Article:
Shulman, H, Birkin, M and Clarke, GP (2015) A comparison of small-area hospitalisation rates, estimated morbidity and hospital access. Health & Place, 36. 134 - 144. ISSN 1353-8292

https://doi.org/10.1016/j.healthplace.2015.10.001

A COMPARISON OF SMALL-AREA HOSPITALISATION RATES, ESTIMATED MORBIDITY AND HOSPITAL ACCESS

Abstract

Published data on hospitalisation rates tend to reveal marked spatial variations within a city or region. Such variations may simply reflect corresponding variations in need at the small-area level. However, they might also be a consequence of poorer accessibility to medical facilities for certain communities within the region. To help answer this question it is important to compare these variable hospitalisation rates with small-area estimates of need. This paper first maps hospitalisation rates at the small-area level across the region of Yorkshire in the UK to show the spatial variations present. Then the Health Survey of England is used to explore the characteristics of persons with heart disease, using chi-square and logistic regression analysis. Using the most significant variables from this analysis the authors build a spatial microsimulation model of morbidity for heart disease for the Yorkshire region. We then compare these estimates of need with the patterns of hospitalisation rates seen across the region.

Keywords

Heart disease, hospitalisation rates, Health Survey of England, spatial microsimulation, morbidity estimates

1. Introduction

Data on hospitalisation rates in most cities and regions reveals widespread geographical variations, especially when mapped at the small-area level. There have been many studies which have explored these variations and a number of alternative explanations have been put forward. The most common set of explanations revolve around 'need' or 'demand': that is, most straightforwardly, that hospitalisation rates will be highest in areas of greatest morbidity. Thus there have been many studies of hospitalisation rates correlated against estimated need by age (highest need coming from the elderly), social class (with lower social class populations generally having a lower health status) and ethnicity (some ethnic groups are much more susceptible to certain types of illnesses than others). In addition, some studies have included other indicators of potential need, including mortality rates, indicators of long-term illness and health deprivation scores. In reality, need is most likely to be driven by a combination of these factors. In fact, a small number of studies have used multiple regression analysis to explore the importance of different individual variables. Whilst exploring health morbidities via regression is useful, it would be valuable if we could combine variables in such a way that we could get a more robust measure of morbidity at the individual level, for example explicitly combining age, social

class and ethnicity for each individual or household in a city or region and estimating potential need on the basis of such joint probabilities.

An essential difficulty in estimating small-area variations in morbidity is the need to combine data from different spatial scales. Sample survey data can be used to develop statistical models of morbidity risk at the individual level based on demographic and socioeconomic characteristics, but such surveys typically are not spatially referenced or include very coarse geographic identifiers. Population data are often available at the small-area scale, but these data rarely include population counts for detailed combinations of age, gender, ethnicity, and socioeconomic characteristics which are needed to estimate morbidity risk.

A spatial microsimulation model (MSM) is a useful technique for reconciling individual survey data with small area demographic estimates. The MSM uses marginal totals from the Census of Population and Households to produce very reliable estimates of individual profiles within a small area. Thus individual profiles can then be linked with equivalent records from the survey data, and then aggregated to produce the desired counts for small geographical areas.

Thus, the aim of this paper is to use microsimulation modelling to produce small-area estimates of morbidity. To test the potential of the methodology we look at heart disease and estimate the number of persons with heart disease at the small-area across a selected region of the county of Yorkshire in the UK. The incidence of the disease can be studied against individual risk factors through a sample dataset, the Health Survey for England. We compare the outputs of the microsimulation model with actual hospitalisation rates provided by the UK Hospital Episode Data.

Note that the model provides estimates of heart disease rather than estimates of hospital admissions. The correlation between these two measures is likely to be a strong one, but is mediated through treatment (hospitalisation) rates. The possibility of a spatial dependence in treatment rates e.g. in relation to access to facilities, cannot be discounted. Thus we argue that building such a model produces a very powerful 'observed' v 'potential' analysis which sheds further light on possible alternative explanations of hospitalisation rates, particularly through a consideration and discussion of the spatial outliers. In areas where estimated need seems low but hospital episodes are high, could some of the high numbers be more a function of accessibility, arguments put forward in a smaller number of geographical studies, in relation to patient locations and GP locations or indeed hospital locations themselves? The contribution of the paper is therefore twofold. First it provides a methodological framework and a specific model for the computation of morbidities, combining a wide variety of individual level demographics. Second, it introduces a case study of heart disease in Yorkshire and provides a substantive analysis of variations between actual rates of hospitalisation and expected use (or 'need' versus 'utilisation'). The outcomes from analyses of this type are important for future health policy, especially in relation to current debates in the UK concerning centralised as opposed to local

service provision and, in particular, with the contemporary policy debate in the UK towards ambulatory care.

The rest of the paper is structured as follows. In section 2 we seek understanding of varying treatment rates in a review of the literature on need or demand for hospital services, especially in light of studies which have analysed variations in hospitalisation rates. The main data sources are described in section 3. In section 4 we use the Health Survey of England to identify the key socio-economic characteristics of individuals who have self-reported with heart disease. This allows us to ascertain the key variables which are needed (in combination) for our microsimulation of morbidity which is described in section 5. The outputs of the model are compared to actual hospitalisation rates in section 6. Concluding comments are offered in section 7.

