Death in England and Wales:

Using a classificatory approach for researching mortality

Mark Alan Green

Submitted for the degree of Doctor of Philosophy

Department of Geography

University of Sheffield

September 2013

To my family, a constant source of inspiration.

Acknowledgements

I cannot thank enough the help and support of my primary supervisor Dr Daniel Vickers. He went beyond what was expected of a supervisor and I will never be able to repay the confidence in my own ability he has given me. Put simply, without his advice and critical eye, this thesis would be of a considerable lower quality.

My second supervisor, Professor Danny Dorling, was also instrumental within this thesis. His support for my work and encouragement to develop it further and get it published has helped me become the researcher that I am today.

Bethan Thomas' encyclopaedic knowledge of the data set was immensely useful throughout the construction of the area classification and I really appreciate all the lunches I ruined accessing the data in her office.

I would like to extend my upmost thanks to Professor Sankaran Subramanian. He organised my visiting studentship at Harvard's School of Public Health and acted like a supervisor to my work, despite not being required to.

Thanks must go to Professor Charles Pattie and Dr Adam Whitworth. Your critical comments during the PhD upgrade process drove me to repay your faith in myself and produce a better thesis than was originally proposed.

I would like to extend my gratitude to my friends and family for their continued support and understanding during my three years of study. To those who showed any interest in my work, this did not go un-appreciated.

Many thanks also to the Economic and Social Research Council (ESRC) for financially backing this project, of which without, I would have not been able to commit to it.

Abstract

The purpose of this thesis is to explore how useful a classificatory research approach can be to researching mortality patterns. This is conducted through creating an area classification of small area mortality patterns for England and Wales (2006 to 2009). The resulting area classification is then applied to research the existence of area effects on health, to assess how useful it is as an analytical tool.

To successfully achieve this, the thesis begins through reviewing the literature to examine the importance of taking an area perspective to researching mortality. This was extended to assess the extent of which an area classification could build upon past research. Data and methodology issues were discussed, to evaluate the best approach required to building a high quality and relevant area classification. The area classification was built and statistical testing was conducted to assess the stability of it.

The area classification was then interpreted to examine the main mortality patterns that dominant England and Wales. Explanations for the clusters were derived from demographic, social and geographical factors. The area classification was then analysed to explore the existence of area effects through a multi-level analysis. This was extended to examine the impact on health as people migrated between the clusters, exploring whether area effects were observed. Benchmarking of these results was performed, comparing it to using GORs instead to group the data, to evaluate how useful the results were.

Contents

List of figures

List of tables

List of equations

List of abbreviations

APHO – *Association of Public Health Observatories*

DHSS – *Department of Health and Social Security*

DOEW – *Death Occurrences in England and Wales, 1981-2009*

DoH – *Department of Health*

ICD – *International Classification of Diseases*

GORs – *Governmental Office Regions*

GRO – *General Register Office*

LSOA – *Lower Super Output Area*

MSOA – *Middle Super Output Area*

NOO – *National Obesity Observatory*

NS-SeC – *National Statistics Socio-economic Classification*

OA – *Output Area*

ONS – *Office for National Statistics*

PHE – *Public Health England*

WHO – *World Health Organisation*

Chapter 1: Introduction

1.1 Background

Everybody dies. Since death is singular and permanent, it provides a useful source of information for evaluating and comparing the health of individuals, especially since it cannot change. Age at death provides one option for analysing differences in a population, given that it is preferable to maximise length of life. However what causes death is not always the same and this provides an additional means of comparison. Cause of death can represent a useful description of an individual's life, affected by the interactions, behaviours and experiences of how an individual chooses to live their life.

Collecting this information provides both a comprehensive, but ultimately complicated data set as common patterns and processes become easily lost in the roughly half a million deaths that occur each year in England and Wales. Furthermore, there are many different causes of death. For example, the latest 'International Classification of Diseases' has over 14,000 possible causes (WHO, 2004). Combining all of this information is far too much for the human mind to effectively analyse (Everitt et al., 2001). Some form of data reduction is required.

Analysing patterns in mortality is effective for formulating effective policy options. As Williams et al. (2004) argue; "The drive to tackle health inequalities and the move to localised policy making have increased interest in small area mortality data." (p958). Previously, one-dimensional measures such as single causes of mortality were used. Whilst they offer useful applications, by themselves they are simply linear variables of which more complex processes and variations remain hidden. The health characteristics of areas are not understood by just single variables, especially given that causes do not operate in isolation.

Typically we tend to think of mortality as being medically defined. However research over the past 30 years has shown the importance of social factors in determining mortality outcomes (Diez-Roux, 2001). For example, Woolf et al. (2007) found that tackling social inequalities in the United States would improve current health (and hence subsequent mortality) more than medical advances could. England and Wales has a long tradition of governmental research formulating the role that social

inequalities play in segmenting our society (the Black and Acheson reports, the Marmot Review).

Whilst health is socially determined, geographical inequalities persist suggesting the importance of the spatial dimension when examining health patterns. Causes of mortality display varying geographical patterns through England and Wales (Shaw et al., 2008), following structures of social, economic and spatial factors. For example, Hacking et al. (2011) found that the North of England displays an excess in deaths of 15 per cent when compared to the South. Tackling geographical inequalities in mortality is a key government aim and therefore producing greater evidence to inform policy development is paramount.

Furthermore, there is evidence showing how geography and particularly neighbourhoods exert an independent effect on health (Pickett and Pearl, 2001; Riva et al., 2007). This has been aided through the development of causal mechanisms to understand how such processes transpire, for example social relations (Kawachi et al., 1999) or deprivation amplification (Macintyre et al., 2008). Research in this area forms an important agenda, with policy implications regarding the effectiveness of targeting individuals or areas. However the field is under-investigated and gaps remain for a more rigorous testing of the hypothesis, through new directions (for example the effect of migration).

The combination of being able to analyse the interactions of vast amounts of data on mortality patterns, whilst maintaining a geographical perspective is therefore an important research agenda. The creation of an area classification represents one option to be able to effectively manage both approaches. This approach takes complex data sets across many different variables of interest and simplifies the data into groups describing the main patterns across the areas (Everitt et al., 2001; Gordon, 1999). Areas are then assigned to the group which they best represent, allowing a simplification of the data through describing the characteristics of the area it generally relates to (Harris et al., 2005). As such, an area classification will bring clarity to the complexities of mortality patterns of England and Wales.

The importance of this approach is linked to the notion that people of similar characteristics tend to live together (Harris et al., 2005). Using areas as the unit of analysis is useful since it will represent the individuals living in them, allowing the description of geographical characteristics in area types. This study puts a focus on the

importance of the neighbourhood at the end of the life course, as an identifier of the type of area a person has lived in throughout.

As Everitt et al. (2001) note, building an area classification "...is essentially about *discovering* groups in data…" (p6). Exploring the underlying structure of mortality patterns across England and Wales adds new information on the understanding of the interaction of causes of mortality, something of which there is little literature on. However its benefits are not restricted to just this, as it can also be applied as a research tool to analyse other fields. The simplification of patterns and processes can herald greater insight and understanding.

An area classification of mortality patterns allows for the simplification of these patterns through incorporating a multi-dimensional framework. This is because it improves the detail available in a measure by adding value through summarising the interactions between variables (Openshaw et al., 1994). It is not restricted to one measure but shows the main patterns and interactions between causes that dominate England and Wales. Knowing and understanding these patterns is paramount to targeting policies more effectively.

This approach has been established in other academic areas for simplifying (large) data sets to help our understanding. Area classifications have been most commonly applied in the field of demography, where they represent big business (Harris et al., 2005; Vickers, 2006). For example the 'People and Places' classification provides a demographic classification of the types of people living in areas, which is useful in examining the characteristics of locations. Yet it has also been applied within the field of public health, for example in the targeting of population groups for the uptake of colorectal screening (Nnoaham et al., 2010).

Within health and mortality research, there have been few applications of the classification of areas. Of those that have (for example Shelton et al., 2006; CACI, 2010), these have been limited in their quality and scope. This has led to calls from both academics (Abbas et al., 2009) and governmental departments (DoH, 2005) for greater application of a classificatory approach. There lies a clear gap for an in-depth development of an area classification of mortality patterns for small areas within England and Wales.

1.2 Aims and objectives

The thesis seeks to show how a classificatory approach is useful for increasing understanding of mortality patterns. The aims of this thesis have been identified as:

I. Create a classification of mortality patterns of small areas for England and Wales, with a clearly justified open methodology.
II. Understand the dominant mortality patterns and why this segmentation exists across the areas within each cluster.
III. Assess the extent that the area classification can help understand the existence of area effects on health.
IV. Benchmark the results to traditional ways of grouping the data, to highlight its usefulness as a research tool.

These aims are not mutually exclusive of each other; rather tackling each aim will help inform the others.

The first aim will form the focus of the thesis. To achieve it, the literature will need to be reviewed to understand how best to implement an area classification into this study. This includes the concept, the data requirements, the methods available and the relevant statistical testing to apply. Evaluating and applying these clearly is important for creating an area classification of high quality.

The second aim seeks to understand the area classification produced in the first aim. This will be conducted through exploring the cluster centres and mapping the area classification to show what each cluster represents. Area statistics will be gathered to analyse the characteristics of the areas that make up each clusters, to explain the segmentation of patterns to form the area classification.

The third aim applies the area classification to assess how useful it is for improving our understanding of area effects. The aim will be achieved through two analyses. Firstly the effect on health of individuals living in particular clusters will be tested for to show evidence for static area effects. Then this concept will extended through examining whether there is any observed effect on health as individuals move between the clusters.

The final aim will evaluate the results from the third aim. This will be achieved through repeating each analysis with a comparable means for grouping the same areas, to assess whether the area classification has greater discriminatory power in the analyses than

equivalent measures. Governmental Office Regions (GORs) and Wales (hereby referred to as a GOR since it is often included in the same statistics) were used since they provide a similar number of groups, as well as being used for reporting mortality statistics.

To achieve the aims, the following objectives have been devised. The aim they tackle is given in brackets, although all the objectives complement each aim:

1. Examine the evidence of how geographical factors can influence health and mortality (III).
2. Critically assess the advantages and disadvantages of creating an area classification for researching health and mortality (I).
3. Review and compile the data available, in preparation for it to be used in building the area classification (I).
4. Evaluate the methods available and select the best method based upon the aims and data of the study (I)
5. Build the classification through detailing, clearly, the steps taken (I).
6. Statistically test the classification to assess robustness and quality (I).
7. Interpret the clusters to understand what they represent (II).
8. Analyse differences in the geographic areas within each cluster to explain why they exist (II and IV).
9. Apply the area classification analytically, within the field of area effects.
    a. Test the impact on health of living in the clusters (III).
    b. Extend the investigation of area effects by incorporating migration through analysing the effect on health as individuals move between clusters (III).
10. Benchmark the results using the classification against using GORs instead (IV).
11. Evaluate the contribution of the research to the literature, critically assess its value and suggesting future extensions to build upon its findings (all aims).

1.3 Thesis structure

To be able to achieve the stated aims and objectives, the thesis is divided up into eight chapters. Table 1.1 presents where each objective will be met across the chapters of the thesis.

Chapter	Title	Objectives
2	Literature review	1 and 2
3	Data and methodology	3, 4 and 5
4	Creating an area classification of mortality patterns	5 and 6
5	Interpreting the area classification	7 and 8
6	Assessing the area classification	9a and 10
7	Internal migration, area effects and health	9b and 10
8	Conclusions	11

Table 1.1: The thesis structure.

Chapter two ('Literature review') begins by arguing for the importance of taking a geographical perspective in researching mortality. This is achieved through reviewing the evidence showing such an approach, as well as the possible causal mechanisms through which geography can affect health and mortality. It briefly touches on common methods for researching mortality within a geographical perspective, before critically evaluating how an area classification would be an useful approach to analysing health and mortality patterns. It ends by discussing gaps in research of which building an area classification would be useful for addressing and building on.

Chapter three ('Data and methodology') discusses the data and methodology involved in the creation of the area classification. It is based upon Milligan and Cooper's (1987) schema for running a cluster analysis (which are often used for building a classification), extending it to be relevant for an area classification. It discusses the data and geographical scale used, which represent the inputs to the area classification. The possible methods are discussed and evaluated to assess which is best applied in this study. A discussion is then provided of the important methodological decisions of the method requiring addressing, as well as the testing procedures to be employed.

Chapter four ('Creating an area classification of mortality patterns') details the creation and testing of the area classification. It is important to provide a detailed methodology, given that particular decisions involved in the process are somewhat subjective. This allows the resulting area classification to be open and hence evaluated fairly. These include the selection of seed points, the choice of the number of clusters in the classification and the calculation of cluster centres. The testing procedures draw upon the few literary suggestions available to assess whether the area classification is statistically stable. This includes performing a replication analysis, evaluating the impact of outliers and assessing variable sensitivity.

Chapter five ('Interpreting the area classification') explores the resulting area classification to understand what it shows. It begins by presenting the characteristics of the clusters, allowing the naming of them to summarise their profiles. The rest of the chapter focuses on explaining the clusters. Health related statistics are compiled to help understand the segmentation of mortality patterns into the clusters. Demographic, social and geographical explanations of differences between the clusters are analysed to help explain the clusters.

Chapter six ('Assessing the area classification') is the first chapter that applies the area classification to assess how useful it is. It is used in a multi-level analysis testing whether the clusters had an independent effect upon the health of individuals living within those areas. This provides one way of showing if the clusters are more than just statistical clusters of similar mortality rates. However to be able to fully evaluate how useful the area classification is, the analysis was compared to using Governmental Office Regions instead since they are often used as a (geographical) grouping tool for the reporting of mortality statistics.

Chapter seven ('Internal migration, area effects and health') extends the idea of neighbourhood effects developed in the previous chapter. If neighbourhood effects are important, then it would be expected that as individuals move between different area types, then there would be an observed effect on health in relation to the change in the environment. The chapter analyses the impact on health of migration between the clusters, using an under-utilised methodology to solve past issues ignored in the literature. The relationship is also conceptualised the other way round, with health influencing migration between the clusters as well. Similarly to Chapter 6, the analysis is compared to using the GORs instead, to be able to evaluate the results.

Chapter eight ('Conclusions') concludes the thesis. A summary of the research findings is firstly outlined, focusing on how the thesis has successfully tackled each of the aims and objectives set out at the start of the thesis. The limitations of thesis are the discussed, to critically evaluate the research findings of this approach. Future extensions of research based upon the findings and the limitations are proposed, to develop and extend the ideas discussed throughout this thesis.

1.4 Outputs

The area classification produced as part of this thesis will be uploaded onto the 'Social and Spatial Inequalities' research cluster's website (www.sasi.group.shef.ac.uk), along with a link to the final version of the PhD thesis itself (via the 'White Rose' website; http://etheses.whiterose.ac.uk/). This provides an open approach to the dissemination of the research tool, encouraging uptake and use of it as an analytical measure. It also presents the decisions and methods used openly, allowing critique and a better understanding of the results.

A paper will be written and submitted to the journal '*Journal of Epidemiology and Community Health*', detailing the construction of the area classification and its interpretation (Chapters 3, 4 and 5). This will reach potential users of the classification in academia, local government and the NHS, encouraging its use as a research tool through highlighting how useful it is.

A second paper will focus on the application of the innovative methodological approach applied to eliminate selection bias in the analysis migration (Chapter 7). It is hoped that encouraging the approach will not just further the analysis of migration and heath, but also the wider geographical field as well. A subsequent article will then be compiled exploring the impact of migration between different area types on health, testing for area effects. The journals '*Health and Place*' and '*Social Science and Medicine*' have been identified as important target journals for publication.

Chapter 2: Literature review

2.1 Introduction

This chapter is split into two sections. The first concentrates upon the theoretical side of the thesis, which informs the later analysis. It begins with a brief introduction highlighting the importance of geography in health-based research, as a dimension which aids and furthers our understanding of patterns and processes. However for geography to be truly important in this research framework there needs to be a theoretical association rather than it simply being a container for processes to occur in. Possible causal links are provided, with a focus on the role of the social environment as this forms a large part of the analytical framework for this thesis. A review of the evidence for possible area effects is examined to explore whether this framework for investigation is useful. The section concludes through making the theoretical link to a different subject, the role of migration on health, to show how this under-utilised area of research would help examine the possible existence of area effects.

The second half of this review concerns the methodological approaches to researching the geographies of health. Firstly, a brief review of the main spatial analytical methods which have been useful for researching geography and mortality is presented. Based upon this, a new research framework is proposed for this purpose, involving the production of an area classification. The justification of both this approach to researching mortality, as well as for advancing our knowledge of area patterns is outlined. Other possible avenues are also briefly suggested, which could achieve a similar result. Some limitations of this approach are presented, with some brief details to address each. It ends with some conclusions gathered from this review, identifying gaps which can be addressed in this thesis.

2.2 A geographical perspective to researching health

There has been long established evidence within Britain of the inherent geographical inequalities in health. Edwin Chadwick (1842) found not just social class inequalities or inequalities based upon place, but that in some places, people of the highest social

class had a lower average age of death than the lowest classes in other places (see Table 2.1). Although only the limited measure of average age of death was used, it still shows early evidence that location was paramount towards influencing health. These differences were also not due to population density (Szreter and Mooney, 1998), instead a factor of urban-rural influences.

	Professional Trades	Tradesmen	Labourers
Rutland	52	41	38
Leeds	33	27	19
Liverpool	35	22	15
Manchester	38	20	17
Bolton	34	23	18

Table 2.1: Average age of death in 1839 calculated for area and social class (after Chadwick, 1842).

Despite these early indications of the importance of geography, most research until the late 20th Century ignored the dimension, merely using it implicitly to display information. As a result, the bulk of health-based research which considered place was found in the sub-disciple of 'Medical Geography' (Bentham et al., 1991). The focus was on aetiology and disease ecology (i.e. just environmental influences), limiting the evidence produced for geographical inequalities. However the 1990s saw the shift away from this, with the name for this research changing to 'Geography of Health' (Kearns and Moon, 2002). Influenced by studies such as the Black Report (DHSS, 1980), this evolutionary shift saw the general subject area move away from strictly the biomedical model towards introducing place and social influences into its analytical understanding of health (Kearns, 1993). As such, most of the evidence provided here has been produced in a short amount of time. A lot of this has been aided through the greater availability of low level data sets and growth in computing power.

Evidence of the importance of geography can be viewed through focusing on a recent study by Hacking et al. (2011) who explored mortality patterns regionally in England. Dividing England into the North (composed of the Governmental Office Regions of the North East, North West, Yorkshire and Humberside, East Midlands, West Midlands) and the South (East of England, London, South East and South West), they found that on average, there was an excess of deaths in the North of around 15 per cent for both genders (2008; although this figure has remained fairly constant since 1965). The

authors acknowledge that the greater levels of poverty in the North may drive this result (a view supported by Woods et al., 2005), however there is evidence discounting this indicating the additional effect of geography (Doran et al., 2004; Whynes, 2009). The North-South divide is an important example of spatial gradients in health. The regional divide does not hold for all causes, with evidence of inverse gradients also observed (for example lymphatic cancer, prostate cancer; Shaw et al., 2008).

The considered importance of geography also can be showed through the exploration of persistent urban-rural inequalities. For example, Erskine et al. (2011) focused on just alcohol-related mortality patterns by deprivation and an urban-rural classification of areas in England and Wales (1999-2003). Using the relative index of inequality, their results showed the continued disadvantage in terms of health in urban areas, with a relative risk of 1.35 for males and 1.13 for females, after controlling for both age and the deprivation level of each area (a decrease from 2.37 (males) and 1.68 (females) without controlling for deprivation). Although the primary reasons behind these inequalities has changed since Chadwick's study from the benefit of a rural lifestyle to mostly one reflecting social factors, there is a still an effect attributed to geography even after accounting for the social dimension.

The final aspect showing how a geographical perspective can improve our understanding of the health patterns, which also underpins the conceptual framework for this study, regards the clustering of health and mortality outcomes. Giggs et al. (1988) highlights the importance of this through analysing acute pancreatitis incidence in Nottingham. There was a cluster representing an area of elevated risk of the disease, however this remained unexplained by social class or age (i.e. risk factors). Rather examining water supply areas, the Burton Joyce reservoir catchment contained the areas of high rates, with the water having high levels of 'hard water' (i.e. high Calcium Carbonate and Magnesium). This helped further understanding of the causal pathway of the disease, tackling the spread of the disease (and hence mortality rates from it).

It is useful to examine the differences in health between places rather than just assume space to be constant or inactive. Assuming that Britain is spatially homogenous would ignore the many important geographical variations that exist within the complex array of health patterns. Many diseases vary in relation to a myriad of social, demographic and geographical factors, all of which display distinct patterns that imprint on the distributions of various diseases (Shaw et al., 2008). Therefore an ecological

perspective within research is important, since analysing individuals alone may ignore these patterns (Macintyre and Ellaway, 2000). It is also paramount in formulating policies, as focusing on the correct areas is just as important in addressing individuals.

2.3 Socio-spatial determinants of health

Whilst there appears to be distinct geographical patterns with regards health, the question arises; does the local environment impact upon health, or instead is space a container for more important processes? It follows the argument of context (the area individuals are exposed to) or composition (risk factors only applicable to the individual) factors as explanations for health patterns (Macintyre and Ellaway, 2000). For geography (context) to be truly important, there needs to evidence of possible causal pathways to justify this research approach. For purposes of what is most relevant to this study, this section will focus just on the social environment and possible causal mechanisms.

There is considerable evidence that poverty/deprivation has a damaging effect upon health. For example, Gregory (2009) analysed patterns of standardised mortality rates across England in 2001 and found the ratio between the 10 percent most and least deprived areas to be 1.36 (i.e. mortality rate in the most deprived areas is 36 percent higher than in the least). Interestingly, when this was replicated for equivalent sized areas from Census data from 1991 (with a deprivation measure based upon the Carstairs index used in 2001), the ratio was 1.39. Despite medical advances and the improvement of the standard of living, geographical inequalities have remained stable. Particular diseases were also associated with deprivation, including respiratory diseases ($r = .545$, $p < .001$) and lung cancer ($r = .584$, $p < .001$).

A compositional explanation on this would focus on poverty being a lack of resources and therefore the poorer health outcomes are a factor of increased numbers of people with poor health living together. There is, however, evidence to show that this is not the case, through the observation of additional contextual disadvantage (see Section 2.4). Evans and Kim (2007) argue that deprived areas present greater exposure to the local population of both physical and social risks, which in turn impact on individual's heart through increased blood pressure levels and stress. Understanding these social risks and their causal mechanism on health is important for justifying a contextual study.

Poverty also leads to knock-on effects through notions such as 'deprivation amplification'. Those areas which are deprived also experience additional disadvantages because of their poverty, leading to double jeopardy effects. For example Tudor-Hart's (1971) inverse care law states that doctors will generally choose to work in less deprived areas, leading to inequalities being reinforced by service provision being lower in areas of greater need. However the evidence for such effects is contested, with Macintyre et al. (2008) finding no significant differences for the location of GPs or Pharmacies across Glasgow.

William Julius Wilson's (1987) social isolation theory argues that neighbourhoods of concentrated poverty can become isolated (both physically and socially) from other areas and wider society. This segregation can lead to localised cultures whereby social and cultural norms have changed as a result of being isolated. These effects upon attitudes and socially acceptable behaviours can impact upon health and mortality, for example diet, alcohol consumption, smoking and unsafe sex (Berkman and Glass, 2000). This highlights how the neighbourhood environment has a structural impact upon health, offering a different causal mechanism for the influence of poverty to operate with.

The social cohesion of an area can influence health. Social cohesion refers to the connectedness of an area. Kawachi and Berkman (2000) identify two main themes for a cohesive area; a lack of social conflict (for example polarisation, ethnic tensions, income inequality) and strong social bonds. There has been focused debate surrounding the importance of income inequality even within neighbourhoods. Direct impacts on health operate through psychosocial channels including feelings of inferiority, impact upon anxiety and stress, possible causing poorer health (Wilkinson and Pickett, 2009). Less cohesive (or unequal) areas will also have lower community participation, as well as fewer controls on the behaviours of individuals. However a review of the literature by Wilkinson and Pickett (2006) showed mixed evidence of the significant effect of inequality on health using smaller geographical areas. Rather, they argue, this process operates at a national level.

How individuals are integrated into their local areas also can affect their health. This association was long theorised by Émile Durkheim (1897) who proposed that suicide, whilst an individual act, was influenced by social phenomena, specifically the social integration (or 'anomie') of individuals into groups. Peter Congdon (1996) tested this

association between anomie and suicide patterns in London. Anomie was measured through an index including one-person households, unmarried people (over 15), population mobility and people privately renting, factors previously shown to be limit social integration. Anomie was found to only have a significant effect for females (especially those aged over 45), where higher levels of factors associated with less integration positively related to suicide rates (Table 2.2). Deprivation was also important with higher rates of suicide in more deprived areas, being a stronger factor for males (especially younger males). The combination of these two factors best understood suicide patterns across London.

	Males		Females	
	Parameter	t-statistic	Parameter	t-statistic
Constant	-0.0221	0.9	-0.0321	0.84
Deprivation	0.1291	7.68**	0.0809	3.07*
Anomie	0.0071	1.01	0.0424	4.05*
Log-likelihood	-1354.9		-947.2	
Deviance	922.9		859.9	

Table 2.2: Results of a Poisson regression analysing suicide patterns in London wards 1990-1992 (after Congdon, 1996).

*Note: $p<0.05$ = *, $p<0.01$ = **, $p<0.001$ = ***.*

Another avenue for which the socially isolated are at a higher risk of poor health is that they have less access to social support (Kawachi et al., 1999). Social support will allow the diffusion of health advice, as well as providing emotional and instrumental aid and also networking may give access to employment, which can indirectly influence health (Berkman and Glass, 2000). Social support and integration may also have a greater impact in poorer areas, as these individuals are more likely to be geographically restricted to the local area (Kawachi and Berkman, 2000).

Where deviant behaviours are more socially acceptable, social integration may not always be beneficial (Wilson, 1987). This was observed by Leventhal and Brooks-Gunn (2004). As part of the MtO (Moving to Opportunity) study, families were randomly chosen and provided the opportunity to relocate from the most deprived areas of five US cities to a better environment. Whilst there were general improvements across most health measures (see Section 2.5 for more details), teenagers aged 14 to 18 reported no improvements in mental health as they maintained contacts from their

previous area and were shown to be returning frequently to visit friends allowing their previous neighbourhood to still be influential.

Neighbourhoods not only have stressors that affect health, but also can provide resources to help improve it as well, through social capital. Whilst similar to both integration and cohesion, Kawachi and Berkman (2000) define social capital as the collective structures that affect relationships within an area, not just the individual ones. Areas with lower social capital will have fewer local ties and hence less community participation. This may be due to lower trust of other residents or fewer community organisations to offer social integration. This in turn will have similar effects, with slower diffusion of health information and less social support, but also communities will be less likely to protest or demand for better services (Kawachi et al., 1999).

Berkman's (1995) review of the evidence of the impact of social capital on (all-cause) mortality shows consistent support of its beneficial effects (which were fairly large themselves). However Berkman goes beyond simply summarising past studies, exploring the evidence for possible biological causal mechanisms of how such an effect can have a direct effect on health and the risk of mortality. Physical, economic and psychological impacts put unnecessary stress on the heart, causing it to increase mortality risk. This has been supported more recently through Muennig et al. (2013) who found that increased integration and social capital leads to reduced mortality risk (for both all-cause and cardiovascular causes) as well as lower blood pressure levels.

2.4 Evidence of area effects

With there being possible routes for areal influences to impact upon health, the question lies of whether there is evidence of area effects. Any examples need to be framed within the context versus composition debates, controlling for the latter so that the former can be tested for. Previously, this research has been restricted by methodological issues, however advances in the field have allow the testing of independent area effects resulting in a growing amount of evidence since the late 1990s (Diez-Roux, 2001). As a result, it is worth beginning with two influential and widely cited literature reviews; Pickett and Pearl (2001) and its update Riva et al. (2007). As they cover the majority of the literature, they provide a useful means to position ourselves in the debates surrounding area effects and health.

Pickett and Pearl (2001) produced the first review of the association between area effects and health, exploring studies published before June 1998. Comparing effects between various different outcome measures (morbidity, mortality and healthy behaviours), the authors report consistent and significant effects. However, the authors appear to downplay the fact that most results tend to only be small, or disappear when other results are controlled for (for example Reijneveld, 1998; Sloggett and Joshi, 1998). The majority of these studies gave little consideration to their use of areas for measuring effects. Without addressing this, it is difficult to conceptualise areas effects and their causal pathways. There were also issues with small numbers affecting the usefulness of results.

Riva et al. (2007) appears to acknowledge the limitations of Pickett and Pearl's discussion and is more conservative in its assertions. It focused on developments in the field and whether models were beginning to improve. More studies were included, using the time period of July 1998 to December 2005, reflecting the growth in available studies exploring this. Similar findings were reported with respect to observed effect sizes, with consistent but small results for both morbidity and mortality studies (although there is still some disagreement; Roos et al., 2004; Veugelers et al., 2001). Riva et al. argue that the poor application of methodologies has limited the current wealth of results. As such it is difficult to assess what is the true effect, especially as there is a fairly wide range of results. Furthermore, no studies have addressed the issue of meaningful boundaries to measure their areas.

There has been little change in the general results found since Riva et al.'s (2007) review, with the majority of studies reporting a significant but small observed effect on health (for example Maheswaran et al., 2010; Morris et al., 2008; Scarborough et al., 2012). However rather than just measure these effects, some studies have refocused their efforts to explore and test for possible causal pathways. Although Section 2.3 presented many different possible associations which areas could impact upon health, these are not always quantifiable. This is important due to selection bias occurring in the data (Bilger and Carrieri, 2013; Riva et al., 2007). Residents find themselves becoming sorted into specific neighbourhoods, which may also be related to their health. As a result, this can result in the mis-specification of the relationship, as effects become inflated. The need for better evidence regarding causal pathways is paramount for avoiding this issue.

Chaix et al. (2007) found the role of social support between individuals in an area to be important for explaining mortality rates. Controlling for individual and area level factors (including health risks such as previous diagnosis of diabetes or hypertension), they ran a multi-level model exploring patterns of Ischaemic Heart Disease (Table 2.3). Where there was less support or more unstable neighbourhoods (measured through population turnover), residential stability was observed as an important factor for explaining mortality. The size of effect between the extremes was equivalent to the size of socio-economic inequalities by neighbourhoods. For incidence, the effect was less than conclusive.

		Ischaemic Heart Disease			
		Incidence		Mortality	
		HR	95% CI	HR	95% CI
Neighbourhood socio-economic position	High	Reference			
	Mid-high	1.3	1.08-1.54	1.21	0.92-1.75
	Mid-low	1.35	1.13-1.63	1.36	1.00-2.02
	Low	1.67	1.39-2.03	1.85	1.37-2.72
Residential Stability	High	Reference			
	Mid-high	1.01	0.84-1.18	1.28	0.94-1.72
	Mid-low	1.06	0.89-1.25	1.47	1.05-1.98
	Low	1.19	1.00-1.41	1.89	1.39-2.52

Table 2.3: Hazard ratios for contextual factors in explaining Ischaemic Heart Disease patterns1996-2002, after controlling for individual factors (Chaix et al., 2007, p108)

Subramanian et al. (2008) analysed the concept of the 'widowhood' effect. This is the well established increased risk of death after an (usually elderly) individual experiences a bereavement of a partner (around 50-90 per cent for the first few months after, declining to on average an increased risk of 15 per cent overall). The authors found that individuals (after controlling for compositional and contextual risk factors) who lost a partner in an area with a low concentration of widowed individuals had an increased risk of mortality of 22 per cent for males and 17 per cent for females. However where there were higher concentrations of widowed individuals in a neighbourhood, the odds ratios decreased to an increase in the risk of death of 17 and 15 per cent respectively. Contextual support can act as a buffer to protect individuals, showing that "...neighbourhoods [can] modify the effects of individual risk factors." (p882). Reasons for this include the greater supply of social support, given that a spouse is typically the

most important part of a social network. Berkman (1995) added to this to note that bereavement or loss affects individuals through increased stress on the heart, as well as a suppressed immune system.

Johnson et al. (2012) analysed the health (the reporting of good health) of individuals late into the life course (aged over 55) and the subsequent risk factors at two points earlier in time (childhood and young adulthood) within an area effects framework. They found that the neighbourhood effect was stronger earlier in life. For example, the size of the effect of living in a medium-poverty (coefficient (β) = -3, $p<.001$) area than compared to a low-poverty area (reference) during young adulthood (aged 20s to 30s) was slightly higher than compared to whether a person smoked (β = -2.7, $p<.001$) during this age as well. For an individual living in a high-poverty area, the effect was higher (β = -6.4, $p<.001$). Whilst the role of neighbourhoods on health overall was not as large as the individual level, it still accounts for a considerable amount of variation, especially when individuals are exposed to it over a longer amount of time. However as Keene et al. (2013) found, living in a neighbourhood longer is not completely bad, as it allows the gathering of social capital and the formulation of support networks, which can positively impact upon health.

2.5 Migration as a possible factor to identify the impact upon health of areas

If area effects are important in influencing our health, then we also need to explore the role of migration within this context. If areas have an effect on health, then migrating to an area with different characteristics would also affect the migrant's health. For example, if subject A lived in an area with a good mortality profile, then we may expect that migration to an area which displays poorer health prospects would be detrimental to subject A's own health.

There is little research outside of the international migration context focusing on the role of internal migration and health. Although the theoretical side has been long hypothesised (for example Farr, 1864), it was not until a study by Bentham (1988) examined the evidence for this. Bentham compared the reporting of people identifying themselves as 'permanently sick' between internal migrants and those who did not migrate between 1980 and 1981. Those who migrated, especially at younger ages,

reported lower proportions of poor health. However this study was limited through its descriptive analytics, focusing less on any migratory effect.

Brimblecombe et al. (2000) analysed the impact of migration on health between birth residence and current residence using the British Household Panel Survey. Migration was split by destination into high and low mortality areas (Local Authorities). They found mixed results with some migrants who migrated from high to low mortality areas reporting significantly better health (through mental health, reporting of poor self-rated health), but with others showing an insignificant improvement in the health outcome measure (SMR, health limits work). However this was largely explained through social characteristics, of which the majority were significantly different, pointing towards people of higher social class migrating from high to low mortality areas. The results for migration from low to high mortality areas was less conclusive with fewer significant differences, a factor of the smaller sample size as the majority of migrations represent an up-scaling.

What Brimblecombe and colleagues found was evidence of healthy selective migration. Migratory patterns act like an internal sorting process, whereby those with the best health migrate into the areas with the best health. It is 'selective' as these migrations are undertaken by certain groups (mainly the young and those of high socio-economic status), who in turn have the best health prospects (Bentham, 1988; Boyle, 2004). At the same time, those with poorer health end up 'drifting' towards similar areas, leading to a concentration of ill health in particular areas (Cox, 2007). This leads to the polarisation of health patterns.

Larson et al. (2004) focused on middle aged females (45-49) in their analysis of the association between health and migration (Table 2.4). For this group of individuals, migration was associated with the onset of poor health (not self-rated health, but chronic diseases). Movements were more common over short distances, usually in relation to housing insecurity and declining incomes due to the onset of poor health. Long distance migration differed as it was due to individuals being closer to health services.

The relationship between migration and health is mostly indirect, more a method for understanding health differences. Research in the early 2000s focused on whether geographical inequalities in health could be part explained by migration. The consideration of (health-selective) migration was often ignored in most prior papers on health inequalities (Boyle, 2004). Simply comparing the change in mortality over time

to explore inequalities is somewhat false, as it assumes that you are comparing the same population, which through migratory patterns is not true (Brimblecombe et al., 2000). It is difficult to assess whether health has actually got worse or rather the people living in the area has changed.

Health measure	Local migration	Migration between postcodes	Rural-to-urban migration
Self-rated good health	0.88	0.93	0.87
Two or more chronic diseases	1.29**	1.24**	1.34*
Poor mental health	1.40**	1.32**	1.07
Current smoker	1.39**	1.35**	1.24
Three of more visits to a specialist in the previous year	1.12	1.30**	1.49*

Table 2.4: Selected odds ratios for migration types (compared to non-migrants) of females aged 45-49 accounting for socio-economic and marital status (after Larson et al., 2004, p2156).

*Note: *p<0.05, **p<0.01.*

These case studies so far have presented the role area effects on health resulting from migration as indirect. However exploring the US policy 'Moving to Opportunity for Fair Housing Demonstration' (MTO) allows us to see a more direct role. Families in five US cities (Baltimore, Boston, Chicago, Los Angeles and New York) between 1994 and 1998 who were in small areas with poverty rates above 40 per cent were approached to take part in a housing lottery (Ludwig et al., 2012). Of those who agreed, they became part of a randomised control study, where individuals were randomly selected to either to received a voucher to enable them to migrate only to an area of low poverty (<10 per cent; through subsidising rent), a traditional housing voucher with no restrictions on where they can migrate, or a control group with no encouragement to migrate (Ludwig et al., 2011). Families were then followed over time.

As such, studies have examined the impact of the policy on individual health outcomes,

an important consideration for the role of neighbourhood effects. Ludwig et al. (2011) analysed obesity (through BMI) and risk of diabetes (through blood glucose levels) variations as part of the study. Those who were migrated to less deprived areas had significantly lower morbid obesity (BMI ≥40; although not for just obese i.e. BMI ≥ 30) and blood glucose levels. There were no significant differences between those who were provided the traditional housing voucher and the control group, showing place (and where people migrated to, to be paramount). Ludwig et al. (2012) showed that the subjective well-being of those who migrated to a low poverty area was higher than compared to the control group. Leventhal and Brooks-Gunn (2003) also found improved mental health for both parents and young children for those migrating to the low poverty areas. Learning from this research design would help frame a useful empirical model of analysis.

The main problem with this small research field is that studies focus on just linear relationships between migration and health. This assumes the relationship to be constant across a range of different types of migration (for example from good to bad areas and *vice versa*). Section 2.1 and 2.2 argued that geography is important for explaining patterns in health and therefore ignoring it in this context would be false. Those few British studies which have considered geography loosely are presented in Table 2.5; however there is much scope for improvement. Even the MtO studies were concerned with policy evaluation, rather than the possible range of different experiences by area types for both origin and destination. There is currently a poor conceptual framework in research for introducing geography into any analysis and the link towards the neighbourhood effects literature remains poorly developed.

Study	Use of geography	Limitations
Brimblecombe et al., 2000	Low vs high mortality areas	Binary divide, restricting possible experiences
Popham et al., 2011	Glasgow vs 'Three Main Cities' vs Elsewhere	Limited to Scotland, large area types lose accuracy and detail
Riva et al., 2011	Rural vs urban	Test of specific process
Wannamethee et al., 2002	North vs South	Binary divide, large scale

Table 2.5: Past research utilising geography in the migratory impact on health.

Data issues form a large part of the problem which limits the quality of the current known evidence. There is little data which combines both migration and health (Boyle,

2004; Larson et al., 2004). Where it does, sample size is important and many studies note that over the course of a year only a fraction will migrate. This leads to the effects either being unobserved or just small. This becomes further problematic once geography is involved, through constraining the sample size further (Brimblecombe et al., 2000). A trade-off ensues, with larger sized areas more likely to give significant results but forego some level of accuracy and detail (Popham et al., 2011).

Data limitations extend to the temporal aspect as well. For example Brimblecombe et al. (2000) compared migration using current area of residence and that at birth to explore the role of migration. However the importance of area effects cannot be drawn out along such a long time period, given that many migrations may take place. This format (and others looking using long time periods) are fairly common within this research area (Cox et al., 2007; Popham et al., 2011; Wannamethee et al., 2002). Few have looked at patterns over small time units. Larson et al. (2004) compared data from a two year interval, however the study was focused only on females aged 45 to 49, restricting the observations that can be drawn out. As such, there are few studies which can really account for the role of the neighbourhood in the migration context.

2.6 Common research methods in health and geography

With the greater availability of small scale health statistics, there has been an increased focus on mapping patterns to explore the spatial variability. Whilst disease mapping is relatively descriptive, it provides a visual exploration of the processes that exist and are easily interpreted (Lawson, 2013). This method provides an effective and easy to understand interpretation of spatial patterns and processes, especially since 'a picture tells a thousand words'. This can be seen in a recent atlas (Shaw et al., 2008) which shows geographical patterns for most types of mortality. The importance of mapping as a method has been helped by the advances in GIS (Geographical Information System), which have allowed spatial data to be displayed quickly. Map overlay functions have made it easier and quicker to compare different patterns, to look for common trends and possible influences. Therefore maps are not just useful for describing the data, but also for generating new hypotheses as well (ibid).

These methods remain mainly descriptive and therefore analytical methods are required to fully explore spatial influences upon health. This is highlighted by the Shaw et al.

(2008) atlas, which just maps each mortality pattern rather than analysing trends, patterns or interactions between causes of death. There is little to link together possible explanatory factors or analyse interactions. Ecological analyses begin to bridge this gap, although they do not fully consider distinct spatial patterns. Studies may instead use correlation or regression to examine the associations of area level variables and their effects. Within regression, simple dummy variables can be included to represent whether a case falls in a certain location/area. The standardised residuals can also be mapped to look for extreme areas, suggesting possible geographical patterns. However the presence of spatial autocorrelation violates the underlying assumptions of regression since the residuals are not independent (Rogerson, 2006).

A method which has grown in use since the late 1990s has been hierarchical linear models (also known as multi-level or mixed-effects models, depending upon discipline). This method allows for the separation of the effects where there is more than one level to the structure of the data (Hox, 2002; Snijders and Bosker, 1999). This makes it useful in this discipline, given that individual level effects can be controlled for, whilst testing for whether there are significant differences by areas (Kreft and De Leeuw, 1998). As such, all of the evidence for the existence of area effects shown in Section 2.4 is taken using this methodology.

Whilst all of these methods analysing the geography of health are useful, there is one issue. All of these methods focus on analysing a singular disease and whilst this is not always a problem, it restricts the analysis to comparing patterns for each disease. Rather when a particular research project has access to data for many different diseases, with varying patterns and co-associated processes, it would be useful to look at how these interact with each other. As such, we can begin to move from a one-dimensional form of investigation, towards a multi-dimensional approach which would help provide a deeper understanding of new patterns and processes. However when multiple causes are analysed, it may be hard to 'see the wood for the trees'. Area classification methods provide a useful technique to solve these issues and therefore it would be useful to explore this further as a means to analyse mortality patterns.

2.7 Area classification and health

At its most basic, classification is the technique of summarising objects through defining clusters which are believed to be similar (Everitt, 1979). Cluster analysis represents one branch of possible methods for classifying data. Cluster analysis itself is also a general term for a family of statistical methods used to group a diverse range of heterogeneous cases into a smaller number of more homogenous clusters (Gordon, 1999). Cases are assigned to clusters based upon their similarities on a number of different dimensions (i.e. internal cohesion) and differences with other cases/clusters (i.e. external isolation) (Everitt et al., 2001).

Classificatory methods are different from other methods of analysis. Rather than test hypotheses (for example, as in regression), cluster analysis seeks to analyse how cases are related (Luke, 2005). The analysis is more exploratory, looking to describe the structure of the data, albeit in a much simpler way which retains most of the information (Cormack, 1971). As Everitt et al. (2001) note; "Cluster analysis is essentially about *discovering* groups in data…" (p6) and hence it can find new and interesting groups within society that can drive future research. Understanding these hidden and unknown clusters with the data can advance our knowledge about particular patterns and processes. The analysis is pre-classificatory, since no pre-conceptions about the number or the characteristics of clusters are known beforehand.

2.7.1 Why classify?

Classifying objects can appear very much a part of human nature (Cormack, 1971; Everitt et al., 2001). We, as a species, seek to bring order to the world that we live in. Classifying is very much a means of simplifying the many complexities that exist (Singleton and Longley, 2009). The human mind cannot think of every encountered object as unique. This would be too difficult to make any sense of what the object means and any patterns or processes occurring, as well as being very inefficient. Classification is therefore required to be able to understand the world around us (Clatworthy et al., 2005). By simplifying the vast array of objects into a number of smaller, more manageable clusters based on similarity (and dissimilarities against other clusters), we can improve our understanding of what is happening. Essentially, we are transforming reality into something more manageable (Dorling, 2012).

At its core, classification provides a method of data reduction. However, this does not mean the method is purely for data organisation, rather the simplification of complex data into more homogenous clusters can help analyse similarities and differences within the data set that would be hard to otherwise see (Cormack, 1971). The mortality data set used within this thesis contains millions of deaths, spread over 30 years for the whole of Scotland, England and Wales (see Chapter three). This is far too much information for the human mind to effectively analyse. Yet if this complex data set can validly be summarised by a smaller range of homogenous clusters, then the clusters may improve our understanding of the mortality patterns in Britain (Everitt et al., 2001). This approach allows us to 'see the wood for the trees'.

The ability of the method to process and reduce large quantities of data to a much smaller number of clusters has been helped by the recent advances in computing power. In the past, powerful computers were needed to produce classifications, one reason why classifications used to be mostly found within the private sector (Harris et al., 2005). Yet now, not just can more data be incorporated into analyses, but they can also be run on personal computers with less technical skill required; it would appear now that just about anyone can classify!

This data reduction does not just apply to the number of cases, but an increasing number of variables can also be included. This is very useful since it allows a multidimensional approach, giving a more detailed analysis through larger numbers of variables describing the characteristics of clusters. For example two areas with the same standardised mortality rate may suggest that they are similar, but a wider analysis may show that the types of mortality causes may be completely different (but just add up to the same standardised mortality rate) suggesting that the areas are in fact very dissimilar. As such, we can begin to move away from a uni-dimensional form of analysis and understanding.

The effectiveness of producing a classification to help the understanding of objects can be shown by the example of the board game Monopoly (Harris et al, 2005). Whilst not a 'true' classification, the board game reflects an urban classification of a city of interest. With no prior knowledge of a city, players know that an area that falls within the 'brown' category will have low rental value and development cost, reflecting the land value of the area. This is in comparison to a 'dark blue' area which has a much higher cost. From this, you can work out that 'dark blue' areas are more upmarket and

make other assumptions (carefully) about the characteristics of the area, especially when comparing it to other areas.

2.7.2 Why classify areas?

Area classifications extend the ideas presented in the previous section by incorporating an increased focus on geography. But why should we consider looking at the health of an area? If we are interested in the patterns concerning human life, then we need to also consider place. Where you live is likely to be a good indicator of your life. It limits, creates and determines your life course (Diez-Roux, 2001). Sections 2.1 to 2.3 highlighted how geography can impact upon your health, but it is not just restricted to these effects; for example income, employment, lifestyle etc (these all have additional effects upon health). By examining similarities and differences between places, we can explore these inter-relationships and their subsequent patterns. It is these patterns, which otherwise may be ignored, which help further our understanding or potentially show up new strands of research direction. It is through place, that the underlying structure of society becomes visible and interpretable (Dorling, 2012).

Place can also hide as much as it reveals. Through aggregating data, detail is lost. A trade-off ensues, given that through aggregating, the data becomes more useful and can be applied (and compared) to much more. The simplest way to solve this is to use the lowest geography possible. Where larger geographies are used, statistics will begin to regress to the mean as variations are lost. For example if an area is heavily polarized by lots of people with high and low life expectancies, the average will not truly reflect the area through giving a middle value. This describes the area as something that it is not (Dorling, 2012). Smaller areas reduce the likelihood of this happening, as areas become more distinct and compact. It becomes less important where the lines are drawn (i.e. the geography used). The underlying structure becomes more visible.

The classification of areas as a research framework has been mostly utilised within ‘geodemographics’. The subject field seeks to explore the types of people who live in different areas. For example we tend to think of areas within a city as containing differing groups. Typically these groups are defined as rich or poor, but having different labels such as ‘Little Italy’ or ‘Chinatown’ helps describe areas much better, improving our understanding of the processes occurring (Harris et al., 2005).

Using small scale geographies as the building blocks for a classification would provide an accurate representation of the geographies of mortality. Although some data quality is lost from aggregating individual level data, at small geographies it is likely that the types of people living in each area are quite similar. Tobler's (1971) first law of geography states that "Everything is related to everything else, but near things are more related than distant things." (p236). Whilst people who live in the same area are not identical, they are likely to have similar characteristics. Using small areas as the building blocks of a classification will therefore be useful. This approach assumes that population and place are intricately linked together (Vickers, 2006).

Area classifications also extent the principle of Tobler's first law further (Harris et al., 2005). People may not just be similar to those closest, but also to others living in the same 'class' or neighbourhood type. Whilst people living at Land's End and John O'Groats are geographically separate, they may share similar characteristics and hence be potentially related. Similar processes may exist in both regions causing a distinct health profile. This could not be easily shown using one-dimensional measure or a different methodological approach.

Socio-spatial classification has a long history within geography itself. Just before the turn of the 20th Century, Charles Booth (1889) published possibly the earliest area classification. Booth produced a socio-economic classification of London, split into 7 groups based mainly on social class. Booth's actual aim was to prove that poverty in London was not as strife as had been claimed, but he actually found it to be higher than he had imagined (Harris et al., 2005). Booth also recognised the importance of place in shaping the distribution of poverty, an early insight into area effects. Another example is William Burgess (1925) and his Concentric Zone Model, which sought to classify the general structure of cities at the time. Cities were divided into zones which each had different land uses. Whilst not intended as a socio-economic classification (it was produced to explain the process of urban growth), the model demonstrates spatial inequalities, with poorer classes constrained to inner city areas and higher classes able to migrate to the outer rings of the city. Both these examples highlight just how useful it can be to classify areas, despite neither being statistically grounded.

Focusing on place appears theoretically grounded to aid research. It helps to simplify patterns and although some detail is lost, it is much easier to see millions of deaths viewed through a few thousand areas. However, this choice is also partly down to data

considerations. It does not just get around issues of confidentially and ease of handling, but can enhance the quality of the study. Using data recorded at a known spatial level allows for comparisons to other sources of information. Relationships can be examined past the initial set of variables used, further improving our interpretation of patterns.

2.7.3 Is it useful for researching mortality?

The application of area classifications within health-based research has been minimal. It is an under-utilised methodology and yet it has the potential to be extremely informative. This is despite calls for increased focus on this approach within governmental policy and research (DoH, 2005). Its use is a relatively new phenomenon for researching health and hence most of the existing evidence and knowledge base has occurred from different sectors (Abbas et al., 2009). Nonetheless it might be first useful to explore some of the small range of examples of its application in past literature to highlight the lack of utilisation of the technique and how this can be improved upon.

Within health research, classification techniques are mostly used to group together individual level data on patients of a specific disease. Fukouda et al. (2007) classified the symptoms presented by elderly cardiac patients with poor recovery after initial treatment to improve diagnosis of further treatment options. Based on a variety of symptoms, three clusters were found representing different responses to treatment; 'weary', 'diffuse symptoms' and 'shortness of breath'. The weary cluster had the poorest recovery, with high levels of shortness of breath, fatigue and sleep disturbance. Diabetes and depression were found to be possible explanatory factors for the cluster. Diffuse symptoms were characterised by just high levels of fatigue, with little else to understand their condition. The shortness of breath cluster was self-explanatory, although it did show worse mental health outcomes. There were no significant demographic differences between the clusters.

There has also been research into classifying infant mortality. The use of the International Classification of Diseases (ICD) for completing death certificates is only limited since the available codes do not relate to cause of death at this period of life, but rather the time period of occurrence. Alberman et al. (1994) attempted to classify infant mortalities into meaningful clusters to help our understanding of what is happening. They found eight clusters in producing a neonatal classification for the

OPCS (Office for Population, Censuses and Surveys). Whilst showing the main causes of neonatal deaths, the study also highlighted how deaths during the neonatal period mainly the result of prenatal conditions. This contrasted with infant deaths post-neonatal where environmental hazards were more important. The Perinatal Society of Australia and New Zealand produced two classifications based on perinatal and neonatal deaths, with the aim to improve surveillance of infant mortality causes and to help improve health care accordingly (Chan et al., 2004).

These types of use, however, ignore the importance of place due to data restrictions. Most previous national low level area classifications have failed to incorporate health dimensions into them. With geodemographic classifications striving to classify the different types of people mainly through a socio-economic classification of Britain, health variables are often ignored (especially due to an often lack of availability at lower levels). For example the Output Area Classification (Vickers, 2006) only included two, albeit subjective, health-related variables ('limiting long-term illness' and 'provision of unpaid care'). Other commercial classifications have some small function of health, although given their commercial nature it is never clear the full extent of this, to protect their intellectual property (Harris et al., 2005). MOSAIC contains a health domain but it is less clear what this constitutes (Experian, 2009). ACORN contains information on 'smokers', 'diabetics' and 'hospital admissions due to heart failure', but together these appear almost random for their inclusion (CACI, 2013).

'HealthACORN' appears to be the only national classification designed with a focus upon health. Nevertheless the variables included in the process tend to concentrate upon health-related behaviours (i.e. surveyed data) rather than actual health outcomes (though some are used; CACI, 2010). This includes variables on exercise routines and diet, such as the number of rashers of bacons eaten. Whilst this may allow researchers to analyse what constitutes poor health (i.e. policy targeting) and help tackle future problems, it ignores the spatial patterns of health which may be important. With the methodology having not been disclosed to preserve the intellectual property rights, it is hard to assess the overall quality of the classification (Harris et al., 2005). Also as a lot of the data comes from modelled survey estimates (collected at a high level), the reliability of the low level clusters may be questionable. Currently, the 'HealthACORN' has been withdrawn from CACI's range of products and is unavailable for use at all, with no plans to replace it (Thurman, 2013).

Although these national classifications mostly ignore health variables, they have still been quite useful within research. A lot of studies have used classifications (rather than produce their own) to explore the relationship between population and health, specifically the characteristics of people at greatest risk of a disease. There is growing awareness within health research of using classifications as an alternative to indices or single explanatory variables to help improve our understanding of the determinants and patterns of health (Abbas et al., 2009).

Nnoaham et al. (2010) highlights this through an exploration of the uptake of colorectal screening. They used the 'People and Places' classification to help gain a deeper understanding of the types of people not using the service, to attempt to improve uptake. The geodemographic classification outperformed using deprivation quintiles instead, finding greater variation in uptake patterns. Three area types had lower quantities of uptake (see Table 2.6). Although useful, targeting gender appeared to provide greater possibilities for increased uptake. Nonetheless the classification helped improve the understanding of the types of people affected, resulting in public health policy targeting being improved.

Cluster	Description
Multicultural centres	Broad ethnic mix, most renting council or housing association property and in the lowest quartile of income for the UK
Disadvantaged households	Single pensioner households, most renting council or housing association property, and in the lowest quartile of income for the UK
Urban challenge	Single pensioner households, most renting council or housing association property; smokers, long-term limiting illness; in the lowest quartile of income for the UK

Table 2.6: Descriptions of the 'People and Places' clusters of increased non-uptake for colorectal testing (Nnoaham et al., 2010, p578).

These classifications however do not deal directly with the health of areas, rather the make-up of who lives there. It is surprising then that despite the existence of wide inequalities within health (for example Thomas et al., 2010), there has been no low level classification of these mortality profiles. Shelton et al.'s (2006) study comes closest

through classifying causes of death within parliamentary districts for England and Wales. The analysis resulted in a ten cluster solution, which are presented in Table 2.7. The clusters highlighted the importance of occupational and lifestyle differences in creating health inequalities. Change over time for cluster membership was assessed and showed a relative stable underlying structure of mortality patterns (82 per cent of districts remained in the same cluster between 1981 and 2000).

Cluster	Geography	Mortality Patterns	
		High	Low
1	London	External causes, HIV	Influenza, senility
2	South East	Air accidents, assault by firearms, accident by electricity	Unknown causes
3	Coastal areas	Water transport, other external causes	Infectious diseases, respiratory diseases
4	Former mining areas	Respiratory diseases, machinery-related	External causes
5	London	HIV, external causes	Machinery-related, neurological
6	London	HIV, external causes	Machinery-related, unknown causes
7	South East (with some Northern)	Suicide, transport accidents	Infectious diseases, external causes
8	London	Assault by firearms and cutting, HIV	Senility, respiratory
9	Former mining areas and areas of high deprivation	Machinery-related, railway accidents, industrial lung-related	External causes
10	Urban (outside London)	Deaths from cutting, respiratory diseases	External causes

Table 2.7: Summary of Shelton et al.'s (2006) classification.

Shelton et al.'s (2006) study was limited, affecting the quality of its results. There was little basis for the selection of variables. For example, there were separate variables for both deaths due to air and sea transport accidents, neither of which account for a significant or useful number of deaths. This is reflected in the results, where the choices

appear to be having a fairly large impact upon the clusters produced (the coastal cluster was strongly driven by the extreme values of deaths due to sea transport accidents). It would have appeared that variable selection was based upon producing a classification which emphasized geographical inequalities, rather than actual mortality patterns. The study was also conducted at a fairly large geographical scale, which may have resulted in rates approaching the mean (Dorling, 2012). Smaller areas allow for the capturing of greater variation in patterns, which are otherwise lost where larger areas capture a heterogeneous population. As a result of both factors, the classification ends up telling a story about a population that does not really exist.

The Department of Health performed a cluster analysis of survey data on attitudes and behaviours to diet and physical activity. This was then applied to other data sources to produce an area based classification to aid policy with regards obesity as part of the 'Change4Life' campaign (NOO, 2009). It formed the basis for creating population profiles of varying sub-groups which could be further targeted based upon their risk to obesity (and diabetes). However, the resulting classification was not robust enough for further analysis, limiting its quality and usefulness to just an exploratory introduction into policy formation. The use of six clusters was not justified and hence appears arbitrary. Having been outsourced to CACI (makers of ACORN), whilst further explanatory variables were added to the area classification, these were unknown limiting any potential gains.

Whilst not a classification, the Department of Health also recently introduced 'Health Profiles' which are to be released each year (APHO, 2013). These summarise inequalities within key health factors (for example life expectancy, early deaths from heart disease and cancers), not just showing local health but also comparisons against national averages. The aims of the reports are to provide greater information on the health of the local region to help identify issues that need to be tackled (ibid). However not all the variables included are related to health, for example deprivation, lifestyle factors, education. Although these are included to help local policy, the irrelevance of a few variables detracts from the overall aim of assessing the health of the region (for example carbon emissions). These reports are currently produced as low as Local Authorities, with the aim to make information available at the MSOA level (not currently available). This example shows the potential for a similar type of study to help aid public health and policy research.

A lot of the examples analysing health variations tend to focus on one disease. Although useful, this approach ignores how diseases are inter-related through examining similarities and differences. Through reducing the data into more manageable clusters based on similarity (and dissimilarity with other clusters), this makes it easier to understanding the patterns and processes that are occurring. For example Shaw et al. (2008) produced an atlas of mortality, however with over 100 maps, it is hard to fully take in what is happening. The construction of a classification condenses this vast amount of information into the most useful parts, allowing us to 'see the wood for the trees'.

It also provides a useful framework for investigating the importance of geography and neighbourhoods in affecting our health. Here we get a better understanding of health inequalities, both through capturing the dominant patterns, as well as the multi-dimensional patterns by cause which have resulted or produced these patterns. This can be linked to other social factors, as both explanatory variables, as well as to highlight the social inequalities in health that exist. By knowing the dominant environment structures and types through mortality, only then can we examine the impact of areas of individuals.

Finally, the classification acts to improve the statistical robustness of any analysis. Investigations of low level mortality patterns are constrained by issues of low numbers. When measuring differences by cause of mortality, this is especially important and often restricts the observations and detail studies can achieve. Producing a classification from a cluster analysis pools data from across the areal unit to create better and more accurate results (Everitt et al., 2001). This allows the study to keep both a good level of detail through an increased number of variables at a low geography, without having to sacrifice either (a common limitation of most health-based research).

2.7.4 Other Directions?

Producing a classification may not be the best approach at summarising a large amount of data into more manageable chunks. One possibility would be to produce an index of mortality. Rather than classifying areas, an index would measure the different characteristics of health and then sort places based upon their rank or scores (Harris et

al., 2005). This involves the selection and amalgamation of several factors into a more useful composite variable (Nando et al., 2008). By using the ranks or scores, areas can also be grouped into differing levels of what is being measured. This allows the relative position of areas to be compared on an ordinal scale to examine patterns with higher or lower scores or ranks. Furthermore, this approach allows the comparison of trends over time a lot more easily than classifications, which are not comparable other than between their own clusters (Vickers, 2011). Indexes are also more easily updated than classifications (Everitt et al., 2001).

Producing an index also bears some similarities with a classification, which would be useful in researching health. An index, like a classification, could be used as a means of data reduction as the process would reduce the visible size of a set of variables without dropping the underlying information (Nando et al., 2008). This in turn allows areas to be compared on the same scale, allowing more meaningful comparisons between areas. The summarisation of the data means that both techniques are multi-dimensional. This allows both measures to provide a greater understanding of the health of areas, through simplifying the complex underlying patterns into an easier to interpret visualisation of what is occurring (ibid). Just analysing the geographies of mortality using one indicator (or dimension) will inadequately capture the true patterns and processes occurring.

Although creating an index may be useful, there are some differences to classifications which suggest that the latter would be better. Indices are created to rank areas based upon poor or bad health. This use is different to area classifications which would look to explore the unique characteristics of health profiles. If two areas were just ranked as 'bad', this ignores the diversity of health in Britain since not all 'bad' places are the same. It is also much harder to know if certain variables are having a greater effect on certain areas or clusters. Therefore it may be suggested that indexes oversimplify trends within the data (Richardson et al., 2010). Furthermore what actually characterises 'bad' health? This would be hard to justify. Classifications just produce clusters that are different, with no suggestion which are better (or can be ranked; ibid). It would also be difficult to interpret what the value of the score actually means and what a one unit change would entail. Thus a classification would be more useful with regards to this thesis.

Performing a principal component analysis could get around these criticisms. This

method also leads to data reduction, however instead of examining how cases are related, this analyses how variables interact with each other. When there are many variables (for example causes of death), the analysis examines how variables are correlated together. The process then creates a smaller set of new, unrelated variables which reflect the relationships between variables in the data set. Yet this approach is still limited in turns of examining which variables are more important. Whilst you can see the spatial patterning of each component, in areas where two components have similar scores it may be difficult to decide which has the greater effect. Cluster analysis does not encounter this issue as areas are assigned one value only.

2.7.5 Limitations of Area Classifications

Whilst these past few sections have shown that conducting an area classification would be useful, the limitations of this approach must be first be taken into consideration. It is questionable whether the clusters produced in an area classification actually reflect reality. Although there may be accuracy issues involved in the methodology, clusters may just instead be statistical artefacts since the methodology may force the data into clusters (Ketchen and Shook, 1996). Thus the method will create clusters even if there are no actually structure to the data.

