A Spatial Microsimulation Analysis of Health Inequalities and Health Resilience

Phil Mike Jones

September 2017

Abstract

Health inequalities persist despite decades of effort to reduce them. Faced with a reduction in public spending, contraction of the welfare state, and rising inequality it is likely that health inequalities will increase for years to come. A better understanding of health resilience, which areas and individuals are resilient, and what factors might 'protect' their health outcomes might help develop policies to break down the link between disadvantage and health.

This research contributes to the understanding of health resilience in the case study area of Doncaster, South Yorkshire. As a former mining town, Doncaster is exposed to significant economic disadvantage reflected in many settlements across the North East, North West, Midlands, and South Wales. Previous geographical research into health resilience has been limited either to small–area information with basic health outcomes, or more sophisticated measures of health outcomes but geographically aggregated to large regions. Using spatial microsimulation, I present the first estimate of health resilience at the small–area level using measures of health previously inaccessible to researchers.

This is complemented by a systematic scoping literature review of measures hypothesised to affect health resilience. I simulate a broad range of these alongside clinical depression and income to explore a more comprehensive range of factors than have previously been possible. This includes small–area and individual–level factors, which are difficult to separate.

I conclude by comparing geographical proximity of a number of health

amenities to resilient and non–resilient areas in Doncaster, and by evaluating local and national policies such as Universal Credit and their likely effect on the residents of Doncaster and their resilience.

For Sam

Acknowledgements

To Hannah and Sam, I can't thank you enough for your patience, love, support, and encouragement, without which this thesis would never have been finished. Please know that I am forever grateful for everything you've done.

To Dimitris and Liddy, thank you for your help and encouragement with my thesis. I consider myself very lucky to have had two excellent and supportive supervisors. I have learned so much that wouldn't have been possible without your guidance and encouragement.

To Mum and Dad, and Richard and Rosemary, thank you for making our move to Sheffield (and therefore this PhD) possible, and for all your support, encouragement, and help keeping us going over the last few years.

To Jo, thanks for encouraging me to start down this path. It's been hard but your 'PhD survival kit' has helped me every step of the way. The last pencil ran out just the other day.

To Tom, Amber, Martin, Becky, Azeezat, and everyone in D5, thanks for your sympathetic ear when I needed to moan, pizza when I needed to eat, and coffee when I needed to function.

Thanks to Laurie Mott of Doncaster Metropolitan Borough Council for providing advice, data, and context on Doncaster, and for the field research that I used to build the case studies. Thank you to Emma Nicholas, also of DMBC and Well Doncaster, for showing me the community development projects in Denaby Main, one of my case study areas.

Thank you to Les Monaghan who kindly allowed me to reproduce the photographs in Chapter 1 from his 'Relative Poverty' project.

I would like to acknowledge the funding provided from Doncaster Metropolitan Borough Council, without which this project would not have been possible.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> [There is a] remarkable robustness of inequalities in mortality across time and space (Mackenbach et al., 2015: 60).

Health inequalities persist despite decades of effort to reduce them. Faced with a reduction in public spending, contraction of the welfare state, and rising inequality it is likely that health inequalties will increase for years to come and the problems associated with these will be exacerbated.

Individuals, communities, and local public agencies are increasingly looking for ways to mitigate the worst effects of widening health inequalities in the absence of policy to reduce them. Health resilience is one such solution thought to mitigate the effects of health inequalities.

Resilience has a history spanning several decades, emerging in psychological literature in the 1970s. In this and later literature resilience is thought of as the "…factors and processes that promote escape from disadvantage" (Schoon, 2006: 1). In the case of health resilience, health inequalities confer real disadvantage in terms of morbidity and mortality for individuals across the social gradient in health.

If health inequalities are the disadvantage, what is the 'escape'? In the health domain reduced morbidity and premature mortality are the 'goal' of resilience, and a number of measures of morbidity are used.

Resilience, then, is having better–than–expected health outcomes given the exposure to disadvantage. For this reason many authors describe the act of overcoming adversity as 'beating' (Bartley, 2006), 'defying' (Cairns, 2013), or 'overcoming' (Werner and Smith, 1992) the odds.

A better understanding of health resilience, which areas and individuals are resilient, and what factors might 'protect' their health outcomes might help develop policies to break down the link between disadvantage and health. I offer evidence to understand health resilience in the case study area of Doncaster, a large town in the Yorkshire and The Humber region in the north of England. Doncaster, as a former mining area, is an ideal site for case study because it is exposed to significant disadvantage associated with areas of former heavy industry, mining, and manufacturing (Doran et al., 2006: 686). Section 1.4 provides context of the Doncaster area and its socio–economic and demographic characteristics.

I achieve this with a systematic scoping literature review and spatial microsimulation of resilient characteristics at the small–area level to answer the research questions below. There are numerous ways 'resilience' is operationalised and equally numerous factors thought to contribute to resilience in the literature. The scoping review allowed me to systematically capture the breadth of measures and factors related to resilience.

Similarly there are numerous data sets that are available to social science researchers in the UK for the study of many aspects of social life, including health inequalities, but there is still a lack of readily–available data at a small–area level (Ballas et al., 2005b: 19). Spatial microsimulation produces *simulated* spatial micro–data, addressing this shortcoming to a large extent. With increasing computational power readily available it is becoming more straightforward to use simulation techniques to fill this void. This thesis contributes to the health resilience literature by using spatial microsimulation to estimate individual–level data at the small–area level, which has previously not been applied in this domain.

1.1 Research questions

With this research I aim to explore health resilience and the factors that influence it in more detail than previous literature has achieved. In particular, the research questions I answer are:

1. How are 'disadvantage', 'good health outcomes', and therefore resilience, operationalised in existing literature?
2. What characteristics are thought to affect resilience in this literature?
3. How does the local area affect health resilience and contribute to positive health outcomes?
4. Where and how will resilience be affected by local and national policy intitiatives over the next four years?

I identify how disadvantage, good health outcomes, and resilience are operationalised in the existing literature, and which characteristics are thought to affect resilience, using a systematic scoping review, described in Chapter 3 (items 1 and 2). I simulate a comprehensive range of these characteristics and discuss their geographical distribution in Doncaster in Chapter 6 (item 2). I also simulate clinical depression and indicators of poverty to examine health resilience in Doncaster which I outline in Chapters 6 and 7 (item 3). I review how resilience could be affected by local and national policy in Section 7.4 of Chapter 7 (item 4). I therefore simulate three sets of information: clinical depression and deprivation (health resilience); resilient characteristics; and a range of indicators of poverty, such as benefit recipients.

1.2 Originality and contribution to knowledge

In many countries, including the UK, few data sets are available with small–area geographies (Birkin and Clarke, 1988). The spatial microsimulation technique I use in this research permits the use of health measures that have not previously been possible to explore at the small–area level, and this is a key strength of this study. It allows the construction "a much richer dataset that would allow us to investigate an individual's health at a very fine level of spatial resolution…" (Morrissey et al., 2013: 221). Existing literature has used health outcomes available in the census—which are limited to self–reported general health and limiting long–term illness or disability—to examine resilience geographically. Tunstall et al. (2007) and Cairns et al. (2012), for example, are limited to English parliamentary constituencies. This research is able to address many of the explicit limitations of Cairns et al. (2012) and similar studies by examining health resilience at a much smaller geographical level than has previously been possible (2012: 932).

A significant benefit of an individual–level simulation at a small–area level is the ability to begin to explore area–level effects of health resilience while taking into account individual characteristics. Without individual–level small–area data it is difficult to separate area–level and individual–level effects on health because individuals with similar characteristics tend to cluster together. For example, most data available at the small–area level is aggregated, such as the census. Because of this it is not possible to determine if differences in health outcomes between areas are because of differences in the area, or differences in the individuals who live in that area. The spatial microsimulation technique can take account of the individual level when exploring differences in health outcomes between areas.

This study is the first to use data from *Understanding Society* to explore health resilience. By using spatial microsimulation I add a geographical dimension to this—largely—aspatial data. *Understanding Society* is a rich data source with comprehensive socio–demographic and health variables available. Previous studies have attempted to articulate concepts thought important to resilience, but have been limited by the quantitative measures available. With *Understanding Society* I am able to offer a more comprehensive analysis of resilient characteristics. *Understanding Society* has a large sample of respondents making it a robust choice to reflect the diversity of an area.

1.3 Thesis structure

This thesis is structured as follows. This chapter concludes with a profile of Doncaster to provide descriptive context about the area and its residents.

Chapter 2 summarises the historical and contemporary resilience literature. This provides the theoretical basis of the analysis and informs the operationalisation of resilience that I use throughout. This chapter identifies a lack of consensus in resilience and resilient characteristics, which I address in Chapter 3, which is a systematic scoping literature review of health resilience measures adopted in contemporary literature. I include a comprehensive range of resilient characteristics and measures in my analyses which are informed by the results of this chapter.

Chapter 4 outlines the spatial microsimulation technique. I summarise the history of spatial microsimulation in the social sciences, demonstrating its rich precedent as a robust method. I describe a number of spatial microsimulation studies in the health domain both domestically and internationally, illustrating the technique's utility in examining health outcomes. There are a number of different ways to create a spatially microsimulated data set, so I also summarise these here and state how I will produce my own model and my reasons for these methodological choices.

In Chapter 5 I begin the spatial microsimulation process. I identify suitable constraints, which can be thought of as analogous to independent variables in regression, and prepare these from census data, matching them with data available in the *Understanding Society* survey. I test these constraints using logistic regression techniques, and create a 'pilot' simulation.

Chapter 6 is the simulation of resilience itself. I expand on the methods in Chapter 5 to include additional constraints to improve the model fit. I simulate health outcomes, socio–economic risk, and resilience characteristics identified in Chapter 3.

My analysis of this data is presented in Chapter 7. With individual–level data at the small–area I am able to explore how the proximity of amenities is associated with resilience. Using the simulated data I describe a rich picture of the charactistics of residents living in four resilient case study areas, and explore narratively the characteristics that might make them resilient compared to their socio–economic peers. I finish this chapter with an analysis of local and national policy and their effects on health outcomes and resilience at the small–area level.

A number of supplementary materials supported the preparation of this thesis, and these can be downloaded in their entirety from `http://etheses.whiterose.ac.uk/id/eprint/19283`. I sometimes refer to some of these files in footnotes, and in these cases the path I specify to the file is found in this zip file.

1.4 Doncaster profile

Doncaster is a unitary authority in the historical county of South Yorkshire, in the Yorkshire and The Humber region of England. Figure 1.1 shows the location of Doncaster local authority (dark grey) within Great Britain with the historical county of South Yorkshire and the Yorkshire and The Humber region shown for context. Figure 1.2 is a close–up of the

Yorkshire and The Humber region, showing Doncaster within the historic county of South Yorkshire and the Yorkshire and The Humber region itself.

1.4.1 Population

The population of Doncaster was $302,402$ on census day 2011 (Nomis, 2013a; Office for National Statistics et al., 2017). The majority of the population lives within the urban area of Doncaster town itself, which includes the suburbs of Balby, Bessacarr, Armthrope, and Bentley. Neighbouring towns of Mexborough, Conisborough, and Thorne also have significant populations. Much of the rest of the borough is made up of less populated rural areas. Figure 1.3 shows the major population centres of Doncaster.

The areas within Doncaster depicted in Figure 1.3 are 'community areas'. These were designed by Doncaster Metropolitan Borough Council to represent a relatively homogoneous area or neighbourhood, and are constructed from statistical output areas. This makes it relatively easy to perform analyses at community level area by obtaining data at output area level and combining to form community areas. Figure 1.4 shows Doncaster community area boundaries with constituent 2011 output area classifications (Office for National Statistics, 2015a). Most community areas are constructed from similar output areas suggesting they are relatively homogoneous units.

Figure 1.5 is a population cartogram of Doncaster. It shows the same population information as Figure 1.3, but is reprojected based on the size of the population in each area (Gastner and Newman, 2004). Areas with larger populations are depicted as larger, and conversely areas that have a smaller population are depicted as smaller. Cartograms generally are useful for depicting a topic of interest—in this case population—on a geographical area which can be resized to reflect the human interest rather than the area of the land (Ballas and Dorling, 2011; Barford and Dorling, 2006).

Figure 1.1: Doncaster local authority shown within the Yorkshire and The Humber region and Great Britain

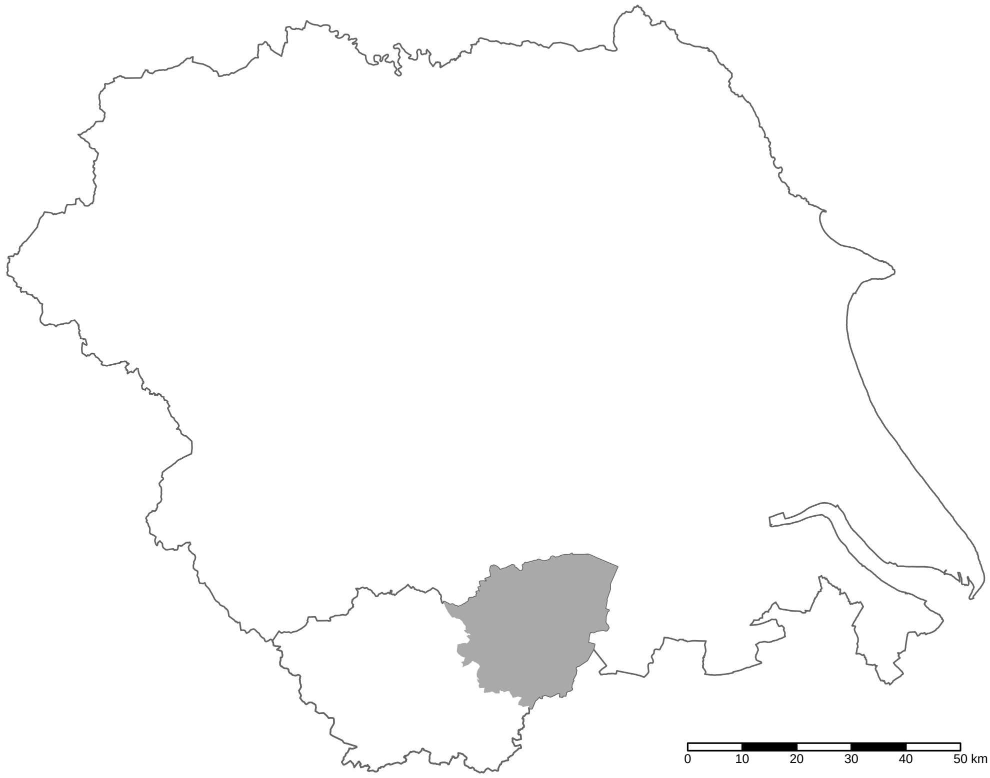

Figure 1.2: Doncaster local authority shown within the historic county of South Yorkshire and the Yorkshire and The Humber region

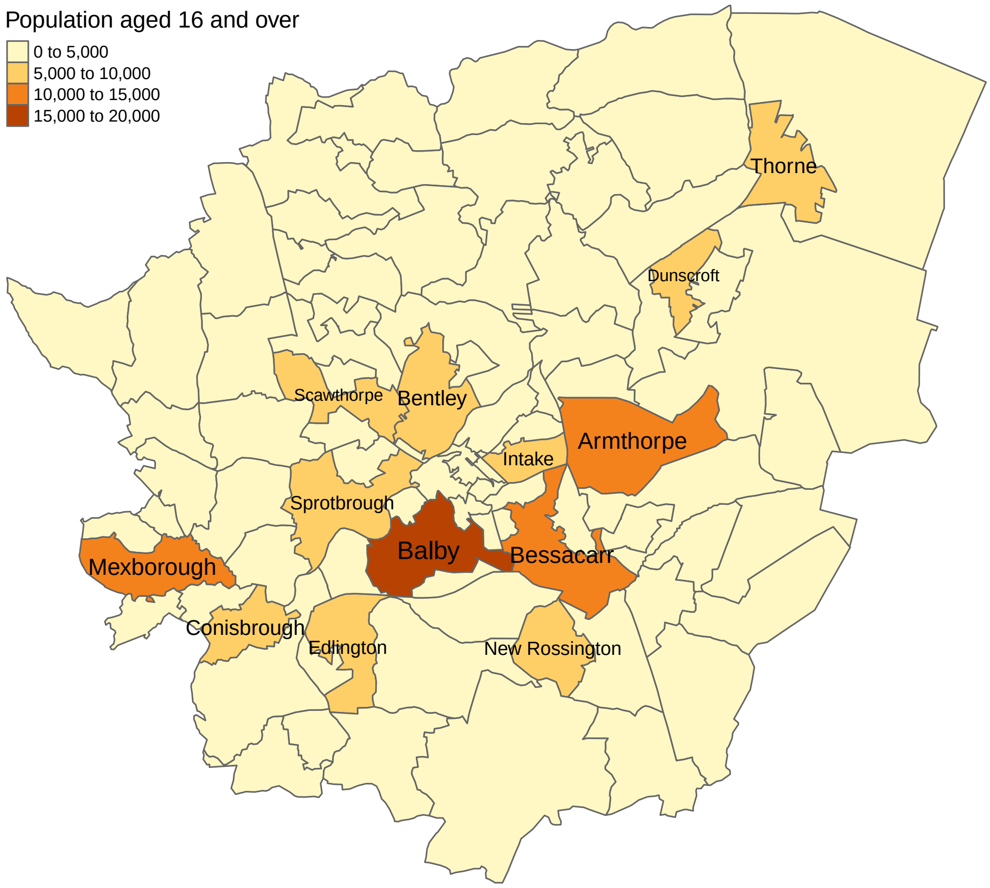

Figure 1.3: Doncaster population aged 16 and over

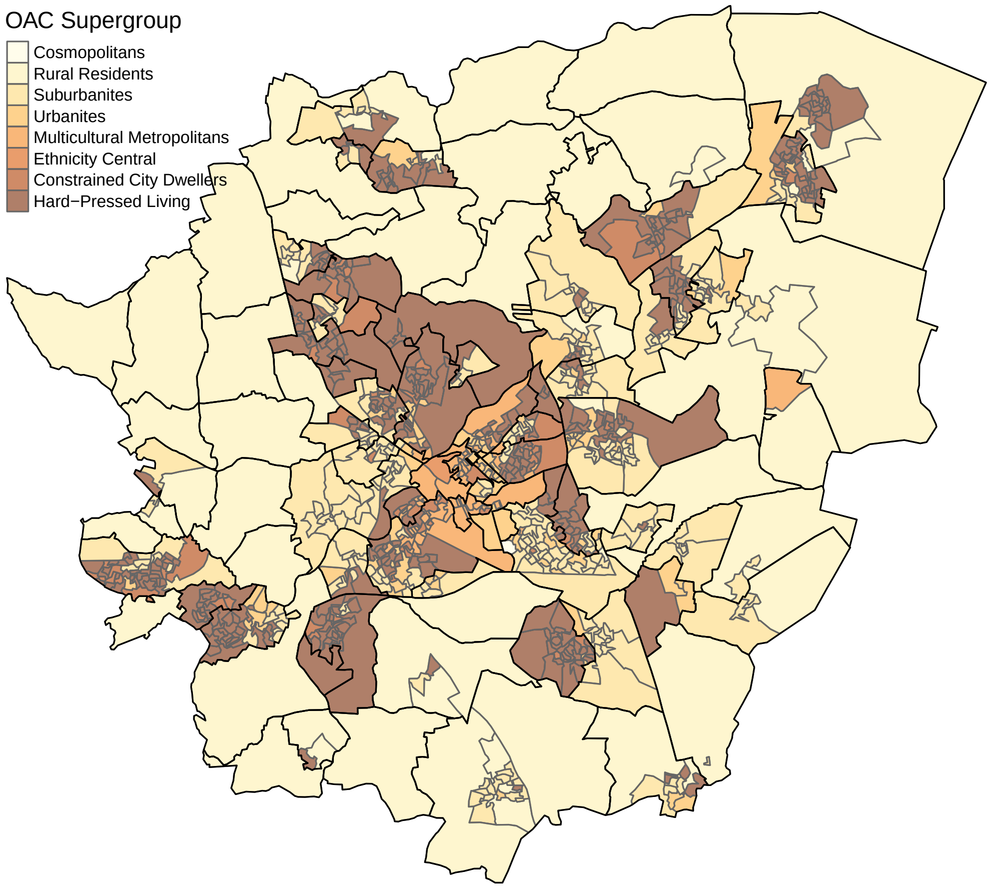

Figure 1.4: Doncaster output area classifications (supergroups) overlaid with community area boundaries

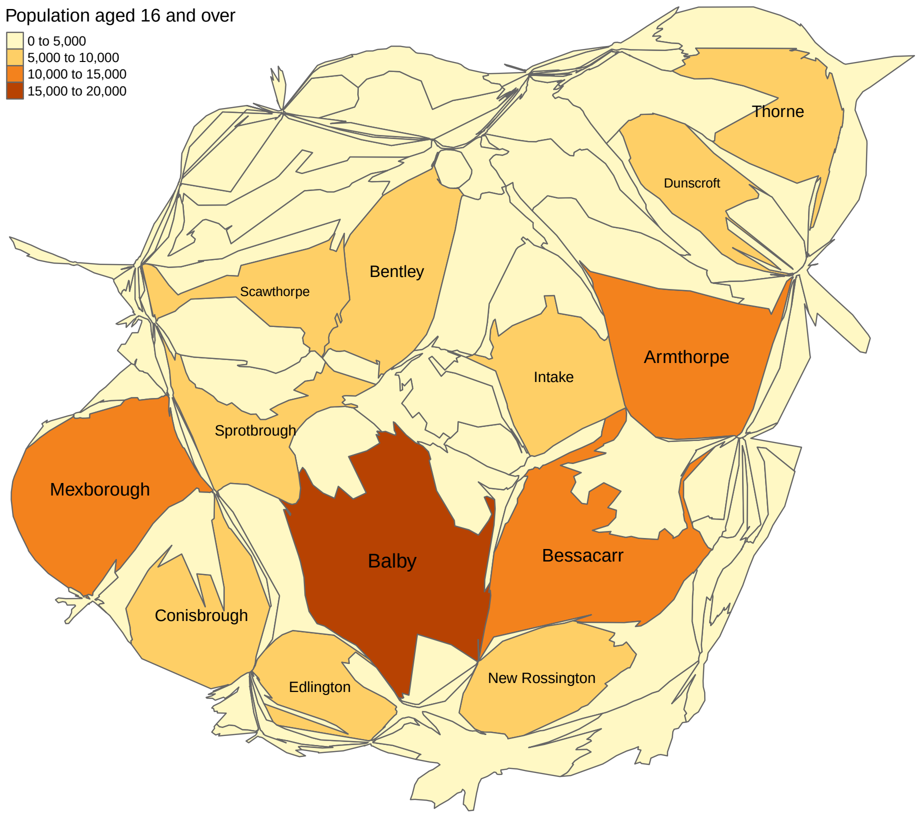

Figure 1.5: Doncaster population aged 16 cartogram

1.4.2 Deprivation

Doncaster is one of the most deprived local authority districts in England overall, based on the English indices of multiple deprivation (IMD) 2015 (Department for Communities and Local Government, 2015). The IMD is a composite measure considering income, employment, education, health, crime, access to housing and services, and living environment. The index combines these individual deprivation domains to produce one overall measure (Smith et al., 2015: 12). Figure 1.6 shows the IMD rank for each local authority in England. Lower ranked areas are more deprived; Doncaster is therefore one of the 50 most deprived local authorities in England.

