
Health : Good	Persons	67.748
Household composition : Lone parent with dependent children	Households	6.539
Qualifications : None	Persons in employment	30.213
Persons per room : >1	Households	1.908
Cars and vans : 2	Households	22.842
Employment status : Unemployed	Persons aged 16 - 74	3.562
Demographic Indicators		
Age : 20-24	Persons	5.998
Aged : 65+	Persons	16.737
Religion : Christian	Persons	71.135
Type of house : Detached	Households	22.184
Social grade : AB	Adults	21.258
Type of house : Terraced house	Households	25.563

Table 3: Demographic criteria used to construct the Mosaic neighbourhood classification system (Ratios have been calculated at postcode level)

Mosaic Neighbourhood Type	Good health	Lone parents with dependent children	No qualifications	Over 1 persons per room	Two cars or vans	Unemployed
Global Connections	115	38	29	146	50	81
Cultural Leadership	114	48	38	71	123	57
Corporate Chieftains	117	33	42	27	207	40
Golden Empty Nesters	110	29	52	17	193	40
Provincial Privilege	108	43	61	28	154	45
High Technologists	116	49	52	25	207	37
Semi-Rural Seclusion	112	41	57	23	190	39
Just Moving In	105	92	80	118	104	99
Fledgling Nurseries	119	72	46	36	179	38
Upscale New Owners	122	57	43	26	233	34
Families Making Good	112	75	67	45	161	43
Middle Rung Families	110	83	81	65	147	51
Burdened Optimists	107	131	84	82	108	67
In Military Quarters	125	56	31	77	112	51
Close to Retirement	107	45	76	27	162	43
Conservative Values	97	41	104	24	124	49
Small Time Business	104	52	87	31	154	49
Sprawling Subtopia	105	69	90	53	129	51
Original Suburbs	110	65	63	60	136	51
Asian Enterprise	103	93	97	410	102	108
Respectable Rows	104	89	80	60	97	61
Affluent Blue Collar	98	65	115	48	113	65
Industrial Grit	100	103	112	69	93	79
Coronation Street	95	155	123	95	49	142
Town Centre Refuge	90	91	101	69	49	155
South Asian Industry	98	145	154	727	42	253
Settled Minorities	102	163	93	347	54	143
Counter Cultural Mix	101	120	70	291	30	146
City Adventurers	113	46	32	147	43	81
New Urban Colonists	113	63	42	86	87	63
Caring Professionals	106	83	58	73	75	79
Dinky Developments	108	82	58	78	80	61
Town Gown Transition	109	69	43	117	57	102
University Challenge	115	55	22	160	62	74
Bedsit Beneficiaries	96	69	74	122	25	161
Metro Multiculture	97	244	114	586	23	240
Upper Floor Families	89	219	142	129	32	244
Tower Block Living	72	140	157	107	10	405
Dignified Dependency	83	320	175	277	15	329
Sharing a Staircase	72	120	168	91	17	287
Families on Benefits	98	393	151	252	39	266
Low Horizons	90	274	173	171	35	252
Ex-industrial Legacy	83	160	170	102	37	184
Rustbelt Resilience	91	147	159	127	63	138
Older Right to Buy	97	149	126	137	82	96
White Van Culture	88	92	149	71	68	101
New Town Materialism	99	216	137	184	74	130
Old People in Flats	57	43	187	41	16	185
Low Income Elderly	79	75	145	67	46	107
Cared for Pensioners	86	97	135	77	58	112
Sepia Memories	81	33	94	27	56	71
Childfree Serenity	98	51	65	44	77	67
High Spending Elders	98	31	78	15	134	49
Bungalow Retirement	83	31	136	20	83	65
Small Town Seniors	94	69	106	41	97	72
Tourist Attendants	92	54	92	41	82	89
Summer Playgrounds	99	46	92	42	129	81
Greenbelt Guardians	110	40	71	30	187	43
Parochial Villagers	102	57	99	46	152	64
Pastoral Symphony	106	46	87	39	172	54
Upland Hill Farmers	106	43	96	43	172	55
Unclassified	100	100	100	100	100	100

Table 4: 'Index' values for Mosaic Neighbourhood Types on six measures of deprivation / affluence