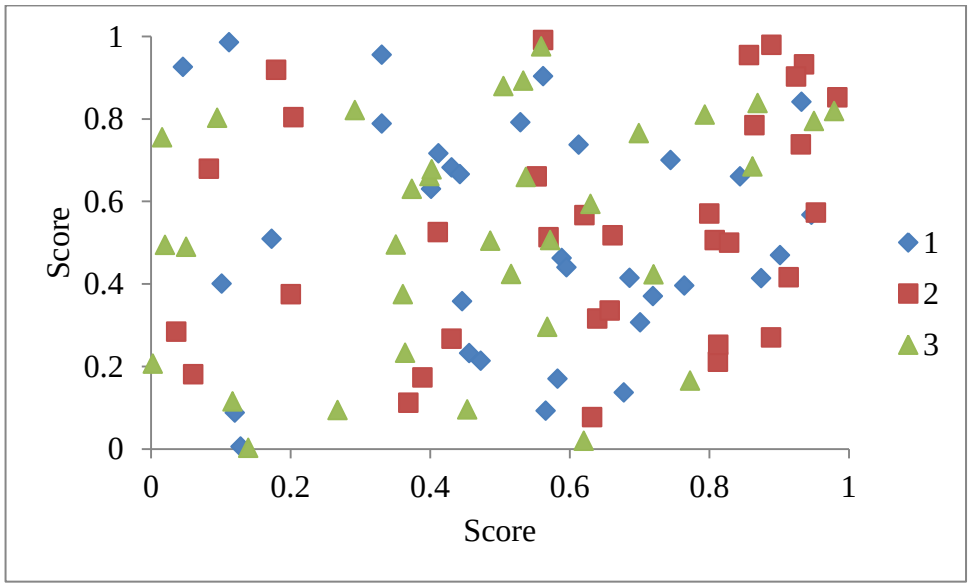


Figure 2.1: A classification of random data into three clusters.

This can be seen by examining Figure 2.1, which shows a three cluster solution of 100 random data points. However with random data, there is no relationship between the

variables. Therefore a random and meaningless result will only ever occur. Real data is not like that, so getting such a false result will not happen. Mortality patterns are not random and display strong geographies (Hacking et al., 2010; Shaw et al., 2008). It will not be an issue.

Most classificatory methods do not to offer a test statistic to check whether the clusters found were significant. This is because clusters are chosen to maximise the between cluster differences and hence any significance tests for this aspect will not be valid given it has been optimised to give the best result possible (Milligan, 1996). This limits the analytical approach when compared to other methods such as regression, which can be used to support or disprove hypotheses. However as Everitt et al. (2001) argue, a classification should be judged upon its usefulness rather than in terms of how 'correct' it is. It is a product of their application and how they can introduce new understanding to society, through a simplified analysis.

Despite the notion that clusters are produced to maximise differences between each cluster, cluster allocation can sometimes be arbitrary. Areas do not always strongly belong to a cluster and may be close to other clusters if they lay on the boundaries of the clusters parameters, especially with regards to broader clusters. Whilst this will always be a small issue, where there are large amounts of cases which could be in either cluster, this may affect the overall results (for example geographical patterns). This problem was present in Shelton et al.'s (2006) study with some less distinctive clusters where cluster membership rested on a 'knife edge'. There were four clusters of London areas, however these clusters had similar mortality profiles suggesting cluster membership could have been arbitrary. However since areas will be placed in the cluster they are most similar to, this should not be seen as false. When interpreting the clusters, this should always be kept in mind to understand the clusters best. What is of most importance is that the clusters represent the main patterns and therefore what they describe will not be affected by this.

An area classification represents an ability to group together data and hence is really a 'family' of (albeit similar) methods. However the choice of method is largely a subjective one which in turn may affect the end product. Therefore it is important to select the correct method, else the findings may not be truly representative results. Yet there is little evidence available that shows one method to be better than the other possible ones, rather each has its own benefits and limitations. There are also further

choices within each methodology, for example means of estimating distance between cases/clusters, reinforcing this notion of the importance of justifying methodological choices. There is a need to carefully detail these choices when conducting research, however as Clatworthy et al. (2005) found, many studies do not disclose this information within health psychology. For example 53 percent of studies failed to mention the similarity measure used and incredibly four (out of 59) did not even include the method used!

Another issue with producing a classification is that since the process is pre-classificatory and no prior knowledge of the clusters exists, there is an issue deciding how many clusters is the correct solution. Too few clusters and results become too broad; too many clusters and it becomes complicated and clusters are more likely to be affected by 'small number effects' (Openshaw and Blake, 1995). Having small sized clusters may offer greater detail but tend to represent outliers (Everitt et al., 2001). It is important that this issue is treated carefully. There exists plenty of test statistics which can help inform any decisions (Milligan and Cooper, 1985), however the subjective element still plays a key role and hence any choice will need to made carefully (and justified).

The naming of the clusters can also be problematic in creating a useful and applicable classification. Although this issue does not directly affect the data or the method, it can cause some issues. Part of the classification involves naming the clusters based upon their characteristics to give a general description of the cases. Nevertheless this can cause labelling issues through stereotyping. Through labelling areas with 'bad' characteristics, this may cause stigmatisation (Butler and Watt, 2007). Although outdated and unlikely to be repeated, one of Charles Booth's (1889) groups was named 'Lowest class, vicious, semi-criminal'. Thus areas may not just suffer from structural disadvantages but also the perception of areas can also be damaging. For example Lupton (2003) notes areas can suffer from postcode discrimination, especially in employment opportunities as employers are less likely to want to employ people from 'bad' areas. Labelling can also have important negative effects even if the characteristics are not truly reflective of the area and can also be linked to ideas surrounding ecological fallacies as well. This will need to be avoided as best possible, whilst still accurately depicting their profile.

There are some issues with comparing different cluster analyses. Each cluster analysis

is an independent analysis of a specific region and data set and therefore slight differences means you may not be comparing like for like (Vickers, 2011). Furthermore with such a range of different methods to utilise within the process, it makes comparing classifications difficult. This means that naming clusters can be difficult since two clusters with very similar names from two different classifications could easily be assumed to represent the same type of area with the same characteristics (Vickers, 2006). For example the Output Area Classification and the MOSAIC classification contain the clusters 'Blue Collar Communities' and 'Blue Collar Owners' respectively. Whilst these clusters share similar characteristics and patterns, they are both independent clusters and hence not truly comparable. This will need to be avoided.

The final issue considered here involves time. Classifications are static and cannot be updated once they have been created. This is because the results are sensitive to the data used and hence through updating the data set this will affect the results, meaning the clusters could be quite different. Thus it would appear that classifications become out of date quite quickly. Nevertheless it is questionable whether areas change their characteristics quickly, instead they may be static. For example Orford et al. (2002) found that by comparing poverty between 1896 and 1991 in London, despite absolute decreases, places remained relatively the same socially suggesting that areas do not change very quickly and thus this temporal issue should not be much of a problem when making short term comparisons.

2.8 Conclusions, literary gaps and directions for research

"Unless we believe that people live in a social vacuum, that physical proximity to others has no effect on our own behaviours and that everyone has an entirely unconstrained choice about where they live, then geography remains important…"

(Harris et al., 2005, p17)

The quote above effectively summarises the theme of this chapter. When researching health and mortality, ignoring the role of geography serves only to restrict any understanding of the patterns and processes that exist. It is only through place that these underlying structures which differentiate the health and death of the population become visible (Dorling, 2012). Geography should not be viewed simply as random noise. The

persistence of inequalities in mortality in particular places highlights this (Thomas et al., 2010).

Geographical research into health has come a long way in such a short amount of time. Despite scholars like Edwin Chadwick and William Farr making the link in the 19th Century, it was largely forgotten until the past 20 years. Recent years have witnessed an explosion in the focus on the geographies of health. Not just is it acceptable to analyse the inequalities that exist between places and areas, but causal pathways have been proposed and tested which strengthen the argument that geography is paramount. The role of geography has moved beyond being just a container for social and biological processes to occur in.

Despite all the research currently being conducted focusing on the geographies of health, there are clear gaps in the literature which could be exploited. Research remains mostly one-dimensional, looking at specific cause and effects for singular causes of mortality (or health outcome). Ignoring the multitude of experiences restricts our understanding of areas; especially in devising effective policies (tackling one disease may have knock-on effects on another). For example it may be found that certain causes which dominate the mortality profile of an area limit the propensity of other causes expected to be more common. Quite simply they are killing people off before others can get to them. However there is no past literature which examines these interactions of all causes of mortality, let alone how it varies geographically. It also builds upon previous studies, moving research towards a multivariate framework of investigation.

The literature review identified the potential for area classifications to improve our understanding of mortality patterns. As Shelton et al. (2006) argue "Whilst area-based geographic inequalities of mortality have been widely analysed and reported… area based classifications of mortality profiles have not been derived." (p558). Despite the popularity and effectiveness in other sectors of research, especially market research (Harris et al., 2005) and psychology (Clatworthy et al., 2005), their application in health-based research has lagged behind. This has led to some researchers to call out for wider implementation of these techniques and methods to help improve analyses where appropriate (Abbas et al., 2009). It has also been backed up by the government with regards policy research (see DoH, 2005; NOO, 2009). There is a clear gap for a detailed analysis using area classifications. Summarising small areas based upon their

mortality is important for targeting resources or locating services.

One reason why there have been few applications of health-related research and area classifications has been a result of data availability. Data is usually restricted in the detail offered, with a trade-off between the number of causes and geographical coverage at lower levels. Studies which contain data with lots of causes tend to be at higher levels (for example Shelton et al., 2006) whereas those at lower levels tend to be restricted to fewer health variables (for example Eames et al., 1993). However the release of a recent data set under less strict guidelines now allows for this project to be achieved. Further discussion of this will be provided in the next Chapter.

The production of a classification of mortality would advance our understanding through the new conceptualization of mortality patterns. Understanding the resulting clusters will further add to our knowledge surrounding mortality patterns, especially given there are few pre-conceptions of the results which would be expected. Although this remains the main focus of this thesis, beyond its interpretation the usefulness of the classification also remains a function of its application. It is important to highlight the usefulness of a classificatory approach to researching mortality through applying it to analyse some facet of research. Given the focus on the importance of geography as a framework for investigating patterns of health, this appears an appropriate start point. A classification of areas based upon their mortality profiles provides an ideal tool for analyzing whether the different types of areas display independent effects upon health. It certainly gives a greater attention to the formation and choice of area level measure, a more relevant selection to improve our understanding of the processes which exist.

Although this analysis addresses some issues with regards to the area effects field, the literature review has also identified a stronger, new and innovative research gap which would better tackle the area effects literature. If area effects are important, then it would be fair to assume that as people migrate between different areas then there should be some aspect of this change observed. Exploring internal migration in this framework improves and builds upon past research, producing a new area for investigation. A classification of mortality would be ideal for researching this as well, as migration between clusters would also be better designed for explaining possible reasons for neighbourhood effects on individuals.

Chapter 3: Data and Methodology

3.1 Introduction

Milligan and Cooper (1987, p331) identified seven critical steps for performing a cluster analysis:

1. "The entities to be clustered must be selected. The sample of elements should be chosen to be representative of the cluster structure in the population.
2. The variables to be used in the cluster analysis are selected. Again, the variables must contain sufficient information to permit the clustering of the objects.
3. The researcher must decide whether or not to standardise the data. If standardisation is to be performed, the researcher must select a procedure from several different approaches.
4. A similarity or dissimilarity measure must be selected.
5. A clustering method must be selected.
6. The number of clusters must be determined.
7. The last step in the clustering process is to interpret, test, and replicate the resulting cluster analysis."

The rest of this chapter is set out to follow and achieve this framework, providing an exhaustive understanding of the inputs and processes within the creation of the area classification. Justification of these decisions is important if others are to assess the quality of the area classification. Whilst it was designed for conducting a general classification of data, it is not completely applicable to the creation of an area classification. Vickers (2006; 45-46) provides a detailed description of the additional gaps requiring and therefore was consulted to ensure a focused and high quality area classification.

3.2 Mortality data

The main data source that is to be used for this study is the 'Death Occurrences in England and Wales, 1981-2009' (DOEW) gathered by the Office for National Statistics

(ONS). Although data can be requested for Scotland as well, due to inconsistencies in geographical scales and additional explanatory variables, it was chosen not to be used in this study. Accessibility to the data was approved by the 'ONS Microdata Release Panel' in December 2010, with 'Approved Researcher Status' also granted. There are two deaths databases; one contains all the information provided about the deceased (i.e. the public records) and the other has only the coded details of deaths (Devis and Rooney, 1999). The second database will be used in this study.

The DOEW is compiled through civil registration records and is the only system in England and Wales that collects individual level data on every death (Griffiths et al., 2005). It is a legal requirement that the General Register Office (GRO) for an area is notified of all deaths that occur in the respective area. Therefore complete coverage for England and Wales is captured.

The DOEW has been under-utilised in research. This is despite the introduction of the 'Statistics and Registration Service Act 2007' giving greater accessibility to the data set. Previously access was only provided to researchers who had received grants to investigate relevant topics. This restriction forced many researchers to use morbidity data instead.

Studies which utilise the DOEW tend to be analysing overall mortality with respect to some explanatory factor (for example Coggon et al., 2010) or are exploring the patterns and trends of a single cause of mortality (for example Brown et al., 2010). These studies are at higher geographies and ignore the inequalities between places. Few studies analyse many different causes of mortality and whilst Shelton et al. (2006) is the exception, other examples such as Shaw et al. (2008) only examine the individual patterns of many different causes, ignoring the connections between them. There is a gap to explore the interrelation of the many causes of mortality at a small geography.

There are, however, other sources of health data that could be useful to this study (i.e. available at small geographical scales). These examples tend to cover current health patterns rather than mortality. For example the ONS regularly produce life expectancy estimates at small geographies, although the lower the scale the greater the confidence intervals due to smaller sample sizes. The Association of Public Health Observatories released data for both incidence and mortality rates across England and Wales at the Local Authority level for a wide variety of cancers. Hospital Episode Statistics can also

be gained which provide morbidity data for a wide range of diseases, however access is restricted. The census also has some health variables ('limiting long-term illness', 'provision of health care' and 'people rating their health as good, fair or not'). However these variables measure perceived, rather than actual, health.

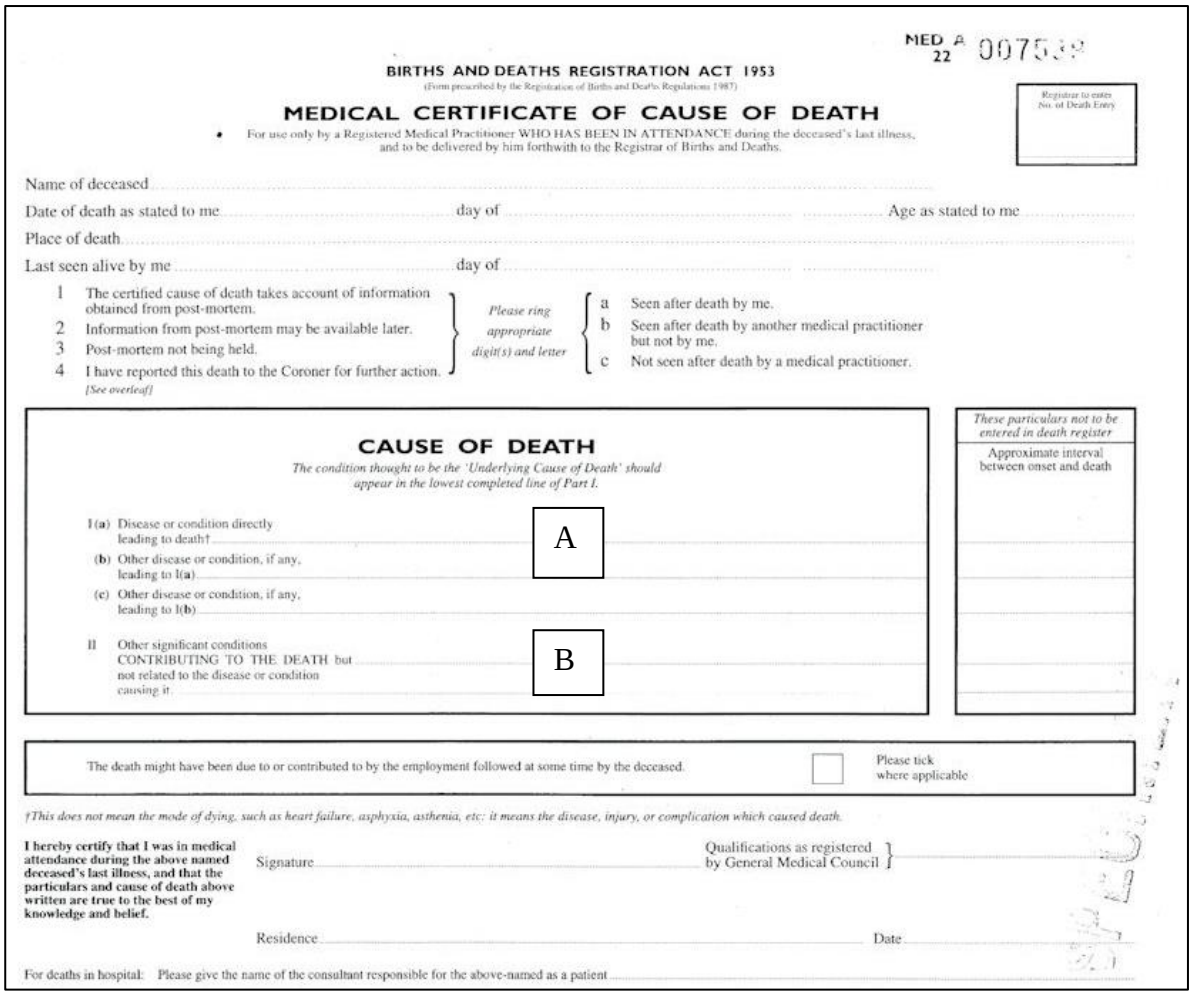
One major issue with these other sources of health data (excluding the census data) is that the data tends not to be released at low geographies. This would affect the overall quality of the research by restricting the analysis to higher levels, which may not fully capture the wide range of patterns. This is not the case with the DOEW, where the data is provided with identifiers for a range of scales including small scales. The DOEW also contains every known cause of death, providing a greater range of data than can be gathered overall from other sources (especially for less common causes). Data on current health ignores the fact that poor health can be treated, unlike mortality (!). There is also data for nearly the last 30 years provided by the mortality data and a lot of other sources only supply more recent years. Furthermore, as identified by the former Health Secretary Frank Dobson, the worst form of inequality is that between the living and the dead (Warden, 1998). Gaining a greater understanding of the similarities and differences between places for mortality may help tackle patterns and processes surrounding the worst health outcome.

3.2.1 Data collection

Death registration occurs through two-steps. Firstly the doctor who treated the deceased last completes a 'Medical Certificate of Cause of Death' (MCCD) (Figure 3.1). With unexpected, sudden and violent deaths or when a doctor was not present (or unable to identify the actual cause), this may be carried out by a coroner (nearly 25 percent of certificates are now conducted by a coroner; Devis and Rooney, 1999).

Medical details are provided detailing what caused the death of the individual. There are two parts to this on the form (labelled 'A' and 'B' on Figure 3.1). The part labelled 'A' details the causal chain of events that resulted in death. The second part ('B') provides any other factors which were of importance in contributing to the death of an individual. This includes diseases which an individual may have been experiencing which led to the death of the person but was not part of the causal chain itself (WHO,

2004). The lowest line filled in this section is then taken as the 'underlying cause of death'. All details are coded by a classification of diseases, as also defined by the WHO.



MED A 22 007539

BIRTHS AND DEATHS REGISTRATION ACT 1953

(Form prescribed by the Registration of Births and Deaths Regulations 1987)

MEDICAL CERTIFICATE OF CAUSE OF DEATH

• For use only by a Registered Medical Practitioner WHO HAS BEEN IN ATTENDANCE during the deceased's last illness, and to be delivered by him forthwith to the Registrar of Births and Deaths.

Registrar to enter No. of Death Entry

Name of deceased .....

Date of death as stated to me ..... day of ..... Age as stated to me .....

Place of death .....

Last seen alive by me ..... day of .....

1 The certified cause of death takes account of information obtained from post-mortem.
2 Information from post-mortem may be available later.
3 Post-mortem not being held.
4 I have reported this death to the Coroner for further action.
[See overleaf]

Please ring appropriate digit(s) and letter

a Seen after death by me.
b Seen after death by another medical practitioner but not by me.
c Not seen after death by a medical practitioner.

CAUSE OF DEATH

The condition thought to be the 'Underlying Cause of Death' should appear in the lowest completed line of Part I.

I (a) Disease or condition directly leading to death† ..... A
(b) Other disease or condition, if any, leading to I(a) .....
(c) Other disease or condition, if any, leading to I(b) .....

II Other significant conditions CONTRIBUTING TO THE DEATH but not related to the disease or condition causing it ..... B

These particulars not to be entered in death register

Approximate interval between onset and death

The death might have been due to or contributed to by the employment followed at some time by the deceased. ☐ Please tick where applicable

†This does not mean the mode of dying, such as heart failure, asphyxia, asthenia, etc: it means the disease, injury, or complication which caused death.

I hereby certify that I was in medical attendance during the above named deceased's last illness, and that the particulars and cause of death above written are true to the best of my knowledge and belief.

Signature ..... Qualifications as registered by General Medical Council .....

Residence ..... Date .....

For deaths in hospital: Please give the name of the consultant responsible for the above-named as a patient .....

Figure 3.1: The 'Medical Certificate of Cause of Death'.

The certificate is then taken by an informant (usually a relative or someone who was present when the deceased was last alive) when they go to register the death in the GRO of the area where the person died (within five days of its occurrence). Additional information is provided about the individual including address of last residence and occupation, as well as basic personal information (date of birth, marital status, name etc). This information is stored locally, with a copy also sent to the ONS which is entered into the DOEW database.

3.2.2 Data quality

It is first worth discussing the possible accuracy issues within the data set which involve the cause of death data recorded. The first ICD was introduced to standardise mortality statistics between countries to allow for better comparisons (WHO, 2004), with the effects also benefiting statistics within countries as well. When this was first introduced in 1900, there were large differences in the causal information recorded. The quality and reliability of the data sets varied both geographically and temporally (Woods, 2000). Mis-specification of cause of death was a common problem, however this was over a hundred years ago and much has changed since.

There were still some evidence of inaccuracies between cause of death recorded and actual cause found in studies during the 1970s and 1980s. For example Hoel et al. (1993) found in a large study that the detection of cancers in Japan should have been 18 per cent higher than were actually reported. Nevertheless, this gap has been reduced through improved technology and diagnosis equipment, as well as better knowledge, improved teaching and the discovery of new diseases (Devis and Rooney, 1999). Furthermore the latest incarnation of the variable has turned the data set into one of the strongest and most reliable available globally (Rooney et al., 2002).

The most effective means of accounting for causal information errors is to carry out an autopsy (Sington and Cottrell, 2002). This can establish the main cause of death, as well as validating the causal chain of events that led to the death. However to be able to test the accuracy of the data set, autopsies would need to be performed on a large proportion of deaths. This would be highly expensive and time consuming to initiate. This has led to many studies which aim to test the accuracy of the system having small data sets, especially since autopsies are usually only carried out on specific cases (where there is an unknown cause, or the cause is thought to be criminally linked or due to an accident). For example Tuffin et al. (2008) questioned the accuracy of junior doctors (suggesting that around half were filled in wrongly) and patients who displayed co-morbidity, despite a sample size of 30 cases. Furthermore Sington and Cottrell (2002) also note that where autopsies are used, the data set becomes over-represented with harder cases which may have multiple causes present or cases which were sent for review by the coroner as the cause unsure (and the data is later updated). The number of autopsies being conducted is also in decline, providing less feedback on diagnosis patterns (Selinger et al., 2007). This method also assumes that coroners views are

correct (Hoel et al., 1993), a similar issue with the inaccuracies of death certification and doctors.

More recent studies which have examined the accuracy of death certificates have found that whilst many contain errors, these are likely to be additional information not relevant to the study (Selinger et al., 2007). Most of these issues arise from a lack of precision rather than accuracy, whereby a more common name is used or information of site are missing on the certificate (Swift and West, 2002). This is the equivalent of a doctor recorded lung cancer but failing to include the specific type or exact location within the lung. These issues will not be much of a problem for the data used, since the specific causes will be aggregated to improve interpretation. Other common errors include specific conditions missed from the causal chain of events that led to death, even though the main cause is correct (which is most important).

3.3 Variables included in the database

Seventeen variables are included in the initial data provided by the ONS, which can be seen in Table 3.1 (also included are various summary notes about the data). Most of these variables are obvious in what they describe. However some further details are provided for the cause of death and social class domains.

Domain	Variable	Coverage	Missing data
Time	Year of Death	1981-2009	0.00%
	Month of Death	1981-2009	<0.01% since 2000
Demographics	Sex of the deceased	1981-2009	<0.01%
	Age in Years of deceased	1981-2009	0.00%
Cause of Death	ICD-9 Cause of Death	1981-2000	<0.01% (above age = 0)
	ICD-10 Cause of Death	2001-2009	<0.01% (above age = 0)
Social Class	Occupation Major Group	1981-2000	64.50%
	Standard Occupation Code 1990	1993-2000	91.30%
	Standard Occupation Code 2000	2001-2009	94.10%
	NS-SeC 1	2001-2009	65.90%
	NS-SeC 2	2001-2009	65.90%
Location	Tract Number	1981-2009	0.00%
	1981 Ward Number	1981-1990	0.00%
	1991 Enumeration District Code	1991-2000	0.00%
	2001 Output Area Code	2001-2009	0.00%
	2001 CAS Ward/Sector Code	2001-2009	0.00%
	Postcode	1981-2009	0.00%

Table 3.1: The original set of variables

3.3.1 Cause of death

There are two variables for cause of death provided. These are both based upon the 'International Classification of Diseases' (ICD) developed by the World Health Organisation. The classification was originally created by Jacques Bertillon in 1900,

based on the works of William Farr, to classify disease based upon organs or body parts affected, as well as other general diseases (WHO, 2004). Its aim was to improve the efficiency of medical statistics in countries, as well as providing a standardised means for comparing between nations. The classification is constantly revised and changed to incorporate medical advances and the discovery of new diseases (Rooney and Smith, 2000). The tenth revision (ICD-10) is the most up-to-date classification and was introduced in England and Wales in 2001. The ninth revision (ICD-9) is also included to cover the earlier years in the database. However back- or forward-mapping is not encouraged as each classification is different, making them incomparable (Rooney et al., 2002; WHO, 2004). Therefore it was chosen to only use data for the years available with the ICD-10 data.

Data coding through the ICD-10 is based upon a hierarchical classification. The ICD groups together individual causes of mortality into broader ICD chapters which reflect their similarities. These ICD chapters are based on diseases of specific organs, pathology, aetiology, as well as more external causes or those related to specific time periods (Griffiths et al., 2005). Although these ICD chapters are based upon similarities between types of diseases, their assignments are not always clear cut. For example influenza highlights the complexities involved with group allocation, since it could both be an infectious or a respiratory disease (ibid). Table 3.2 shows the ICD chapters used for the ICD-10 variable, along with the size and prevalence of each.

Over time the way that death is reported by doctors will change through improvements in knowledge, technology and teaching (Janssen and Kunst, 2004). These changes alter how certain causes are identified and coded, for example before the discovery of HIV/AIDS, deaths under the disease would be coded as something else. These mistakes are not just down to unknown disease, since diagnosis technology may limit what can be detected. Although an issue, these changes cannot be accounted for since they represent gradual changes which would be complicated to identify and alter. This misclassification of some deaths also extends to cases like that of Dr Harold Shipman. Deaths caused by the infamous doctor are still in the data set, classified as caused by ‘old age’ (or natural causes), and have not been changed to their actual cause (i.e. homicide). Although a famous example, it also highlights potential errors both known and unknown.

Chapter	Blocks	Title	Number of causes with death(s)	Number of deaths
I	A00-B99	Certain infectious and parasitic diseases	249	31400
II	C00-D48	Neoplasms	555	621597
III	D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism	102	4416
IV	E00-E90	Endocrine, nutritional and metabolic diseases	182	32680
V	F00-F99	Mental and behavioural disorders	102	80320
VI	G00-G99	Diseases of the nervous system	181	72451
VII	H00-H59	Diseases of the eye and adnexa	13	32
VIII	H60-H95	Diseases of the ear and mastoid process	10	92
IX	I00-I99	Diseases of the circulatory system	282	744223
X	J00-J99	Diseases of the respiratory system	185	305869
XI	K00-K93	Diseases of the digestive system	267	420717
XII	L00-L99	Diseases of the skin and subcutaneous tissue	77	7941
XIII	M00-M99	Diseases of the musculoskeletal system and connective tissue	191	18404
XIV	N00-N99	Diseases of the genitourinary system	173	50679
XV	O00-O99	Pregnancy, childbirth and the puerperium	67	189
XVI	P00-P96	Certain conditions originating in the perinatal period	113	1362
XVII	Q00-Q99	Congenital malformations, deformations and chromosomal abnormalities	247	5393
XVIII	R00-R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	66	43528
XIX	S00-T98	Injury, poisoning and certain other consequences of external causes	0	0
XX	V01-Y98	External causes of morbidity and mortality	1201	71620
XXI	Z00-Z99	Factors influencing health status and contact with health services	0	0
XXII	U00-U99	Codes for special purposes	1	2088

Table 3.2: ICD-10 Chapters and additional death information (2006-2009).

Coverage for the cause of death variable was good, with the majority of cases containing data. However when inspecting those that were missing, they were mostly for individuals aged zero. For deaths at age zero, only 30 per cent contained ICD-10 data. This is because neonatal and stillbirth deaths use a different death certificate (Rooney et al., 2002). This contains two lines for conditions occurring in the

foetus/infant, two for those related to the mother and another for other factors contributing to the death (Rooney and Smith, 2000). This system does not allow for a sequence of events like the death certificate (i.e. MCCD) and therefore no single code can be input into the data set. Instead a modified ICD-10 coding system is used based upon a classification of the infant mortality types (usually by time period of occurrence; Alberman et al., 1994), which is not applicable to this study. There was only 639 cases missing an ICD-10 code over the age of zero and these were found only for the years 2001 to 2005.

3.3.2 Social class

With the decision to focus on solely the period covered by the ICD-10, only two social class variables are relevant here. Social class is recorded from the deceased's occupation, which is provided by a relative when the death is registered. There is however no check on this information required and hence is dependent upon the honesty of the individual registering the death.

'Standard Occupation Code 2000' provides data on the exact job the deceased was employed in (through 496 unique codes), of which social class can be derived from. Codes are grouped similarly to the Registrar General's Social Class groupings, based upon an updated version compiled after the results of the 1991 Census. However coverage is particularly poor, due to the measure being replaced at the beginning of the period by the National Statistics Socio-economic Classification (NS-SeC).

The NS-SeC gave an updated system for deriving social class, drawing not only on occupation but also employment conditions, relationship to other workers and size of the company (ONS, 2013a). The social groupings can be seen in Table 3.3, as well as their split into a linear three group system as advised by the ONS (otherwise comparisons between groups three and four become difficult as there is not a definite hierarchy; *ibid*). The group 'Not Classified' contains full time students, as well as those unclassifiable. Deaths where no NS-SeC value is included are recorded as "0".

There are two variables for NS-SeC in the DOEW. 'NS-SeC 1' records the social class of the deceased and 'NS-SeC 2' gives the value for the individual's spouse. These variables differ for those aged below 16, where 'NS-SeC 1' gives the class of the mother of the deceased and 'NS-SeC 2' the class for the father. Unlike 'Standard

Occupation Code 2000', there is greater coverage in the dataset. However there is still a majority of the data which contains missing data. A closer inspection shows that this is age dependent, with only one case missing social class aged below 75. This is because social class data is not recorded if the individual has not worked in the last ten years. Whilst the ability to accurately analyse social class variations in premature mortality is good and would be useful in this study, the possibilities for understanding patterns for the majority of deaths is limited (65.3 per cent of deaths in this period occurred at 75 and over, of which 0.19 percent of these contained data on social class).

Social Class	Name	Occupation Categories	Three Class Breakdown
1	Higher managerial and professional occupations	L1, L2, L3	High
2	Lower managerial and professional occupations	L4, L5, L6	
3	Intermediate Occupations	L7	Intermediate
4	Small employers and own account workers	L8, L9	
5	Lower supervisory and technical occupations	L10, L11	Low
6	Semi-routine occupations	L12	
7	Routine occupations	L13	
8	Never worked and long-term unemployed	L14	Not used
9	Not Classified	L15, L16, L17	

Table 3.3: NS-SeC groupings (after ONS, 2013a).

There are some issues with the use of social class by occupation in this data. An approach used in prior studies has been to use father's social class and whilst this may improve the consistency of the data, it is an unfair reflection when the mother has a higher job. To avoid any problems when used, the highest social class is taken where both are provided. The social class provided may not be a complete reflection of actual social class, as it ignores temporal changes throughout the life course. For example, a person who became unemployed may have a lower social class job before death. Given that the data set covers the beginning of the economic recession, this could be problematic, although little can be done to solve this. Finally, some detail is lost through not being able to compare those not classified as part of the various measures (for example, students). Although this represents a fairly large group of people, there is little basis for incorporating effects into the analysis since they represent a diverse range of people.

3.3.3 Data considerations and issues

Before 2006, the DOEW statistics were published in the data year. This compiles all deaths registered in a year, with the 31st of December being the cut off point. However this records all deaths that occur in a year, since late registration would be excluded. This design choice was to allow for quicker compilation of statistics. Since 2006, this has been changed to contain all deaths registered as happening in a particular year. Nevertheless due to the small numbers involved, this will only have a minimal effect overall (ONS, 2010a). These years in question should not affect the classification itself (see Section 3.7.1)

Not everyone gets the option to die at their home and therefore there are some points which need raising about this location data. A lot of people die whilst in hospital, however the death is recorded with the address of their last residence (since all this would tell us is that there are disease hotspots where hospitals are!). This is also the same for those people who die whilst at her Majesty's pleasure. Those who die overseas, but usually reside in Britain, are not included in the DEOW (Shaw et al., 2008). Students who are living away from home are registered back at their home address. For nursing homes, the residency address is only taken if the deceased had been living there for at least six months (ibid). Anything less, the choice of address is left to whoever registers the death. The six months tenancy rule has always applied to those living in psychiatric hospitals and hospices.

3.4 What geography?

Since the raw data is recorded at the individual level, this source is of the highest quality. Nonetheless analysing every death at this level would be hard to explain. Therefore aggregation will be necessary to help make sense of the patterns and processes occurring. It is also required due to confidentiality issues, to avoid being able to identify specific deaths. Furthermore aggregating to a geographical level will incorporate the analysis of place-based effects by examining similarities and differences between areas, as well as the spatial patterning of mortality profiles. Despite this, there are many different geographical scales that exist within Britain, each with their specific benefits and disadvantages.

Zone	Number	Mean Population (2001)
Output Areas	175,434	297
Lower Super Output Areas	34,378	1500
Middle Super Output Areas	7194	7200
Statistical wards	8800	5914
Census Area Statistics wards	8850	5880

Table 3.4: Information about possible geographic zones.

Table 3.4 presents some of the possible geographical scales which could be applied in this study to capture small scale variations in mortality. These are all taken from census statistical geographies, to allow for subsequent demographic information compiled by various sources to be derived and used with the mortality data. To decide which level would be best for the classification, a balance is needed between maximising both the detail (i.e. size) and the number of deaths captured. The number of deaths needs to be big enough to produce a meaningful classification, especially with the large range of variables used.

The smallest geographies would, ideally, be the best option (i.e. OAs or LSOAs). This would preserve greater data quality from the raw data, presenting the intricacies between small areas in their mortality profiles. This is especially the case given that they were both designed to be socially homogenous (Martin, 2002). However, with such few people living in these areas, the amount of people dying over the time period used would be low. For example, using LSOAs between 2006 and 2009 would have given an average of 58 deaths per area (a range of 1 to 427), smaller than the number of variables selected to be used in the classification. Hence the time period required to pool together enough deaths per area would have to be larger, which may miss out changing trends and patterns. Furthermore at the lower levels, issues of confidentiality arise due to the problem above, as individual deaths may become easily recognised in such small areas. Therefore a higher level is required.

Although the ward level would be useful with regards data dissemination, as people have a better understanding of electoral boundaries, they would be less useful for creating accurate clusters. MSOAs would improve comparisons between areas since they are more similar in terms of population size (ONS, 2011). This variation creates issues with regards contrasting areas, since you are not always comparing like for like. Wards are also more unstable, frequently changing unlike the other geographies which were designed to alter less often, making them more useful for future research or comparisons to other data sets. Furthermore there is a larger set of data (i.e. for

explanatory variables) released at the MSOA level than compared to the ward level for each country. Therefore MSOAs will be used for England and Wales. The average number of deaths in each MSOA between 2006 and 2009 was 275.09 (with a median of 262, suggesting a slight positively skewed data set). The number of deaths ranged from 46 to 1096, with a standard deviation of 105.04.

3.5 Choice of method

There is no one set methodology for producing an area classification. Advances in modern computing have led to the explosion of possible methods, as newer computers allowed more information to be processed (Everitt et al., 2001). Even within established methods, there is variation within them (cluster analysis is itself more a 'family' of methods) and it is not restricted by discipline either (Gordon, 1999). Each potential method has both benefits and problems involved, and thus it is hard to recommend a single method. However most algorithms are relatively similar to each other, only changing small parts of the process to refine the methodology. Therefore it is more useful to concentrate on those frequently used research methods.

It is important to select a robust methodology. The results found in this study should be due to the variations and patterns within the data, not because of the method used. Therefore experimental methods will not be considered. Whilst even the most robust methods will produce different solutions, these are likely to be only slight (for example, differences in cases found at the edges of clusters). The selected method needs to be the technique which can be implemented best within this study across a range of issues, for example outputs, efficiency, producing compact solutions etc (Everitt et al., 2001).

Supervised classificatory methods (for example discriminant analysis) have not been considered since they are not applicable. Supervised methods work through teaching a model how to classify data through rules and examples. However as the underlying structure is not known *a priori*, an area classification cannot be created to adhere to certain rules. Unsupervised methods instead require no training sets since they group together the data without need for human input (Duda and Hart, 1973). These approaches will be more useful since they are data-driven through looking for patterns in the data, unlike supervised which are theory-driven (Halpern et al., 2004). Although

this may lead to some error, unsupervised methods allow the study to find 'natural clusters' in the data which may otherwise by ignored.

It is also important to define what is understood as being a cluster, as this will determine which methodology is chosen. A well-defined cluster within the data would display two important characteristics; cohesion and isolation (Gordon, 1999). Therefore it will be defined as a sequence of points in *k*-dimensional space (where *k* is the number of variables) which lie near to each other and also far from other clusters, separated by empty space (Milligan and Cooper, 1987). The areas in each cluster display similar characteristics (mortality rates) to each other, also being distinctly different to other clusters and their respective areas (in terms of mortality rates). Therefore the selected methodology will need to take this into consideration and, preferably, emphasise these qualities.

The examples used in the following sections will be drawn from studies which use similar data sizes (i.e. within five thousand cases of around seven thousand). The main classificatory approaches will now each be described, along with a general discussion of the benefits and issues involved with them to assess which would be best to use.

3.5.1 Hierarchical clustering

Hierarchical methods (also known as stepwise or top-down methods) are the most common cluster analysis method for producing a classification within research (Everitt et al., 2001). Generally these produce non-overlapping clusters of data, which are part of a pyramidal structure of clusters. There are two main types of hierarchical clustering approaches; agglomerative and divisive. Agglomerative methods begin with *n* clusters (where each individual case is its own cluster) and the two most alike clusters are joined together. This process is repeated through a series of steps until there is just one cluster (containing all clusters/data), creating a hierarchy (Duda and Hart, 1973). Divisive methods are similar but the process of clustering occurs in reverse, starting with one cluster (containing all cases) which is then split into clusters until there are *n* clusters left.

The choice between these two hierarchical methods may have an impact upon the results, even though the process of clustering appears similar for each. Within the literature, agglomerative methods are more commonly used. Divisive methods tend to

be more effective when clustering binary data, but also generally produce more evenly sized clusters (Gordon, 1996). However they also can be more computationally demanding when dealing with larger datasets, although agglomerative methods are not particularly quick themselves (Everitt et al., 2001).

Agglomerative clustering also has different methods of determining which cases/clusters to join together, each with their own strengths and weaknesses (Cormack, 1971; Gordon, 1996; Milligan and Cooper, 1987; Everitt et al., 2001). The most common techniques include:

- Single linkage (or nearest neighbour) method joins together the clusters which have the smallest distance between individual cases (between two clusters) based upon their explanatory variables.
- Complete linkage (or furthest neighbour distance) is the opposite of the previous method, through considering the largest distances between any two cases when comparing clusters (joining together the match which minimises this figure).
- Centroid clustering method is also similar to single linkage however instead of searching for the smallest distance between any cases in clusters, the centroid of a cluster (or mean vector for that cluster) is used to calculate distances linking clusters.
- Group average method uses the mean distance of all cases that make up a cluster with each other cases in another group to calculate distances. Median values can also be calculated instead of the mean for these previous two methods to give equal weighting to each case in a cluster.
- Ward's (1963) method is the most common of all these approaches. It joins together the clusters which minimise the increase in error (using sum of squares) within the cluster that they are added to.

The choice of method is likely to affect the results gained and therefore choosing the right measure is important. Milligan and Cooper (1987) note that throughout the literature, Ward's method appears to perform best, as well as producing more even sized clusters. This was also shown by Blashfield (1976), though it is difficult to generalise these results. Nevertheless it is sensitive to outliers and may give a spherical structure to clusters, when there is not one present (Everitt et al., 2001). The group average method also performs well and is less affected by outliers than the other methods, albeit it is occasionally erratic. Centroid clustering can be affected by large case values within

clusters, although median values could be used instead to give equal weighting to the cases in the cluster. Complete linkage and single linkage both ignore the cluster structure by selecting just the nearest or furthest away cases. Furthermore Milligan and Cooper (1987) note that single linkage method usually performs least well in sensitivity studies, producing the most amount of errors and creating unbalanced clusters.

The hierarchical nature of these methods allows the process to be viewed through a dendogram. Described by Everitt et al. (2001, p56) as "an evolutionary tree", the dendogram is a diagram which shows a graphical visualisation of the clustering process (Ketchen and Shook, 1996). This shows the order of the joining process (which cases were joined together and when), and the subsequent distance between each join (i.e. separation). This is very useful when deciding upon the optimal number of clusters to use, as it allows the researcher to make a judgement on which solution would be best, through seeing which number of clusters have the greatest distance between them (i.e. the most different) (Duda and Hart, 1973; Gordon, 1999). Nonetheless this can be quite subjective, for example Day et al. (2008) proposed a 12 cluster solution to their data set. However the dendogram shows that possibly an eight, four or three cluster solution may have been better since these solutions each had greater distances between clusters, with some clusters of 12 cluster solution being relatively similar.

The advantage of using this methodology is that it splits the data into a hierarchy. Through presenting a range of solutions, you can gain a better understanding of the structure of the data (especially through the dendogram). This means that the method is ideal for showing the optimal step or number of clusters that are present in the data (Gordon, 1996). However a focus on finding the ideal scale means less attention is given to the quality of the clusters (Everitt et al., 2001). This makes the chosen solution more restrictive unlike other methods which may choose to optimise a solution rather than produce the whole range of solutions. This is because once two objects or clusters have been joined together, this cannot be undone. A case could be assigned to a cluster, only to find it closer to a different cluster later in the process and unable to be reassigned (Duda and Hart, 1973).

Other important issues include that the method (or at least many of the algorithms) are sensitive to random noise through chaining. Where outliers or random points lie in between two distinct clusters, this may result in the clusters being joined together into one larger cluster (Duda and Hart, 1973; Everitt et al., 2001). The outlier point becomes a 'bridge' to join the two clusters, through reducing the distance between either. Both

distinct clusters then become lost, with the result unrepresentative of the data it symbolises. The results produced can vary when presented with non-spherical data sets (Duda and Hart, 1973). Since the methods produce a multiple solution classification, this procedure takes a long time to compute with more than a thousand cases since it compiles the range of solutions between 1 and *n* (Harris et al., 2005). This is an importance issue within this study, however Jensen et al. (2001) produced a hierarchical classification with a similar amount of cases, though the amount of time taken to run the classification is not noted. Finally there is an issue regarding how the method chooses between two (or more) joins which are just as close, especially since the decision can have a large impact on the results (Gordon, 1996).

Hierarchical methods are quite popular within research and hence have been applied to many different topics. However the majority of these applications only use small samples of individual level data, especially in the health literature (*c.f.* Fitzpatrick et al., 2011; Franceschini et al., 2012; Raleigh et al., 2012). Parfitt et al. (2001) used Ward's method hierarchical cluster analysis to produce a segmentation of waste management strategies for Local Authorities (376 large areas) in England and Wales. An eight cluster solution was created, allowing for an evaluation of different combinations of policy strategies. There were significant observations in recycling performance, with wheelie bin implementation producing higher recycling rates and lower waste. As such, policy recommendations and areas to tackle could be identified.

Whilst there is the potential for building an area classification with this method, the sample size issue has resulted in researchers often opting for other methods. However there are a few examples using a similar sample size. For example Jensen et al. (2001) classified the biophysical features (for example climate, vegetation, elevation) of the American part of the Columbia River Basin to assess ecosystem management. This saw the classification of 7462 areas into 13 clusters, which required different strategies for environmental conservation.

3.5.2 Partitional clustering

> "In essence, you are groping your way almost blind through an unimaginable complex hyper-dimensional space towards (you hope) a set of usable results."
>
> (Openshaw, 1983, p257)

Partitional methods (also known as optimisation methods) are also quite common within research. These are different to hierarchical since they are iterative processes which seek to refine a set number of clusters (Harris et al., 2005). Like hierarchical, this is also a group of methods, with slightly varying algorithms including; k-means/medians, simulated annealing, hill climbing/descending.

The process generally begins with the specification of a set number of clusters for the method to group the data into. The data is then split into this number of clusters, with membership assigned randomly. Cases are then individually relocated (through random sampling, although systematic could also be used) to the cluster which it is closest to (compared to the centroids mean). This move is checked against some measures that assess whether there has been an improvement in the model (usually with regards to internal homogeneity). Cases are reassigned to the cluster which experiences the greatest improvement, thus optimising the initial 'solution' suggested. The reassignments of cases are then repeated until no more cases can be moved that result in an improvement in the model, suggesting a stable classification (Everitt et al., 2001). This approach creates non-overlapping clusters that are not hierarchical, since only one partition of the data set is created (Milligan and Cooper, 1987).

The main benefit of using this method is that since the process only produces one partition of the data set (or level of the data structure), this saves computing time (Harris et al., 2005). This does not just mean that the classification is much quicker, but also the process can also use additional resources to focus on optimising the specific solution. This allows it to produce the ideal cluster solution for the given level, through rearranging cases to improve the internal cohesion of clusters and external differences between clusters. It also makes it more stable when handling larger data sets. It is not so restrictive when it comes to the shape of the clusters, such as non-spherical data points (Everitt et al., 2001). Harris et al. (2005) also notes that it retains a higher

proportion of the variance from the variables. The method also tends to produce more evenly sized clusters.

This most important problem with this approach regards the selection of the number of clusters. Since the method only seeks to optimise a solution, a problem is that the number of clusters must be specified prior to the analysis. This number may not be known, meaning that the chosen solution may not be correct and could have a large impact on the results (Bezdek and Pal, 1998; Everitt et al., 2001). Some measures and approaches exist to estimate what the value to use is but they are relatively ad hoc in their creation. Whilst classificatory approaches as a whole suffer from the problem of deciding what the true number of clusters is, other methods such as hierarchical get around this issue through producing a solution with all possible levels. The data used can only be continuous (interval or ratio) else other methods should be used (this is not an issue with this study however). Whilst producing evenly sized clusters can be a benefit, it could also be a limitation where the true structure is not (Gordon, 1999).

Similarly to hierarchical methods, all of the applications of the method in the health literature have been using individual level data (for example Halpern et al., 2004; Rozumalski and Schwartz, 2009; although both have large sample sizes). There has been greater usage of the method for larger data sets when creating an area classification. Gould et al. (2012) examined food deserts in Gatineau, Canada. Using a variety of measures of accessibility (in terms of distance, proximity to a wide selection of fruit and vegetables etc), they found six clusters. Although three clusters were of poor accessibility, the classification showed that as the level of deprivation for an area increased, so did the level of accessibility.

3.5.3 Neural networks

Artificial neural network studies were originally conducted to study how the brain functioned, through modelling a highly connected series of neurons (Everitt et al., 2001). Inspired by this, the methodology became increasingly applied to other functions that involved a large, interconnected network of phenomena, especially those involving a flow of data (for example telecommunications). There are three parts to a neural network; the inputs (i.e. the data), the hidden layers (i.e. the parameters) and the output (i.e. clusters). The hidden layers apply rules and variables to test the inputs and this

value is then assigned a cluster depending upon reaching a threshold (Ahmed, 2005). The process is iterative, which allows the model to adjust weightings of data and learn the best pattern recognition procedure, minimising errors. This process can be either supervised (i.e. learns through training data) or unsupervised (similar to cluster analysis methods).

Waller et al. (1998) found that artificial neural networks performed just as well a k-means or hierarchical classification (Ward's, complete and average linkage models were all ran). Nevertheless there are many problems with this approach and Waller et al. note that there is conflicting evidence over its effectiveness compared to other methods throughout the literature. The biggest issue is that it poses a black box issue. With models containing a hidden layer (which also changes as the model learns), it is difficult to understand how the classification was produced (Ahmed, 2005). The more hidden layers used, the more complicated the process becomes (though little evidence that more than one has many benefits). Most uses of the method are supervised, requiring prior knowledge about the cluster structure which would not be useful here (though not restricted to). Neural networks can also take a long time to compute, due to the large amount of calculations and the learning process involved (ibid). This may be why many examples involve small data sets.

Most examples of neural networks used for classificatory purposes involve small amounts of data, although they are capable of handling larger sets. Openshaw et al. (1994) produced an area classification of 1991 Enumeration Districts for Great Britain (GB Profiles '91). Despite creating a six cluster solution, the method proved computationally intensive. Otherwise there are few area classifications created using the method, other than the classification of remote sensing data (Knorn et al., 2009; Melchiorre et al., 2008). The lack of examples would suggest that it is less useful for the creation of an area classification. Many of the uses of neural networks require supervised learning, for the computer to be able to understand and learn from the processes.

3.5.4 Fuzzy classification

Fuzzy classification refers to an increasingly popular methodological type, rather than a specific technique and is based upon fuzzy logic. Mathematics is traditionally based on

Boolean logic; either something is true or it is false. Classifications function with this idea as cases can only belong to one cluster (i.e. in or out). Fuzzy logic extends this mathematical principle (and at the same time dismisses many other theories which depend upon Boolean logic!) through introducing the idea of partial truths. Within a classification context, cases are no longer assigned to a single cluster, since this is only true to a certain extent (Everitt et al., 2001; Voas and Williamson, 2001). Rather cases are assigned a measure which represents the strength of membership to each cluster (based upon a distance or similarity index. With a smooth transition between clusters, there are no resulting 'knife edge' issues, meaning the classification is less affected by 'noise' (Nauck and Kruse, 1999).

A fuzzy classification can be more effective than a traditional classification. It could be argued that a classification produces an 'oversimplification' of the data. Cluster membership is unlikely to be black or white. This is especially the case where case membership falls on a 'knife edge'. Values towards the edge of a cluster may fall more into a 'grey' area rather than be distinctly one value. This will be more problematic where the cluster structure is less distinct. Therefore it may be more realistic to not restrict areas to just one class, rather areas are examined as proportions of each cluster. This helps our understanding as if there was another cluster that equally explains what is happening for certain areas, then this would be otherwise hidden away (Everitt et al., 2001). Although providing greater detail in cluster membership for areas can add to our understanding how what is happening, it goes against the aims of producing a classification. Rather than simplifying the data set, to summarise the main patterns and relationships between areas, a fuzzy classification would bring added complexity (Gordon, 1999). Cluster membership may not always be clear through including competing factors.

Fuzzy classifications are not as popular as the previous methods. The method has been important within health and medical research. Nauck and Kruse (1999) notes that due to the smooth transition between clusters, cases which lie close to other clusters can be assessed to avoid issues of misclassification, making it useful for diagnosis. However its application in producing area classifications or using geographical data is severely limited with no real examples past remote sensing studies (Burrough et al., 2001; Lorette et al., 2000). Vickers (2006) created an ad-hoc fuzzy classification of the ONS Output Area Classification however this was methodologically limited. The

development and application of this methodology is, however, still very much a continuing part of research and as such, many examples are from supervised approaches.

3.5.5 Which method?

An issue with producing a classification is that no one method can be recommended (how do you even attempt to measure which is best?), since each method is different and they all have their own applications. Yet this is not necessarily a bad thing as it provides a wealth of choice to hand pick the most effective method for the job. All methods could achieve the aims of this thesis. However it is important to select the method which can be implemented both most effectively and efficiently into building an area classification here. Table 3.5 shows those factors of which the methods were judged on.

	Hierarchical	Partitional	Neural Networks	Fuzzy
Handle large data	Becomes unstable and computationally intensive as size increases	Yes	Computer intensive	Yes
Applied to geographic data successfully	Yes	Yes	Yes	Yes (but difficult to visualise)
Speed	Slow	Fast	Very slow	Medium
Robustness	Yes	Yes	Conflicting evidence	Methodology still in development
Iterative development of clusters	No	Yes	Yes	Mixed
Specification of number of clusters prior	No	Yes	Mixed	Mixed

Table 3.5: Factors important to building an area classification by method.

Hierarchical methods were mainly rejected based on their ability to handle large data sets. The hierarchical methodology becomes unstable with larger data sets both in time and potential (or probability) for errors through incorrect joins (Milligan, 1996). Whilst in practice, the process was not too long (just over five minutes), the process became

unstable when handling this size of data. This led to outputs taking longer to be compiled, with some not even materializing (for example the dendogram). This approach also does not create the optimal solution, as clusters are not refined iteratively. This leads to erroneous cluster membership and less accurate clusters (Everitt et al., 2001).

There is also the time problem for artificial neural networks and though it was not tested, it is an important issue in the literature (Ahmed, 2005). Since the process allows the computer to learn about cluster membership (through a series of rules and processes), this iterative step is both time and computer intensive when compared to the other methods (Openshaw et al., 1994). There is also a black box issue, meaning it is difficult to interpret the findings as we are unsure about what is happening within the data or what will be found by the classification. Conceptually, it is designed to analyse flows of data, making it less relevant in this application.

A fuzzy classification approach in its entirety was not chosen since as this is just an initial investigation into the data set, simplifying the unknown complexities makes more sense than providing a more complicated answer at first (Openshaw et al., 1994). With areas being able to belong to more than one cluster, geographical visualisation can be problematic, limiting the interpretation of the classification. The methodology is also still being developed and therefore is not always completely robust.

The partitional methodology shall be used to classify mortality in this study. With such a large data set used within this study, it was important to choose the method that was most efficient to save processing time. With only one partition carried out, there is a significant saving in computational power (Harris et al., 2005) which is very important when dealing with 7194 objects, across a large range of variables. Although the number of clusters must be specified prior to running the method, this can be investigated through various test statistics (see Section 3.8.4) to assess which solution is the best. As such, the method can produce the most optimal solution based upon this.

3.6 More detail of how the partitional methodology works

Section 3.5.2 provided a brief description of how a partitional clustering method works. Here this explanation will be expanded on to provide a little more information as to how the methodology operates. Some detail will be provided, however there are plenty of

references which give a better explanation with greater coverage (for example Duda and Hart, 1973; Gordon, 1999; Everitt et al., 2001).

There are two main parts to the partitional process, which affect how the methodology works. One of these is how the clustering criteria are selected. This is how the method assesses the quality of the clusters it has created within the data, during the iterative stage of the process (Everitt et al., 2001). The total variance (for a given solution), *T*, found here as the total sum of squares (i.e. the error) can be found using the following equation:

$$T = (\Sigma W_i) + (\Sigma B_i) \quad (1)$$

T can be split into *W* (the within cluster sum of squares) and *B* (the between cluster sum of squares). It is these variables which represent the concepts of homogeneity (*W*) and separation (B) found within the model. These are summed individually across each variable (*i*) included in the model, to assess similarity and isolation of data points through multivariate dimensions (Duda and Hart, 1973).

The within cluster sum of squares measures how close the objects in each cluster lie to each other (i.e. homogeneity). Therefore an optimisation methodology looks to mainly minimise the value for *W* (Duda and Hart, 1973). Everitt et al. (2001) notes three common measures of homogeneity. One index calculates the sum of all squared dissimilarities between two objects for a cluster. Another looks for the maximum value between two objects for a cluster (effectively finding the diameter of the cluster). The third ('the star index') finds the minimum value between all values to a single value in the same cluster.

The between cluster sum of squares measures the difference in distances between each cluster (i.e. separation). Isolated clusters result in a distinct classification and hence the methodology looks to maximise *B*. Everitt et al. (2001) also gives two examples of separation measures. These either find the sum of the (squared) dissimilarities between an object found in a cluster and another outside of it, or the minimum value for this is used instead. These concepts seek to find a clear cluster structure within the data through these processes. Maximising *B* is less commonly used in research than minimising *W* as the clustering criteria.

Whilst this equation is only applicable for analyses where one variable is used, it is much easier to think about the process like this. How the equation exists is not clear

when the number of variables begins to exceed one, but it still continues to employ the same mechanics. It highlights how the method assesses the partitions created through minimising and/or maximising numerical criteria (Everitt et al., 2001). A multivariate perspective only differs through looking at how the data points are scattered in k-dimensional spaces (Duda and Hart, 1973). Clusters are formed based upon this similarity across many different variables.

The other major part to the process is the choice of algorithm for running the methodology. This creates the (specified number of) clusters through splitting the data up into the best solution possible, through optimising the clustering criteria. There are thousands of algorithms that exist, each performing different techniques to clustering the data. For the analysis in this study, the main types will only be considered. This is because a robust methodology will mean that the findings are more likely to reflect variations in the data, rather than issues in the clustering procedure.

It is inefficient to compare the cluster criteria for all possible cluster combinations. For example Duda and Hart (1973) note that that a data set with 100 objects with five clusters would require roughly 6.6×10^{67} calculations. Even with advances in computing power, this is still going to be computationally heavy especially with the size of data used in this study. Since global searches of cluster solutions are impracticable, algorithms have been developed to perform local searches instead. These usually work by making an initial partition of the data into the specified number of clusters. Each individual object from a cluster is then moved into each other cluster and then the change in the solution is assessed against the clustering criteria. The move which results in the best value based upon the criteria becomes permanent (though not if no gain). This process is repeated until no more moves can further improve the clustering criterion (Gordon, 1999).

The most common algorithm used in research is the k-means method. Objects are relocated to the cluster whose mean (i.e. cluster centroid) it is closest to. This move is then assessed against the clustering criteria to see if there was an improvement in the model. If so, the move is kept and the cluster mean is recalculated. The equation it uses is:

$$J = \sum_{j=1}^{k} \sum_{n \epsilon Sj} (x_n - \mu_j)^2 \qquad (2)$$

Here k is the number of clusters being partitioned, n is the total number of data points, S_j is one of the partitions and the data points found within, x_n is a data point from all cases

and μ_j is the mean centroid value for S_j (Duda and Hart, 1973). The equation splits the data into clusters with the aim of minimising the clustering criteria (J is the total squared error). This algorithm will be used due to its widespread use, helping ease its understanding by other users (Jain, 2010). As a result of its popularity, it can be found in most statistical packages, making the process of creating the classification more efficient (Everitt et al., 2001). Milligan and Cooper (1987) also argue that it is the best partitioning algorithm available.

3.6.1 An example of the k-means methodology

To illustrate how the k-means methodology works in classifying data, a simplified example is presented. Fewer data points are used to improve the visualisation of the process, showing the approach clearly. Data for Governmental Office Region's (GORs) in England and Wales was collected. Two variables were used and these were the percentage of people who said that their health was 'good' at the 2001 Census and the life expectancy for the region for the period 2000 to 2002.

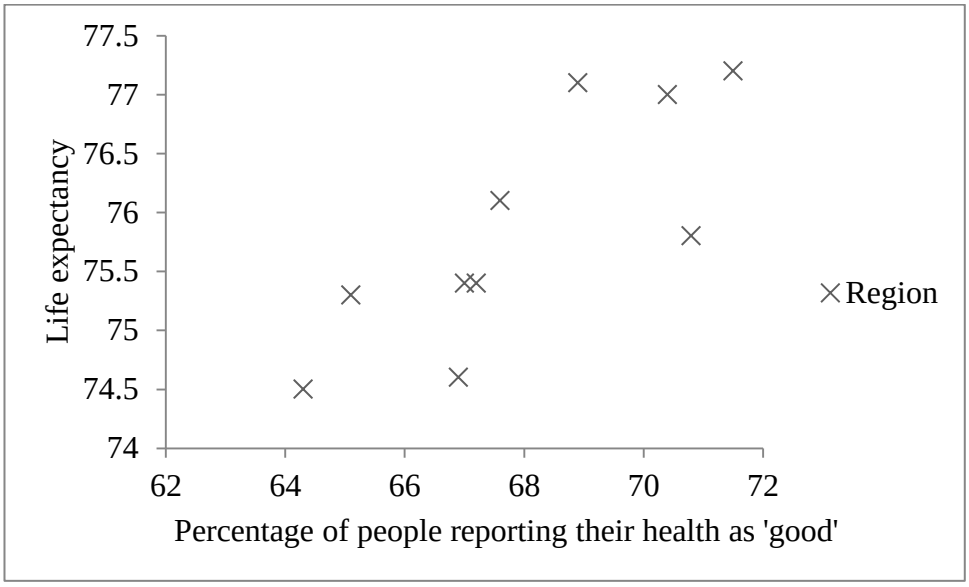


Figure 3.2: The relationship between people reporting their health as good and life expectancy in England and Wales.

Figure 3.2 presents a scatter plot of the variables, which shows a strong positive relationship between the two (a Pearson's correlation gives the value 0.8, with p=0.005). The pattern makes sense, since if the self-reported variable is taken as a proxy for health then those areas with greater health needs are going to be less likely to have people

living longer. To explore if there are clusters within the data, a k-means cluster analysis was run on the data for a two cluster solution (to keep things simple).

First the algorithm selects two data points as 'seed points'. These represent an initial partition into the data, the centres of the first clusters. From this, data points are assigned to their nearest cluster centre (i.e. classifying the data). The initial starting points for determining the clusters were selected as the North East and the South East. These have been highlighted on Figure 3.3 as triangles (the seed points remain separate from the data point).

Figure 3.3 also shows the allocation of data points into these first two clusters. The interesting case is the East Midlands value, which lies far away from both initial centres but slightly nearer to cluster 2. Cluster 1 contains Wales, North East, North West, Yorkshire and Humberside and West Midlands. Cluster 2 contains the regions South East, South West, London, East and East Midlands. The clusters are evenly sized. Based upon these clusters, the cluster centres are recalculated.