Breaking down the overall index, Doncaster is within the 50 most deprived local authorities for the income, employment, education, health, and crime domains. Doncaster is less deprived in the living environment domain and one of the least deprived areas in the country based on access to housing and services (Department for Communities and Local Government, 2015).

Doncaster overall is clearly relatively deprived for multiple domains, but there is significant variation within Doncaster (Department for Communities and Local Government, 2015). Many of the urban areas in and immediately around Doncaster town centre, Mexborough, Connisborough, and Thorne are among the most deprived small areas—lower layer super output areas (LSOAs)—in England. Some smaller urban areas and some rural areas are also significantly deprived, such as Askern, Stainforth, and rural areas north of the town centre. There are some large areas that are among the least deprived in the country, for example Finningley and Blaxton to the east of the district, Tickhill to the south, and hamlets around Sprotbrough to the west of the town centre. Figure 1.7 shows the index of multiple deprivation for small areas in Doncaster.

Output area classifications (OACs) show a similar message to the indices of deprivation, but at an even smaller geographical level. Output area classifications group together output areas that share similar population

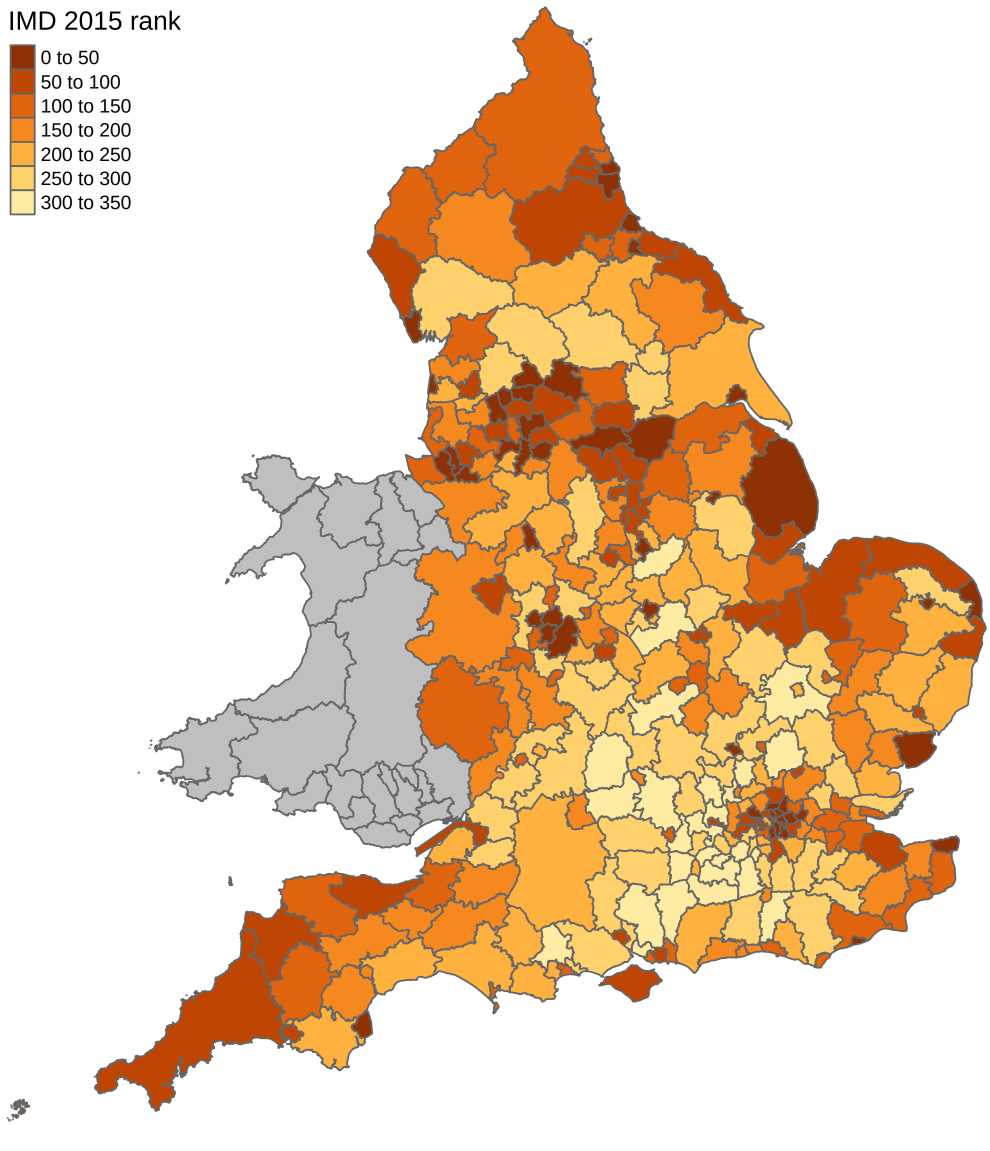

Figure 1.6: IMD 2015 average rank (lower rank is more deprived)

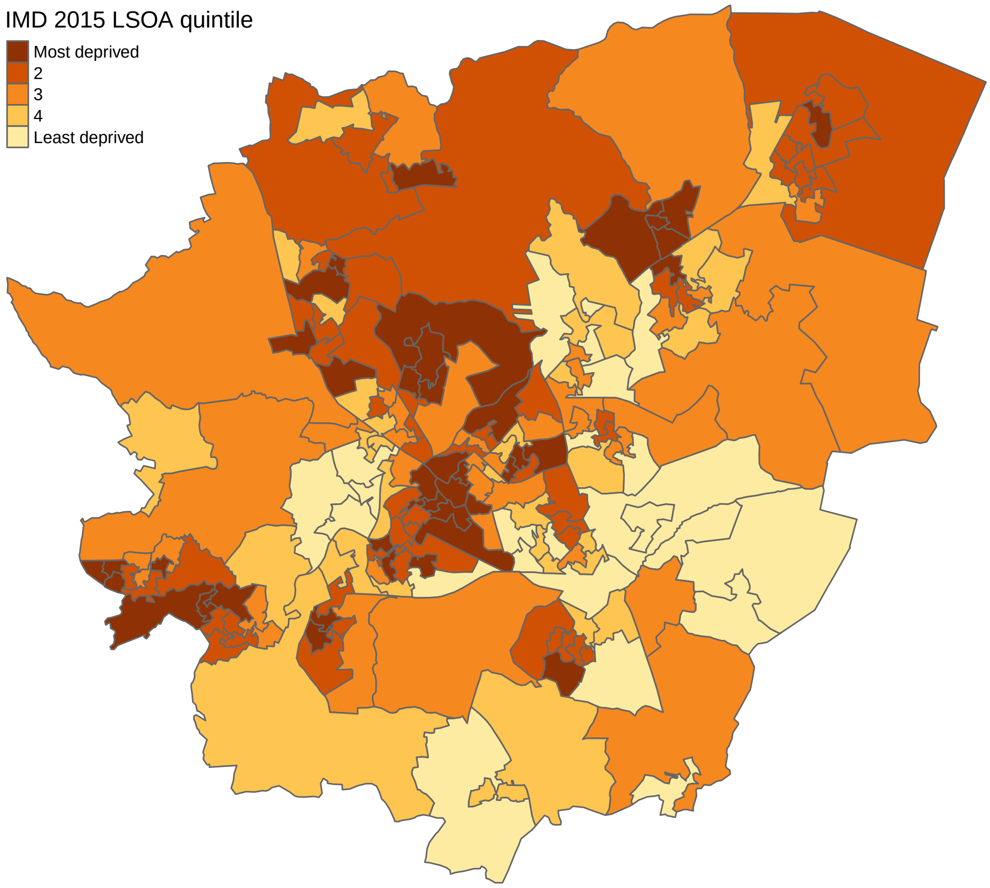

Figure 1.7: IMD 2015 quintile by LSOA in Doncaster

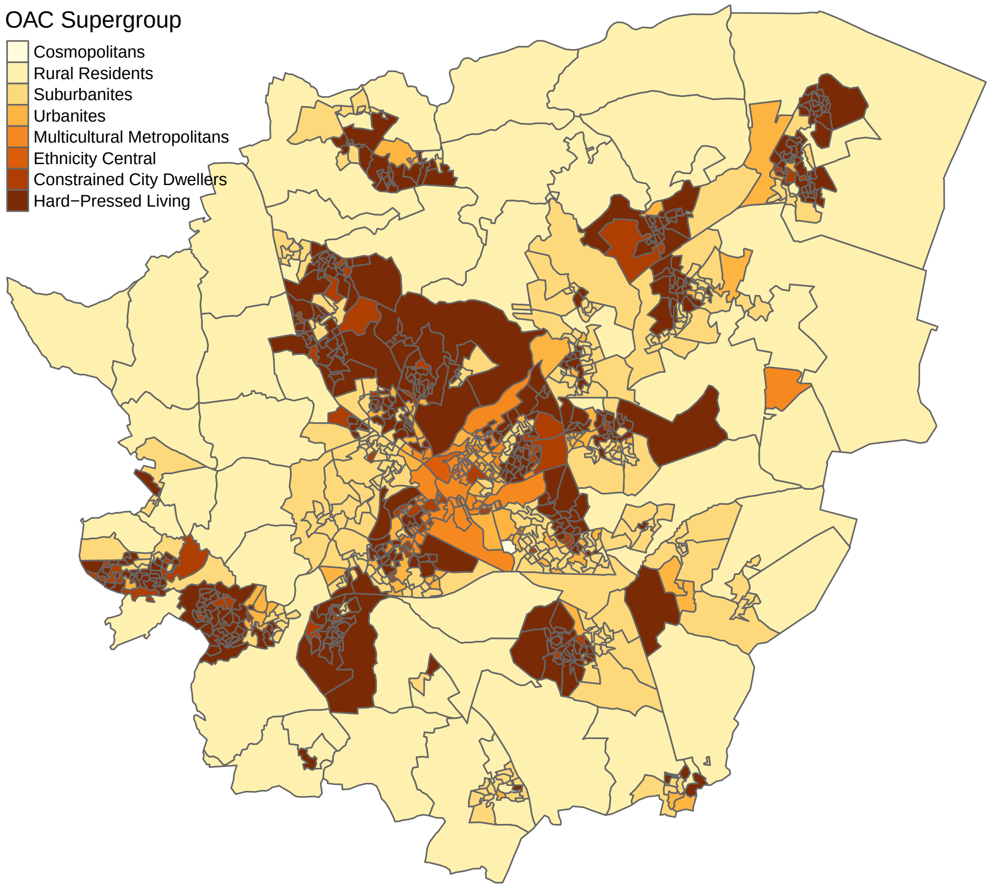

Figure 1.8: Doncaster output areas with output area classification supergroup

characteristics (Office for National Statistics, 2015a). The output areas classifications are not a measure or index of deprivation, but nevertheless do illustrate some areas that are primarily characterised by ‘hard–pressed’ and ‘constrained’ living environments. Figure 1.8 show the supergroup output area classifications for Doncaster.

Areas of hard–pressed living and constrained dwelling can be seen in the familiar places; suburbs around Doncaster town centre, Mexborough and Conisborough to the West, and Thorne to the north east. ‘Rural Residents’ and ‘suburbanites’ of the borough surround the central urban area of Doncaster itself.

Photographs taken by Doncaster photographer Les Monaghan for the

Figure 1.9: ‘Dave’s’ cupboard

Relative Poverty project reflect some of the extremes of poverty in Doncaster. Figures 1.9 and 1.10 shows the conditions and hardships faced by two of the participants of the project. These photographs are reproduced here with the kind permission of the author, and additional material can be found at `http://www.relativepoverty.org`.

Doncaster has a relatively high proportion of people economically active unemployed. Economically active unemployed individuals are those who are not in employment or self–employed, but are: “seeking work and ready to start work within two weeks”; or are “waiting to start a job already obtained and available” (Office for National Statistics, 2014a: 15). Figure 1.11 shows the proportion of individuals economically active unemployed in each local authority district in England and Wales. At census day 2011 there were $12,697$ economically active unemployed people in Doncaster, or approximately 4.2% of the population aged over 16, higher than the other districts in South Yorkshire and one of the highest in the Yorkshire and The Humber region.

Figure 1.10: ‘Niamh’s’ bathroom

Figure 1.12 shows the proportion of individuals economically active unemployed in each community area in Doncaster. While the mean proportion of unemployed for Doncaster was 0.04 overall, it can be seen that the proportion of unemployed people in some areas is up to 0.13.

Many of the areas that were identified as deprived—using the index of multiple deprivation or output area classification—have a high proportion of unemployment. Mexborough and Denaby Main to the west, suburbs around the town centre such as Bentley or Balby, and Moorends to the north east have an unemployment rate greater than 6% (Nomis, 2013a; Office for National Statistics et al., 2017). Community areas that are labelled in Figure 1.12 have an unemployment rate of greater than 6% and a working–age population greater than $5,000$.

Figure 1.13 shows the number of people who are long–term unemployed or who have never worked. Community areas with greater than 500 long–term unemployed or never worked are labelled. Mexborough, Conisbrough, Balby, Bentley, and Thorne have relatively high number of long–term

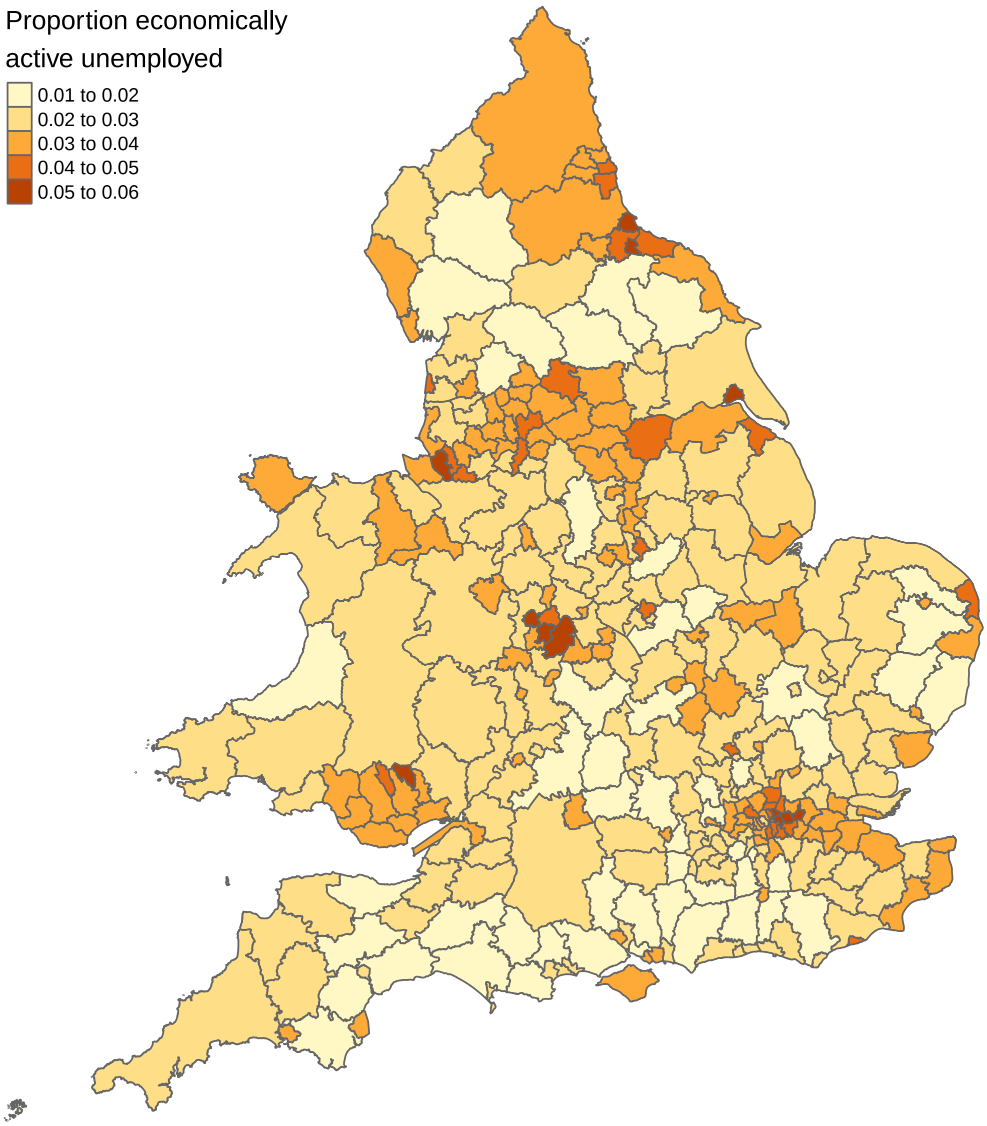

Figure 1.11: Proportion of economically active unemployed (source: 2011 Census)

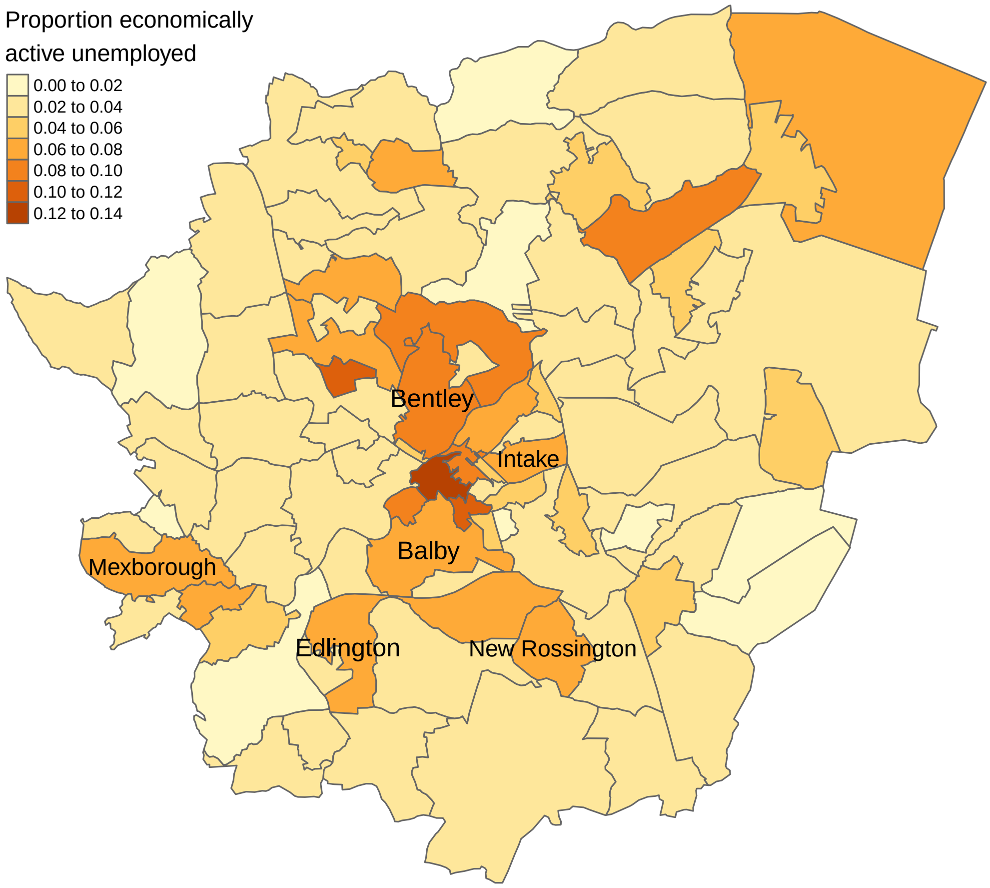

Figure 1.12: Proportion of economically active unemployed in Doncaster community areas (source: 2011 Census)

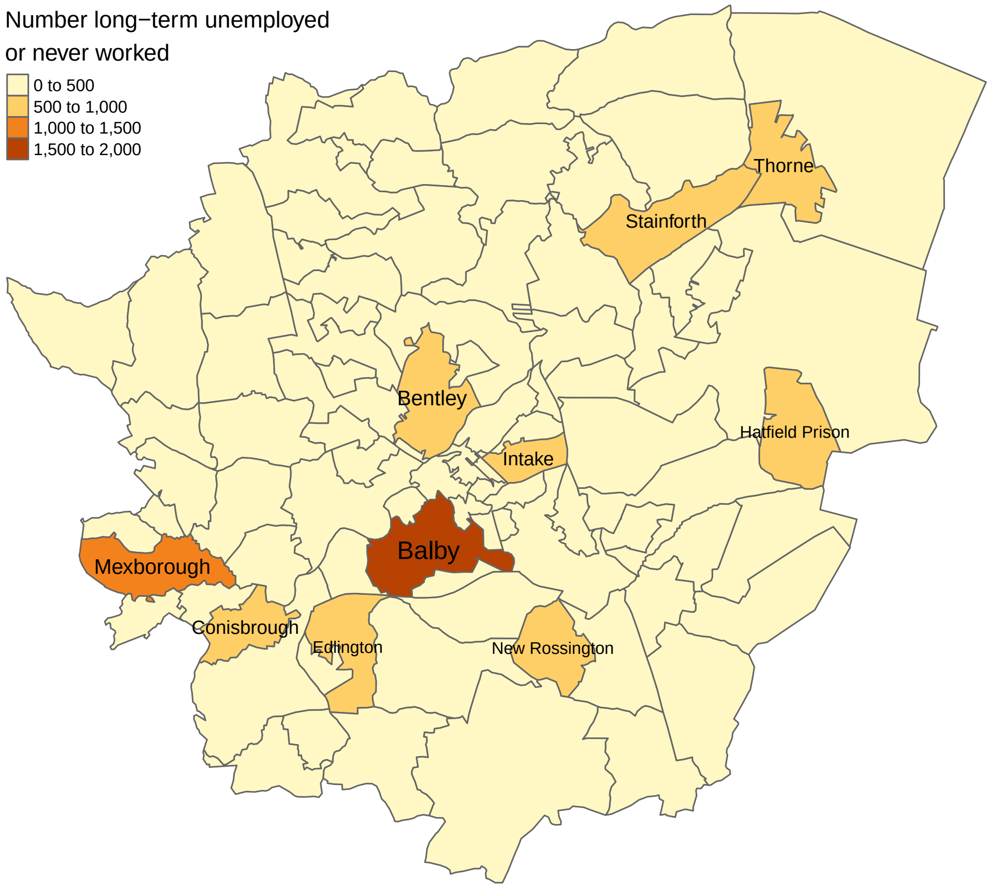

Figure 1.13: Number of people long-term unemployed or never worked by community area

unemployed.