2. Understanding variations in hospitalisation rates

There have been a number of studies which have explored variations in hospitalisation rates at a variety of different spatial scales, for a variety of different illnesses. The majority of this literature has explored variations in hospital usage against different socio-economic or geodemographic factors about individuals and/or where they live. A core argument is that hospitalisation rates are driven by variations in the geodemographics of the population itself. The first major factor is age – clearly we would expect more hospital episodes within a community of elderly residents (Tseng et al., 2013; Webb et al., 2006, Meade and Earickson, Bay et al., 1997; Morris and Carstairs, 1991). As Meade and Earickson (2005) note "The age structure of a population in large part determines consequences as diverse as the spread of an infectious agent and the severity of the illnesses it causes…and the need for health services" (2005, p35). de Andrade et al. (2013) studied the relationship between age and ischaemic heart disease in Brazil. They created an 'elderly index' (the ratio between the population over 65 and the population under 15) and compared it to mortality rates for ischaemic heart disease. They found that 53.07% of the deaths from this condition occurred for people between 60 and 79 years of age. Walker et al. (2006) plotted hospitalisation admission rates for the elderly segments of the population in Australia from 1996 to 2001, clearly showing the importance of admissions for persons over the age of 70.

The second major explanatory factor has been social class. A common line of argument throughout the literature is that the lower socio-economic status of residents, the higher hospitalisation rates tend to be, particularly for what is deemed to be ailments associated with poor diets and lifestyles, such as diabetes, asthma, pneumonia (Andrulis, D.,1998; Pappas et al, 1997; Haynes et al. 2003, Macintyre et al. 2008). Andrulis et al. (1998), Pappas et al. (1997), Bindman et al. (1995) found that poverty in particular was associated with high hospitalisation rates for these avoidable conditions. Pappas et al (1997) argued that

those who were below the poverty level in the United States (having an income of less than $20 000) had 2.1 to 2.6 times the hospitalisation rates of the highest income group, which they defined as (income of $40,000+). In addition, there are a number of studies which have used different proxies for income or social class to show the relationship between well-being and morbidity, including housing tenure and car availability (Macintyre et al 1998), mortality ratios (Haynes et al 1999), neighbourhood environment (Cummins et al 2005), education or IQ levels (Batty et al 2006).

A third major explanatory factor for variations in hospitalization rates is differences in ethnic composition. In the UK, Gilthorpe et al. (1998) studied the relationship between ethnicity and hospitalisation rates. They were able to identify that age-standardized admission rates were higher for those of African-Caribbean heritage compared to those of Caucasian background. It has been suggested that ethnic variations in hospitalisation rates could simply reflect socio-economic differences (age and income). In some instances, however, higher hospitalisation rates could be due to a greater burden of disease affecting that minority (Mathieu et al., 2012; Wang et al., 2012; Underwood et al., 2012; Lip et al., 2007; Nazroo, 2003). For example, asthma in the US affects those of African Caribbean heritage more so (closer to three times higher hospitalisation rates) than those of Caucasian descent (Getahun, 2005).

The literature has also explored variables which might explain differences in hospitalisation rates other than those associated with socio-economic characteristics. For example, some authors have explored whether certain GPs routinely refer more patients to hospitals than others. The study of variations in GP referral rates have to date been inconclusive. A systematic review, presented by O'Donnell (2000), outlines some of these ambiguous results. Delnoij et al. (1997) and Kerssens et al. (1990) found that GP referrals increased as practice size increased in their Dutch studies, whilst Anthony (2003) found higher GP consultation rates and referral rates from those from lower socio economic groups. However, Anthony also argued that whilst those from lower socioeconomic groups may consult more, they do not get referred more compared to those from higher socio economic groups.

This leads on to a key question: do variations in hospitalisation rates largely reflect different access to care rather than need or morbidity itself? So for example, do those individuals with greater access to primary or secondary care have lower or higher hospitalisation rates than we might expect? It might be argued that greater access to primary care would lead to lower hospitalisation rates for 'avoidable or preventable conditions' in particular, as these people could obtain the treatment they need before it would become necessary to visit a hospital. Bindman et al. (1995) found that "Individuals living in areas where residents had difficulty receiving medical care had high rates of preventable hospitalisations for chronic medical conditions." (1995, p305). They found a strong relationship between access and hospitalisation rates. Similarly, Haynes et al (2006, p432) suggest 'that in some circumstances the

difficulties patients experience in travelling to see their GP can deter contacts which might have resulted in a planned inpatient episode'.

In relation to access, there are many studies examining access to health care for those in rural communities. Sherwood and Lewis (2000) found that physical distance to a GP and the nearest secondary care institution was a problem for rural residents. They studied the issue of access to the Byfield Medical Centre in the UK West Midlands, as this institution became the centralized hub for medical services in the immediate area. This had a significant impact on the outlying areas due to the physical distance between them and this new location of centralized services, especially as this hub was not in the largest settlement in the area (Sherwood and Lewis, 2000:342). There have been many other studies of rural accessibility problems. Roovali and Kiivet (2006) for example, suggest children living 30 minutes from a hospital in Estonia were 50% less likely to be hospitalised. Haynes et al (1999, 2006) have shown a marked distance effect in urban/rural patterns. There are also some interesting papers which evaluate rural accessibility against socio-economic factors. van Hooijdonk et al (2007) for example, showed that areas with lower hospitalisation rates than expected are mainly rural areas (with few non-western migrants) but also that access can override the importance of other variables, such as income levels.

Figures 1 and 2 show hospitalisation counts for heart disease in Leeds in 2007. Leeds is a city of more than 750,000 inhabitants which sits at the centre of a major conurbation in the north of England. In common with many British cities, the population is served by two major hospitals. Alongside the obvious split into East and West, there is clear evidence here of a distance-decay effect in hospital patronage. The number of in-patients at St James's hospital in the centre of Leeds from East Leeds declines quickly the further from its location. The same is the case in West Leeds the further from the Leeds General Infirmary.

While the overall pattern still reflects the traditionally different catchment areas of the two hospitals, previously allocated to different health authorities, the distance decay effect could be the outcome of a number of processes. In the first place, the overall concentration of the population is highest in the city centres. Secondly, certain characteristics of the population such as both average age and deprivation might increase hospitalisation rates. Indeed, one might expect these patterns given that in UK cities many deprived communities are still located in inner city areas where the first Victorian hospitals tended to be located. Thus we might expect high usage of hospitals from these areas based on socio-demographics only. In addition, people living close to hospitals are more likely to use Accident & Emergency departments as walk-in centres (an alternative to waiting for a GP appointment). Pappas et al. (1997) found that those of a lower socio-economic status are more likely to not have a regular source of primary care leading to poorer access to care and hence higher hospitalisation rates. However, some

of the poorest areas of UK cities can be found at the suburban edge (especially the case in east Leeds), and these have low hospital referrals to the two main Leeds hospitals.