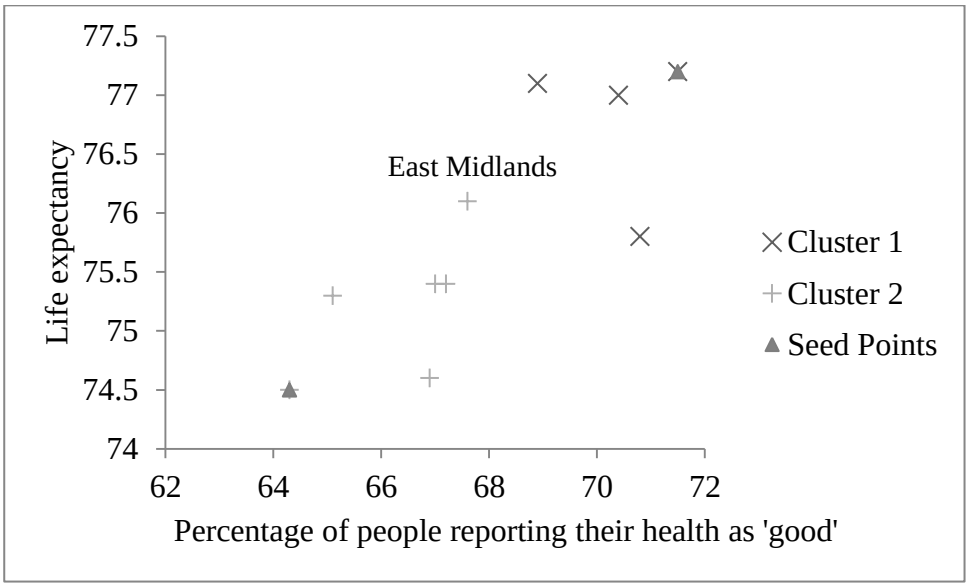


Figure 3.3: The initial classification of the data.

The next iteration of the model seeks to examine the impact of moving each case into the other cluster and looking at the impact on the average distances of cases from the cluster centres. Where there is an improvement in the model, the move is made permanent. In the model, the only change which improves the model is moving the East Midlands into cluster 1. The region is closer to Northern characteristics in terms of health, than compared to the South. The cluster centres are then recalculated. No more moves can further improve the model, therefore the iterations stop and the final solution

can be seen in Figure 3.4. Whilst this example is simpler than trying to think about 7194 cases across multiple dimensions, it visually shows how the process works.

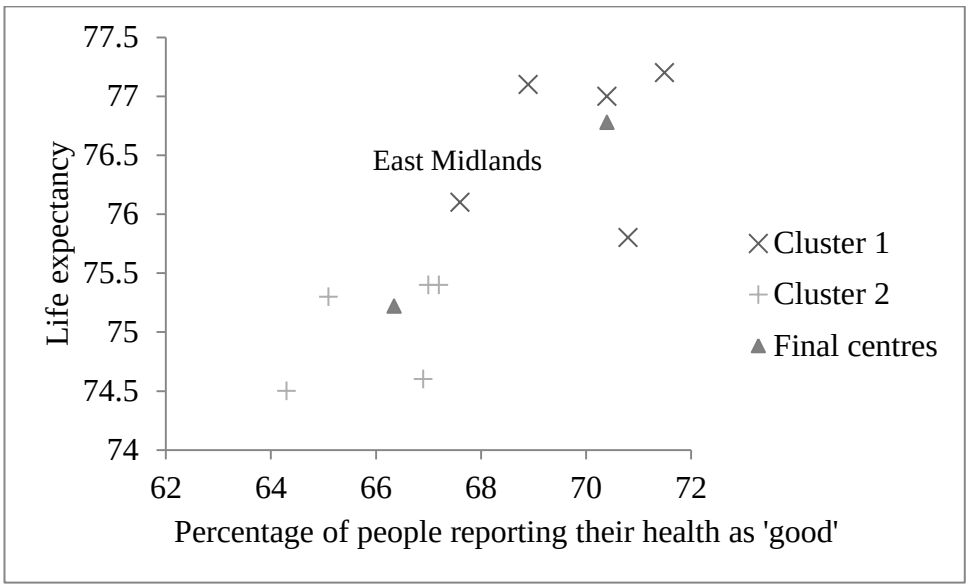


Figure 3.4: The final classification.

3.7 Creating the inputs of the classification

With both the data and geographical scale of analysis selected, the variables to be used in creating the classification can now be built.

3.7.1 Time periods

With MSOAs chosen as the geographical unit to use for the analysis, these are fairly small. Therefore using just the latest year of data will not suffice in providing enough deaths to produce stable and useful statistics, especially for rarer causes. Instead the individual years of the DOEW will need to be pooled together to avoid any small number issues, as well as reducing the effect of fluctuations in a year which would otherwise falsely represent a pattern. However too many years combined will make it difficult to accurately present a picture of what is happening, as temporal trends become smoothed. It was chosen that the latest four years (2006-2009) would be used for constructing the classification, since this would achieve enough deaths whilst not being too wide.

3.7.2 Variable selection

The ICD-10 forms a hierarchical classification of causes of mortality. Diseases are grouped initially into 22 Chapters (see Table 3.2) and within each ICD-Chapter codes are further grouped into types of causes. At its lowest level, which gives the specific type or site of a particular cause, there are over 14,000 possible codes. This is not practical for any useful and interpretable form of analysis. Therefore a useful form of data reduction is required, which maintains the differences and detail offered in the data.

The number of variables chosen is important since it needs to be large to truly capture the variation in mortality between places, especially to gain a meaningful separation of areas (Cohen et al., 1977). The chosen method may fail to capture the true underlying structure of the data if only a subset is analysed (Everitt et al., 2001). However more does not always result in better. A larger number of variables can hinder interpretation. Selection of irrelevant variables can mask the true underlying structure through increasing the effect of random noise which may impose a false cluster structure (Cheng and Milligan, 1996).

It was decided that all deaths should be aimed to be included. As no other study has carried such an investigation out at such a low geography and little is known about how a large range of mortality variables truly interact with each other, there is less theoretical basis for just focusing on a few (and hence limiting our ability to describe areas; Voas and Williamson, 2001). Whilst principle components analysis may be an option for reducing a large set of variables down (Gordon, 1999), this approach used in this situation may be problematic through making interpretation more difficult. It can also distort clusters, when reducing a large number of variables (Milligan, 1996). The right balance between the two needs to be found.

Variable selection was chosen to be grounded by only incorporating causes which accounted for at least 0.5 per cent of the total of deaths throughout the study period (between 2006 and 2009 there were 1,993,407 deaths). This choice was due to statistical reasons, since it gives a figure greater than the number of areas. Therefore an even distribution would at least provide more than one death in each area, in line with recommendations from the literature (Everitt et al., 2001; Gordon, 1999; Milligan and Cooper, 1987). By implementing this cut-off point, the most prevalent diseases are selected. Since a classification of mortality has not been done before, this analysis has

decided not to initially focus on specific causes. This allows for a comprehensive analysis, exploring how the main causes of mortality interact and cluster with each other.

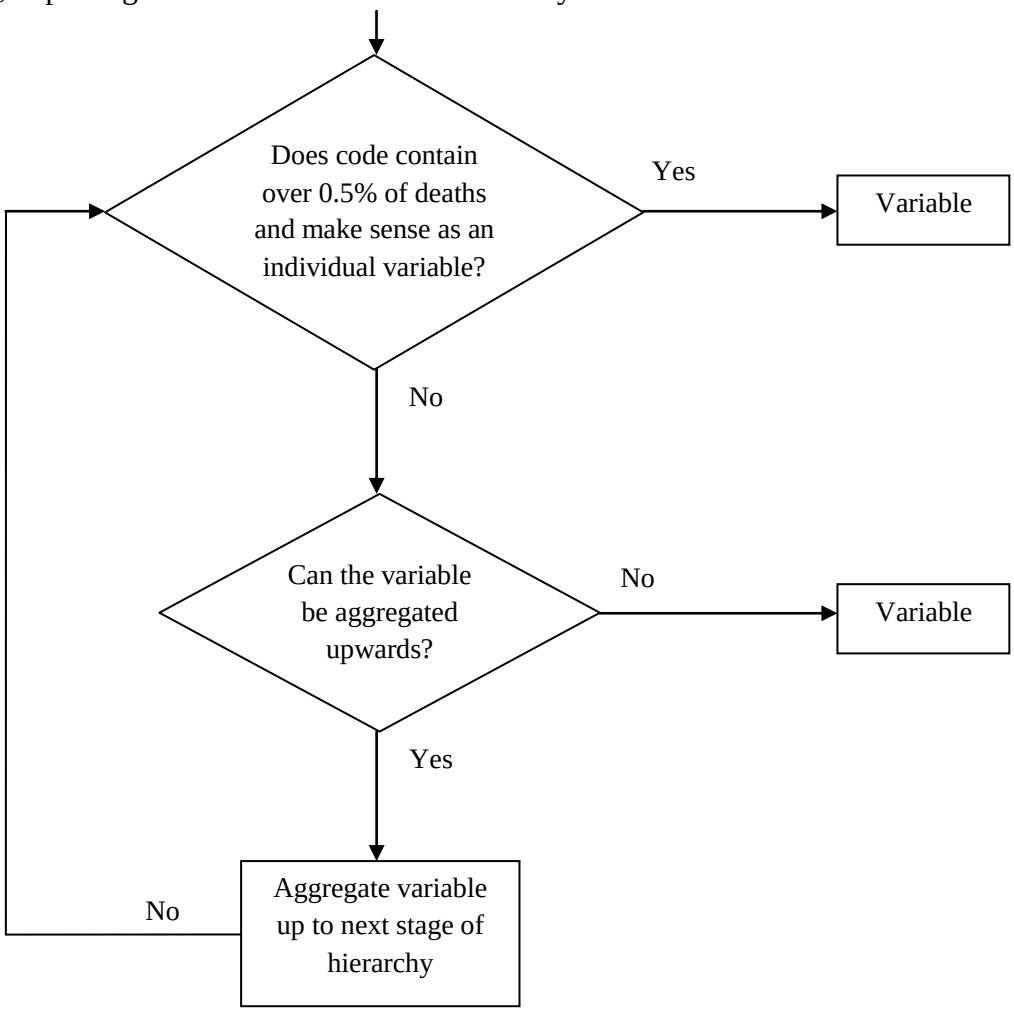


Figure 3.5: Conceptual model for variable selection.

The ICD-10 codes were first aggregating up one level (i.e. A001 becomes A00) since this last digit refers to the specific type or site of a cause and therefore is unnecessary for this study. Nevertheless this resulted in 1136 possible codes and a further means for splitting the variables up was needed. Figure 3.5 shows a flow chart which presents the decision making process for variable selection (although there was some flexibility in its design; for example where causes accounted for 0.49% of all deaths). Variables are first checked to see if they are prevalent enough and whether they should be kept as they are. For example, some codes were for the "other..." part of a cause of death (i.e. usually where there has been insignificant information above the case) and hence were not included since their names are somewhat unhelpful. This first step resulted in 25 variables, accounting for 49.6 per cent of all deaths.

The remaining variables were aggregated to their next level. For example, 'Hansen's disease' (leprosy) was aggregated with the other relevant codes into its collective group

'Other Bacterial diseases'. The prevalence of these groupings was assessed and each checked to see if was useful and interpretable in the analysis. If so, this became a variable itself. Else the process repeated itself, until no further aggregations within a chapter could occur. Even if these did not meet the previous requirements for a variable, they were kept so that all deaths could be used. They were not aggregated further since an all-encompassing "other causes" section would be irrelevant since each ICD chapter is formed by completely different conditions, with different needs, policy options and causes.

As noted in Section 3.3.1, the ICD-10 data for individuals aged zero is inconsistent, due to a different coding system for cases. With 8979 cases at age zero missing cause of death information, this is a fairly large number of deaths to ignore. Rather all deaths at age zero (including those with ICD-10 codes) were combined into one variable to represent the infant mortality of an area. This provides a more interpretable variable with better policy relevance and applications. It is also fairly prevalent, meeting the required specification of 0.5 per cent. The full list of the 67 variables chosen can be seen in Table 3.6.

Although the aim was to use all deaths, not all cases were included in the data set produced. Cases with the code U50 (n=2072) were not included since these refer to cases which have been sent to the coroner pending further investigation (Rooney and Smith, 2000). These codes are from cases involved in inquests and therefore the full registration of the death is held until the verdict is given (Devis and Rooney, 1999). Some of these inquests were still ongoing when the data was released by the ONS (to be updated in subsequent versions). Whilst the ONS suggest that these are most likely homicides, due to their unknown state they will not be included. This meant that there were 1,991,335 deaths in the period and variable percentages were adjusted for this.

No.	Name	ICD10 Codes	Total Cases	Percentage
4	Cancer of the Gullet	C15	26098	1.3%
5	Stomach Cancer	C16	17994	0.9%
6	Colon Cancer	C18	35393	1.8%
7	Rectum Cancer	C20	14328	0.7%
8	Liver Cancer	C22	11507	0.6%
9	Pancreatic Cancer	C25	27390	1.4%
10	Lung Cancer	C34	118885	6.0%
11	Breast Cancer	C50	42824	2.2%
12	Ovarian Cancer	C56	14909	0.8%
13	Prostate Cancer	C61	36811	1.9%
14	Kidney Cancer	C64	12252	0.6%
15	Bladder Cancer	C67	17556	0.9%
16	Cancer of the Brain	C71	12738	0.6%
17	Leukaemia's	C91-95	15588	0.8%
18	Other Lymphatic Cancers	C81-90, 96	26748	1.3%
19	Other Cancers	Rest of C's, D00-48	127490	6.4%

Table 3.6: Cancer related variables.

No.	Name	ICD10 Codes	Total Cases	Percentage
23	Dementias	F00-03	61315	3.1%
24	Other Mental and Behavioural Disorders	F04-99	6362	0.3%
25	Parkinson's Diseases	G20-22	18040	0.9%
26	Alzheimer’s	G30	23015	1.2%
27	Other Diseases of the Nervous System	G00-13, 23-26, 31-99	24893	1.3%

Table 3.7 Mental and nervous system causes of death variables.

No.	Name	ICD10 Codes	Total Cases	Percentage
44	Pneumonia	J12-18	112279	5.6%
45	Chronic Lower Respiratory Diseases	J40-47	103135	5.2%
46	Lung Diseases due to External Agents	J60-70	12031	0.6%
47	Other Diseases of the Respiratory System	J00-11, 20-39, 80-99	48838	2.5%

Table 3.8: Variables included representing respiratory causes of death.

No.	Name	ICD10 Codes	Total Cases	Percentage
30	Hyperintensive Diseases	I10-15	17266	0.9%
31	Acute Myocardial Infarction	I21	120378	6.1%
32	Chronic Ischaemic Heart Disease	I25	188496	9.5%
33	Pulmonary Heart Disease and Diseases of Pulmonary Circulation	I26-28	13852	0.7%
34	Atrial Fibrillation and Flutter	I48	12878	0.7%
35	Heart Failure	I50	33275	1.8%
36	Other Heart Diseases	I00-09, 20, 22-24, 30-47, 49, 51-2	40630	2.0%
37	Intracerebral Haemorrhage	I61	17878	0.9%
38	Cerebral Infarction	I63	17917	0.9%
39	Stroke	I64	88897	4.5%
40	Other Cerebrovascular Diseases	I60, 62, 65-69	59867	3.0%
41	Aortic Aneurysm and Dissection	I71	29921	1.5%
42	Diseases of Veins, Lymphatic Vessels and Lymph Nodes, Not Elsewhere Classified	I80-89	15489	0.8%
43	Other Circulatory Diseases	I70, 72-79, 95-99	13598	0.7%

Table 3.9: Heart related causes included as variables.

No.	Name	ICD10 Codes	Total Cases	Percentage
48	Ulcers	K25-28	11582	0.6%
49	Vascular Disorders of the Intestine	K55	9455	0.5%
50	Other Diseases of Intestines	K56-63	19629	1.0%
51	Alcoholic Liver Disease	K70	18270	0.9%
52	Other Liver Diseases	K71-76	11166	0.6%
53	Diseases of Gallbladder, Binary Tract and Pancreas	K80-86	10666	0.5%
54	Other Diseases of the Digestive System	K00-24, 29-54, 64-67, 90-93	21311	1.1%

Table 3.10: Variables related to the digestive system.

No.	Name	ICD10 Codes	Total Cases	Percentage
1	Infant Mortality	All cases where age = 0	12909	0.7%
2	Septicaemia	A40-41	8929	0.5%
3	Other Infectious and Parasitic Diseases	A00-39, 42-B99	18546	0.9%
20	Diseases of the Blood	D50-89	3926	0.2%
21	Diabetes Mellitus	E10-14	21649	1.1%
22	Other Endocrine, Nutritional and Metabolic Diseases	E00-07, 15-90	6993	0.4%
28	Diseases of the Eye and Adnexa	H00-59	27	0.0%
29	Diseases of the Ear and Mastoid Process	H60-99	77	0.0%
55	Diseases of the Skin and Subcutaneous Tissue	L00-99	7359	0.4%
56	Diseases of the Musculoskeletal System and Connective Tissue	M00-99	16936	0.9%
57	Renal Failure	N17-19	11903	0.6%
58	Other Diseases of the Genitourinary System	N00-16, 20-99	33912	1.7%
59	Causes Related to Pregnancy and Childbirth	O00-99	160	0.0%
60	Conditions Originating in the Perinatal Period	P00-99	63	0.0%
61	Congenital Malformations, Deformation and Chromosomal Abnormalities	Q00-99	3887	0.2%
62	Senility	R54	35176	1.8%
63	Other Symptoms, Signs and Abnormal Findings	R00-53, 55-99	5872	0.3%
64	Falls	W00-19	12801	0.6%
65	Other Accidents	V01-99, W20-X59	30352	1.5%
66	Intentional Self-Harm	X60-84	12361	0.6%
67	Other External Causes	X85-Y98	6632	0.3%

Table 3.11: Other Causes selected.

The inclusion of variables ‘Senility’ and ‘Other Symptoms, Signs and Abnormal Findings’ could also be questioned. These are both from ICD-10 chapter XVIII ‘Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified’. This ICD chapter is reserved for cases where there is no consensus over a

cause which causes the death of a person. The chapter accounts for 2.06 per cent of all deaths and since it is not much different to ICD-9, it can be seen that rates have been starting to rise since the 1980s, especially with regards the use of 'old age' (Griffiths et al., 2005). Nevertheless, this may just reflect population dynamics, with people living much longer now, rather than being a large proportion of wrong cases. As people get older, their body is more likely to just fail through 'wear and tear' and so their inclusion is justified. However, the issue of deaths recorded as unknown is not just confined to these two variables. For example C80 ('malignant neoplasm without specification of site') accounted for 1.96 per cent of all deaths however as just a variable by itself, it is not useful. Such large numbers of deaths simply coded as this masks the true patterns of more specific cancers.

An issue with this selection process is that choosing the variables solely based on numerical reasoning may ignore important distributions. Whilst some causes are less common, they may still feature important social and/or geographical patterning. For example this is especially the case within 'Other Accidents' variables which contains both water transport and air traffic accidents. These both have opposite social distributions, as well as differing spatial patterns (Shaw et al., 2008). This issue was also identified by Griffiths et al. (2005) who found that aggregating accidental mortality rates can be problematic. Whilst most other sections of the ICD-10 have been split up to reveal more specific causes, there may still be issues with the aggregation process for the broader variables. These will occur where differing processes relating to causes occur, which may hinder the understanding of the classification, as well as future policy usage. However where aggregation has occurred with such small numbers, they are unlikely to have a large impact on the overall model.

3.7.3 Geo-referencing the data

A look-up table compiled by the ONS matched OAs provided in the DOEW to MSOAs, allowing each individual case could be assigned to its correct geographical area quickly (using ArcView and the 'table join' function). This simple process however led to 34 areas becoming corrupted and these were manually altered. Next a look-up table was created for the variables used to convert the ICD10 codes to each variable to be included in the analysis. This was then exported into SPSS and the data was compiled to its geography through using the 'aggregate' option. The total was then found for

each variable (adding together the values for each MSOA) and this was then compared to the totals to validate the process and check for errors (Table 3.12). Those variables which were different from the totals gathered had missing (or erroneous) geographical variables.

Variable no.	1	2	3	4	5	6	7	8
Sum of Variables	12902	8928	18547	26098	17995	35391	14329	11507
Total of database	12902	8929	18546	26098	17994	35393	14328	11507
Difference	0	-1	1	0	1	-2	1	0

Variable no.	9	10	11	12	13	14	15	16
Sum of Variables	27390	118883	42824	14909	36810	12252	17555	12737
Total of database	27390	118885	42824	14909	36811	12252	17556	12738
Difference	0	-2	0	0	-1	0	-1	-1

Variable no.	17	18	19	20	21	22	23	24
Sum of Variables	15588	26746	127492	3926	21649	6993	61315	6362
Total of database	15588	26748	127490	3926	21649	6993	61315	6362
Difference	0	-2	2	0	0	0	0	0

Variable no.	25	26	27	28	29	30	31	32
Sum of Variables	18040	23015	24893	27	77	17266	120378	188493
Total of database	18040	23015	24893	27	77	17266	120378	188496
Difference	0	0	0	0	0	0	0	-3

Variable no.	33	34	35	36	37	38	39	40
Sum of Variables	13852	12878	33275	40667	17878	17917	88896	59867
Total of database	13852	12878	33275	40630	17878	17917	88897	59867
Difference	0	0	0	37	0	0	-1	0

Variable no.	41	42	43	44	45	46	47	48
Sum of Variables	29921	15489	13598	112279	103135	12031	48837	11582
Total of database	29921	15489	13598	112279	103135	12031	48838	11582
Difference	0	0	0	0	0	0	-1	0

Variable no.	49	50	51	52	53	54	55	56
Sum of Variables	9455	19629	18270	11165	10666	21311	7360	16935
Total of database	9455	19629	18270	11166	10666	21311	7359	16936
Difference	0	0	0	-1	0	0	1	-1

Variable no.	57	58	59	60	61	62	63	64
Sum of Variables	11904	33911	160	63	3887	35177	5872	12801
Total of database	11903	33912	160	63	3887	35176	5872	12801
Difference	1	-1	0	0	0	1	0	0

Variable no.	65	66	67
Sum of Variables	30372	12359	6632
Total of database	30352	12361	6632
Difference	20	-2	0

Table 3.12: Validation check for the variable aggregation process.

3.7.4 Standardisation of mortality

The aggregation of the total number of deaths by cause for each MSOA tells us very little about the mortality of England and Wales. Since certain causes dominate certain types of people (for example Prostate Cancer; Majeed et al., 2000), this is a poor measure. Therefore there is a need to standardise rates to allow for a meaningful data set. The standardised mortality ratio (SMR) was chosen to initially account for the different factors that could cause a certain amount of deaths in an area. This controls for the age and sex make-up of each MSOA, therefore not creating a data set which shows that the most deaths occur where there are more elderly people for example. By having all variables measured on the same scale, this avoids the results becoming misleading. An indirect approach was chosen to be taken since it provides a more stable variable when dealing with small numbers, as for many variables used here.

Population estimates by age and sex bands for the years after the 2001 Census were requested from the ONS. Whilst these do exist on Neighbourhood Statistics website, the age bands at the MSOA level are broader than required. The requested data gave a more accurate and useful set of age bands. The following bands were used; 0, 1-15, 16-24, 25-34 and then ten year bands up to 85 and over. This was chosen as it is what is

used throughout the literature (for example Shaw et al., 2008). The SMR is calculated by:

$$SMR_{\chi} = 100(\alpha_{\chi} / \Sigma(\beta_{\chi\gamma\upsilon}\delta_{\gamma\upsilon})) \quad (3)$$

χ represents the cause of death variable, α is the number of observed deaths for an area, β is the mortality rate for England and Wales, γ is the age band, υ is sex and δ is the population size. The equation calculates the percentage difference in the total number of deaths in an area for a particular cause against the expected number of deaths given the age and sex make-up of an area's population (i.e. based on the mortality rate for the whole population).

Further standardisation could also be employed, to normalise the data, as has been conducted in other classifications (for example Vickers, 2006). There are plenty of different standardising techniques, with standardising using the range of a variable often viewed as the best (Milligan and Cooper, 1987). However with all of the variables measured on the same scale, it is not as important (Everitt et al., 2001; Gordon, 1999). As Milligan and Cooper (1987) argue, standardisation is not always required, as it can obscure patterns (through producing false model dependent results) through transforming the data (which itself becomes less interpretable than compared to the SMR score). The inputs will not be standardised additionally.

The variable standardisation process compares the data in each MSOA to the national average. This means that you cannot really directly compare two areas against each other, rather only to the national average. Comparisons between clusters will need to be made carefully in response to this. It may have been better (or more accurate) to standardise regionally, to take account of known geographical variations (Hacking et al., 2010). This would allow for stronger or unexplained mortality patterns to be visualised. However few (if any) past studies have tackled this and instead always choose to compare nationally, as it provides easier to understand comparisons.

SMRs have also been known to be fairly sensitive, with small fluctuations in trends having large changes in values. As a result, this can have a strong effect on the classification process. The aggregation of time periods should help stabilise and reduce such errors. However some variables, such as 'Diseases of the Ear and Mastoid Process', only witnessed few deaths. Therefore the expected amount of deaths for each area is going to be very low and areas with few deaths may skew the distribution. This problem will be addressed in the next section.

3.7.5 Weighting the variables

The process of standardisation has made each cause of death equal. This does not relate to the true geography of mortality as the prevalence of causes is ignored. For example 'Conditions Originating in the Perinatal Period' caused 63 deaths between 2006 and 2009, yet it is given equal importance to 'Chronic Ischaemic Heart Disease', even though the latter kills 188,433 more people in the same period. As the classification methodology groups together cases based upon the distribution of data, this will be problematic. For the study to be representative of the true structure of England and Wales, this needs to be accounted for. The SMRs were multiplied by the percentage of the total deaths (see Table 3.6) that they make up within the data set. This method retains the structure of the variable (for example a SMR of 100 which has been multiplied by 0.5 per cent means that 50 is now the average value). Other statistical weighting procedures are available (Everitt et al., 2001), however this approach is theoretically more relevant.

There may be some issues resulting from this process. The more prevalent diseases may end up dominating the classification, with clusters determined based mainly on their patterns only. As a result, any interactions between variables could become obscured. Where less prevalent diseases contain higher values (or even a cluster themselves), this may become lost. Furthermore, some disease which are fairly prevalent but represent amalgamations of ICD-Chapters can be problematic due to this. For example, the variable 'Other Cancers' accounts for 6.4 per cent of all deaths, but weighting in its favour is unhelpful to our understanding. These possible effects can only be investigated when interpreting the clusters. As Sneath and Sokal (1973) argue, weighting without a strong and consistent basis is a (subjective) classification itself. Therefore the choices of weighting will be kept.

3.7.6 Inter-related variables

With such a large amount of variables used in the analysis, it is important to decide whether all the variables are relevant and not unnecessary. If there is a strong relationship between any two variables, then it could be said that one variable mostly predicts the other (Rogerson, 2006). Therefore one could be dropped, since the variation in mortality will be mostly accounted by the other. It also served to briefly

examine any interactions between causes, aiding future interpretation of the classification.

Variables	Correlation
Lung Cancer Acute Myocardial Infarction	0.433***
Lung Cancer Chronic Ischaemic Heart Disease	0.424***
Lung Cancer Chronic Lower Respiratory Diseases	0.644***
Lung Cancer Alcoholic Liver Disease	0.443***
Dementias Alzheimer's	0.510***
Dementias Other Cerebrovascular Diseases	0.446***
Alzheimer's Other Cerebrovascular Diseases	0.419***
Acute Myocardial Infarction Chronic Lower Respiratory Diseases	0.474***
Chronic Ischaemic Heart Disease Chronic Lower Respiratory Diseases	0.464***
Chronic Lower Respiratory Diseases Alcoholic Liver Disease	0.448***

Table 3.13: Moderately Correlated Variables.

*Note: Significance levels; * is p<0.05, ** is p<0.01, *** is p<0.001.*

Using a Pearson's correlation to examine the relationship between each variable combination, the majority of cases were either insignificant or poorly correlated. Only ten relationships (0.44 per cent) could be considered moderately strong and significant (i.e. greater or less than ±0.4). These could present early indications of causes which 'travel' together and are shown in Table 3.13.

The largest of these was only 0.644 ($p<0.001$). This was between the weighted variables 10 ('Lung Cancer') and 45 ('Chronic Lower Respiratory Diseases'). This may be expected, since the diseases are found in variable 45 (Asthma, Bronchitis and Emphysema) have similar causes to 'Lung Cancer' (primarily smoking but also air pollution; Cornfield et al., 2009). The correlations for four other relationships (10 and 31, 10 and 32, 31 and 45, 45 and 32) may also be linked to smoking. The negative effects of smoking are not just confined to the Lungs, but also are detrimental to the heart through the effects of carbon monoxide (Aronow, 1973).

The moderate relationship between Lung Cancer and Alcoholic Liver Disease (as well as 45 and 51) may suggest that those areas which have people engaging in excessive alcohol intake may also contain high numbers of people smoking also. There is much evidence of the strong association between the use of these two substances with detrimental effects on health (Room, 2004; Hart et al., 2010), as well as the link between their combined usage. Therefore these two cultural practices may 'travel' with each other.

The final three associations are linked to brain dysfunctions. Dementia and Alzheimer's are similar diseases (of old age), especially since many Dementias eventually turn into Alzheimer's. This suggests similar causes, or that the types of people more susceptible to it (for example the elderly) live together. The relationship between 'Other Cerebrovascular Diseases' with these two diseases may also reflect similar conditions due to them all being related to the declining quality of the Brain (Rocca et al., 2011). As no variable displays a strong relationship (taken here using the rule of thumb of greater or less than ±0.7), no variables were dropped from the analysis.

3.7.7 Confidentiality

Confidentiality is important with the use of the DOEW data set, as each individual death is provided within the data set. This is more of an issue for the least prevalent causes of death, which may have sparse counts (mostly where counts are fewer than 5). Also with this study conducted at fine geographical scales (defined by the ONS as local authority level or below), individual deaths may become identifiable. Any disclosure of personal information can be punishable under the 'Statistics and Registration Service Act 2007'. These issues must be taken into account to avoid any potential harm or distress caused by identification of deaths to any living relatives (ONS, 2010a). Due to the methodological approach conducted in this thesis, these confidentiality problems will not be an issue. Since a classification will reduce the data set containing every cause of death into a much smaller number of clusters summarising the data, individual records will not be able to be identified. The raw data is also stored on one computer in a locked room within the Geography Department at the University of Sheffield, which prevents unlawful access.

3.8 Choices in the k-means method

There are various decisions which must be considered when running a k-means methodology which may affect the overall results. These are detailed in the following sections.

3.8.1 Software

Whilst it may not initially seem to be part of the process, this choice is one of the earliest decisions made in creating a classification. There are many different programs and applications which could be used to conduct a cluster analysis on the data. Since many pieces of software use varying algorithms and conditions to conduct a cluster analysis, it is an important decision to make. PASW version 18 (formerly SPSS) was used due to the author's extensive knowledge and experience of working with the program. Whilst this is not suggestive that one piece of software is better than another, it is important to note. This is especially the case for any replication of the analysis in a different program as this may yield slightly different results (especially if it is found that the classification is particularly unstable). The program allows for a k-means method to be selected. The seed points can be specified and the analysis can save the distance of each object to its cluster centroid. These last two options are particularly useful for applying testing procedures to the classification.

PASW will not be used for all parts of the analysis. Although the software can perform most of what is proposed, it is limited in the amount of tests that can be run to test for the number of clusters to be taken. Therefore the software 'R' will be incorporated into parts of the thesis which require more specific processes. Although it could have been used for the whole thesis, the author's unfamiliarity with it, coupled with the fact it is less easy to use than PASW meant it would only be used when necessary. Whilst the k-mean algorithms between programs may be unlike each other, these should only be slight differences which may result in cases on the edge of groups being differently classified (i.e. least distinct objects).

For example, a simplified two cluster solution was run using the final data set compiled for this study to illustrate this point. When the k-means algorithm was run in SPSS, it gave clusters with the sizes of 2900 and 4294. Running the same algorithm in R produced clusters containing 2899 and 4295 cases highlighting the similarity of

algorithms between programs. Equivalent cluster membership was equal for all data points bar one, which was located near the boundary of the cluster centres in each, the reasons for this difference. As it is being used to specify the underlying cluster structure, exact locations of data points are less important than compared to finding the optimal solution for the number of clusters. The results of the classification will be driven by the patterns in the data rather than the choice of software.

3.8.2 The starting points for clustering

As the methodology begins by dividing the data points into *k* (specified by the user) number of clusters which are later refined, how these initial partitions are chosen will play a large role in the classification produced (Harris et al., 2005). A decision is made to select the initial cluster centres, for which the *k* clusters will be clustered around (i.e. each case is moved into the cluster which it lies nearest to). This step can result in the production of a classification that achieves local optimum (a small cluster) rather than a global optimum (the underlying structure of the data) (Gordon, 1999). A starting (or seed) point may select a case which leads to a small, specific cluster that does not accurately reflect a major cluster in the whole data set.

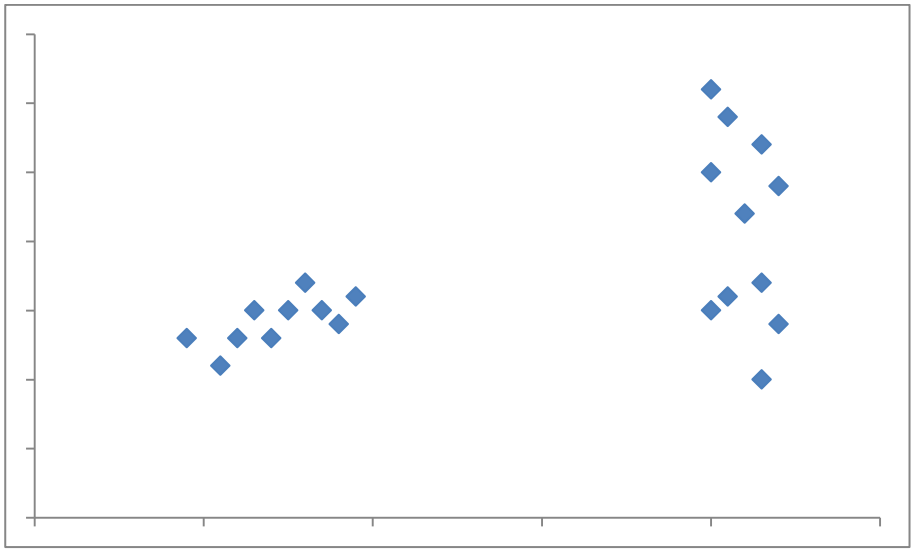


Figure 3.6: A hypothetical three cluster example data set.

The issue is illustrated by a hypothetical example, using made up data across two variables. Suppose we have identified that we expect the data to display a three cluster solution. Figure 3.6 shows the data and this appears correct visually. The clusters have been emphasised in the data creation process, so that the issue can be better visualised.

Real data will not be as well separated, with the cluster analysis required to divide up the data into something more interpretable and useful. Running a cluster analysis on the data may, however, find an erroneous solution.

Figure 3.7 shows a hypothetical scenario where the starting points were erroneous selected. Although three clusters have been identified, it does not reflect what is actually happening in the data. A different problem occurs in Figure 3.8, where outliers are present and taken as an independent, yet erroneous, cluster (although not in the eyes of the algorithm, as this cluster is very compact). The effect of outliers will additionally be considered later through the statistical testing of the classification (see Section 3.9.2).

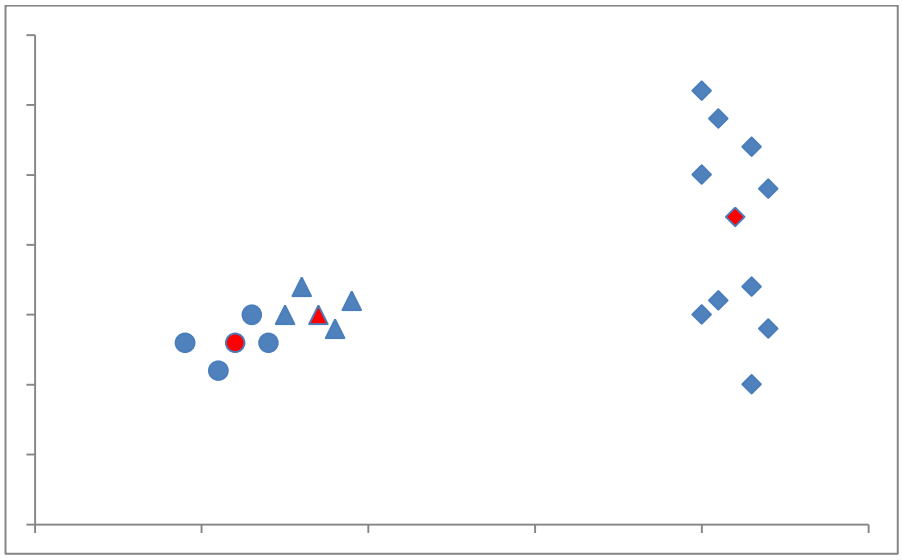


Figure 3.7: A hypothetical classification of the data set into three clusters.

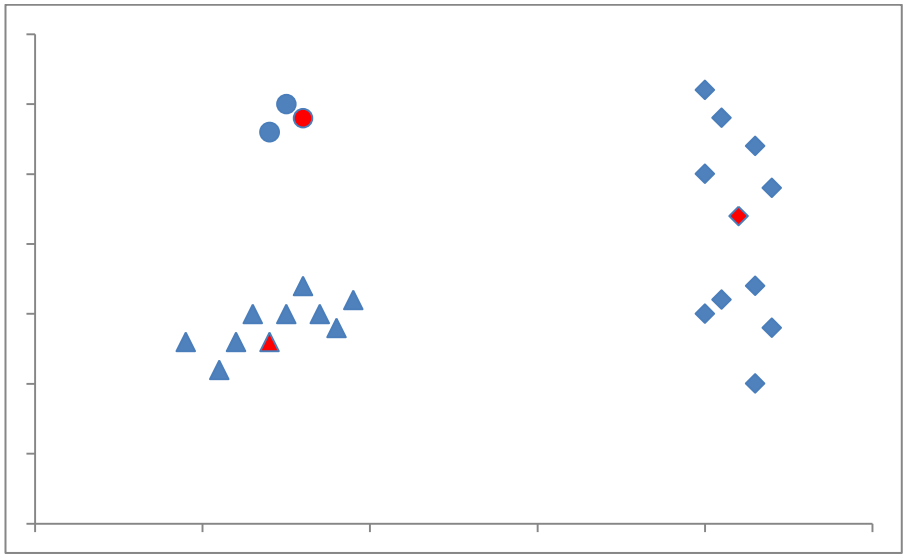


Figure 3.8: The possible impact of outliers selected as seed points.

The default approach of PASW 18 is to choose the initial cluster centres through selecting the first k cases in the data set. This seems problematic, especially as the data is sorted by geography. Ordering of the cases is important and the classification will be sensitive to this possible error. Instead the initial seed points will need to be specified. One approach would be to randomly select *k* cases (Duda and Hart, 1973; Gordon, 1999), so that the choice of starting points are fairer as each case has an equal chance of being selected. This could also be altered, to make it more systematic through randomly sorting the data and then choosing *k* cases at even intervals (i.e. every *n/k* case). These methods, however, still could be affected by the localisation issue.

An alternative would be to run a hierarchical analysis and then assign the centroid values for the chosen solution (Everitt et al., 2001). The results from the hierarchical method would not be the same as a k-means cluster analysis due to the different methodologies involved. Rather this would provide a base of which the k-means analysis could later refine to create the optimal solution. Krieger and Green (1999) note that use of random starting points is less effective than this approach. This problem was also found by Milligan (1980) who found that random seed points were much less effective than using 'rational' points (i.e. performing a hierarchical analysis and taking the points from that). However with a large data set, the results may become unstable and take a long time to be calculated. Yet the k-means analysis afterwards would refine and improve upon these results, reducing any impact. This approach is not completely safe from the localisation problem either, though it is less likely to occur.

These two possible methods will be considered in Chapter four to explore which produces the more optimal solution.

3.8.3 Measuring 'closeness' of objects

There are two ways to measure how close objects are to each other. Similarity (homogeneity) measures look at the proximity of objects, with a high value meaning two objects are close together. Dissimilarity (separation) measures examine the distance between objects, meaning that a low value represents two objects close together. Both methods are useful, yet dissimilarity measures are more commonly used. The exact choice will also depend upon the data types used. Since the classification

only includes continuous (ratio) data, the relevant measure will be chosen in accordance to optimise this.

There are many different measures for continuous data, however most are similar or are irrelevant to this study (Gordon, 1999). The most commonly used measure is Euclidean distance, which is calculated through:

$$d_{ij} = \{ \sum w_k^2 (x_{ik} - x_{jk})^2 \}^{1/2} \quad (4)$$

Here d_{ij} represents the distance between objects *i* and *j*, with x_{ik} being the value *x* for object *i* of variable *k* and x_{jk} being the same for object *j*. This calculates the straight line distance between objects. w_k^2 is the weight for the variable *k*. The classification then constructs a *n* x *n* matrix of the physical distances of all objects, to see which are closest together and where best to join objects. A variation of this measure is the squared Euclidean distance. The equation is similar, rather you do not square root the sum of distances between objects. There is little difference between the two measures. Squared Euclidean distance is used where there are large data sets, since with an iterative process the method runs quicker (Everitt et al., 2001). It also puts increased importance on data points which lie further apart.

Other useful methods that measure the dissimilarity of (continuous) objects include Minkowski distance, City Block distance (both provide variations on the Euclidean distance) and Canberra distance (this appears more favourable to binary variables than continuous) (Everitt et al., 2001; Gordon, 1999). A modified Pearson's correlation coefficient can also be used to estimate the similarity of two objects, although this is seen as slightly controversial as it ignores size (Gordon, 1999). Euclidean distance shall be used in this study as the software being used to conduct the analysis restricts the methodology to this measure.

3.8.4 Selecting the number of clusters

Most cluster analysis methods do not tell us about the actual number of clusters present in the data. This leaves the researcher with the dilemma of what solution should be used. This issue is even more important when conducting a k-means analysis since selecting the number of clusters is a pre-requisite for the cluster analysis to take place. It is commonly used as the key criticism of the specific methodology, since it can have a

large effect upon the results as other parts of the process are determined by it (Bezdek and Pal, 1998; Everitt et al., 2001).

Selecting the number of clusters to cluster can therefore present two errors; producing a solution with either too many or too few clusters. The latter error is considered more problematic since it results in information lost through cluster merges (Milligan and Cooper, 1985). Where two clusters have been combined together into one cluster to force a specific solution, the end result is a cluster unrepresentative of the data it is supposedly summarising. Whilst with unsupervised techniques, we can never truly know the correct number of clusters present in the data (and cluster structure), we can at least test for which solution is best for the model and aid our interpretation of the data.

One common technique to judge which solution may be better is to compare the average distance of each individual case from its cluster centre (Everitt et al., 2001). This shows how compact the clusters are for a particular solution. This statistic is gathered for a range of solutions and then can be compared graphically. A trade-off ensues, as increasing the number of solutions leads to a more detailed classification, but too many clusters can make the data reduction confusing and lead to local solutions/clusters (Gordon, 1999). The solution to be used may be the one where there are no large reductions in the value afterwards, as the downward trend begins to smooth off. Scree plots are ideal for the former statistic, visually showing where the 'natural break' occurs. The plot will usually decline over range of solutions, with each one improving over the previous, until it reaches a point where this decline slows as the clusters have become more compact (Duda and Hart, 1973).

Another test concerns a different aspect of the k-means method, by comparing average cluster size. This is useful, as ideally you are looking for similarly sized clusters. Small clusters may be outliers (Vickers, 2006). This detracts from the interesting information elsewhere, as they become lost in other larger clusters as a result. For example with Figure 3.8, the impact of the outliers results in the variation and split of the two clusters on the far right is lost. Simply looking at average size will always fall since it will always be a case of n/k (i.e. n (sample size) remains constant, as k (number of solutions) continues to rise). The average difference from what is expected (i.e. n/k) will be used instead to compare the difference for whether the model produced evenly sized clusters or not.

Nevertheless this approach of comparing these measures through the use of graphs is still relatively subjective. The decision of which solution is still left to human judgement (for example 'what is a small value?') and therefore relatively informal. It may be useful to supplement them with some other statistical measures as well to help advise this choice. Yet there are simply too many different possible measures and hence only the most commonly used ones should be considered (Bezdek and Pal, 1998). Milligan and Cooper (1985) produced a comparison of 30 possible methods which are more formal and this provides a useful point of reference. Choosing some of the more effective measures to apply in this thesis would improve the strength of this subjective choice, especially as it moves away from relying on just one or two measures.

Using measures chosen from this comprehensive study could be problematic. The data sample size is low (50 points), making it less comparable. The tests are only examined for a small range of clusters (up to five) and a larger range of solutions may be wanted to be explored here. The results of simulation studies may not always allow for generalised knowledge (Gordon, 1999; Milligan, 1996). The findings may be dependent upon the cluster structures, cluster sizes, number of solutions or chosen algorithm, meaning that they may not be as useful for this thesis. However as Milligan and Cooper (1985) argue, it is difficult to justify applying a method which performs poorly on well-defined data to test a less structured data set. It should also always be remembered that "Since the stopping rules are heuristic, ad hoc procedures, an applied researcher must critically examine the suggested solution provided by any such index." (Everitt et al., 2001, p160). This also applies to all other tests carried out in this thesis.

The index which performed the best in Milligan and Cooper's (1985) study was the C-index (Hubert and Levin, 1976). This displayed excellent cluster recovery, making a low amount of errors, especially with regards to producing two few clusters. Of the few errors encountered, these tended to be close to the true solution. The index looks at the homogeneity of clusters. It is calculated through the equation:

$$= d_w - \min(d_w) / \max(d_w) - \min(d_w) \qquad (5)$$

Here d_w is the sum of within cluster distances for all the clusters, with $min(d_w)$ the smallest value for any cluster and $max(d_w)$ the largest. Since a cluster analysis seeks to create a solution with compact clusters, a smaller value on the index will indicate a better solution. Similarly with this index, it is not so much about the exact value but rather the overall plotted trend and how it changes between values of *k*.

Another variable which performs well in Milligan and Cooper's (1985) study is the Davies and Bouldin (1979) Validity Index. Whilst it did not perform as well as the previous indicator, Milligan and Cooper still present the index as favourable despite producing more solutions with too many clusters. It examines the similarity of clusters, by looking at the ratio of within to between cluster distances. The index is calculated by:

$$= \frac{1}{n}\sum_{i=1}^{n} \max_{i \neq j} \left\{ \frac{S_n(Q_i)+S_n(Q_j)}{S(Q_i,Q_j)} \right\} \quad (6)$$

Where *i* and *j* are individual clusters, S_n is the average distance of cases to the centre of cluster *Q*, $S_n(Q_i, Q_j)$ is the distance between the centroids for clusters Q_i and Q_j and n is the number of clusters present in the solution. The index calculates an average value of the largest values involving each cluster. This finds the ratio of dispersion against separation. A lower value would suggest a better solution.

These tests however are only a few of many possible tests that could have been chosen. Indeed Milligan and Cooper (1985) only consider 30 indices and yet many hundreds more exist. Milligan (1996) notes that until a more complete measure can be developed, at least two or three should be used. If there is only partial agreement, then a larger set should be taken. An issue that arises is for what range of solutions these tests should be conducted for.

There is not much direction provided in the geographical literature as to the maximum value which should be tested for. In the creation of OAC, Vickers (2006) used the literature to decide to test between two and 12 possible solutions, before settling on seven. Other area classifications include ACORN (CACI, 2013) and GB Profiles '91 (Openshaw et al., 1994) with six clusters and MOSAIC with 15 (Experian, 2009). However there is very little past evidence equivalent research on health and mortality classifications. Shelton et al.'s (2006) study contained ten clusters whilst 'HealthACORN' (CACI, Unknown Date) only has four. Measures for displaying mortality data for England and Wales include GORs (10), NS-SeC group (seven, with the other two often ignored) and Indices of Deprivation (quintiles).

Bezdek and Pal (1998) suggest an ad hoc test to calculate this:

$$c_{max} \leq \sqrt{n} \quad (7)$$

c_{max} represents the highest possible number of clusters which could be used in the classification, a function of the square root of the sample size (*n*). However applying

this to the data set gives a value of 84.82. This figure presents hardly much of a data reduction and certainly makes the process inefficient. Although it only gives the maximum number of clusters to be incorporated into the classification, it does not aid our analysis.

This figure needs to be relatively low to aid interpretation, but also wide enough to allow for much variation. The number must also not be too large so that the interpretation of the solution is compromised. When visualising the classification, too many different clusters (and hence colours) can harm our understanding as colours may be less distinct from each other. The eye works through searching for breaks in common patterns and therefore the more colours used, means that greater effort is required, which makes interpretation confusing (Dorling, 2012). There are also further considerations involving future use of the classification. If there are too many clusters then this may become problematic when applying the classification to statistical tests, which may lead to confusing results. Whilst this figure is arbitrary, the range of two to 16 was chosen to fully capture the expected range where the true solution lies.

Another issue that arises through this process is related to Section 3.8.2. In theory, when performing each test for every possible solution value, the seed points should be specified more accurately. Nevertheless this would take far too much time to compute. Whilst R uses random seed points, it also allows for multiple runs to be conducted. This will ensure that the findings are more robust and not due to localised errors. The average score will be taken from this.

3.9 Testing

With no prior expectations as to what the classification will find, evaluating the results can become difficult (Openshaw and Wymer, 1995). Given the nature of how the classification is constructed, how do you objectively measure how 'good' it is? A common critique of cluster analysis is that the methodology will always produce an answer (Harris et al., 2005). Thus where there may not be an underlying structure to the data, the method will force a structure to it (see Section 2.7.5). The results should be driven by the data, rather than being an artefact of the methodology (Everitt et al., 2001). It is straightforward to create a classification, yet it is less simple to actually create something that is accurate, representative and, more importantly, useful.

Testing is not about identifying how 'good' the classification is. Rather it requires assessing through statistical testing to check whether the model is robust. This is all that can be done to ensure a high quality set of results (Everitt et al., 2001). It is less about having each area correctly classified in its actual cluster, but that the relationships of the clusters overall fit a stable, applicable and useful structure. As such, accuracy is more important here for testing than precision.

There are many hundreds of possible methods which could be applied here, however many are dubious in their application and others are irrelevant. Most of these tests involve making slight changes to the rules and parameters of the process, since a strong classification would display similar results when faced with small changes. This represents a clearly structured data set, moving towards the statistical concept of 'generalisability' (Milligan, 1996). This section will present the most common measures which are of most use to the study. Only once the classification has been fully tested can it finally be described and analysed further (Gordon, 1999).

3.9.1 Replication

A different validation technique for checking whether the clusters found are robust is through conducting a replication analysis. The most common approach is the 'split-sample' method which was developed by Blashfield and Macintyre (1980). The data (cases) are split in half and the first sample is clustered. The centroids gained from this are then used to assign the cases in the other sample to determine which each data point it is nearest to. Then a cluster analysis is performed, using exactly the same rules and parameters, on the second sample. These results are then compared to the cluster membership produced earlier by applying the centroids from the cluster analysis of the first sample.

Krieger and Green (1999) note that there is much evidence showing that this has been successful in evaluating the stability of clusters. Nevertheless, they also note that there are many issues with the approach, especially for unsupervised methods. The approach does not seem to work when clusters lack definition (or are not very unique). This is more problematic if the true number of clusters is unknown. Everitt et al. (2001) also notes there are issues where there exist uneven cluster sizes, especially where there are many different variables. Milligan and Cooper (1987) suggest that the analysis should

be able to be replicated with different data. This is not relevant here as it is harder to achieve, especially as this kind of study can only be carried out now that this data is available.

3.9.2 Influence of individual points

This will look at whether certain areas are having more of an impact on certain clusters than others. However the question arises as to how first identify an outlier. A typical approach would be to look at the standard deviation from the mean for the Standardised Mortality Ratio values and identify areas which lie outside the bounds of -3< $\bar{x}$ <3. However with 67 variables, does an area which has just one extreme value count as extreme? This is seen in the data, as 2974 areas (41.3 percent) contain at least one variable as an outlier.

Instead Cheng and Milligan's (1996) framework will be followed (although it was not performed using a k-means method, the approach is still relevant). They applied a different approach by considering the location of cases in relation to their cluster centres. These are not outliers in the case of extreme data values but rather outliers in relation to cluster membership or in how well the classification fits the data.

The influence of a data point can be examined by comparing two cluster analyses. The classification is used as a reference point (since the true underlying structure is unknown, this is used instead). Then the classification is re-run (with the same rules and parameters) with a (single or series of) data point(s) removed and cluster membership of cases is then compared. Cheng and Milligan also argue in favour of using cluster criteria (for example within sum of squares) or more formal measures to examine whether the point was beneficial or damaging to the model. It is likely that there will be more influential cases where there is a less well defined underlying cluster structure.

If some points are having a negative impact upon the outcome, should they be deleted? This debate is not just found within classification studies but also applies to other forms of analysis (for example linear regression and the impact of leverage points). As Cheng and Milligan (1996) argue, common logic would suggest that any points which inhibit the overall quality of the classification should be dropped from the final model. The classification is about the exploration of relationships and patterns found between the

causes of death. However it is also a summarisation of mortality patterns across England and Wales and therefore excluding data since it has some slight detrimental effect on the classification reduces the ability to compare patterns geographically (as well as look for possible explanatory spatial factors). There are also questions as to how do we define when a point should be dropped? What happens when you have many points having an effect (is this also independent)? Testing is about looking at how stable the classification is and therefore the exclusion of points should depend on this.

Whilst the previous section looked at the impact of data points on the stability of the model, it would be useful to explore the extreme values within each cluster. Through examining the strength of the assignments of cases within each cluster, it is possible to evaluate how successful the cluster procedure is. However how do you define what is an extreme value? Openshaw and Wymer (1995) suggest that any points which lie past ±3 standard deviations from the cluster centre should be considered. Cheng and Milligan (1996) argue for a stricter cut off, using ±1 standard deviation from the cluster centre. Whilst this would appear false (when compared to the normal distribution), it provides a more rigorous testing procedure (especially as cluster analysis do not require the assumption that the data is normally distributed; Milligan, 1996). Both approaches will be adopted, removing values in accordance to either and then re-running the classification to assess their impact.

3.9.3 Impact of variables

A sensitivity analysis will be used to explore the individual impact of the variables. This will shows the impact each variable has on the formation of the clusters. It can both help understand what is driving the segmentation of the data, as well as identifying those variables having little impact in the classification other than adding random noise, obscuring patterns (Milligan, 1996). This approach was useful in the creation of the Output Area Classification (OAC) as part of its testing (see Vickers, 2006). The procedure begins by removing each variable individually. The analysis is then re-run with the same rules and parameters as the original classification for each possibility, containing all variables bar the one excluded at that stage. The results of each are then compared to the main classification, through statistical measures (see Section 3.8.4) to

assess their effect on the segmentation process. Comparisons can also be made between the solutions of cluster membership and subsequent changes in them.

3.9.4 Interpretation

Where there is data that we do not understand what to expect when a classification processes is applied to it, ground truthing is paramount to validating any results (Duda and Hart, 1973). The preceding tests will highlight whether the classification is statistically stable, however this tells us very little about whether the classification accurately portrays the true underlying structure of mortality patterns across England and Wales. It is very difficult to work out if the clusters found actually make sense or have meaning. Rather a different approach is required.

Interpreting the classification is the most exciting part of the process. Here we learn about the clusters and what they mean. However interpreting these clusters also forms part of testing, as it allows us to assess whether the results make theoretical sense. Since the classification is an exploratory form of analysis, this is required (Gordon, 1999). Chapter five will be dedicated to this, analyzing and examining the patterns in relation to the literature to ensure that the clusters are not just statistically robust. It is the final stage of testing and its importance is best summarized by Milligan (1996, p365): *"The bottom line is that any clustering or classification will be useful only if the results can be substantively interpreted."* However as Gordon (1999, p183) argues, we should not just rely solely on this; *"...the human brain is quite capable of providing post hoc justifications of results of dubious validity."*

3.10 Conclusion

This chapter has reviewed and detailed the methodological and data related decisions required to build an area classification. Building a classification does not equate to a strict set of rules and procedures, unlike other methodologies. Rather there is freedom to pick and choose different approaches. Although this allows for greater customization to build a classification to a desired specification, it increases the possibility of erroneous choices. This is especially important given that the method is data driven. Poor data quality and choices will only produce results of low value.

It is important to use a framework to ensure that the correct direction is taken, with each choice justified to benefit the model. These decisions were based upon the approach set out by Milligan and Cooper (1987), but also adapted to accommodate the additional needs of building an area classification. The clear structure of development is important for any future user or reader to be able to evaluate the quality of the classification (Milligan, 1996).

Chapter 4: Creating an area classification of mortality patterns

4.1 Introduction

This chapter begins the process of building a classification of mortality patterns across England and Wales. The chapter is split into two main sections, although there is some overlap between the two. It begins through detailing the initial considerations and decisions made, which affect how the classification is constructed. With these rules and parameters decided, the initial results will be statistically tested. This will allow the study to examine whether the clusters found are useful, stable and reflective of an underlying structure throughout the data. Based on these findings, the model may be modified or improved. The chapter is informed by the methodological discussion throughout Chapter 3.

It is relatively easy, with the right knowledge and access to software, to build a classification. It is, however, less simple to create a classification that is useful. However it is important to note that the processes detailed here are not necessarily about making the best classification mathematically. There are not a series of metrics which require optimising, since there is no 'best' solution (Everitt et al., 2001). Rather the process of building and testing a classification requires focus to concentrate upon how useful it is. Each stage needs to justify why the approach was taken was valid and the most appropriate given the data and objectives (Gordon, 1999). This is especially the case where decisions can impact upon the results, meaning any choices should be grounded.

Given that some choices are subjective (albeit informed decisions), this is important. Even with objective measures, there is little indication within the literature as to key figures to compare against. Every classification is different and hence it is not possible to make a formal assessment (Everitt et al., 2001). What we are looking for instead is variability, especially relative rather than absolute changes. The classification needs to be shown to be stable, so that the clusters are consistent. It is about showing that the underlying structure of the data found in the analysis exists and hence the analysis is

useful. Detailing each decision clearly, whilst testing for stability in the results is paramount for achieving a valid classification.

Building the classification

4.2 Method of selecting seed points

Seed points refer to the choice of the initial partition (i.e. clusters) made into the data, of which the method refines to produce the solution. This is an important and unavoidable part of the start of the k-means methodology and the effect of approaches requires evaluating to ensure that the model results are not in part influenced by the method chosen. As such, the resulting classification is more stable and valid.

4.2.1 Random sampling

As detailed in Chapter 3, the k-means cluster analysis methodology begins by dividing the data into *k* clusters. These *k* clusters can be derived by randomly selecting *k* cases from the data set and then using the values of each variable for that case as the initial starting cluster centres. Therefore which cases are selected can lead to varying results (Milligan, 1980). Even when the clusters in the data are well defined and distinct, the clustering process may not always refine the solution to define these clusters (i.e. the global solution).

An important issue involves the creation of local clusters (Everitt et al., 2001). For example if these initial starting points were to land on outliers and extreme values (with few other cases which are similar), clusters may form of just these cases. Since the few points are very alike and yet much different from the rest of the data, they fulfil the clustering processes aims of creating clusters which are separate and compact (Gordon, 1999). This means the cluster will not be changed, since separating it or adding other values will not lead to an improved classification.

The first approach proposed in Section 3.8.2 was random sampling (the default setting in PASW). An ad-hoc Monte Carlo approach has been adopted to efficiently assess the stability of a classification produced using random seed points. The classification shall

be re-run with different starting points, selected by the same method (Gordon, 1999). This will show the sensitivity of the choice of starting point on the classification's clusters (i.e. changes in cluster sizes, cluster membership). This was done 20 times, with the number of clusters set to eight to allow for the possibility of local clusters to form.

Iteration	Cluster							
	1	2	3	4	5	6	7	8
1	2533	380	3	77	1115	1117	1151	818
2	653	2014	1108	73	773	1276	1071	226
3	2492	3	647	968	391	71	1279	1343
4	1270	71	2081	1403	785	384	254	946
5	767	835	67	1491	2098	379	1312	245
6	1078	593	1249	2327	1117	193	69	568
7	81	775	1387	367	316	1246	962	2060
8	1481	838	775	1305	2098	71	255	371
9	368	1368	315	960	1242	2058	88	777
10	333	403	1276	71	2084	1096	1080	851
11	2550	1058	3	70	668	1245	410	1190
12	668	225	71	2038	923	1332	1432	505
13	316	1251	2058	1390	774	967	70	368
14	3	81	1124	844	387	2507	1092	1156
15	2526	958	626	1290	3	68	1326	397
16	1299	1264	72	2443	3	401	667	1045
17	564	1290	985	69	545	187	1228	2326
18	2422	1329	71	707	1035	1213	414	3
19	1318	2559	1298	390	946	3	611	69
20	1470	71	855	261	777	369	1292	2099

Table 4.1: Cluster sizes through multiple iterations using random starting points.

Note: Cluster number is arbitrary and just the order in the output; they do not have equivalent characteristics.

The results from this analysis (Table 4.1) showed that the choice of using random seed points was problematic. Eight out of the 20 times the classification was re-run, there was a cluster created which contained only three cases. This is a localised solution,

which limits the quality of the classification as this cluster tells us very little of what is happening in the mortality patterns throughout England and Wales. This cluster was always made up of the same areas (parts of Wandsworth, Milton Keynes and Wrexham). These cases seemed to be outliers on variables 32 ('Chronic Ischaemic Heart Disease') and 44 ('Pneumonia'), though only the former was greater than one standard deviation above the rest of the data. They also had lower than average values for the majority of cancer related variables. Although these would appear statistically improbable (the random starting point being one of these areas out of 7194 areas eight times), it is likely that the iterative refinement of clusters caused cases to be moved elsewhere, only leaving these outliers together.

This issue was compounded by the fact that there was always a large cluster (usually over 2500 cases, much greater than the other clusters and what would be expected). It is likely that since these outliers form their own cluster, a lot of variation is lost from the larger cluster which otherwise would have been split up. Nonetheless in the other classifications that did not contain the localised solution, there was still a large cluster (above 2000 cases). There was always a small cluster with around 70 cases in it which may also contribute to this problem.

This use of random seed points appears problematic. The approach produces an unstable and inconsistent set of results. The classification should be the product of the dominant patterns in the data (Milligan, 1980), rather than a product of small methodological choices. The solutions offered are not very even as well, showing large differences in the range of cluster sizes. This is problematic as ideally, the classification should have evenly sized clusters which avoid possible local solutions or large clusters which are less distinct. Therefore it should not be used.