1.4.3 Mortality

Life expectancy at birth in Doncaster is lower than the national average (Office for National Statistics, 2015c; Public Health England, 2015). Life expectancy at birth in Doncaster for males is 77.5 years while the median nationally is 79.8, a difference of 2.3 years. For females life expectancy at birth in Doncaster is 81.6 years compared to the national median of 83.4 years, a difference of 1.8 years.

Figure 1.14 shows the life expectancy at birth for local authority districts

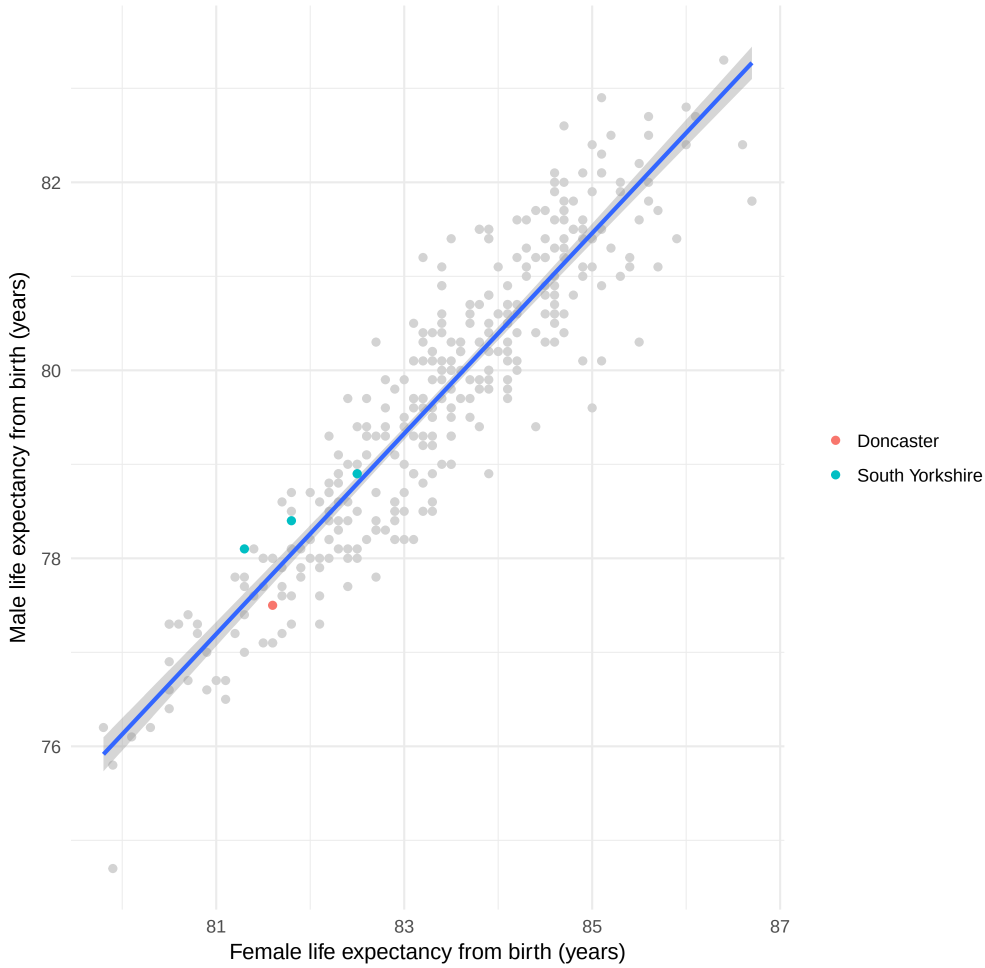

Figure 1.14: National life expectancy at birth in years by Local Authority District (LAD)

in England and Wales, with local authorities in South Yorkshire and Doncaster itself highlighted in blue and pink respectively for comparison.

Doncaster has one of the highest rates of premature mortality of any local authority in England, where a death is considered premature if the individual died aged less than 75. In the years 2013–2015 Doncaster had a premature death rate of 400 premature deaths per 100,000 population (Public Health England, 2015). Doncaster has a higher rate of premature mortality than both Barnsley and Rotherham, both local authorities in South Yorkshire that are considered statistically 'similar' to Doncaster

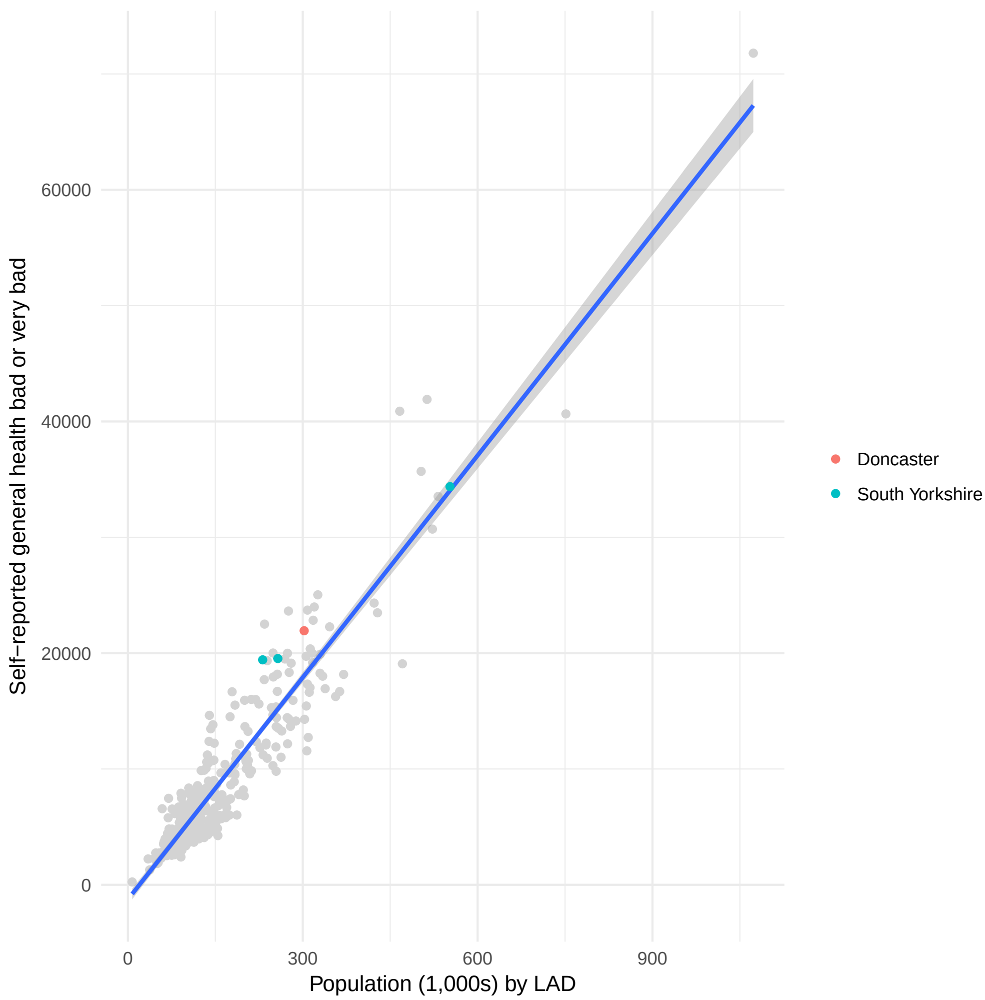

Figure 1.15: Self-reported general health bad or very bad by LAD

by Public Health England in terms of deprivation.

1.4.4 Morbidity

Doncaster has a slightly higher number of people with bad or very bad self–reported general health than local authority districts (LADs) with similar population sizes (Source: Nomis (2013a); Office for National Statistics et al. (2017); Figure 1.15).

Doncaster has a high number of people with a limiting–long term illness or disability that limits their day–to–day activities a lot (source: Nomis

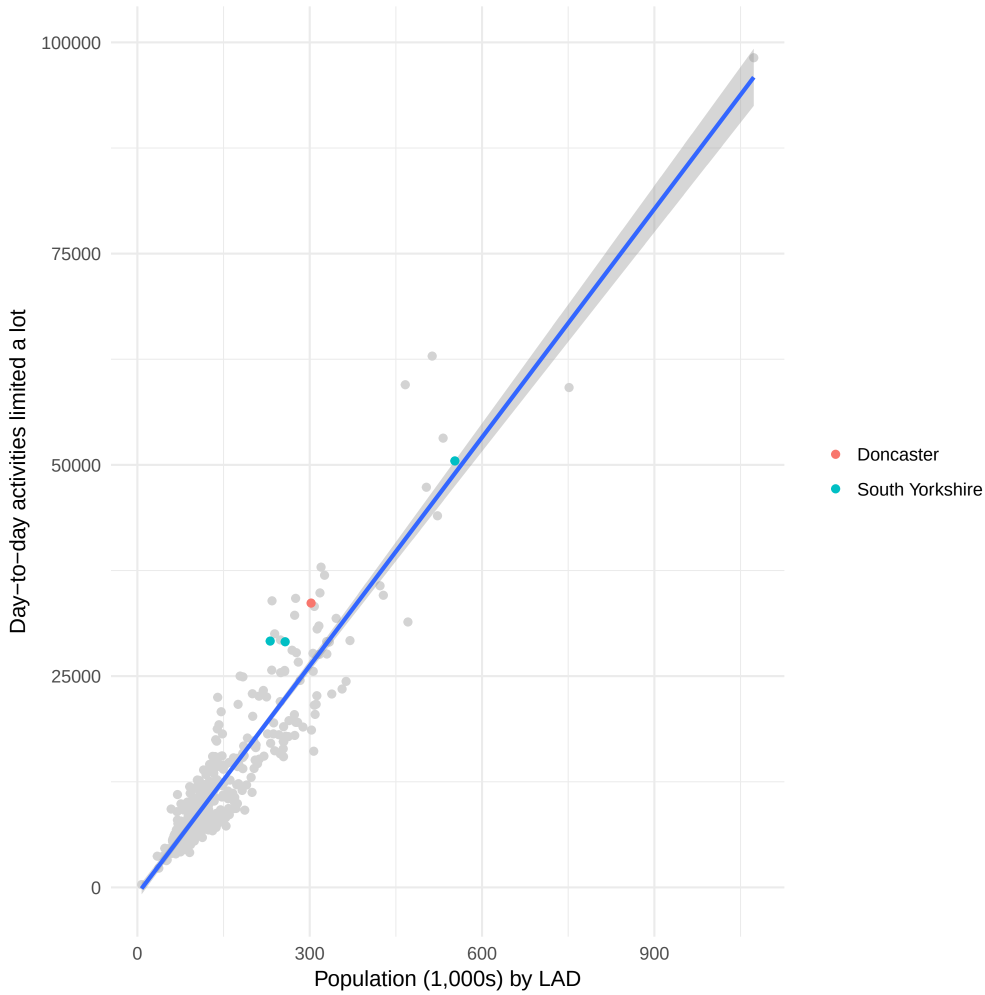

Figure 1.16: Limiting long-term illness or disability by LAD

(2013a); Office for National Statistics et al. (2017); Figure 1.16).

The prevalence of depression and anxiety in Doncaster in 2015–16 was approximately 15.0% based on the GP patient survey (Public Health England, 2016). This is higher than the England average of 12.7%. Similarly the prevalence of long–term mental health problems is higher than the England average, at 5.9% compared to 5.2%, respectively. Access to treatment in Doncaster also appears to be poorer than England overall. Two–thirds (66.1%) of patients referred to Improving Access to Psychological Therapies (IAPT) services wait less than six weeks for their first appointment, less than the England average of 84.8%.

Chapter 2

Health resilience literature

2.1 Introduction

In this chapter I summarise the available literature on resilience, and health resilience more specifically. Early literature on resilience tended to be psychological in nature, and this still has important implications for contemporary resilience literature. Most notably, it is this early literature that first used the notions of risk exposure, positive outcomes, and protective factors which still frame the majority of resilience research today.

As part of this summary I describe some of the measures of risks, positive outcomes, and protective factors used in this research. I expand on this list in Chapter 3, which is a systematic scoping review of health resilience literature I use to identify the range of measures used to articulate risks and positive health outcomes.

I conclude by outlining some of the determinants of health that affect clinical depression which inform my selection of independent variables for the spatial microsimulation model, which I describe further in Chapter 4.

2.2 History of resilience

Research into resilience is generally considered to have first emerged in the 1970s (Luthar et al., 2000; Sameroff and Seifer, 1983; Schoon, 2006). Examples of resilience research from this era include research into the outcomes of children recruited for 'Project Competence' based in Minnesota (Garmezy et al., 1984; O'Dougherty and Wright, 1990), children born in 1955 in Kauai, Hawaii (Werner and Smith, 1977, 1992), and children in the 1970 Rochester longitudinal study (Sameroff and Seifer, 1983, 1990).

The hallmark of these studies was their investigation of high–risk children and young people who exhibited positive outcomes. These young people at high–risk who nevertheless experienced positive outcomes were considered 'resilient'. The researchers studied and hypothesised about what factors might have encouraged positive outcomes in these high–risk young people. These were deemed to be 'protective factors', and they existed in the middle of the pathway between risk exposure and positive outcome, forming a barrier preventing the transition from high–risk status to negative outcome seen in some individuals.

A common methodological approach (Schoon, 2006; Werner and Smith, 1977, 1992) was to use longitudinal data, allowing the researchers to track the choronology of risk, protective factors, and outcomes. For example, Werner and Smith used data from the 1955 birth cohort study of children born on the island of Kauai, Hawaii. As well as data available in the birth cohort—which provided information on the cohort at birth, infancy, age two, and age ten—the authors performed their own follow–ups at age 17 to 18 beginning in 1972 (Werner and Smith, 1977) and age 31 to 32 beginning in 1985 (Werner and Smith, 1992). Figure 2.1 shows the location of Kauai in the Hawaii archipelago (boundary data from Biogeo (2016)).

This pioneering research tended to focus on the psychological and psy-

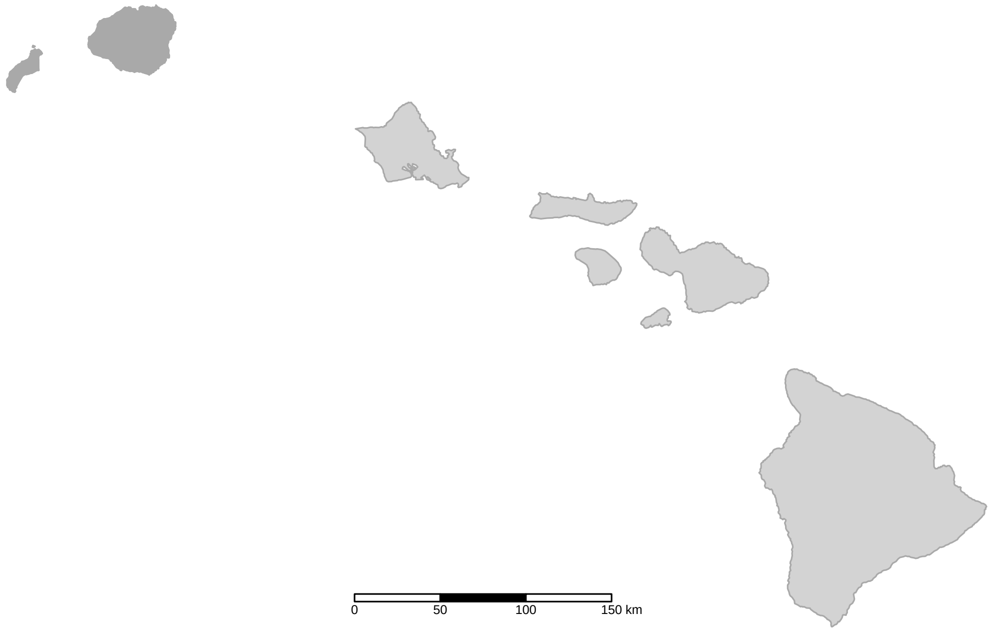

Figure 2.1: Kauai is north west in the Hawaii archipelago

chosocial domains. Perhaps because of the focus on the psyche, high–risk status tended to be identified by psychological measures, such as a mental illness diagnoses. Similarly, at first protective factors that were internal to the child or young person were considered, such as their temperament or social skills. “As work in the area evolved, however, researchers increasingly acknowledged that resilience may often derive from factors external to the child” (Luthar et al., 2000: 544). This developed into the theory that three groups of protective factors affected the development of resilience in such children and young people: attributes of the children themselves; attributes of their families; and the availability of external sources of support and the ability of the family to obtain them (Garmezy and Masten, 1986: 511; Luthar et al., 2000: 544).

As research in this field developed so too did the understanding of protective factors. Instead of a barrier that resilient individuals had—and non–resilient individuals did not—a more nuanced understanding of ‘protective mechanisms’ or ‘processes’ developed. These enable *some* individuals under *some* circumstances to be resilient, that is it is the combination of the individual's own abilities and their current circumstances that allows them to remain resilient in the face of risk (Rutter, 1987: 317).

Not dissimilar to attempts to understand 'the causes of the causes' (Rose, 1992; The Marmot Review Team, 2010) of poor health outcomes in health inequalities research, the focus of resilience research moved on to try to understand the 'causes of the causes' of positive outcomes. Self–esteem, for example, is undeniably beneficial for individuals who have it as it helps them to achieve and maintain good mental and physical health. But rather than seeing 'self–esteem' as the source of resilience, authors (Rutter, 1987; Schoon, 2006) asked what enables some individuals to develop and maintain self–esteem when others in similar circumstances do not?

> . . . we need to ask why and how some individuals manage to maintain high self–esteem and self–efficacy in spite of facing the same adversities that lead other people to give up and lose hope. *How* is it that some people have confidants to whom they can turn? *What* has happened to enable them to have social supports that they can use effectively at moments of crisis? (Rutter, 1987: 317, emphasis added)

2.3 Risk and positive outcomes

Despite similarities in this literature there were differences in how each study operationalised resilience (Luthar et al., 2000), with different measures of 'high–risk' and positive outcomes. Some measure of high–risk status and some measure of positive outcome are necessary to articulate resilience, but there is no standard definition or criteria for deciding what these should be. I return to this issue in Chaper 3 but outline some of the measures used for each below.

Given the psychological origin of this research many psychological and psychosocial measures and definitions were used to identify 'at risk' or ‘high risk’ children and young people in these studies. Garmezy and Streitman (1974) considered a child to be ‘at risk’ “if there is a greater likelihood that he [sic] will develop a mental disorder than a randomly selected child from the same community” (Garmezy and Streitman, 1974: 17). Similar measures of at–risk children included those with a mother with a diagnosis of schizophrenia (Sameroff and Seifer, 1990), affective disorder, or personality disorder (Garmezy et al., 1984). Studies using the 1955 Kauai birth cohort considered high–risk children as those with: a learning disability diagnosis and recommendation to attend special educational classes; a need of long–term—greater than six months—mental health services; a need of short–term—less than or equal to six months—mental health services; or those with a ‘new behavioural problem’ when followed–up at age 10 (Werner and Smith, 1977: 26).

In time later studies used more general risks that extended beyond the psychological. These risks included: socio-economic disadvantage; urban poverty; community violence; chronic illness; and catastrophic life events (Luthar et al., 2000: 554). In a later follow–up Werner and Smith (1992) considered 72 cohort members (42 females and 30 males) who were born into poor families—as measured by the ‘breadwinner’s occupation’, income, and condition of housing—and who experienced additional ‘potent’ risk factors before age two to be high–risk. In studies that examined medical risk, severe heart defects—cyanotic congenital heart defects (O’Dougherty and Wright, 1990), moderate or severe perinatal stress, low–birth weight (<2,500g), physical handicaps, and alcoholic parents (Werner and Smith, 1992: 55) were taken as the risk exposure.

The positive outcomes resilient children achieved—and researchers measured—included: educational achievement; low unemployment; higher employment grade; lower reported work–related stress; marriage or entering a long–term commited relationship (females); placing a high value on parenting and caring for their children; remaining law–abiding (especially when compared to non–resilient, high–risk peers); and an absence of significant mental health problems (Werner and Smith, 1992:

59–63).

2.4 Protective factors

Early research into resilience began by examining the types of factors that enabled children and young people to achieve positive adaptation and positive social adjustment despite being considered ‘high–risk’ or ‘at risk’ (Garmezy and Streitman, 1974; Garmezy et al., 1984; Luthar et al., 2000; Werner and Smith, 1977, 1992). Werner and Smith (1977), for example, found the resilient children in their study differed from their non–resilient peers in a number of ways.

They found that most of the resilient children grew up in a family with a maximum of three other siblings (four children in total) and that there were at least two years between the resilient child and any other siblings (Werner and Smith, 1992: 56). The resilient children had not experienced any prolonged separation fom a primary caregiver in their first year of life, and they formed a close bond with one or more caregivers who could be either parental or ‘substitute parents’, for example a grandparent or older sibling (Werner and Smith, 1992: 56).

As infants they had temperaments that “elicited positive attention”, were considered active, affectionate (females), and good-natured (males), and also had “fewer eating and sleeping habits that distressed their parents” (Werner and Smith, 1992: 56).