To shed more light on the potential reasons for these intra-urban variations in hospitalisation rates it is useful to try and estimate morbidity directly, especially given the variations in geodemographics of households within a city. Despite the considerable amount of research noted above, there are fewer examples of trying to explore variations in hospitalisation rates against estimated individual or household models of potential need. Haynes et al (1999) built a suite of regression models of potential need factors which included many socioeconomic factors as well as mortality rates. They then added indicators of service provision such as distance to GP and distance to hospital. These types of study are useful and help to show the multi-causality effect.

A different approach, microsimulation modelling (MSM), is adopted here in which individual household members are explicitly represented using attributes associated with heart disease. The MSM approach has been widely used within economics and public finance for more than 50 years. The models are typically static (run in comparative static mode) as dynamic models, whilst possible to build, are complex and even harder to calibrate. MSM is particularly powerful as a means for combining evidence between different scales, yielding important benefits in flexible aggregation. A classic example would be that it is much more sensible to calculate the effect of a 1% rise in income tax by applying an appropriate rule to all households that are exposed to the increase. Trying to model this change from an aggregate analysis of household structure and income would be messy and inaccurate. Here the considerations are similar in our desire to compute the effect of changes in the risk profile for individuals rather than populations. The use of MSM is increasingly popular for health care applications (e.g. Tomintz et al 2008, Morrissey et al 2012, Clark et al, 2014). The model will be articulated in Section 5.

Figure 1 here

Figure 2 here

3. Data sources and Methodology

Fig 3 shows the study area for this research, incorporating the key urban areas of Leeds, Bradford, York, Huddersfield alongside more rural districts around Harrogate to the north. The data utilised in this research comes from three sources: the Hospital Episode Statistics (HES), the Health Survey of England (HSE) and the UK Census of Population. The Health Survey of England (HSE) is an annual survey that combines not only questions about health and health related activities, but also records physical

measurements and blood analysis. The HSE asks particular 'core' questions each year on such topics as: general health, smoking, alcohol and demographic characteristics. In some years, the HSE focuses on a particular type of ailment to provide more in-depth information. The 2006 HSE focused solely on cardiovascular disease with a total sample size of 21, 399 individuals, hence the 2006 HSE was selected for this research. Despite the fact that the HSE has a wealth of information on the attributes of individuals with health problems, the spatial resolution in the HSE is poor –it is not possible to know exactly where these persons live.

Thus to investigate potential spatial variations in heart disease it is necessary to reweight the survey using small area variations in geodemographics from the 2001 Census of Population. Hence the strategy for generating small area morbidities is as follows. A number of key risk factors for the morbidity are identified from a survey dataset. A spatial microsimulation model is used to generate synthetic estimates of each of these risk factors for the population of a small area. Then the individuals in the microsimulation are linked to the individuals in the survey data to provide morbidities which can then be aggregated to higher level geographies.

This two stage process is articulated in Sections 4 and 5 of the paper. In Section 4 we use an analytic approach to identify the major risk factors for heart disease and to determine the relative incidence associated with each of these factors. In Section 5 we show how a spatial MSM can be constructed to represent the distribution of these risk factors in a small area population, and we describe how local morbidity rates can be estimated from the geodemographic analysis. Once small area estimates of morbidity have been computed, we want to examine the variation in admission rates by small area which are recorded by the HES. This is achieved in Section 6, where hospitalisation rates are investigated alongside the estimated small area morbidity rates from the MSM.

Fig 3 here

The Hospital Episode Statistics (HES) are a data source that record detailed information on all patient admissions to NHS based hospitals in England. HES data is based on admitted patients, whether inpatients or day cases. The HES data not only contains information on the patient related to diagnosis but also in regards to limited geographic and socio-demographic characteristics. For instance, age, ethnicity and gender are recorded as well as the postal sector in which each patient lives.

The final data set used for understanding the small-area variations in geodemographics is the 2001 Census of Population. This provides data at the postal sector level on age, gender, social-class, ethnicity,

self-assessed general health, variables that are likely to be important for estimating heart disease (see next section).

4. Correlating heart disease with geodemographics

Microsimulation is a complex modelling technique which requires simultaneous estimation of different characteristics, in this case the risk factors. This task is explosively difficult for large numbers of variables (see for example van Imhoff and Post, 1998, who show how a relatively simple problem of estimating variations in maternity rates can generate millions of combinations from a small number of underlying factors). The aim of this section is to use data from the Health Survey for England to identify a restricted number of key factors which underpin the variations in heart disease, and to parametrise these factors in terms of their influence.

Table 1 here

Both a chi-squared analysis and logistic regression were utilized to examine the relationship of the risk factors for heart disease. The dependent variable are those individuals who responded 'yes' to having heart disease in the Health Survey of England 2006. Table 1 shows the results of the chi-squared analysis, whilst table 2 shows the results from the logistic regression. Both tables show the importance of age, social-class, gender, ethnicity and self-assessed general health. [In table 2 the Wald statistic tests the significance of the coefficients of each independent variable in the logistic regression (Garson, 2009): the higher the value, the greater the significance. The exp(b) is the odds ratio for each independent variable. If the odds ratio is above one, that factor leads to increased risk. If the exp(b) is below one, the independent variable leads to less risk.]