4.2.2 Hierarchical cluster analysis

The alternative approach from Section 3.8.2 was to use a hierarchical methodology. An agglomerative methodological approach was used instead of a divisive one, since the latter is mostly used when binary data is involved and all the variables here are continuous (Everitt et al., 2001; Gordon, 1999). Within this cluster, the main algorithms have been briefly described in Section 3.5.1. All these methods have different uses and applications, though as Blashfield (1976) notes, Ward's method

appears most effective when using interval or ratio data (like in this study). It is also the most commonly used method (Everitt et al., 2001) and therefore this was used. All other factors were kept as similar as to the process described in Chapter 3 for the k-means approach where possible (for example Euclidean distance was selected here) to maintain a fairer starting point.

Cluster	Size
1	931
2	1005
3	681
4	1035
5	746
6	905
7	615
8	1286

Table 4.2: Differences in cluster sizes through running a hierarchical cluster analysis.

This method, however, cannot be assessed in the same way. As each step of the clustering process occurs by joining together the closest cases or cluster of cases, this approach will always happen in the same order since the values used to measure distance (or similarity) will always be the same (Everitt et al., 2001; Gordon, 1999). Therefore the initial cluster centres will not vary. This does not mean that a local solution cannot be achieved, however as Table 4.2 shows, this was not the case. This method is the better option for the construction of the classification.

4.3 How many clusters?

Unlike other classificatory methods, a partitional approach requires the selection of the number of clusters to be nominated before the analysis takes place. Therefore it is important to ground this choice through testing a range of solutions to assess which is most useful in this study. However the findings used to inform this decision may be slightly misleading as neither potential solution has been optimised (Everitt et al., 2001; Milligan and Cooper, 1985). Instead they provide an indication, rather than the exact correct figure, to make an informed decision on which will be the most useful direction for this data set.

To decide which number of clusters is optimal for the classification, a k-means cluster analysis will be run on a series of possible solutions between two and 16 (Section 3.8.4 provides a discussion for this choice). It is expected that the true solution will lie between these limits, as well as being effective with dissemination of results. With the data gathered from the solutions, each can be assessed through a series of different measures to see which is the most appropriate (also detailed in Section 3.8.4).

The measures are designed in the way that they naturally improve as the number of clusters increases. Therefore the best solution mathematically will always be a 16 cluster solution. However assessing the measures purely based on the metric's value is not effective for building a useful classification. Rather it is important to select the solution which balances the extra detail offered by a larger number of clusters and the simplification of patterns gained from a smaller number. Since the point of classifying a data set is to simplify patterns to those which dominate the data (Everitt et al., 2001), the classification needs to offer depth, without being too confusing.

To be able to apply this understanding, the results for each measure are plotted graphically (Figures 4.1 to 4.4), in an approach similar to Scree plots (Duda and Hart, 1973). This improves our visual interpretation of the data. To be able to assess what is efficient for capturing the patterns in the data, the focus is in the change in the gradient of the metric. Where there is a large improvement in the model, which is followed by a flattening of the trend (or just smaller changes), this would show that further improvements in the number of clusters are not adding much more understanding (Milligan and Cooper, 1985). Where this trend of a large improvement in the metric ends, this would suggest that this solution is useful for efficiently capturing patterns in the data.

Whilst this approach may be argued as being subjective since it involves user interpretation of changes in the gradient of the graph, as Milligan and Cooper (1985) argue there is no theoretical justification (or literary comparisons) for evaluating the metric value otherwise. This is especially the case when comparing between different measures, each capturing various parts of the cluster analysis process (Everitt et al., 2001). Every classification is different in terms of parameters and data, making this the most useful.

4.3.1 Analysis

The first measure assessed is the mean Euclidean distance of each point to its cluster centre (Figure 4.1). The measure shows how compact (or similar) the data in the cluster solutions are, by showing the average dispersal of points (Everitt et al., 2001). Larger values will show clusters where points lay further from the centres of the clusters. With the interpretation of clusters derived from their cluster centres (i.e. average characteristics of variables for a cluster), these reflect clusters which are less accurate of the data contained within them as many points are not representative of the cluster centres. Therefore a minimal score is favourable here.

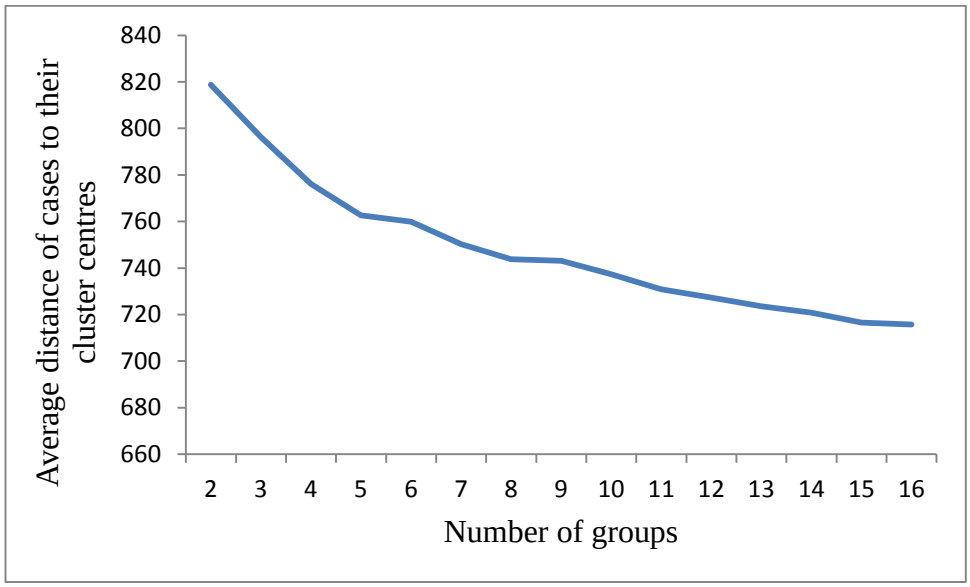


Figure 4.1: Differences in the mean similarity of cases to their respective clusters.

Beginning at a two cluster solution, the graph shows a rapid decline (i.e. improvement) with each increase in the number of solutions used. This continues until the fifth solution where there is a 'kink' in the graph showing a flattening of the curve, suggesting fewer gains from solutions. The line then falls again after the sixth solution (albeit at a slower rate), flattening once again at the eighth solution. The graph begins to fall at an even lower rate after this, however there is no noticeable kink in the curve for the rest of the possible solutions. The line begins to level off at the fifteenth solution, suggesting that further increases do not gain much in improving the model (supporting the decision not to include any more solutions). In summary, the graph would suggest either a five or eight cluster solution as the rate of improvement gained from increasing the number of solutions noticeably changes (i.e. slows down) after each.

Ideally a useful classification would produce clusters that are fairly evenly sized (Everitt et al., 2001; Vickers, 2006). This would indicate a solution which has been not influenced by outliers. For example for a two cluster solution, if the clustering process decides to make one cluster of just a few data points and the rest as one huge cluster this tells us very little, as most of the variation in the data is lost in the large cluster. Having a solution which captures the dominant patterns in the data is ideal, so that the underlying data structure is not lost or mis-specified. Figure 4.2 examines this issue through showing the mean difference in cluster sizes from what would be expected if each cluster was evenly sized (i.e. n/k).

The first two solutions (two and three) are much higher than the rest showing these not to be effective choices. However there is little to distinguish between the rest of the solutions which little difference between them. With the added complexity that comes from having increased solutions when interpreting the results, this may suggest that a smaller cluster solution may just as useful for capturing patterns in the data. The findings from this measure remain mostly inconclusive.

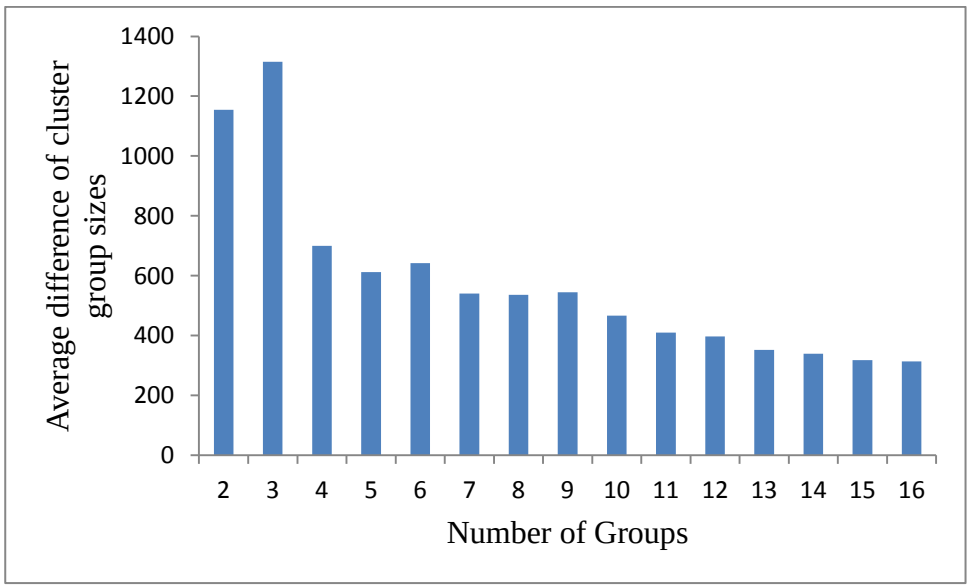


Figure 4.2: The mean difference in cluster size for different numbers of clusters.

A similar issue that became apparent when conducting this analysis concerns the creation of local solutions. The probability of this occurring increases as the number of partitions in the data also increases. This was seen when calculating the measure used in Figure 4.2, as each cluster after the eighth solution contained a cluster with a single digit in total membership. These small clusters could easily represent outliers,

obscuring more important patterns. This would suggest that the eighth solution may be the largest number of clusters which are effective for this measure.

The next two tests (detailed in Section 3.8.4) give more formal statistics for evaluating the most useful cluster solution to use. Figure 4.3 shows the results of running the C-index test (Hubert and Levin, 1976). It examines the homogeneity and compactness of clusters, with a lower value showing a better model. The resulting curve is smoother than Figure 4.1, with there being several distinct 'kinks' in the graph. These are found at solutions three, five, eight and ten. The graph evens off after the tenth solution, suggesting very little gains in having more clusters.

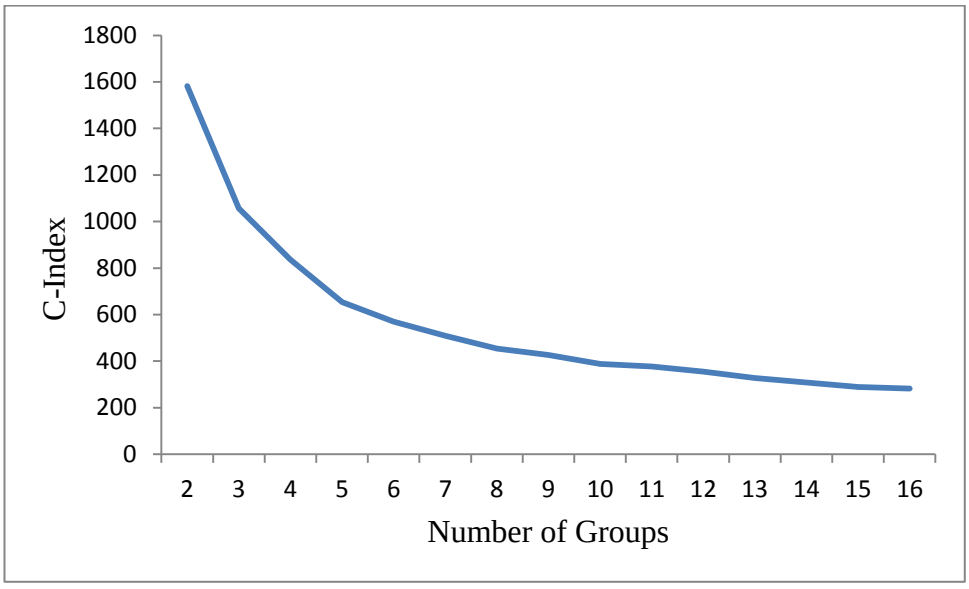


Figure 4.3: The variation of the C-index for different cluster solutions.

The final statistic applied is the Davies and Bouldin (1979) validity index (Figure 4.4). It analyses the ratio of within cluster distance to the between cluster distance (Section 3.8.4). A lower value represents a better solution, with compact and distinct clusters.

The graph shows a slight rising trend at the early values. However the values are quite unstable at first, before levelling off. This small difference in the later values would suggest little improvement through increasing the number of clusters. A two cluster solution performs the most effective on the metric. Interestingly, a five cluster solution performs the worst on this statistic which contrasts with previous variables which have shown it to be favourable. The graph would suggest a six cluster solution may be the most useful, since although it is not the lowest it is much lower than most other values and also gains more information from added number of clusters than the other low values. However it should be noted that due to how the statistic is calculated, the lowest

value is not always the most useful solution, as other solutions may reveal more information (Davies and Bouldin, 1979; Milligan and Cooper, 1985).

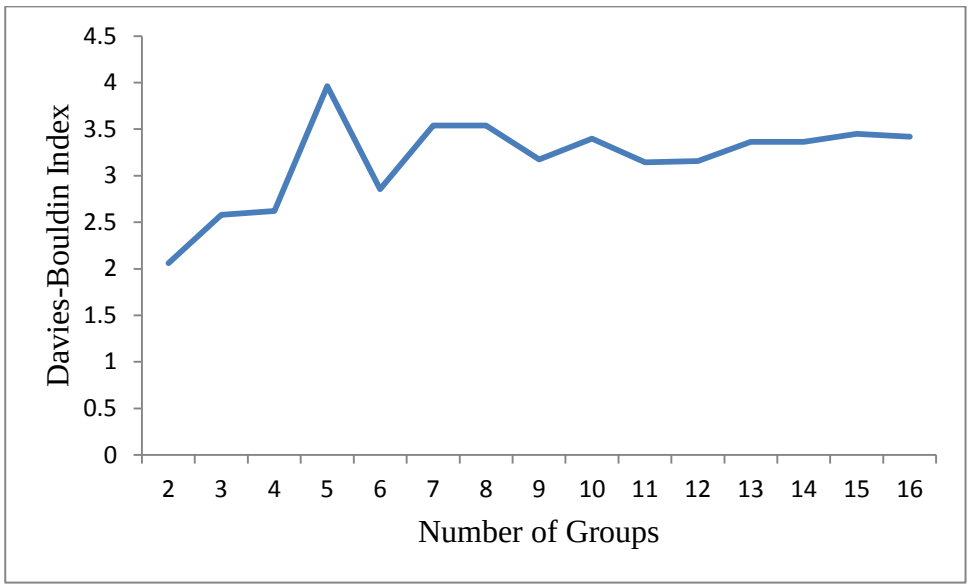


Figure 4.4: The change in the Davies-Bouldin validity index across multiple cluster solutions.

4.3.2 Which solution?

As Everitt et al. (2001) note, it is not effective to rely on a single statistic for deciding on what number of clusters to use. It is an important choice that will determine the structure of the classification. The solution needs to perform well across these four measures, with each capturing a different aspect of the methodology. Comparing the performance of various solutions will inform which is most useful in this study.

For most of the measures, the highest number of clusters consistently performed best on the metric scale. However they may be less useful for future analyses, as patterns and relationships may become too confusing to find. This is especially the case where the gains in extra information provided by extra clusters do not always present a better model across the different measures. Having a large number of clusters may end up hiding more than it reveals (Gordon, 1999; Vickers, 2006). Therefore a balancing act is needed, to choose the most efficient number.

An eight cluster solution has been chosen for the model. The solution performs well across all measures and it is this consistency which shows that it is the most useful application. In both Figures 4.1 and 4.3, the solution is the last 'kink' in the graph, with

the curve starting the slightly level off afterwards suggesting limited gains in other solutions. Whilst it did not stand out in Figure 4.2, it does not perform poorly. It does however not perform well through using the Davies-Bouldin index (Figure 4.4). Higher solutions than this were shown across the various measures not to add much more detail, showing them to be less useful through their added complexity of larger number of clusters (reducing our simplification of the data).

A five cluster solution also performs well however this was not chosen. This is shown in both Figures 4.1 and 4.3, where there is still considerable information gained by having a few more clusters after it. It also does not perform well for the Davies-Bouldin index, showing it to be less consistent than an eight cluster solution.

4.4 Calculating the cluster centres

There are two commonly used approaches for calculating cluster centres in the k-means methodology. Normally in the methodology, once cases have been split into their initial clusters, cluster centres are derived (see Section 3.6.1; Duda and Hart, 1973; Gordon, 1999; Everitt et al., 2001). Cases are then moved into the other clusters and these moves are assessed to see if they improve the model. If there is an improvement, then the move is made permanent, else it remains where it was. After all cases have been moved and checked, then the cluster centres are re-calculated and the process is repeated.

Running means present an alternative technique to re-calculating cluster centres. Instead of waiting until all moves have been completed, running means re-calculate the centres after each move has been made (Everitt et al., 2001). This may produce a more accurate decision for moving cases, as the cluster centres are constantly updated throughout the process. However it is sensitive to the ordering of cases, since the means are updated after each move.

A cluster analysis was run using both options to allow a brief comparison of what is the most effective approach for this study. These classifications were compared using two measures used in Section 4.3; average distance of cases to their respective cluster centres and the mean difference in cluster sizes compared to if they were evenly sized. The use of running means gave a classification which was both less compact (average distance was 742, being 739 when the standard algorithm was used) and less evenly

sized (a mean difference of 447, compared to 367 when not used). The use of running means will not be applied for this study, given that its performance was less suited to the aims of the methodology.

4.5 An area classification is born

With all decisions about the clustering process made, the classification model was run. First some initial details concerning the model are provided, to help inform subsequent analyses and observations in the testing process. The initial compilation of the classification iterated 67 times to a convergence criterion of 0, before halting at its final solution where no changes could be made to further improve the model. Eight clusters were produced, with varying sizes (Table 4.3). The average distance of cases from their cluster centres was 738.7 (Euclidean distance), with the average difference in cluster sizes being 367.25.

Cluster	N
1	1562
2	1149
3	1309
4	854
5	656
6	296
7	322
8	1046
Total	7149

Table 4.3: The number of areas found within each cluster.

It is not efficient to begin by interpreting the results. The testing process may end up flagging something as problematic and if this requires modification, the time spent interpreting becomes wasted. Instead the rest of this chapter will focus on statistically testing the classification to assess whether what is produced is stable and hence useful. If these clusters in the data are found to be stable through varying scenarios, then the existence of this underlying structure in the data can be said to useful and less likely to be due to random chance (Everitt et al., 2001; Gordon, 1999). Once this has been justified (or improvements made), then the classification can be explored.

4.6 Replication analysis

Replication analysis is important to be conducted when compiling a classification, because it provides confidence that the clusters exist and the results can be generalised (or at least moving towards this), a key issue in statistics in general (Milligan, 1996). Through randomly splitting the data set, we would expect that if the clusters are distinct then under the same rules and parameters, we would not get inherently different results. As such, we can infer stability in the classification.

The following procedure was followed (in accordance with Blashfield and Macintyre's (1980) split sample method) to test the classification:

1. The data set was split in half, with 3597 cases being randomly selected to form the first sample and the rest being the second sample.
2. A k-means cluster analysis was conducted on the first sample of data. The method kept all the same rules and parameters that were used to compile the main classification (for example an eight cluster solution was used, a hierarchical methodology was used to find the seed points, Euclidean distance was used etc).
3. The final cluster centres of the cluster analysis were then taken and applied to the second sample, assigning each point to its nearest cluster (using the 'classify only' option in PASW).
4. Step two was then repeated using the second data sample instead.
5. The output from steps three and four were then compared.

An issue with this approach is that randomly splitting the data in half may lead to some bias in the data found in either subset of data. It is possible (albeit unlikely) that the random selection of data points places one cluster only within one sample. Although extreme, it highlights the uneven selection of data points which may falsely show the cluster structure to be chaotic and not replicable, when in fact comparisons are not fair. Therefore the first part of Blashfield and Macintyre's (1980) procedures has been modified to solve this. Random sampling was performed within each cluster, to give a fairer set of two samples (i.e. random stratified sampling).

When comparing the two solutions for the second sample, each classification produced a similar level of compactness, with the average distance of cases to their cluster centres being 739.6 (step 3) and 739.8 (step 4). These were not much different from the value for the main classification (738.7). There were differences in the range of cluster sizes, with the third step producing less even clusters (the average difference of cluster sizes from what would be expected if they were completely even was 188.1 at step three and 162.6 at step four). Yet it would be expected that the solution from the third step may produce less even clusters as it has not been iteratively refined. Comparing these results to the value for the main classification (which when halved to account for the fact that there are twice as many cases is 183.6) shows neither of these values to be too worrying. There were no clusters of particularly small size.

Exploring the cluster centres for each variable across the eight clusters allows an understanding of the clusters produced. However, since a full interpretation of the main classification has yet to be conducted, comparisons are only limited. The clusters created through the replication analysis were not too dissimilar to those of the main classification (results not shown), showing the clusters to be distinct in the data. There was little difference between the results of step three and four in the replication analysis. Although they are not exactly the same, this is not what the replication analysis seeks. With the overall patterns generally similar, this shows stability of the overall patterns/clusters found which is important in showing that the results are useful.

Table 4.4 shows a cross-tabulation of cluster membership for the cases from the second sample for the two classifications. With the profiles of the clusters produced being similar, the numbers represent equivalent clusters to be able to examine stability of the results. The majority of cases are transferred to the equivalent cluster, suggesting the same core underlying relationships/patterns exist in the data. Where there are higher numbers of data moving differently, this is likely to represent those cases which lie at the edges of clusters. This is not necessarily bad, as small differences in the clusters produced can affect cluster membership (Everitt et al., 2001; Gordon, 1999). These movements are always to other clusters with similar mortality profiles, highlighting this, showing stability in the patterns observed (Blashfield and Macintyre, 1980). Overall, stability and replication can be inferred from these results.

		Step 4 Clusters								Total
		1	2	3	4	5	6	7	8	
	1	587	145	77	0	0	0	0	0	809
	2	0	461	82	50	59	0	0	0	652
	3	0	0	289	7	163	52	0	0	511
Step 3	4	45	70	101	363	0	0	0	0	579
Clusters	5	0	2	0	57	222	44	0	105	430
	6	0	0	0	0	3	221	4	48	276
	7	0	0	1	112	2	7	117	11	250
	8	0	0	0	0	0	1	55	34	90
Total		632	678	550	589	449	325	176	198	3597

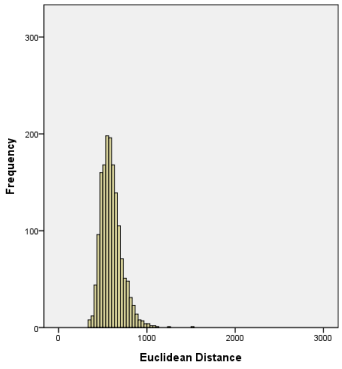
Table 4.4: A cross-tabulation of cluster membership in the replication analysis.

4.7 Impact of outliers

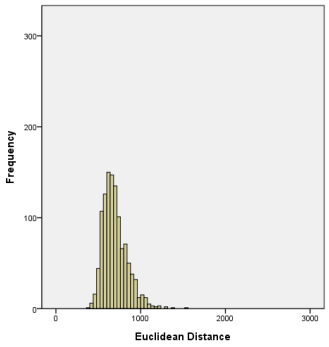
The next testing procedure is to examine the impact of extreme data points (i.e. areas) on the classification (see Section 3.9.2 for a discussion). This is important for evaluating stability of clusters to see if they are influenced by extreme data values. These affect the results by creating cluster profiles which do not reflect the true structure of the data, by skewing the results.

Cluster	1 S.D.	3 S.D.
1	1.0	0.1
2	4.7	0.1
3	2.4	0.0
4	15.7	0.6
5	18.8	0.3
6	63.5	11.5
7	66.1	17.7
8	10.8	1.1
Total	12.1	1.5

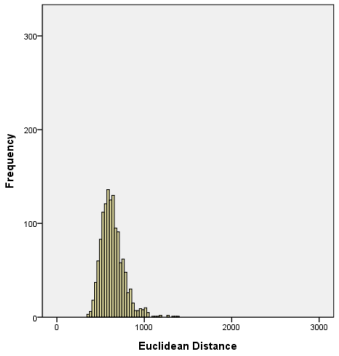
Table 4.5: The percentage of areas identified as outliers through using one and three standard deviations from the mean as the threshold value.



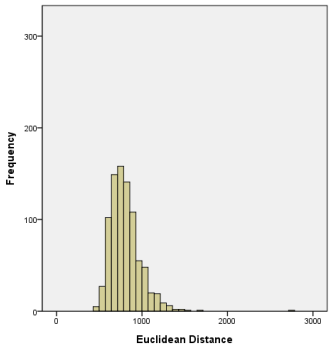
Cluster 1



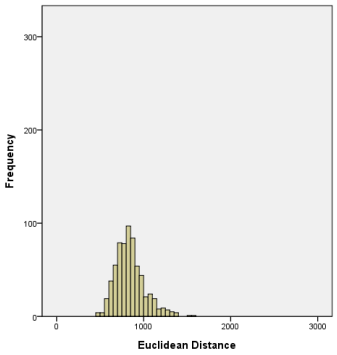
Cluster 2



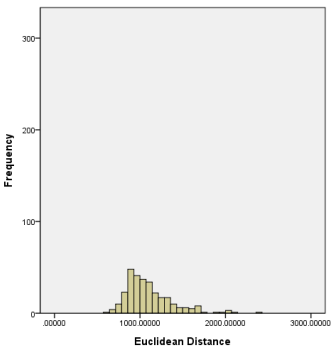
Cluster 3



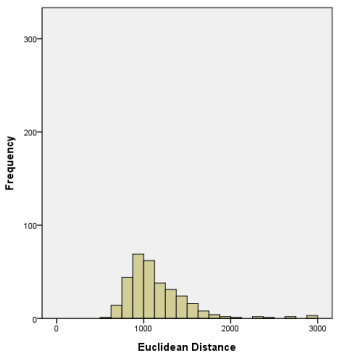
Cluster 4



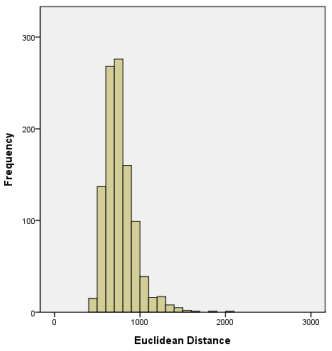
Cluster 5



Cluster 6



Cluster 7



Cluster 8

Figure 4.5: The distribution of distance of data points for each cluster.

Table 4.5 shows the percentage of 'outliers' as defined in Section 3.9.2 using both Cheng and Milligan (1996) test (±1 standard deviations) and a more conventional

schema (±3 standard deviations), split by cluster. Cheng and Milligan's criteria showed 12.1 per cent of areas reflecting outliers, which would indicate an issue in the data. However data which is normally distributed would be expected to have 32 per cent of its values lying one standard deviation further from the mean (Rogerson, 2006). Plotting a histogram for each cluster of distance of each case to its cluster centre shows that the majority of the clusters appear normally distributed (see Figure 4.5). The two smallest clusters (six and seven) are, however, slightly positively skewed. Therefore the result would suggest that the clusters are fairly compact. This is shown with the more stringent value of using three standard deviations beyond the mean showing only 1.5 per cent of areas as outliers.

The number of outliers varied by the clusters as well. Clusters six and seven have higher proportions of areas that could be considered as outliers by either measure. This appears to be related to the size of the clusters (see Table 4.3), where smaller clusters have higher proportions. This shows that these clusters are less compact, with the areas contained within less similar. However further explanation can only be gathered once the classification has been fully interpreted, to understand whether this affects stability (Section 5.1.1).

Although this testing procedure may appear problematic, Cheng and Milligan's approach examines the stability of the model. If the classification was good, then the removal of these points should have little effect (even if they are not true 'outliers'). By using their criteria and hence removing more points (than compared to using three standard deviations measure), it also provides a more rigorous test for the classification. The outliers were removed from the data set and the analysis was re-run with the same parameters and rules.

There were few changes in the overall cluster profiles captured by the classification, indicating stability through little effect of the outliers. This allowed the examination of change in equivalent cluster membership (Table 4.6). The results show a largely stable classification, with 4554 cases (72 percent) remaining in their 'equivalent' cluster after the outliers were removed. This was fairly consistent between the clusters, with all retained the majority of their data. Although the two smaller clusters (six and seven) retained the highest proportions, they experience the highest number of cases removed, leaving just areas which are similar.

		"Outliers" removed								Total
		1	2	3	4	5	6	7	8	
All cases	1	1145	103	156	0	0	0	0	143	1547
	2	0	818	1	123	142	0	2	9	1095
	3	0	29	890	233	0	0	0	125	1277
	4	0	0	0	454	52	151	63	0	720
	5	0	0	0	0	384	123	26	0	533
	6	0	0	0	0	0	104	4	0	108
	7	0	0	0	0	0	0	109	0	109
	8	0	2	0	56	0	0	225	650	933
Total		1145	952	1047	866	578	378	429	927	6322

Table 4.6: Change in equivalent cluster membership when the 'outliers' (one standard deviation above the mean) are removed from the analysis.

Where there were movements to a different cluster, these were mostly between clusters with similar mortality profiles. This is less problematic, as it will capture those cases which lie close to the boundaries of the clusters and therefore slight changes to the clustering process lead to them being classified differently (Everitt et al., 2001). The underling pattern of mortality is still captured as patterns remain fairly stable. Cluster eight is of particular interest as it loses and gains a fairly high share of areas for many different clusters suggesting a mixture of experiences. Further interpretation will have to consider this carefully.

The cases which were removed could be argued not to be actual outliers. Therefore the test was also re-run with any cases falling three standard deviations from the mean removed instead. The results are shown in Table 4.7, showing a more stable model, with fewer changes. This shows a set of more distinct clusters, although with much fewer data points excluded from the analysis this result was always likely. Overall, the findings of this testing procedure have shown a stable classification, with extreme values that have little effect on the clusters produced.

With the outliers identified and their impact on the model described, the question that arises is should these data points be removed from the analysis? The answer is probably not. Testing the impact of their removal shows that the model remains relatively stable with its clusters. Although there is an improvement on two measures, this should be expected since you are removing extreme data points which would otherwise skew these

measures. Leaving them out may also result in lost information or interesting patterns lost. This is important, as it is questionable whether these areas can be defined as outliers themselves. Furthermore Cheng and Milligan's study is more relevant to studies which use sample-based data sets. This study uses data for all areas in England and Wales, rather than a sample. It is important to keep all areas to show what is happening, to remain representative. It would be a poor classification of England and Wales if some areas were missing!

		Outliers removed								Total
		1	2	3	4	5	6	7	8	
All Cases	1	1428	112	0	0	0	0	0	21	1561
	2	1	819	88	3	232	0	0	5	1148
	3	73	37	1142	3	0	0	0	54	1309
	4	0	0	25	680	102	7	32	3	849
	5	0	0	0	7	464	171	12	0	654
	6	0	0	0	7	0	202	53	0	262
	7	0	0	0	0	0	0	265	0	265
	8	0	14	2	6	25	0	50	938	1035
Total		1502	982	1257	706	823	380	412	1021	7083

Table 4.7: Change in cluster membership when the 'outliers' (three standard deviations above the mean) are removed from the analysis.

4.8 Variable sensitivity

With 67 variables included in the analysis, it is important to assess the impact of each. This is not just important for testing stability of the classification, but also shows what is driving the formation of the classification. The process is detailed in Section 3.9.3, but essentially involves the removal of each variable individually and then assessing the difference to the main classification. Variables can be identified on the x-axis by their number from Tables 3.6 to 3.11.

Change in the average distance of cases to their respective cluster centre after the removal of a particular variable was first evaluated (Figure 4.6). The expected value is presented by the line, which represents the mean value for the main classification multiplied by 0.985 (i.e. 66/67). This accounts for how Euclidean distance is

calculated, since the other results all contain one variable less. It assumes each variable contributes an even amount to the classification.

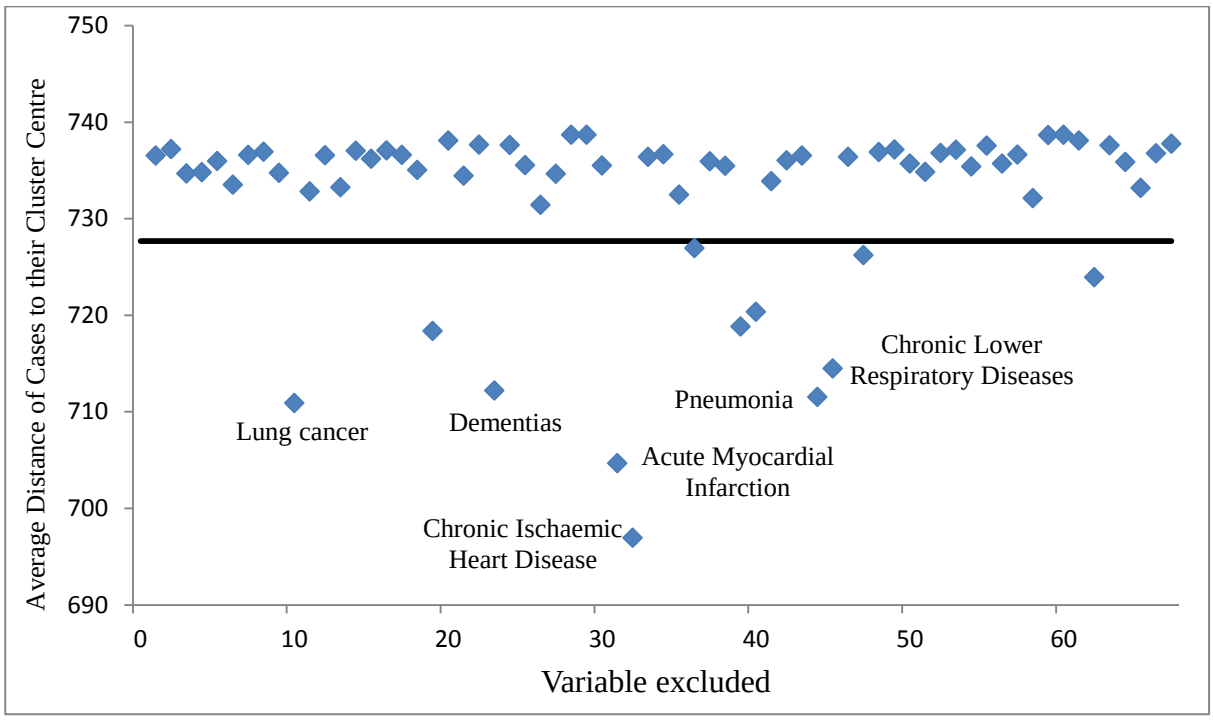


Figure 4.6: Changes in the average distance of cases from their cluster centres when a variable is removed from the classification.

The majority of variables, when removed, lead to an increase in the mean distance of cases to their cluster centres. Leaving them out of the classification leads to the formation of less compact clusters than what would be expected, showing that their inclusion is useful in adding value to the classification. None of these values were higher than the mean distance for the main classification (738.7).

There are fewer variables (12) whose exclusion from the classification leads to a more compact set of clusters. The variable names for those furthest from the expected value are shown, however there is no discernible pattern by cause of death. They are important prevalent causes of death across a range of different types of diseases. Whilst they are having a large negative impact on the model, their exclusion from the main classification would not be useful. These are key prevalent cause and ignoring them would limit the observations found. It would lead to a constrained and restricted view on the underlying patterns of mortality in England and Wales, representing a reality which does not exist.

An explanation for the patterns presented concerns the use of weightings. As Euclidean distance is calculated by adding up the differences between the value for each variable for an area and the centre of the cluster it belongs to (i.e. the average of all cases), weighting has an important effect on impacting on this difference. Higher weights bring a larger range of possible values, as the distribution becomes stretched out through multiplying the values. Removal of variables with larger weights will naturally decrease this value, leading to more 'compact' clusters.

Those variables having the largest impact through this test have the highest weightings (*c.f.* Tables 3.6 to 3.11). Calculating the correlation between the weight of the variable removed and the mean distance of cases to their cluster centre (after the removal of that variable) was -.957 ($p<.001$), showing a strong and significant association. Where the size of the weighting increased, the mean distance value was lower (i.e. having a larger change compared to the value for the main classification).

An issue that arises is that it is difficult to infer stability of the classification and the true impact from this assessment. For example, the removal of 'Chronic Ischaemic Heart Disease' leads to the average distance of cases from their respective cluster centre declining by a value of 5.7 per cent of the value for the main classification. Although Figure 4.6 shows that the variable leads to the largest change, the difference is only fairly small and hence evaluating the sensitivity analysis is problematic.

Although the approach was useful in Vickers (2006), it has been shown to be less here. This is because the variables used in Vickers classification were all evenly weighted and therefore this is not a fair form of testing. For this classification, the variables were weighted differently to account for their different importance, as measured through prevalence. As such, variables are not comparable here, given that they do not have the same weightings. Rather a new approach must be taken to solve this issue, one that has not been applied before in the literature. The size of the individual weightings must be accounted for. This will show whether particular variables are having a larger or smaller impact than would be expected.

To calculate this, the expected change needs to be re-estimated. Instead of just comparing to a standard value, the mean distance of cases to their cluster centres for the main classification is multiplied by the weighting for each variable. This shows the true expected change in the mean distance, since weighting has been shown to be important for determining the Euclidean value. To evaluate the performance and impact of each

variable once removed from the classification, the relative change can be calculated through dividing the actual change in Euclidean distance (from the value for the main classification) by the expected change. Figure 4.7 shows the results from this standardised measure.

Overall, the results show a more stable classification. For the majority of the variables, their removal does not have a large relative change on the model results. The recorded actual change is consistently lower than what we would have expected from their weighting. However there are some variables which have led to a greater change in mean distance than would be expected (i.e. a value over one).

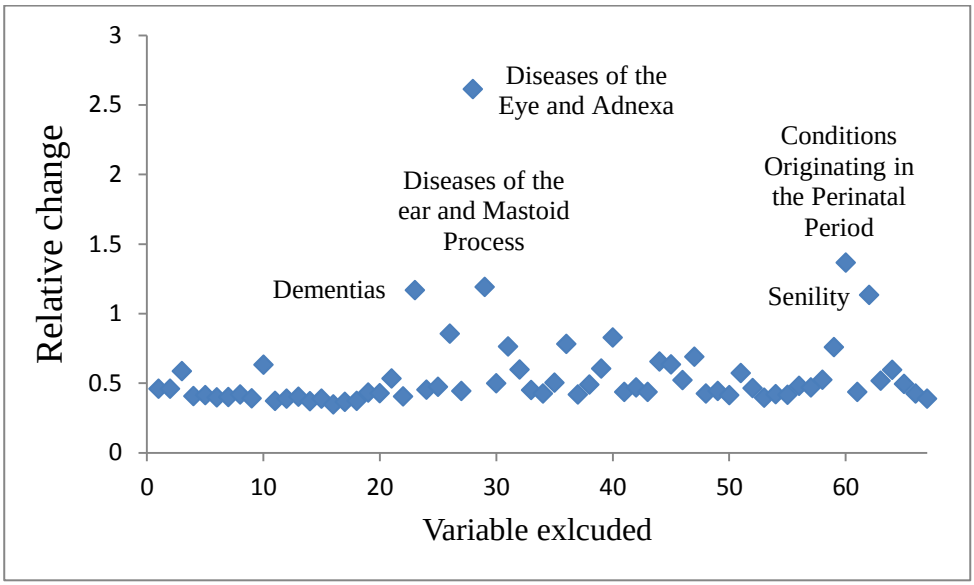


Figure 4.7: Relative change in average unweighted distance of cases to their cluster centres once variables are removed from the classification.

Three of these variables ('Diseases of the Eye and Adnexa', 'Diseases of the Ear and Mastoid Process', 'Conditions Originating in the Perinatal Period') contain low prevalence levels. These are problematic and their low levels of deaths captured would suggest that they adding little to the classification. These are the three smallest variables used in the study, accounting for 0.008 percent (167) of all deaths in the study. With such small changes, their inclusion in the classification could be questioned.

With this study being an initial investigation taking a different approach that has not been applied before, it is argued that including all deaths would useful at first. This would create a complete classification of all areas and all possible causes. Having all deaths covered would allow for the most variation to be captured by the classification. However this test has shown that it may be beneficial to drop these variables which

cover such few deaths. More is not always better, especially since they could be restricting or masking other effects (i.e. adding noise to the model).

Vickers (2006) offers an explanation of why having such small variables is unhelpful when compiling a classification. For variables to be useful and add to the classification, there needs to be enough deaths to allow for variations to occur between the 7194 MSOAs used in the study. Applying an eight cluster classification for just these three variables combined (into one variable) highlights this problem. With 167 deaths in the three variables in question, only this information can be classified. Thus would result in seven clusters made up by dividing up these 167 deaths. The final cluster will represent areas with no deaths in them, since with this approach you cannot decide how to divide areas with nothing in them.

To avoid the classification containing variables that add little value to the classification, it would be useful to drop these variables from the analysis. Another variable ('Causes Related to Pregnancy and Childbirth') contains a similar level of deaths (n=160) to these variables and was included here as well for the same reason. Exploring the output produced from running the classification without these variables shows absolutely no change in the classification (results not shown). There is no change in cluster membership and the cluster profiles remained identical. This highlights the lack of information these variables have added to the model.

The other two variables shown to be problematic in Figure 4.7 are 'Dementias' and 'Senility'. However both of these causes are fairly prevalent (3.08 per cent and 1.77 per cent of all deaths between 2006 and 2009 respectively). Dropping these would therefore not be useful for creating a representative classification. Rather what this highlights is that these two variables are driving the classification, being important in cluster formation.

Removing these from the classification and re-running the analysis, showed that the results of the main classification were stable. There is not much difference in the mortality profiles of each cluster, with the same overall characteristics prevailing. There is some change in cluster membership, however this is only involves small numbers, relating to those cases which lie on the edge of the clusters moving to clusters with similar profiles. Therefore whilst they are important in the analysis, the classification is not solely capturing their patterns. Rather they reflect the underlying

structure of the data being similar. Stability through variable selection can be inferred through this section.

4.9 Conclusion

This chapter has focused on the creation of a new type of classification for England and Wales, exploring the underlying structure of mortality patterns. The decisions made in creating this have been detailed, with justification of the important steps to ensure that the final product is both relevant and of high quality. These have been made to create an efficient tool in terms of the methodology employed, as well as be appropriate to achieving the aims of the study. As such, the procedures have focused on showing the classification produced to be valid given the data and objectives available, creating a useful tool.

The resulting classification was tested fully to ensure that what is presented here is useful and stable. Whilst there is no strict testing schedule or metrics to optimise, various tests showed the clusters to be stable, showing that the underlying structure of the data was captured effectively. The lack of variability in results shows that the classification is useful, warranting further investigation of what it shows in terms of mortality patterns. The testing process also helped inform the interpretation and understanding of the clusters found.

This does not spark the end of the classificatory process. So far, the emphasis has been on creating the classification. But what is next? The classification will be explored to examine what clusters have been created. This interpretation will help form further analyses to gain a better understanding of how these clusters have come to exist. Whilst the testing conducted in this chapter informs us whether the classification is statistically stable, analysing the relationships that form (and explain) the clusters further will aid in ground truthing. Having clusters which relate to the processes and patterns previously identified in the literature is paramount to assessing the usefulness of the classification.

Chapter 5: Interpreting the area classification

5.1 Introduction

This chapter seeks to improve the understanding of the clusters which constitute the area classification of mortality patterns for England and Wales. It begins through detailing the cluster centres, showing what each cluster captures with regards to the inputs to the model (i.e. the causes of mortality). There is limited literature assessing how the different causes of mortality interact together and hence the classification will provide an overarching insight into the dominant structure that exists throughout England and Wales. The area classification was then mapped to view the geographical spread of the clusters, to see if this could add to the understanding of what the clusters represent. Based upon these findings, the individual clusters were named to improve their use as a research tool. Basic statistics to evaluate the clusters based upon isolation, compactness and the impact of variables were calculated to show how useful the area classification is.

To further understand the clusters produced, other statistics were gathered to add value to the area classification. Life expectancy and premature mortality were calculated for each cluster to see if this could provide additional information not provided through the individual mortality rates for causes. A social and spatial analysis was then applied to the classification to explore the characteristics of the areas which form each cluster. The analysis focuses upon demographic, social and geographical factors. This aids the understanding of how the clusters have manifested and developed throughout England and Wales. It helps move towards ground truthing the classification through assessing their relevance based upon previous research, as well as showing its use as a research tool for further applications.

5.2 The characteristics of the clusters

Exploring the 'cluster centres' allows the examination of each cluster to be able to understand what they are showing. The cluster centres show the mean value of each input variable (i.e. cause of death) of the areas which make-up each cluster. These are presented across Tables 5.1 to 5.6, with the cluster centres split into different types of

causes, serving to highlight interaction effects between causes. The centres were also converted from their weighted scores to SMRs to aid interpretation.

Conditional formatting was applied to each variable individually (using Excel). This illustrates each variable in relation to the range of values (represented through a colour scale), where the lowest value is coloured in green and the highest is red. The rest of the values are automatically coloured proportionally, dependent upon their position in relation to the range (where yellow represents the mean). Using colours simplifies patterns, making it easier to understand, given that the human eye is excellent at pattern recognition (Dorling, 2012; Everitt et al., 2001). As such, it allows the visual inspection of the dominant patterns, relationships and interactions between the causes of mortality, within and between each cluster.

Variable	Clusters							
	1	2	3	4	5	6	7	8
Cancer of the Gullet	84	106	96	110	123	130	98	96
Stomach Cancer	75	109	91	124	146	160	125	90
Colon Cancer	93	98	98	107	104	114	105	100
Rectum Cancer	87	100	95	107	122	132	114	97
Liver Cancer	84	106	92	123	137	161	123	97
Pancreatic Cancer	96	102	99	106	107	116	98	97
Lung Cancer	68	113	83	127	173	190	124	83
Breast Cancer	95	97	100	100	101	106	115	104
Ovarian Cancer	100	97	102	96	99	93	97	102
Prostate Cancer	96	96	100	99	100	102	111	106
Kidney Cancer	89	99	95	107	111	130	112	104
Bladder Cancer	84	105	96	111	112	135	115	99
Cancer of the Brain	102	94	98	94	94	95	105	106
Leukaemia's	99	100	97	100	107	101	99	98
Other Lymphatic Cancers	98	100	99	102	100	111	101	98
Other Cancers	84	105	94	112	123	132	111	96

Table 5.1: The mean SMRs for all cancer-related mortality by cluster.

Key: Green is a low SMR, red a high SMR, yellow the mean and the other colours scaled accordingly to their respective position.

Variable	Clusters							
	1	2	3	4	5	6	7	8
Hyperintensive Diseases	84	107	91	115	119	141	131	108
Acute Myocardial Infarction	82	132	82	112	169	180	136	102
Chronic Ischaemic Heart Disease	67	85	107	148	114	169	128	96
Pulmonary Heart Disease and Diseases of Pulmonary Circulation	84	104	97	121	116	134	121	97
Atrial Fibrillation and Flutter	84	96	90	101	113	120	128	114
Heart Failure	84	100	92	103	118	138	134	111
Other Heart Diseases	92	95	102	106	103	113	103	103
Intracerebral Haemorrhage	90	99	97	112	118	127	110	99
Cerebral Infarction	76	90	89	110	113	143	148	114
Stroke	76	92	83	99	106	149	173	128
Other Cerebrovascular Diseases	70	71	82	100	95	154	210	134
Aortic Aneurysm and Dissection	86	104	97	116	114	116	111	94
Diseases of Veins, Lymphatic Vessels and Lymph Nodes, Not Elsewhere Classified	77	98	99	129	116	129	137	101
Other Circulatory Diseases	74	94	83	107	122	167	146	112

Table 5.2: Differences by cluster of average the SMR values for causes related to the digestive system.

Key: Green is a low SMR, red a high SMR, yellow the mean and the other colours scaled accordingly to their respective position.

Variable	Clusters							
	1	2	3	4	5	6	7	8
Dementias	57	60	63	87	78	170	349	159
Other Mental and Behavioural Disorders	56	96	72	138	159	217	147	88
Parkinson's Diseases	83	71	83	87	70	130	198	145
Alzheimer's	66	62	71	88	79	151	284	157
Other Diseases of the Nervous System	84	91	90	102	103	139	141	115

Table 5.3: Variations by average SMR score for mental and nervous system related causes of death across the classification.

Key: Green is a low SMR, red a high SMR, yellow the mean and the other colours scaled accordingly to their respective position.

Variable	Clusters							
	1	2	3	4	5	6	7	8
Pneumonia	72	90	85	117	112	165	167	122
Chronic Lower Respiratory Diseases	61	109	79	134	176	214	138	90
Lung Diseases due to External Agents	72	97	83	111	141	195	162	100
Other Diseases of the Respiratory System	73	91	83	107	115	160	176	120

Table 5.4: Average SMR score for respiratory causes of death by cluster.

Key: Green is a low SMR, red a high SMR, yellow the mean and the other colours scaled accordingly to their respective position.

Variable	Clusters							
	1	2	3	4	5	6	7	8
Ulcers	79	105	90	124	132	158	124	94
Vascular Disorders of the Intestine	73	101	92	124	140	160	133	98
Other Diseases of Intestines	82	100	92	118	117	135	132	105
Alcoholic Liver Disease	58	105	76	143	182	241	140	81
Other Liver Diseases	71	114	85	131	153	188	131	87
Diseases of Gallbladder, Binary Tract and Pancreas	72	104	95	124	145	173	126	96
Other Diseases of the Digestive System	81	102	88	112	125	143	135	102

Table 5.5: The cluster centres for causes related to the digestive system.

Key: Green is a low SMR, red a high SMR, yellow the mean and the other colours scaled accordingly to their respective position.

Variable	Clusters							
	1	2	3	4	5	6	7	8
Infant Mortality	75	103	85	109	123	145	110	84
Septicaemia	78	100	88	119	130	166	124	103
Other Infectious and Parasitic Diseases	83	117	90	118	139	152	128	100
Diseases of the Blood	85	110	84	108	130	143	123	111
Diabetes Mellitus	66	105	82	117	132	177	178	117
Other Endocrine, Nutritional and Metabolic Diseases	77	101	93	110	124	150	152	96
Diseases of the Skin and Subcutaneous Tissue	79	103	88	111	117	134	154	112
Diseases of the Musculoskeletal System and Connective Tissue	86	100	94	101	109	124	133	107
Renal Failure	80	104	87	105	124	147	149	114
Other Diseases of the Genitourinary System	76	97	85	117	109	145	169	118
Congenital Malformations, Deformation and Chromosomal Abnormalities	80	105	97	105	126	142	132	105
Senility	77	62	71	68	75	148	191	156
Other Symptoms, Signs and Abnormal Findings	65	106	78	136	160	197	149	81
Falls	76	92	91	128	133	178	122	96
Other Accidents	83	96	93	109	118	140	116	106
Intentional Self-Harm	86	100	94	119	117	139	101	94
Other External Causes	69	106	85	123	127	158	124	92

Table 5.6: The cluster centres for the rest of the causes that formed the classification.

Key: Green is a low SMR, red a high SMR, yellow the mean and the other colours scaled accordingly to their respective position.

The tables show varying patterns emerging throughout each cluster, with each capturing different profiles of mortality patterns. The main factor dividing England and Wales into the clusters appears to be levels of prevalence. However variations by cluster are

not just dissimilar scales of the SMR scores, rather there are differences by cause as well. For example, clusters six and seven are fairly similar, containing high mortality rates across causes. Yet the causes which dominate clusters are different.

With the sixth cluster, there are high rates for causes related to unhealthy behaviours (for example respiratory, digestive and some heart-related causes; Tables 5.4 and 5.5). The seventh cluster is not a function of slightly lower rates, being characterised by higher rates for mental and nervous system related causes (Table 5.3). There are also differences by the cancer related diseases between the clusters (Table 5.1). They are capturing different patterns and processes.

There was no change in cluster size (or group membership) of the figures reported in Table 4.3. There is an uneven spread of cluster sizes, with fairly large differences (both absolutely and relatively) in the range of values. The smaller clusters had the profiles which showed higher mortality rates and *vice versa*. Poor health appears to be more differentiated across England and Wales, with more clusters of these patterns. This is compared to the larger and more homogenous areas of good health characteristics which there are fewer clusters of.

5.2.1 Naming the clusters

With each of the clusters better understood, they can be allocated a name rather than just referring to them by individual numbers. Just as the clusters are summaries of the main mortality patterns in England and Wales, each name must articulate them well in a few words. Whilst they are subjective, they also need to be accurate (Harris et al., 2005). The names need to be effective to aid with the dissemination of the classification (as well as subsequent analyses), improving any outputs produced using it through providing a quick reference point.

Naming the clusters is a difficult (but important) task. Names need to make sure they are not the same to those used in other classifications, to avoid any issues of users assuming the clusters were capturing the same types of areas (Vickers, 2006). Ecological fallacies could also arise if the names are too detailed (not quite describing those cases located on the edges of the clusters), but at the same time they need not be too ambiguous (Harris et al., 2005). Possible stigmatisation of areas is inevitable

(Butler and Watt, 2007) and therefore needs to be minimised, especially with the low geographies involved. Areas should not also be insulted by the name attached to them.

Tables 5.1 to 5.6 provide a wealth of information for which the names are based upon. However there has been some influence through the analysis presented later in the chapter, helping to shape their names. The names are presented below, with a short summary to highlight why through their main traits:

Cluster One: "*Best Health and Most Desirable*"

- The majority of variables (57) contain the lowest value between clusters, showing that it represents good health areas.
- Two variables have slightly high scores; 'Ovarian Cancer' and 'Cancer of the Brain' (Table 5.1) although only the latter is above 100 (i.e. average).
- Within the cluster, rates are highest relatively amongst the cancers, and the lowest are found in those variables with the highest variation across the clusters (for example neurodegenerative diseases; Table 5.3).

Cluster Two: "*Average Mortality Profiles*"

- The cluster contains low to average rates, with most fluctuating around 100.
- It has the lowest rates for 'Senility' and 'Cancer of the Brain' (Tables 5.1 and 5.6), with low neurodegenerative disease both between and within (Table 5.3).
- 'Acute Myocardial Infarction' is the only variable which could be considered slightly high (Table 5.2).

Cluster Three: "Good Health Areas"

- This cluster appears to fall in-between 'Best Health and Most Desirable' and 'Average Mortality Profiles', being mostly low rates. This is its most distinctive trait.
- Ovarian cancer is the only variable which contrasts the trend, although the SMR is only 102 (Table 5.1).
- Within the cluster, heart-related causes and cancers are quite high, with low neurodegenerative diseases (Tables 5.1 to 5.3). It is similar to 'Best Health and Most Desirable', just the SMRs are slightly higher.

Cluster Four: "The Middle"

- The cluster is dominated by rates found mostly in the middle, when compared to the other clusters. These values are generally just above average.
- There are higher rates for heart-related diseases, especially 'Aortic Aneurysm and Dissection' (Table 5.2). Digestion- and accident-related variables also display slightly high rates (Tables 5.5 and 5.6).
- Some cancers and 'Senility' are quite low, as are neurodegenerative diseases (Tables 5.1, 5.3 and 5.6).

Cluster Five: "Poor Health Experiences"

- The rates in this cluster are mostly above average, with high values for respiratory- and digestive-related causes (Tables 5.4 and 5.5). Some cancers are also high, especially those linked to respiratory and digestive causes (Table 5.1). Unhealthy behaviours may be more important in explaining this cluster, an important policy consideration.
- 'Parkinson's Disease' rates are the lowest (the other neurodegenerative diseases are low when compared within clusters, but not between) and 'Cancer of the Brain' is also quite low (Tables 5.1 and 5.3).

Cluster Six: "Poorest Health and Least Desirable"

- It contains the worst health of all the clusters, containing the highest cluster centres for 40 variables.
- Only two variables fall below 100, 'Cancer of the Brain' and 'Ovarian Cancer' (Table 5.1). This contrasts with those clusters which display better mortality profiles.
- Within the cluster, cancers tend to be lower (Table 5.1), with highest rates within for respiratory, liver- and accident-related causes (Tables 5.4 to 5.6).

Cluster Seven: "Poorest Neurodegenerative Health"

- Mortality rates are consistently high rates across the majority of variables.
- A few cancer-related causes are low, however never below 97 (Table 5.1).
- Within the cluster, the rates for neurodegenerative disease are much higher than the other means (Table 5.3).

Cluster Eight: "Mixed Experiences"

- The rates mostly fall slightly lower than average, although there are many which are just above average as well, shaping a cluster with varied outcomes.
- Most cancers, as well as digestive- and heart-related causes contain the lower values (Tables 5.1, 5.2 and 5.5).
- Gender specific cancers are slightly high and 'Cancer of the Brain' is the highest value when compared to the other clusters (Table 5.1). Neurodegenerative diseases are also quite high, as is 'Senility' (Tables 5.3 and 5.6).

5.2.2 Visualising the area classification

An advantage of producing an area classification is that it allows the mapping of cluster membership, to show how the clusters vary over space. Visualising the classification brings it alive, giving meaning to it. The classification is no longer just a set of (boring) tables and text; rather you can begin to see what is happening. Quite simply put; 'a picture can say a thousand words'. It also helps improve the understanding of the classification, through illustrating potential processes which create this structure to the data. The boundaries of the GORs (Governmental Office Regions) have been overlaid to aid the interpretation of the maps, providing some reference points.

The explosion of colours which constitutes the spatial distribution of the classification does not appear to present a clear geographical pattern itself (Figure 5.1). It is difficult to see common patterns from just this; however this does not necessarily mean that no geographies are present for particular clusters (a dimension which is explored in more detail later; Section 5.6.2). Particular colours dominate the map however this is because they tend to be found in MSOAs which are larger in size. These reflect rural areas and therefore rurality is a factor which should be further examined later (see Section 5.6). Nonetheless the main urban conurbations also stand out on the map, highlighting wide variety of area types within each.

London was included to give a case study of the area classification within a city. The classification captures the East-West divide of the city (Orford et al., 2002), showing it to reflect social patterns rather than geographical variations. Whilst the more affluent West is more homogenous in its pattern of mortality profiles, the deprived East-End of London shows a wider range of experiences. This shows the importance and relevance

of the area classification within a policy targeting framework, as assuming the deprived area to be similar in health outcomes is false.

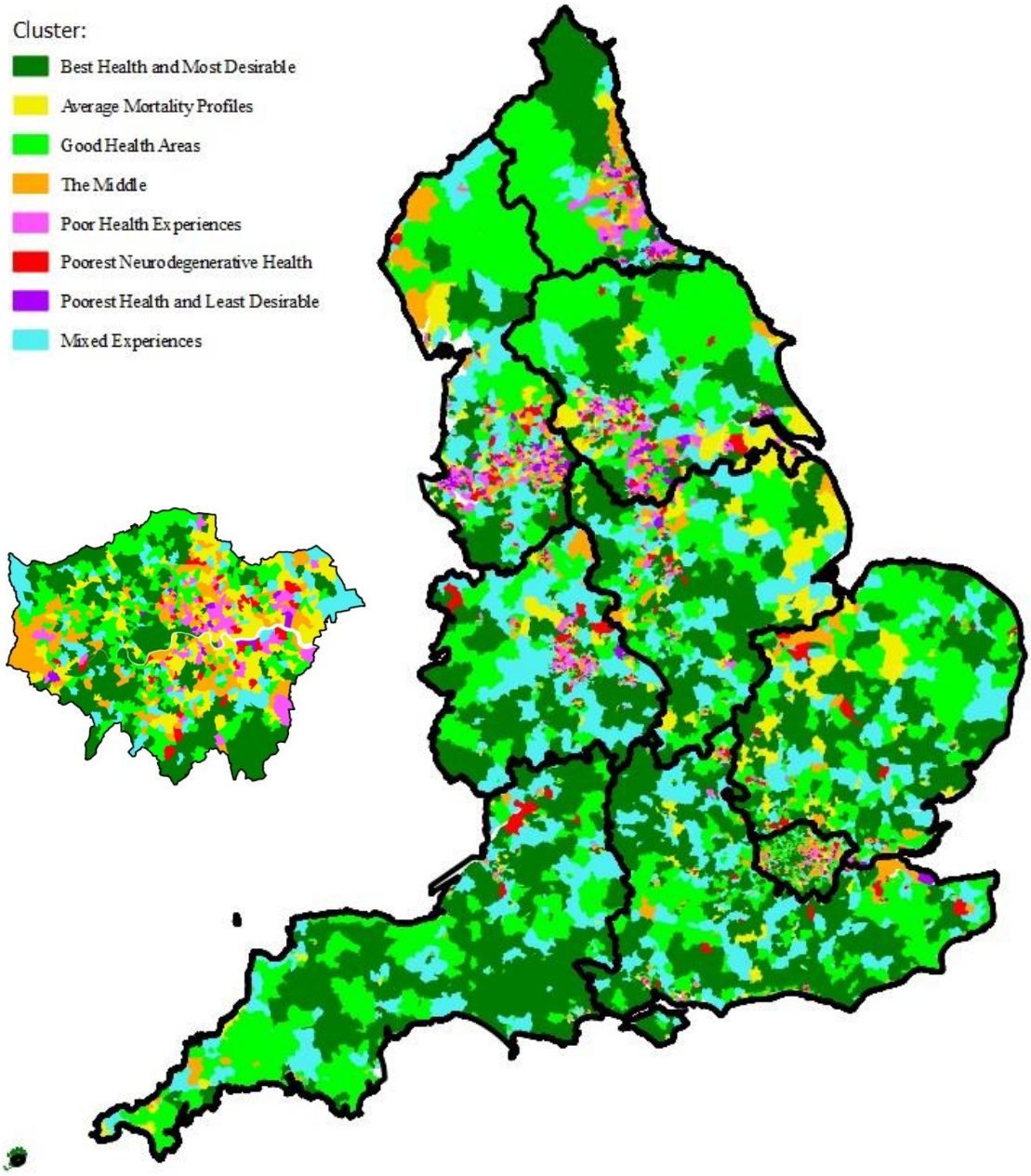


Figure 5.1: The geography of the classification

5.2.3 Compactness and isolation of clusters

A k-means cluster analysis aims to produce a set of clusters which increase the similarity of areas within each cluster, whilst at the same time maximising the dissimilarity between each cluster (Everitt et al., 2001; Gordon, 1999). This is to ensure that the clusters are distinct and therefore the classification is useful by capturing

different patterns. Applying this framework is important for evaluating an area classification, as well as contributing to its interpretation.

Examining the compactness of each cluster improves the understanding of the classification through highlighting how stable each is internally. Clusters which are homogeneous will have mortality profiles that are more representative of the areas contained within (Everitt et al., 2001). This assesses how precise Tables 5.1 to 5.6 are for detailing each cluster.