In difference to their low–risk peers, the high–risk resilient group tended to withdraw from troubled relationships with parents, but it could be argued the low–risk group did not *need* to as they did not experience the same problems with their relationships with their parents. Many of the children, by nature of being from high–risk families, had parents who divorced, had illnesses, or lived in households with ‘chronic family discord’ (Werner and Smith, 1992: 65). The resilient children tended to cope by becoming detached or withdrawn from these situations, in comparison to their high–risk peers who continued to be involved (Werner and Smith, 1992: 65).

Despite this the resilient cohort's family played an important role in the positive outcomes of many of the resilient children, despite contributing to or causing their high–risk status. Educational level of an opposite–sex parent was strongly associated with positive adaptation in adulthood (Werner and Smith, 1992: 177). The males tended to have older fathers, had more positive interactions with caregivers, and had higher ratings of family stability (Werner and Smith, 1992: 179).

In addition to family, community also played an important role in the resilient cohort's positive adaption. Resilience was associated with having additional caring adults, including grandparents, uncles and aunts, neighbours, parents of partners and boy– or girlfriends, youth leaders, church leaders, and, in adolescence, teachers (Garmezy and Streitman, 1974: 64; Werner and Smith, 1992: 178).

These factors all suggest that, despite their adverse beginnings, these resilient children learned how to establish and maintain important social relationships with family, peers, and elders who they are able to draw on for support and encouragement. This is in contrast to their peers who were high–risk but not resilient, who struggled to achieve positive outcomes and who may have lacked the social skills necessary to form such bonds.

In cases where children had mental health problems the ability of the family to obtain psychological and psychiatric support from professional and community services—and the knowledge of the existence of such support—helped the children to manage their condition (Garmezy and Streitman, 1974: 63; Werner and Smith, 1977: 216).

Perhaps one of the most powerful protective factors among the resilient young people was "… faith that life made sense, [and] that the odds could be overcome" (Werner and Smith, 1992: 177), an 'internal locus of control'. Locus of control and competence had a positive effect even on high–risk youths with more severe needs, such as learning disability or long–term mental health problems:

> The degree to which youth had faith in the effectiveness of their own actions was related not only to the effectiveness with which they used their intellectual resources in scholastic achievement but also to positive change in behaviour in adolescence. An internal locus of control was a significant correlate of improvement (Werner and Smith, 1977: 220)

Work ethic ('hard work' and 'persistence') was mentioned by young people with more severe mental health issues in childhood who later improved (Werner and Smith, 1977: 221).

Werner and Smith argued that one of the biggest differences separating the high–achieving and low–achieving high–risk individuals was their goal setting and aspirations. Career and employment success was the most important goal for the resilient cohort but the lowest priority for their non–resilient peers (Werner and Smith, 1992: 69).

The resilient cohort faced many of the same difficulties as their high–risk peers, but took opportunities when presented with them to recover from these difficulties. These opportunities presented at 'major life transitiions', and included marriage or entering a long–term committed relationship, the birth of a child, employment and establishment of a career, graduating from high school, going to and graduating from college (university), joining the military, and becoming an active member of a church group (Werner and Smith, 1992: 178). The authors contrasted the resilient cohort's focus on taking opportunities with their high–risk peers who instead discussed life events that limited opportunities, including divorce or the break–up of a long–term relationship, the death of a parent (women), and moving away from home (men) (Werner and Smith, 1992: 178).

The resilient cohort overall had greater work satisfaction, measured by self–rated satisfaction with work or school achievement at age 31 or 32 when asked in a structured interview or self–completion questionnaire

(Werner and Smith, 1992: 179). The resilient cohort had higher self–rated satisfaction with their state of life (Werner and Smith, 1992: 181).

In adulthood the resilient cohort tended to have positive relationships with their parents–in–law or the parents of their long–term partner, and many resilient women in particular sought emotional support from their parents–in–law (Werner and Smith, 1992: 66). The resilient cohort had more 'satisfying' relationships with their siblings as adults based on self–rated responses, and this was most notable among siblings who had alcoholic or mentally ill parents (Werner and Smith, 1992: 67). They also had more satisfying relationships with parents, spouses or partners, and children at age 31 or 32 (Werner and Smith, 1992: 180).

Their relationships with friends were more complex; they shared a similar number of friends with their high–risk peers and had similar satisfaction with their relationships, but tended to be more self–reliant and rely on friends less for financial support and counsel than their high–risk peers (Werner and Smith, 1992: 68–69). As outlined above these resilient individuals tended to have an internal locus of control and be more self–confident, so it is perhaps not surprising that they relied on themselves more to address problems. In addition, they may simply have had greater financial resources as a result of their employment—which tended to be of a higher grade—or simply be better at managing their own money.

2.4.1 Sex and gender

The authors noted that resilient girls tended to have increased autonomy and responsibility in households where the mother worked and the father was absent, for example by providing care to younger siblings (Werner and Smith, 1992: 57). For women, having a mother who had steady employment also had positive results (Werner and Smith, 1992: 177).

Resilient women were significantly more likely to have had "regular household chores and domestic responsiblity during adolescence" while resilient men had higher self–rated temperament and activity scores (Werner and Smith, 1992: 177). It seems unlikely that such a result is genetic given what is known about gendered role profiles but, nevertheless, perhaps the resilient young people found their lives somehow easier if they conformed to these gendered expectations.

Resilient boys tended to have a positive male role model, although this was not necessarily the child's father (Werner and Smith, 1992: 57). Both resilient boys and resilient girls had additional role models outside of the family, including close friends, teachers, neighbours, youth leaders, ministers or faith leaders, or elders (Werner and Smith, 1992: 57).

In females, internal protective factors—self–esteem, for example—had the biggest effect on resilience. For males, outside sources of support—for example from caregivers, friends, and family—had the biggest effect on resilience.

2.4.2 Development stage

Because of the longitudinal nature of these studies the authors were able to explore which factors affected the participants's resilience at any given developmental stage. By comparing groups of children and young people with different socio-economic and familial circumstances in early life and comparing their trajectories into adulthood these studies were able to explore which factors led to positive outcomes.

In the 1972 follow–up Werner and Smith were successful in tracking down 88% of the original Kauai cohort. This included an 'at–risk' and a control group of young people, matched for age, sex, socio-economic status, and ethnicity. In the 1985 follow–up the researchers managed to obtain responses from 82% of the original cohort, for which data was available at birth, infancy, age two, age 10, and age 18 (Werner and Smith, 1992: 34). All participants were surveyed for education and health outcomes, ability, achievement, and personality using standard instruments contemporary for the time (Werner and Smith, 1977: 24).

Werner and Smith (1992) suggest that during infancy and early childhood, constitutional factors such as health and temperament played the biggest part in effecting resilience. This changed as the resilient children matured, and by middle school their verbal and reasoning skills played a bigger part in their positive development. By late adolescence and adulthood their personality characteristics—self–esteem and internal locus of control—most helped to reinforce their resilience and positive adaptation (Werner and Smith, 1992: 57).

At age two the resilient cohort displayed alertness and autonomy, sought out experiences, had a 'positive social orientation', and had better communication, locomotion, and self–help skills than their high–risk but non–resilient peers (Werner and Smith, 1992: 56). The coping style of the primary caregiver at age two, observed by psychologists and paediatricians, was linked with positive adaptation, as was the presence of rules and structure in the household.

In infancy the resilient group had good sleeping and eating habits, and were considered 'affectionate and cuddly' (girls) or 'very active' (boys) (Werner and Smith, 1992: 173). At age two assessments by paediatricians and psychologists found the resilient children to be more agreeable, relaxed, responsive, self–confident, and sociable. In comparison, their high–risk non–resilient peers were characterised by anxiety, fearfulness and suspicion and were more frequently withdrawn (Werner and Smith, 1992: 176).

At age ten, teachers (for boys) and parents (for girls) noted fewer behavioural problems, and at age 17 and 18 the resilient cohort enjoyed greater popularity among their peers (Werner and Smith, 1992: 176). In grade four (approximate age nine to ten) the resilient children had higher reading achievement scores, especially among the boys (Werner and Smith, 1992: 176). In elementary school the resilient children got along well with classmates, had better reasoning skills, better reading skills, and had many interests including "activities and hobbies that were not narrowly sex–typed" (Werner and Smith, 1992: 56).

Between age 10 to follow–up at age 17–18 'perception of parental understanding', peer support, the young person's belief in their own abilities, hard work, persistence, and ability to communicate in the first language of the island ('standard English') were associated with improvement and positive change (Werner and Smith, 1977: 216).

The authors compared characteristics of these high–risk children who had positive outcomes with high–risk children who did not fare so well by ages 10 and 18, matched for age and sex. These young people by contrast had learning problems, mental health problems, and 'serious delinquencies' (Werner and Smith, 1992: 56).

In senior year of high school (approximate age 18) the resilient group considered their school experience to be more positive and had higher—and more realistic—expectations for their future (Werner and Smith, 1992: 176). In addition, their interviewers considered them to have higher self–esteem.

In adulthood (ages 31 and 32) the resilient group had lower self–rated distress and emotionality using the EAS temperament survey instrument, and women had higher self–rated sociability and lower anger (Werner and Smith, 1992: 176). The authors also found a significant association between the resilient groups' problem solving skills (PMA IQ) at age ten and successful adaptation in adulthood (Werner and Smith, 1992: 176).

2.4.3 Validation

Using discriminant function analysis, Werner and Smith were able to enter these protective factors chronologically into their model. In 94.4% of cases they were able to identify the correct, resilient, individuals by entering all protective factors. The authors were still able to identify the majority—87.5%—of individuals correctly by entering protective factors the cohort were exposed to between birth and age two (Werner and Smith, 1992: 182–183). Crucially the success rate was similar, even higher, for high–risk non–resilient peers (96.8%, all protective factors) suggesting validity in the measures (Werner and Smith, 1992: 183).

2.5 Geography and health resilience

The historical resilience literature I described in Sections 2.2 to 2.4 was based on individual, and sometimes family, experiences of risk and positive outcomes which rarely discussed environmental, area–based, or geographical factors external to the subjects that may have contributed to their resilience. As studies of resilience in other disciplines began to appear, the range of measures expanded beyond the individual and included geographical resilience literature which took account of area–based factors for the first time.

> In psychology literature 'resilience' describes the process whereby people avoid the negative outcomes associated with risks. Related processes may operate at the population level, with some deprived communities resisting the detrimental health effects of adverse socioeconomic conditions, while others succumb (Doran et al., 2006: 686).

As well as broadening the range of measures of risk and positive outcomes to the health domain, geographical resilience literature also used a range of geographical units to assess area–based effects (Cairns et al., 2012; Cairns–Nagi and Bambra, 2013; Doran et al., 2006; Mitchell et al., 2009; Tunstall et al., 2007).

Doran et al. (2006) found several English local authorities that had better than expected life expectancy for their level of deprivation. They identified deprived local authority districts using the Townsend material deprivation index (Townsend et al., 1988) using data from the 1991 census and sociodemographic context from Office for National Statistics *classification of local and health authorities of Great Britain* (Doran et al., 2006: 686). They found these strong predictors of life expectancy using Spearman's rank correlation coefficients (Doran et al., 2006: 688).

Taking the highest standardised residuals to be 'resilient', Doran et al. (2006) found 'education centres' to have comparatively high life expectancy given their level of deprivation (Doran et al., 2006: 682), but mining, manufacturing, and industrial areas often underperformed in terms of life expectancy (Doran et al., 2006: 687). They argue that most outliers are not atypical and may be part of a long–term pattern of inequality where the north is more deprived.

Tunstall et al. (2007) examined the relationship between age–specific mortality and long–term economic adversity using data from 1971–2001 in parliamentary constituencies in Britain. They used a bespoke index of adversity constructed from census variables to primarily identify areas of low labour market activity which was strongly correlated with common deprivation measures (Tunstall et al., 2007: 338). Using this measure they identified 54 'persistently disadvantaged' areas. For these 54 areas the authors calculated a 'resilience score' based on the mortality distribution compared to deprivation through time (Tunstall et al., 2007: 338) and identified 18 above–average resilient areas. Barnsley East and Mexborough was one of the resilient areas identified (Tunstall et al., 2007: 340). This parliamentary constituency no longer exists but did overlap the Doncaster local authority district. Figure 2.2 shows where the Barnsley East and Mexborough parliamentary constituency (grey polygon) overlapped with the Doncaster local authority district (black outline).

Mitchell et al. (2009) also examined the relationship between low mortality rates and 'persistent economic adversity' from 1971–2001 in 54 parliamentary constituencies using mixed methods. Barnsley East and Mexborough parliamentary constituency was identified as a resilient area using their criteria of "significantly lower mortality rates" (Mitchell et al., 2009: 19).

They argue that one of their case study areas with a large South East

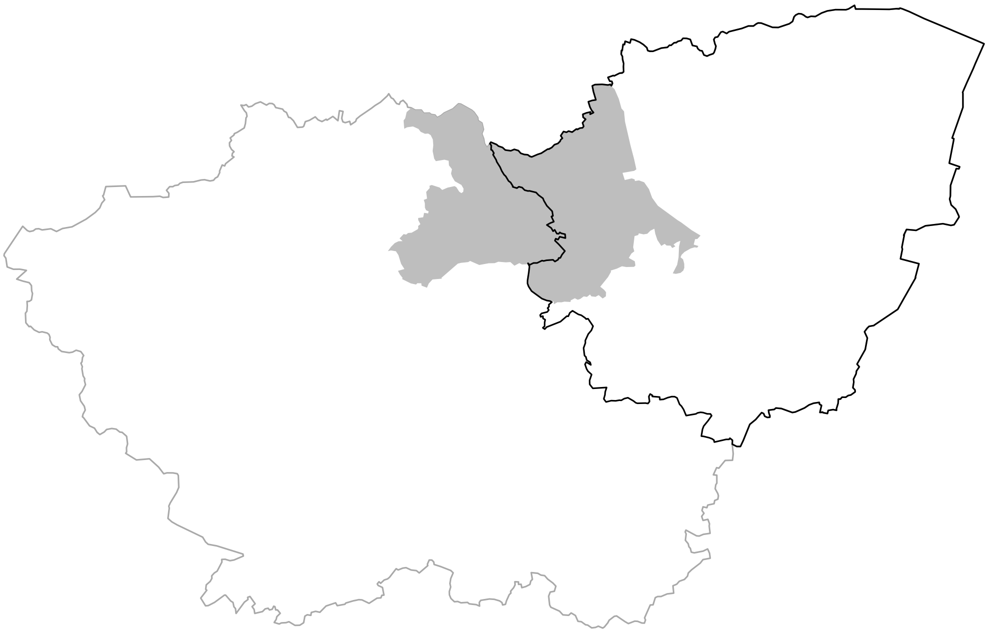

Figure 2.2: Barnsley East and Mexborough Parliamentary Constituency

Asian and Caribbean population had lower than expected mortality, perhaps because people with these ethnic backgrounds have lower rates of cancer mortality and cardiovascular mortality, respectively (Mitchell et al., 2009: 19). Howevever, this did not hold for other resilient areas: "[s]ome of the resilient areas were among Britain's most ethnically mixed, yet other very mixed areas were not resilient" (Mitchell et al., 2009: 19). They found little difference between the availability of green space between resilient and non–resilient areas (Mitchell et al., 2009: 20), but this could not take account of the quality of these green spaces. Similarly levels of social capital were not significantly different between resilient and non–resilient areas, using political participation—measured by voter abstention rates in general elections from 1979–2001—as "a valid proxy for the degree of social capital in a community" (Mitchell et al., 2009: 21). They did find that "resilient constituencies were significantly better at retaining or attracting population in the face of economic adversity than the non–resilient areas" (Mitchell et al., 2009: 20).

The lack of difference between resilient and non–resilient areas could be attributed to the ‘crudeness’ of the quantitative measures used—something that subsequent studies have attempted to address (Cairns et al., 2012; Cairns–Nagi and Bambra, 2013), or evidence that another attribute must also be present in areas that acts as a ‘catalyst’ to improve resilience (Mitchell et al., 2009: 21).

Because of the crudeness of health outcome measures in these studies Cairns et al. (2012) expanded this research to include morbidity indicators. These included self–reported general health and limiting long–term illness or disability, but were also able to include emergency hospital admissions and chronic heart disease (CHD) hospital admissions from the 2011 hospital episode statistics (Cairns et al., 2012: 928–929) which they combined into a composite mortality index. They used multiple correspondence analysis (MCA) to test the association between area resilience and ethnic density—living in an area with a high proportion of people with the same ethnic background, residential mobility—the rate of moving in and out of an area, employment type, housing tenure, and social cohesion—using a proxy index of social fragmentation (Cairns et al., 2012: 928–929).

They found 15 mortality resilient parliamentary consituencies, nine morbidity resilient areas, and four were resilient for morbidity and mortality (Cairns et al., 2012: 930). Areas were considered resilient if they were in the highest quartile for the composite mortality index (Cairns et al., 2012: 930). Doncaster was not among the resilient areas identified in this study. Factors the authors suggest help improve health resilience were availability of social housing, higher quality employment—in higher occupational grades, and relatively high ethnic density (Cairns et al., 2012: 932).

Cairns later extended her research to examine outliers of the relationship between socio–economic deprivation and poorer population health at the area level (Cairns–Nagi and Bambra, 2013). They measured deprivation using the Townsend material deprivation score (Townsend et al., 1988) and used a combination of morbidity and mortality health measures, including self–reported general health, self–reported limiting long–term illness or disability, and premature mortality at the area level (Cairns–Nagi and Bambra, 2013: 229). ‘Areas’ examined were England and Wales local authority districts (LADs) and census area statistic wards (CASWARDs). Using these measures, Cairns–Nagi and Bambra (2013) used regression tree classification—alternatively known as recursive partitioning—to separate resilient from non–resilient areas, using a standardised residual of less than -1.96 to signify a resilient area (Cairns–Nagi and Bambra, 2013: 231). Aggregate tables from the 2011 census were not yet available at the time this study was published, so there is an opportunity to update analyses of resilient areas with 2011 data.

The authors found the only resilient local authority districts, using their criteria, were in London and the East of England. There was greater variation among CASWARDS with resilient areas found in all regions except the North West, suggesting resilience could be predominantly a small–area phenomenon (Cairns–Nagi and Bambra, 2013: 231).

Cairns–Nagi and Bambra (2013) used Townsend material deprivation scores to articulate risk, with high scores over a long period of time (1971–2011) associated with high risk. Using this measure the authors studied *areas* rather than *individuals*. They choose not to assess the whole spectrum of deprivation, instead concentrating on the bottom quintile (most deprived 20%) in their analysis. To measure health outcomes they analysed self–reported general health and limiting long–term illness or disability from the census, and premature mortality—defined as those who died below the age of 75 years—after standardising for age and sex using the England population as reference (Cairns–Nagi and Bambra, 2013: 231).

2.6 How many are resilient?

One of the areas of contention about health resilience is, exactly what proportion of the population are resilient? The psychological literature tended to identify an outcome that must be met to consider an individual resilient. For example, Werner and Smith (1992) considered children to be resilient if they were exposed to a defined risk but achieved a positive outcome. "... one out of every three of these high risk children (some 10% of the total cohort) had developed into a competent, confident, and caring young adult by age 18" (Werner and Smith, 1992: 2).

Geographical health resilience literature tends to specify a threshold over which *areas* are resilient because this literature tends to deal with information about individuals aggregated to a the area level. For example Cairns–Nagi and Bambra (2013) considered a standardised residual of ≤ 1.96 to indicate 'health resilient' areas (Cairns–Nagi and Bambra, 2013: 231). They found between three and five health resilient LADs for each of the morbidity and mortality measures they analysed, all of which were within London or East of England regions only. This equates to approximately 1.4% of LADs, based on 354 LADs examined (Cairns–Nagi and Bambra, 2013: 230).

When analysing CASWARDs they found between 62 and 90 health resilient areas depending on the health outcome measure used. In addition they found 36 health resilient areas common to all three measures of morbidity and premature mortality. The 36 common CASWARDs equates to less than 0.5% of CASWARDs identified as health resilient, while the maximum 90 health resilient CASWARDs (identified by self–reported health) equates to only 1.1%.

This second approach works well on an area–level basis, as the number of people who have a positive health outcome is a continuous variable so a standard deviation or other numerical threshold is a useful approach. One of the main advantages of using the spatially microsimulated data set, though, is that it is possible to work with individual–level data.

I will use a combination of these two approaches, essentially specifying individual–level and area–level criteria. I have chosen to use clinical depression as the health outcome measure to indicate resilience (see Section 6.2). As this outcome is binary—does or does not have clinical depression—a numerical threshold will not work. Instead I will consider individuals with a high–risk exposure but not clinical depression to be resilient. Alongside this I use area–level aggregate measures of deprivation. I examine these empirically after simulating the data set in section 6.7.

2.7 Factors affecting mental health

As with psychological resilience, geographical health resilience literature uses measures of risk, positive outcomes, and protective factors to articulate resilience. For health resilience, the risk or exposure is usually a measure of deprivation, and the positive outcome is usually an indication of positive health or wellbeing. I outline these more comprehensively in Chapter 3.

By articulating health resilience as the relationship between deprivation (risk) and positive health outcomes (positive outcome) this can be thought of as an expression or function of the social determinants of health, and of health inequalities more generally.

The study of health inequalities and, later, the social determinants of health emerged after careful observation of cardio–vascular health and premature mortality of government employees in Westminster in the 1980s. The study by Marmot and his colleagues demonstrated that civil servants of lower employment grade had poorer health than their contemporaries of higher employment grade overall (Brunner and Marmot, 2006; Marmot et al., 1984). The social determinants of health attempt to explain the causes and mechanisms underpinning these health inequalities.

Here I outline a few key determinants of health using, where possible, results of systematic literature reviews that considered clinical depression. These either informed my theoretical selection of independent variables that I used to construct my model in Chapters 5 and 6), so that I could control for them, or for the analysis I present in Chapter 7, or both.

2.7.1 Age

Age is an important determinant of health in its own right, as people tend to experience an increased number and range of detrimental health outcomes as they grow older. These can include sensory loss, musculo–skeletal conditions, diabetes, chronic obstructive pulmonary disease (COPD), depression, coronary heart disease (CHD), and stroke (World Health Organization (2015), Age UK (2017)). Onset of these, and other conditions, are often a result of biological damage in older age, but this varies between individuals and is affected by other factors such as socio–economic position (McMunn et al., 2006).