Table 2 here

As demonstrated in Table 1, all demographic predictors are significant in the chi-squared analysis. In Table 2, the logistic regression, certain predictors are not as significant as others. Age is a significant predictor of heart disease. The importance of age has been determined in other studies of heart disease (American Heart Association, 2010; Tidy and Willacy, 2009; Gottdiener et al., 2000). Interestingly, according to the American Heart Association (2010) over 85% of all people who have experienced coronary heart disease were over 65 years of age. Table 1 shows that gender is also a significant variable, with men more at risk of heart disease than women. Gender has also been found to possess different risk factors for heart disease in the literature (American Heart Association, 2010; Tidy and Willacy, 2009; Viil-Kajander et al., 2003). Viil-Kajander et al. (2003) focussed specifically on Finland but also found that men were more at risk of coronary heart disease, especially middle aged men. These same

results were also found by Franco et al. (2011); Fodor and Tzerovska (2004); Gottdiener et al. (2000); Esrey et al. (1996); Kannel et al. (1986); and Cullen et al. (1983).

Table 1 shows social class to be a significant explanatory variables for heart disease. Table 2 shows that four social class groups: managerial technical, semi-skilled manual, unskilled manual and other were more at risk in comparison to the other more professional or skilled worker categories. It has been found throughout the literature that social class has a strong association with heart disease (Brownstein, 2008; Viil-Kajander et al., 2003; Marmot et al., 1997; Esrey et al., 1996). Part of the explanation lies with the fact that those of a lower economic status are often also associated with poor diet, increased smoking, increased alcohol intake and increased physical inactivity (Gaziano et al., 2010; Fodor and Tzerovska, 2004; Gottdiener et al., 2000; Marmot et al., 1997; Esrey et al., 1996).

The chi-square analysis, as seen in Table 1, clearly shows the significance of ethnicity as a constraint variable. The fact that for certain ethnic groupings, the exp(b) in table 2 is also above one is important, specifically, Asian or Asian British and Black or Black British. This corroborates the existing literature and shows that ethnicity is an important variable to be included as a key factor in the microsimulation model. (American Heart Association, 2010; Tidy and Willacy, 2009; Brownstein, 2008; McKeigue et al., 1989; Gillum, 1982). Self-assessed general health is also significant at the 95% confidence level in the chi-squared analysis (table 1). However, the logistic regression (table 2) shows that in comparison to 'good' general health, those with 'fair' or 'not good' general health are not necessarily more at risk. However it does make sense that persons with poor general health (even if self-assessed) may have illnesses that themselves might promote heart disease: i.e. higher blood pressure, smoking, higher cholesterol, diabetes, and obesity. The American Heart Association (2010) stated "Smokers risk of developing coronary heart disease is 2-4 times that of non-smokers" (2010). Obesity is another factor that can lead to increased risk of heart disease, but again is very difficult to measure. Homer et al. (2008) stated "The literature points clearly to adverse direct effects of inadequate physical activity on the onset of hypertension, high cholesterol, and diabetes" (2008, p3).

The five key variables identified above – age, gender, ethnicity, occupation, and general health - will be used to build the microsimulation model in the next section. All have been noted in the literature as important risk factors, and all have been found to be significant in the HSE data using chi-square analysis, whilst some elements of each were found to be significant in the logistic regression analysis. Other potential risk factors were tested, including marital status, tenure and educational attainment, but found not to be independently significant. The final model yielded a goodness-of-fit with $R^2 = 0.64$, suggesting that this combination of factors provides a basis for understanding spatial variations which is good but by no means complete. Global goodness-of-fit measures, such as Cox and Snell R^2 or Nagelkerke R^2 can also be used to assess the explanatory power of the model. The Cox and Snell Pseudo R^2 value was 0.13 (the smaller the value the better the fit) whilst the Nagelkerke R^2 was 0.36 (1.00

representing a perfect fit). Even though, this pseudo R^2 is lower than we would like, it was the highest out of all the logistic regression models run.

5. Building a spatial microsimulation model for heart disease

In this section, microsimulation is used to simulate the estimated spatial distribution of residents in the study area with heart disease. The primary strength of a microsimulation model is that small area level information is generated that did not exist before. As such they have already been widely used in health geography. Clarke and Spowage (1984) first proposed a multi-factored model of hospital need that included: population demographics, morbidity (demand), type and amount of care offered (provision) and a hospital allocation model. The microsimulation and allocation model focused specifically on geriatric care and suggested that the use of these models increased the ability to make more informed decisions regarding the allocation of resources.

Smolen et al. (2007) used a microsimulation model to predict mortality from strokes for patients with asymptomatic carotid stenosis over a five year time span. They validated the microsimulation output with data from clinical trials, using a total of eleven characteristics to estimate the specific population most liable to have a stroke. Similarly, there have been a number of applications of microsimulation to estimate the location of smokers in order to provide stop-smoking services. Tomintz et al. (2008) found that smoking was significantly associated with age, social class and ethnicity. Importantly, it was the inclusion of all three of these variables (not just one) which led to a more robust measure of demand or need (see also Tomintz et al 2012 for similar work in Austria). Smith et al. (2011) also built a microsimulation model to estimate the location of smokers, this time in New Zealand. Given that New Zealand asks smoking related questions in their Census, the estimates of smoking based on the microsimulation model could be validated against real world data. Encouragingly, Smith et al. (2011) found that microsimulation can accurately estimate smoking prevalence with minimal error (see also Hermes and Poulson 2012).

Procter and Smith (2008) used a microsimulation model to estimate childhood obesity specifically to identify intervention methods and health policies to reduce childhood obesity. (see also Edwards and Clarke, 2009). This work, using 'SimObesity', showed that estimated obesity was generally higher in lower socio-economic areas (poorer diets, less playground space etc.) but could also be found in higher income areas where more sedentary lifestyles could be contributing to higher obesity levels. In work similar to that reported here Morrissey et al. (2010, 2012) used a spatial microsimulation model to estimate mental health at the small-area level in Ireland. Although not possessing individual hospital

records (as we have in this study) they were still able to signpost areas where estimates of mental health problems were high, but access to acute hospital services was poor.