The mean Euclidean distance of cases to their respective cluster centre was calculated. This shows the similarity of areas within each cluster based upon their location to the average for all areas, across every variable. Where areas are different from the cluster centres, they will be located further from it. A greater range of area types will therefore increase this measure, showing a less compact cluster.

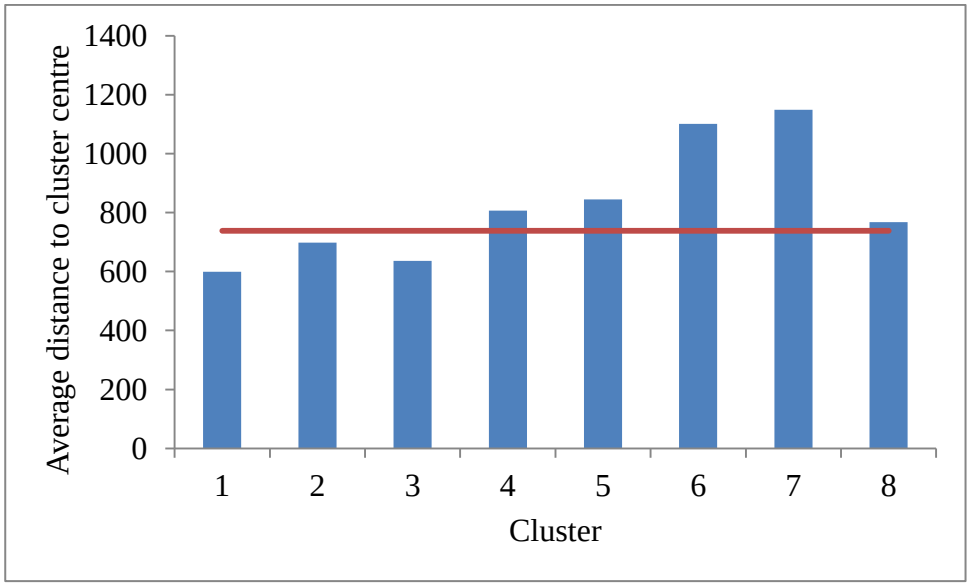


Figure 5.2: Mean distance of areas to the centre of each cluster.

Key: Clusters; 1 = Best Health and Most Desirable, 2 = Average Mortality Profiles, 3 = Good Health Areas, 4 = The Middle, 5 = Poor Health Experiences, 6 = Poorest Health and Least Desirable, 7 = Poorest Neurodegenerative Health, 8 = Mixed Experiences. The line gives the mean value for the classification as a whole.

Figure 5.2 presents a bar chart of the mean distance of areas to their cluster centre by cluster. The line included represents the average distance for all cases. All of the clusters vary from the value for this overall value. Applying confidence limits shows no overlap between any of the values (not shown as too small to see).

The clusters which were most compact were 'Best Health and Most Desirable' and 'Good Health Areas'. The clusters with the mortality profiles which display better health outcomes contain areas that are most similar to each other. This pattern is mirrored at the other end of the distribution. Those clusters with mortality profiles displaying worse health are less concise, with a greater range of experiences in the areas contained within them.

An explanation for this can be derived from exploring the mortality profiles of the clusters. The dissimilar clusters 'Poorest Health and Least Desirable' and 'Poorest Neurodegenerative Health' consist of higher mortality rates. Similarity of areas within this cluster is restricted, since all causes cannot be high. This is because if there is a higher concentration of deaths for a particular cause, then this limits the ability of other causes to also be high as well. Therefore areas within these clusters will be dissimilar to the cluster centres. It does not mean that they are wrong, as it is their extremities which makes them similar (i.e. high rates), even if the distance measure shows otherwise. The clusters with lower mortality rates are less affected by this, given that better health cannot be extreme, as the bottom end of the distribution is bounded and cut off at zero.

The pattern of similarity also appears to be related to the size of the clusters. To test for this possible effect, a Pearson's correlation was run. This will evaluate the strength of the association between these two factors and whether this is significantly different from random chance (Rogerson, 2006). The result gave a significant result ($r = -.967$, $p < .001$), showing a strong negative association. The clusters which were larger in size reported a lower mean distance of cases to their cluster centres (i.e. were more similar).

Exploring the concept of isolation focuses upon the separation of the clusters. It evaluates how dissimilar each is. If the clusters are to be distinct and different from each other, then each cluster should lie far from each other within the data. To calculate this, the Euclidean distance between each cluster (i.e. based upon their cluster centres) was computed. A larger value shows greater dissimilarity in the cluster centres between different clusters. The results are presented in Table 5.7, which was conditionally formatted to improve interpretation (see Section 5.2 for description of interpretation).

The clusters which are most similar (i.e. located nearest to each other) are 'Best Health and Most Desirable' and 'Good Health Areas'. Those clusters which displayed good health outcomes are all similar to each other. These clusters are less distinct, reflecting only slightly different traits for areas with low rates. 'Mixed Experiences' is more

similar to these types of areas, compared to the poor health clusters. Whilst it displays varying mortality rates between causes, the cluster is more representative of areas with lower mortality rates.

Cluster	Best Health and Most Desirable	Average Mortality Profiles	Good Health Areas	The Middle	Poor Health Experiences	Poorest Health and Least Desirable	Poorest Neuro-degenerative Health	Mixed Experiences
Best Health and Most Desirable		2138	1237	3756	4601	7682	6685	2958
Average Mortality Profiles			1482	1915	2523	5629	4674	2288
Good Health Areas				2532	3389	6477	5487	1996
The Middle					1771	3958	3548	2539
Poor Health Experiences						3123	4083	3377
Poorest Health and Least Desirable							3574	4816
Poorest Neuro-degenerative Health								3733
Mixed Experiences								

Table 5.7: Euclidean distance between cluster centres.

Key: Green is a low distance, red a high distance, yellow the mean and the other colours scaled accordingly to their respective position.

The most dissimilar clusters are 'Best Health and Most Desirable' and 'Poorest Health and Least Desirable', which given their mortality profiles is not surprising. Both 'Poorest Health and Least Desirable' and 'Poorest Neurodegenerative Health' are isolated from the clusters with good health characteristics, showing them to be distinct

with this regards. However these two clusters were shown not be too similar to each other. This is different when compared to the clusters with good health, which are all fairly similar to each other. The poor health clusters (including 'Poor Health Experiences') are dissimilar, showing distinct clusters that capture different patterns in the data. There is greater variation in area types for poor health patterns.

5.2.4 Impact of the variables

Calculating the range of each input variable across the cluster centres (Tables 5.1 to 5.6) allows an assessment to be made regarding the impact of the variables in the classification (Table 5.8). A large difference across a variable would show that the variable was more important in driving the clusters through offering greater variation in patterns (Everitt et al., 2001). These variables are more geographically and socially determined, being picked up by the classification through similar distributions. A lack of variation would suggest that there is less underlying structure to the data for that particular variable.

17 variables (27 per cent) had a range between the cluster centres of over 100. Those which had the largest range were neurodegenerative diseases, particularly 'Dementias' and 'Alzheimer's' (Table 5.3). This group of causes are having the largest impact upon the classification. This is in part due to their distinct geographical and social patterning (Shaw et al., 2008). However their extremities are mostly due to the values for the clusters 'Poorest Health and Least Desirable' and 'Poorest Neurodegenerative Health'.

Table 5.9 also shows patterning in the types of causes with low ranges. Eight out of the bottom ten of these were cancer related diseases. It shows that these types of cancers are less geographically and socially determined, with less distinct patterns throughout the data. Instead, they are more evenly distributed across England and Wales. This suggests that the use of the gender-related cancers and 'Cancer of the Brain' in the descriptions given of the clusters in Section 5.2.2 is less useful.

Their low values do not mean that they do not add understanding to the area classification though. They still provide additional information to the model, capturing how they vary within the clusters as well. Ignoring them would just produce an area classification representing social divisions, an artificial result that does not reflect reality. This restrained the quality of Shelton et al.'s (2006) mortality classification,

with variables selected based upon their social variations despite being less important for understanding the underlying structure of mortality patterns overall (for example water transport deaths had a fairly large impact on one cluster; see Table 2.7).

Rank	Variable	Range
1	Ovarian Cancer	9
2	Leukaemia's	10.1
3	Cancer of the Brain	12.1
4	Other Lymphatic cancers	12.5
5	Prostate Cancer	14.9
6	Pancreatic Cancer	19.5
7	Breast Cancer	20.2
8	Other Heart Diseases	21.1
9	Colon Cancer	21.2
10	Aortic Aneurysm and Dissection	30.2
...	...	...
53	Lung Disease Due to External Agents	123.3
54	Parkinson's Disease	128.1
55	Senility	128.7
56	Other Symptoms, Signs and Abnormal Findings	131.9
57	Other Cerebrovascular Diseases	139.7
58	Chronic Lower Respiratory Diseases	152.8
59	Other Mental and Behavioural Disorders	161.4
60	Alcoholic Liver Diseases	183.3
61	Alzheimer's	221.9
62	Dementias	292

Table 5.8: The ten highest and ten smallest ranges for variables in the classification.

Section 4.8 highlighted that the weightings had an effect on variable sensitivity. Therefore it would appear useful to assess whether the prevalence of the variables was associated in accounting for this patterning by the range. If it is just the more dominant causes of death affecting the classification, then this adds little to our understanding and the usefulness of the classification. A Pearson's correlation was run to assess this association between the total number of deaths by a cause and its reported range in values for the cluster centres. This resulted in a weak correlation, which was insignificant ($r = .191$, $p < .134$). Prevalence does not mask geographical variation.

Rather social and spatial factors are more important in explaining the structure of the data found by the area classification.

5.3 Health-related statistics

Assigning each individual death to its corresponding cluster allows further health-related statistics to be compiled. These extend our understanding of the classification through revealing different dimensions and characteristics of the clusters captured. A high mortality rate does not necessarily mean that the health of an area is poor. Since everyone will die, this can stop any misinterpretation of the clusters. They also offer greater comparability and applications outside of this study due to their standardised approaches (ONS, 2010b; Thomas et al., 2010). This allows the evaluation of how useful the area classification is at capturing variation in the underlying structure of health inequalities for England and Wales as well.

5.3.1 Life expectancy

Life expectancy estimates at birth were calculated using the Chiang II method (Chiang, 1972). The statistic gives the amount of life a person (controlling for age and sex) can expect to live on average if a person was born and remained in each cluster (as well as the mortality rates remaining the same). As such, the inequality in terms of life that can be expected to live (a key health dimension) can be compared between clusters. Estimates are regularly used both for compiling national health measures and within academic research (for example ONS, 2010b; Smith et al., 2010), highlighting how useful the measure is in this context. The results are presented in Table 5.9.

The classification has captured wide variations between areas across England and Wales. Geography is clearly important when researching public health and mortality variations, with the classification appearing to be a useful research tool for exploring and explaining this. The observed pattern largely follows the mortality profiles shown in across Tables 5.1 to 5.6. Those areas which displayed better mortality profiles (i.e. lower SMRs), have higher life expectancies at birth. However this may be expected given that the measure is derived from current mortality rates (Chiang, 1972), it highlights the inequalities captured by the classification. Nonetheless the gap is fairly

large, with a difference of nine years of life on average at birth for males, and eight for females between the highest and lowest clusters. There are considerable injustices in which people can expect to live fewer years on average dependent upon where they live (Diez-Roux, 2001; Woods et al., 2005).

Cluster	Males	Females	Difference
Best Health and Most Desirable	81.3	85.1	3.8
Average Mortality Profiles	77.9	82.5	4.6
Good Health Areas	79.4	83.5	4.1
The Middle	75.9	80.6	4.7
Poor Health Experiences	74.5	79.7	5.2
Poorest Health and Least Desirable	72.2	77.2	5
Poorest Neurodegenerative Health	75	78.6	3.6
Mixed Experiences	78.5	81.7	3.2
Total	78	82.1	4.1
Range	9.1	7.9	

Table 5.9: Life expectancy variations by cluster and gender.

Overall, females overall can expect to live, on average, four more years than males in England and Wales. The classification builds upon this established relationship to show how this varies by area type. The clusters which displayed higher mortality rates and poorer health also had a larger difference between males and females in terms of life expectancy at birth. The negative impact of these areas on health has a greater effect on the health of males, than compared to females. This is further shown through the range of values between clusters, which is also larger for males. Males are more susceptible to social and spatial processes, reflecting the protective biological characteristics of females (see Christensen et al., 2001). Poor health for males has a stronger impact upon explaining the poor health of an area. However this relationship is not completely consistent, as 'Poorest Neurodegenerative Health' shows.

To evaluate how useful this result was, the range of values captured using the classification was compared to a variety of different area-based measures which have been used to analyse life expectancy variations (Table 5.10). For those measures with a small number of categories, the classification captures a wider range of values than equivalent measures. It is a useful tool for analysis, capturing greater information in the data. Although local authorities have a larger range of values, this is not much higher

and therefore with the larger number of areas involved, the classification is shown to be still useful.

Measure	Deprivation Quintiles	Governmental Office Regions	Local Authorities
Range for males	8	2.8	10.7
Range for females	5.6	2.7	9.9
No. of areas	5	10	347
Source	Smith et al., 2010	ONS, 2010	ONS, 2010

Table 5.10: Differences in the range of life expectancy values by area measure, 2007-2009.

5.3.2 Premature mortality

Premature mortality was also calculated (Table 5.11). It gives a different aspect of the mortality patterns captured, given that like life expectancy, a high mortality rates does not necessarily mean a poor outcome since everyone dies. These were calculated as standardised mortality ratios (see Section 3.7.4 for details on method). It gives the percentage of deaths in respect to the expected trend for the total population (i.e. a value of 100). Total deaths under the age of 75 were used since this is both below the life expectancy for England and Wales (i.e. premature; see Table 5.10) and is comparable to other research (for example Thomas et al., 2010).

The patterns for premature mortality follow that of life expectancy. Those clusters which displayed higher mortality rates also had higher than expected rate of premature mortality. The range shows wide inequalities in premature mortality captured by the classification, indicating it to be a useful research tool. As with life expectancy, the range was greater for males than females, showing males to be greater affected by social and geographical factors.

The difference between males and females varied by cluster as well. For those clusters with better mortality profiles, males performed better. This pattern reversed for the clusters with poorer mortality profiles. Male health is affected more in those clusters which had higher mortality rates, highlighting its greater contribution to the formation of these clusters. Although this pattern reverses in the clusters with better health outcomes, the excess of male premature mortality in the other clusters would naturally

decrease the SMR elsewhere showing males not to be more susceptible to good environments.

Cluster	Males	Females	Difference
Best Health and Most Desirable	67	70	3
Average Mortality Profiles	96	97	1
Good Health Areas	81	83	2
The Middle	116	112	4
Poor Health Experiences	132	127	5
Poorest Health and Least Desirable	157	151	6
Poorest Neurodegenerative Health	118	118	0
Mixed Experiences	84	87	3
Range	90	81	

Table 5.11: The standardised mortality ratio for premature mortality (deaths under the age of 75) split by cluster and gender.

A comparison of these results to other analyses allows the classification's impact to be assessed. In Thomas et al. (2010), the range of SMR for both genders between deprivation deciles (of areas) was 65 (a ratio of 1.88), showing it to offer less detail than the classification's eight clusters (including even as a ratio) albeit only for 2006-2007. A recent report by PHE (2013) showed the relative difference between the best and worst Local Authorities to be 227 per cent. Although this was only using the crude rate, converting this for the data used here showed a relative difference of 194 per cent for males and 176 per cent for females. However it still performed well considering it uses far fewer areas.

5.4 Demographic variations by cluster

Having focused so far on medical and health based explanations of the clusters, it now seems useful to examine the types of areas within each cluster. This is the benefit of creating an area classification, as whilst it is both easier and more useful to cluster together areas, their information can be further analysed to gain additional information (Harris et al., 2005; Openshaw et al., 1994; Vickers, 2006). This adds value to the clusters, as well as developing further explanations for them.

5.4.1 Age

Age is an important demographic factor for understanding health differences. Risk of mortality and poor health increases throughout the life course (see Lindeboom and van Doorslaer, 2004). This operates at the individual level of analysis and its importance was why the input variables controlled for it (see Section 3.7.4). Rather the examination here focuses on improving the understanding of the characteristics of the areas captured in each cluster. This is important in explaining the differences in the mortality profiles between the clusters.

Data was taken from the ONS population estimates by age (2006-2009), as discussed in Section 3.7.4. The focus was on the differences by old age (identified here as the percentage of the population aged over 65), a key factor in explaining health (Table 5.12).

Cluster	Aged 65 and over (%)
Best Health and Most Desirable	17.67
Average Mortality Profiles	15.41
Good Health Areas	17.45
The Middle	14.27
Poor Health Experiences	14.03
Poorest Health and Least Desirable	13.47
Poorest Neurodegenerative Health	14.23
Mixed Experiences	17.86
Total	16.23

Table 5.12: Variations in people aged 65 and over (2006-2009) between the clusters.

There is only a small variation in the reported levels of the elderly population captured by each cluster. This is shown by the relative difference in values, with the largest value being only 1.33 times larger than the smallest. The distribution of values across the classification reflects the mortality profiles of each cluster, whereby there are fewer elderly people in those clusters with higher mortality rates. However it is not a strict linear process, as the cluster 'Mixed Experiences' contains the largest proportion despite not representing the cluster with the best health outcomes.

A Pearson's correlation was run to show the degree of association between this explanatory variable and the health statistics calculated in Section 5.3. This was used to

show whether age was a useful addition to the understanding of the area classification. The percentage of elderly people was a strong predictor of life expectancy for males ($r = .916$, $p = .001$) and females ($r = .849$, $p = .008$) within the classification, as well as the SMR for premature mortality for both sexes ($r = -.924$, $p = .001$ and $r = -.918$, $p = .001$ respectively). Age is a useful variable to explain the differences between clusters.

5.4.2 Communal homes

When compiling small area estimates of mortality, the effect of communal housing needs to be considered. Nursing homes have been shown to have higher mortality rates compared to the rest of the population, even after accounting for age (Raines and Wight, 2002). This concentration of frailer members of the population geographically will impact upon mortality measures. For example, Williams et al. (1995) showed that nursing homes bias SMR estimates when calculated using electoral wards. With the area classification created using SMRs across MSOAs (which are similarly sized to wards), evaluating the impact of this factor is important for understanding the clusters.

Data for communal homes was collected from the 2011 Census through the 'Neighbourhood Statistics' website. Communal establishments were defined as buildings containing supervised accommodation for its residents (ONS, 2012a). Those identified as 'medical or care' establishment types were also included to differentiate between them, due to their added focus on health. These included NHS related, local authority owned and social housing associations that provided some form of care due to health needs. The mean number of these establishments per area (MSOA) was calculated from these statistics for each cluster.

The results are presented in Table 5.13. There does not appear to be a common pattern which relates to the mortality profiles for all the clusters. There is little variation between the majority of the clusters, with patterns remaining mostly even. This holds for both variables.

For both 'Poorest Neurodegenerative Health' and 'Mixed Experiences', it appears useful as an explanatory variable. Both clusters report higher averages for both variables. The range of values is also relatively wide, being more than twice as large for each variable.

Cluster	Mean number of communal establishments	Mean number of medical or care establishments
Best Health and Most Desirable	7.6	2.6
Average Mortality Profiles	6.5	2.7
Good Health Areas	7.9	3.2
The Middle	8.3	3.3
Poor Health Experiences	5.6	2.5
Poorest Health and Least Desirable	11.5	4.2
Poorest Neurodegenerative Health	11.0	6.2
Mixed Experiences	10.6	5.4
Total	8.2	3.4

Table 5.13: Differences in average number of communal homes per MSOA between the clusters.

The location of these establishments would appear a useful explanatory factor in explaining their respective mortality profiles. For example both of the clusters displayed high mortality rates of neurodegenerative diseases; diseases of old age. It supports past research that has shown it to be important with regards small area health analysis (Williams et al., 1995; 2004). Yet it is only a good predictor for these two clusters. For example this is not the case for 'Poorest Health and Least Desirable', which demonstrates high mortality rates for these diseases as well but only a slightly increased number of establishments.

5.4.3 Migration

Areas are not static; rather they constantly evolve and change. Population change is the most important factor in how they change (Champion, 2012). Migration patterns capture this best and therefore form important dimensions of understanding the demographic characteristics of areas. Migration varies by social and demographic factors (Boyle, 2004). This in turn, impacts upon patterns of health and mortality (Brimblemcombe et al., 2000; see Section 2.5).

Migration is typically analysed using socio-economic area measures (Brimblecombe et al., 2000; Vickers, 2006). Applying the classification here will help reveal new

dimensions to this analysis. However since migration is important for understanding inequalities in mortality (Bentham, 1988; Brimblemcombe et al., 2000; Section 2.5), it can also help explain how the clusters have formed.

Statistics on population turnover were collected from the 'Neighbourhood Statistics' website for July 2008 to June 2009, the most recent year of the classification. The focus was on the net change (per 1000) of individuals per area. Migration was also split to examine the pattern for just those aged over 65, in accordance to the findings of Sections 5.4.1 and 5.4.2. This provides additional details to those factors which have been shown to be important so far.

Cluster	Net change (all)	Net change of those aged 65 and over
Best Health and Most Desirable	1.61	-4.28
Average Mortality Profiles	-1.84	-6.78
Good Health Areas	1.10	-3.02
The Middle	-2.48	-5.49
Poor Health Experiences	-5.12	-7.33
Poorest Health and Least Desirable	-5.58	1.77
Poorest Neurodegenerative Health	-0.64	17.81
Mixed Experiences	4.07	10.44
England and Wales	-0.17	-1.49

Table 5.14: Net change in the population of the clusters (per 1,000) mid-2008 to mid-2009.

The pattern for overall net migration rates appears to be useful (Table 5.14). The distribution of this variable relates to each cluster's mortality profile, where a cluster which displays better health outcomes attracts people to those areas. A Pearson's correlation was run for overall net change to analyse the association between it and life expectancy (by both sexes). However there was only partial support for this, with a significant positive association for males ($r = .811$, $p = .015$), but not for females ($r = .67$, $p = .069$). There was stronger support for the association between this factor and the SMRs for premature mortality ($r = -.872$, $p = .005$ for males, $r = -.855$, $p = .007$ for females).

The mortality profile also acts as a proxy for the social desirability of an area (see Section 5.5.1), as migration to a different area is not solely based upon health

considerations (Boyle, 2004). However migration does not just reflect the quality of the areas which consist of each cluster. Rather as Popham et al. (2011) argue, population instability and particular decline impacts negatively on health through the loss of community and social cohesion, as well as the provision of services (due to less demand).

The pattern for the migration of those aged over 65 was less clear. There was no discerning pattern between the clusters. This was reflected through correlating this to both the life expectancy of males and females which gave no significant relationships ($p = .617$ and .286 respectively). This result was mirrored for premature mortality as well ($p = .865$ for males and $p = .787$ for females).

Instead, the measure was more useful for explaining the overall net migration patterns for both 'Poorest Neurodegenerative Health' and 'Mixed Experiences'. Both clusters gave higher than expected overall net changes in the migration trends for the areas within them. This was accounted for by the inflow of elderly people.

Combining the data from Table 5.14 to Table 5.13 shows that this inflow would partly explain the higher rate of communal homes in these clusters. These clusters are partly capturing individuals who are moving to their areas near the end of their lives, requiring communal living arrangements to manage their health needs (Williams et al., 2004). Calculating the Pearson's correlation showed this association between these two factors to be both strong and significant ($r = .981$, $p < .001$).

This concentration of people with higher mortality risk has led to their respective morality profiles. Both 'Poorest Neurodegenerative Health' and 'Mixed Experiences' have higher than expected (as well as compared to the other clusters) mortality rates for diseases associated with old age, especially those related to the mental and nervous systems. Migration was useful for explaining these clusters.

5.5 Social patterns across the classification

There is a wealth of evidence which has shown the detrimental impact of socio-economic factors on health, both at the individual and area level (Evans and Kim, 2007; Gregory, 2009, Riva et al., 2007). As such, mortality patterns by cause begin to vary by social phenomenon, which itself displays sharp divisions geographically across England

and Wales (Dorling, 2012). This, in turn, results in mortality patterns reflecting this underlying structure (Shaw et al., 2008; see Section 5.2.4). Analysing the social dimensions of the clusters is important for understanding them.

5.5.1 Poverty

The first feature considered is the degree of poverty experienced within an area. Whilst it may be argued that this refers to an individual-level process, the literature review (especially Section 2.3) showed evidence that neighbourhoods can contain additional disadvantage beyond simply reflecting a resource issue (Evans and Kim, 2007). The level of poverty across areas has been shown to be strongly related towards mortality rates, through a detrimental effect (Gregory, 2009; Shaw et al., 2008; Yen et al., 2009). Therefore it would be useful to explore the level of poverty between the clusters.

Data was collected on household poverty levels from the 'Neighbourhood Statistics' website. These were modelled estimates of the percentage of households in an area with a combined income of less than 60 per cent of the median income for England and Wales (April 2007 to March 2008). This value has been commonly applied elsewhere as a key measure for poverty (Townsend and Kennedy, 2004). Income had been equivalised to account for differences in housing types (for example size, welfare, family structure) to improve comparability between households. The results are shown in Table 5.15.

Cluster	Poverty (%)
Best Health and Most Desirable	15.18
Average Mortality Profiles	23.20
Good Health Areas	19.12
The Middle	27.09
Poor Health Experiences	30.43
Poorest Health and Least Desirable	31.97
Poorest Neurodegenerative Health	24.92
Mixed Experiences	18.32
England and Wales	21.56

Table 5.15: Variations in average household poverty level between clusters.

There are large differences in the level of household poverty between the clusters, with there being over twice as much poverty found in 'Poorest Health and Least Desirable' than compared to 'Best Health and Most Desirable'. The classification is useful for capturing different types of areas. The distribution of values reflects each mortality profile of their respective cluster, showing the classification to capture social divisions throughout its clusters of mortality patterns. Where there are profiles associated with higher mortality rates, there is also a greater concentration of poverty.

The detrimental effect is further shown through correlating this with the health outcome measures from Section 5.3. Statistically significant associations (at the 99 per cent level) were found for both the life expectancy variables, as well as the premature mortality measures as well. The associations were always strong (each above ±0.8), showing it to be useful explanatory factor of health.

The clusters 'Poorest Neurodegenerative Health' and 'Mixed Experiences' contain lower levels of poverty than their mortality profiles would have suggested. As shown in Sections 5.4.2 and 5.4.3, this is the effect of the elderly migrating into these areas to communal homes. This in turn lowers the level of poverty, leading to this slight mismatch.

5.5.2 Social class

In response to Goldthorpe's (2010) call that health research (especially that based around inequality) needs to refocus back around the importance of social class, this study will explore the influence of this additional factor.

Social class plays an important role in England and Wales in creating inequality in health outcomes. Research into this field is well established, with the negative effect of social class on health being observed throughout the life course (Doran et al., 2004; Johnson and Al-Hamad, 2011; Langford and Johnson, 2010; White and Edgar, 2010; Young et al., 2010). High social class limits exposure to health damaging factors throughout the life course, influencing how and when people die. Though research has begun to focus on the impact of deprivation instead, social class should not be forgotten, especially in England and Wales where society has longstanding structural divisions based upon class (Goldthorpe, 2010).

The original mortality data included information regarding the social class (using the National Statistics Socio-economic Classification; NS-SeC), with near perfect coverage under the age of 75 (see Section 3.3.2). Applying this data allowed the premature mortality rate to be calculated, albeit only for individuals aged 16 to 74 due to limitations in the population data available (social class at the area level was only collected for these years). The area statistics were gathered from the 2011 Census, to allow the standardisation of mortality rates. The NS-SeC group was collapsed into its three group classification (high, intermediate and low; see Table 3.3). Although some quality of data was lost through collapsing to three groups, the other means of collapsing or analysing the groupings cannot be assumed to be a direct ordinal scale, restricting any comparisons (ONS, 2004).

Firstly, variation in the proportion of individuals by their social class in the areas contained within each cluster is explored (Table 5.16). There are clear differences by the clusters, reflecting similar patterns to previous analyses. Those clusters which displayed worse health outcomes had higher proportions of people with low social class living in them and *vice versa*. This corresponds to past research, which has shown the detrimental impact of low social class on health, as well as the existence of social gradients by area types (Doran et al., 2004; White and Edgar, 2010; Young et al., 2010). Greater detail was captured at both extremes of the measure, which are more important to understanding processes between the clusters.

Cluster	High	Intermediate	Low
Best Health and Most Desirable	40.1	24.4	24.3
Average Mortality Profiles	28.8	21.9	34.3
Good Health Areas	34.0	24.0	30.0
The Middle	24.2	19.8	38.1
Poor Health Experiences	20.0	18.4	42.3
Poorest Health and Least Desirable	17.7	16.8	42.7
Poorest Neurodegenerative Health	27.3	20.2	33.9
Mixed Experiences	34.7	23.7	29.5
England and Wales	31.1	22.2	32.2
Range	22.4	7.5	18.5

Table 5.16: Variations in the percentage of the population by social class groups.

The cluster 'Mixed Experiences' performs particularly well, considering its varied mortality profile. Those areas which may exhibit the best social structures do not always display the best health. 'Poorest Neurodegenerative Health' also displays better social factors than its mortality profile would suggest. These patterns can be explained with the inflow of elderly people to communal homes, which increased the proportion of people with high social class in their areas. 'Mixed Experiences' had a larger percentage of high social class, showing the cluster to be more desirable (and therefore attract more elderly people of a higher social class).

In this context, social class has been used as a proxy of the quality of an area. However this approach may ignore the true mechanisms which affect health. The impact of social class on health is more likely to be at the individual level, having a direct impact upon health (Goldthorpe, 2010; Young et al., 2010). With the mortality database containing individual level information on social class, this would be useful in showing social effects within and between clusters. This ability to explore the actual impact of social class on premature mortality provides a huge benefit to research, rather than relying on solely aggregated (area) statistics.

Exploring the differences between clusters within the NS-SeC schema shows a divided and unequal society (Figure 5.3). Each cluster displays a social gradient, with those in the lower social classes experiencing higher premature mortality. This follows the findings of past research (Johnson and Al-Hamad, 2011; Langford and Johnson, 2010). However differences between the clusters results in varying social gradients, showing the classification to be useful for discriminating between patterns.

The gradient for the first cluster 'Best Health and Most Desirable' is the smallest. The advantages of living in this cluster apply for all social classes, being distinctly lower for each social class. Unlike the other clusters, social class has less influence on those individuals within this cluster. Death is less socially determined here, with other factors being more important.

The opposite is experienced in the cluster 'Poorest Health and Least desirable'. There is a sharp gradient, with the rate almost doubling (196 per cent) between the top and bottom classes, showing a higher level of inequality. Not only does the effects of social class on health affect those of low social class greatly, but even impacts on those of high social class. Similarly for 'Poor Health Experiences' as well, people of high social class

have a higher premature rate than compared to those of lowest social class in the cluster 'Best Health and Most Desirable'.

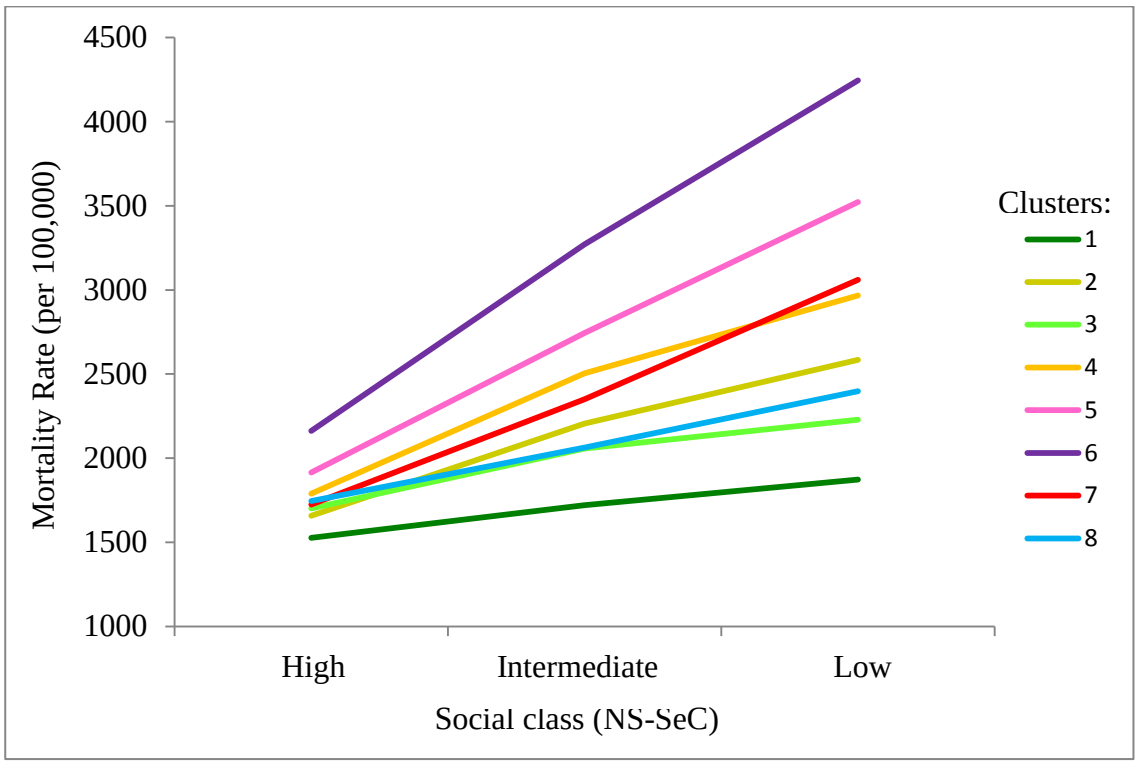


Figure 5.3: Differences in the premature mortality rates (16-74 year olds) of each cluster across social class.

Key: Clusters; 1 = Best Health and Most Desirable, 2 = Average Mortality Profiles, 3 = Good Health Areas, 4 = The Middle, 5 = Poor Health Experiences, 6 = Poorest Health and Least Desirable, 7 = Poorest Neurodegenerative Health, 8 = Mixed Experiences.

There is less variation observed amongst the other clusters. Rather they follow their respective mortality profiles in both their position in the graph and the size of their social gradients. The clusters with worse health outcomes and higher mortality rates have more distinct social gradients, with greater levels of social inequality in health. They are affected more by social class than compared to those areas which have better health.

Whilst the analysis has shown the classification to be useful in this context, it does not mean that it is a better tool than the social class measure. Rather using both measures enhances understanding of patterns and processes, complementing each other. The cluster each individual finds themselves in is also influenced by social class itself.

5.6 Geographical variation of the clusters

Visualising the classification earlier (Section 5.2.1) showed a complicated geographical pattern. However this does not necessarily mean that no spatial pattern exists. Chapter 2 showed the importance of taking a spatial approach to understanding mortality patterns and therefore examining this dimension is important for extending the understanding the clusters created.

5.6.1 Rural-urban differences

Mapping the classification (Figure 5.1) showed that particular clusters dominated the map, whereas others appeared less prominent. However MSOAs are not designed to be equal in geographical size, rather reflecting population sizes (Martin, 2002; ONS, 2011). As such, rural areas are larger in area size. Analysing rural and urban differences would appear useful for differentiating between the clusters.

Rural-urban measures have been long established as important for understanding differences in health patterns. For example, both Erskine et al. (2011) and Riva et al. (2011) have shown evidence that rural areas have significantly lower risk of mortality than compared to urban areas. This is due to the mixture of a concentration of affluence in rural areas, as well as the beneficial environment experienced in these areas also (for example lower levels of pollution). As such, it is an important factor for understanding the areas captured by the area classification.

Cluster	Rural	Urban
Best Health and Most Desirable	30.3	69.7
Average Mortality Profiles	12.2	87.8
Good Health Areas	29.7	70.3
The Middle	9.3	90.7
Poor Health Experiences	3.8	96.2
Poorest Health and Least Desirable	3.0	97.0
Poorest Neurodegenerative Health	7.8	92.2
Mixed Experiences	28.2	71.8
England and Wales	20.0	80.0

Table 5.17: Cross-tabulation of the classification with the ONS urban-rural classification of MSOAs

Table 5.17 presents the differences in cluster membership of rural and urban areas. Areas were split by using the official ONS classification, grouping areas based mostly upon population size (Pateman, 2010/2011). Supporting past research, the clusters which had lower mortality rates and better health had higher proportions of areas classified as rural (and *vice versa*). A rural-urban divide is observed through the classification. However this is not a strict division in health outcomes. There is still a large proportion of the clusters with good mortality profiles found in urban areas as well. The classification is capturing more than simple geographic divisions.

Although there are clear spatial differences between the clusters, this does not show whether health is better in rural areas. To test this, life expectancy by sex was calculated using the same method as in Section 5.3 (see Chiang, 1972). Estimates were calculated for each cluster, split by the identified rural and urban areas within each. The results are shown in Table 5.18.

For both rural and urban areas, life expectancy at birth varied similarly to Table 5.10, with the clusters which displayed poorer mortality profiles having lower life expectancy. Although life expectancy is higher in rural, the difference between urban and rural areas is only small. This difference is smaller for the clusters with better health, however the variation is not particularly large. Inequalities between sexes offer less variation in rural areas between the clusters compared to urban areas. Variation is larger in those disadvantaged clusters, with males particularly affected.

Cluster	Male		Female	
	Urban	Rural	Urban	Rural
Best Health and Most Desirable	80.2	80.3	82.9	83
Average Mortality Profiles	77.2	78.2	80.9	81.2
Good Health Areas	78.6	79.3	81.9	82.2
The Middle	75.5	76.8	79.7	80.3
Poor Health Experiences	74.3	76.1	78.8	79.5
Poorest Health and Least Desirable	72.4	74.3	77.1	78.1
Poorest Neurodegenerative Health	75.1	77.5	78.8	79.6
Mixed Experiences	78.1	79	81.1	81.7

Table 5.18: Life expectancy at birth split both by cluster and gender for urban and rural areas.

The classification is accounting for the variation in health usually observed by simply comparing urban and rural areas. A similar result was found by Gartner et al. (2011), who showed that the effect of rurality was attenuated by social factors. The classification is acting as a useful proxy measure for the social conditions of areas and accounting for geographical patterns (albeit not all social patterns; see Section 5.5).

5.6.2 Regional patterns

The wide social divisions that occur across England and Wales have resulted in a distinct geographical patterning of mortality patterns (Hacking et al., 2011; Shaw et al., 2008; Walters et al., 2011). The main pattern concerns the North-South divide whereby the North experiences poorer health outcomes (Doran et al., 2004; Hacking et al., 2011), although variations are found within each region. This is mostly accounted by spatial differences in deprivation (Woods et al., 2005). Since it is an area classification, it would be expected that some artefact of these patterns would persist through the clusters (Vickers, 2006).

Table 5.19 presents a cross tabulation of cluster membership between the 'Governmental Office Regions' (GORs) of England and Wales. GORs were used to analyse regional patterns due to their use in prior research (Doran et al., 2004; Hacking et al., 2011), as well as MSOAs being self-contained within them.

A Chi-squared test was performed on the data to test whether there are significant differences across the observed data (by region) compared to if there were no pattern at all (Rogerson, 2006). The test gave a highly significant relationship (χ^2=1363.957, $p<0.001$), indicating the existence of regional patterns across the classification.

The standardised residuals were calculated from the Chi-squared test to show the degree of difference in the observed value compared to the expected value given no pattern (Rogerson, 2006). Absolute values greater than three should be viewed as important, where cluster membership is more or less common than expected. The residuals show a geography to the classification, something which is lost in the map (Figure 5.1). The most common pattern appears to be a North-South division. Those clusters which had better mortality profiles are more commonly found in the Southern areas, with the opposite found in the North. This follows past research (Doran et al., 2004; Hacking et al., 2011), showing the classification to be useful in capturing spatial patterning.

Cluster		GOR								
		NE	NW	Y	EM	WM	Wales	London	SE	SW
1	Count	40	98	93	112	142	43	240	366	217
	Std. Residual	-4	-7.2	-4.7	-1.1	-1.4	-4.9	1.8	8.1	5.4
2	Count	44	154	137	79	151	67	196	131	74
	Std. Residual	-1.4	0.6	2.5	-1.3	3.1	0.1	3.1	-3.4	-3.5
3	Count	33	92	76	119	90	85	201	256	182
	Std. Residual	-3.7	-5.9	-4.5	1.5	-3.8	1.1	1.6	3.8	4.9
4	Count	66	132	77	75	72	87	136	100	51
	Std. Residual	4	2.2	-0.6	0.9	-1.6	5.4	1.8	-2.7	-3.5
5	Count	87	161	122	32	92	48	66	21	12
	Std. Residual	10	8.4	7.4	-2.8	3.1	1.7	-2.5	-7.9	-6.4
6	Count	31	122	57	19	19	24	12	6	3
	Std. Residual	4.5	13.6	5.3	-0.9	-2.1	1.7	-4.5	-5.9	-4.8
7	Count	13	55	53	23	34	14	38	41	16
	Std. Residual	-0.6	2.1	3.9	-0.5	0.2	-1	-0.9	-1.2	-2.7
8	Count	28	108	79	112	135	45	94	185	140
	Std. Residual	-3.1	-2.3	-2.2	3.2	2.7	-1.9	-4.1	1.9	3.9

Table 5.19: A cross-tabulation of cluster membership and geographical location.

Key: Clusters; 1 = Best Health and Most Desirable, 2 = Average Mortality Profiles, 3 = Good Health Areas, 4 = The Middle, 5 = Poor Health Experiences, 6 = Poorest Health and Least Desirable, 7 = Poorest Neurodegenerative Health, 8 = Mixed Experiences. GORs; NE = North East, NW = North West, Y = Yorkshire and Humberside, EM = East Midlands, WM = West Midlands, W = Wales, E = East, SE = South East, SW = South West.

Not all the clusters display distinct spatial gradients with regards to regional membership. The second cluster 'Average Mortality Profiles' is less common in the South, but does not show a clear pattern for where it is more commonly found. Similarly the pattern for 'Poorest Neurodegenerative Health' does not show any distinct patterning other than the 'Yorkshire and Humberside' GOR. What these clusters are capturing is likely to be independent of spatial factors and consistent across England and Wales.

Cluster	GOR										Range
	NE	NW	Y	EM	WM	W	E	L	SE	SW	
Best Health and Most Desirable	80.4	80.2	80.4	80	79.8	79.9	80.2	80.4	80.3	80.4	0.6
Average Mortality Profiles	77.4	77.7	77.8	77.5	76.7	77.1	77.7	77	77.4	77.5	1.1
Good Health Areas	78.6	79.1	79	78.8	78.2	78.6	79.1	78.3	79.1	78.8	0.9
The Middle	76.1	75.5	76.1	75.3	74.8	75.9	76.1	75.5	75.5	75.5	1.4
Poor Health Experiences	74.8	73.9	74.7	74.5	73.8	74.5	75.8	74.4	73.9	74.6	2
Poorest Health and Least Desirable	72.2	72.1	72.4	72.3	73.1	73.3	73.5	73.4	72	71.7	1.8
Poorest Neuro-degenerative Health	74.9	74.3	75	75.3	75.1	74.8	75.9	75	76.3	76.5	2.2
Mixed Experiences	78.2	78.2	78.4	77.9	77.8	77.6	78.7	78.3	78.9	78.6	1.4
Range	8.2	8.1	8	7.7	6.7	6.6	6.7	7	8.4	8.7	

Table 5.20: Geographical differences in life expectancy at birth for males by cluster.

Key: GORs; NE = North East, NW = North West, Y = Yorkshire and Humberside, EM = East Midlands, WM = West Midlands, W = Wales, E = East, SE = South East, SW = South West.

Cluster	GOR										Range
	NE	NW	Y	EM	WM	W	E	L	SE	SW	
Best Health and Most Desirable	82.8	82.8	83	82.8	82.5	82.8	82.8	83.2	83	83.1	0.7
Average Mortality Profiles	80.8	81	81	80.9	80.5	80.8	81.2	81	81.1	81.1	0.7
Good Health Areas	81.5	82	81.9	81.7	81.6	81.8	82.1	82	82	82.2	0.7
The Middle	79.8	79.6	79.9	79.5	79.3	79.7	80.1	80.1	79.9	79.9	0.8
Poor Health Experiences	78.7	78.5	78.9	78.7	78.8	79	79.6	79.3	78.9	79.7	1.2
Poorest Health and Least Desirable	77	76.7	77.1	77	78.1	77.6	78.5	78.6	77.6	78.4	1.9
Poorest Neuro-degenerative Health	77.8	78	78.8	78.8	78.8	78.5	79.3	79.1	79.6	79.7	1.9
Mixed Experiences	81.2	81.1	81.3	80.9	81.1	80.9	81.5	81.3	81.6	81.5	0.7
Range	5.8	6.1	5.9	5.7	4.5	5.3	4.3	4.6	5.4	4.8	

Table 5.21: Geographical differences in life expectancy at birth for females by cluster.

Key: GORs; NE = North East, NW = North West, Y = Yorkshire and Humberside, EM = East Midlands, WM = West Midlands, W = Wales, E = East, SE = South East, SW = South West.

Calculating life expectancy (see Chiang, 1972) at birth for both sexes split by both cluster and GOR allows the examination of geographical inequalities across both measures (Tables 5.20 and 5.21). The differences in life expectancy (measured by the range) are larger within GORs, rather than between clusters. Whilst the classification is capturing different mortality profiles, there is less spatial variation within these clusters

across England and Wales. Once again the classification has accounted for the majority of spatial variation in mortality.

Spatial inequalities exist but mainly between clusters. These remain fairly wide within each region, for example in 'Yorkshire and Humberside', males at birth can expect to live eight years different based upon which cluster they live in. The range of inequalities by cluster does not vary by region either, as shown by males as the range is just as high in the more deprived (the Northern areas) regions as it is in the more affluent ones (those in the South). Compared to the variation within GORs, this is much lower although there is some difference by cluster with those with poorer health characteristics presenting a greater range of values in a GOR. Inequalities are also smaller for females than males, reflecting previous findings. The classification adds more to our understanding of mortality in England and Wales than by using regions alone.

5.7 Conclusion

Mortality has been shown to vary geographically by cause (Shaw et al., 2008), a factor of underlying social processes. An area classification was built to capture and summarise the interactions between the most prevalent causes of death. This chapter has analysed the resulting classification to explain what has been captured, to highlight how useful it is.

The division of mortality patterns into a set of eight clusters initially appears to have occurred based upon the degree of mortality rates. Despite clear (and differing) geographies in the causes of mortality throughout the literature (see Chapter 2), this is less important with a single level area classification. However this is not the only factor in explaining differences between clusters and cause of death does have some importance, especially when comparing 'Poorest Health and Least Desirable' and 'Poorest Neurodegenerative Health'. Large inequalities exist between the clusters in terms of mortality rates and life expectancy, showing that it captures a useful amount of detail.

Neurodegenerative diseases appear to play an important role in the formation and understanding of the classification. The variation in detail captured between the clusters was highest for these types of cause of death (Table 5.8). This was accounted by the

location of communal homes and the migration of the elderly, which were important factors in explaining two of the clusters. However it would not be useful to drop these deaths or areas from the analysis as it would produce a result that does not reflect the true underlying structure of mortality patterns for small areas across England and Wales.

The understanding of the classification was aided through a social and spatial investigation of the characteristics of the areas contained within the clusters. Social and demographic explanatory factors were important for explaining patterns between clusters, particularly poverty. Spatial inequalities were better captured through using the classification than compared to other standard measures, showing the classification to be a useful analytical tool.

Chapter 6: Assessing the area classification

6.1 Introduction

The clusters that make up the classification were explained in the previous chapter. Although the divisions in mortality patterns across England and Wales showed differences in health and social factors, how useful these clusters are can only be inferred descriptively. Through introducing a focused analytical framework, the classification can be evaluated to assess how useful it is as a research tool.

This chapter will apply the classification to assess the existence of neighbourhood effects upon health. This is an important research area identified in the literature review (Chapter 2), with policy implications. Applying the classification in this context will show the degree that the types of areas identified by the clusters can be shown to be important, through exhibiting patterns and processes beyond simply being areas of similar mortality rates. It is also important for justifying and introducing future analyses set around this theme (Chapter 7).

Utilising the classification will help explore and extend previous research within this theme. The clusters act as a proxy for the damaging or beneficial impact of that area type. Chapter Five showed how the classification adds value to a traditional analysis (Openshaw et al., 1994), through applying a measure which not just divides England and Wales by dominant mortality profiles, but also incorporates social and demographic correlates as well. Therefore the analysis allows for a more focused investigation of how an area can influence health. It offers a different and better designed perspective than previous studies, which tend to focus on just social and demographic factors (Riva et al., 2007).

To be able to assess the evaluation of the classification, the analysis needs to be compared to an equivalent measure. This will show how useful the results are for the performance of the classification under this particular approach. The nine Governmental Office Regions (GORs) of England were chosen for this, along with Wales as well (hereby referred to as a GOR, for ease of description since it is usually included within them in official statistics). With ten areas in total, this provides a number of categories that allows a fair analytical comparison. Since it allows the same

geographical areas to be grouped differently, hypothetically it could be a classification itself with the main 'clusters' being the regions themselves.

The use of GORs was selected since they are often used to display key health and mortality information by the ONS (to allow some geographical breakdown of variables). This does not appear appropriate, since the GORs are just arbitrary location-based categories. There has been little change in how statistics have been reported since William Far advocated using geography to report mortality and morbidity rates (Dorn, 1956). If the classification can be shown to improve on these, there would be a strong case for releasing the statistics with the clusters as well, since they improve our understanding of the processes that exist (than compared to GORs which contain a multitude of area types within them; Dorling, 2012; Vickers, 2006).

This analytical approach was useful for Parfitt et al.'s (2001) classification of waste management strategies. An eight cluster solution was created to analyse how local authorities operated through waste collection, with the aim of targeting areas for greater implementation of recycling strategies. The conclusions of the approach showed that their classification captured greater variance than compared to a regional categorisation of recycling patterns. Therefore it can be inferred that their classification is a better measure for data dissemination and analysis than existing methods.

This rest of this chapter is set as following: Firstly, details are provided of a new data set and methodology for the analysis. This includes the model inputs, as well as the testing of model assumptions. Next a comparison is carried out between the GORs and the classification to examine which is more useful for capturing variations in health. This is conducted through a descriptive investigation, as well as fitting competing models. The analytical framework is then developed to incorporate small areas to test the existence of area effects, incorporating the classification to further the understanding of it as well.

6.2 Data and methodology

Health is a complex mixture and interaction of individual and area level factors (Macintyre and Ellaway, 2000). To assess the importance of the clusters, then both of these levels at which risk factors operate require controlling for. Otherwise building an analytical model for just one of these levels will lead to a falsely specified model. It is

paramount that both 'Simpson's paradox' (analysing individual level data without accounting for the true (multi-level) structure of the data) and the 'ecological fallacy' (assuming the relationships of area level variables are equivalent at the individual level) are avoided (Hox, 2002; Rogerson, 2006).

The importance of this direction is not just theoretical, but also statistical. If relationships differ by cluster, logistic regression techniques will ignore these differences leading to statistical issues through a larger sample size, which falsely leads to a drop in the variance and hence standard errors (Hox, 2002; Kreft and De Leeuw, 1998; Snijders and Bosker, 1999). This could lead to a set of misleading results.

A multi-level logistic regression model was used to evaluate the usefulness of the classification within this context. The methodology incorporates the above mentioned conceptual framework, as well as being commonly used in previous research to analyse neighbourhood effects (Riva et al., 2007; Andersson and Musterd, 2010). This is because the method allows for the modelling of the data at multiple levels within the same model (Snijders and Bosker, 1999). This is allowed through 'nested data', where there are hierarchical levels of data self-contained within other data units (for example individuals within areas).

The method works through fitting multiple regression models, which interact with each other (Hox, 2002). At the lowest level, a simple regression model is fitted to analyse the relationship between a set of predictors and an outcome variable. Then for subsequent levels, the model allows the relationship to vary by that particular data unit. For example letting the individual level relationship vary by each area, to view if some areas are having influential effects through differing relationships. As such, it captures both the between- and within-group variance allowing the understanding of how much effect is accounted by a level (Kreft and De Leeuw, 1998).

Past studies have only shown low levels of variance captured by multi-level models investigating neighbourhood effects, with models where areas account for anything over five per cent after controlling for covariates deemed impressive (Pickett and Pearl, 2001; Riva et al., 2007). It is because the causal link for area effects tends not to be direct in their impact upon health (Diez-Roux, 2001), so their effect size will be minimised in such an analysis. Therefore the results of this chapter need to be viewed in context to past results. There are also few studies which apply classifications in a

multi-level analysis, showing little expectation of how useful this is as a methodological application, as well as benchmarking effect sizes.

An additional data set is therefore required to analyse individual effects on health. The British Household Panel Survey (BHPS) was chosen, as this contains a wide range of social, economic, health and demographic information. The BHPS was collected annually, between 1991 and 2008. It is a large survey, with 9667 individuals located in England and Wales in its most recent year. For this analysis, the focus is on the most recent wave; 2008. This was because some of the social variables of interest were only collected at specific waves, denying the ability to incorporate a longitudinal framework into the analysis.

Special licence access was applied for and approved by the Economic and Social Data Service, who provided an additional data set which matched individuals in the data to their geographical area through the LSOA (Lower Super Output Area) of where they lived. This allows MSOAs (Middle Super Output Areas) to be joined, since LSOAs are nested within MSOAs (ONS, 2011). Therefore individuals could be assigned to their respective cluster, allowing a multi-level modelling approach to be adopted.

6.2.1 Outcome variable

To analyse the impact of areas on individuals, current health was chosen as the theme to focus on. This was based upon the literature review (Sections 2.3 and 2.4), in which evidence was presented to show that areas have an influence on current health. It was important to analyse current health rather than mortality, to avoid any double counting of mortality when using the classification as an explanatory factor. Morbidity and mortality are not the same and this is important to stress.

Using the BHPS, the dependent variable chosen for the analysis was self-reported health status. It was selected since it has been applied in a multi-level analysis frequently (for example; Jen et al., 2009; Kondo et al., 2009; Lopez, 2004; Malmström et al., 1999; also see Pickett and Pearl, 2001 and Riva et al., 2007), showing it to be a useful measure to apply. Knowing that it has demonstrated a consistent effect of areas on health is useful for evaluating the results using the classification.

Self-reported health status asks respondents to rate their current health on a Likert scale ('Excellent', 'Good', 'Fair', 'Not Good' or 'Not Very Good'). To allow for a more concise analysis, the measure was collapsed into a binary measure of poor health ('Fair', 'Not Good' or 'Not Very Good') and good health ('Excellent', 'Good'). This approach has been commonly used elsewhere (Jen et al., 2009; Kondo et al., 2009; Lopez, 2004; Malmström et al., 1999).

Splitting the variable into the binary measure was also useful for improving the accuracy of observations. This is because the categories of the measure become less abstract (Jylhä, 2009), making any judgements more reliable (i.e. it is difficult to quantify the difference between 'Fair' and 'Not Good' health, but combined together they incorporate poor health which is more important). The analysis modelled risk of poor health, with 30.3% of the dataset reporting this.

Self-reported health status is, however, a subjective assessment made by the individual on their actual health. Therefore there may be error in the data, if individuals falsely record their health. Nevertheless the application of the self-reported health has been shown to be useful for accounting for actual health. For example, the international systematic review conducted by Idler and Benyamini (1997) showed consistent evidence of the link between self-reported health and mortality, even after controlling for explanatory factors of mortality. This review has been updated by Jylhä (2009), who found the same result across more recent literature, incorporating a greater range of health measures as well.

There are other issues with the measure limiting its usefulness in an analysis of health. How it is filled in can vary by different factors not related to an individual's health including age, gender (see Section 6.2.2) and national differences (Dorling and Barford, 2009). This is partly because of a lack of a point of 'reference' for an individual to be able to assess their own health (Jylhä, 2009). However, this is more important for international comparisons using the measure, to account for cultural effects which cannot be controlled for. Roos et al. (2010) also note that self-reported health usually gives a stronger association when examining area effects (through multi-level analyses) than compared to mortality or actual health variables.

6.2.2 Explanatory variables

The focus of the analysis is whether where an individual lives (context) has an independent effect on health beyond who they are (composition). With strong evidence that health is socially determined (for example Congdon, 1996; Gregory, 2009; Muennig et al., 2013), it is important to control for the main influential factors to ensure that any observed areas effects cannot be otherwise explained.

The selection of variables to control for the main social processes which affect health was dictated by the conceptual framework outlined in the literature review (section 2.3). This included social-economic status, poverty, social capital, local support, cohesive environments and social integration. Variables from the BHPS were taken to best account for each of these factors. These operate at both the individual and area levels, to account for the varying interacting processes occurring (Macintyre and Ellaway, 2000). The variables chosen are summarised in Table 6.1 and were all added as fixed effects in the model.

Name	Level	Description	Type
Age	Individual	Age (Years)	Continuous
Male	Individual	Gender (male or not)	Nominal
Income	Individual	Annual income (£)	Continuous
Trust	Individual	Trust of others	Nominal
Civic	Individual	Civic participation	Nominal
Meet	Individual	Meeting other people	Nominal
Belong	Individual	Belong to the local area	Nominal
Advice	Individual	Get advice from the local area	Nominal
Poverty	Area	Poverty rate of area	Continuous
Turnover	Area	Population turnover	Continuous

Table 6.1: Explanatory variables incorporated in the multi-level analyses.

Age and gender were included to control for personal characteristics. These were important, as both variables have been shown to affect how people report their health. Females and the elderly are often found to be more likely to report their health as poorer (Idler and Benyamini, 1997). However, Lindeboom and van Doorslaer (2004) find that females and younger people are more likely to give fairer and more accurate assessments of their health. Controlling for these differences is therefore important for

allowing fairer comparisons. Age was included as a continuous variable and gender was coded as a binary measure of whether the individual is male or not.

Social disadvantage is an important factor that operates at both the area and individual level. Individual socio-economic status has a direct effect on health through limiting resources that could prevent health damaging processes (for example housing quality and location; Jin et al., 2011). To account for this, annual income was selected as it conceptualises the resources issue effectively, as well as being used for the same purpose in previous multi-level models (for example Jen et al., 2009).

Whilst income accounts for compositional factors, social disadvantage can also operate through a contextual effect as well. Concentrated poverty and deprivation can lead to additional effects through the exposure to other physical and social risks (for example the diffusion of unhealthy behaviours; Evans and Kim, 2007; Wilson, 1987). The poverty rate for an area was therefore introduced as well, using the same measure used previously (Section 5.5), which was shown to be important for understanding the classification. Controlling for this risk factor has been shown to be important in past multi-level analysis (for example Chaix et al., 2007).

Social capital is an important factor in influencing health. It represents the relationships between individuals in a community, which has been associated with lower levels of stress (Berkman, 1995), better support networks and greater demand for services (Kawachi and Berkman, 2000). Robert Putnam's seminal works provide the original basis for measuring social capital (Putnam et al., 1993, Putnam, 1995). Putnam defined social capital as partly (albeit most importantly) through trust and civic participation. Although they were used to evaluate economic success, social capital has also been applied within a health setting as well (Kawachi et al., 1999; Muennig et al., 2013).

For measuring trust, the variable 'trustworthiness of others' was converted into a binary measure of 'most people can be trusted' for a value of one and both 'can't be too careful' and 'depends' for zero. The split is more useful for analysis, as well as theoretically. The variable 'Do unpaid voluntary work' was used for civic participation, recoded as a binary variable between whether an individual had in the past year or not. These variables have been applied elsewhere for the same purpose of measuring social capital (Li et al., 2005; Snelgrove et al., 2009).

The level of support provided locally was also identified in Section 2.3 as important for the diffusion of advice about health issues, as well as emotional support (Berkman and

Glass, 2000). To measure this, the individual level question 'Advice obtained locally?' was incorporated into the analysis. The variable is measured using a Likert scale to show how much an individual agrees with the question. It was re-coded to a binary response, with respondents answering either 'Strongly Agree' or 'Agree' coded as one, and all other responses zero. The variable was used by Li et al. (2005) in their analysis of social capital, who notes that it also shows behavioural responses to local attachment, linking it to social capital as well. Although not specifically about health, it implies the range of advice diffused through communities of which health information will form a part of.

The integration of individuals into society can affect health (Congdon, 1996). The variable 'frequency of meeting people' was used to measure how well people are integrated through focusing on social interactions. This was influenced by Muennig et al.'s (2013) analysis of the different dimensions of social capital on heart related disease incidence and mortality.

An issue with this variable is how to divide it up to be useful, given that there is not an even distribution of responses. For Muennig et al. (2013), variables should be split to cover a meaningful number of visits. To be integrated, it was decided that a value of one was allocated to those meeting people at least once a week, with anything longer assigned to be zero.

The cohesive nature of the local environment an individual resides in can influence health both indirectly since social bonds and conflict affect social capital and support networks (Kawachi and Berkman, 2000), and also directly through anxiety and stress as well (Wilkinson and Pickett, 2009). The variable 'do you belong to your neighbourhood' was used to measure the tie to the local area which would reflect a cohesive area (Berkman and Glass, 2000). It was re-coded to a binary response, with respondents answering either 'Strongly Agree' or 'Agree' coded as one, and all other responses zero. The variable was included in Li et al.'s (2005) analysis for a similar purpose.

The final variable included was the population turnover (the net change of the population for all ages as a rate per 1000) for an area, which was used in Section 5.4. It was chosen as it was used in Chaix et al. (2007) as a measure for residential stability. Those areas which were losing population were denied the ability to build strong social bonds, which would impact upon social capital and support within an area. It was also

important in Congdon (1996) in forming part of his anomie index (i.e. barriers to the social integration of individuals), also at the area level. As such, it is a useful addition to account for the various social processes detailed previously, albeit this variable occurs at the area level instead.

6.2.3 Testing the assumptions of the model

Three multi-level logistic regression models were run in the analysis. The assumptions of these models were tested prior to presenting the results of the analyses. As the methodology is an extension of multivariate logistic regression, many of the diagnostic tests remain the same (Hox, 2002; Kreft and De Leeuw, 1998; Snijders and Bosker, 1999).

Checks for linearity and homoscedascity were first performed. These were conducted using the predicted values of *y* from each of the models (Tables 6.6 to 6.8). The variables included all appeared linear and therefore are fine for inclusion in the model (results not shown). Multi-collinearity was then tested through producing a correlation matrix of all the explanatory variables (results not shown). Associations were mainly weak, with the strongest being 0.37. Therefore this is not an issue within the data.

The last assumption tested for involves assessing the impact of outliers (i.e. extreme values) within the dataset, since these can affect the results gathered (Snijders and Berkhof, 2007). Residuals were created based upon the difference between the predicted and actual value of *y* (from each model). These were standardised by z-scores to assess whether they lie outside of the expected range of values (the rule of thumb used is ±3; Rogerson, 2006). The range of values across all the models was -2.08 to 3.4, showing that the extreme data points were only for cases which the model predicted poor health when there was not. Only 0.98 per cent cases were classified as outliers, which is lower than would be expected with a normally distributed data set.