Late–life depression can differ from depression among younger sufferers. Depression among older people can be associated with a loss of physical functioning in addition to other risk factors, and anxiety is higher among older people with depression (Pruckner and Holthoff–Detto, 2017: 662). There may also be additional risks for older people with depression. A study of participants in the Swedish National Study on Aging and Care found that older people aged 60 and over showed cognitive decline overall, but that respondents transitioning into a depressed state showed expedited cognitive decline (Pantzar et al., 2017: 681). It was not clear if younger people transitioning into a depressed state were also at greater risk of cognitive decline, but this was nevertheless a risk for older people.

While people of any age can be depressed, or transition into depression, there are additional risk factors and outcomes for older people with depression.

2.7.2 Sex

A systematic review of 47 primary studies compared the prevalence of post–stroke depression (PSD) among men and women. The majority of studies found that the prevalence of PSD was higher among women (Poynter et al., 2009: 565–566). Furthermore, women are at risk of antenatal depression, the prevalence of which is approximately 10% in the United States (Mukherjee et al., 2016). I am not aware of any systematic reviews in the last 30 years that consider the prevalence of clinical depresion by sex. However, these studies suggest there may be a difference in depression prevalence between men and women, and that women are both at greater risk of depression and have additional risk factors that could mean they transition into depression.

2.7.3 Ethnicity

A systematic literature review and meta–analysis of clinical depression prevalence among minority and majority ethnic groups reported mixed findings (Tarricone et al., 2012). Of the twenty–five included studies, ten reported a significant difference between ethnic minority and ethnic majority groups. However, of these ten, four reported higher prevalence rates of depression for ethnic majority groups, and the remaining six reported higher prevalence among ethnic minority groups (Tarricone et al., 2012: 102).

Another systematic review of pregnant women in the United States found that clinical depression prevalence was higher among non–Hispanic Blacks (NHB) and Hispanics, compared to non–Hispanic White (NHW) individuals (Mukherjee et al., 2016: 1793). It is not clear why the two systematic reviews reached different conclusions, but it may be associated with the different populations in the study, and ethnic–minority women or ethnic–minority women who are pregnant are at greater risk of depression. The evidence for an association between ethnicity and clinical depression in general, however, is ambivalent, and may be modified by other socio–demographic characteristics or context.

2.7.4 Economic activity

Studies have demonstrated that unemployment is associated with poor mental health and greater depression prevalence (Jefferis et al., 2010; Khlat et al., 2004). In addition to the greater risk from being unemployed, there is evidence that the *duration* of unemployment is also associated with increased prevalence of depression. Stankunas et al. (2006), for example, found that people unemployed for twelve months or more in Lithuania had higher depression scores on the Beck Depression Inventory (DBI) instrument than those who were unemployed for less than twelve months, and both groups had higher depression than their employed counterparts. In a longitudinal study of young people in the United States Mossakowski (2009) again found that unemployment was associated with greater prevalence of depressive symptoms. This study did not find that the duration out of the labour market was associated with depressive symptoms, but that past unemployment duration did have a negative effect on depression (Mossakowski, 2009: 1829).

2.7.5 Education

Higher education status is argued to protect against mental health issues and cognitive decline (McLaren et al., 2015). In a representative study of Finnish adolescents, education was found to protect against the risk of depression, even among adolescents at high–risk of depression based on their parental income and socio–economic position (Korhonen et al., 2017).

2.7.6 Income, deprivation, and poverty

Area–level income and poverty is associated with higher prevalence of depression (Hiilamo, 2014; Poetz et al., 2007). Studies have also demonstrated an association between individual–level income and clinical depression (Aranda and Lincoln, 2011; Pickett and Wilkinson, 2015). For example in a study of the Danish workforce, Cleal et al. (2017) found that those with diabetes were more likely to have depression than those without, but that this was even higher among Danes in the lowest socio–economic position and with lowest incomes. Income inequality may also exacerbate an individual's depression. It is hypothesised that income inequality can 'get under the skin' (Deurzen et al., 2015), and that countries with higher income inequalities report greater numbers of people with depressive symptoms (Deurzen et al., 2015: 485–486).

2.7.7 Housing

Poor quality housing is associated with poorer physical health outcomes, but a number of recent studies also suggest an association between housing quality and mental health and clinical depression. For example in a study of US mothers Corman et al. (2016) find an association between depressive mothers and poorer housing quality in terms of heating and energy insecurity, and food insecurity (Corman et al., 2016: 82). The affordability of housing is also an important consideration to health outcomes. Using data from the UK Annual Population Survey Reeves et al. (2016) demonstrated that a reduction in housing benefit was associated with an increase in clincial depression in the UK.

2.7.8 Marital status

Hosseinpour et al. (2012) demonstrated that people who were single and had never married experienced the best self-reported health. Respondents who were married or cohabiting reported slightly worse health than their single peers, while those who were divorced, separated, or widowed reported the worst health overall (2012, p. 3). Sacker et al, however, found no relationship between cohabitation status and general health independent of other factors (2009, p. 132).

2.7.9 Social isolation

A recent systematic review of social isolation and loneliness demonstrated overwhelmingly that they are associated with detrimental physical and mental health (Courtin and Knapp, 2017), although in most cases the studies were unable to identify a causal mechanism or pathway (Courtin and Knapp, 2017: 804–805). The most commonly used measures of physical and mental health were cardiovascular health and depression, respectively. The studies were typically of older adults, although a study of primary and secondary school age children also found an association between social isolation or loneliness and poorer mental health (Matthews et al., 2015).

2.8 Conclusion

In this chapter I have briefly outlined the history of the health resilience literature, from its beginnings in psychological research in understanding positive outcomes in the face of risk. Over the course of the last forty to fifty years the field has developed, and a geographical health resilience literature has emerged that takes into account area–based positive outcomes. Throughout these literatures there are a range of measures used as health outcomes, risks, and protective factors or characteristics which I explore in more detail in Chapter 3. My choice of outcome variable is nevertheless informed by this initial literature review.

A key feature of any health resilience literature is how resilient individuals or areas are to be categorised. Some require an approach based on categorical information, while some—typically aggregated area–level data—use a numerical threshold as a 'cut–off'. I will need to use a combination of both, and I explore this emprically in Chapter 6.

The rest of the thesis proceeds as follows. Chapter 3 expands on many of the issues identified in this chapter, namely the measures used as health outcomes, risk, and protective factors. Chapter 4 then discusses the spatial microsimulation technique, how it has been used in similar fields successfully before, and what it can offer the understanding of health resilience. Further chapters then test the spatial microsimulation method, simulate and describe the results of the health resilience simulation in Doncaster, and offer policy analyis and suggestions based on these results.

Chapter 3

Systematic review

3.1 Introduction

A preliminary review of relevant literature, which I described in Chapter 2, revealed a broad range of ways in which 'health resilience' was conceptualised and operationalised. That is to say, exactly what health resilience meant and how it was measured varied from study to study. This ambiguity extended to both the characteristics thought to influence resilience as well as the health outcome deemed to be affected by resilience.

The variety suggested there was little consensus about how health resilience was defined and measured, limiting its practical use in empirical research and policy making. It therefore became apparent that a comprehensive understanding of the range of concepts and definitions used in this research field would be essential in order to understand the diversity of evidence and to choose appropriate concepts and definitions to apply in future work. To ascertain the range of uses of the term in empirical literature I undertook a systematic scoping review of relevant studies.

This chapter: describes the methods used to select, appraise, and synthesise relevant empirical studies; presents a summary of the individual papers; and provides a synthesis of the various ways health resilience was operationalised in this literature. It is presented using the recommendations stated in the Preferred Reporting items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher et al., 2009; PRISMA, 2013), adapted to suit a social science study. These recommendations were followed to allow readers to assess the quality of the review and its findings, and replicate it if desired (Booth et al., 2012, ch. 9).

3.2 Rationale

A systematic scoping review was undertaken to explore the ways in which health resilience was used in empirical literature. A systematic scoping review was ideally suited to this task because it is a tool intended to gather the breadth of information available about a topic efficiently and comprehensively. Booth et al. (2012) describe the scoping review as a "preliminary assessment of potential size and scope of available research literature [which] [a]ims to identify [the] nature and extent of research evidence" (2012: 27 [adapted from table 2.3]). In addition the scoping review does not typically appraise the quality of the evidence, which fits the requirements of my review as this was not appropriate. The synthesis and output of the scoping review can also be a narrative, which again fits the requirements of this review. A properly conducted systematic review will also mitigate issues of bias, which will help to ensure the results of the review are a valid reflection of the state of the existing literature on health resilience.

3.3 Objective

The objective or research question that initially defined this systematic review was stated as:

> What literature is available about the associations between the socio–economic position of working–age individuals (aged

> 16–74) in England and Wales and their health, and their ability to recover from, or avoid, poor health?

During the review process it also became apparent the objective needed to be more specific. Instead of simply *identifying* papers about health resilience, it became clear I needed to focus on the *measures* used to operationalise health resilience. Because of this I updated the objective to be:

> How are 'socio–economic position', 'health' and, therefore, 'resilience' operationalised in literature which mentions these—or synonymous—terms which is based on research of individuals aged 16 and over conducted in the United Kingdom?

3.4 Methods

3.4.1 Eligibility criteria

The eligibility criteria were designed to favour breadth, rather than sensitivity or specificity (Booth et al., 2012: 70), since the aim of this review was to explore the extent of literature available.

Studies that only included children, specifically aged less than 16, were excluded. This was done so that the results of the systematic review would be based on the same age group as that in *Understanding Society* which the simulations in chapters 5 and 6 are based on.

Studies that were not based on data or respondents from the United Kingdom, or one of its constituent countries, were excluded. I made this decision to match the geography of the data sources I used in the spatial microsimulation. England is the geographical focus of my case study area. I use census data from England and Wales to constrain the spatial microsimulation. Respondents in *Understanding Society* are from any of the four UK countries.

Some empirical papers and systematic reviews were based on international data or data from multiple countries, often as a comparison. It was necessary to make a judgement about whether to include such papers or not, balancing the desire for a comprehensive review with maintaining applicability to the UK. Often I opted to ensure applicability by only considering results from UK studies where possible, or by excluding the paper where it was not possible to determine the applicability to the UK[1].

English language papers were included. Papers in other languages were excluded as it seemed highly unlikely any relevant articles would be published in other languages given the geographical area of interest.

The exposure of interest is resilience or resistance to the 'usual' detrimental effects of poor socio–economic position on health. That is, where respondents have better health than expected for their economic, social and environmental circumstances.

No comparison group is included in this study. 'Negative outliers'— such as the 'Glasgow effect' (Livingston and Lee, 2014)—are not included because of the increase in scope this would introduce. Furthermore it does not necessarily follow that the mechanisms creating worse health than expected would lead to any insight in to the mechanisms creating health resilience when reversed.

Outcomes of interest were anticipated to include, but not be limited to, measures of life expectancy, mortality, healthy life expectancy, morbidity, self–reported illness and limiting long-term illness, prevalence of disease, and incidence of disease.

Dates of coverage included empirical research published within the most recent five years, that is 2012–2017. Earlier research was excluded to capture the 'state–of–the–art' of empirical research on resilience. Useful concepts and measures developed prior to this period are assumed to be updated or used by literature created during this more recent period.

[1]For transparency I have documented all such decisions in `inst/systematic_review/2017-systmatic-review-results.csv`

I excluded identified titles that were not peer–reviewed articles, PhD theses, or systematic reviews themselves. I did not include 'grey literature', such as policy literature, in this systematic review. This was decided to help ensure the included titles met a minimum quality standard based on their peer–reviewed status.

During the course of the review it became apparent that a small number of studies had been identified by the search strategy that needed to be excluded because they were not relevant to the resilience of the general population. The general characteristic of these papers was: they studied a population with a specific, usually physical, usually chronic, pathology; and tested an intervention to improve this pathology. For example, Knott (2013) tested an intervention to help patients with Type–2 diabetes improve their self–management of their condition, while Blickem et al. (2013) test an intervention called 'BRIGHT' to help patients with stage 3 chronic kidney disease (CKD).

Deciding to exclude such papers was problematic but in general I excluded papers that focused on a physical pathology—such as diabetes or CKD—but left in mental or psychological pathologies—such as depression. An important component of resilience is psychological coping and adaptation (see Section 2.2; Schoon (2006)) so mental or psychological pathologies did not necessarily make a study irrelevant in the same way that management of a physical condition did.

As the nature of this review is to scope out the breadth of literature on the subject of health resilience I defaulted to including a paper unless it clearly did not benefit the review. Nevertheless, because of the subjectivity of this decision I solicited a second opinion from a colleague to maintain validity and reproducibility of the review. I asked if they would include or exclude the affected papers based on my inclusion criteria and relevance to the resilience of the general population. In all cases their second opinion matched my original decision[2].

[2] See `inst/systematic_review/2017-systematic-review-results.csv`.

3.4.2 Information sources

I used the ProQuest[3] and Web of Science[4] portals to conduct my search. Relevant social science and health science databases were selected, and others excluded, based on the taxonomies provided by the respective portal. ProQuest databases searched were: ProQuest Dissertations & Theses UK & Ireland; ProQuest Dissertations & Theses Abstract and Indexing (A&I); Physical Education Index; Applied Social Sciences Index & Abstracts (ASSIA); Education Database; Social Services Abstracts; Sociological Abstracts; and Worldwide Political Science Abstracts. The Web of Science search included the Social Science Citation Index (SSCI).

I excluded some social science or health science databases, for example the Education Resources Information Centre (ERIC), because they were based in a country other than the United Kingdom.

3.4.3 Search terms

To ensure comprehensive results were obtained I used the academic thesauri available through ProQuest[5] to obtain synonyms for 'resilience' and 'socio–economic position'. All relevant social science, medicine and general science thesauri were searched, while those not relevant—for example Aquatic Sciences & Fisheries Abstracts (ASFA) thesaurus—were excluded.

Only 'matched terms' from relevant databases were included. These were terms at the same level in the topic hierarchy. For example, when searching for resilience, 'personality' was found as a parent term but was too broad to be a synonym for resilience and was excluded.

Where there are differences in the British English and American English spelling of these terms, suitable adjustments were made to the search

[3] http://search.proquest.com

[4] http://webofknowledge.com

[5] http://search.proquest.com/thesaurus/browsepage/

syntax to account for these. For example, 'behaviour' was searched for as `|behavio*r|` taking account of the spelling both with and without a 'u'.

3.4.4 Search

Synonyms for resilience and socio–economic position were searched for in the title and abstract, increasing the likelihood of obtaining only relevant papers (the Web of Science portal uses a 'Topic' field to refer to both title and abstract). The search query was structured so resilience and its synonyms were searched for within three words of the word 'health', ensuring terms such as 'health resilience' or 'resilience to poor health' were both included. Finally, socio–economic position was added to the search query as an 'AND' term, such that both concepts were included.

I used the 'Advanced Search' or 'Command Line' search functionality to specify the inclusion and exclusion terms. Syntax and full inclusion criteria are included in `systematic_review/`. The ProQuest[6] and the Web of Science[7] searches can be replicated in full by users with appropriate log–in credentials.

Most articles were not tagged with a specific geography or country of study so no results were narrowed by location at this stage. Instead I opted to do this manually after obtaining the abstract or full–text as applicable.

3.4.5 Search results

Using the search strategy outlined above, 190 papers were identified by Web of Science and 116[8] were identified by ProQuest, or 306 papers in

[6] `https://goo.gl/vxoahW`

[7] `https://goo.gl/mUq41J`

[8] ProQuest erroneously reports this as, variously, 156, 128, or 122. I assume this is because the search algorithm returns an estimate of the number of results based on its index. I have based the final number of 116 on the number of records available for

total. Paper details were downloaded to a spreadsheet programme for further review.

In total 50 articles were identified in both the ProQuest search and the Web of Science search, suggesting a degree of reliability to the search criteria. After removing these duplicates 256 articles remained to be reviewed for inclusion.

3.4.6 Study Selection

Figure 3.1 summarises the process used to select relevant research papers. Titles were reviewed and papers excluded in cases where the study unambiguously did not meet the eligibility criteria, for example because the study was based on data obtained in a country other than England or Wales. Abstracts for all remaining articles were then obtained. I reviewed these and included or excluded these based on the eligibility criteria specified in Section 3.4.1. In cases where the abstract did not provide enough information, the full–text of the article was obtained for further scrutiny.

Of the original 256 papers, 99 papers were unambiguously not relevant based on their title and excluded. Abstracts were obtained for the remaining 157 papers and reviewed. 89 papers were removed after reviewing their abstract, leaving 68 papers for which the full–text was obtained. Of these 48 did not meet the eligibility criteria after reviewing the appropriate section—typically the methods section—of the paper. This left 20 papers which met the eligibility criteria for inclusion in this review.

I manually added one paper to the review in addition to those found by the systematic search strategy. These papers were found through the bibliographies of the included studies. In total 21 papers were included in this review.

---

export.

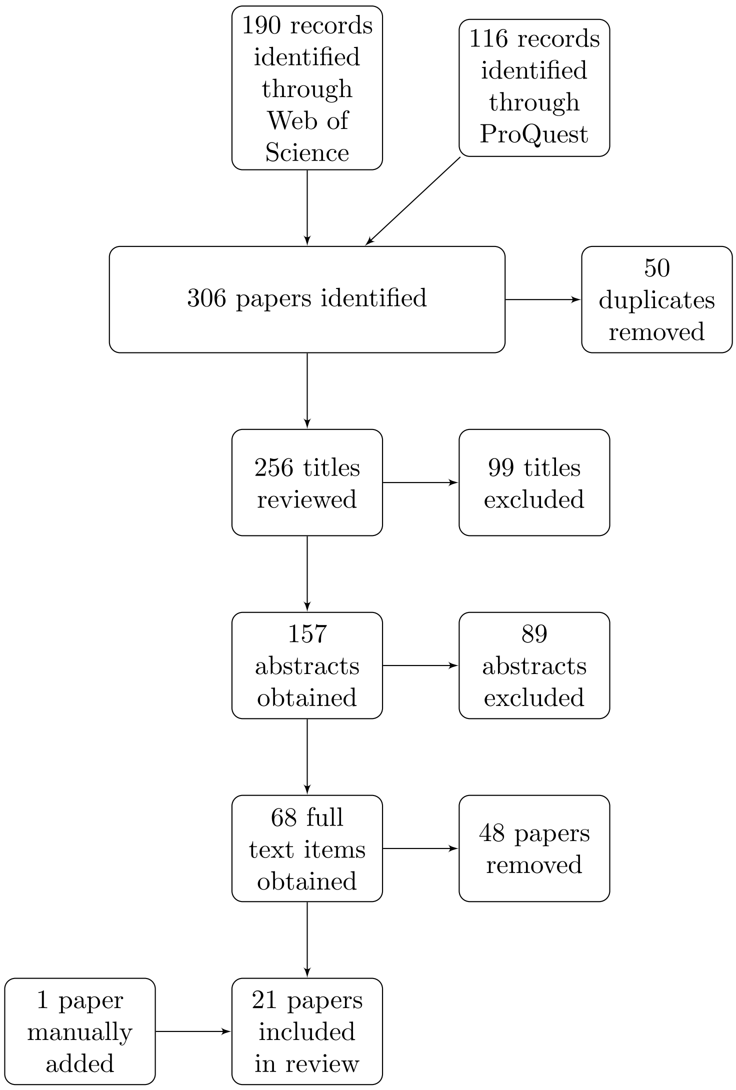

Figure 3.1: Review and selection of articles

3.4.7 Data collection process

The data collection instrument is found in `inst/systematic_review/`. For each paper the: population of study; method or methods used; 'outcome' or 'dependent' measure, usually a measure of (good) health; concept or measure of resilience; limitations; and key results and summary were extracted and recorded.

The population or sampling frame, methods, limitations, and key results and summary were generally explicit and could be extracted and summarised easily. The outcome measures and sources of resilience were not always so orderly, typically in papers that were only incidentally concerned with resilience. In these cases I have extracted all empirical variables used in the study and exercised my judgement as to how to best to record these.

3.5 Results

Here I present a 'narrative' or 'textual' analysis (Booth et al., 2012: 145–149) of all the studies collectively to identify areas of homogeneity or similarity. Statistical meta–analysis are not possible or necessary given the disparate nature of the measures used in each study.

Twenty–one papers were included in this review. Ten papers were predominantly quantitative (Albor et al., 2014; Bambra et al., 2015; Bellis et al., 2014; Erskine et al., 2016; Johnston et al., 2013; Mackenbach et al., 2015; Mõttus et al., 2012; Poortinga, 2012; Sull et al., 2015; Wel et al., 2015). Five were qualitative in nature (Cameron, 2013; Fenge et al., 2012; Haycock and Smith, 2014; Mastrocola et al., 2015; Matthews and Sykes, 2012). Four were mixed–methods studies (Cairns, 2013; Cairns–Nagi and Bambra, 2013; Reeves et al., 2014; Robinson et al., 2015). Two were systematic literature reviews (Glonti et al., 2015; Lai and Oei, 2014).

Three papers were longitudinal in nature, comparing data from at least two time points. These papers found that conditions at baseline were significantly associated with health outcomes (Erskine et al., 2016; Mõttus et al., 2012); or that parental (and even grandparental) conditions affected child outcomes (Johnston et al., 2013).

Many of the papers were not primarily concerned with health resilience, but this is not surprising given the goal of the search strategy was breadth rather than sensitivity or specificity (see Section 3.4.1). Nevertheless, all papers discussed some measure, concept, or intervention that could improve the health outcome of study even if this was not explicitly conceptualised as a source of resilience. For example Mõttus et al. (2012) establish a 'typical' relationship between neighbourhood deprivation (Scottish Indices of Multiple Deprivation (SIMD)) and quality of life as an outcome measure, and tested to see if cognitive ability mediates this relationship.

The papers covered a broad range of populations or used broad sampling frames. Of those that were empirical—that is, excluding the systematic reviews—two were international with a focus on Europe (Mackenbach et al., 2015; Wel et al., 2015), while the remainder sampled individuals from various areas of the United Kingdom. Of these, seven were broadly representative of at least England (Bellis et al., 2014; Johnston et al., 2013; Poortinga, 2012), or used census data (Bambra et al., 2015; Cairns, 2013; Cairns–Nagi and Bambra, 2013) and so were nationally representative.