The type of microsimulation model used in this research is a simulated annealing model. Simulated annealing is a global optimization method that has been used regularly in recent years for solving difficult combinatorial optimisation issues (Harland et al., 2012; Hermes and Poulsen, 2012a; Hermes and Poulsen, 2012b; Cullinan et al., 2011; Hynes et al., 2008; Hynes et al., 2009a; Hynes et al., 2009b). The simulated annealing methodology works by matching the population in the sample (HSE) to the Census data using the constraint variables. The model decides if a person in the HSE with heart problems should be allocated to a household in an individual census zone based on the match between the constraint variables in both data sets.

The simulated annealing microsimulation model code employed and revised for this research was initially created by Harland et al. (2012). The steps inherent in this procedure can be listed as:

1 Configure the thresholds = starting value at t_0 and steps of t (t_0 >>t).
2 Generate a random sample where all weights are set to zero (1 if the person is included and 0 if they are not).
3 Select individuals at random
4 Repeat steps 3 until the required number of individuals has been selected
5 Compute goodness-of-fit for the current solution
6 Replace one individual at random
6a From the current selection, pick an individual and flip the weight from 1 to 0
6b Pick a new individual at random from the sample
7 Recalculate goodness-of-fit
8 Update the weights if the threshold has been exceeded
9 Repeat steps 4 to 6 until the threshold is zero

(adapted from Harland et al. (2012) and Hynes et al. (2009a).

The first step in the simulated annealing methodology was to randomly select a sample of the survey to be matched with the Census population for each census output area (middle super output areas: MSOAs). Once this step was completed, the goodness-of-fit tests were calculated to see how close the matches appeared. Next, an individual within each area was replaced with another individual. The goodness-of-fit tests were recalculated to see if the fit was improved (i.e. the error between the two data sets decreased or lessened). If so, that individual was kept and another random individual was replaced and tested again. If the fit was found to be worse, then that individual was not replaced and the algorithm moved on to the next iteration. This process was repeated until the annealing threshold, set at the beginning of the simulation, reached zero.

The variables estimated by the model can be checked against actual data, but only where that data exists. SRMSE is a general average error measure that measures the difference between the synthesised and actual values. This is a frequently used measure to evaluate how well the synthetic population matches

the real population. The closer SRMSE is to zero, the better the model is at estimating the population. A perfect fit of SRMSE only occurs if the predicted and the actual values match exactly.

$$\text{SRMSE}=\frac{\sqrt{\sum_i \sum_j \left(T_{ij}-\hat{T}_{ij}\right)^2/m \times n}}{\sum_i \sum_j T_{ij}/m \times n} \qquad (1)$$

Where T_{ij} are the observed values and $\hat{T}_{ij}$ are the predicted values.

Total Absolute Error (TAE) and Standardised Absolute Error (SAE) can also be used for calibration**.** TAE is the number of people in the population that have been misclassified (Harland et al., 2012; Voas and Williamson, 2001). TAE can be used to evaluate the degree of error of a constraint variable in a microsimulation analysis. The formula used is:

$$TAE = \sum_i \sum_j \left|T_{ij} - E_{ij}\right| \qquad (2)$$

Where T_{ij} are the observed counts for the item ij of a data table, and E_{ij} are the expected counts.

TAE measures the absolute number of people who have been misclassified. However, TAE can produce larger error counts than actually occur, as it double counts each person. It counts them once if they were in the 'wrong' category and counts them again if missing from the 'correct' category (Harland et al., 2012, Voas and Williamson, 2001). Thus, each misclassification occurrence is counted twice. Thus, a relative measure of the absolute error is needed and that is where the Standardised Absolute Error (SAE) is utilised. The SAE is a statistic that is easy to interpret, especially as it gives equal weight to each table regardless of the size of that table (Voas and Williamson, 2001, p191). The formula is given as:

$$SAE = \frac{\sum_i \sum_j \left|T_{ij} - E_{ij}\right|}{N} \qquad (3)$$

The lower the SAE is, the less error that is present. .

Table 3 here

Table 3 details the goodness-of-fit statistics for the microsimulation models using the two alternative calibration methodologies. It can be seen that the model predicts the constraint variables almost

perfectly in both cases. Thus we can have confidence that the model is performing well in terms of predicting the distribution of the key constraint variables we need to understand heart disease.

The estimated morbidities for heart disease were calculated from the microsimulation model using a Monte Carlo or 'roulette wheel' sampling procedure. For each individual, the odds ratios are combined for each of the observed risk factors to provide a probability of heart disease. Probabilities are drawn at random to determine the incidence – for example, if the combined risk is 0.1 and a random number is drawn less than 0.1 then the condition would be assigned. In this way, the occurrence of heart disease in the HES are merged with the spatial MSM.

6. Comparing hospitalisation rates to estimates of heart disease

The individuals who had heart disease were totalled for each small census tract (the middle super output area or MSOA) and divided by the total population above 16 to provide a rate. The simulated data was compared to the actual hospitalisation rates to see how similar the patterns are.

Fig 4 here

Figure 4a maps the simulated rate of heart disease for over 16s in the study region. Figure 4b shows the hospitalisation rates for heart disease for the same population. Figures 4a and 4b were both mapped in the quantile classification scheme so the same proportion of the population could be identified. Fig 4a shows areas with high estimations of individuals with heart problems to be found in Bradford, Leeds and the semi-rural Harrogate and York areas. The different combinations of variables can be seen at work here: the high rates in Bradford are largely driven by low social class and high non-White ethnic populations. The semi-rural areas of Harrogate and York contain many elderly residents. The patterns in Leeds probably reflect more of a mixture of all these factors: a combination of elderly, low affluence and high numbers of non-Whites

It is the areas that have high proportions of the population estimated to have heart disease but where low hospitalisation rates are evident that are the most interesting. Fig 5 shows hospitalisations for heart disease (as derived from the Hospital Episode Statistic data 2006/2007) for those above 16 years of age, divided by the total expected need (from the spatial microsimulation model). Areas with a rate closer to one indicate those areas where the simulation had created a synthetic population that was extremely close in expected heart disease numbers to the actual hospitalisation rates in those areas. The closer the rate is to zero the greater the difference between hospitalisations and expected disease. As can be seen many areas have a close fit – Bradford is a good example.