The outliers from their respective models were excluded, with the analysis re-run. There was little effect on the all of the models, experiencing a small improvement from the loss of data points unrepresentative of the model conditions. The individual estimates also improved slightly both in strength and significance due to a similar effect. Examining the outliers shows that, on average, they were young males with high incomes, living in good areas who reported their health as poor. These are low risk

groups (Congdon, 1996) which would not be expected to have poor health. However, the overall model remained stable, showing the results to be useful.

The issue of centering is also important within multi-level modelling. Centering is the transformation of predictor variables along different scales to improve the interpretability of the model, mostly through the understanding of the intercept term (Hox, 2002; Snijders and Bosker, 1999). This can be achieved through two different methods based around either using the grand or group mean.

The grand mean transforms individual observations based upon their relation to overall mean, whilst the group mean method uses the mean for all observations by their respective level-2 location (Snijders and Bosker, 1999). The choice depends mostly upon theoretical considerations. However, there is little evidence regarding which is better and centering can significantly affect the relationships found unexpectedly, making it somewhat a subjective decision (Paccagnello, 2006). Therefore for this analysis, no centering will be applied (making it a 'raw score model').

6.2.4 Theoretical considerations on neighbourhood size

An issue that should be considered is whether the MSOA level is too large to truly capture and test for area effects. As size increases, variables begin to regress to the mean and any effects may become less pronounced. However exploring the number of quantitative studies analysing this topic would suggest the contrary. Pickett and Pearl's (2001) review of neighbourhood effect studies found that evidence of effects was found at a variety of geographical scales and sizes. Andersson and Musterd (2010) found that effects were stronger at lower levels, although most tests for every level were significant. Electoral wards have been used most often in the past within similar studies (Flowerdew et al., 2008). These represent similar sized areas to MSOAs (see Section 3.4), suggesting that MSOAs should not be much of a problem.

Areas can be too small as well. Having smaller areas may be more accurate in terms of area characteristics and their associated effects. However it can lead to biased data estimates, as there becomes fewer cases per area. As such, it becomes more important where the lines of the area are drawn in the results produced (Flowerdew et al., 2008). This use of MSOAs reflects the compromise between scale and data.

The use of MSOAs is due to the data used. The classification was designed and created for the MSOA level and hence this is the only choice of areas to use. It is the only means for evaluating its usefulness. However data limitations often restrict most neighbourhood studies analysis, forcing researchers to test hypotheses at levels outside of their theoretical framework due to data availability (Flowerdew et al., 2008).

MSOAs are also less arbitrary geographical levels. They were designed to be socially homogenous, being built from LSOAs (ONS, 2011). The classification was designed to capture the mortality pattern of small areas and therefore using geographic units which are fairly socially homogenous is useful in this analysis. The social and demographic patterns captured indirectly through the mortality classification was shown in Chapter Five. The analysis showed the different neighbourhood characteristics and conditions captured which will be useful in the analysis.

6.3 Descriptive comparisons between the classification and GORs

To investigate whether a multilevel analysis of area effects would be useful in this new data set, the percentage of people reporting 'poor' and 'good' health was calculated for both the clusters and the GORs (Tables 6.2 and 6.3). Exploring the initial differences in health between the two grouping options allows an initial examination of how effective either approach is to discriminating variations in health.

Table 6.2 shows the split of self-reported health between each cluster. Clusters which displayed worse mortality profiles (i.e. higher mortality rates across the range of causes) also had greater proportions of people reporting poor health in them. This would indicate that self-reported health is a useful measure for capturing the geographical divisions in mortality patterns.

Table 6.3 presents the percentage of people reporting either health option split by GOR. GORs in Southern areas contain a lower proportion of poor health than compared to those Northern areas (with the North East particularly high). Wales also reports the third highest percentage of poor health reported, highlighting the notion that the Welsh report their health worse than it actually is (Dorling and Barford, 2009).

Cluster	Poor Health	Good Health
Best Health and Most Desirable	24	76
Average Mortality Profiles	33	67
Good Health Areas	30	70
The Middle	34	66
Poor Health Experiences	34	66
Poorest Health and Least Desirable	40	60
Poorest Neurodegenerative Health	34	66
Mixed Experiences	28	72
Total	31	69
Range	16	16

Table 6.2: Differences in the percentages of self-rated health within the classification.

GOR	Poor Health	Good Health
North East	37	63
North West	30	70
Yorkshire and The Humber	33	67
East Midlands	30	70
West Midlands	31	69
East of England	26	74
London	25	75
South East	25	75
South West	28	72
Wales	32	68
Total	31	69
Range	12	12

Table 6.3: Differences in the percentage of people reporting their health as good or poor within in the BHPS split by Governmental Office Regions of England and Wales.

Comparing Tables 6.2 and 6.3 allows an assessment between the two measures. The range shows the distribution of values across each measure, providing a comparable statistic of the information captured by each. The range of values for the percentage of poor health across the GORs is 12 percentage points. However for the classification, there is a wider range of values (16 percentage points). The classification is a more discriminating measure, accounting for a third more detail in variations in poor health.

The lower amount of variation in health patterns captured by the GORs reflects their construction. GORs are large administrative areas that contain a variety of areas (Dorling, 2012; Vickers, 2006). As such, there is a mixture of both poor and good health within each (APHO, 2013; Doran et al., 2004; Woods et al., 2005). The area classification separates out these types of areas, allowing a more accurate analysis through capturing the important processes which influence health. GORs fail to do this, only providing a regional description of the data.

This was reflected in Tables 5.20 and 5.21. Calculating life expectancy, variation in values was higher within GORs than compared to within the clusters. It highlighted the range in types of areas found within each GOR, whereas the classification was more evenly spread, in comparison, throughout England and Wales (also see Table 5.19). Although this is rather simplistic, given that this is how morbidity and mortality statistics are generally reported (i.e. using GORs), the classification is a better fit for this purpose to improve understanding of divisions in health.

The usefulness of the classification is further shown due to the GORs containing two more categories. This should otherwise capture greater variation within the data, through being able to distinguish between different groups in the data. To ensure that the comparison is fairer here, the range value for GORs was multiplied by 0.8 to account for the two extra categories. This gave a value of 9.6, showing the classification to capture 40 percent more variation in the data. The classification is more useful for discriminating the data in terms of health

It could be argued that due to the segmentation process, as areas based upon mortality are clustered together, this difference may be manufactured by the clustering of certain types of people. As such, it may have artificially higher variation captured. With Chapter 5 showing that the clusters varied in their membership by age distinctly, this factor could easily be driving this difference. Nevertheless this issue also applies to GORs, as there are also slight variations in age regionally (Dorling, 2012).

To account for this, the age and sex standardised expected number of people reporting their health as poor for both the clusters and the GORs were calculated individually using the BHPS data. This is similar to the construction of standardised mortality ratios (SMRs), where a value of 100 is average and values above and below represent the percentage difference from this expected average (see Section 3.7.4).

Cluster	AS-SP	Population in the BHPS (%)
Best Health and Most Desirable	81.3	20.1
Average Mortality Profiles	113.6	15.7
Good Health Areas	101.2	18.2
The Middle	119.0	13.0
Poor Health Experiences	127.5	9.5
Poorest Health and Least Desirable	140.2	5.4
Poorest Neurodegenerative Health	118.4	3.9
Mixed Experiences	92.7	14.2
Range	58.9	16.2

Table 6.4: The age and sex standardised percentage (AS-SP) of actual against expected numbers of people reporting their health as poor between each cluster in the classification.

GOR	AS-SP	Population in the BHPS(%)
North East	132.2	3.7
North West	106.8	10.8
Yorkshire and The Humber	116.6	8.6
East Midlands	110.9	7.5
West Midlands	112.6	7.4
East of England	92.0	8.7
London	89.5	6.4
South East	90.1	12.1
South West	94.1	8.4
Wales	106.0	26.5
Range	42.7	22.8

Table 6.5: The age and sex standardised percentage (AS-SP) of actual against expected numbers of people reporting their health as poor between the GORs.

Tables 6.4 and 6.5 show the proportion of the total number of people reporting their health as poor as a ratio of what is expected given the age and sex make-up of that group (i.e. cluster or region). The general patterns follow those seen in Tables 6.2 and 6.3, showing that age and sex had little effect on the findings. The range of values captured by the classification is still greater than when the GORs were used, despite containing fewer groups. The classification captures a slightly higher amount of detail

through the relative difference in the range when the data is no standardised as well (38 per cent). Accounting for the extra groups, sees the revised range for the GORs being 34.16, an increase in the relative difference to 72.4 per cent. The classification is a better discriminating measure even after accounting for age and sex characteristics.

Population was also included in each table. Whilst it has been demonstrated that the classification has greater discrimination within the data, this could be influenced by the smaller cluster sizes, producing biased results through capturing smaller extreme sub-groups with greater differences in health. However comparing the two shows that this does not appear to be a complete explanation of the pattern, with small sizes in the GORs as well (indeed the North east is the smallest category across both measures). There is still a fairly large proportion of the sample within the clusters demonstrating poor health. The GORs offer a more even regional spread throughout just England, with the larger proportion of Welsh respondents a factor of the region receiving a booster data set (and why the range for England and Wales as a whole is larger for the GORs).

6.4 A multi-level analysis

This section extends the current examination of health differences through analysing the risk of poor health in a multi-level analysis. The focus of this approach is to assess whether if the area type (cluster) that individuals live in has an independent effect upon health. The existence of area effects using the classification is one direction which would show how useful it is as a research tool. The results are replicated using GORs instead to provide a reference set of results to be able to evaluate the performance of the classification.

6.4.1 Applying the classification

Firstly, the clusters were used as the level-2 unit in the multi-level model (with the level-1 unit being the individual). The dependent and independent variables detailed previously (Sections 6.2.1 and 6.2.2), which act as controls for the main social processes that affect health set out in the literature review (Section 2.3). As such, a logistic regression model was fit for the individual level data, with the intercept being

allowed to vary for each cluster (leading it to be a multi-level model). If the clusters are important containers of effects on health, then the model intercept and pattern of the data would differ by each cluster. The model was fitted in stages following the analytical framework discussed in both Andersson and Musterd (2010) and Hox (2002).

The reporting of results from a multi-level analysis begins with the unconditional model (Andersson and Musterd, 2010). Only the intercept is fitted in the model to explain self-reported health, allowing it to vary randomly by the level-2 variable (Hox, 2002; Kreft and De Leeuw, 1998). This shows whether the level-2 variable is important for explaining differences in self-reported health within the data, without controlling for any other covariates.

With only the intercept fitted, there are few results to present. The -2 Restricted Log Likelihood (-2 RLL) for this model is -5921, showing the pseudo-deviance of the model (Snijders and Bosker, 1999). It gives a value which represents how well the data fits the model. However by itself, it is meaningless (rather it is important for evaluating relative changes in a model and hence is stated only for future comparisons).

Table 6.6b presents the important part of the results from this model, showing the between-group variance accounted by the level-2 variable. This shows the degree of difference between the clusters that accounts for variations in health status (Hox, 2002). This result was significant, showing that which cluster individuals live in has an effect upon their health. This would indicate that the classification is a useful measure.

Calculating the ICC (Intra-Class Correlation) shows the percentage of variance accounted between groups as a total of all variance (i.e. the total of within and between group variance; Hox, 2002). This gives a percentage figure of how well the clusters account for differences in health status. It shows that only 1.2 per cent of the variance in self-reported health is accounted by differences between clusters.

Compared to past analysis, studies analysing the size of area effects tend to report low levels of variance for the level-2 objects as explanatory variables. The analysis presented here shows though that the classification generally accounts for a lower proportion of the variance than other studies have shown. The classification does not add much detail of the data based upon this type of analysis. Nevertheless it does display a significant, albeit small, effect on health.

(a) Fixed effects model

Variable	Model 1		Model 2	
	OR	SE	OR	SE
Age	1.024***	0.001	1.025***	0.001
Male	1.021	0.052	1.021	0.052
Income	0.99999***	2.05E-06	0.99999***	2.05E-06
Trust	0.689***	0.038	0.699***	0.039
Civic	0.791***	0.048	0.794***	0.048
Meet	0.874*	0.059	0.858*	0.058
Belong	0.711***	0.042	0.726***	0.043
Advice	0.998	0.055	0.999	0.055
Poverty			1.029***	0.004
Turnover			0.999	0.002

(b) Random effects parameters

Model	Variance	SE	P	ICC
Unconditional	0.04	0.023	<0.001	0.012
1	0.027	0.0162	<0.001	0.008
2	0.001	0.003	>0.999	0

Table 6.6: Results from the multi-level logistic model using the classification as the level-2 unit.

*Key: OR = odds ratios, SE = standard errors, P = significance, ICC = Intra-Class Correlation. Significance levels: $p < 0.05$ = *, $p < 0.01$ = **, $p < 0.001$ = ***.*

The small effect size may be explained by methodological issues. The classification only has eight types of areas within it. This is fairly small and could be affecting the ability for a significant proportion of variance to be significantly captured at the second level of analysis (Hox, 2002; Bell et al., 2010). A lack of level-2 sample size is one of the most common reasons why these types of studies fail (Snijders and Bosker, 1999). This is especially the case given that both Tables 6.2 and 6.4 showed the classification to effectively capture differences within the data.

This does not mean that a model should not be fitted. It is important to fit the correct model within the specified conceptual and analytical framework. Health is influenced

by both individual and area level risk factors (Macintyre and Ellaway, 2000; Pickett and Pearl, 2001; Riva et al., 2007). Ignoring this multi-level structure to how health is affected would falsely account for such relationships between variables. Furthermore there have been examples where multi-level models with a similar number of level-2 units were used, showing it to be a useful analytical approach. For example Langford and Bentham (1996) used the ACORN classification of Local Authorities in their analysis. They found significant area effects despite only having nine clusters at their level-2 unit.

The individual level fixed effects were then introduced into the analysis (Model 1). There are few statistics to assess the quality of a multi-level model. A statistic to look at is the relative change in the -2 RLL figure, to see whether adding certain conditions and variables improve the model (Snijders and Bosker, 1999; Hox, 2002). This allows comparisons to be made within model, albeit not to models with different outcome variables or level-2 units (and therefore cannot be used to compare to using GORs). The -2 RLL value was -5038, a decline of 15 per cent. The individual level variables are a useful addition to understanding the differences in self-rated health.

The level-1 variables mostly behaved as expected based upon the literature review (see Section 2.3; Table 6.6a). The age variable showed that increasing age was associated with a greater probability of poor health. Gender, however, was not significant, producing a result contrary to other research (Lindeboom and van Doorslaer, 2004). Income gave a significantly negative result, where those with higher incomes were less likely to report their health as poor.

The majority of the variables representing social interactions showed their beneficial effects on health (Kawachi et al., 1999). The variable for whether an individual got advice from the local community was the only one of these variable types to be non-significant, not being a factor in influencing health directly. This effect is likely to have been accounted for by the other social relations variables. The result may also be due to the variable being poorly defined, as it does not cover only health advice.

Table 6.6b shows the variance accounted by allowing the clusters to vary randomly. Even after controlling for personal risk factors, the area that individuals' lived in still produced a significant impact on health. However the size of this effect has diminished (partly accounted for by the individual variables), being less than one per cent, showing its contribution to health to be only small.

Kreft and De Leeuw (1998) offer a statistic that can allow a quick evaluation of model change. They suggest calculating the percentage of variance explained at each stage of the model through reporting the change in the unexplained variance by the total variance reported at the unconditional model. This shows how each stage has accounted for the effect of the level-2 variable. However, this measure can be problematic, as the statistic could show negative variance explained which is not strictly true (ibid).

Calculating this, gave a figure of 32.5 per cent. This shows that the introduction of the individual level variables decreased the effect that the clusters had on health by just under a third of its total effect. This is a fairly large effect size for accounting for between area variance through the introduction of variables largely unrelated to areas (i.e. just individual characteristics). However when the classification is accounting for such a small amount of the total variance, this is not particularly insightful (highlighting the small effect of individual level factors on area level processes).

The final model introduced the area level variables covariates the analysis (Model 2). The -2 RLL value was -5015, representing only a small model improvement, suggesting that they were less useful for explaining the variation in health in the data. There was little change in the individual level variables, with all remaining within their respective confidence limits (Table 6.6a). It is not surprising that the area level variables had little impact upon the individual level variables, given that they represent independent processes.

Population turnover was insignificant, a result contrary to that found elsewhere (Chaix et al., 2007). Neither variable which examined social support has shown any significant association towards how people view their health. It would appear that this mechanism is indirect (Berkman and Glass, 2000). The role of social relations on health appears to only exist at the individual level.

The poverty measure was highly significant. A one per cent increase in the percentage of households classed as in poverty also increased the probability of people reporting their health as poor by 2.9 per cent. It shows that similar processes of lack of resources captured at both the individual and area levels co-exist, and are independent of each other, suggesting different mechanisms (Evans and Kim, 2007).

Examining the random effects parameters (Table 6.6b) shows that the effect accounted by the classification has now become non-significant. There is little variance left which is explained by the classification (the figure is reported to three decimal places). It

would appear that poverty accounts for the effect of the classification on health. This is an important finding, showing the poverty is both an important dimension of area effects, as well as explaining the segmentation that occurs through the classification.

6.4.2 Introducing GORs as the level-2 unit

The analysis was re-run using the GORs as the level-2 variable instead. This is important to be able to effectively assess the results from when the classification was used. With the small unit size issue possibly hampering the evaluation of the classification, comparing it to a similar sized measure which is used to analyse and disseminate mortality statistics is required to assess the importance of the classification as a useful analytical measure. The results are presented in Table 6.7. Whilst some statistics are presented here, they are only used as guidelines since they do not (methodologically) apply to comparing models.

The unconditional model was first run, producing a -2 RLL of -5935. Table 6.7b shows that this model resulted in a significant effect observed for the GORs as the level-2 unit. The region of England and Wales that individuals lived in had an effect on their health. Similarly to the classification, only a small effect size was reported, with the ICC value showing that the between group variance accounted for less than one per cent of the total variance in the model. Region only has a small impact upon health. Importantly, this effect size was smaller than for the classification at the same stage.

'Model 1' and 'Model 2' represent the same model stages as described previously (Section 6.4.1). The change in the pseudo deviance measure was relatively the same as when the classification as used (declining by 15 per cent for 'Model 1' and then a further one per cent for 'Model 2'). The fixed effects added as control variables also display similar odds ratios, with no difference in the conclusions drawn from them. This is useful, as the results are less model dependent as they are capturing the main patterns in the data. They are important controls for effects on health for assessing the level-2 variable.

The difference between the use of the GORs and the classification in the analysis lies in the random effects parameters. The introduction of the individual level covariates saw the variance captured between GORs increase, with the ICC showing a rise from 0.9 per

cent to 1.1 per cent. This resulted in the GORs capturing a greater proportion of the variance than compared to the same stage when the classification was used.

(a) Fixed effects model

Variable	Model 1		Model 2	
	OR	SE	OR	SE
Age	1.024***	0.001	1.025***	0.001
Male	1.021	0.052	1.014	0.052
Income	0.99999***	2.05E-06	0.99999***	2.05E-06
Trust	0.682***	0.038	0.698***	0.038
Civic	0.775***	0.047	0.798***	0.048
Meet	0.859*	0.058	0.845*	0.057
Belong	0.692***	0.041	0.723***	0.043
Advice	0.99	0.055	0.988	0.055
Poverty			1.029***	0.004
Turnover			0.999	0.002

(b) Random effect parameters

Model	Variance	SE	P	ICC
Unconditional	0.03	0.017	<0.001	0.009
1	0.036	0.021	<0.001	0.011
2	0.025	0.016	<0.001	0.008

Table 6.7: Results from the multi-level logistic model using Governmental Office Regions as the level-2 unit.

*Key: OR = odds ratios, SE = standard errors, P = significance, ICC = Intra-Class Correlation. Significance levels: $p < 0.05$ = *, $p < 0.01$ = **, $p < 0.001$ = ***.*

This pattern continued once the area level variables were added to the model, with the GORs still reporting a significant effect (with the ICC value showing that their introduction declined their effect to 0.9 per cent). The equivalent model using the classification gave a non-significant effect on health, showing the GORs to be more important here. Unlike the classification, the regional effects observed through the GORs are independent of poverty in this type of analysis.

A comparison of the results from this methodological approach shows that the GORs appear to be more useful for capturing geographical effects on health. Even after accounting for individual and area risk factors, the region of England and Wales where individuals lived had a significant and independent effect on health, unlike the classification. However the absolute difference between the ICC measures at each stage of the multi-level models was only ever particularly small. As such, it may be that the results do not really matter much. This is especially the case since both variables account for very small proportions of the variance in health at each stage of the model.

The significance of the model using the GORs after controlling for individual and area level risk factors could be due to methodological differences. A small sample size of level-2 variable is an important factor in determining the significance of a model (Bell et al., 2010; Snijders and Bosker, 1999). The GORs contain two extra groups which aid in discriminating a greater level of variance in health of individuals in the data set. Although in absolute terms, this does not appear large, due to the small numbers involved, the relative difference is 20 per cent showing it is an important factor. As such, the significant effect captured by the classification could be unobserved. This is especially the case given that both Tables 6.2 and 6.4 showed the classification to effectively capture differences within the data. From this methodological approach, a comparison of the two factors is not fair.

6.5 MSOAs as containers for neighbourhood effects

The results so far show only small evidence of area effects from the two measures. These are, however, smaller than the evidence shown throughout the literature (Pickett and Pearl, 2001; Riva et al., 2007). Methodological restrictions have limited the ability to accurately test the effect, constraining the true impact of neighbourhoods. Instead it would be better to use the areal units themselves (i.e. the MSOAs) as the level-2 unit to more effectively explore the impact of areas on health.

MSOAs are the containers which area effects are proposed to occur in (within in this study) and therefore are conceptually better for testing effects. This falls more in line with the majority of multi-level research into this topic (Flowerdew et al., 2008), which tend to use the areas themselves, rather than area types, in their analyses. With a larger degree of units at the second level (2257), if there are any area effects observed then this

approach should capture them (as it will be more statistically robust; Snijders and Bosker, 1999; Riva et al., 2007).

Previously the clusters were included as random effects, however this may have mis-specified the model specification (possibly a reason in their lack of significant effect on health). The clusters could have been added as fixed effects instead, since they cannot be regarded as a random sample from a wider population of units (Kreft and De Leeuw, 1998; Snijders and Bosker, 1999). As all the areas of England and Wales were included in the classification, using the clusters as fixed factors would be statistically sound (Hox, 2002). It helps to understand what the clusters consist of through exploring the effect of controlling for particular factors on the each individual clusters, rather than as a whole. Whether each has an effect on health as a fixed effect can also be explored.

This approach was applied in Nnoaham et al. (2010), who analysed colorectal cancer screening uptake of individuals, with LSOAs as the level-2 unit and the 'People and Places' geodemographic classification clusters as fixed effects. Particular area types were more or less likely to uptake the service and therefore the analysis was useful for the targeting policy (for more detail, see Table 2.6).

This would also account for the small sample issue regarding the level-2 variable. The results of Andersson and Musterd (2010) and Riva et al. (2007) would appear to indicate that where a greater number of areas are used in the analysis, a stronger area effect is also observed (beyond a simple function of power size due to low numbers). Therefore allowing the unit of analysis to increase will present a fairer analysis of the existence of area effects. This is also important for justifying subsequent analyses (Chapter 7).

The issue with using the MSOAs as the primary focus of analysis is that they are only useful in this example. Since there are 7194 different areas, their application to research is limited. The classification benefits through its simplification of patterns and interactions across a multitude of factors (Openshaw et al., 1994). This allows for a better understanding of patterns and processes occurring across England and Wales, which would otherwise become confusing if MSOAs were used only. It is likely that the MSOAs will have a larger effect on health simply because there are far more areas, which will capture a greater level of detail in the data. As such, it is better to apply the clusters as fixed effects and utilise the MSOAs to explore the concept of area effects as well so that each complements the other best.

An issue that arises from this change is that as the number of areas increases, the share of cases between them will fall as a result. This was not an issue when both the classification and the GORs were used, however the MSOAs have far fewer individuals per area. Where there are many areas which only contain one case (i.e. an individual) in them, this can become problematic. This is because they are adding nothing to the variance of the model, since the area effect cannot be prominent or consistent. Similarly low numbers of cases per area also limit the power of the analysis (Bell et al., 2010). Investigating the extent this is an issue through a sensitivity analysis will give us an indication of whether the results from the analysis are useful.

6.5.1 Analysis

With the MSOAs used as the level-2 unit, the unconditional model was fitted. This gave a -2 RLL value of -5911. Table 6.8b shows that the MSOAs displayed a significant effect on self-rated health, showing evidence of neighbourhood effects. The area people live in has an effect on health. The MSOAs also captured a higher level of variance, with the ICC value showing that differences between MSOAs accounted for eight per cent of the total variance in the model. It is a useful spatial scale in this analytical investigation of area effects.

The first model fitted with explanatory variables included the clusters as dummy variables (Model 1). The first cluster was used as the reference cluster, since both its mortality profile and average score for self-reported health were at one extreme of the scale, making it useful for comparisons to other clusters. The -2 RLL value here is -5881, a decline of 0.5 per cent, suggesting that their introduction was not particularly useful in explaining patterns in self-rated health of individuals.

Examination of their odds ratios, however, shows that each dummy is significant. The differences in odds ratios between the clusters follow their respective mortality profiles, with larger effects where clusters performed worse. There are particularly large effect sizes across the clusters. For example those who resided in the 'Poorest Health and Least Desirable' cluster were more than twice as likely to report their health as poor than compared to those who reside in the 'Best Health and Most Desirable' cluster. Although these represent the two clusters at the extremes of the classification, the smallest odds ratio was 1.28, which is still a fairly large effect. The classification is

useful for capturing differences, appearing to operate as fixed effects rather than random effects.

The model also continues to display a significant random effect. The between group variance attributed to letting the MSOAs vary randomly has fallen by 22.4 per cent. Whilst the introduction of the classification only reported a small decline in the pseudo-deviance, it was useful for accounting for differences between MSOAs. The classification was useful for geographical patterns in poor health, as well as an individual fixed effect. As such, the ICC shows that MSOAs now account for 6.3 per cent of the differences across the whole data set.

The next stage of the analysis was to introduce the individual level variables to control for their effects (Model 2). The -2 RLL fell to -5009, 15.3 per cent less than the figure for the unconditional model showing that their introduction was useful. The covariates had little change compared to past models, being found in the same direction and of similar strength. None of the significant odds ratios changed greater than their confidence intervals from the previous models (Tables 6.6 and 6.7), showing their relationships to be distinct in the data. The classification variables all remained significant, showing them to account for greater patterns beyond individual level factors. They lost some strength, a factor of the individual level variables accounting for some of their effect.

The variance accounted between MSOAs has also remained significant. The figure shows that it has decreased by 29.7 per cent since the unconditional model. This is not much more than at the previous stage of the model, although since the variables introduced are not measuring area level factors this is not surprising. The percentage of variance accounted for between the MSOAs is still fairly large, at 5.8 per cent

The final model added the two area variables into the analysis (Model 3). There was a small decrease of the -2 RLL statistic to -4995, suggesting that their introduction was not particularly useful. The two area level variables were found to be similar to the previous analysis and there was little change in the personal risk covariates either.

(a) Fixed effects model

Variable	Model 1		Model 2		Model 3	
	OR	SE	OR	SE	OR	SE
'Best Health and Most Desirable' reference for clusters						
Average Mortality Profiles	1.6***	0.144	1.491***	0.144	1.248*	0.126
Good Health Areas	1.287***	0.114	1.248*	0.118	1.136	0.108
The Middle	1.582***	0.152	1.57***	0.161	1.202	0.136
Poor Health Experiences	1.618***	0.172	1.548***	0.176	1.131	0.143
Poorest Health and Least Desirable	2.146***	0.28	1.869***	0.258	1.317	0.2
Poorest Neuro-degenerative Health	1.616***	0.228	1.61**	0.241	1.314	0.202
Mixed Experiences	1.19***	0.112	1.236*	0.124	1.136	0.115
Age			1.026***	0.001	1.026***	0.001
Male			1.022	0.534	1.02	0.053
Income			0.99999***	2.11E-06	0.99999***	2.11E-06
Trust			0.69***	0.398	0.698***	0.04
Civic			0.797***	0.05	0.802***	0.05
Meet			0.86*	0.06	0.849*	0.06
Belong			0.703***	0.432	0.717***	0.044
Advice			1.007	0.058	1.002	0.058
Poverty					1.025***	0.005
Turnover					0.999	0.003

(b) Random effect parameters

Model	Variance	SE	P	ICC
Unconditional	0.286	0.049	<0.001	0.08
1	0.222	0.045	<0.001	0.063
2	0.201	0.048	<0.001	0.058
3	0.179	0.046	<0.001	0.052

Table 6.8 Results from the multi-level logistic model using Middle Super Output Areas as the level-2 unit.

*Key: OR = odds ratios, SE = standard errors, P = significance, ICC = Intra-Class Correlation. Significance levels: p < 0.05 = *, p < 0.01 = **, p < 0.001 = ***.*

Only one of the clusters remained significantly different from 'Best Health and Most Desirable', after accounting for individual and area level risk factors. This supports the previous models which showed how poverty patterns explained the variations between clusters. Differences between the clusters are mainly a product of disparities in the level of poverty, which once accounted for, present no significant differences between area types.

Nevertheless those who lived in the cluster 'Average Mortality Profiles' were 24.8 per cent more likely to have reported their health as poor than compared to the reference cluster. This is different to the past analysis, where it was assumed that poverty attenuated the effect of the classification. However the effect of poverty is clearly not consistent. The cluster is measuring an effect that we cannot otherwise account for, unlike the other clusters. Clearly, location and area type is important to exploring health patterns and inequality.

Examination of the characteristics of the clusters ('Best Health and Most Desirable' and 'Average Mortality Profiles') in Chapter 5 helps to aid the understanding of why this is. There are not huge social variations between the clusters. What is being reported here instead are differences in health captured in the data which are beyond what would be expected once these social factors are controlled for. Despite being similar, the classification is capturing additional (and different) disadvantage in those areas part of the cluster 'Average Mortality Profiles' when compared to 'Best Health and Most Desirable'. The classification is useful, adding value as a research tool. Furthermore, introducing them as fixed effects rather than random effects was important since the attenuated effect of poverty is not completely consistent.

There is still a significant effect captured by MSOAs. Controlling for individual and area risk factors, as well as the classification, only sees a decline of 37.4 per cent in the variance from the unconditional model. This is quite a large decline and since population turnover was insignificant, it highlights the importance of poverty in explaining patterns.

The neighbourhoods themselves exhibit a fairly large effect, with the ICC statistic showing that MSOAs account for 5.2 per cent of the total variance, after accounting for individual and area risk factors. This level supports past research (Pickett and Pearl, 2001; Riva eta l., 2007). Neighbourhoods (measured here through MSOAs) have a significant and large impact on health. Indeed given their large geographical size, the

true effect size may be underestimated. Geography is paramount in health research, especially within a policy context as these results highlight the need for co-ordinated action between both levels (Chaix et al., 2007; Lupton, 2003).

6.5.2 A (pseudo-) sensitivity analysis

With MSOAs containing far more areas than compared to the classification and the GORs, the issue arises of an uneven spread of cases between areas. Where there are few cases per area, this can affect the performance of the MLM since areas with only one individual cannot add anything to the random effects part of the model (Bell et al., 2010; Hox, 2002). Few points give poor estimates. Explaining the impact of this is important for the generalisability of the results.

Figure 6.1 shows the frequency distribution of the number of cases per MSOA. The pattern is strongly positively skewed (a Poisson distribution), which appears problematic. You would expect a higher frequency in the smaller values, as the BHPS aims to provide a representative sample and thus is less likely to concentrate in collecting data in certain areas. There is a long tail showing a wide range of values, a shadow of the sampling strategy employed in the BHPS.

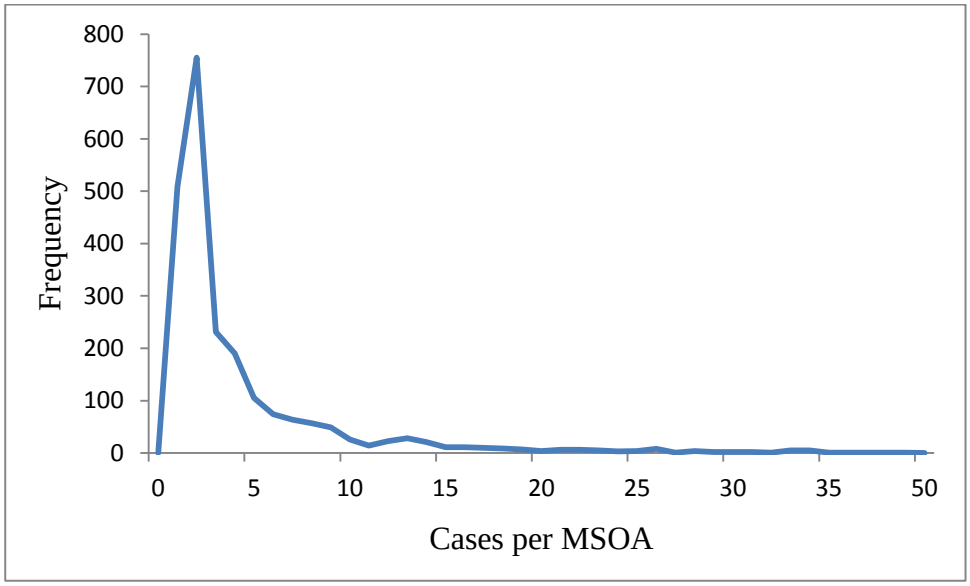


Figure 6.1: The distribution of individuals by MSOAs within the BHPS.

The mean number of cases per MSOA in the data was 4.3. However a rule of thumb would suggest that you should not have a mean of less than five (Bell et al., 2010), suggesting that this may be problematic. 22.6 per cent of MSOAs only contained one

case and 74.7 per cent of areas had fewer than five cases within them. With this being fairly prevalent, any effects observed using this variable are likely to be under-estimating the actual total effect. Nevertheless as Bell et al. (2010) find, as long as the number of level-2 units is large, a small number of cases per unit should be fine for analysis.

To assess whether the large number of MSOAs containing few data points was having an influence on the model and results, a pseudo-sensitivity analysis was performed. Two models were run and the subsequent results were compared to that of the final model (Table 6.8). The first model included identifying those areas which only contained a single case and removing them (and their respective cases), before re-running the analysis. The second was more stringent, excluding the data for those areas (and cases) which contained fewer than five cases. The results are presented in Table 6.9.

The removal of all cases that were the only individual within their MSOA saw the mean number of cases per MSOA rise to 4.8. This small improvement reflects the large number of cases still lower than the mean (52.1 per cent of the MSOAs contained only two to four cases). The -2 RLL figure for this model was -4734, showing the large improvement in the understanding of health status through the removal of those single case MSOAs (and their cases as well). The level of variance captured by the model remains fairly similar to the final model, being only slightly smaller.

All of the individual and area level control variables remained similar, not changing beyond the confidence limits of the final model. The only difference is seen in the classification dummy variables. The dummy variable for the cluster 'Average Mortality Profiles' has now become insignificant, following the same pattern for the rest of the clusters. This would show that additional information it was capturing as suggested in the previous section, is not particular strong.

The next model removed all data which belonged to a MSOA that contained less than five cases within it. This led to an increase of the mean number of cases per area to 9.7, a more useful sample for a multi-level analysis. Running the model saw a large improvement in the -2 RLL value to -3234, although this is not surprising given the large amount of data excluded. There was little change in the explanatory variables, although the cluster 'The Middle' has become significant, showing that individuals

which lived in this cluster compared to the cluster 'Best Health and Most Desirable' were 35.3 per cent more likely to report their health as poor.

(a) Fixed effects model

Variable	Cases per MSOA > 1		Cases per MSOA > 4	
	OR	SE	OR	SE
'Best Health and Most Desirable' reference for clusters				
Average Mortality Profiles	1.172	1.22	1.229	0.158
Good Health Areas	1.091	0.108	1.26	0.153
The Middle	1.156	0.14	1.353*	0.192
Poor Health Experiences	1.124	0.147	1.313	0.204
Poorest Health and Least Desirable	1.269	0.199	1.399	0.253
Poorest Neurodegenerative Health	1.257	0.201	1.344	0.263
Mixed Experiences	1.091	0.113	1.232	0.161
Age	1.026***	0.002	1.026***	0.002
Male	1.034	0.056	1.02	0.066
Income	0.99999***	2.19E-06	0.99998***	2.96E-06
Trust	0.689***	0.041	0.677***	0.048
Civic	0.779***	0.051	0.748***	0.058
Meet	0.848*	0.612	0.802*	0.072
Belong	0.717***	0.046	0.758***	0.06
Advice	0.999	0.059	0.932	0.067
Poverty	1.027***	0.005	1.027***	0.007
Turnover	0.9996	0.003	0.999	0.004

(b) Random effects parameters

Model	Variance	SE	P	ICC
Cases per MSOA > 1	0.177	0.045	<0.001	0.051
Cases per MSOA > 4	0.115	0.038	<0.001	0.034

Table 6.9: Results from the sensitivity analysis of the multi-level model.

*Key: OR = odds ratios, SE = standard errors, P = significance, ICC = Intra-Class Correlation. Significance levels: p < 0.05 = *, p < 0.01 = **, p < 0.001 = ***.*

These fixed effects relationships are established in the data and removal of data points as part of the sensitivity analysis (since we can assume that the areas with few cases within them are random) has little impact on them. This is different to the patterns for the clusters, which are less strong as they have been accounted for by the other variables in the analysis. The significance of 'The Middle' in the second model reflects this (but the cluster 'Average Mortality Profiles' may also apply here); an artefact of the model conditions.

There was greater change in the degree of variance captured between areas. This fell by a larger proportion, with the ICC statistic showing that MSOAs accounted for 3.4 per cent of the total variance in the model. This would suggest that the size of neighbourhood effects may not be as large as previously demonstrated. There is overlap for the level of variance reported for both of these models compared to the final model, suggesting that this change may not be too important.

Overall, the sensitivity analysis has shown little variation in the results. Stability can be therefore be inferred.

6.6 Conclusion

This chapter has applied the classification to evaluate how useful it is as a research tool within a particular field of analysis. The focus was on whether the clusters displayed independent effects on health, following the area effects hypothesis. The results were compared to using GORs instead, which provided a fair and relevant measure to assess the usefulness of the clusters.

The analysis showed mixed results. The final multi-level models showed that the classification no longer significantly explained health, unlike GORs. Although this would show the classification to be less useful, the GORs included two extra units. Given the methodological limitations of such small sample sizes for the level-2 measures (Bell et al., 2010), this may be aiding the model through making it significant. This is especially the case since descriptive comparisons showed that the classification captured greater variations in self-rated health. With the GORs only accounting for less than one percent of the total variance in the model, they are not that more useful than the classification.

The multi-level model using the classification as the level-2 unit also aided the understanding of the classification. The model remained significant once the individual level factors were controlled for, showing the classification to be useful in capturing patterns independent of these. However the introduction of the poverty measure at the area level attenuated any effect captured by the classification in terms of between cluster variance. Divisions in mortality patterns found by the classification are largely a product of the underlying structure of poverty and deprivation. However this attenuation was not consistent across all clusters, showing it to add value to the analysis.

The examination of possible area effects was not completely lost. Rather, changing the model specification to use MSOAs as the level-2 unit appears more correct. Area effects were significantly and consistently found, even after accounting for possible confounding or explanatory effects. The size of this effect observed was also fairly large. Geography clearly plays an important role in understanding health.

A final consideration in posed in Riva et al. (2007), regarding 'self-selection'. Over the life course, people become 'sorted' into various different residential areas (and types) in relation to their health. We can control for certain confounders, but there still exists mis-specification, resulting in inflated findings for area effects (which are, themselves, an artefact of such processes). This links to the 'health selective migration' literature, of which there is sufficient evidence for (for example Bentham, 1988; Brimblecombe et al., 2000). If area effects are to be important, then they should still be observed once this is accounted for (i.e. when people migrate between areas). This issue forms the basis for the next chapter.

Chapter 7: Internal migration, area effects and health

7.1 Introduction

In the previous chapter, significant neighbourhood effects were identified on individual health, independent of both individual and area characteristics. A natural extension of this neighbourhood effects research would be to examine how the role of migration between areas alters this association. If areas have an effect upon the health of individuals, then as people migrate to a different area type it would be expected that this effect would also be observed.

This chapter is set out as following: It begins through discussing the methodological limitations facing a 'traditional' analysis of migration and health, proposing an alternative framework to help improve the accuracy of the findings. This is then applied to the British Household Panel Survey (BHPS), exploring the impact of migration to and from different areas on health. Firstly a demographic and social examination of the migration in the data is shown to help inform subsequent analyses. Then the role of migration between different area types is analysed, with the focus on areas as the explanatory factor for the impact on health. The chapter ends through exploring the role of health selective migration in explaining inequalities in health.

Similarly to the previous chapter, the results produced applying the classification are compared to using the Governmental Office Regions (GORs) of England and Wales (hereby referred to as a GOR). This allows a fairer evaluation and assessment of how useful the classification is when applied in a research setting.

7.2 Employing a pseudo-experimental design

Research has shown that there are distinct types of people who are more likely to migrate (Catney and Simpson, 2010; Champion, 2012; Dorling, 2012; Evandrou et al., 2010). Migration is not randomly assigned, rather selective based upon personal circumstances. Therefore the covariate distributions of the data for those who migrate

and those who do not (based upon their individual characteristics) will be different. It is not possible to simply fit a regression model to analyse the effect migration has on health, as this would break the assumption of the model that there is no selection bias present in the data (Ho et al., 2007). Put simply, results of any comparisons made between those who migrate and those who do not may just be an artefact of the differences in the group composition (as you are not comparing like-for-like). However this has been ignored in a lot of past migratory research (for example Bradley and van Willigen, 2010; Brimblecombe et al., 2000; Cox et al., 2007; Kahlmeier et al., 2001; Larson et al., 2004), other than the 'Moving to Opportunity' studies which were randomised control trials (Leventhal and Brooks-Gunn, 2003).

A new methodological approach is required in this research area to be able to conduct accurate analyses which accounts for these differences. Matching methods offer a solution to the issue. Iacus et al. (2011b) offers a definition of this approach; "Matching is a non-parametric method of controlling for the confounding influence of 'pre-treatment' control variables in observational data." (p1). Essentially you are measuring a change in status to a group of individuals through the comparison of a 'control group'. Despite the advantages of this approach, the methodology has few applications, especially within epidemiological and geographical research.

The aim of this methodology is to balance a data set to allow for a more accurate level of causal inference to be estimated. Where there is a variable of interest measuring a change in status (or participation in a program) within observational data, it allows the comparison of differences in a variable between those who changed (referred to as the treatment group) and those who did not (Blackwell et al., 2009; Iacus et al., 2011a).

The method pairs data that experiences a change in status to the rest of the data to create an equivalent control group based on a set of confounders (i.e. balancing the data), allowing fairer comparisons (King et al., 2011). The benefit of this approach is that selection bias is reduced and the 'treatment' and control variables become (or very close to) identical in relation to individual characteristics (Iacus et al., 2011b; Ho et al., 2007). There are fewer assumptions with this approach, therefore any analyses post-matching will be less model dependent and hence results will be less affected by any underlying assumptions.

This approach would appear useful here when applied to this study. Migration represents a change in status for individuals that can be tested. Between two time

points, some individuals will migrate. Therefore what this chapter will focus on is whether those who migrate have significantly different health than compared to those who did not migrate and whether this effect varies by the type of area of origin and destination. The pairing of individuals is conducted at a time point prior to anyone migrating, so that the change in health status after migration has occurred can be analysed in comparison to those who remained the same area.

The effect that this has can be seen in Figure 7.1, which presents a histogram for some pseudo data. This was chosen to emphasis the effect the model has regarding selection bias, to better understand this approach. The distribution of this variable for migrants and non-migrants is not equal, meaning that any comparisons between these two groups would be unfair. Essentially the matching process pairs individuals between the two groups so that the resulting distributions of the data across the variable are equivalent. As such, this allows fairer and stronger comparisons, since it eliminates the effects of the variables as the data is matched on to focus on the change in status (i.e. migration). This process can occur across multiple variables, matching data to fit a multi-dimensional distribution of data.

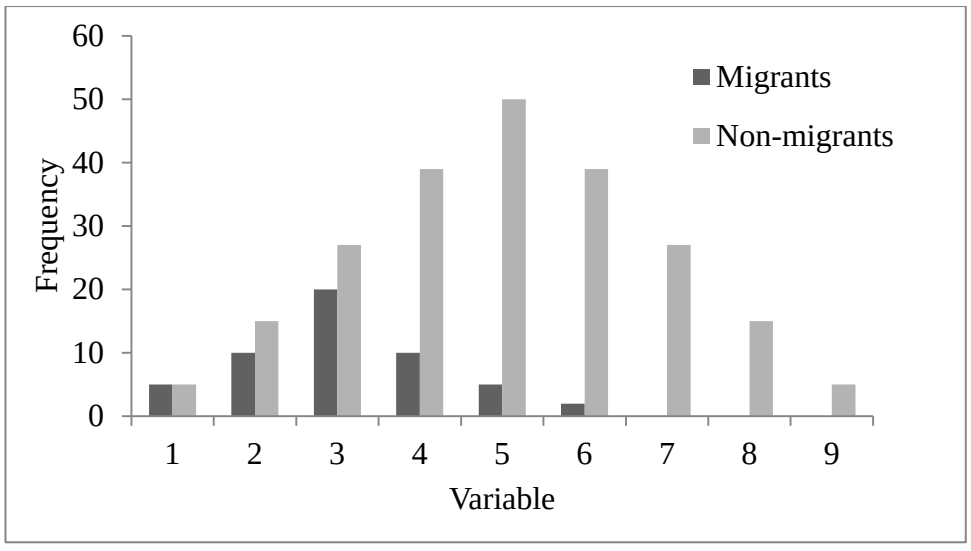


Figure 7.1: A histogram for pseudo data split by migration status.

There are two approaches to the methodology; 'one-to-one' or 'inexact' matching. Based upon a set of characteristics, one-to-one matching pairs together data points so that the 'treatment' and 'control' groups are equal in size, with all other unmatched cases discarded (Iacus et al., 2011a). However this approach involves the loss of much information, which can be problematic depending upon sample size (Ho et al., 2007).

Inexact matching solves this through attempting to match all of the data. The method remains mostly the same apart from all of the non-treatment data (i.e. those who did not migrate) are paired to their most similar case in the treatment group (i.e. the migrants). Each data point is then individually weighted based upon how 'close' it lies to its equivalent treatment case (Iacus et al., 2011b). Where the number of characteristics used to match data is greater, this becomes a more effective tool as it becomes difficult to find exact matches. However the matching of the data with this approach can be less precise. Due to the larger sample size involved, all cases are attempted to be matched. However this can lead to cases less accurately matched where they lay slightly far away. One-to-one matching would otherwise drop these, reducing any slight bias that may arise.

Matching methods represent a family of varying types of methods, all with the same aim. The main methods use 'equal per cent bias reducing' approaches through propensity scores or Mahalanobis matching (Blackwell et al., 2009; Iacus et al., 2011b). These rely on measures of similarity to assess how 'good' matches are in creating a useful control group (i.e. how close a data point is to another). However these methods are iterative and user-reliant approaches, resulting in a time laden process. They must be continuously checked to examine how good the matches are in creating a balanced data set between the treatment and control groups; with the method refined in accordance to improving the pairing of cases and the process re-estimated many times in order to effectively match a data set (Blackwell et al., 2009; Caliendo and Kopeinig, 2008). The metric to decide how close a match should be must also be decided and there is no clear indication in the literature as to what is a useful value.

Instead, the recently developed Coarsened Exact Matching (CEM) method is utilised here (see Iacus et al., 2011a). The method is more robust than the other available methods, requiring less user interaction and iterative steps (it is mostly automated), as well as satisfying all assumptions of the method (Blackwell et al., 2009). This results in the method requiring less time and processing power, as well as a reduced likelihood of user errors. Further Iacus et al. (2011b) showed the method leads to a greater reduction in the imbalance of matched data sets than other methods (i.e. the creation of a more similar control group).

Figures 7.2 to 7.5 help highlight how the method operates. A subset from the British Household Panel Survey (2006-2008) was taken and is displayed in Figure 7.2 across two variables (split by migration status). Age and income are used as variables to

match cases in this example, given that they have been shown to vary by migration status (Catney and Simpson, 2010; ONS, 2013). The sample was kept small to highlight the methodological process.

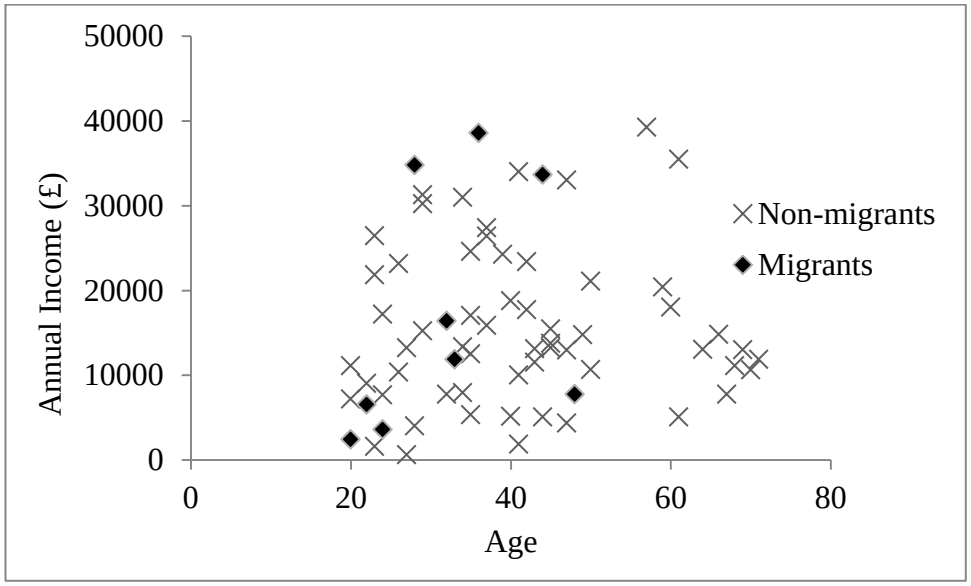


Figure 7.2: A subset of the BHPS data for age and income, split by migration status.

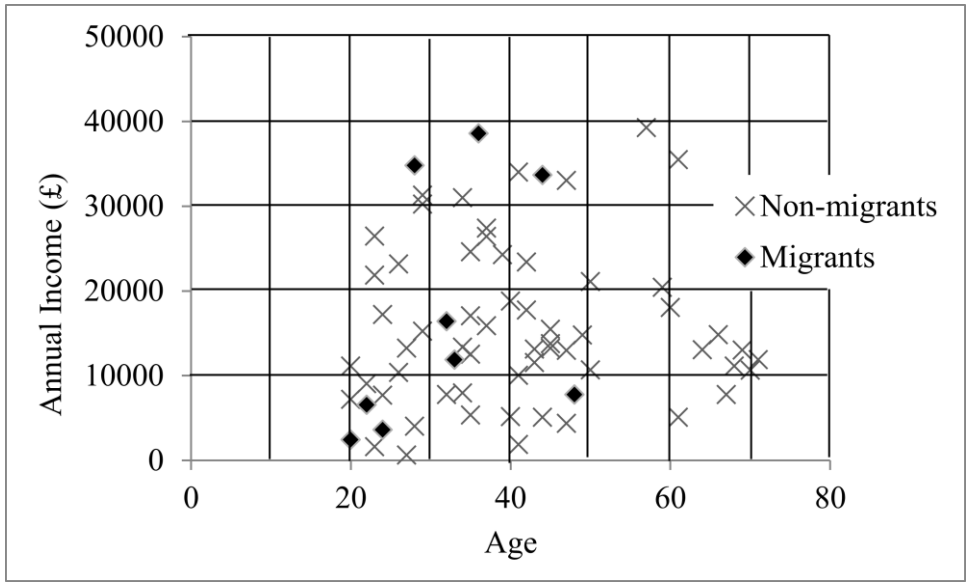


Figure 7.3: Data set coarsened into small groups for matching.

Rather than using a measure of similarity to assess which individuals to match together, variables are temporarily transformed into a series of (meaningful) categorical groups to match data more efficiently (Blackwell et al., 2009). Figure 7.3 shows these groupings, with each line representing the boundary of how a variable has been split up (in ten year bands for age and £10,000 bands for income). Due to the nature of this, the method works best when using categorical rather than continuous variables (Iacus et al., 2011a).

However the method splits up continuous variables into groups to allow them to be used. There is a trade-off between a greater number of groups providing more accurate matches, yet also being less likely to get exact matching for all cases (this is, of course, more an issue for one-to-one matching).

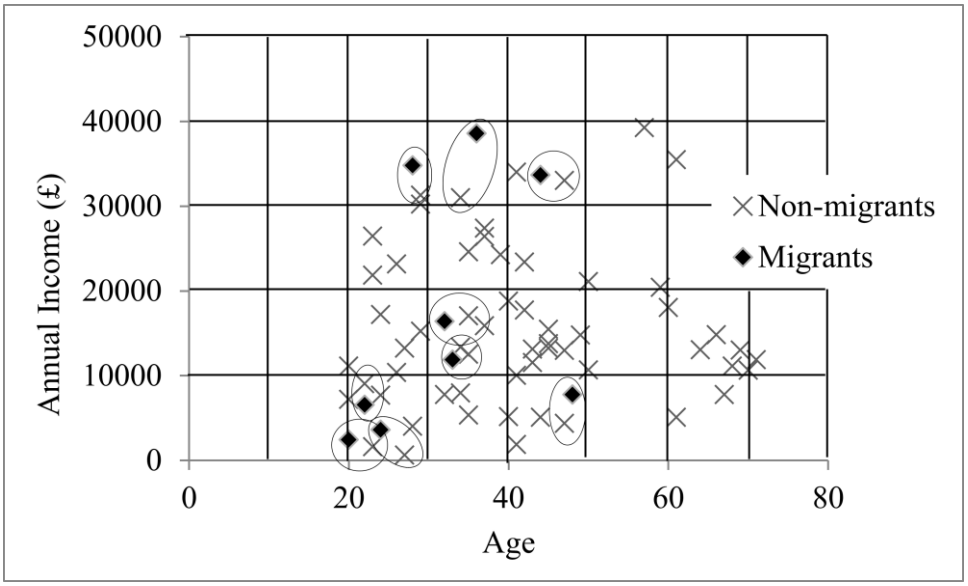


Figure 7.4: Matching of data (circles showing the joins).

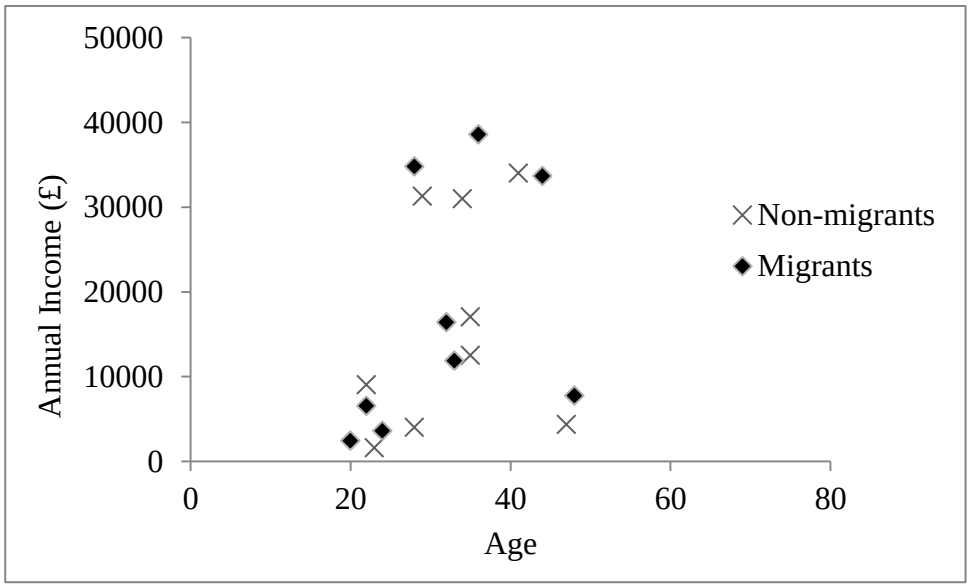


Figure 7.5: The final data set.

The categories from Figure 7.3 are then used to pair cases through looking at each datum of 'migrant' status, finding an equivalent datum of the opposite status in each box (category). This is shown in Figure 7.4, with the circles highlighting the joins made through the method. The result is that the data set becomes pruned to allow for fairer comparisons through the creation of a control group, as shown in Figure 7.5. When

inexact matching is used, the data for non-migrants would be weighted to show their importance.

A drawback of this method is that in its one-to-one matching, the methodology randomly prunes observations to choose which exact matches to use when there are more than one match that could be assigned (Blackwell, 2012). This random element can lead to slightly different results each time the method is run. Therefore inexact matching is to be used only, especially due to the small sample size involved once origin and destination are accounted for.

As a result, a simple comparison of means cannot be used to test the difference, as often used with one-to-one matching (Iacus et al., 2011a; Blackwell et al., 2009). This is because not all the selection bias has been eliminated. Therefore parametric tests are required to find the causal estimate of migration's effect on health, which will also help control for any low level of bias remaining.

Most previous research has failed to account for selection bias when analysing migration. Through matching the data, this approach has been shown here to be a more robust and relevant method. Therefore the results from this chapter are useful in building upon and furthering research in this area, providing a new and more accurate insight into the effect of migration on health.

7.3 Data considerations

To analyse the influence of migration between areas on health, the British Household Panel Survey data set was used. As Larson et al. (2004) note, there are few data sets which combined information on migration, location and health. The BHPS allows for the tracking of individuals over time, with information on both individual social characteristics and health. The classification was assigned to each individual previously (Section 6.2) to be able to explore where people were migrating to and from.

Table 7.1 shows the numbers of people migrating within the BHPS sample. The proportion of migrants remains even for the first two years, before falling in 2008. This drop may show the effect of the recession, which led many households to delay buying a house, causing a drop in sales and prices (Campos et al., 2010/2011). The data here is fairly representative to national patterns, for example Randall (2011) reporting on behalf

of the ONS (using the English Housing Survey), notes that the number of migrants was nine per cent in 2008/09 making the sample fairly representative.

Wave	Number Migrated	Percentage
2006	625	9
2007	642	9.3
2008	522	7.5

Table 7.1: Number of people migrating by wave in the BHPS, 2006-2008.

The data years 2006 to 2008 were pooled together to improve both the stability and the power of analyses. These years were chosen as they correspond with the classification. Data was kept to identify variables for the same individual for both the year prior to (referred to as year 'A') and including the year where migration was taken from (year 'B'). This allowed a more stable and accurate model to be built and tested. This approach has been useful in past migration related studies (Evandrou et al., 2010).

Through pooling the panel data, individuals interviewed in each wave will appear multiple times. This creates the potential for slight issues to arise in the model, as correlation exists within the cases of the data. For example 76 (4.2 per cent of all migrants) people between 2006 and 2008 migrated in each wave, with 529 (29.6 per cent) migrating more than once. However, the small issues this will create are likely to be offset by the advantages of having a greater sample size (Evandrou et al., 2010). Effectively tripling the data set gives greater power when running models, keeping the standard errors lower and providing more accurate results.

7.4 Conceptual approach

The interpretation of the classification in Chapter 5 showed the varying experiences in the types of mortality profiles throughout England and Wales. This coupled with the longstanding evidence of the social and medical inequalities that exist (Thomas et al., 2010) highlights that not all migrations are equal (at least in terms of area characteristics). If neighbourhood effects and the role of geography to be important, then we would expect differences to exist in respect to area type.

The classification is a useful tool for analysing the effect of migration by area type. With each cluster representing a distinct and different mortality profile, it allows for a

deeper analysis of differences in area effects, rather than comparing two or three types of areas like past research (*c.f.* Brimblecombe et al., 2000; Popham et al., 2011; Riva et al., 2011). There are eight areas, providing a more focused analysis for health than using other determinants such as just rural and urban areas (for example Riva et al., 2011). It forms a proxy for the area characteristics, compiled of greater information and therefore adding value to the analysis (Openshaw et al., 1994). If area effects are important, then we would expect that migration with an up- or down-scaling in the mortality profile of an area to have an observed effect on health.

To evaluate the usefulness of the classification, the analysis is once again compared to applying the GORs. This will evaluate whether the classification is more discriminating as a research tool, showing if it can add more to our understanding than traditional measures for analysing health.

A theoretical issue questions whether neighbourhood effects can exist in this context. It is often theorised that neighbourhood effects occur over a long time frame (Hedman, 2011; Musterd et al., 2012; Quillian, 2003) and therefore would not occur in this context. Yet analysing and testing such an effect is likely to be difficult once a long time scale is involved as it assumes total knowledge of all areas lived in and their respective characteristics (as they constantly change). Furthermore both Briggs and Keys (2009) and van Ham and Manley (2012) argue that greater consideration is needed to the temporal aspect in the area effects literature. Short term effects on health after migration have been shown elsewhere (for example Kahlmeier et al., 2001; Leventhal and Brooks-Gunn, 2003; Ludwig et al., 2011; 2012; Sanbonmatsu et al., 2012).

7.5 Migration in England and Wales

Firstly, geographical migration within the data is first examined. This will help aid our understanding of future analyses by understanding the patterns in the data that may be driving processes identified later. It will also examine the existence of selective migration (socially) within the sample as well.

It is worth starting by drawing out, briefly, patterns from the literature which would be expected to be found in this data. Whilst migration occurs throughout the life course, it is more pronounced at earlier ages. For example, the ONS (2013) reports that 19 is median age of all migrants. This is the result of a variety of factors including the

transition to work or going to university, allowing the young to leave the parental home. However, as Chapter 5 highlighted, migration is also prominent later in the life course, due to changes in economic and health situation (see Evandrou et al., 2010).

Economic geography also plays an influential role in attracting migrants. Areas that have a greater share of employment opportunities attract migrants (Champion, 2012). In England and Wales, the South East is traditionally seen as an escalator region (ibid). Being the economic powerhouse of England and Wales, it attracts the young in search of employment. It is used for social mobility to occur, through allowing them to accumulate economic resources allowing them to move out later in life (ibid).

Wealth accumulation drives the housing market through a continued upward spiralling. This reflects social desirability, as individuals seek to improve the social standing of where they live (Catney and Simpson, 2010). However the rental market allows those with few resources to forego accumulating wealth to access desirable areas (especially the young). This, however, consistently reflects a social gradient, helping to maintain inequalities (ibid).

7.5.1 Migration between the clusters

Table 7.2 presents the pattern of migration in the BHPS, between the clusters. Any combination of migration involving less than 20 individuals were made bold to highlight them as any further observations may be problematic and biased due to small numbers.