The remaining papers sample from more specific, less representative, populations. Some studies sample from various geographical regions of the UK. Mõttus et al. (2012) sample from older people in Edinburgh, Haycock and Smith (2014) sample adults aged 30–35 years from north west England, Sull et al. (2015) recruit employees from an NHS Hospital Trust in northern England, and Fenge et al. (2012) sample older people from south England who are 'asset–rich, income–poor.' Other samples are from 'at–risk' populations. Matthews and Sykes (2012) recruit care leavers or Looked After Children (LAC), Robinson et al. (2015) sample unemployed men aged 45–60, and Mastrocola et al. (2015) carry out interviews with sex workers.

Given the disparate nature of the studies the same concept was sometimes considered a source of resilience by some studies, and sometimes an outcome by others. For example Mõttus et al. (2012) consider quality of life an outcome, which includes physical, psychological, social, and environmental domains of life quality. Cairns (2013) and Cairns–Nagi and Bambra (2013), on the other hand, hypothesise that the natural environment is a source of resilience. I suggest this is further evidence that the nature, and especially the causal nature, of health improvement and resilience is not well known. In general, however, outcome variables tended to be a measure of individual mental or physical health, not area or spatial health, so this is how I will operationalise my simulation in Chapter 6.

Outcome measures included, variously: self–reported general health (Bambra et al., 2015; Cairns, 2013; Cairns–Nagi and Bambra, 2013; Poortinga, 2012; Wel et al., 2015); quality of life or quality adjusted life–years (Mõttus et al., 2012; Reeves et al., 2014); psychological functioning or well–being (Matthews and Sykes, 2012); absence of risk–taking behaviours or participating in healthy behaviours (Bellis et al., 2014; Matthews and Sykes, 2012; Reeves et al., 2014); low premature mortality (Bambra et al., 2015; Cairns, 2013; Cairns–Nagi and Bambra, 2013); limiting long–term illness (LLTI) (Bambra et al., 2015; Cairns, 2013; Cairns–Nagi and Bambra, 2013); sports participation (Haycock and Smith, 2014); lower depression, anxiety, anger, or negative mood (Albor et al., 2014; Johnston et al., 2013; Lai and Oei, 2014, systematic review); low mortality (Mackenbach et al., 2015); mental well–being (Cameron, 2013; Erskine et al., 2016; Fenge et al., 2012; Reeves et al., 2014; Wel et al., 2015); effective management of chronic conditions (Mastrocola et al., 2015; Reeves et al., 2014); self–esteem (Albor et al., 2014); low social isolation (Fenge et al., 2012); and bespoke resilience scores (Robinson et al., 2015; Sull et al., 2015). Table 3.1 summarises the outcome measures.

Table 3.1: Outcome variables measured

Paper	Outcome	Variable(s)
13	Health	Self-reported general health
16	Quality of Life	WHOQOL-BREF
23	Motivation	Self-reported
	Self-esteem	Self-reported
	Non-smoker	Self-reported
32	Low premature mortality	Mortality <75
	Low morbidity	Self-reported general health
		Limiting long-term illness
37	Mental health	Self-reported
46	Low premature mortality	Mortality <75
	Low morbidity	Self-reported general health
		Limiting long-term illness
67	Sports participation	Self-reported
78	Lower depression/anxiety	Self-rated
		Hospital Anxiety and Depression Scale (HADS)
		Depression Anxiety Stress Scale (DASS-21)
		Questionnaire on Resources and Stress (QRS)
		Anxiety inventory
	Lower anger	Profile of Mood States (POMS)
	Lower negative mood	Profile of Mood States (POMS)
	Self-esteem	Rosenberg self-esteem scale
	Various	Various bespoke measures
90	Preventable mortality	All-cause mortality
96	Well-being	Self-reported
		WHO Wellbeing Index
		Self-reported general health
98	Well-being	Bespoke
	Self-efficacy	Bespoke
	Social support	Bespoke
173	Well-being	General Health Questionnaire (GHQ)
	Mental control	Thought Control Questionnaire (TCQ)
195	Management of conditions	Self-reported
204	Morbidity	Self-reported general health
		Limiting long-term illness
	Mortality	All-cause premature mortality
206	Resilience	25-item resilience scale
208	Physical health	Cardiovascular disease
		Respiratory illness
	Mental health	Depression
	Sleep quality	Sleep disturbance
241	Health self-management	Health Education Impact Questionnaire (HEIQ)
	Healthy behaviours	Summary of Diabetes Self-Care Activities Scale (SDSCA)
	Physical health	Short-form 12 (SF-12)
	Emotional well-being	Two items taken from ESS 2010
	Health economics	Quality Adjusted Life Years (QALYs)
242	Depression and anxiety	Self-reported
	Social Support	Two items from MCS
	Self-esteem	Abridged Rosenberg Self-esteem Scale
250	No health harming behaviours	Various behaviours included
272	Mental health	Malaise Inventory
		Strengths and Difficulties Questionnaire (SDQ-25)
307	Mental well-being	Self-reported
	Social well-being	Reduced social activities

Table 3.2: Sources of resilient characteristics

Paper	Resilience concept	Variable(s)
13	Bonding social capital	Neighbourhood cohesion
		Neighbourhood trust
		Neighbourhood belonging
		Civic participation
	Bridging social capital	Social cohesion
		Mutual respect
		Heterogeneous relationships
	Linking social capital	Political participation
		Political activism
		Political efficacy
		Political trust
16	Cognitive ability	Moray House Test no. 12
23	Support/encouragement	Self-reported
32	Place attachment	Self-reported
	Social capital	Self-reported
	Natural environment	Self-reported
37	Employment	Self-reported
	Finances/income	Self-reported
	Social isolation	Self-reported
	Occupational capital	Skills learned through occupation
	Social support	Self-reported
46	Place attachment	Self-reported
	Social capital	Self-reported
	Natural environment	Self-reported
67	Sport involvement in youth	Self-reported
78	Coping strategy	Ways of Coping scale (WOC)
		Multidimensional Coping Inventory (MCI)
		Brief Social Support Questionnaire
		Coping Orientations to Problems Experienced (COPE)
		Social Support Index
		Bespoke measures
90	Behaviour	Smoking, alcohol consumption, diet, exercise
96	State social support	Sickness benefit provision
98	Peer support	Self-reported
173	Repressive coping	Spielberger State-Trait Anxiety Inventory (STAI)
		Marlowe-Crowne Social Desirability Scale (MC)
195	Access to healthcare	Self-reported
204	Greater distance to brownfield	Previously Developed Land Index (PDL)
	Low environmental deprivation	MED-Ix database
206	Resilience	Resilience Scale (RS-25)
208	Gender	
	Age	Self-reported
	Education level	Self-reported
	Employment	Self-reported
	Income	Financial problems in last year
	Area of residence	Area-level deprivation
241	Network member characteristics	Number of 'nodes' within 5 minutes
		Percent nodes giving support within 5 minutes
		Number of frequent contacts (>1/week)
		Number of cohabitants
		Binary: network include spouse/partner
	Social network characteristics	Number of different relationship 'types'
		Number of network pairs who know each other
		Support given to others
		Social resources measure
		Involvement in groups or organisations
	Network change	Binary: network member lost in last 12 months
		Total network members lost in 12 months
242	Individual's similarity with area status	Education and occupation
250	Low Adverse Childhood Experiences (ACE)	Various, incl.: abuse, parental separation, etc.
272	No exposure to familial mental health	Parental and grandparental mental health
307	Budgeting skills	Self-reported
	Money management skills	Self-reported

Sources of resilience—or at least opportunities to improve the health outcome measure—included: social capital or social networks (Cairns, 2013; Cairns–Nagi and Bambra, 2013; Poortinga, 2012; Reeves et al., 2014; Robinson et al., 2015); cognitive ability in childhood (Mõttus et al., 2012); a mentor or someone to provide support (Matthews and Sykes, 2012); place attachment (Cairns, 2013; Cairns–Nagi and Bambra, 2013); natural environment (Cairns, 2013; Cairns–Nagi and Bambra, 2013); being in or returning to employment, income, or social class (Cameron, 2013; Glonti et al., 2015; Reeves et al., 2014); involvement in sports in childhood and youth (Haycock and Smith, 2014); 'problem–focused' coping (Lai and Oei, 2014, systematic review); behaviour change (Mackenbach et al., 2015); sickness benefit provision (Wel et al., 2015); repressive coping (avoidance) (Erskine et al., 2016); access to—especially primary—healthcare (Mastrocola et al., 2015); nearby–environment (Bambra et al., 2015); demographics such as gender, age, ethnicity, and education level (Glonti et al., 2015; Reeves et al., 2014); congruity between individual circumstances and neighbourhood or area circumstances (Albor et al., 2014); absence of Adverse Childhood Experiences (ACE) (Bellis et al., 2014); parental and grandparental mental health (Johnston et al., 2013); budgeting and money management skills (Fenge et al., 2012); and bespoke resilience scale (Sull et al., 2015). Table 3.2 summarises the sources of resilience.

In reviewing the sources of resilience I have found it useful to categorise these as 'internal' or 'external'. I consider internal sources means of resilience that the individual has direct control over, and is typically psychological or emotional in nature. I suggest cognitive ability, playing sports, and coping strategies could be considered 'internal' to the individual. External sources, on the other hand, are those which the individual does not have direct control over, although they can often exert some influence, such as in the choice of friendships or in choosing where to live. I do not suggest these as definitive distinctions, and what is within the locus of control will vary between individuals depending on their resources available. Instead I use these as a shorthand to summarise many of the shared characteristics of the disparate measures.

Many of the papers shared similar limitations. Papers using secondary data often did not have access to the 'ideal' variable for their chosen measure, and so relied on the measures available in the data set. These same papers typically used self–reported data, which is vulnerable to degrees of recall bias and respondent's willingness to share sensitive information. My own simulations share these issues, although I have selected my data set—*Understanding Society*—because it has a broad range of applicable variables.

The majority of the papers were not representative of the UK, or even of England, overall because of their sampling method. Nevertheless they offer valuable insight into how individuals manage and seek to improve their physical or mental health (or not). Most of the papers were also cross–sectional, so were unable to pick apart the causal nature of health and sources of resilience.

It should be said that not all of the included papers found evidence for resilience, and many found only limited evidence. This was compounded by the fact that some papers did not find evidence while others did, despite using similar or identical measures. For example Poortinga (2012) states, "... no support was found for the hypothesis that the different aspects [of social capital] help buffer against the detrimental influences of neighbourhood deprivation" (2012: 286). On the other hand, Cairns (2013) and Cairns–Nagi and Bambra (2013), and Robinson et al. (2015), found evidence that peer support *can* be supportive. While Poortinga's (2012) study is quantitative, the studies by Cairns (2013) and Cairns–Nagi and Bambra (2013), and Robinson et al. (2015) are qualitative in nature. It may be that they are able to explore nuances of social capital that are important at the local level that the quantitative instruments used by Poortinga (2012) cannot capture. Alternatively it may be that qualitative respondents view their own health outcomes favourably, these favourable views are not consistent when compared quantitatively with other respondents.

3.6 Discussion

In this scoping review of relevant, peer–reviewed literature I have identified the breadth of measures used to operationalise health outcomes and sources of resilience. From twenty–one papers of data for respondents from the United Kingdom in the period 2012–2017 I have identified over 53 health outcome measures and a similar number of measures thought to be sources of resilience. Many of these are similar or overlap, illustrating the ambiguity surrounding the currently very broadly defined concept of 'health resilience' and the need for more specific and well–defined definitions and measures.

To the best of my knowledge this study is the first to systematically and comprehensively identify the wide range of measures used to operationalise the concept of resilience. Similar systematic reviews of health resilience, such as Glonti et al. (2015), have been conducted, but these have tended to focus on demographic explanations for differing outcomes, rather than psychological sources of resilience that explain differences in outcomes *within* demographic groups. I hope that, in addition to my own use of these measures in Chapter 6, that researchers may be able to articulate health resilience in empirical and policy literature with greater confidence.

3.6.1 Limitations

The nature of a scoping review means that the search strategy used is unlikely to be exhaustive. Nevertheless it is comprehensive, providing a broad overview of the range of studies that operationalise resilience in some manner.

Statistical meta–analyses or quantitative combination of results is not possible for this review given the disparate nature of the measures used to operationalise resilience, health outcomes, and deprivation when applicable. Instead this review has produced a narrative summary of the included papers. Future, more narrowly focused, systematic reviews could potentially include meta–analyses of studies using consistent definitions and measures of health resilience.

The geographical restriction to the UK, and the related restriction to English–language papers, is arguably the most significant limitation of this review. While this decision was appropriate for the aims of this particular review, it will limit the generalisability of its findings for other countries with different socio–economic circumstances, cultures, and health and social care systems.

3.7 Conclusion

This scoping systematic review has identified the breadth of measures used by relevant literature to operationalise and study sources of resilience. I will use these measures in Chapter 6 to simulate resilient characteristics in my case study area of Doncaster, to investigate if and where health resilience if prevalent at the small–area level. I will also explore if these characteristics are associated with small areas that are resilience, as defined by lower than expected prevalence of clinical depression.

Chapter 4

Spatial microsimulation literature

4.1 Introduction

Spatial microsimulation is the technique that underpins the substantial empirical work in this thesis, and is what permits my analysis of resilience at the small–area level. The family of microsimulation, and later spatial microsimulation, techniques have a long history in the social sciences and there are numerous precedents of its use in the analysis of health care and health–related factors, as well as other domains.

Using this technique it is possible to perform spatial analyses that would otherwise be impossible at the small–area level. Researchers often have access to small–area aggregated data, such as the census, and to individual–level data available in surveys. The small–area data does not contain disaggregated information about individuals in that area, and surveys often use extremely coarse geographies, often region at best. Spatial microsimulation is a family of statistical techniques that estimate individuals at a small–area level by combining these two data sources (Ballas et al., 2005; Tanton and Edwards, 2013a).

Once simulated, some of the main uses of this data are to assess the likely effect proposed policy changes may have on the population and to perform GIS analyses, for example to assess service provision (see Section 4.3 later in this Chapter).

> Spatial microsimulation methods can be used to examine the changes resulting from public policy change in the lives of individuals within households at different geographical levels... (Campbell and Ballas, 2013: 285).

In some cases the data set produced by the spatial microsimulation is not comprehensive or ideal for the decision–making purpose. Nevertheless it is usually richer than the data available from other sources and so provides additional evidence for the purpose of policy planning that would be missing without the spatial microsimulation. For example, the rules surrounding the caps to housing benefit were complex and the data available to Campbell and Ballas (2013) were not sufficient to identify only individuals affected by these caps. Nevertheless they did have sufficient data to identify individuals who received housing benefit and to plot the proportion of recipients in each area. Even though it was not possible to identify only those affected by the cap, it was still useful to identify households that might be affected.

In this chapter I outline a brief history of microsimulation and spatial microsimulation. I outline a number of spatial microsimulation studies that have been developed in the health domain. I then move on to discuss some of the specific methods of spatial microsimulation that can be adopted, including the types of models that can be constructed, the reweighting techniques that can be applied, the process of creating a model, and validating the model once constructed. Throughout this chapter I outline my own choices for my analysis and my reasoning for choosing these specific methods.

4.2 History of microsimulation

Microsimulation is a technique used to estimate data about individuals when this data is not readily available. Individuals can be people, organisations, businesses, or any other discrete entity. Microsimulation has been used in the social sciences since at least the 1950s (Orcutt, 1960, 1962), so has a long history of application: "…it can be argued that microsimulation modelling methodologies have long become accepted tools in the evaluation of economic and social policy" (O'Donoghue et al., 2013: 3).

Spatial microsimulation is an extension of the microsimulation technique to include a spatial or geographical dimension. The goal of spatial microsimulation is frequently to "…simulate the *distributional* impact of different socio/economic policies or a change in those policies at the micro–level" (Ballas et al., 2013: 36, emphasis added).

Spatial microsimulation is a more recent refinement than microsimulation. The first uses of this technique can be traced back to the 1960s (see Ballas et al. (2013), p. 39) but the technique gradually became more common during the 1970s and 1980s as computers became more powerful and accessible (see, for example, Birkin and Clarke (1988)). Because of this precedent of microsimulation and spatial microsimulation, crucially, "…its behaviour is relatively well known" (Anderson, 2013: 53).

> There is now a growing body of evidence showing that the technique provides robust estimates of health–related variables in particular (Campbell and Ballas, 2016: 4).

4.3 Spatial microsimulation of health

Spatial microsimulation lends itself well to the analysis of health outcomes and policy:

> Health is an area which lends itself to spatial microsimulation techniques as there are many surveys but few comprehensive data bases in this field (Ballas et al., 2013: 41)

There are numerous examples of spatial microsimulation in the health domain, including a number of recent UK–based (Tomintz et al. (2008), Procter et al. (2008), Edwards and Clarke (2009), Edwards et al. (2010), Campbell (2011), Campbell and Ballas (2016)) and international (Morrissey et al. (2010), Morrissey et al. (2013)) models.

Tomintz et al. (2008) present a spatial microsimulation model of smoking prevalence at the small–area level in Leeds, UK. After identifying smoking prevalence at the small–area level, they use this information to assess the performance of local stop–smoking services and suggest where to place centres to best meet the needs of local residents by solving the *p*–median problem based on their simulated data (Tomintz et al., 2008: 348–351).

Procter et al. (2008), Edwards and Clarke (2009), and Edwards et al. (2010) describe a spatial microsimulation model of obesity at the small–area level in Leeds, called SimObesity. To construct their model they used primary care trust records of routinely collected data for children born since 1995, data from the '*Trends*' study, and data from the '*RADs*' study, combined with data from the 2001 census (Procter et al., 2008: 324). With their model they examined obesogenic environments in Leeds, identifying factors that were associated with small–areal–level obesity such as expenditure on food, number of household televisions, internet access, school meals, and level of physical activity (Procter et al., 2008: 330), social capital and poverty (Edwards and Clarke, 2009). Crucially they found that key factors associated with childhood obesity varied across different wards selected for case study, so different interventions to reduce childhood obesity might be required depending on the needs of the local area (Edwards et al., 2010: 13). They also find that self–reported, or 'subjective', measures of the safety of the neighbourhood and the quality of transport are more strongly associated with obesity than 'objective' measures of actual crime or transport quality. This suggests perception is at least as influential in an individual's activity and consumption decisions (Procter et al., 2008: 336).

Campbell (2011) and Campbell and Ballas (2016) present a spatial microsimulation of model of health behaviours and outcomes in Scotland called SimAlba. It was built using data from the Scottish Health Survey 2003 and the 2001 census for Scotland at the output area level. Using the model Campbell (2011) assess well–being and happiness, smoking, alcohol consumption, and obesity at the small–area level.

Morrissey et al. (2010) is a pioneering spatial microsimulation model of depression in Ireland. At the time of publication, the authors reported there was no research on the accessibility of mental health services for individuals with depression in Ireland, nor any data on the small–area incidence of depression (2010: 11). They developed their model, the Simulation Model of the Irish Local Economy (SMILE), using small–area population statistics and the Living In Ireland (LII) survey. Using their model they were able to identify small–area levels of depression, and indicate areas that had high demand but poor access (2010: 23).

Morrissey et al. (2013) extended the SMILE model to include long term illness (LTI). They compared the spatial distribution of individuals with long term illnesses with the accessibility of acute hospitals. They found higher rates of LTI overall in the west of Ireland but that areas with high LTI tended to have lower access scores to acute hospitals (2013: 225–227).

4.4 Static and dynamic models

There are two broad categories of spatial microsimulation, static and dynamic (Ballas et al., 2013; Mertz, 1991). Static spatial microsimulation, as the name suggests, does not update the population after initial simulation. Static spatial microsimulation models have been used in the health domain to model many health–related behaviours and outcomes.

Examples of static models include Tomintz et al. (2008), Procter et al. (2008) and Edwards and Clarke (2009); Morrissey et al. (2010).

It is also possible to 'age' the individuals in the spatial microsimulation. Mertz (1991) argues that the units in a static model can be 'aged'—'static with aging'—and that this technique can be useful for modeling short–term population changes if the underlying demographics are unlikely to change much (1991: 81).

Fully dynamic models incorporate population changes such as fertility, mortality, and migration to project the population into the future, as well as policy changes. As such they are useful for exploring 'life–course' impacts of social and environmental policies and change (Ballas et al., 2013: 38). Both of these techniques can provide an insight into the temporal movement and fluidity of the population.

The advantage of a static (spatial) microsimulation model is that it is less 'expensive' to produce (Mertz, 1991: 84), while still allowing the researcher to evaluate the effects of policy on individuals and areas. "Static model simulations allow the researcher to vary policy rules and produce estimates of gains or losses for an individual or household resulting from the policy change... and to examine the distributional impacts of policy change" (Ballas et al., 2013: 37).

I have opted to create a static spatial microsimulation model for Doncaster, as I am interested in distributional effects of policy change on residents as a snapshot. Resilience research is still relatively preliminary, and as such I felt it was more important to understand a snapshot of resilience than attempt to model life events and their effects on resilience without a more solid understanding of the basis of resilience.

4.5 Deterministic and probabilistic models

There are a number of methods for constructing a static spatial microsimulation model. Adjustment can be probabilistic where individuals are sampled from microdata to match aggregated small–area totals. Combinatorial optimisation is a probabilistic method which makes use of various algorithms such as hill–climbing, simulated annealing, and genetic algorithms (Williamson et al., 1998). One of the few recent examples of a probabilistic spatial microsimulation model in the health domain is Morrissey et al. (2010), which used a simulated annealing (SA) method (2010: 11).

Deterministic reweighting is instead a process in which individuals are reweighted to reflect the aggregate population (Ballas et al., 2005b: 9). Arguably the first study to distinguish between deterministic and probabilistic simulation approaches is Ballas et al. (2005). In the simulation of SimBritain—and its pilot, SimYork—the authors used a deterministic reweighting algorithm to produce small–area estimates using 1991 Census Small Area Statistics data and the British Household Panel Survey (BHPS). These formed input data to a dynamic model to estimate the population of Britain to 2021 (2005: 19).

Because probabilistic approaches make use of random selection and optimisation, these approaches will produce a different outcome each time they are run. This makes it more difficult to know if a change to the results is due to a change in initial conditions or just an artefact of the probabilities. The advantage of a deterministic approach is that, all things being equal, the same results will be generated each time the simulation is run (Procter et al., 2008: 325). Error tracking is much more straightforward if it is known in advance that any changes to the model can only be the result of changes to the inputs:

> This determinism meant that variations in input data coding, constraint ordering or small-area table recoding were the

> only source of variation in the small-area estimates. This proved extremely useful because it allowed the testing of different combinations of constraints and data coding options without the additional uncertainty caused by a probabilistic reweighting method (Anderson, 2013: 64).