Figure 5 here

However, in some areas of the study region, particularly Huddersfield, York, and Leeds, the values of this rate show a marked difference. An important feature of these variations is the existence of clear spatial patterns (or 'autocorrelation'). This suggests that the low rates which can be seen consistently across north and east Leeds, the outskirts of York, to the north and west of Harrogate, and between Huddersfield and Dewsbury, are not the product of random variations in a statistical model. It is much more likely that such strong geographical patterns are associated with an underlying spatial process. The areas of North East and East Leeds provide a good example of where differences are large. In north Leeds the patterns of high heart disease are driven largely by age – a very high elderly population is resident here. Like Harrogate and York where low rates can also be seen this is also mostly quite an affluent area, and one possible explanation is that an interaction between affluence and heart disease is not fully captured in the model as it currently stands (for example, because the affluent have been able to enjoy healthier lifestyles or better diet). However, in East Leeds there are much higher numbers of lower income residents (including council estates such as Gipton, Whitkirk and Seacroft). Given that we would expect higher hospital rates in such areas, access may well be a key explanatory factor to understand the low rates as these are areas furthest from major hospital for treatment of heart conditions.

Areas in which the simulated rates of disease are low relative to admissions are less widespread but also of interest. Here the area of North-West Bradford adjacent to the Airedale NHS Trust Hospital at Keighley is a good example, in counterpoint to its neighbour at Wharfedale (North Leeds/ Otley). In this instance, recent cuts have resulted in the restriction of Wharfedale to basic services, comprising Angiography, Arrhtyhmia, Hypertension, Ischaemic Heart Disease, General Cardiology and Pacemaker Implantation . In contrast, Airedale continues to offer a much wider range of facilities, including Congenital Heart Disease, Heart Failure and Valve Disorders. There is a sense here, then, that not only does poor access lead to reduced levels of uptake, but good access to high quality services can boost utilisation.

In addition to service provision, the waiting time between hospitals also varies substantially from place to place. For example, according to BMI Healthcare (2012), Bradford NHS Trust has average waiting times of 32.4 weeks, in comparison to Harrogate District Hospital at 'only' 18.4 weeks. The relative ease or difficulty in obtaining a hospital bed is another factor which could exacerbate variations in physical access, for example if a long waiting time is combined with poor access.

7. Conclusions

In this paper a spatial microsimulation model has been utilised to estimate morbidity for heart disease based on statistical analyses of the HSE data. The spatial microsimulation model was built using a combination of age, gender, ethnicity, self-assessed general health and social class. Certain MSOAs demonstrated a high simulated need compared to low hospitalisation rates (as derived from HES data). Could these areas be under-served for heart treatment? If facilities were located more closely would more residents receive hospital treatment? Could more lives be saved if variations in access to hospitals were reduced? The modelling of health care need and provision is an important part of any study in the health care field. By studying the geographical components of health care need in relation to service provision, it is possible to understand more about the complexities present in the health care market. The conclusion that access is as important as need has important implications for health care policy, especially contributing to the debate over the benefits of ambulatory care (or dispersed versus centralised health care provision).

The microsimulation model which has been presented here is not capable of fully disentangling these complex cause and effect relationships – more detailed work for example on GP referral patterns, patient attitudes and behaviour would be needed for this purpose. Also the final morbidity patterns seen in the maps cannot be validated externally without access to further small-area survey analysis. However the model has made clear that the level of unexplained variation is substantial, it has demonstrated a clear geography, and has highlighted those places in which the effects are strongest and therefore where the search for further clues might most profitably be targeted.

Acknowledgements

The authors are very grateful for the very helpful comments and suggestions made by the editors and two anonymous referees on earlier drafts of this paper.

8 Work Cited

Andrulis, D. P. (1998). Access to care is the centerpiece in the elimination of socioeconomic disparities in health. Annals of Internal Medicine, 129(5), 412-416.

Anthony, D. (2003). Changing the nature of physician referral relationships in the US: the impact of managed care. Social Science & Medicine, 56(10), 2033-2044.

Association, American. Heart. (2010). Heart disease and stroke statistics: supplement to heart and stroke facts. Dallas, Texas: American Heart Association.

Batty, G. D., Der, G., Macintyre, S., & Deary, I. J. (2006). Does IQ explain socioeconomic inequalities in health? Evidence from a population based cohort study in the west of Scotland. BMJ, 332(7541), 580-584.

Bay, K., Long, M. & Kerr, J. (1997). Utilization of hospital services by the elderly: geriatric crisis in one Canadian single payer system. Health Services Management Research, 10, (1-2), 42-57.

Bindman, A. B., Grumbach, K., Osmond, D., Komaromy, M., Vranizan, K., Lurie, N., & Stewart, A. (1995). Preventable hospitalizations and access to health care. Jama, 274(4), 305-311.

BMI Healthcare (2012) Patient Information, bmihealthcare.co.uk

Brownstein, J. (2008). Addressing heart disease and stroke prevention through comprehensive population-level approaches. Preventing Chronic Disease 5(2), A31 Print

Clark, S.D., Birkin, M. & Heppenstall, A.J. (2014). Sub regional estimates of morbidities in the English elderly population. Health and Place, 27, 176-185

Clarke, M. & Spowage, M. (1984). Integrated models for public policy analysis: an example of the practical use of simulation models in health care planning. Papers of the Regional Science Association, 55(1), 25-45.