There is wide variation in the numbers of people migrating between each cluster. Given the varying types of areas, assuming migratory effects to be consistent across space would appear false. Therefore analysing variations between area types appears a useful extension of the analysis.

The total number of migrations within each cluster corresponds to the overall size of each cluster. Coverage across the sample appears representative of our population. The single largest trend in destination and origin of migration was within cluster type, with 34.9 per cent of all migrations to the same cluster.

		Cluster migrated to								
		1	2	3	4	5	6	7	8	Total
Cluster migrated from	1	203	51	76	32	30	**6**	23	84	505
	2	62	144	54	58	26	27	**14**	55	440
	3	74	55	171	42	**17**	20	**19**	44	442
	4	49	43	67	132	28	22	**17**	39	397
	5	27	33	20	**19**	107	27	**11**	20	264
	6	21	**12**	**16**	33	**11**	48	**5**	26	172
	7	20	29	20	**15**	20	**3**	22	**14**	143
	8	76	38	65	37	25	**11**	**13**	140	405
	Total	532	405	489	368	264	164	124	422	

Table 7.2: A cross tabulation of migration to and from clusters.

Key: Clusters; 1 = Best Health and Most Desirable, 2 = Average Mortality Profiles, 3 = Good Health Areas, 4 = The Middle, 5 = Poor Health Experiences, 6 = Poorest Health and Least Desirable, 7 = Poorest Neurodegenerative Health, 8 = Mixed Experiences.

Table 7.3 shows the variation in how well each cluster retained the proportion of individuals who migrated from the cluster type. 'Poorest Neurodegenerative Health' is a particularly unstable cluster, retaining fewer than half the proportion otherwise found for the other clusters. Given the mortality and social characteristics of the cluster, it is less desirable and thus fails to attract people remain within. Since migration if usually in relation to a social and geographical upgrading (Catney and Simpson, 2010; Champion, 2012), migrants are more likely to seek to leave this area type. This can be seen in Table 7.2, where there are higher numbers of migrants moving into areas which displayed good mortality profiles rather than poorer profiles.

It is also influenced by the findings of Chapter 5. Section 5.4.3 showed the importance of the migration of elderly, possibly in search of communal homes at the end of their life, in explaining this cluster. This continued turnover of people in these areas would partly help explain the lack of population retained (Evandrou et al., 2010). It is not completely due to the desirability of the area, with both of these negative factors combining to keep population retention low.

Whilst 'Poorest Health and Least Desirable' also retained a low percentage of migrants, the proportion was higher than for 'Poorest Neurodegenerative Health', despite worse

social and mortality characteristics. It is not completely a linear relationship between desirability and migration. This is further shown by 'Poor Health Experiences', which retains the highest proportion of migrants. Those with the best mortality profiles also perform well, highlighting the influence of their desirability.

Cluster	Migrants remaining in the same cluster (%)	Change in overall size (%)
Best Health and Most Desirable	40.2	5.3
Average Mortality Profiles	32.7	-8.0
Good Health Areas	38.7	10.6
The Middle	33.2	-7.3
Poor Health Experiences	40.5	0
Poorest Health and Least Desirable	27.9	-4.7
Poorest Neurodegenerative Health	15.4	-13.3
Mixed Experiences	34.6	4.2

Table 7.3: Change in cluster sizes as a result of migration.

Table 7.3 also presents the overall change in membership of the clusters as a result of migration. Comparing Table 7.3 to Table 5.14 (net migration of the clusters for all areas within them) shows that the trends are similar. Those areas with positive population gains in the BHPS, also display the same pattern when using ONS area data. Therefore the results are representative and useful.

The clusters with the best mortality profiles ('Best Health and Most Desirable' and 'Good Health Areas') also have seen the highest relative growth. People are migrating to those clusters with better health characteristics. It is not surprising that they are viewed as the most attractive areas, given the social characteristics these areas also exhibit.

The higher change of 'Good Health Areas' reflects this as being an 'escalator cluster'. As they are not the most socially desirable but still are good, the areas that consist of the cluster are slightly cheaper. They therefore are more attractive to people first, a stepping stone towards getting to the cluster 'Best Health and Most Desirable' after wealth accumulation (see Champion, 2012). Although past research has focused on escalator regions (with the whole region attracting people; ibid; ONS, 2013), this would indicate that neighbourhood type is influential, as the whole effect in a region will not be consistent.

This pattern is not strictly linear, as it does not quite repeat itself at the other end of the distribution. 'Average Mortality Profiles', whose mortality profile was fairly good, has a high population loss. Examining the destinations of migrants (Table 7.2) would suggest that this change may represent people 'upgrading' (Catney and Simpson, 2010), with a large number migrating to those cluster with better mortality profiles. They are less attractive areas, with individuals seeking to migrate from despite not being displaying poor health or high poverty characteristics. This also relates to the 'escalator cluster' concept as well.

The highest relative decline is found in 'Poorest Neurodegenerative Health', a factor of it being unable to retain its population. Whilst this makes sense (as they are less socially desirable; Catney and Simpson, 2010), the relationship is not consistent with the cluster 'Poorest Health and Least Desirable', who did not report as high a negative change. The destinations of migrants from these two clusters is similar, showing that this is not necessarily an up-scaling (although if you live in the worst, migrating to any other cluster is still up-scaling even if it is not to a desirable one).

Given that the majority of the theory surrounding escalator areas and migratory patterns are based around regional migrations (Champion, 2012), Table 7.4 presents an exploration of retention of individuals through migration by region. The regions have higher retention of individuals, showing them to be less discriminating in discerning between differences in migration. The majority of migrations are within regions, although as Table 7.2 shows, they are not necessarily to the same area types, which is better captured by the classification. The idea of an escalator region is less accurate, with escalator cluster (or area types) being more useful in understanding patterns and processes.

London differs from this trend, retaining a slightly lower proportion of individuals than compared to the other regions. This follows previous analysis, which has shown London to be characterised by the outflow of migrants (ONS, 2013). Exploring the destination of these migrants shows that they are migrating to the GORs East (11 per cent) and South East (8.4 per cent). These are likely to represent people migrating out of the city to commuting settlements (Champion, 2012). Rather, the majority of inflow for London is now of international migrants, showing that it is still attractive (ONS, 2012b).

GOR	Migrants remaining in the same cluster (%)	Change in overall size (%)
North East	86.6	-1.5
North West	88.9	0.9
Yorkshire	87.3	3.7
East Midlands	82.0	-2.3
West Midlands	85.1	-4.5
East	86.5	5.4
London	71.0	-11.6
South East	82.0	3.8
South West	84.6	1.7
Wales	93.9	-0.8

Table 7.4: Changes in the size of regions due to migration.

7.5.2 Characteristics of migrants

It would be useful to examine the variations in the characteristics of the people migrating between clusters given that migration by area type is not consistent. This will help improve the understanding of processes occurring, which may explain future analyses.

Age and income are examined in Tables 7.4 to 7.5 due to their importance in previous research (Bentham, 1988; Catney and Simpson, 2010; Chapter 6). Also included in Table 7.6 is the change in income between years. Given that negative major life events are an important determinant of migration (Dorling, 2012) and especially on affecting health (Bradley and Van Willigen, 2010), it is a useful addition to our analysis.

Similarly to Tables 5.1 to 5.6, the results in each table have been conditionally formatted to aid interpretation of the results. The value in each table that is highest is coloured red and the lowest is coloured green. The rest of the values are then coloured in respect to their position in the range of values, with yellow representing the mean. Those combinations of migrations that were bold in Table 7.2 have been suppressed to avoid any misleading associations due to small numbers.

There were relatively wide variations in the average age of migrants between clusters (Table 7.5). The young migrated into clusters of higher mortality rates and *vice versa*. Generally those who migrated within a cluster, rather than to a different destination type, had a higher age as well. These patterns can be explained by the accumulation of

resources over the life course, which allows people to migrate to more socially desirable areas (i.e. the clusters with better health profiles), which in turn demand higher prices (Catney and Simpson, 2010). Those who migrated, however, were younger on average than the underlying populations of the clusters they were migrating to. This is reflecting the higher prevalence of migration at younger ages (ONS, 2013), a factor of the young leaving home through the transition to work.

		Cluster migrated to									
		1	2	3	4	5	6	7	8	Total	
Cluster migrated from	1	36.7	34.6	34.7	26.2	33.6		31	37.2	35.1	(49.4)
	2	34.3	34.8	31	31.6	33.7	27.6		36.4	33.4	(47.7)
	3	36.5	30	36.9	30		29.1		33.9	33.9	(48.9)
	4	35.3	31.8	30.1	35.2	26.8	31.8		31.8	32.8	(45.9)
	5	28.9	30	32.9		32.2	30.8		27.6	30.9	(46.4)
	6	27.7			36.6		32.9		38.2	33.1	(48.2)
	7	36	30.1	33.1		27		36.9		32.2	(44.9)
	8	35	31.5	36.2	30.9	28			38.5	35.1	(57)
	Total	35.2 (50.4)	32.7 (48.7)	34.4 (49.9)	32.6 (46.9)	30.7 (47.2)	31.1 (49)	31.2 (45.9)	36.1 (48.2)	33.6	(48.2)

Table 7.5: Variations in the average age of migrants between the clusters (mean value for all individuals in the cluster given in brackets).

Key: Clusters; 1 = Best Health and Most Desirable, 2 = Average Mortality Profiles, 3 = Good Health Areas, 4 = The Middle, 5 = Poor Health Experiences, 6 = Poorest Health and Least Desirable, 7 = Poorest Neurodegenerative Health, 8 = Mixed Experiences. Colours; Green is a low SMR, red a high SMR, yellow the mean and the other colours scaled accordingly to their respective position

The clusters which displayed better health characteristics attracted those people with higher incomes (Tables 7.6). The pattern of mean income followed the respective mortality profile of each cluster fairly well (and their area characteristics; Chapter 5). This was the same for both those leaving a cluster and those arriving to one. Furthermore those who migrated from a good to a bad cluster (in terms of mortality) had lower income levels, than compared to those making the other direction. This is reflecting the social upgrading of individuals over space (Catney and Simpson, 2010; Champion, 2012). Migrants though had lower incomes than the underlying populations, showing migration to be more opportunistic than just a simple function of increasing income over time. It also reflects the greater number of younger people migrating, lowering these estimates.

		Cluster migrated to								
		1	2	3	4	5	6	7	8	Total
Cluster migrated From	1	18373	17730	18960	12796	12675		13099	19040	17418 (18555)
	2	20932	13621	14253	13835	11887	14529		17183	15159 (15394)
	3	19095	14854	15847	13239		12351		12471	15154 (16097)
	4	15720	13370	14309	14169	12909	11236		13247	13819 (13929)
	5	12485	15183	9466		11613	12093		13414	12456 (12682)
	6	16157			13349		10761		14131	13610 (12501)
	7	22118	14250	11794		11134		12417		14091 (15503)
	8	20319	13464	13082	12134	11131			17400	15605 (17286)
	Total	18549 (19480)	14464 (16050)	15158 (16191)	13641 (14698)	11784 (13303)	11880 (13249)	12527 (16177)	16197 (18294)	15029 (16277)

Table 7.6: The distribution of mean income for migrants, cross-tabulated between the cluster of area migrated to and from (mean value for all individuals in the cluster given in brackets).

Key: Clusters; 1 = Best Health and Most Desirable, 2 = Average Mortality Profiles, 3 = Good Health Areas, 4 = The Middle, 5 = Poor Health Experiences, 6 = Poorest Health and Least Desirable, 7 = Poorest Neurodegenerative Health, 8 = Mixed Experiences. Colours; Green is a low SMR, red a high SMR, yellow the mean and the other colours scaled accordingly to their respective position

These patterns were further reflected with the income change variable (Table 7.7). On average, a migration into a cluster with a worse mortality profile was also associated with a fall in income. Those who migrated to the better clusters experienced a rise in income. Clearly this represents social mobility taking place (i.e. the accumulation of resources), where improvements in income levels are allowing people to migrate to more desirable areas (Champion, 2012), of which they also reap health benefits as well (Brimblecombe et al., 2000).

Those migrating from 'Poorest Health and Least Desirable' witnessed the largest average increase in income, suggesting that higher increases are required to escape the cluster with the worst mortality profile (although not necessarily to the best areas). Internal migration within clusters show the most interesting pattern, as generally there is a rise in income levels facilitating migration. Rises in income are allowing people to migrate, but not all choose to leave their area type (but still may be up-grading within the cluster), even amongst the areas with poor health outcomes.

		Cluster migrated to								
		1	2	3	4	5	6	7	8	Total
Cluster migrated From	1	1725	953	491	-931	-4247		633	-3541	-68 (679)
	2	-2717	112	2106	-2786	-2105	-1338		2790	-363 (718)
	3	1395	1231	365	-624		519		-3231	194 (818)
	4	333	-1281	-24	1479	-1639	377		-304	197 (851)
	5	2553	4740	-2108		660	-559		-1496	543 (689)
	6	4355			1280		277		-390	1893 (700)
	7	6984	-702	1056		-916		-806		647 (654)
	8	1995	-3827	-1709	-797	-3677			1766	116 (815)
	Total	1448 (781)	359 (681)	383 (816)	-117 (908)	-972 (661)	-475 (642)	-238 (708)	-310 (687)	208 (754)

Table 7.7: The average change in annual income levels of migrants, with cluster origin and destination (mean value for all individuals in the cluster given in brackets).

Key: Clusters; 1 = Best Health and Most Desirable, 2 = Average Mortality Profiles, 3 = Good Health Areas, 4 = The Middle, 5 = Poor Health Experiences, 6 = Poorest Health and Least Desirable, 7 = Poorest Neurodegenerative Health, 8 = Mixed Experiences. Colours; Green is a low SMR, red a high SMR, yellow the mean and the other colours scaled accordingly to their respective position

A multinomial regression was run using these variables to explain whether they differed significantly by cluster destination. This was important for identifying the existence of selection bias in the data and hence whether matching is required. Gender was also included to control for this factor as well. Multicolinearity was first tested for. With the change in income variable constructed from the income variable itself, this might be an issue. However there was only a moderate association between the two variables, with a correlation of .366 ($p<.001$). The rest of the associations were all weak, meaning that this model assumption has been checked for. The other assumptions (linearity, influence of extreme data points, normality etc) were not violated.

The results of the regression model can be seen in Table 7.8. The model was significantly different from the null model ($\chi^2=109.25$, $p<.001$) and the pseudo-r^2 value was 0.01. Few relationships are significantly different, when compared to migrating to the cluster 'Best Health and Most Desirable' (selected as it displayed the best health outcomes, which is useful to make comparisons against). Any significant relationships were made bold to improve the interpretation of the table.

Cluster	Variable	Coefficient	Std. Error	P
Reference: Best Health and Most Desirable				
Average Mortality Profiles	**Age**	**-0.01**	**0.005**	**0.039**
	Sex	0.216	0.144	0.133
	Income	**-2.10E-05**	**5.91E-06**	**<0.001**
	Income Change	2.07E-06	6.33E-06	0.743
	Constant	0.279	0.192	0.148
Good Health Areas	Age	-0.002	0.004	0.689
	Sex	0.218	0.136	0.11
	Income	**-1.62E-05**	**5.14E-06**	**0.002**
	Income Change	1.20E-07	5.37E-06	0.982
	Constant	0.125	0.181	0.49
The Middle	Age	-0.009	0.005	0.073
	Sex	0.274	0.147	0.062
	Income	**-2.69E-05**	**6.32E-06**	**<0.001**
	Income Change	4.60E-07	6.40E-06	0.943
	Constant	0.228	0.196	0.244
Poor Health Experiences	**Age**	**-0.018**	**0.006**	**0.002**
	Sex	**0.478**	**0.163**	**0.003**
	Income	**-3.96E-05**	**7.91E-06**	**<0.001**
	Income Change	-1.76E-06	7.22E-06	0.808
	Constant	0.262	0.221	0.236
Poorest Health and Least Desirable	**Age**	**-0.017**	**0.007**	**0.017**
	Sex	0.166	0.195	0.396
	Income	**-3.81E-05**	**9.86E-06**	**<0.001**
	Income Change	2.07E-06	9.62E-06	0.83
	Constant	-0.164	0.261	0.53
Poorest Neurodegenerative Health	**Age**	**-0.017**	**0.008**	**0.03**
	Sex	0.145	0.215	0.501
	Income	**-3.27E-05**	**1.06E-05**	**0.002**
	Income Change	2.19E-06	1.06E-05	0.836
	Constant	-0.46	0.288	0.11
Mixed Experiences	Age	0.003	0.004	0.424
	Sex	0.056	0.142	0.692
	Income	-8.20E-06	4.75E-06	0.084
	Income Change	-4.16E-06	4.22E-06	0.324
	Constant	-0.265	0.187	0.158

Table 7.8: A multinomial logit regression explaining the characteristics of people migrating to each cluster (Coefficients significant at the 95 per cent level are in bold).

Only income was consistently significant across the results, with the level of income for individuals being significantly lower when individuals migrated to a cluster other than 'Best Health and Most Desirable' (with 'Mixed Experiences' the only insignificant result). The size of the respective coefficients for income reflects the mortality profile of its cluster (in comparison to 'Best Health and Most Desirable'). Age also was significant for four clusters, showing that those who migrated to these clusters (in comparison to 'Best Health and Most Desirable') were younger. Overall these results reflect the previous findings. The significance of sex for those migrating to 'Poor Health Experiences' is likely a facet of the small numbers involved, given that there is little theoretical basis for this.

The evidence in this section shows the social selection that exists through migration. Selective migration occurs between the clusters, across social and demographic characteristics (Catney and Simpson, 2010; Champion, 2012; Evandrou et al., 2010). Migration allows for social inequalities to become reciprocated and maintained geographically. The reinforced divisions of social groups, which research shows to have better health (Gregory, 2009; Woods et al., 2005), impacts on spatial patterns of health through polarisation. It is important that these differences are accounted for in subsequent analyses. Therefore taking a matching methods approach to control for selection bias is a useful and necessary approach.

The analysis was repeated, using regional destination of migrants instead of clusters (Table 7.9). However it was a particularly poor model, with only two coefficients statistically significant. Using the GORs adds little to the understanding of patterns in the characteristics of internal migrants. The classification is more useful for discriminating variations in the data.

Cluster	Variable	Coefficient	Std. Error	P
Reference: North East				
North West	Age	0.01	0.01	0.335
	Sex	-0.032	0.313	0.92
	Income	-5.72E-06	1.44E-05	0.69
	Income Change	-1.03E-05	2.43E-05	0.67
	Constant	**1.044**	**0.419**	**0.013**
Yorkshire	Age	0.004	0.011	0.685
	Sex	0.032	0.317	0.92
	Income	4.60E-06	1.43E-05	0.748
	Income Change	-1.74E-05	2.43E-05	0.473
	Constant	**0.939**	**0.427**	**0.028**
East Midlands	Age	0.011	0.011	0.321
	Sex	-0.088	0.322	0.785
	Income	-2.65E-06	1.47E-05	0.857
	Income Change	-2.24E-05	2.46E-05	0.362
	Constant	0.0774	0.432	0.073
West Midlands	Age	0.006	0.011	0.593
	Sex	-0.193	0.328	0.557
	Income	1.96E-05	1.42E-05	0.166
	Income Change	-1.63E-05	2.44E-05	0.504
	Constant	0.509	0.442	0.25
East	Age	0.017	0.011	0.116
	Sex	-0.331	0.32	0.302
	Income	2.18E-05	1.38E-05	0.116
	Income Change	-1.96E-05	2.37E-05	0.409
	Constant	0.351	0.43	0.414
London	Age	-0.012	0.012	0.34
	Sex	**-1.01**	**0.346**	**0.003**
	Income	**4.54E-05**	**1.37E-05**	**0.001**
	Income Change	-3.67E-05	2.26E-05	0.105
	Constant	0.822	0.459	0.073
South East	Age	0.005	0.01	0.632
	Sex	-0.323	0.307	0.293
	Income	2.43E-05	1.35E-05	0.07
	Income Change	-3.83E-05	2.25E-05	0.089
	Constant	1.114	0.413	0.007
South West	Age	0.005	0.01	0.664
	Sex	-0.372	0.309	0.228
	Income	3.72E-06	1.40E-05	0.791
	Income Change	-2.38E-05	2.36E-05	0.313
	Constant	**1.405**	**0.413**	**0.001**

Wales	Age	0.013	0.01	0.197
	Sex	-0.375	0.299	0.21
	Income	-1.38E-05	1.38E-05	0.318
	Income Change	-7.43E-06	2.33E-05	0.75
	Constant	**1.736**	**0.399**	**<0.001**

Table 7.9: A multinomial logit regression explaining the characteristics of people migrating to each GOR (Coefficients significant at the 95 per cent level are in bold).

7.6 Intra-cluster migration effects

This section examines the effects of migrating to and from the clusters and how whether the areas involved impact upon health. Unlike past research (Bentham, 1988; Brimblecombe et al., 2000; Larson et al., 2004), migration is not viewed as the same effect independent of areas and hence the analysis looks for variations split by origin and destination.

7.6.1 Analysing the role of cluster origin

To explore the role of migration on health between different area types, the analysis begins by looking at the effect of migrating out of particular clusters. As such, matching was performed using whether an individual migrated or not. Subsets of data were created, splitting up individuals by cluster location at time point 'A' (i.e. origin). This allowed the pairing of migrants against those which remained in the same area, making fairer comparisons by accounting for area types.

To match individuals, the covariates age, sex and income were used since they presented differences in migration patterns earlier (Section 7.5) and also in past research (Chaix et al., 2007; Jen et al., 2011). No further variables were included to minimise the amount of noise added to the process, as it becomes more difficult to find exact matches across a greater range of variables (Iacus et al., 2011b). As the methodology works through categorising variables, age was split into year bands (16-24, 25-34, 45-54, 55-64, 65-74, 75-84, 85+) in accordance to previous usage (Section 3.7.4). Income was divided into ten thousand pound bands up to £50,000, where everything above was included as a group. Sex did not need to be altered.

Prior to matching the data, it is useful to statistically evaluate how balanced the data is (how similar the created control group is to the characteristics of migrants; i.e. internal validity), to be able to gauge how effective this approach has been. This can be evaluated using the L_1 global statistic developed in Iacus et al. (2011a). This is the difference in the (multivariate) distribution of data. Data is automatically coarsened to examine the initial imbalance exhibited across the data, as well as the individual imbalance of each variable. It is measured on a scale of zero to one, which gives the separation of cases from matches (where one is complete separation and zero is perfect matching). For a balanced data set, this should be less than a value of 0.5. The results in Table 7.10 show that matching was effective.

After matching the data, a logistic regression was run to explore whether there was an observed effect of people migrating from each cluster on their health for each of the subsets of data. The results for all these regressions are summarised in Table 7.10.

Origin Cluster	CEM	Effect of Migration		
	L_1	Odds ratio	Standard Error	Significance
Best Health and Most Desirable	0.372	1.005	0.133	0.97
Average Mortality Profiles	0.419	1.081	0.143	0.554
Good Health Areas	0.413	1.516	0.193	<0.001
The Middle	0.418	1.026	0.143	0.852
Poor Health Experiences	0.407	1.411	0.242	0.044
Poorest Health and Least Desirable	0.379	1.104	0.228	0.631
Poorest Neurodegenerative Health	0.399	0.97	0.227	0.897
Mixed Experiences	0.355	0.821	0.125	0.195

Table 7.10: Results of a series of logistic regressions exploring the differential effect of migration on health by cluster.

Migration for only two clusters was found to be significant, showing little existence of area effects captured through migration. The significant relationships were found for the clusters 'Good Health Areas' and 'Poor Health Experiences', both yielding positive relationships. People who originally resided in either of these clusters were found to have greater probability of reporting their health as poor if they had migrated from (or within) these areas (as opposed to those who remained). These effect sizes were large; being 52 and 41 per cent more likely to report poor health if they were in clusters 'Good Health Areas' and 'Poor Health Experiences' respectively.

To explore whether these significant effects were possibly neighbourhood effects, their respective models were extended further. Firstly health status prior to migration was included in the regression, to control for whether a person had already poor health (prior to migrating). This essentially adds to the model the health selective migration hypothesis, testing whether any effect is independent of this. Although this may be suggestive of introducing multicolinearity into the model, correlating the two covariates showed this not to be too problematic despite moderate association ($r=0.58$, $p<0.001$). Results of this model are shown in Table 7.11. Both of these significant relationships for the clusters held after controlling for this effect, with little change in the values.

Origin Cluster	*Variable*	*Odds ratio*	*Standard Error*	*Significance*
Good Health Areas	Migration	1.402	0.208	0.023
	Health	11.412	1.018	<0.001
Poor Health Experiences	Migration	1.581	0.322	0.024
	Health	12.242	1.691	<0.001

Table 7.11: Controlling for prior health status to migrating on the hypothesised impact of migration on health.

The next stage is to split the migration variable further to include the migratory destination (cluster). As such, this controls for the effect of the area as well. The results of these two models are presented in Tables 7.12 and 7.13. The reference cluster for comparing area types for each model is the cluster 'Best Health and Most Desirable', since it displays the best mortality profile and thus show the effect on health of not migrating to area with the best health outcomes. The results are, however, the same for either model. All of the dummy variables are insignificant, indicating no observable variations by area types.

The absence of significant findings in Tables 7.12 and 7.13 would point towards a lack of evidence of area effects between the clusters. The type of area you migrated to does not have a differential effect on your health, at least measurable soon after migrating. However, as the literature argues, area effects are not usually direct (Berkman and Glass, 2000). Therefore it is unlikely that any effect would become apparent in the short term (Johnson et al., 2012). Any observable effect would have most likely been psychological. If you suddenly migrated into an area where more people are suffering from poor health, then it is plausible that you start seeing your own health more critically.

Variable	Odds ratio	Standard Error	Significance
Cluster:	('Best Health and Most Desirable' is the reference)		
Average Mortality Profiles	1.247	0.696	0.693
Good Health Areas	1.16	0.41	0.675
The Middle	0.872	0.461	0.795
Poor Health Experiences	4.282	3.424	0.069
Poorest Health and Least Desirable	1.754	1.464	0.501
Poorest Neurodegenerative Health	1.704	1.432	0.526
Mixed Experiences	1.338	0.758	0.607
Health	11.505	1.031	<0.001

Table 7.12: The analysis of self-rated health variations between individuals who migrated from 'Good Health Areas'.

Variable	Odds ratio	Standard Error	Significance
Cluster:	('Best Health and Most Desirable' is the reference)		
Average Mortality Profiles	0.218	0.194	0.088
Good Health Areas	0.603	0.781	0.696
The Middle	0.818	0.803	0.838
Poor Health Experiences	0.46	0.324	0.27
Poorest Health and Least Desirable	1.045	0.947	0.961
Poorest Neurodegenerative Health	0.66	0.762	0.719
Mixed Experiences	0.221	0.253	0.187
Health	12.478	1.736	<0.001

Table 7.13: The results of the analysis of migrants from 'Poor Health Experiences'.

These results would indicate that the original significant effects (Table 7.10) could be migratory effects instead. Destination of migration is less important, rather it is the process of migration which impacts upon health. Given that they show that migrating from these areas increases the probability that an individual reports their health as poor, migration is of importance. With the direction of the effect being the same for either cluster, despite different mortality and social profiles, this appears more plausible when

combined with the lack of evidence for differences by destination. Yet such an effect is not consistent across all the clusters, showing it is not that dominant as a process.

As the data becomes split into smaller and smaller subsets, it may not be surprising that no significant results were found. The small sample sizes involved with those who migrated between clusters limit our ability to test any relationships. Migration between the extremes is not common (Brimblemcombe et al., 2000; Table 7.2). Therefore area effects should not be completely written off, until more rigorous analysis can be conducted. Given that the few studies which examined a similar approach, found varying results by area types (albeit usually only comparing few types), this points towards the importance of the neighbourhood (Leventhal and Brooks-Gunn, 2003; Popham et al., 2011; Riva et al., 2011).

This is reflected through the relationships which are approaching an acceptable level of significance (i.e. which can be accepted at the 90 per cent level; Tables 7.12 and 7.13). They appear to follow expected pathways, with migration from 'Good Health Areas' to 'Poor Health Experiences' resulting in an increasing probability of reporting poor health (and *vice versa* in the other model, with migration from a poorer health cluster to one with a better mortality profile). It follows the results of the 'Moving to Opportunity' studies, showing how area effects imprint upon the migration process (Leventhal and Brooks-Gunn, 2003). However, with their wide standard errors, these should not be applied further and are only commented on to provide an exploratory indication of possible avenues requiring a larger sample size.

7.6.2 Accounting for cluster destination

The analysis was repeated using cluster destination, rather than origin, to compare the observed effect of migration to explore if this effect is consistent (using the same procedure for matching). The results can be seen in Table 7.14, with the majority of relationships being insignificant. However there was still a significant effect for the cluster 'Good Health Areas'. People who migrated into the cluster 'Good Health Areas' were 31.5% more likely to report their health as poor than compared to those who did not migrate. Controlling for health status prior to migration into the model however results in this effect becoming insignificant (Table 7.15).

Cluster Destination	CEM	Effect of Migration		
	L_1	Odds Ratio	Standard Error	Significance
Best Health and Most Desirable	0.377	1.054	0.139	0.693
Average Mortality Profiles	0.444	1.149	0.157	0.31
Good Health Areas	0.381	1.29	0.167	0.049
The Middle	0.466	1.315	0.191	0.06
Poor Health Experiences	0.427	1.162	0.204	0.392
Poorest Health and Least Desirable	0.359	1.127	0.245	0.583
Poorest Neurodegenerative Health	0.329	0.951	0.242	0.842
Mixed Experiences	0.335	1.083	0.155	0.577

Table 7.14: Results of logistic regression models assessing the impact of migrating to various clusters.

	Odds ratio	Standard Error	Significance
Migrated	1.176	0.176	0.279
Health Status	11.2	1.019	<0.001

Table 7.15: Controlling for health status prior to migrating to the cluster 'Good Health Areas'.

Splitting the migration variable to differentiate by origin however (Table 7.16), provides greater detail than the previous analysis (Tables 7.12 and 7.13). There are two clusters which observed significant differences when compared to the cluster with the best mortality profile ('Best Health and Most Desirable'). This shows that there is some evidence that area effects exist through migration. It supports the results of the 'Moving to Opportunity' studies as well (Leventhal and Brooks-Gunn, 2003). Indeed, the odds ratio for 'Poorest Health and Least Desirable' is particular large, with migrants from this cluster to 'Good Health Areas' over 12 times more likely to have poor health than the reference cluster at the other extreme of the classification. The results highlight the inequalities that exist and are observed through migration. This would indicate the possible role of (large) area effects.

Low sample size also plays a role here, restricting the observations that can be drawn. Rather than the lack of significance reflecting the lack of strong relationships, it is also partly explained by the low sample size in particular migrations (as shown in Table 7.2). This limits the power of the model and such it would be useful for further research with larger datasets to explore this possibility further to see if this can explain the results.

The standard errors are fairly wide due to this, meaning that the large effect recorded may not necessarily be true, possible over-estimating the importance of area.

Variable	Odds ratio	Standard Error	Significance
Origin cluster: ('Best Health and Most Desirable' = reference)			
Average Mortality Profiles	4.828	2.957	0.01
Good Health Areas	2.154	0.876	0.059
The Middle	2.81	1.506	0.054
Poor Health Experiences	3.201	3.637	0.306
Poorest Health and Least Desirable	12.454	11.359	0.006
Poorest Neurodegenerative Health	1.708	1.418	0.519
Mixed Experiences	1.962	1.106	0.232
Health Status	11.37	1.04	<0.001

Table 7.16: Exploring the impact of accounting for cluster origin on migrants to the cluster 'Good Health Areas' on health.

7.6.3 Application of GORs to the analysis

To be able to put these results in context, GORs were introduced in the same analysis to compare whether it can add greater understanding to the impact of migrating between areas on health. This aids the evaluation of the effectiveness of the classification as an analytical tool for discriminating between patterns. The analysis was conducted similarly to the previous two sections. Matching was performed using whether an individual migrated or not, pairing them to individuals based upon both region of origin (Table 8.19) or destination (Table 8.20). Individual logistic regressions were then run analysing the impact of migration to or from each region upon health.

The analysis showed little evidence of an impact upon health across the geographical regions of England and Wales. This was independent of whether region of origin and destination was used. The classification was more useful here, presenting more significant relationships and greater detail from subsequent investigations. Applying the classification within this analysis is more effective at discriminating differences in the data and analysing patterns.

GOR Origin	CEM	Effect of Migration		
	L1	Odds Ratio	Standard Error	Significance
North East	0.467	1.356	0.398	0.299
North West	0.374	1.062	0.171	0.710
Yorkshire	0.373	0.978	0.166	0.896
East Midlands	0.435	1.115	0.194	0.532
West Midlands	0.404	1.057	0.211	0.783
East	0.382	1.068	0.195	0.731
London	0.313	0.887	0.205	0.605
South East	0.346	1.265	0.191	0.119
South West	0.296	1.058	0.174	0.733
Wales	0.437	1.433	0.161	0.001

Table 7.17: The effect on health after migration from GORs.

GOR Destination	CEM	Effect of Migration		
	L1	Odds Ratio	Standard Error	Significance
North East	0.432	1.580	0.463	0.118
North West	0.358	1.148	0.183	0.388
Yorkshire	0.372	0.873	0.151	0.433
East Midlands	0.441	1.066	0.189	0.718
West Midlands	0.405	1.021	0.207	0.917
East	0.384	0.942	0.173	0.745
London	0.292	1.067	0.252	0.782
South East	0.371	1.130	0.173	0.426
South West	0.356	1.167	0.190	0.341
Wales	0.430	1.446	0.163	0.001

Table 7.18: Migrating to a GOR and the impact upon health.

Wales is the only GOR to be significant and this is independent of whether origin or destination was used. Individuals who migrated were over 40 per cent more likely to have poor health than compared to the underlying population. This may be accounted by the sample size of Wales. As reported previously (see Table 6.5), Wales contains a larger population share in the BHPS through receiving a booster sample. The greater sample size involved, compared to the other GORs, helps it to achieve significance.

The result reflects a migratory effect, which is otherwise hidden in the smaller GORs. This relationship falls consistent with the previous findings of this chapter.

An alternative explanation for this effect would be that since the Welsh population rate their health as worse than it actually is, it represents a geographical effect (Dorling and Barford, 2009). Geography imprints upon the interaction between health and migration. Those who migrated out of Wales carry with them this negative understanding of their health, with individuals who migrated inwards being influenced by their geographical location in how they view their own health. Further research to explore this hypothesis would be useful to be able to test it.

Adding self-rated health status prior to migration to control for the effect of unhealthy migrants however saw this effect become non-significant, highlighting it not to be strong. Furthermore introducing greater geographical detail to mirror the analyses in Tables 7.12, 7.13 and 7.16 gave no significant results either. Applying the GORs in the same analysis is less useful.

The lack of significant effects once both origin and destination were accounted for in the same model was due to the small numbers of migration recorded between GORs. For example 6.1 per cent of migrations when destination was used were outside of Wales (5.4 per cent when origin was used). Migration is fairly self-contained. This is because most migrations are at short distances (Larson et al., 2004; Riva et al., 2011). Any differences in migration type (at least with regards the area effects hypothesis) are lost when GORs are used. This would support the sample size explanation, rather than the Wales effect explanation.

Geographical and environmental effects will not be captured, especially as there is a diversity of social conditions within each GOR (Doran et al., 2004; Vickers, 2006; Woods et al., 2005). They are invisible administrative boundaries which are arbitrary in their location. Using GORs misses out on the more important factors, something the classification better captures. The classification focuses more on the important dimensions, capturing differences in social inequalities and divisions in mortality patterns (a proxy for the health damaging or beneficial impact of an environment). This is why it captures greater detail in this analysis.

7.7 Health selective migration

So far there has been a lack of consistent evidence surrounding the impact of area effects as measured through migration. However it may be that the relationship has been mis-specified in the analysis. Migrating between different areas does not impact upon health; rather the relationship would be the other way round in that health influences individuals to migrate. Given that Section 7.5 showed selective migration by social and demographic characteristics, it appears a natural step in the analysis to explore whether this is occurring in terms of health as well. It may also help explain the significant relationships found in the previous section.

The investigation of health selective migration has gained the most attention in England and Wales (for example Brimblecombe et al., 2000; Wannamethee et al., 2002). However most of this analysis has occurred using simple linear/binary comparisons and there are few examinations using a more detailed set of area types. There are some possible mechanisms which could operate for this approach to be useful; for example migrating to be nearer to services or family, downsizing through loss of income or the ability to work (Larson et al., 2004).

7.7.1 Matching the data

To examine the existence of health-selective migration, the analysis focuses only on those who migrated. With the relationship reversed, it is no longer fair to match based upon migration, as it is now the outcome variable. Rather matching was conducted using health status instead. Research has shown that self-rated health status varies socially and demographically (Lindeboom and van Doorslaer, 2004; Dorling and Barford, 2009). Therefore matching is required to control for selection bias, to be able to accurately assess differences by location (measured using location at time point 'B' i.e. the destination). Matching was employed to allow the correct model specification, controlling for age, sex and income.

Figure 7.6 shows the distribution of data for the matching variables by health status of those who migrated between waves. Poor health extends further into the life course, being less prevalent in earlier ages than good health. Income shows a greater concentration of poor health in lower incomes. However the differences by status for migrants are not wide. Sex was not included in the graphs, but instead can be seen in

Table 7.19. There is a clear difference, with females more likely to report poor health than males. This mismatch (since female health is generally better) is because women tend to be more concerned with their health and as a result, use health services more often (Lindeboom and van Doorslaer, 2004; Young et al., 2010). Overall, selection bias is evident and therefore matching is warranted to reduce its effect on estimates.

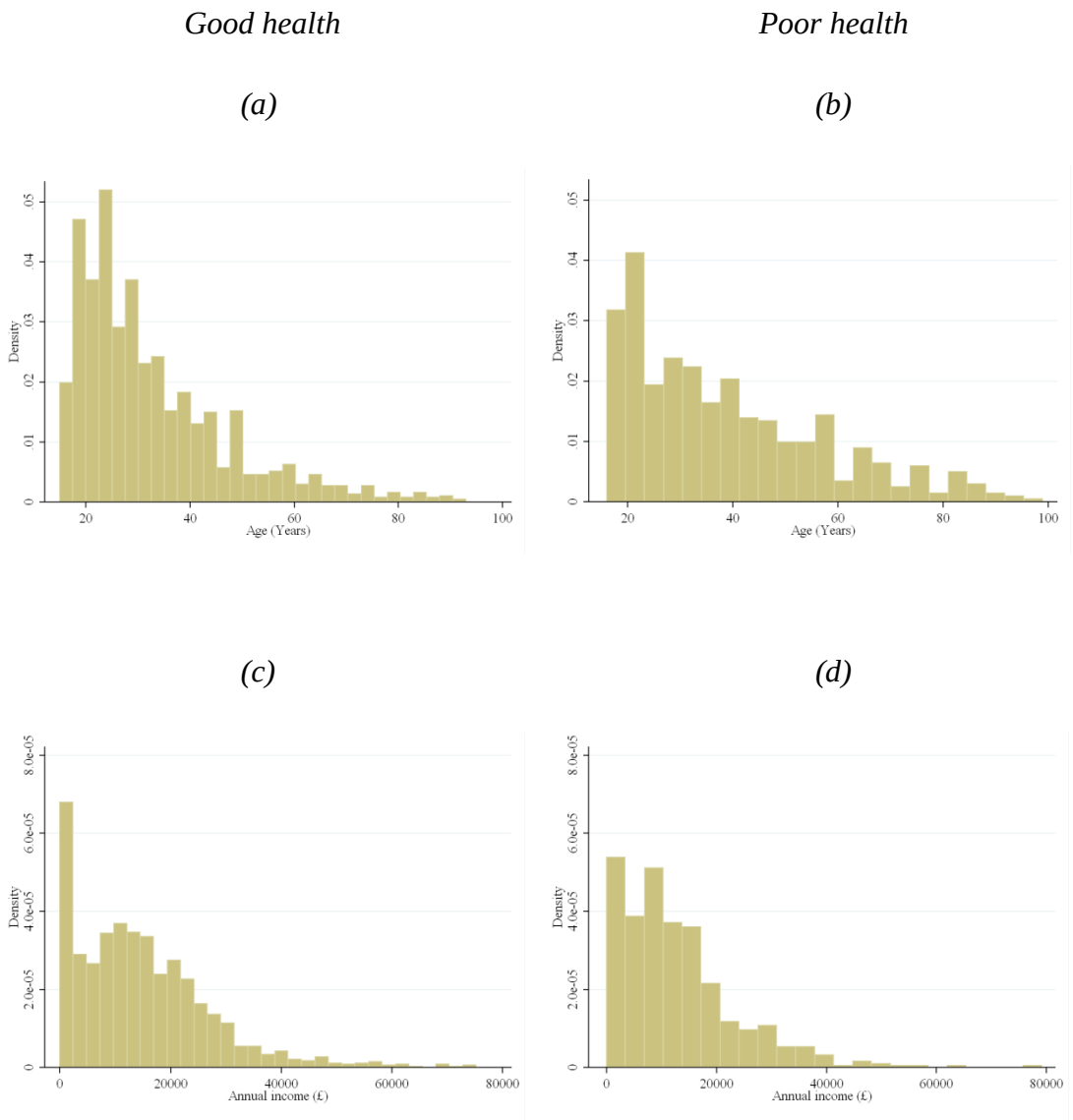


Figure 7.6: The distribution of people who migrated in year-B of the pooled data split by health status across age and annual income. (a) Age and people with good health; (b) age by people with poor health; (c) the range of income for those with good health; (d) annual income levels of people with poor health.

Note: Annual Income was cut off at £80,000 for those with good health to improve the comparison in the distribution. There were only eight other cases, of which four were below £100,000 and the largest was £470,259.

	Good Health	*Poor Health*
Males	75	25
Females	69.7	30.3

Table 7.19: The percentage variation by gender of people reporting poor and good health of internal migrants.

The dataset contained 557 people who reported poor health, with 1459 who reported their health as good. The overall multivariate L_1 statistic (.337) showed that matching produced a fairly balanced data set. Table 7.20 shows the improvement in the balance of the variables after matching, each declining to become near perfect balanced. Although the variables were not particularly imbalanced before, it is less about producing a vastly different model, rather about specifying the correct model for analysis (Iacus et al., 2011b). It is also useful in controlling for their effects as well.

Variable	Before matching	After matching
Age	0.164	0.092
Sex	0.104	<0.001
Income	0.124	0.051

Table 7.20: Change in the L_1 statistic through the matching process by variable.

7.7.2 Analysis

To test whether there is evidence of health selective migration, a multi-nomial regression was run using the matched data. Health status prior to migration was used to explain cluster membership. 'Best Health and Most Desirable' was used as the reference cluster, since it displayed the best mortality profile making comparisons easier. The pseudo-r^2 value of this was 0.005, which may indicate that health status is a poor predictor of migratory location by itself. Yet this matters less, as this is not the objective of our analysis. The model was significantly different from the null model (χ^2=34.63, p<0.001), showing it to be useful.

The results of the analysis (Table 7.21) showed evidence of health selective migration. With the reference cluster being the cluster with the best mortality profile (i.e. lowest mortality rates across the majority of the variables), for all bar the cluster 'Mixed Experiences' there was a significant effect found. The significant coefficients were always positive, showing that people with poorer health were more likely to migrate to each cluster when compared to the cluster with the best health outcomes. The strength

of each coefficient was generally related to the mortality profile of that cluster, with the effect larger where the cluster represented higher mortality rates. If the analysis examines only people who actually migrated to a different cluster (i.e. no intra-cluster migration), there is little change in the result showing it to be a strong relationship (results not shown).

The results support the findings of past research (for example Bentham, 1988; Brimblecombe et al., 2000; Wannamethee et al., 2002), who found evidence of health selective migration across England and Wales. Unlike the other health selective migration literature, this finding is using data over single years rather than comparing two points in time separated a long time apart. These processes are not occurring over the life course overall, but play out over the short term. It is also consistent over more area types than compared to a simple binary split of the data as in previous research.

Cluster	Variable	Coefficient	Std. Error	P
Best Health and Most Desirable	(base outcome)			
Average Mortality Profiles	Health Status	0.660	0.182	<0.001
	Constant	-0.563	0.095	<0.001
Good Health Areas	Health Status	0.356	0.176	0.043
	Constant	-0.215	0.085	0.012
The Middle	Health Status	0.652	0.184	<0.001
	Constant	-0.603	0.096	<0.001
Poor Health Experiences	Health Status	0.581	0.206	0.005
	Constant	-0.943	0.108	<0.001
Poorest Health and Least Desirable	Health Status	1.079	0.238	<0.001
	Constant	-1.687	0.144	<0.001
Poorest Neuro-degenerative Health	Health Status	0.685	0.257	0.008
	Constant	-1.654	0.142	<0.001
Mixed Experiences	Health Status	0.173	0.181	0.341
	Constant	-0.198	0.085	0.020

Table 7.21: Results of the multinomial regression.

Migration (to an extent) helps reinforce inequalities in health. With those of good health migrating to the better health areas, and those with poor health left 'drifting' to the worst areas (Riva et al., 2011), health patterns begin to polarise. With the poorer

health areas also being more likely to contain higher levels of poverty (and *vice versa*), the resulting polarisation is also indirectly in terms of social conditions as well, with those of poorer health ending up in those less socio-economically disadvantaged areas. This relationship was shown earlier indirectly though the observed social mobility (Section 7.5), helping to explain this process.

7.7.3 A comparison to GORs

To evaluate how useful the classification is, the analysis was repeated replacing the clusters with the GORs. The same matching procedures were used as detailed in 7.7.1, since only the outcome variable has changed.

To test for health selective migration, a multi-nomial regression model was also fitted on the data. The GOR 'North East' was selected as the reference region to make comparisons to, since it has been shown in both Chapters 5 and 6 to perform worst, making it useful as the base outcome. Although the pseudo-r^2 was particularly low (0.003) suggesting a poor model, it was significantly different ($p = 0.008$) from the null model showing it to add value to the analysis.

Table 7.22 presents the results from the model. There are few significant relationships across the model. Only two regions present significant relationships, with people who reported their health as poor being less likely to migrate to either the 'South East' or the 'South West', in comparison to the 'North East'. The South East and South West Regions would reflect a possible escalator region effect (Champion, 2012), attracting individuals with better health towards them (although not socially as shown in Table 7.9).

Whilst this represents health selective migration, the lack of consistent significant relationships across the whole GORs in the analysis (compared to using the classification) does not present strong evidence for health selective migration in this context. The insignificant relationship for Wales also shows that this is not an explanatory factor for the previous evidence of a regional effect of migration on health.

This result is contrary to other work such as Wannamethee et al. (2002) who showed evidence of health selective migration geographically in England. This is due to a simplified concept of place, using only a North-South split over a long time period. The

analysis presented here fits the correct model specification, showing the hypothesised relationships to be less strong.

GOR	Variable	Coefficient	Std. Error	P
North East	(base outcome)			
North West	Health Status	-0.412	0.301	0.17
	Constant	1.278	0.176	<0.001
Yorkshire	Health Status	-0.181	0.301	0.547
	Constant	1.098	0.180	0.012
East Midlands	Health Status	-0.194	0.307	0.528
	Constant	0.967	0.183	<0.001
West Midlands	Health Status	-0.418	0.320	0.191
	Constant	0.878	0.186	<0.001
East	Health Status	-0.543	0.308	0.078
	Constant	1.194	0.178	<0.001
London	Health Status	-0.558	0.332	0.093
	Constant	0.781	0.188	<0.001
South East	Health Status	-0.784	0.299	0.009
	Constant	1.594	0.171	<0.001
South West	Health Status	-0.664	0.301	0.027
	Constant	1.457	0.173	<0.001
Wales	Health Status	-0.137	0.279	0.622
	Constant	1.827	0.168	<0.001

Table 7.22: The results of a multinomial regression analysing GOR location of migrants by health status.

Adding greater detail to the analysis shows that there is a less clear definition of the pattern of health selective migration. It is only significantly found at the extremes of the country, rather than being a consistent pattern. This would explain why it was found between the North and the South. More importantly, the lack of significant relationships shows the classification to capture greater detail at testing for health selective migration than the GORs. The classification is a more useful tool in discriminating patterns in the data.

7.8 Conclusion

The final analytical Chapter of this thesis concerns the role that internal migration plays in affecting health. The conceptual framework used was that if area effects were important, then there may be an observed effect as people migrated between different area types. This approach has not been taken before, as the focus of past research has been on static neighbourhood environments (Van Ham and Manley, 2012) or just the impact upon health of migration itself (Brimblecombe et al., 2000; Larson et al., 2004).

The analysis has shown no consistent evidence that migration by area type modifies any impact upon health. Table 7.16 showed some evidence of area effects in the likelihood of migrants reporting their health as poor, however this was only significant for two area types. These results are contrary to the neighbourhood effects hypothesised, supporting other research which argues that neighbourhood effects operate over long time scales (Hedman, 2011; Johnson et al., 2012; Musterd et al., 2012; Quillian, 2003).

Although there are some observational effects found, it appears more plausible that the relationship exists in the other direction. Evidence showing the existence of both health and social selective migration was more consistent and stronger, showing it to be the more dominant process operating. These results could quite easily be driving the significance of any results suggestive that migration to an area has an effect on health.

The classification was shown to capture greater detail in the analysis of migratory patterns. The inclusion of GORs showed few significant relationships at all, showing it to be less useful. This is because the GORs represent large geographical regions, in which the majority of migration remained within. The classification adds value to the analysis through capturing social and mortality dimensions, whilst maintaining a geographical perspective as well.

Chapter 8: Conclusions

8.1 Introduction

The final chapter of this thesis looks to summarise and present the useful contribution of this thesis to the fields of research it draws from. It begins through discussing the research findings of the thesis, with respect to the aims and objectives set out at the start of thesis. This will assess the quality of the results produced throughout the study to evaluate what it has achieved. Then the limitations of the research are discussed to outline how the thesis could be improved upon. Next, a series of possible future extensions to the research are presented which could build upon the understanding introduced through the findings of this thesis. The chapter ends with a concluding statement.

8.2 Research findings

> *"The availability of computer packages of classification techniques has led to the waste of more valuable scientific time than any other "statistical" innovation"*
>
> (Cormack, 1971, p321)

In a speech to the Royal Statistical Society in 1971, the mathematician Cormack attacked the use of classificatory techniques in research, proclaiming them to be nothing but a bit of fun. However much has changed in the 42 years since his talk. Examples of research, particularly area classifications, have increased our understanding of how individuals are clustered across populations (Harris et al., 2005; Openshaw et al., 1994; Vickers, 2006). This thesis has shown how such an approach is useful for researching health and mortality. This section evaluates the success of the thesis against the aims set out in Chapter 1 (Section 1.2), to show the main findings of the thesis.

I. Create a classification of mortality patterns of small areas for England and Wales, with a clearly justified open methodology.

This is the core aim of the thesis. Achieving the aim began in Chapter 2, where the literature was reviewed to assess the advantages and disadvantages of creating an area classification for researching health and mortality (Objective 2). The main methods used to analyse the geographical patterning of mortality in the literature were limited. They focused on analysing single causes of mortality, restricting comparisons and interactions between causes. This limits the overall understanding of patterns and processes.

An area classification was identified as a response to tackle this gap in research. It allows the development of a multi-dimensional tool through summarising areas by defining a structure to the data (Everitt, 1979). This helps manage the complex patterns in the data, to give an efficient understanding of reality that may have remained hidden otherwise (Dorling, 2012). It also maintains a geographical perspective to the explorative analysis.

Despite presenting evidence that showed the importance of a geographical understanding of mortality patterns (Section 2.3), there had only been minimal applications throughout the literature of the use of area classifications in health and mortality research. Of those that had, they were limited in their quality (CACI, 2010; NOO, 2009; Shelton et al., 2006). This was despite calls from both researchers and the Government (Abbas et al., 2009; DoH, 2005). Rather most evidence showing how they could add a useful dimension to research in these fields has been gathered from other fields. A gap existed for a high quality and fine scale application of this approach.

Other directions were suggested, however these failed to offer the same benefits of an area classification (especially the simplification of patterns). Limitations of the approach were outlined to help develop the approach, as well as evaluating if it was useful.

Chapter 3 discussed the data and methodological decisions made in creating the classification. It was based upon Milligan and Cooper's (1987) schema, which was adapted for the creation of an area classification (Section 3.1). The data were reviewed and manipulated to prepare it for inclusion in an area classification (Objective 3). This included outlining the data set used, discussing the variables to be included and how they would be useful in building the area classification, and assessing data quality and

issues. The geographical scale of analysis was chosen after evaluating which would be best for this study (Section 3.4). Based upon these steps, the inputs for the classification were compiled. This also included the standardisation and weighting of variables to improve the quality of the inputs. These steps allowed the creation of high quality inputs through clear decisions.

Following this, the methodological options were reviewed and evaluated. Hierarchical clustering, partitional clustering, neural networks and fuzzy cluster were all evaluated. A partitional k-means methodology was chosen since it would be implemented most effectively given the aims and data involved in the study (Section 3.5.5; Objective 4). With the large data size, the method would run most efficiently, saving computational time (Gordon, 1999). Furthermore, although the method requires the selection of the number of clusters prior to the analysis, it does produce the optimal solution for that number and therefore gives a better result than other methods (Everitt et al., 2001; Gordon, 1999).

The steps to running a k-means analysis on the data set were detailed clearly throughout both Chapters 3 and 4, including the selection of the number of clusters in the model, the calculation of seed points and the choice of measure for measuring similarity. This was important to allow the resulting classification to be evaluated. Based upon these stages, the classification was created (Objective 5).

The final task to ensure a high quality area classification was to statistically test the resulting model, to assess whether it is stable (Objective 6). There are few methods for testing the robustness of an area classification. It was identified through the literature that performing a replication analysis, examining the impact of outliers and assessing variable sensitivity were the most important factors to test (Section 3.9). Each test showed relative stability of the main underlying structure of mortality patterns that the area classification captures (Chapter 4).

In summary, an area classification was created based upon observations from the literature review showing that it would be a useful. Data was collected and compiled to be entered in the model. The most relevant methodology was selected and implemented, detailing clearly each step in its process. Testing showed a stable and robust classification. These steps led to the successful creation of an area classification analysing mortality patterns for small areas in England and Wales, produced with an open and detailed methodology.

II. Understand the dominant mortality patterns and why this segmentation exists across the areas within each cluster.

To achieve this aim, firstly the individual clusters were interpreted in Chapter 5 to examine what they represent (Objective 7). The cluster centres were explored, showing the average characteristics of each cluster (Tables 5.1 to 5.6). They summarised the main mortality patterns found across England and Wales. The main differentiation between the clusters was in terms of prevalence, each generally showing a different degree of mortality rates. There was little interaction of causes shown through a single level classification.

There was some interaction of causes. Whilst both 'Poorest Health and Least Desirable' and 'Poorest Neurodegenerative Disease' initially appeared to be very similar (clusters with high mortality rates), they were slightly concentrated in different groups of causes. This is important with regards policy implementations, as these differences require varying approaches to improve the health of an area. Knowing the health needs is more efficient than basing policy on demographic factors, which ignore variations in mortality experiences.

Health related statistics were calculated to extend the understanding of the clusters beyond the inputs. This was because a high mortality rate is not necessarily bad, given that everyone dies. Life expectancy and premature mortality rates were calculated and showed wide inequalities captured between the clusters. For example, there was a gap of 9 years in male life expectancy between the highest and lowest cluster values (Table 5.10). This was better than compared to equivalent measures for analysing life expectancy variations (ONS, 2010; Smith et al., 2010).

The fifth chapter progressed to analysing the geographic areas that made up each cluster, to further the understanding and explanation for why each cluster of areas exists (Objective 8). Beginning with demographic factors, the migration of elderly people to communal establishments was useful for explaining the clusters 'Poorest Neurodegenerative Health' and 'Mixed Experiences'. The elderly are gravitating to the same areas at the end of life, whether this is to nursing homes, downsizing their home or migrating elsewhere (for example retirement villages). However this is not universal and there is slight split between the two clusters, following the social desirability of the cluster.

Social factors were shown to have a stronger explanation of the other clusters. Where poverty was higher, clusters displayed worse mortality profiles. This reflected observations made in the literature review (Section 2.3). There were wide relatively differences in the measure, showing the classification to have been useful at capturing detail in the data. This pattern was reflected when social class was introduced. Furthermore the area classification also showed a varying level of inequality by cluster, with the social gradient less prominent where the mortality profile was better.

Visualising the area classification showed no discernable geographical pattern, showing it to be capturing patterns related to mortality and social phenomenon's (Figure 5.1). However there were some geographical differences aiding the understanding of the clusters. Clusters with good health profiles were more commonly found in rural areas and Southern regions (and *vice versa*), highlighting a slight geographical split to the area classification. Nevertheless, calculating life expectancy split by both geographical measures and the clusters showed the clusters to negate variation in life expectancy. The area classification accounts for more variation in mortality patterns than can be attributed to geographical factors.

Tackling Objectives 7 and 8 have led to a greater understanding of what the individual clusters represent and has begun to explain why their differences exist. As such, this has achieved Aim II. This understanding allowed each cluster to be named accurately, to describe its characteristics to aid future applications.

III. Assess the extent that the area classification can help understand the existence of area effects.

As Everitt et al. (2001) argue, to be able to assess how useful an area classification is, it needs to be applied to evaluate what it can add to the understanding of a topic. This aim seeks to highlight the usefulness of the area classification through using it to analyse the existence of area effects.

Tackling this aim began in Chapter 1, through reviewing the association between geography and health (Objective 1). Evidence of area effects were evaluated to show that the environment individuals live in can have an independent impact upon their health. This was shown to be further extended to migration research, as if area effects were to matter, then any effect would be observed as individuals migrate. However,

there was little evidence for this approach being taken in research. Possible causal mechanisms were reviewed, which later informed the individual- and area-level risk factors that were controlled for in the subsequent analysis.

The analysis of area effects began through a multi-level analysis of whether the cluster individuals lived within had an effect on health (Objective 9a). This would show whether the clusters were more important than simply areas of similar mortality patterns. A significant effect was observed, which remained after accounting for individual covariates. However the level of variance accounted for by the area classification was lower than past research has found for area effects (Pickett and Pearl, 2001; Riva et al., 2007). Furthermore, this effect became insignificant once poverty was introduced into the model, showing the importance of this factor in explaining the area classification. The low number of clusters however, may have driven this insignificance given the variations shown descriptively (see Table 6.2).

The analysis was then extended to include the clusters as fixed effects, with the Middle Super Output Area (MSOA) individuals resided in, left to vary as the random effect. This showed stronger evidence for area effects, with MSOAs significantly accounting for 5.2 per cent of the total variance in the model after controlling for individual and area covariates (Table 6.8). This represents a similar sized effect to previous research (Pickett and Pearl, 2001; Riva et al., 2007). The level of poverty in an area accounted for the effects of the clusters again, bar 'Average Mortality Profiles' which remained significant, showing the area classification to add extra detail to the analysis.

Building upon this analysis in relation to gaps identified in the literature review (Sections 2.5 and 2.8), the analysis of area effects was extended in Chapter 7 to explore the impact on health as individuals moved between clusters (Objective 9b).

The chapter successfully introduced a new method which has been under-utilised in research and applied it to address issues surrounding selection bias, which have previously been ignored in past research. This allowed for a more accurate and appropriate set of findings.

The results did not show consistent evidence of area effects. Applying the area classification mostly resulted in insignificant results. Moving out of the clusters 'Good Health Areas' and 'Poor Health Experiences' showed significant impacts upon health. However once destination was accounted for, this effect disappeared. Moving into the cluster 'Good Health Areas' added the most to the model, as introducing cluster origin

into the model (Table 7.16) showed the effect of migration varied by cluster. Moving from a cluster with a worse mortality profile resulted in a greater likelihood of poorer health. Whilst only two out of seven origin variables were significant, this is a useful exploratory result as the other coefficients made sense even though they were insignificant.

Of the significant relationships found throughout the analysis, each showed that migration was associated with an increased likelihood of poor health (as opposed to not migrating). As the direction did not vary, this would show that the findings reflected a migratory effect. Nevertheless, with the results from Table 7.16, this would indicate that area type can modify this migration effect on health.

The patterns were stronger when health selective migration was tested. Individuals with poor health were found to be significantly more likely to migrate into the clusters with poorer mortality profiles than compared to moving to the cluster 'Best Health and Most Desirable'. It highlights how health inequalities are maintained and exacerbated, through the social and spatial sorting of the population. This would appear to be the dominant relationship when analysing migration and health, supporting previous research on this topic (Brimblecombe et al., 2000; Wannamethee et al., 2002).

Conducting the analysis for Objectives 9a and 9b achieved Aim III. Applying the area classification has developed the understanding of area effects, particularly with the new inclusion of migration into area effects research.

IV. Benchmark the results to a relevant means of grouping the data, to highlight its usefulness as a research tool.

Whilst the area classification improved the understanding of mortality patterns across England and Wales (Aim II) and area effects on health (Aim III), this does not show how useful it is as a research tool. Rather it needs to be compared to another approach for categorising areas, to be able to assess the importance of the results.

The area classification needs to be better than the status quo and the GORs were selected since they are currently used to report mortality statistics beyond the national level. The results applying the area classification do not need to be perfect, rather they need to add further understanding to analysing patterns and processes than compared to

using the GORs. The analyses were repeated, replacing the use of the clusters for discriminating variations in the data with GORs instead (Objective 10).

Table 8.1 shows all of the main analytical results conducted for both the area classification and the GORs. On eight out of nine of the tests, the area classification was found to be more useful compared to using the GORs instead. The difference in performance was greatest when the analysis was descriptive (measuring life expectancy, health status and change in size of groupings). This would indicate that it is better for displaying data, as it presents greater understanding of patterns and processes.

Test	Evidence	Result
Life Expectancy I	Tables 5.11	The area classification gave a range of values of 9.1 (males) and 7.9 (females). ONS (2010) reported the range for the GORs as 2.8 (males) and 2.7 (females).
Life expectancy II	Tables 5.21 and 5.22	The range of values within clusters (0.6 to 2.2) was less than within GORs (6.6 to 8.7).
Health status	Tables 6.2 and 6.3	A wider range of values using the clusters (16) than using GORs (12).
Health status standardised	Tables 6.4 and 6.5	Clusters had a greater range of values (58.9) than the GORs (42.7).
Multi-level analysis	Tables 6.6 and 6.7	Using the GORs gave a significant effect once individual and area level factors were controlled for. The area classification was insignificant.
Change in size of groupings	Tables 7.3 and 7.4	Less population retained using the clusters than compared to the GORs.
Characteristics of migrants	Tables 7.8 and 7.9	The multinomial regression saw 11 coefficients significant using the clusters, compared to the GORs where there were only two significant.
Impact of migration by area on health	Tables 7.10, 7.14, 7.17 and 7.18	Three significant relationships using the area classification, with two using the GORs.
Health selective migration	Tables 7.21 and 7.22	Using the classification resulted in seven out of eight of the coefficients being significant, whereas using GORs only gave two out of nine significant.