Because of this key advantage I opted to use a deterministic reweighting algorithm over a probabilistic approach. Examples of deterministic simulations include: Ballas et al. (2005); Tomintz et al. (2008); Procter et al. (2008), Edwards and Clarke (2009), and Edwards et al. (2010); and Campbell (2011). Deterministic reweighting has been demonstrated to accurately predict health behaviours and outcomes (Smith et al., 2011: 623).

4.6 Iterative Proportional Fitting (IPF)

Iterative proportional fitting (IPF) is an example of a deterministic reweighting algorithm (Birkin and Clarke, 1988, 1989; Johnston and Pattie, 1993), and is the method I use in this simulation. IPF reweights each individual from the survey in each zone, using constraints in both the census and the survey to determine the resultant weights. The formula used the calculate the resultant weights is given by:

$$n_i = w_i(\frac{c_{ij}}{m_{ij}}) \tag{4.1}$$

where n_i is the new weight to be calculated for individual i, w_i is the previous weight (initially set to 1), c_{ij} is index $_{ij}$ in the constraint (census) table, and m_{ij} is index $_{ij}$ of the survey microdata table (adapted from Ballas et al. (2007), p. 51–52)).

Table 4.1: Example constraints

zone	age_0_49	age_gt_50	sex_f	sex_m
a	8	4	6	6
b	2	8	6	4
c	7	4	8	3

4.6.1 Initial weights

With deterministic spatial microsimulation, and therefore IPF, it is necessary to set an initial weight from which the algorithm reweights individuals. There are two main options for setting these initial weights. One is to use, or calculate, a weight for each individual so that the survey is representative, typically at the regional or national level. For example, a survey with too few respondents aged under 25 might weight those individuals so that they are used more frequently to make the overall survey representative of the population. The other option is to simply set all initial weights to 1 (Ballas et al. (2005)).

As IPF weights each individual so that the reweighted data becomes representative of the area it is, by definition, creating representative weights. Furthermore it has been proven that the IPF algorithm converges to a single result as long as there are no empty cells (Fienberg, 1970), so the initial weight will not affect the final weight provided the reweighting algorithm is reiterated over a sufficient number of times.

4.6.2 IPF: a simple example

To illustrate the procedure the following is a simple example using two constraints to simulate income, which has been adapted from Lovelace and Dumont (2016). Table 4.1 shows the constraints which shows the total number of individuals in each zone by demographic, while Table 4.2 shows the individual–level data which represents a sample of individuals such as that taken in a survey.

For example, eight individuals live in zone a who are aged 49 and below,

Table 4.2: Example individual-level data

id	age	sex	income	initial.weight
A	age_gt_50	sex_m	2868	1
B	age_gt_50	sex_m	2474	1
C	age_0_49	sex_m	2231	1
D	age_gt_50	sex_f	3152	1
E	age_0_49	sex_f	2473	1

Table 4.3: Aggregated individual-level data

Age.sex	Male	Female	Total
Under 50	1	1	2
50 and over	2	1	3
Total	3	2	5

and four individuals live in zone a who are aged 50 and above, making a total of twelve individuals living in zone a. This can be confirmed by observing that there are six females and six males, again making a total of 12. It should be noted at this point we know there are 12 individuals in the population, and we know how these individuals are distributed by age or by sex, but not by both age and sex simultaneously. The individual level data contains five individuals for whom we know their age, sex, and income, but not where they live.

Discarding income from the individual–level data set for now, aggregating the individuals produces the marginal totals in Table 4.3. This represents the same data as that in Table 4.2, simply restructured.

For zone a, we know that we have eight individuals aged 49 or less from Table 4.1. This is c_{ij} in Equation (4.1). Individuals C and E are aged 49 or less, so the margin total is 2. This is m_{ij} in Equation (4.1). Taking w_i as 1.0 initially and inserting these values into Equation (4.1), we thus have a new weight for individuals C and E of $\frac{8}{2}$, or 4. Similarly, for individuals aged 50 and above the new weight, n_i, is $\frac{4}{3}$, or approximately 1.33. Table 4.4 shows the individual–level data with the new weights applied. We can confirm that these are correct by summing the weights and observing that this total (12) matches the total population from the census (Table 4.1).

Table 4.4: Individuals with new weights

id	age	sex	new.weight
A	age_gt_50	sex_m	1.3
B	age_gt_50	sex_m	1.3
C	age_0_49	sex_m	4.0
D	age_gt_50	sex_f	1.3
E	age_0_49	sex_f	4.0

Table 4.5: Individual marginal weights after re-aggregating

Age.sex	Male	Female	Total
Under 50	4.0	4.0	8
50 and over	2.7	1.3	4
Total	6.7	5.3	12

To calculate the next weights based on sex, the individual–level data must first be re–aggregated, taking into account the weights calculated above. For example, the weight calculated for males aged 50 or over is $\frac{4}{3}$, taken from Table 4.4. To re–aggregate, this weight is multiplied by the number of individuals who fit this criteria. As there are two individuals who meet this criteria (individuals A and B), this weight is multiplied by 2 to give $\frac{4}{3} \times 2 = \frac{8}{3}$, or approximately 2.67. Re–aggregating all margin total produces the values in Table 4.5, which we can again verify sums to 12.

These re–aggregated margin totals can then be used to produce the next iteration of weights using the sex constraint. From Table 4.1 we know we have six males and six females in zone a, and these are now the values for c_{ij} in Equation (4.1). m_{ij} is given by the new marginal totals in Table 4.5, so is 6.67 for males and 5.33 for females. These fractions are now multiplied by the weights in Table 4.4 (in the previous iteration the fractions were effectively multiplied by 1, the initial weight). For individual A in zone a, the new weight is therefore calculated by $1.33 \times \frac{6}{6.67} = 1.2$, while the new weight for individual E is $4 \times \frac{6}{5.33} = 4.5$. Table 4.6 shows the final weights for all individuals in zone a, which can again be verified by ensuring the sum of the weights matches the initial population of 12.

Table 4.6: Individuals with weights including sex constraint

id	age	sex	new.weight
A	age_gt_50	sex_m	1.2
B	age_gt_50	sex_m	1.2
C	age_0_49	sex_m	3.6
D	age_gt_50	sex_f	1.5
E	age_0_49	sex_f	4.5

At this point producing the final weights is a matter of iterating sufficient times for the weights to converge, and repeating this procedure for all zones.

4.6.3 Integerisation

Deterministic reweighting, and hence IPF, produces fractional weights for each area, unlike probabilistic approaches which return randomly selected cases. These fractional weights can be used, and are more precise, but some applications require integer cases where fractional weights are problematic, such as for dynamic models (see Williamson et al. (1998)). Integer weights are also more intuitive to use, as each case represents a (simulated) individual, which is a unit of measurement that most social scientists are familiar with.

There are a number of methods available for integerisation, including simply rounding the weights, using a threshold, using a counter–weight, using proportional probabilities, and the more modern 'truncate, replicate, sample' method (Lovelace, 2013: 4–5).

Rounding, using a threshold, or using a counter–weight would not be useful to my simulation. These will drop individuals with a weight below a certain threshold (0.5, for example with rounding). Because I have a relatively large sample size with *Understanding Society*, the fractional weights that are generated are typically very small, and certainly less than 0.5 in many cases. If I were to use one of these approaches many of my cases would be dropped by the integerisation procedure, dramatically reducing the diversity of the model (and negate the point of maximising this diversity).

Proportional probabilities treats the resulting fractional weight as a probability. As the method samples with replication, cases with higher weights are more likely to be selected than those with smaller weights, which is the intended behaviour. Nevertheless there is a small possibility that cases with smaller weights will be selected more times than those with higher weights, owing to the random nature of the selection process, which is not ideal.

The truncate, replicate, sample method tries to address the problems associated with these other methods, and provide integerised results that are nearly as accurate as the fractional weights. It achieves this by using the fractional weight in two ways. First, any weights greater than 1 indicate the case should be *replicated*. For example, a weight of 9.2 would mean the case should be replicated nine times in the area. The remaining fractional weight is separated ('truncated') and used as a more traditional probability for sampling (Lovelace, 2013: 6). By separating these steps Lovelace (2013) is able to produce results that fit known data more accurately than the other integerisation techniques (Lovelace, 2013: 9).

For my simulation I opted to retain fractional weights for the analysis of single simulated variables such as clinical depression (see Chapter 6), but to integerise when using multiple simulated variables. The integerised weights were highly consistent with the fractional weights (see Section 5.12), but are more intuitive to work with given that they reflect the structure of a typical data set arranged with one row per case.

For the integerisation I used the truncate, replicate, sample method given its ability to return integerised weights that are nearly as accurate as the fractional weights. In practice, many of the fractional weights returned for this model were less than 1, so there was little difference between this method and the proportional probabilities approach for many cases. Nevertheless, I did have cases with a fractional weight greater than 1, so the truncate, replicate, sample method was still likely to be more accurate, without any disadvantages over other methods.

4.7 Constraint configuration

To perform either probabilistic or deterministic spatial microsimulation, it is necessary to use constraint variables. In the example given in Section 4.6.2, age and sex were the constraints. The constraints should have a conceptual and statistical relationship to the target variable or variables and should be informed by theory, or an empirical test, or both (Anderson, 2013; Smith et al., 2009). In the case of health conditions, the constraint variables should help to predict the presence of health conditions using standard regression techniques (Edwards and Clarke, 2009: 1129; Smith et al., 2009: 1253).

Selecting appropriate constraint variables requires finding a balance between having too many constraints and too few (Tanton and Edwards, 2013b: 163). Too many constraints increases the chance the simulation will not converge, and too few increases the chance the simulation will produce results that are not valid. As with regression, a parsimonious selection of constraint variables should be made.

Ultimately the choice of constraints can only be made from variables that are present in both data sources: "…the choice of the constraints, though informed by the literature and other empirical research, must be pragmatic" (Campbell and Ballas, 2016: 3).

I made an intital selection of constraints based on a theoretical understanding of the social determinants of health (see Section 2.7), and tested these empirically using appropriate regression techniques (see Sections 5.5 and 6.3).

Once selected, some authors argue the order the constraints are input into the model can, in some circumstances, affect the outcome of the simulation:

> "The order of constraints is important as the first variable is reproduced most accurately" (Tomintz et al., 2008: 344).

Some suggest the relative contribution of the constraint to the model should determine the order (Smith et al., 2009: 1252), with the most 'powerful' constraint entered last:

> ...because the IPF technique iteratively reweights a series of constraints, the last constraint is necessarily fitted perfectly. It is therefore important that the constraints are used in an order that represents their increasing predictive power so that the 'best' constraint is fitted last... (Anderson, 2013: 56).

Lovelace et al. (2015) argue that the constraint with the most categories should be entered first, as placing at the end of the process detrimentally affected the model (Lovelace et al., 2015: 6.7). In Sections 5.6 and 6.4.1 I test the order of constaints, comparing a random order with an order specified by the β values.

4.8 Validation

Validating the results of a spatial microsimulation can be challenging (Smith et al., 2011), because to comprehensively validate the results would require data that is not available (otherwise it would not be necessary to perform the simulation!). Nevertheless a number of approaches are available which can broadly be described as either internal validation or external validation (Ballas et al., 2005b, 2005; Edwards et al., 2011).

Internal validation is used to assess how well the simulated constraint data match the 'real' constraint data. This process uses data that is 'internal' to the simulation to provide an indication of how well the reweighting algorithm has matched survey respondents to the aggregate counts in each area. This does not provide a direct assessment of the accuracy of the simulated target variable, but it does at least give an indication if the 'right' individuals are being simulated in each area.

Internal validation methods include: correlation of simulated weights against constraints; two–sided, equal variance t–test to test if simulated values are statistically from the same distribution as the census values; and total absolute error (`TAE`) and standardised absolute error (`SAE`), both for the whole area simulated and for each zone.

Correlation is used to assess how accurately the simulated population in each zone match the known population for that zone, and as such is a general indicator of model fit.

A two–sided, equal variable t–test is used to determine if the simulated variables differ statistically from the constraint variables. If the result of the t–test is not statistically significant we accept the null hypothesis and determine that the distributions are not different and the simulated variable is the same as the observed variable.

Total absolute error give an indication of the level of deviation between the simulated data and the known data (Williamson et al., 1998). It is calculated by:

$$\text{TAE} = \sum_{ij} |U_{ij} - T_{ij}| \tag{4.2}$$

where U_{ij} is the observed count of area i in category j, and T_{ij} is the simulated count for the same area and category.

While there is no fixed level of acceptable error, Smith et al. (2009) advise that "...error thresholds need to be chosen on the basis of the intended usage of the model" (2009: 1256).

From total absolute error, the standardised absolute error can be calculated, which is arguably the simplest method of internal validation to interpret because it provides one standardised measure to assess that is comparable across models. SAE is constructed from the total absolute error (TAE) divided by the population for each area (Smith et al., 2009).

Smith et al. (2009) suggest that for estimating the incidence of rare events—as health resilience is—the model should have an SAE of less than 0.1 (10%) in 90% of simulated areas for the constraints, and an SAE of less than 0.2 (20%) in 90% of simulated areas for unconstrained variables (2009: 1256).

The internal validation results do not tell us about the fit of the simulation variable to real target data, however. Assessing the accuracy of the simulated target data is challenging and relatively few spatial microsimulation models are able to include such validation. There are two methods which can provide confidence in the accuracy of the simulation.

One way to test this is to aggregate the simulated target variable up to a larger geography and compare it to comparable external agreggated data. For example, if an external data source can be found with the target data provided at a regional geography, this can be compared to the simulated target variable if this is aggregated up to the region. Smith et al. (2011) take this approach to determine if a simulation of individuals in New Zealand provides reasonable estimates of small–area smoking prevalence.

Similarly, it is possible to compare to a correlated variable (Edwards and Tanton, 2013). This is one reason why I simulated limiting long–term illness or disability as a 'pilot'; to provide an indication of how the constraints are simulating with the *Understanding Society* dataset (see Chapter 5).

I use both approaches in this model which is relatively uncommon (Smith et al., 2011: 618) in the spatial microsimulation literature, and provides confidence in the accuracy of the model, something that was essential as I was simulating relatively rare events. The pilot simulation indicates how well the simulation estimates a correlated target variable (see Section 5.11). I was also able to obtain known data on clinical depression aggregated to Doncaster overall, so I aggregated the results of my simulation and compared this against the known data (see Section 6.6.4).

4.9 Limitations

Spatial microsimulation can be very accurate when producing small–area estimations of target variables. It is, nevertheless, an approximation and simulated counts can, and do, differ compared to known counts. The purpose of validating the model is to minimise these errors and ensure the simulation is accurate enough for its purpose (Section 4.8).

The issue of error can be especially important for small areas with 'extreme' population demographics, for example areas that are exceptionally wealthy or exceptionally poor (Smith et al., 2009: 1253). If an individual or household has a characteristic that is extreme, there may not be any individuals in the survey that can accurately represent them. If this is only one or two individuals in an area it might not be a problem depending on the application of the model, but if such extremes are more typical of an area the simulation will not be able to model them from the pool of available individuals in the survey.

Doncaster is diverse in many socio–demographic dimensions, but no area of Doncaster is typified by extremes of characteristics. It does, however, have a large prisoner population across four prisons and youth offender institutions which cannot be characterised by respondents in *Understanding Society*. I discuss this issue, and how I mitigated it, in Section 5.3.2.

4.10 Conclusion

Spatial microsimulation is a geographical technique to estimate individual–level data at the small–area level where this information does not exist. It is common to have access to small–area aggregated data, such as the census, and disaggregated, individual–level data but without information at the small–area level. Spatial microsimulation allows us to combine the two data sets to create a synthetic estimate of this data, typically for policy analysis.

Spatial microsimulation has a history in the social sciences stretching back at least thirty years, so there is a clear precedent of its applicability. Spatial microsimulation models of health behaviours and outcomes are no exception, and there are examples of its use in the health domain both domestically and internationally.

In all cases the spatially microsimulated data sets are an estimation of 'real' data, but a number of studies attest to the accuracy of the simulated results, and therefore their real–world utility for policy analysis (Ballas et al., 2005b; Edwards and Clarke, 2009; Morrissey et al., 2013; Procter et al., 2008; Tomintz et al., 2008).

There have been a number of methods developed to validate the accuracy of a spatially microsimulated data set, including t–tests and absolute error, which I have incorporated into my model.

It is therefore entirely possible to create a robust, accurate spatial microsimulation model of the target variable of interest to the researcher. Using the techniques described in this chapter I produce a robust spatial microsimulation model of health resilience in Doncaster in the following chapters. In Chapter 5 I create a pilot model to develop and test code to produce a spatial microsimulation model, and to externally validate my model against a correlated variable. In Chapter 6 I extend this model to simulate health resilience which I validate and test to ensure it is robust, using the techniques outlined in this chapter.

Chapter 5

Data and Methods

I have measured out my life with coffee spoons.

T. S. Eliot

5.1 Introduction

This chapter descibes the data and methods used to produce a spatially microsimulated data set. Broadly the steps involved in producing the simulated data were: identify suitable data sets; prepare these data sets; select appropriate constraint and target variables; create the simulation; and validate the simulation.

Because producing a spatially microsimulated dataset requires both a lot of time and computational power—both to obtain and prepare the data properly and to perform the actual reweighting—I chose to produce a minimal 'pilot' spatial microsimulation model before completing the full model. By producing a minimal model I was able to quickly produce a reweighted dataset which offered me a number of advantages.

First, by producing a minimal simulation I was able to quickly develop and improve the code base used to produce the reweighted data, as iterations can be faster when working with smaller data. This helped me to improve my own understanding of the process and to develop code known to produce accurate simulations. Between completing the pilot simulation and producing the final simulation model I developed this template into a CRAN (Comprehensive R Archive Network)[1] package called `rakeR`.[2] This eliminated many of the manual steps involved in microsimulation, reducing the possibility of human error creeping in to the process.

Second, by simulating a target variable that is also in the census I was able to externally validate the quality of the simulation. Externally validating the results of a spatial microsimulation model is challenging—if it was easy, we would not need to produce the simulation in the first place!—so this is a good way to test the performance of the model. This is not to say that the accuracy of the final target variables will be the same as the pilot, but it does identify any errors introduced by the code and does give an indication that the simulation will converge.

Third, being able to compare simuated data with known values made it possible to test how the size of the target zones affected the quality of the simulation. Small areas in the UK are typically considered to include, in increasing size, output areas, lower layer super output areas (LSOAs), middle layer super output areas (MSOAs), and sometimes wards. Depending on the size of target area the simulation performs differently, so I was able to assess the quality of simulations with different zone sizes by comparing known values with simulated values. I was also able to compare the quality of results obtained by just reweighting (fractional weights) against those obtained by reweighting and integerising (integerised weights).

[1]CRAN is analguous to the standard library in many other computer languages such as `C++` or `Python`.

[2]`https://cran.r-project.org/package=rakeR`

Table 5.1: Example survey data

id	age	sex	qualification	long-term illness
22445	age_25_29	sex_female	qual_3	llid_no
29925	age_30_44	sex_female	qual_4_plus	llid_yes
280165	age_30_44	sex_female	qual_2	llid_no
333205	age_20_24	sex_female	qual_2	llid_no
387605	age_25_29	sex_female	qual_3	llid_no

5.2 Data

Deterministic IPF requires two input data sources. First, it requires data about individuals. This is typically taken from survey data in which each row is a case or individual. The geographical origin of each case or individual is not known at the small area level in this data set; this is what we will be simulating. This data is commonly referred to as the individual–level data, microdata, or the survey data. Table 5.1 illustrates what typical survey data looks like. Each row is an individual, and each column is a variable.

Second, it requires aggregate data about small areas. This is typically taken from the census, in which each row is a small area and each column is an aggregated count. This aggregate data is used to 'constrain' the weighting so that when the simulated individual–level data is aggregated it matches the actual aggregated data as closely as possible (Cassells et al., 2013). This data is commonly referred to as the aggregate data, constraint data, or even census data, reflecting the fact that the census is the most common source of area–level aggregate data. Table 5.2 illustrates what typical constraint or census data looks like. Each row is a small area (in the case of this example, a sample of output area codes in Doncaster), and each column is an aggregate count of a variable.

I will use the terms 'survey data' and 'census data' to refer to the two data sources respectively. This reflects the sources of data I use rather than more abstract terminology.

Table 5.2: Example census data

code	female	male	no cars	1 car	2 or more cars
E00037829	117	101	53	92	73
E00037830	125	131	45	116	95
E00037831	132	128	51	131	78
E00037832	122	128	44	105	101
E00037833	127	125	115	116	21

5.2.1 Survey data

I required a survey data source to contain a variety of health outcome measures as these would become target variables. The survey data also needed to include a number of socio–economic and demographic factors which match those in the census as these would form the basis of the simulation constraints. Examples of these include age, sex, ethnicity, education or qualifications, socio–economic position, and housing tenure.

The survey sampling frame should include respondents from at least England and Wales. Wider sampling frames, such as Great Britain or the United Kingdom, were considered. Sources sampling only one country, for example just Wales or Scotland, were not. This is necessary so that the sample in the survey matches the population from the census (see Section 5.3).

The survey does not necessarily need to be representative, for example by being based on an entirely random sample. This is because the reweighting algorithm matches individuals to areas based on known data, in effect producing its own survey weights. This allowed me to include surveys that were comprehensive but not necessarily representative without the need to apply corrective weights (see, for example, Lumley (2010), chap. 7).

The survey data should contain as large a sample as possible to maintain the diversity of respondents. Health resilient individuals, as we have seen in Chapter 2, are outliers by definition, and the proportion of individuals considered resilient ranged from 20% to 2.5% (1.96 standard deviations). As I am effectively attempting to understand between 2.5–20% of the original sample size, a large initial sample leaves me with a reasonable sample size to investigate.

I considered a number of data sources, many available from the UK Data Service's *health and health behaviour* data discovery service[3]. I excluded aggregate data sets and data sets only collected in one country of the United Kingdom. This left a number of potential data sets to use, including: the 1946, 1958, 1970, and 2000–01 birth cohort studies; Health and Lifestyle Survey (HALS); health surveys for England, Wales, Scotland, and Northern Ireland; Infant Feeding Survey; Life Opportunities Survey; Living Costs and Food Survey; National Diet and Nurtition Surveys (NDNS); Opinions and Lifestyle Survey; Surveys of Psychiatric Morbidity; UK Time Use Survey; and Understanding Society.