Cullen, K., Stenhouse, N., Wearne, K. & Welborn, T. (1983). Multiple regression analysis of risk factors for cardiovascular disease and cancer mortality in Busselton, Western Australia—13-year study. Journal of Chronic Diseases, 36, 371-377.

Cummins, S., Stafford, M., Macintyre, S., Marmot, M., & Ellaway, A. (2005). Neighbourhood environment and its association with self rated health: evidence from Scotland and England. Journal of Epidemiology and Community Health, 59(3), 207-213.

De Andrade, L., Zanini, V., Batilana, A. P., De Carvalho, E. C. A., Pietrobon, R., Nihei, O. K. & De Barros Carvalho, M. D. (2013). Regional Disparities in Mortality after Ischemic Heart Disease in a Brazilian State from 2006 to 2010. PloS one, 8(3), e59363.

Delnoij, D. M., & Spreeuwenberg, P. M. (1997). Variation in GPs' referral rates to specialists in internal medicine. The European Journal of Public Health, 7(4), 427-435.

Edwards, K. & Clarke, G. P. (2009). The design and validation of a spatial microsimuation mode of obesogenic environments for children in Leeds, UK: SimObesity. Social Science & Medicine, 69(7)), 1127-1134

Esrey, K. L., Joseph, L. & Grover, S. A. (1996). Relationship between dietary intake and coronary heart disease mortality: lipid research clinics prevalence follow-up study. Journal of Clinical Epidemiology, 49, 211-216.

Fodor, J. G. & Tzerovska, R. (2004). Coronary heart disease: is gender important? The Journal of Men's Health & Gender, 1, 32-37.

Franco, M., Cooper, R., Bilal, U. & Fuster, V. (2011). Control of coronary heart disease risk factors and evidence-based therapies: joint efforts for coronary heart disease prevention in Spain. Revista Española de Cardiología, 64, 962-964.

Garson, G. D. (2009). Statnotes: Topics in Multivariate Analysis–Logistic Regression.

Gaziano, T. A., Bitton, A., Anand, S., Abrahams-Gessel, S. & Murphy, A. (2010). Growing epidemic of coronary heart disease in low-and middle-income countries. Current Problems in Cardiology, 35, 72-115.

Getahun, D., Demissie, K., & Rhoads, G. G. (2005). Recent trends in asthma hospitalization and mortality in the United States. Journal of Asthma, 42(5), 373-378.

Gillum, R. F. (1982). Coronary heart disease in black populations I. Mortality and morbidity. American Heart Journal, 104, 839-851.

Gilthorpe, M. S., Lay-Yee, R., Wilson, R. C., Walters, S., Griffiths, R. K., & Bedi, R. (1998). Variations in hospitalization rates for asthma among black and minority ethnic communities. Respiratory Medicine, 92(4), 642-648.

Gottdiener, J. S., Arnold, A. M., Aurigemma, G. P., Polak, J. F., Tracy, R. P., Kitzman, D. W., Gardin, J. M., Rutledge, J. E. & Boineau, R. C. (2000). Predictors of congestive heart failure in the elderly: the Cardiovascular Health Study. Journal of the American College of Cardiology, 35, 1628-1637.

Harland, K., Heppenstall, A., Smith, D. & Birkin, M. (2012). Creating realistic synthetic populations at varying spatial scales: a comparative critique of population synthesis techniques. Journal of Artificial Societies and Social Simulation, 15(1), 1-15.

Haynes, R., Bentham, G., Lovett, A. & Gale, S. (1999). Effects of distance to hospital and GP surgery on hospital inpatient episodes, controlling for needs and provision. Social Science & Medicine, 49(3), 425-433.

Haynes, R., Jones, A. P., Sauerzapf, V., & Zhao, H. (2006). Validation of travel times to hospital estimated by GIS. International Journal of Health Geographics, 5(1), 40.

Haynes, R., Lovett, A. & Sunnenberg, G. (2003). Potential accessibility, travel time, and consumer choice: geographical variations in general medical practice registrations in Eastern England. Environment and Planning A, 35(10), 1733-1750.

Hermes, K., & Poulsen, M. (2012a). A review of current methods to generate synthetic spatial microdata using reweighting and future directions. Computers, Environment and Urban Systems, 36(4), 281-290.

Hermes, K., & Poulsen, M. (2012b). Small area estimates of smoking prevalence in London. Testing the effect of input data. Health & Place, 18(3), 630-638.

Homer, J., Milstein, B., Wile, K., Pratibhu, P., Farris, R. & Orenstein, D. (2008). Modelling the local dynamics of cardiovascular health: risk factors, context, and capacity. Preventing Chronic Disease: Public Health Research, Practice and Policy, 5, 6.

Hynes, S., Farrelly, N., Murphy, E. & O'Donoghue, C. (2008). Modelling habitat conservation and participation in agri-environmental schemes: A spatial microsimulation approach. Ecological Economics, 66, 258-269

Hynes, S., Morrissey, K. & O'Donoghue, C. (2009a). Building a static farm level spatial microsimulation model: statistically matching the Irish national farm survey to the Irish census of agriculture. International Journal of Agricultural Resources, Governance and Ecology, 8, 282-292.

Hynes, S., Morrissey, K., O'Donoghue, C. & Clarke, G. (2009b). A spatial micro-simulation analysis of methane emissions from Irish agriculture. Ecological Complexity, 6, 135-146.

Kannel, W., Neaton, J., Wentworth, D., Thomas, H., Stamler, J., Hulley, S. & Kjelsberg, M. 1986. Overall and coronary heart disease mortality rates in relation to major risk factors in 325,348 men screened for the MRFIT. American Heart Journal, 112, 825-836.

Lip, G. Y. H., Barnett, A. H., Bradbury, A., Cappuccio, F. P., Gill, P. S., Hughes, E., Imray, C. & Jolly, K. (2007). Ethnicity and cardiovascular disease prevention in the United Kingdom: a practical approach to management. Journal of Human Hypertension, 21, 183-211.