Table 8.1: Comparative results from the analysis.

With the more rigorous analytical techniques, the results favoured the area classification. The analysis of the characteristics of migrants showed the area classification to capture greater detail in who was moving between area types. This result was further replicated in the analysis of health selective migration. Whilst the comparative result for the analysis of the impact of migration by area type of health showed the area classification to be only just better, prior health negated the effect using

the GORs and the classification offered detail when both origin and destination were measured. The only analytical comparison where the GORs outperformed the area classification was in the multi-level analysis. Yet with only accounting for 0.8 per cent of the total variance, this shows that it is not particularly better than the area classification (unlike the other tests).

The successful application of the area classification is further highlighted since the GORs contained two more categories. Whilst in absolute terms, this may not appear much, proportionally the extra categories offer 20 per cent extra detail for helping to discriminate patterns in the data. This may explain why the GORs were significant in the multi-level analysis. It might have been better from the outset to have a ten cluster solution to ensure a fairer comparison, however this is not the point of building the area classification since it was shown to be less optimal (Chapter 4).

The GORs are less useful as an approach for grouping areas. This is because they are invisible administrative boundaries representing arbitrary geographical regions. The area classification focuses on drawing out the divisions in mortality patterns. Greater detail is therefore captured as social conditions vary greatly within GORs (Vickers, 2006; Woods et al., 2005), whereas the area classification groups them together into clusters of similar characteristics. The clusters focus on the factors that matter more for understanding patterns and processes.

Achieving this aim has shown the area classification to be a useful research tool, through comparing it to an equivalent measure to benchmark how good the results were. However the GORs are still useful. They offer a different dimension, focusing solely on geography. It is important to examine both measures when researching health and mortality to gain the most understanding.

8.3 Limitations

It is important to identify the main limitations of this thesis, so that future research can address and develop these issues. These are outlined in this section, discussed in chronological order to how they arose in the thesis structure.

The area classification created in this thesis is a static classification. It presents a snapshot of the mortality patterns across England and Wales between 2006 and 2009.

As a result, the classification is already out-of-date, remaining only applicable to that particular period (Harris et al., 2005; Vickers, 2006). However this may not be too problematic, given that changes in mortality patterns occur slowly, since the main causes of mortality operate over a long time period (Musterd et al., 2012). This is highlighted in Figure 8.1, which shows that mortality trends are more affected by changes to coding rules, than societal influences, with trends occurring on long time scales (Rooney and Smith, 2000). Whilst there may be changes in the cluster some areas lie in, it would not be expected that the patterns captured by the main clusters would have changed since its developed, meaning that any application would still be useful.

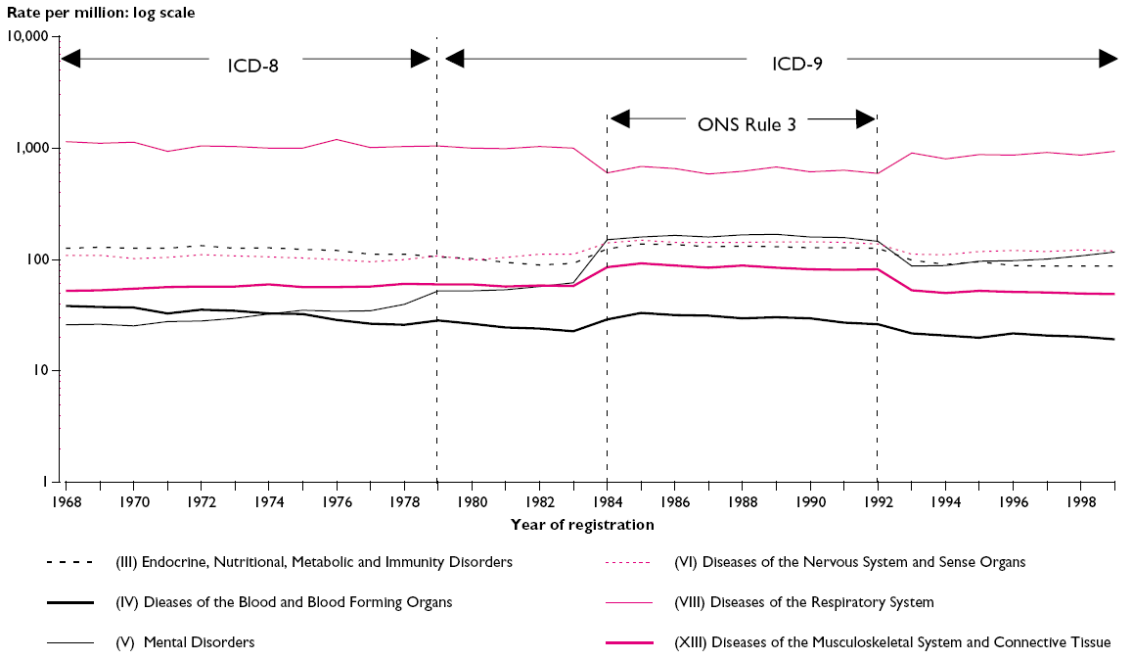


Figure 8.1: The age-standardised mortality rates for women in England and Wales 1968-99 by cause (ICD-Chapter) (Rooney and Smith, 2000).

Some decisions made throughout both the creation of the classification and its subsequent testing (Chapter 4) could be argued to be subjective. There are no set statistics which can objectively answer each step to building an area classification (Milligan and Cooper, 1987). For example, there is no set approach to deciding the number of clusters to include (Milligan and Cooper, 1985). Whilst a limitation, it is also a benefit of the methodology, giving a good level of freedom and variation in how an area classification is constructed. To minimise any issue throughout the study, any decision which is not objectively made has been made with clear and justified reasoning, based upon a variety of factors. This ensures that the choices made are

shown to be the most useful within the situation, not relying on a single measure. This is, however, a wider criticism of the field (Everitt et al., 2001; Gordon, 1999).

The testing procedures applied in Chapter 4 are all fairly ad hoc. This is because they are often testing the optimising of a solution (which is thought to be the best). This solution (the rules and parameters) may not apply once a variable has been removed, rather you may be comparing a less optimal (or even wrong) solution to the main classification to assess the assumed impact (Everitt et al., 2001). However there are no set tests to perform when creating a classification (Gordon, 1999). Testing is important and the literature review of the methodology was important in identifying the few tests which could be applied in the study.

True stability cannot always be achieved (or inferred) with unsupervised approaches (Everitt et al., 2001). The majority of the tests conducted have shown that the results do not vary much, especially in terms of scale as well. Stability can be inferred, albeit with some caution. Yet even with distinct clusters, there is likely to be some change, especially in those cases which lie towards the edges of groups. This is because the parameters and conditions that exist in the process change with each test and hence something different will always be created.

The influence of communal homes on the clustering processes was shown and this could be problematic in the results produced. With their higher death rates, they can lead to bias in the mortality measures input into the classification (Williams et al., 1995; 2004). They were particularly important in explaining both 'Poorest Neurodegenerative Health' and 'Mixed Experiences'. However ignoring them would give a false geography which does not reflect reality (Dorling, 2012). Instead, communal homes should be viewed as helping to understand the geographical patterns of mortality through the area classification. They reflect the changing nature of death, as their concentrated impact has become more prevalent with an ageing population (Williams et al., 2004). A possible extension of the analysis to tackle this issue is proposed in Section 8.4.1.

The use of self-rated health is also limited in its application within the analyses presented. Whilst the measure usefully captures variations in actual health (Idler and Benyamini, 1997; Jylhä, 2009), the issue lies in how it is measured. As a binary measure, the data becomes bounded between 'good' and 'poor' health. There is no possibility of health improving if it was already 'good' (and *vice versa*). As such, there

is lost variation at both ends of the distribution, which cannot be accounted for in the analysis, limiting the results. It would be useful to extend the analyses using a different variable that does not suffer from this problem to truly test any effects. However this is limited by the use of the British Household Panel Survey, which has no other variable which would solve this problem. It has also been used in past research in this form, showing that it was a useful measure (Jen et al., 2009; Kondo et al., 2009).

The lack of significant area effects observed through migration may be due to temporal limitations. For example Johnson et al. (2012) showed that the neighbourhood effect was more pronounced later in the life course, a factor of individuals being exposed to neighbourhood effects longer. As such, any effect by migration using yearly data may not have been expected to materialise in this analysis. However, the 'Moving to Opportunity' studies have shown evidence of the health benefit of migrating away from deprived areas within a year of migration (Leventhal and Brooks-Gunn, 2003; Ludwig et al., 2011; 2012). Therefore such an approach does appear warranted and provides additional understanding of the process within the short term.

The matching of individuals to create a comparable control group was largely successful. However not all migrations are equal, as people migrate for different reasons (for example employment, health, upgrading to new house). Furthermore, Turnstall et al. (2010) showed that where migration was for negative reasons (such as divorce or bereavement), there was a stronger observed effect of migration upon health. There is some data for this in the BHPS, however it is patchy and it is not clear how best to group together the different reason available as there are many (and even then, not all possible reasons are covered). Since we do not always know why people migrate, we do not really have an experimental design. This is something that future research would need to address to improve the quality of the results.

Finally, there is also an issue with the sample size in the migration analysis. Table 7.1 showed that only 8.6 per cent of the total data set migrated. This limited the power of the analyses once origin and destination of movers was considered. Pooling together the data helped the stability of the data set, however the results are limited since there may be a significant effect observed that is otherwise hidden by the sample size issue. The analysis should be conducted using a large sample size, although there are few which measure both migration and health (Larson et al., 2004). The Understanding Society data set would offer a possible solution (40,000 households), although the second wave was not available at the time of analysis to be used here.

8.4 Future extensions

In relation to the research findings and limitations identified throughout this chapter, a few examples of possible directions for further analysis and investigation are outlined in this section. These are expected to extend the mortality classification and its subsequent analysis to further research.

8.4.1 Bayesian modelling

The classification's data set was compiled through using indirect Standardised Mortality Ratios (SMRs). This gave the relative risk of mortality for a particular cause in an area compared to what would be expected based upon national trends (Section 3.7.4).

SMRs can be particularly unstable, especially for small area estimates (Williams et al., 1995). Since the measure operates as a ratio, where a cause has a low total of deaths nationally, its respective expected mortality will be low across a large number of small areas. Therefore small differences in the observed data can easily lead to large SMRs (Lawson, 2013).

Examination of the classification showed that there were not local clusters (i.e. small, extreme groups of areas). However there was some influence of communal homes, which partially explained the clusters 'Poorest Neurodegenerative Health' and 'Mixed Experiences'. Extreme data points may have had an impact upon the findings (Williams et al., 2004).

The smoothing of the mortality data would be a useful extension, given that some variables selected in the analysis had low prevalence once split into 7194 areas. Bayesian modelling (through 'Markov chain Monte Carlo' methods) offers a useful technique for accounting for extreme data points (Lawson, 2013). It works through calculating the probability of an event occurring in an area (i.e. mortality for a particular cause), based upon the distribution of the data observed (Ntzoufras, 2009). This gives a fairer account of the data, being less affected by spurious data. Spatial correlation can be introduced, improving the geographical accuracy of variables.

This approach has been useful in past research for smoothing SMRs for areas. For example, Strong et al. (2012) used Bayesian modelling to smooth alcohol-related mortality in South Yorkshire to reduce the effects of bias with the small counts

involved. The model was not just used to smooth the data, rather it was also used to analyse the distribution of data within explanatory factors (i.e. fixed effect covariates). This showed that deprivation had a greater impact upon males, than compare to females (a finding shown similarly using the classification and life expectancy previously; Chapter 5). Extending the model to incorporate a random effect by electoral ward produced a significant area effect on alcohol-related mortality, independent of explanatory factors.

There are limitations involved with introducing such an approach into building a classification of mortality. The main issue regards computational power. Bayesian modelling is an iterative and intensive methodology (Lawson, 2013; Ntzoufras, 2009). It is mainly used for the analysis of few variables (although usually only one outcome variable) across a relatively small number of areas. This could be problematic for 7194 areas, especially when introducing the spatial element of analyses (i.e. allowing for the distribution of variables to vary by geography).

8.4.2 Subsequent levels of the area classification

The area classification summarises the main mortality patterns of England and Wales. An extension to this understanding would be to create a second level to the area classification, by creating sub-groups for each cluster. Essentially, an area classification is created for each cluster, resulting in a hierarchy. It shows the area types within each cluster.

As Voas and Williamson (2001) argue, this is important for examining the diversity of area types within each cluster. For example, cause of death was shown to have less influence with a single level classification. By extending the area classification, it would examine whether it becomes more important for explaining differences within clusters. Examining the diversity of mortality profiles by social, demographic and geographical factors would also be useful for furthering the understanding of patterns and processes. This would improve the usefulness of the area classification to policy targeting and planning.

This extension does not just further the understanding of mortality patterns across England and Wales. A possible explanation of the insignificant result from the multi-level analysis of area effects was the low number of clusters. Extending the area

classification to a second level would address this issue and allow a fairer analysis to examine whether the clusters exhibit an independent effect on health.

To exploratory test this, a k-means cluster analysis was performed individually on the areas within each cluster to create sub-groups for each (following the same methodology to the main area classification). A range of solutions were examined for each cluster between 2 and 5 sub-groups, based upon previous area classifications (CACI, 2013; Vickers, 2006). However, a future in depth study might look at a wider range of values. The decision of the number of sub-groups was made based upon Scree plots of mean distance of cases to their cluster centers and also cluster sizes (see Section 3.8.4). Where clusters did not have a second level, this was because further levels produced local clusters (i.e. a few areas as a cluster) and therefore were left as a single level to avoid any issues. The hierarchical structure of the data is shown in Table 8.2.

Cluster	Number of sub-groups
1	3
2	4
3	3
4	1
5	5
6	2
7	3
8	1

Table 8.2: The number of sub-groups found within each cluster.

The same logistic multi-level regression model was then fitted (as found in Section 6.4.1), controlling for individual and area level risk factors. The results are presented in Table 8.3. However there was little difference when compared to the results in Table 6.6. The significant random effect of the area classification attenuated once area level factors were included (albeit capturing slightly more variance). A more refined second level would not appear to be useful for the analysis of area effects. Nevertheless, it may be that the area classification accounts for any area effects through how it segments the population. However the greater detail captured from this approach and its ability to be more discriminating within data would show this extension to be useful.

(a) Fixed effects model

Variable	Model 1		Model 2	
	OR	SE	OR	SE
Age	1.024***	0.001	1.025***	0.001
Male	1.022	0.052	1.02	0.052
Income	0.99999***	2.05E-06	0.99999***	2.05E-06
Trust	0.687***	0.038	0.698***	0.039
Civic	0.788***	0.048	0.794***	0.048
Meet	0.878	0.059	0.858*	0.058
Belong	0.709***	0.042	0.726***	0.043
Advice	1.001	0.055	0.993	0.055
Poverty			1.029***	0.004
Turnover			0.999	0.002

(b) Random effects parameters

Model	Variance	SE	P	ICC
Null	0.044	0.017	<0.001	0.013
1	0.035	0.016	<0.001	0.011
2	0.003	0.005	0.275	0.001

Table 8.3: Results from the multi-level logistic model using the second level of the area classification as the level-2 unit.

*Key: OR = odds ratios, SE = standard errors, P = significance, ICC = Intra-Class Correlation. Significance levels: p < 0.05 = *, p < 0.01 = **, p < 0.001 = ***.*

8.4.3 A Scottish area classification

As part of the mortality database used, there includes all deaths registered in Scotland. These were originally dropped from the study because of a lack of a comparable geographical scale, as well as few comparable explanatory area-level variables to help with understanding the analysis (Section 3.2).

It would be useful to produce a mortality area classification of Scotland, especially with the data available. Research has shown that Scotland displays different mortality patterns, including higher rates of premature mortality when compared to England and Wales particularly for lung cancer, ischaemic heart disease, alcohol-related causes, stroke and external causes (Griffiths and Fitzpatrick, 2001; Young et al., 2010). This is

referred in the literature as the 'Scottish effect', whereby the health of Scotland is poorer than England and Wales, independent of social and demographic factors (Hanlon et al., 2005). There are also wide variations within it, with Glasgow accounting for most of this poor health (Macintyre et al., 2008).

An independent area classification would be constructed to explore the geographical patterns that dominate Scotland. The following patterns and socio-economic correlates would provide an interesting comparison to the findings of this study, as well as the analytical approaches of Chapters 6 and 7 being applied with this new classification to explore whether this can add new understanding. Furthermore, the cluster centres for the area classification of England and Wales could be applied to the Scottish data to see how well it fits the data and whether this gives a varying result to an independent area classification.

8.4.4 Area classifications of affluence and poverty

The classification summarises the main mortality patterns for small areas within England and Wales. However taking small subsets of the population and classifying these would be a useful extension of this thesis through having a greater focus. This is because specific groups in the population may become lost in the area classification, as they become joined with other areas (although ones not dissimilar). Extracting this information can enable us to ask and answer concise questions, supplementing the results from the area classification of England and Wales. It would be useful to create separate area classifications of the poorest and most affluent areas.

Although the theme of poverty has been heavily research and is well established, research usually considers its effect to be linear (Keene et al., 2013). Whilst poverty has a negative effect on health, it assumes that all poor areas are the same. This ignores the multitude of experiences found in poorer areas (Dorling et al., 2007; Jin et al., 2011). Furthermore, often this research examines the direct impact of poverty against one cause of mortality, ignoring the interactions of all causes in an area. With the ability to now account for all causes of mortality, these issues can be addressed.

Whilst the poor represent a commonly researched group, the affluent do not (Hajat et al., 2010). Often forgotten in research, with the focus on tackling the negative effects of poverty, it may become easy to forget that the rich die too! This dominance of research

has led to calls that to fully understand the bottom of society, we need to consider what is happening at the top (Phillips, 2001). Although this call occurred over a decade ago, little research has addressed this imbalance, especially in health-based research. Certain causes of mortality dominate more affluent areas, for example malignant melanoma (Shaw et al., 2008). Yet whether there is a diversity of health throughout these areas is unknown.

Creating an area classification for each of these two area types would therefore be useful. The dominant mortality patterns within each are not well known and therefore this study would be innovative in its findings. Identifying the varying experiences of specific groups is also important for policy targeting and resource allocation.

An issue that would arise would be how best to measure these two competing factors. Although many measures of poverty exist (Townsend and Kennedy, 2004), there has been less consideration of affluence. Affluence is less well defined, consisting of different dimensions including income, wealth and social status (Butler and Watt, 2007; Dorling et al., 2007; Hajat et al., 2010). Reviewing and producing a measure to account for this would be useful, as well as necessary to producing the area classification.

8.5 Concluding statement

The fable 'The Fox and the Cat' begins with a fox and cat discussing the tricks they have learnt. The fox proclaims that it knows a wide variety of tricks, whereas the cat says that it knows only one trick very well. Upon the arrival of a hunter, the cat darts up a tree to safety (its one trick), whereas the fox is eventually caught as it cannot execute any of its tricks well enough. The moral of the story is to focus on being good at one skill, rather than spreading yourself too thinly across many different ones.

Academia has become full of cats, focusing on ever decreasing small areas of study. This has led to greater detail in studies and a deeper understanding of patterns and processes. However there is nothing wrong with being a fox. The classificatory approach taken in this thesis chooses to focus on clustering all causes of mortality, rather than on a small set of causes. The result produced a useful area classification, showing the interaction of causes and their varying distributions across England and Wales. Its application into the field of neighbourhood effects was useful in extending the understanding of patterns and processes. It highlights that a co-ordinated approach

is the only way to effective policy formulation, since causes do not happen in isolation. The ideas presented here, as well as the area classification itself, present a useful research approach for analysing mortality patterns across England and Wales. Hopefully it can spur on the growth of such techniques seen after their introduction into the field of geo-demographics.

References

Abbas, J, Ojo, A, Orange, S, 2009, Geodemographics – a tool for health intelligence, *Public Health*, Vol 123, No 1 (e-supplement), e35-e39.

Ahmed, FE, 2005, Artificial neural networks for diagnosis and survival prediction in colon cancer, *Molecular cancer*, Vol 4, No 29.

Alberman, E, Botting, B, Blatchley, N, Twidell, A, 1994, A new hierarchical classification of causes of infant deaths in England and Wales, *Archives of Disease in Childhood*, Vol 70, No 5, 403-409.

Andersson, R, Musterd, S, 2010, What scale matters? Exploring the relationships between individuals' social position, neighbourhood context and the scale of neighbourhood, *Geografiska Annaler: Series B, Human Geography*, Vol 92, No 1, pp23-43.

APHO (Association of Public Health Observatories), 2013, About Health Profiles, http://www.apho.org.uk/default.aspx?QN=HP_ABOUTUS2012, (last accessed 18th April 2013).

Aronow, WS, 1973, Smoking, Carbon Monoxide, and Coronary Heart Disease, *Circulation*, Vol 48, No 6, pp1169-1172.

Bell, BA, Morgan, GB, Kromrey, JD, Ferron, JM, 2010, The impact of small cluster size on multilevel models: A Monte Carlo examination of two-level models with binary and continuous predictors, *JSM Proceedings; Section on Survey Research Methods*, pp4057 –4067.

Bentham, G, 1988, Migration and morbidity: implications for geographical studies of disease, *Social Science and Medicine*, Vol 26, No 1, pp49-54.

Bentham, G, Haynes, R, Lovett, A, 1991, Introduction, *Social Science & Medicine*, Vol 33, No 4, ppix-x.

Berkman, L, Glass, T, 2000, Social Integration, Social Networks, Social Support, and Health, pp137-173, in Berkman, L, Kawachi, I, (eds), Social Epidemiology, New York: Oxford University Press.

Berkman, LF, 1995, The Role of Social relations in Health Promotion, *Psychosomatic Medicine*, Vol 57, No , pp245-254.

Bezdek, J, Pal, N, 1998, Some New Indexes of Cluster Validity, *IEEE Transactions on Systems, Man and Cybernetics - Part B: Cybernetics*, Vol 28, No 3, pp301-315.

Blackwell, M, 2012, *Personal Correspondence*.

Blackwell, M, Iacus, S, King, G, Porro, G, 2009, CEM: Coarsened Exact Matching in Stata, *The Stata Journal*, Vol 9, No , pp524-546.

Blashfield, RK, 1976, Mixture model tests of cluster analysis. Accuracy of four agglomerative hierarchical methods, *Psychological Bulletin*, Vol 83, No 3, pp377-385.

Blashfield, RK, Mcintyre, R, 1980, A nearest-centroid technique for evaluating the minimum-variance clustering procedure, *Multivariate Behavioural Research*, Vol 2, No , pp225-238.

Booth, C, 1889, *Life and Labour of the People of London*, London: Williams and Norgate.

Boyle, P, 2004, Population geography: migration and inequalities in mortality and morbidity, *Progress in Human Geography*, Vol 28, No 6, pp767-776.

Bradley, DE, Van Willigen, M, 2010, Migration and Psychological Well-being Among Older Adults: A Growth Curve Analysis Based on Panel Data From the Health and Retirement Study, 1996-2006, Journal of Aging and Health, Vol 22, No 7, pp882-913.

Briggs, XDS, Keys, BJ, 2009, Has Exposure to Poor Neighbourhoods Changed in America? Race, Risk and Housing Locations in Two decade, *Urban Studies*, Vol 38, No , pp429-458.

Brimblecombe, N, Dorling, D, Shaw, M, 2000, Migration and geographical inequalities in health in Britain, *Social Science & Medicine*, Vol 50, No , pp861-878.

Brown, G, Fearn, V, Wells, C, 2010, Exploratory analysis of seasonal mortality in England and Wales, 1998 to 2007, *Health Statistics Quarterly*, Vol 48, pp58-80.

Burrough, PA, Wilson, JP, van Gaans, PFM, Hansen, AJ, 2001, Fuzzy *k*-means classification of topo-climatic data as an aid to firest mapping in the Greater Yellowstone Area, USA, *Landscape Ecology*, Vol 16, No 6, pp523-546.

Butler, T, Watt, P, 2007, *Understanding Social Inequality*, London: SAGE.

CACI, 2010, HealthACORN User Guide, http://www.caci.co.uk/HealthACORN.aspx, (last accessed 7th September 2010).

CACI, 2013, The ACORN User Guide, http://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf, (last accessed 8th May 2013).

Caliendo, M, Kopeinig, S, 2008, Some practical guidance for the implementation of propensity score matching, *Journal of Economic Surveys*, Vol 22, No 1, pp31-72.

Campos, C, Dent, A, Fry, R, Reid, A, 2010/2011, Impact of the recession, *Regional Trends*, Vol 43, pp1-69.

Catney, G, Simpson, L, 2010, Settlement area migration in England and Wales: assessing evidence for a social gradient, *Transactions of the Institute of British Geographers*, Vol 35, No 4, pp571-584.

Chadwick, E, 1842, Report on the sanitary condition of the laboring population of Great Britain, London: W. Clowes and Sons.

Chaix, B, Rosvall, M, Merlo, J, 2007, Neighbourhood Socioeconomic Deprivation and Residential Instability: Effects on Incidence of Ischemic Heart Disease and Survival After Myocardial Infarction, *Epidemiology*, Vol 18, no 1, pp104-111.

Champion, T, 2012, Testing the return migration element of the 'escalator region' model: an analysis of migration into and out of south-east England, 1966-2001, *Cambridge Journal of Regions, Economy and Society*, Vol 5, No 2, pp255-270.

Chan, A, King, J, Flenady, V, Haslam, R, Tudehope, D, 2004, Classification of perinatal deaths: Development of the Australian and New Zealand classifications, *Journal of Paediatrics and Child Health*, Vol 40, No 7, pp340-347.

Cheng, R, Milligan, G, 1996, Measuring the Influence of Individual Data Points in a Cluster Analysis, *Journal of Classification*, Vol 13, No , pp315-335.

Chiang, CL, 1972, On Constructing Current Life Tables, *Journal of the American Statistical Association*, Vol 67, No 339, pp538-541.

Christensen, K, Ørstavik, KH, Vaupel, JW, 2001, The X Chromosome and the Female Survival Advantage, *Annals of the New York Academy of Sciences*, Vol 954, pp175-183.

Clatworthy, J, Buick, D, Hankins, M, Weinmann, J, Horne, R, 2005, The use and reporting of cluster analysis in health psychology: A review, *British Journal of Health Psychology*, Vol 10, No 3, pp329-358.

Coggon, D, Harris, EC, Brown, T, Rice, S, Palmer, KT, 2010, Work-related mortality in England and Wales 1979-2000, *Occupational and Environmental Medicine*, Vol 67, No 12, pp816-822.

Cohen, A, Gnanadesikan, R, Kettenring, J, Landwehr, J, 1977, Methodological developments in some applications of clustering, pp141-162, in Krishnaiah, P, (ed), *Applications of Statistics*, North-Holland Publishing: Oxford.

Congdon, P, 1996, The Epidemiology of Suicide in London, *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, Vol 159, No 3, pp515-533.

Connolly, S, O'Reilly, D, 2007, The contribution of migration to changes in the distribution of health over time: Five-year follow-up study in Northern Ireland, *Social Science & Medicine*, Vol 65, No 5, pp1004-1011.

Cormack, R, 1971, A Review of Classification, *Journal of the Royal Statistical Society. Series A (General)*, Vol 134, No 3, pp321-367.

Cornfield, J, Haenszel, W, Hammond, E, Lilienfeld, A, Shimkin, M, Wynder, E, 2009, Smoking and lung cancer: recent evidence and a discussion of some questions, *International Journal of Epidemiology*, Vol 38, No 5, pp1175-1191.

Cox, M, Boyle, P, Davey, P, Morris, A, 2007, Does health-selective migration following diagnosis strengthen the relationship between Type 2 diabetes and deprivation?, *Social Science & Medicine*, Vol 65, No 1, pp32-42.

Davies, D, Bouldin, D, 1979, A Cluster Separation Measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 1, No 2, pp224-227.

Day, P, Pearce, J, Dorling, D, 2008, Twelve worlds: A geo-deomographic comparison of global inequalities in mortality, *Journal of Epidemiology and Community Health*, Vol 62, No 11, pp1002-1010.

Department of Health (DoH), 2005, Choosing Health: Making health choices easier, http://webarchive.nationalarchives.gov.uk/+/www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/Browsable/DH_4097491, (25th September 2011).

Department of Health and Social Security, 1980, *Inequalities in health: report of a research working group (The Black Report)*, London: DHSS.

Devis, T, Rooney, C, 1999, Death certification and the epidemiologist, *Health Statistics Quarterly*, Vol 1, pp21-33.

Diez-Roux, A, 2001, Investigating Neighborhood and Area Effects on Health, *American Journal of Public Health*, Vol 91, No 11, pp1783-1789.

Doran, T, Drever, F, Whitehead, M, 2004, Is there a north-south divide in social class inequalities in health in Great Britain? Cross sectional study using data from the 2001 census, *BMJ*, Vol 328, No 7447, pp1043-1045.

Dorling, D, Rigby, J, Wheeler, B, Ballas, D, Thomas, B, Fahmy, E, Gordon, D, Lupton, R, 2007, Poverty, wealth and place in Britain, 1968 to 2005, Policy Press: Bristol.

Dorling, D, 2012, *The Visualisation of Spatial Social Structure*, Chichester: Wiley.

Dorling, D, Barford, A, 2009, The inequality hypothesis: Thesis, antithesis, and synthesis?, *Health and Place*, Vol 15, No 4, pp1166-1169.

Dorn, H, 1956, Some Problems for Research in Mortality and Morbidity, *Public Health Reports*, Vol 71, No 1, pp1-5.

Duda, RO, Hart, PE, 1973, *Pattern Classification and Scene Analysis*, New York: Wiley.

Durkheim, E, 1897 [1951], *Suicide: a study in sociology*, Glencoe: The Free Press.

Erskine, S, Maheswaran, R, Pearson, T, Gleeson, D, 2010, Socioeconomic deprivation, urban-rural location and alcohol-related mortality in England and Wales, *BMC Public Health*, Vol 10, No 99.

Evandrou, M, Falkingham, J, Green, M, 2010, Migration in later life: evidence from the British Household Panel Survey, *Population Trends*, Vol 141, pp77-94.

Evans, GW, Kim, P, 2007, Childhood Poverty and health: Cumulative Risk Exposure and Stress Dysregulation, *Psychological Science*, Vol 18, No 11, pp953-957.

Everitt, B, 1979, Unresolved Problems in Cluster Analysis, *Biometrics*, Vol 35, No 1, pp169-181.

Everitt, B, Landau, S, Lesse, M, 2001, *Cluster Analysis*, London: Arnold.

Experian, 2009, MOSAIC UK – the consumer classification of the United Kingdom (Brochure), http://www.experian.co.uk/assets/buisness-strategies/brochures/mosaic-uk-2009-brochure-jun10.pdf, (last accessed 28th January 2013).

Farr, W, 1864, *Supplement to the 25th annual report of the registrar General*, HMSO.

Fitzpatrick, Am, Teague, WG, Meyers, DA, Peters, SP, Li, X, Li, H, Wenzel, SE, Aujla, S, Castro, M, Bacharier, LB, Gaston, BM, Bleecker, ER, Moore, WC, 2011, Heterogeneity of severe asthma in childhood: Confirmation by cluster analysis of children n the National Institutes of health/national Heart, lung, and Blood Institute Severe Asthma Research Program, *Journal of Allergy and Clinical Immunology*, Vol 127, No 2, pp382-289.

Flowerdew, R, Manley, D, Sabel, C, 2008, Neighbourhood effects on health: Does it matter where you draw the boundaries?, *Social Science & Medicine*, Vol 66, No 6, pp1241-1255.

Franceschini, J, Jardim, JR, Fernandes, AL, Jamnik, S, Santoro, IL, 2013, Relationship between the magnitude of symptoms and the quality of life: a cluster analysis of lung cancer patients in Brazil, *Jornal Brasileiro de Pneumologia*, Vol 39, No 1, pp23-31.

Fukouka, Y, Lindgren, T, Rankin, S, Cooper, B, Carroll, D, 2007, Cluster analysis: a useful technique to identify elderly cardiac patients at risk for poor quality of life, *Quality of Life Research*, Vol 16, No 10, pp1655-1663.

Gartner, A, Farewell, D, Roach, P, Dunstan, F, 2011, Rural/urban mortality differences in England and Wales and the effect of deprivation adjustment, *Social Science & Medicine*, Vol 72, No 10, pp1685-1694.

Giatti, L, Barreto, SM, Comini César, C, 2008, Household context and self-rated health: the effect of unemployment and informal work, *Journal of Epidemiology and Community Health*, Vol 62, No 12, pp1079-1085.

Giggs, JA, Bourke, JB, Katschinski, B, 1988, The epidemiology of primary acute pancreatitis in Greater Nottingham, *Social Science & Medicine*, Vol 26, No 1, pp79-89.

Gordon, AD, 1996, Hierarchical classification, pp65-121, in Arabie, P, Hubert, L, De Soete, G, (Eds), *Clustering and Classification*, World Scientific: London.

Gordon, AD, 1999, Classification (2nd Edition), Chapman and Hall/CRC: London.

Gould, AC, Apparicio, P, Cloutier, M-S, 2012, Classifying Neighbourhoods by Level of Access to Stores Selling Fresh Fruit and Vegetables and Groceries: Identifying Problematic Areas in the City of Gatineau, Quebec, *Canadian Journal of Public Health*, Vol 103, No 6, ppe433-e437.

Gregory, IN, 2009, Comparisons between geographies of mortality and deprivation from the 1900s and 2001: spatial analysis of census and mortality statistics, *BMJ*, Vol 339, b3454.

Griffiths, C, Fitzpatrick, J, (eds), 2001, Geographic variations in health, http://www.ons.gov.uk/ons/rel/subnational-health3/geographic-variations-in-health--ds-no-16-/2001/index.html, (last accessed 20th July 2012).

Griffiths, C, Rooney, C, Brock, A, 2005, Leading causes of death in England and Wales – how should we group causes?, *Health Statistics Quarterly*, Vol 28, pp6-17.

Hacking, JM, Muller, S, Buchan, IE, 2011, Trends in mortality from 1965 to 2008 across the English north-south divide: comparative observational study, *BMJ*, Vol 342, d508.

Hajat, A, Kaufman, J, Rose, K, Siddiqi, A, Thomas, J, 2010, Do the wealthy have a health advantage? Cardiovascular disease risk factors and wealth, *Social Science & Medicine*, Vol 71, No 11, pp1935-1942.

Halpern, C, Hallfors, D, Bauer, D, Iritani, B, Waller, M, Cho, H, 2004, Implications of Racial and Gender Differences in Patterns of Adolescent Risk Behaviour for HIV and Other Sexually Transmitted Diseases, *Perspectives on Sexual and Reproductive Health*, Vol 36, No 6, pp239-247.

Hanlon, P, Lawder, R, Buchanan, D, Redpath, A, Walsh, D, Wood, R, Bain, M, Brewster, D, Chalmers, J, 2005, Why is mortality higher in Scotland than in England and Wales? Decreasing influence of socioeconomic deprivation between 1981 and 2001 supports the existence of a 'Scottish Effect', *Journal of Public Health*, Vol 27, no 2, pp199-204.

Harris, R, Sleight, P, Webber, R, 2005, *Geodemographics, GIS and Neighbourhood Targeting*, Chichester: Wiley.

Hart, C, Davey Smith, G, Gruer, L, Watt, G, 2010, The combined effect of smoking tobacco and drinking alcohol on cause-specific mortality: a 30 year cohort study, *BMC Public Health*, Vol 10, No 789.

Hedman, L, 2011, The Impact of Residential Mobility on Measurements of Neighbourhood Effects, *Housing Studies*, Vol 26, No 4, pp501-519.

Ho, DE, Imai, K, King, G, Stuart, EA, 2007, Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference, *Political Analysis*, Vol 15, No , pp199-236.

Hoel, D, Ron, E, Carter, R, Mabuchi, K, 1993, Influence of Death Certificate Errors on Cancer Mortality Trends, *Journal of the National Cancer Institute*, Vol 85, No 13, pp1063-1068.

Hox, JJ, 2002, *Multilevel Analysis. Techniques and Applications*, Mahwah, NJ: Lawrence Erlbaum Associates.

Hubert, L, Levin, J, 1976, A general statistical framework for assessing categorical clustering in free recall, *Psychological Bulletin*, Vol 83, No 6, pp1072-1080.

Iacus, SM, King, G, Porro, G, 2011a, Multivariate Matching Methods That Are Monotonic Imbalance Bounding, *Journal of the American Statistical Association*, Vol 106, No 493, pp345-361.

Iacus, SM, King, G, Porro, G, 2011b, Causal Inference without Balance Checking: Coarsened Exact Matching, *Political Analysis*, Vol 20, No 1, pp1-24.

Idler, E, Benyamini, Y, 1997, Self-rated health and mortality: A review of twenty-seven community studies, *Journal of Health and Social Behavior*, Vol 38, No 1, pp21-37.

Jain, AK, 2010, Data clustering: 50 years beyond K-means, *Pattern Recognition Letters*, Vol 31, No 8, pp651-666.

Janssen, F, Kunst, A, 2004, ICD coding changes and discontinuities in trends in cause-specific mortality in six European countries, 1950-99, *Bulletin of the World Health Organisation*, Vol 82, No 12, pp904-913.

Jen, M, Jones, K, Johnston, R, 2009, Compositional and contextual approaches to the study of health behaviour and outcomes: using multi-level modelling to evaluate Wilkinson's income inequality hypothesis, *Health and Place*, Vol 15, No 1, pp198-203.

Jensen, ME, Goodman, IA, Bourgeron, PS, Poff, NL, Brewer, CK, 2001, Effectiveness of biophysical criteria in the hierarchical classification of drainage basins, *Journal of the American Water Resources Association*, Vol 37, No 5, pp1155-1167.

Jin, W, Joyce, R, Phillips, D, Sibieta, L, 2011, Poverty and Inequality in the UK: 2011, *IFS Commentary C118*, http://www.ifs.org.uk/comms/comm118.pdf, (last accessed 29th October 2012).

Johnson, B, Al-Hamad, A, 2011, Trends in socio-economic inequalities in female mortality 2001-08. Intercensal estimates for England and Wales, *Health Statistics Quarterly*, Vol 52, pp3-32.

Johnson, RC, Schoeni, RF, Rogowski, JA, 2012, Health disparities in mid-to-late life: The role of earlier life family and neighborhood socioeconomic conditions, Social Science & Medicine, Vol 74, No 4, pp625-636.

Jylhä, M, 2009, What is self-rated health and why does it predict mortality? Towards a unified conceptual model, *Social Science & Medicine*, Vol 69, No , pp307-316.

Kahlmeier, S, Schindler, C, Grize, L, Braun-Fahrlander, C, 2001, Perceived environmental housing quality and wellbeing of movers, *Journal of Epidemiology and Community Health*, Vol 55, No , pp708-715.

Kawachi, I, Berkman, L, 2000, Social Cohesion, Social Capital, and Health, pp174-190, in Berkman, l, Kawachi, I, (eds), Social Epidemiology, New York: Oxford University Press.

Kawachi, I, Kennedy, BP, Glass, R, 1999, Social capital and self-rated health: a contextual analysis, *American Journal of Public Health*, Vol 89, No 9, pp1187-1193.

Kearns, R, 1993, Place and Health: Towards a Reformed Medical Geography, *Professional Geographer*, Vol 45, No 2, pp139-147.

Kearns, R, Moon, G, 2002, From medical to health geography: novelty, place and theory after a decade of change, *Progress in Human Geography*, Vol 26, No 5, pp605-625.

Keene, D, Bader, M, Ailshire, J, 2013, Length of residence and social integration: The contingent effects of neighborhood poverty, *Health & Place*, Vol 21, No 1, pp171-178.

Ketchen, DJ, Shook, CL, 1996, The application of cluster analysis in strategic management research: An analysis and critique, *Strategic Management Journal*, Vol 17, No 6, pp441-458.

King, G, Nielsen, R, Coberley, C, Pope, JE, Wells, A, 2011, Avoiding Randomization Failure in Program Evaluation, with Application to the Medicare health Support Program, *Population Health Management*, Vol 14, No S1, ppS11-S22.

Knorn, J, Rabe, A, Radeloff, VC, Kuemmerle, T, Kozak, J, Hostert, P, 2009, Land cover mapping of large areas using chain classification of neighboring Landsat satellite images, *Remote Sensing of Environment*, Vol 113, No 5, pp957-964.

Kondo, N, Sembajwe, G, Kawachi, I, van Dam, R, Subramanian, S, Yamagata, Z, 2009, Income inequality, mortality, and self rated health: meta-analysis of multilevel studies, *BMJ*, Vol 339:b4471.

Kreft, I, De Leeuw, J, 1998, *Introducing Multilevel Modelling*, London: SAGE.

Krieger, A, Green, P, 1999, A cautionary note on using internal-cross validation to select the number of clusters, *Psychometrika*, Vol 64, No 3, pp341-353.

Langford, A, Johnson, B, 2010, Trends in social inequalities in male mortality, 2001-08. Intercensal estimates for England and Wales, *Health Statistics Quarterly*, Vol 47, pp1-28.

Langford, IH, Bentham, G, 1996, Regional variations in mortality rates in England and Wales: an analysis using multi-level modelling, *Social Science & Medicine*, Vol 42, No 6, pp897-908.

Larson, A, Bell, M, Young, AF, 2004, Clarifying the relationships between health and residential mobility, *Social Science & Medicine*, Vol 59, No 10, pp2149-2160.

Lawson, A, 2013, *Statistical Methods in Spatial Epidemiology*, Wiley: Chichester.

Leventhal, T, Brooks-Gunn, J, 2003, Moving to Opportunity: an Experimental Study of Neighborhood Effects on Mental Health, *American Journal of Public Health*, Vol 93, No 9, pp1576-1582.

Li, Y, Pickles, A, Savage, M, 2005, Social Capital and Social Trust in Britain, *European Sociological Review*, Vol 21, No 2, pp109-123.

Lindeboom, M, van Doorslaer, E, 2004, Cut-point shift and index shift in self-reported health, *Journal of Health Economics*, Vol 23, No 6, pp1083-1099.

Lopez, R, 2004, Income inequality and self-rated health in US metropolitan areas: A multi-level analysis, *Social Science and Medicine*, Vol 59, No 12, pp2409-2419.

Lorette, A, Descombes, X, Zerubia, J, 2000, Texture Analysis through a Markovian Modelling and Fuzzy Classification: Application to Urban Area Extraction from Satelittle Images, *International Journal of Computer Vision*, Vol 36, No 3, pp221-236.

Ludwig, J, Duncan, GJ, Gennetian, LA, Katz, LF, Kessler, R, Kling, JR, Sanbomatsu, L, 2012, Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults, *Science*, Vol 337, No , pp1505-1510.

Ludwig, J, Sanbonmatsu, L, Gennetian, L, Adam, E, Duncan, GJ, Katz, LF, Kessler, RC, Kling, JR, Lindau, ST, Whitaker, RC, McDade, TW, 2011, Neighborhoods, Obesity, and Diabetes – A Randomized Social Experiment, *New England Journal of Medicine*, Vol 365, No , pp1509-1519.

Luke, D, 2005, Getting the Big Picture in Community Science Methods That Capture Context, *American Journal of Community Psychology*, Vol 35, No 3-4, pp185-200.

Lupton, R, 2003, 'Neighbourhood Effects': Can we measure them and does it matter?, CASE paper 73, London: CASE, London School of Economics.

Macintyre, S, Ellaway, A, 2000, Ecological Approaches: Rediscovering the Role of the Physical and Social Environment, pp332-348, in Berkman, l, Kawachi, I, (eds), Social Epidemiology, New York: Oxford University Press.

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*, Vol 67, No 6, pp900-914.

Majeed, A, Babb, P, Jones, J, Quinn, M, 2000, Trends in prostate incidence, mortality and survival in England and Wales 1971-1998, *British Journal of Urology International*, Vol 85, No 9, pp1058-1062

Malmström, M, Sundquist, J, Johansson, S-E, 1999, Neighbourhood Environment and Self-Reported Health Status: A Multilevel Analysis, *American Journal of Public Health*, Vol 89, No 8, pp1181-1186.

Martin, D, 2002, Geography for the 2001 Census in England and Wales, *Population Trends*, Vol 108, pp7-15.

Melchiorre, C, Matteucci, M, Azzoni, A, Zanchi, A, 2008, Artificial neural networks and cluster analysis in landslide susceptibility zonation, *Geomorphology*, Vol 94, No 3-4, pp379-400.

Milligan, G, 1980, An estimation of the effects of six types of error perturbation on fifteen clustering algorithms, *Psychometrika*, Vol 45, pp325-342.

Milligan, G, 1996, Clustering Validation: Results and Implications for Applied Analyses, pp341-375, in Arabie, P, Hubert, L, De Soete, G, (Eds), *Clustering and Classification*, World Scientific: London.

Milligan, G, Cooper, M, 1985, An examination of procedures for determining the number of clusters in a data set, *Psychometrika*, Vol 50, No 2, pp159-179.

Milligan, G, Cooper, M, 1987, Methodological Review: Clustering Methods, *Applied Psychological Measurement*, Vol 11, No 4, pp329-354.

Muennig, P, Cohen, AK, Palmer, A, Zhu, W, 2013, The relationship between five different measures of structural social capital, medical examination outcomes, and mortality, *Social Science & Medicine*, Vol 85, No 1, pp18-26.

Musterd, S, Galster, G, Andersson, R, 2012, Temporal dimensions and measurement of neighbourhood effects, *Environment and Planning A*, Vol 44, No ,pp605-627.

Nauck, D, Kruse, R, 1999, Obtaining Interpretable Fuzzy Classification Rules from Medical Data, *Artificial Intelligence in Medicine*, Vol 16, No 2, pp142-169.

Nnoaham, K, Frater, A, Roderick, P, Moon, G, Halloran, S, 2010, Do geodemographic typologies explain variations in uptake in colorectal cancer screening? An assessment using routine screening data in the south of England, *Journal of Public Health*, Vol 32, No 4, pp572-581.

Ntzoufras, I, 2009, *Bayesian Modeling Using WinBUGS*, Wiley: Hoboken, USA.

NOO (National Obesity Observatory), 2009, Healthy Weight, Health Lives: Market Segmentation and Mapping, http://www.noo.org.uk/securefiles/130130_1442//HWHL_Market_segmentation.pdf, (last accessed 30th January 2013).

ONS (Office for National Statistics), 2004, NS-SeC classes and collapses, http://www.ons.gov.uk/ons/guide-method/classifications/archived-standard-classifications/ns-sec/categories--sub-categories-and-classes/ns-sec-classes-and-collapses/index.html, (last accessed 16th April 2012).

ONS (Office for National Statistics), 2010a, Briefing Note: ONS policy on protecting confidentiality within birth and death statistics (Revised 2010), http://www.statistics.gov.uk/downloads/theme_health/ConfidentialityBirth&Death.pdf, (last accessed 18th January 2011).

ONS (Office for National Statistics), 2010b, Life expectancy at birth and at age 65 by local areas in the United Kingdom, 2007-09, http://www.ons.gov.uk/ons/rel/subnational-health4/life-expec-at-birth-age-65/2004-06-to-2008-10/statistical-bulletin.html, (last accessed 14th March 2013).

ONS (Office for National Statistics), 2011, Super Output Areas (SOAs): Frequently asked questions, http://neighbourhood.statistics.gov.uk/dissemination/Info.do?page=aboutneighbourhood/geography/superoutputareas/soafaq/soa-faq.htm, (last accessed 25th January 2011).

ONS (Office for National Statistics), 2012a, Estimation and Adjustment for Communal Establishments, http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-data/2011-first-release/first-release--quality-assurance-and-methodology-papers/coverage-within-communal-establishments.pdf, (last accessed 3rd July 2013).

ONS (Office for National Statistics), 2012b, International Migrants in England and Wales 2011, http://www.ons.gov.uk/ons/dcp171776_290335.pdf, (last accessed 17th July 2013).

ONS (Office for National Statistics), 2013a, The National Statistics Socio-economic Classification (NS-SEC rebased on the SOC2010), http://www.ons.gov.uk/ons/guide-method/classifications/current-standard-classifications/soc2010/soc2010-volume-3-ns-sec--rebased-on-soc2010--user-manual/index.html#1, (last accessed 28th April 2013).

ONS (Office for National Statistics), 2013b, Internal Migration by local Authorities in England and Wales, Year Ending June 2012, http://www.ons.gov.uk/ons/dcp171778_315652.pdf, (last accessed 14th July 2013).

Openshaw, S, 1983, Multivariate analysis of census data: the classification of areas, pp243-264, in Rhind, R, (ed) *An Census User's Handbook*, London: Methuen.

Openshaw, S, Blake, M, 1995, Geodemographic segmentation systems for screening health data, *Journal of Epidemiology and Community Health*, Vol 49, No 2, pps34-38.

Openshaw, S, Blake, M, Wymer, C, 1994, Using Neurocomputing Methods to Classify Britain's Residential Areas, Working Paper 94/17, School of Geography, University of Leeds.

Openshaw, S, Wymer, C, 1995, Classifying and regionalizing census data, pp239-270, in Openshaw, S, (ed), *Census Users' Handbook*, Cambridge: GeoInformation International.

Orford, S, Dorling, D, Mitchell, R, Shaw, M, Davey Smith, G, 2002, Life and death of the people of London: a historical GIS of Charles Booth's inquiry, *Health & Place*, Vol 8, No 1, pp25-35.

Paccagnella, O, 2006, Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects, *Evaluation Review*, Vol 30, No 1, pp66-85.

Parfitt, JP, Lovett, AA, Sünnenberg, G, 2001, A classification of local authority waste collection and recycling strategies in England and Wales, *Resources, Conservations and Recycling*, Vol 32, No 3-4, pp239-257.

Pateman, T, 2010/2011, Rural and urban areas: comparing lives using rural/urban classifications. *Regional Trends*, Vol 43, pp1-77.

Pearce, A, Elliman, D, Bedford, H, Law, C, 2008, Residential mobility and uptake of childhood immunisations: Findings from the UK Millennium Cohort Study, *Vaccine*, Vol 26, No 13, pp1675-1680.

Pickett, K, Pearl, M, 2001, Multilevel analyses of neighbourhood socioeconomic context and health outcomes: a critical review, *Journal of Epidemiology and Community Health*, Vol 55, No 2, pp111-122.

Popham, F, Boyle, PJ, O'Reilly, D, Leyland, AH, 2011, Selective internal migration. Does it explain Glasgow's worsening mortality record, *Health & Place*, Vol 17, No 6, pp1212-1217.

Public Health England (PHE), 2013, Longer Lives, http://longerlives.phe.org.uk/, (last accessed 20th June 2013).

Putnam, RD, 1995, Bowling Alone: America's Declining Social Capital, *Journal of Democracy*, Vol 6, No 1, pp65-78.

Putnam, RD, Leonardi, R, Nanetti, R, 1993, *Making democracy work: civic traditions in modern Italy*, Princeton University Press: Chichester.

Quillian, L, 2003, How long are exposure to poor neighbourhoods? The long-term dynamics of entry and exit from poor neighbourhoods, *Population Research and Policy Review*, Vol 22, No , pp221-249.

Raines, JE, Wight, J, 2002, The mortality experience of people admitted to nursing homes, *Journal of Public Health Medicine*, Vol 24, No 3, pp184-189.

Raleigh, VS, Frosini, F, Sizmur, S, Graham, C, 2012, Do some trusts deliver a consistently better experience for patients? An analysis of patient experience across acute care surveys in English NHS trusts, *BMJ Quality and Safety*, Vol 21, No 5, pp381-390.

Raleigh, VS, Kiri, V, 1997, Life expectancy in England: variations and trends by gender, health authority, and level of deprivation, *Journal of Epidemiology and Community Health*, Vol 51, pp649-658.

Randall, C, 2011, Housing, pp1-30, in (ed.) Beaumont, J, *Social Trends 41*, Newport: ONS.

Richardson, E, Pearce, J, Mitchell, R, Day, P, Kingham, S, 2010, The association between green space and cause-specific mortality in urban New Zealand: an ecological analysis of green space utility, *BMC Public Health*, Vol 10, No 240

Riva, M, Curtis, S, Gauvin, L, Fagg, J, 2009, Unravelling the extent of inequalities in health across urban and rural areas: Evidence from a national sample in England, *Social Science & Medicine*, Vol 68, No 4, pp654-663.

Riva, M, Curtis, S, Norman, P, 2011, Residential mobility within England and urban-rural inequalities in mortality, *Social Science & Medicine*, Vol 73, No 12, 1698-1706.

Riva, M, Gauvin, L, Barnett, T, 2007, Toward the next generation of research into small area effects in health a synthesis of multilevel investigations published since July 1998, *Journal of Epidemiology and Community Health*, Vol 61, No 10, pp853-861.

Rogerson, PA, 2006, *Statistical Methods for Geography*, London: SAGE.

Room, R, 2004, Smoking and drinking as complementary behaviours, *Biomedicine & Pharamcotherapy*, Vol 58, No , pp111-115.

Rooney, C, Griffiths, C, Cook, L, 2002, The implementation of ICD-10 for cause of death coding – some preliminary results from the bridge coding study, *Health Statistics Quarterly*, Vol 13, pp31-41.

Rooney, C, Smith, S, 2000, Implementation of ICD-10 for mortality data in England and Wales from January 2001, *Health Statistics* Quarterly, Vol 8, pp41-50.

Roos, LL, Magoon, J, Chateau, D, 2010, Does It Matter What You Measure? Neighbourhood effects in a Canadian Setting, *Healthcare Policy*, Vol 6, No 1, pp47-63.

Roos, LL, Magoon, J, Gupta, S, Chateau, D, Veugelers, PJ, 2004, Socioeconomic determinants of mortality in two Canadian provinces: multilevel modelling and neighbourhood context, *Social Science & Medicine*, Vol 59, No 7, pp1435-1447.

Rozumalski, A, Schwartz, M, 2009, Crouch gait patterns defined using k-means cluster analysis are related to underlying clinical pathology, *Gait & Posture*, Vol , No , pp155-160.

Sanbonmatsu, L, Marvokov, J, Porter, N, Yang, F, Adam, E, Congdon, WJ, Duncan, GJ, Gennetian, LA, Katz, LF, Kling, JR, Kessler, RC, Lindau, ST, Ludwig, J, McDade, TW, 2012, The Long-Term Effects of Moving to Opportunity on Adult Health and Economic Self-Sufficiency, *Cityscape*, Vol 14, No 2, pp109-136.

Selinger, C, Ellis, R, Harrington, M, A good death certificate: improved performance by simple educational measures, *Postgraduate Medical Journal*, Vol 83, No 978, pp285-286.

Shaw, M., Thomas, B., Dorling, D., Davey Smith, G., 2008, *The grim reaper's road atlas: an atlas of mortality in Britain*, Bristol: Policy Press.

Shelton, N, Birkin, M, Dorling, D, 2006, Where not to live: a geo-demographic classification of mortality for England and Wales, 1981-2000, *Health & Place*, Vol 12, No 4, pp557-569.

Singleton, A, Longley, P, 2009, Geodemographics, visualisation, and social networks in applied geography, *Applied Geography*, Vol 29, No 3, pp289-298.

Sington, J, Cottrell, B, 2002, Analysis of the sensitively of death certificates in 440 hospital deaths: a comparison with necropsy findings, *Journal of Clinical Pathology*, Vol 55, No 7, pp499-502.

Smith, MP, Olatunde, O, White, C, 2010, Inequalities in disability-free life expectancy by area deprivation: England, 2001-04 and 2005-08, *Health Statistics Quarterly*, Vol 48, pp1-22.

Sneath, PHA, Sokal, RR, 1973, *Numerical Taxonomy*, San Francisco: W.H. Freeman.

Snelgrove, JW, Pikhart, H, Stafford, M, 2009, A multilevel analysis of social capital and self-rated health: Evidence from the British Household Panel Survey, *Social Science & Medicine*, Vol 68, No 11, pp1993-2001.

Snijders, TAB, Berkhof, J, 2007, Diagnostic Checks for Multilevel Models, pp139-173, in De Leeuw, J, Meijer, E, (eds), *Handbook of Multilevel Analysis*, New York: Springer.

Snijders, TAB, Bosker, RJ, 1999, *Multilevel Analysis*, London: SAGE.

Strong, M, Pearson, T, MacNab, YC, Maheswaran, R, 2012, Mapping gender variation in the spatial pattern of alcohol-related mortality: A Bayesian analysis using data from South Yorkshire, *Spatial and Spatio-temporal Epidemiology*, Vol 3, No 2, pp141-149.

Subramanian, SV, Elwert, F, Christakis, N, 2008, Widowhood and mortality among the elderly: The modifying role of neighbourhood concentration of widowed individuals, Social Science & Medicine, Vol 66, No 4, pp873-884.

Swift, B, West, K, Death certification: an audit of practice entering the 21st century, *Journal of Clinical Pathology*, Vol 55, No 4, pp275-279.

Szreter, S, Mooney, G, 1998, Urbanization, Mortality, and the Standard of Living Debate: New Estimates of the Expectation of Life at Birth in Nineteenth-Century British Cities, *The Economic History Review*, Vol 51, No 1, pp84-112.

Thomas, B, Dorling, D, Davey Smith, G, 2010, Inequalities in premature mortality in Britain: observational study from 1921 to 2007, *BMJ*, Vol 341, c3639.

Thomas, B, Pritchard, J, Ballas, D, Vickers, D, Dorling, D, 2009, A Tale of Two Cities: The Sheffield Project, http://www.sasi.group.shef.ac.uk/research/sheffield/a_tale_of_2_cities_sheffield_project_final_report.pdf, (last accessed 27th March 2012).

Thurman, I, 2013, *Personal correspondence*.

Tobler, W, 1970, A computer movie simulating urban growth in the Detroit region, *Economic Geography*, Vol 46, No 2, pp234-240.

Townsend, I, Kennedy, S, 2004, Poverty: Measures and Targets, *House of Commons Research Paper 04/23*.

Tudor Hart, J, 1971 [2000], The inverse care law, pp312-39, in Davey Smith, G, Dorling, D, Shaw, M, (eds.), *Poverty, inequality and health in Britain 1800-2000: A reader*, Bristol: Policy Press.

Tuffin, R, Quinn, A, Cramp, P, 2009, A review of the accuracy of death certification on the intensive care unit and the proposed reforms to the coroner's system, *Journal of the Intensive Care Society*, Vol 10, No 2, pp134-137.

Turnstall, H, Pickett, K, Johnson, S, 2010, Residential mobility in the UK during pregnancy and infancy: Are pregnant women, new mothers and infants 'unhealthy migrants'?, *Social Science & Medicine*, Vol 71, No 4, pp786-798.

Van Ham, M, Manley, D, 2012, Neighbourhood Effects Research at a Crossroads: Ten Challenges for Future Research, *IZA Discussion Paper Series*, No 6793.

Vanwambeke, SO, van Benthem, BHB, Khantikul, N, Burghoorn-Maas, C, Panart, K, Oskam, L, Lambin, EF, Somboon, P, 2006, Multi-level analyses of spatial and

temporal determinants for dengue infection, *International Journal of Health Geographics*, Vol 5, No 5.

Veugelers, PJ, Yip, AM, Kephart, G, 2001, Proximate and contextual socioeconomic determinants of mortality: multilevel approaches in a setting with universal health care coverage, *American Journal of Epidemiology*, Vol 154, No , pp725-732.

Vickers, DW, 2006, Multi-Level Integrated Classifications Based on the 2001 Census, PhD thesis, University of Leeds.

Vickers, DW, 2011, England's Changing Social Geology, pp37-52, in, Stillwell, J, Norman, P, Thomas, C, Surridge, P, (eds.) *Spatial and Social Disparities, Understanding Population Trends and Processes, Vol 2*, London: Springer.

Voas, D, Williamson, P, 2001, The diversity of diversity: a critique of geodemographic classification, *Area*, Vol 33, No 1, pp63-76.

Waller, NG, Kaiser, HA, Illian, JB, Manry, M, 1998, Cluster analysis with Kohonen neural networks, *Psychometrika*, Vol 63, No 1, pp5-22.

Wannamethee, SG, Shaper, AG, Whincup, PH, Walker, M, 2002, Migration within Great Britain and cardiovascular disease: early life and adult environmental factors, *International Journal of Epidemiology*, Vol 31, No 5, pp1054-1060.

Ward, JH, 1963, Hierarchical groupings to optimize an objective function, *Journal of the American Statistical Association*, Vol 58, No 301, pp236-244.

Warden, J, 1998, Britain's new health policy recognises poverty as major cause of illness, *British Medical Journal*, Vol 316, No , p493.

White, C, Edgar, G, 2010, Inequalities in healthy life expectancy by social class and area type: England, 2001-03, *Health Statistics Quarterly*, Vol 45, pp28-56.

WHO (World Health Organisation), 2004, International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Volume 2 (2nd Ed), http://www.who.int/classifications/icd/ICD-10_2nd_ed_volume2.pdf, (last accessed 31/1/2011).

Whynes, DK, 2009, Deprivation and self-reported health: are there 'Scottish effects' in England and Wales, *Journal of Public Health*, Vol 31, No 1, pp147-153.

Wilkinson, RG, Pickett, KE, 2006, Income inequality and population health: a review and explanation of the evidence, *Social Science & Medicine*, Vol 62, No 7, pp1768-1784.

Wilkinson, RG, Pickett, KE, 2009, *The spirit level: why more equal countries almost always do better*, London: Allen Lane.

Williams, E, Dinsdale, H, Eayres, D, Tahzib, F, 2004, Impact of nursing home deaths on life expectancy calculations in small areas, *Journal of Epidemiology and Community Health*, Vol 58, No 1,, pp958-962.

Williams, ES, Scott, CM, Scott, SM, 1995, Using mortality data to describe geographic variations in health status at sub-district level, *Public Health*, Vol 109, No 1, pp67-73.

Wilson, WJ, 1987, *The truly disadvantaged: the inner city, the underclass, and public policy*, Chicago: University of Chicago Press.

Woods, L, Rachet, B, Riga, M, Stone, N, Shah, A, Coleman, M, 2005, Geographical variation in life expectancy at birth in England and Wales is largely explained by deprivation, *Journal of Epidemiology and Community Health*, Vol 59, No 2, pp115-120.

Woods, R, 2000, *The Demography of Victorian England and Wales*, Cambridge: Cambridge University Press.

Yen, IH, Michael, YL, Perdue, L, 2009, Neighbourhood Environment in Studies of Health of Older Adults: A Systematic Review, *American Journal of Preventive Medicine*, Vol 37, No 5, pp455-463.

Young, H, Grundy, E, O'Reilly, D, Boyle, P, 2010, Self-rated health and mortality in the UK: results from the first comparative analysis of the England and Wales, Scotland, and Northern Ireland Longitudinal Studies, *Population Trends*, Vol 139, pp11-36.