Mosaic Neighbourhood Type	Aged 20-24	Aged 65+	Christian	Detached house	Social grade - AB	Terraced house
Global Connections	132	78	80	10	213	52
Cultural Leadership	96	94	84	85	216	78
Corporate Chieftains	68	96	100	332	233	16
Golden Empty Nesters	56	131	110	338	202	18
Provincial Privilege	68	126	107	192	171	30
High Technologists	75	60	109	319	193	22
Semi-Rural Seclusion	58	99	109	272	187	42
Just Moving In	138	79	94	111	125	85
Fledgling Nurseries	99	29	102	177	164	100
Upscale New Owners	55	30	108	343	203	28
Families Making Good	77	57	107	169	132	78
Middle Rung Families	94	52	108	99	104	89
Burdened Optimists	111	52	100	61	86	163
In Military Quarters	181	8	109	55	128	153
Close to Retirement	76	99	112	204	139	31
Conservative Values	62	146	115	157	100	31
Small Time Business	65	107	111	224	120	51
Sprawling Subtopia	74	96	109	61	101	78
Original Suburbs	83	90	99	81	154	95
Asian Enterprise	131	68	56	35	93	150
Respectable Rows	100	82	99	49	105	168
Affluent Blue Collar	82	100	112	91	76	69
Industrial Grit	90	81	106	48	66	179
Coronation Street	126	76	97	17	52	270
Town Centre Refuge	140	116	94	29	69	93
South Asian Industry	165	51	34	24	49	239
Settled Minorities	139	61	76	17	94	163
Counter Cultural Mix	153	66	76	7	125	55
City Adventurers	216	47	75	6	192	42
New Urban Colonists	116	65	85	30	191	145
Caring Professionals	227	67	81	24	134	215
Dinky Developments	163	52	90	27	118	118
Town Gown Transition	559	50	72	19	106	157
University Challenge	761	25	71	36	91	132
Bedsit Beneficiaries	256	66	74	7	90	29
Metro Multiculture	129	62	77	6	65	40
Upper Floor Families	137	82	89	17	41	117
Tower Block Living	137	117	88	4	34	17
Dignified Dependency	117	69	91	8	24	39
Sharing a Staircase	105	152	97	12	32	68
Families on Benefits	115	40	86	25	36	156
Low Horizons	100	74	100	23	30	159
Ex-industrial Legacy	86	130	105	22	32	138
Rustbelt Resilience	89	92	104	28	38	148
Older Right to Buy	91	99	99	31	62	151
White Van Culture	73	140	108	44	48	133
New Town Materialism	102	59	96	29	48	175
Old People in Flats	49	321	109	19	26	53
Low Income Elderly	74	201	105	32	50	95
Cared for Pensioners	80	164	105	43	55	111
Sepia Memories	65	282	107	58	92	48
Childfree Serenity	101	160	96	51	137	70
High Spending Elders	46	201	113	280	139	29
Bungalow Retirement	43	237	116	186	69	26
Small Town Seniors	70	146	108	113	86	90
Tourist Attendants	72	182	106	111	88	96
Summer Playgrounds	54	152	105	264	95	49
Greenbelt Guardians	58	100	110	287	153	32
Parochial Villagers	61	109	109	241	97	49
Pastoral Symphony	57	102	108	299	109	33
Upland Hill Farmers	57	101	107	336	87	23
Unclassified	100	100	100	100	100	100

Table 5: 'Index' values for Mosaic Neighbourhood Types on six demographic variables from the 2001 census.

Correlation between actual level and estimated level : postcode sectors	Variable
0.968	Cars or vans : 2
0.955	AB Higher & intermediate manager/admin/prof
0.929	Type of house : Detached
0.920	Qualifications : none or unknown
0.913	Age : 20-24
0.898	Age : 65+
0.898	Occupancy : Over 1 person per room
0.887	Economic status : Unemployed
0.879	Lone parents with dependent chidren
0.870	Health : Good
0.826	Religion : Christian
0.787	Type of house : Terraced
0.894	***Average***

Table 6: Correlation between estimated and actual values : postcode sectors.

[1] The variables which were used as the basis for calculating similarities between police forces were selected on the basis of their known correlation with local levels of crime.

[2] The form of the Standardised Ratio as used in this paper is 100 plus the actual rate (expressed in the form of an index value to UK average = 100) less the expected rate (expressed in the form of an index value to UK average = index). This is not the same as the alternative definition, actual rate as a percentage of the expected rate.

[3] It is worth noting that whereas in the use of regression the residual error (difference between actual and predicted levels) has no correlation with the actual level, it is quite possible for residuals to be correlated with predicted values when using the geodemographic composition method. On the evidence of the twelve variables used in this analysis it is clear that whilst the method tends to be highlight effective in ranking areas according to the actual levels on the twelve distributions, there is a greater tendency to underestimate the range of values than might be expected just from the measurement of the correlation coefficients of actual and predicted values. These patterns point to the existence of 'higher area effects' on a number of variables, of which aged 65+ is the most striking example.

[4] Although each of the variables used in the analysis are just one of nearly 400 used altogether, each variable used to build Mosaic is given a weight or influence which may or not be greater than the average. Out of a total weight of 8050, a weight of 80 was assigned to Detached Houses, a weight of 60 to Two car ownership, a weight of 24 to Social Grade AB and a weight of 0 to Christians. All eight other variables were given a weight of 40, just under one half a percentage of the overall weight.