Macintyre, S., Ellaway, A., Der, G., Ford, G., & Hunt, K. (1998). Do housing tenure and car access predict health because they are simply markers of income or self esteem? A Scottish study. Journal of Epidemiology and Community Health, 52(10), 657-664.

Macintyre, S., Macdonald, L. & Ellaway, A. (2008). Do poorer people have poorer access to local resources and facilities? The distribution of local resources by area deprivation in Glasgow, Scotland. Social Science & Medicine 67, 900-914.

Marmot, M. G., Bosma, H., Hemingway, H., Brunner, E. & Stansfeld, S. (1997). Contribution of job control and other risk factors to social variations in coronary heart disease incidence. The Lancet, 350, 235-239.

Mathieu IV, R. A., Powell-Wiley T. M., Ayers, C. R., McGuire, D. K., Khera, A., Das, S. R. & Lakoski, S. G. 2012. Physical activity participation, health perceptions, and cardiovascular disease mortality in a multiethnic population: The Dallas Heart Study. American Heart Journal, 163, 1037-1040.

McKeigue, P., Miller, G. & Marmot, M. (1989). Coronary heart disease in South Asians overseas: a review. Journal of Clinical Epidemiology, 42, 597-609.

Meade, M. & Earickson, R. (2005). Medical Geography 2nd Edition, London, The Guilford Press.

Morris, R. & Carstairs, V. (1991). Which deprivation? A comparison of selected deprivation indexes. Journal of Public Health Medicine, 13, 318-326.

Morrissey, K., Hynes, S., Clarke, G., & O'Donoghue, C. (2010). Examining the factors associated with depression at the small area level in Ireland using spatial microsimulation techniques. Irish Geography, 43(1), 1-22
.
Morrissey, K., Clarke, G., O'Donoghue, C. (2012). Linking static spatial microsimulation modelling to macro models: The Relationship between Access to GP services & Long Term Illness. In: Edwards, K. and Tanton, R. ed(s). Microsimulation Methods and Models. London, Springer

Nazroo, J. (2003). The structuring of ethnic inequalities in health: economic position, racial discrimination, and racism. American Journal of Public Health, 93, 277-284.

O'Donnell, C. A. (2000). Variation in GP referral rates: what can we learn from the literature? Family Practice, 17(6), 462-471

Pappas, G., Hadden, W. C., Kozak, L. J., & Fisher, G. F. (1997). Potentially avoidable hospitalizations: inequalities in rates between US socioeconomic groups. American Journal of Public Health, 87(5), 811-816

Procter, K. & Smith, D. (2008). Editorial: size matters: the role of scale in geographies of health. Area, 40, 303-305.

Rooväli, L., & Kiivet, R. A. (2006). Geographical variations in hospital use in Estonia. Health & Place, 12(2), 195-202.

Sherwood, K. B., & Lewis, G. J. (2000). Accessing health care in a rural area: an evaluation of a voluntary medical transport scheme in the English Midlands. Health & Place, 6(4), 337-350.

Smith, D. M., Pearce, J. R. & Harland, K. (2011). Can a deterministic spatial microsimulation model provide reliable small-area estimates of health behaviours? An example of smoking prevalence in New Zealand. Health & Place, 17, 618-624.

Smolen, H. J., Cohen, D. J., Samsa, G. P., Toole, J. F., Klein, R. W., Furiak, N. M. & Lorell, B. H. (2007). Development, validation, and application of a microsimulation model to predict stroke and mortality in medically managed asymptomatic patients with significant carotid artery stenosis. Value in Health, 10, 489-497.

Tidy, C. & Willacy, H. (2009). Prevention of cardiovascular disease [Online]. Available: www.patient.co.uk/doctor/prevention-of-cardiovascular-disease.

Tomintz, M. N., Clarke, G. P. & Rigby, J. E. (2008). The geography of smoking in Leeds: estimating individual smoking rates and the implications for the location of stop smoking services. Area, 40, 341-353.

Tseng, K. C., Hsu, C.-L. & Chuang, Y.-H. (2013). Designing an Intelligent Health Monitoring System and Exploring User Acceptance for the Elderly. Journal of Medical Systems, 37, 1-18.

Underwood, J. M., Townsend, J. S., Tai, E., Davis, S. P., Stewart, S. L., White, A., Momin, B. & Fairley, T. L. (2012). Racial and regional disparities in lung cancer incidence. Cancer, 118, 1910-1918.

Van Hooijdonk, C., Droomers, M., van Loon, J. A., van der Lucht, F., & Kunst, A. E. (2007). Exceptions to the rule: healthy deprived areas and unhealthy wealthy areas. Social Science & Medicine, 64(6), 1326-1342.

Van Imhoff, E., Post, W. (1998) Microsimulation Methods for Population Projection, Population: An English Selection, 10, 97-138.

Viil-Kajander, M., Moltchanova, E., Salomaa, V., Tuomilehto, J., Ketonen, M. & Plalmaki, P. (2003). Geographical variation in the incidence of acute myocardial infarction in eastern Finland - a Bayesian perspective. Annals of Medicine, 35, 8.

Voas, D. & Williamson, P. (2001). Evaluating goodness-of-fit measures for synthetic microdata. Geographical and Environmental Modelling, 5, 177-200.

Walker, A., Thurecht, L. & Harding, A. (2006). Changes in Hospitalisation Rates and Costs in New South Wales, 1996–97 and 2000–01. Australian Economic Review, 39, 391-408.

Wang, O. J., Wang, Y., Chen, J. & Krumholz, H. M. (2012). Recent trends in hospitalization for acute myocardial infarction. The American Journal Of Cardiology, 109, 1589-1593.

Webb, P., Bain, C. & Pirozzo, S. (2006). Essential Epidemiology: An Introduction for Students and Health Professionals, Cambridge, University Press.