Many of the surveys available would have met my criteria and would have been suitable for use, and subsequent researchers may wish to attempt to replicate the spatial microsimulation using different data sets. Ultimately I chose *Understanding Society* because it fulfilled my criteria for health, socio–economic, and demographic variables, and had the largest sampling frame with 40,000 households sampled in the first wave (UK Data Service, 2016).

Understanding Society had the added benefit that it is relatively recent and waves were contemporaneous with the 2011 census data, making it appropriate to use with the most recent census data and related geographical boundaries. The years of data collection do not match exactly, but this is a common problem in spatial microsimulation and a solution necessarily involves a 'pragmatic compromise' (Campbell and Ballas, 2016: 3).

I used cases from Understanding Society waves `a_` through `f_` (University of Essex et al., 2016). The file serial number for *Understanding Society* is 6614. They were obtained from the UK Data Service under usage number 75413. Data for these waves were collected between the years 2009–2015.

[3] `https://www.ukdataservice.ac.uk/get-data/themes/health`

Table 5.3: Years of data collection by wave

Wave	y2009	y2010	y2011	y2012	y2013	y2014	y2015
e					2013	2014	2015
d				2012	2013	2014	
c			2011	2012	2013		
b		2010	2011	2012			
a	2009	2010	2011				

Table 5.3 shows the years each wave of data were collected.

`data-raw/0-prep-understanding-society.R` (see Section 1.3) prepares a 'tidy' (Wickham, 2014) dataset and saves it for use in the simulation. The aim of the processes was to obtain the most recent data for each individual in Understanding Society and treat this collated data as one large cross–section. The longitudinal nature of the data is lost, but this is not the focus of this research and the amalgamation of waves effectively increases the diversity from which to choose individuals for the area. This approach can reduce error size (Edwards and Clarke, 2009: 1133) which is beneficial in this case. Subsequent research may wish to consider the longitudinal nature of health resilience, and Understanding Society may be a useful data set for this. The cross–section was constructed to include the most recent information known about each individual.

Individual responses from each wave were loaded in turn. These files are named `x_indresp`, where `x_` is a one–character prefix identifying each wave individually. For example the most recent wave is identified by the prefix `f_`. Variables within each wave were prefixed with the appropriate wave identifier. For example, age at wave `f_` might be labelled `f_age`, at wave `e_` would be labelled `e_age`, and so on. These individual files were joined using `FULL` joins on person ID (`pidp`) which are unique to each respondent and consistent across waves. A `FULL` join—equivalent to a SQL `FULL OUTER JOIN`—ensures cases are not dropped if the respondent is not present in all waves. Household IDs are not consistent across waves, so cannot be used for this join[4].

[4]`https://www.understandingsociety.ac.uk/support/issues/481#change-1638`

Household responses (`x_hhresp`) were loaded and joined to each individual using the appropriate household identifier from each wave (`x_hidp`). For example, individuals in `f_indresp` are joined to their household data in `f_hhresp` using `f_hidp`, present in both files. `LEFT` joins are used for this purpose.

Answers coded as 'missing', 'proxy', 'inapplicable', 'refusal', or 'don't know' in Understanding Society were recoded as an explicit missing value (`NA` in `R`). 'Refusal' and 'don't know' are not technically a missing value as they tell us something about the respondent, but they have been coded as such and removed because such information is not required in this instance. Missing values could then be removed prior to analysis if necessary.

Not all variables were measured in all waves. Because of the longitudinal nature of the dataset many variables, such as ethnicity, did not need to be repeated and were typically present in the first wave the respondent participated in. This was handled by taking the most recent known data for each individual in building the cross–section of data to use. For example, if individuals had ethnicity data missing in wave `f_` I looked back as far as necessary to obtain this information. In this way I was able to create a complete cross–section with each respondent's most recent information, which was then validated and checked before use.

After completing the final data frame $81,540$ individuals were eligible for inclusion in my analysis. Using estimates for health resilient individuals of between 2.5–20%, this equates to between approximately $2,038$ and $16,308$ individuals.

5.2.2 Census data

The 2011 census tables (Nomis, 2013a; Office for National Statistics et al., 2017) were used as constaints. The data contained in the census is now several years old. For example economically active unemployment in

Table 5.4: Census variables and table IDs

ID	Variable
KS102EW	Age structure
KS601UK	Economic activity
...	...

Doncaster has risen from 5.5% as identified in the census in 2011 (Nomis, 2013a; Office for National Statistics et al., 2017) to 6.1% in 2016 (Nomis and Office for National Statistics, 2017). However the census is the most suitable data available to serve as constraints because it provides accurate aggregate data capturing nearly the whole population in each small area. The 2011 census data has the advantage of being the most recent census data available, and it is contemporaneous with a number of the waves used for the survey data.

Census constraint data were downloaded from Nomis using the Nomis RESTful API[5] and prepared prior to the simulation itself in `data-raw/2-simulate-pilot.R`. Table 5.4 shows the census tables used and their respective IDs.

5.3 Sampling frame

For the spatial microsimulation to produce valid results the sampling frame of the survey data must match the population of the constraint data. The simulation will not fail, but the quality and validity of the results will be affected and incorrect results could be produced.

Section 5.2.1 stated that the survey data should include England and Wales. This is to match the population given in the census tables available by default from Nomis. There are, however, some differences between the sampling frame used in the survey data with the population data gathered in the census.

[5] `https://www.nomisweb.co.uk/api/v01/help`

5.3.1 Age

Only individuals aged 16 and over are asked to complete a full survey in *Understanding Society* (Knies, 2015: 8). Therefore the *Understanding Society* data only contains individuals aged 16 and over, while the census includes all individuals, including those aged 0–15. To ensure the populations from *Understanding Society* and the census matched, individuals aged 15 and below were removed from the census.

The precise process to remove these individuals from the census varied by measure, and these are documented in `data-raw/2-simulate-pilot.R`. Typically the census tables either included individuals aged 16 and over—for example economic activity questions, or contained age information as either individual year or as bands so the appropriate age groups could be removed (Office for National Statistics, 2011).

5.3.2 Communal establishment residents

The census contains responses from individuals in non–private accommodation—such as nursing homes, hospitals, student halls of residence, military accommodation and barracks, and prisons (Office for National Statistics, 2014b)—while *Understanding Society* does not[6] (Buck and McFall, 2012: 9).

To perform the spatial microsimulation without making allowance for residents in communal establishments could introduce a systematic bias to the results. The problem arises because residents in communal establishments may differ in important ways to residents living in private

[6] In a private communication, a member of the *Understanding Society* sampling team clarified that only residents living in private households were included in the inital sample in 2007/8 but that if respondents moved into communal establishments they were followed up if possible in their new residence in subsequent waves (Kaminska, 2015). Communal residents are not, however, systematically sampled so *Understanding Society* is still best considered a survey of residents living in private accommodation only.

accommodation, and these can't be simulated if they do not exist in the individual–level data.

For example, (Cassells et al., 2013) argue that if there are a large number of care home residents in an area who are present in the census but not the individual–level survey, the spatial microsimulation will use the survey respondents to populate the area erroneously (Cassells et al., 2013: 12). In this example, survey respondents who are the same age as those in the care home will be used to populate the area as they will match the requirements of the constraints, but these individuals are likely to have better health than their contemporaries living in care homes so the health of this age group will be inflated in these areas.

There are essentially three potential ways to mitigate this issue. One is to impute these individuals and add them to the individual–level survey data. Another is to remove communal establishment residents from the census. The final option is to acknowledge and document the problem and the effects it might have on the simulation results.

Imputing responses is problematic; although an individual sample of records from the census is available that could be used to weight the samples there is no way to relate this to the target variable. Imputation was therefore discounted as an option to address this issue. Instead, removal of communal establishment residents was attempted, with varying degrees of success.

A relatively unique concern to this research is the fact that Doncaster is home to four prisons and young offender institutions: HMP & YOI Doncaster; HMP & YOI Hatfield; HMP Lindholme; and HMP & YOI Moorland (Ministry of Justice, 2015). The total offender population is 2522, or just over one percent of the Doncaster population. Table 5.5 shows the population of each prison based on the 2011 census (Office for National Statistics, 2011).

The proportion of prisoners in each of these output areas varies from 44% to 87%. In these output areas in particular, the characteristics of

Figure 5.1: Doncaster output areas with prisons (source: 2011 Census tables; Ministry of Justice)

Table 5.5: Doncaster prison population 2011.

Output area	Prisoner population
E00038161	599
E00038290	221
E00172382	1702

non–prisoners will be attributed to a large number of prisoners who are not represented in *Understanding Society*. This is problematic because prisoners have been demonstrated to have, on average, poorer mental and physical health than non-prisoners (Fazel and Baillargeon, 2011).

Two approaches were tested to remove prisoners from the census. The first, more nuanced, approach was to estimate the appropriate number of individuals to remove based on characteristics of prisoners in the sample of census microdata (Office for National Statistics, 2015b). This approach did not work because in some cases too many individuals were removed. This left negative numbers of individuals for some characteristics in some zones. To correct this would have involved a largely arbitrary decision about which demographic category to re-assign these individuals to, which would be difficult to justify. Code for this approach is documented in `data-raw/prisoners-deprecated.R` for reference, but was not implemented.

The second approach to remove prisoners was simply to remove the zones affected. For prisoners this meant removal of three output areas in which the prisons are located. Using this approach did mean that some non-prison residents were removed from the model, but this is arguably not a significant issue because the prisoners made up the majority of the population in these areas, and people in similar areas will have similar characteristics so such individuals are not 'lost'. This is the approach that I ultimately implemented, leaving 985 output areas in the simulation.

Doncaster also has a number of care homes. Removing care home residents and other community establishment residents was less feasible than removing prisoners. This was because the location of residential and nursing homes was not focussed on a small number of output areas as it was for the prisoners, and were instead dispersed throughout the district. Figure 5.2 shows the location of nursing and residential care homes in Doncaster, of which there are 92 unique establishments (Care Quality Commission, 2016), with the number of people aged 65 and over

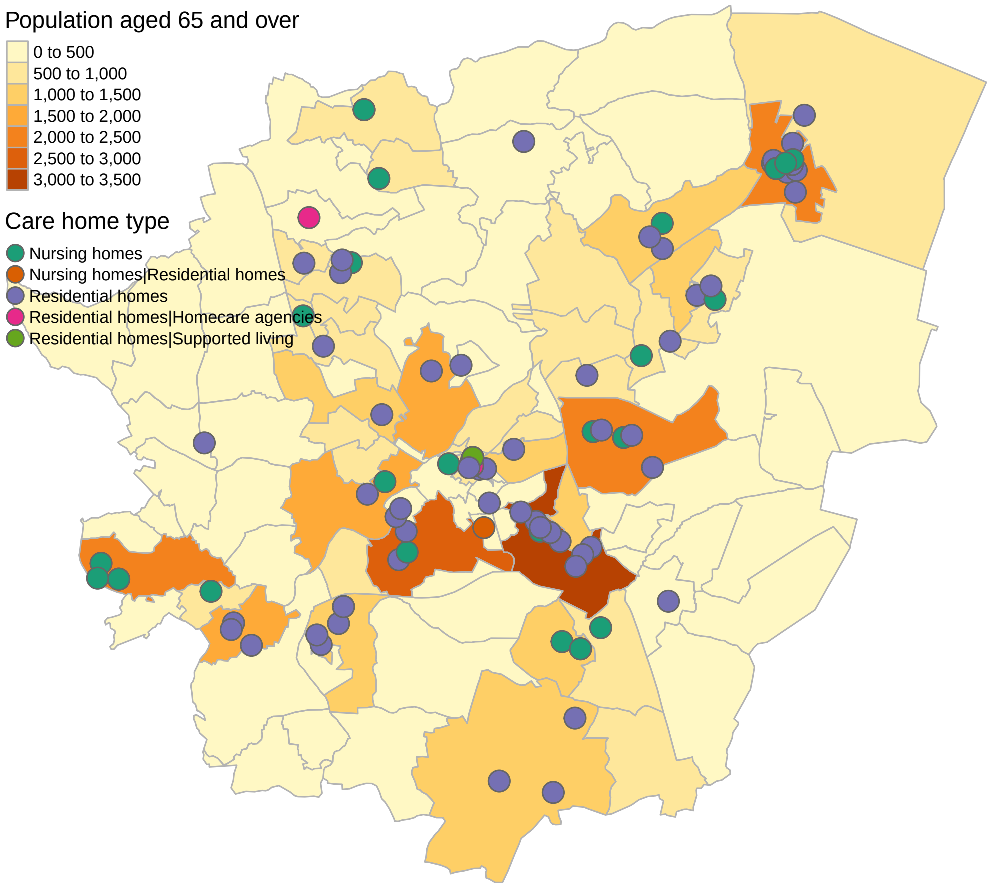

Figure 5.2: Location of nursing and residential care homes in Doncaster community areas (source: Care Quality Commission)

by community area.

Note, some care home establishments are in zones without any care home residents. This may be due to record swapping as part of the anonymisation procedure if these zones only include a small number of care home residents.

The number of residents in residential and nursing care homes in Doncaster on census day 2011 was 1,906, or just less than one per cent of the overall Doncaster population (Office for National Statistics, 2011). This is similar to the number of prisoners, but because they are spread across a significantly larger number of areas, they are less likely to have a large effect on these areas. Additionally, only age groups 65 and above in the zones are typically affected. Thus these individuals were left in the constraints.

Leaving residential and nursing care home residents and other community establishment residents in the model *will* affect the results. Specifically the model will likely over–estimate the health of older residents, which should be remembered and considered when examining the results of the microsimulation. Studies of care home migration in Sheffield have indicated that the effect this has on the over 65 population mortality is minimal (Maheswaran et al., 2014), so it seems likely the effect of care home residents on the model will be small and not significant.

5.3.3 Regional sampling

All eligible individuals from *Understanding Society* were used, rather than only individuals from Yorkshire and The Humber. Some authors have suggested that better results can be obtained by sampling from within regions only, as "[t]his avoids filling… Sheffield with Londoners" (Anderson, 2007: 15). Other suggested approaches include use of a 'geographical multiplier' or including a geographical constraint (Ballas et al., 2005b: 22–23, 2005), essentially so that 'local' individuals are more likely to be included in a zone than those living further away. Conversely (Tanton and Edwards, 2013b) compared the difference between sampling nationally and from within regions and concluded there was not a significant difference in their results (2013b: 166).

Analysis of area classifications produced by the (Office for National Statistics, 2015a) suggested regional sampling was not necessary and could be detrimental to the results because it would reduce the sample size significantly. These area classifications use a wide range of information from the census to group together similar areas based on a number of characteristics (Office for National Statistics, 2015a). These characteristics include age, ethnicity, number of children in the household, marital status, car or van ownership, employment, number of rooms and other characteristics of the household's accommodation. Using these classifications, figure 5.3 shows LADs in England and Wales belonging to supergroup 8, which includes Doncaster. There are a large number of areas similar to Doncaster outside Yorkshire and The Humber, including clusters in Wales and the North East, so to restrict the simulation to use only individuals from the Yorkshire and The Humber region would exclude individuals from these areas with similar characteristics. Similarly, local authority districts in Yorkshire and The Humber region are not homogeneous so to restrict the sample would not necessarily populate the model with similar individuals.

5.4 Target (dependent) variable

A number of suitable health variables were available in *Understanding Society* to serve as a target variable for this pilot. I wanted to simulate a target variable that was also available in the census data to help with externally validating the model (see 5.11). Health outcome variables that are available in both *Understanding Society* and the census were limiting long–term illness or disability and self–reported general health.

In *Understanding Society* limiting long–term illness or disability (referred to as 'long–standing illness or disability' in this data source) is asked as follows (University of Essex et al., 2016):

> Do you have any long–standing physical or mental impairment, illness or disability? By 'long-standing' I mean anything that has troubled you over a period of at least 12 months or that is likely to trouble you over a period of at least 12 months.

Responses can be either yes or no. Figure 5.4 shows the prevalence of long–standing illnesses or disabilities in *Understanding Society* after removing missing responses.

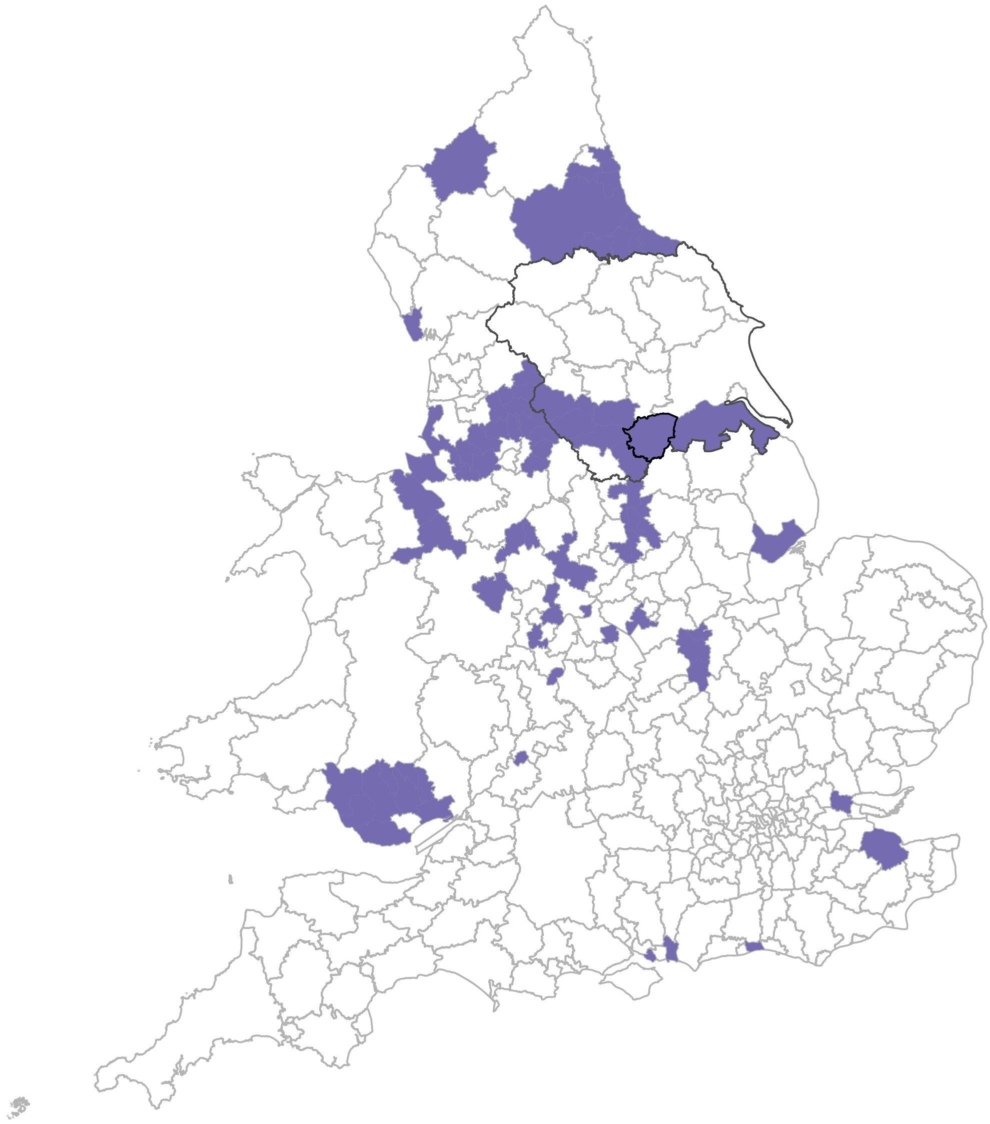

Figure 5.3: Local authority districts (LADs) classified as supergroup 8. Doncaster and Yorkshire and The Humber are outlined for context

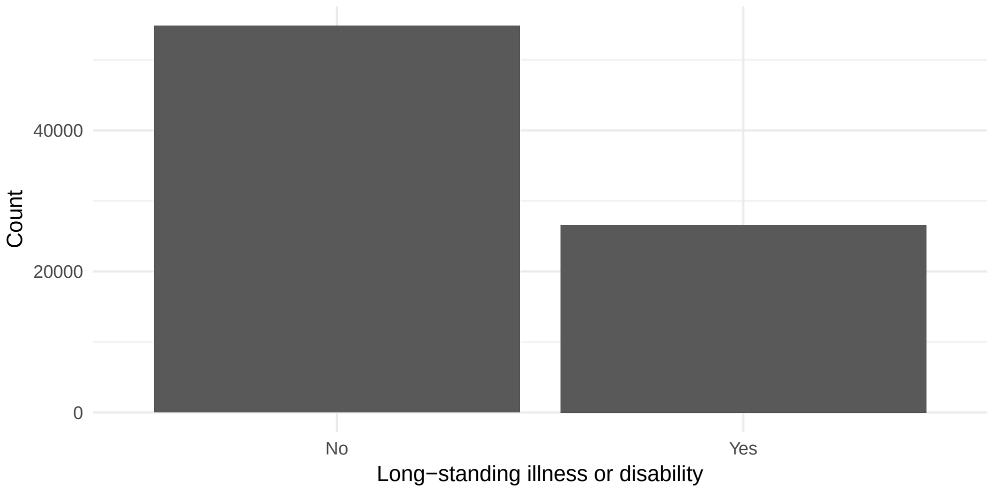

Figure 5.4: Long-standing illness or disability prevalence in Understanding Society

The census similarly asks:

> Are your day–to–day activities limited because of a health problem or disability which has lasted, or is expected to last, at least 12 months? Include problems related to old age (Office for National Statistics and National Archives, 2016).

Possible responses are: 'Yes, limited a lot'; 'Yes, limited a little'; and'No'. To make the two data sources compatible it is necessary to 'collapse' the census responses to match those in *Understanding Society*. We can add 'Yes, limited a lot' and 'Yes, limited a little' responses together to match 'Yes' so the two data sources can be compared. Figure 5.5 shows the prevalence of limiting long–term illness or disabilities in Doncaster based on data from the 2011 census.

The semantics of the two questions are slightly different, but both are conceptually similar and are attempting to determine if the respondent is affected by an illness or disability over a period of 12 months or more.

Self–reported general health is also asked in both data sources but the possible responses differ. In *Understanding Society* respondents are asked if their health in general is excellent, very good, good, fair, or poor. In the

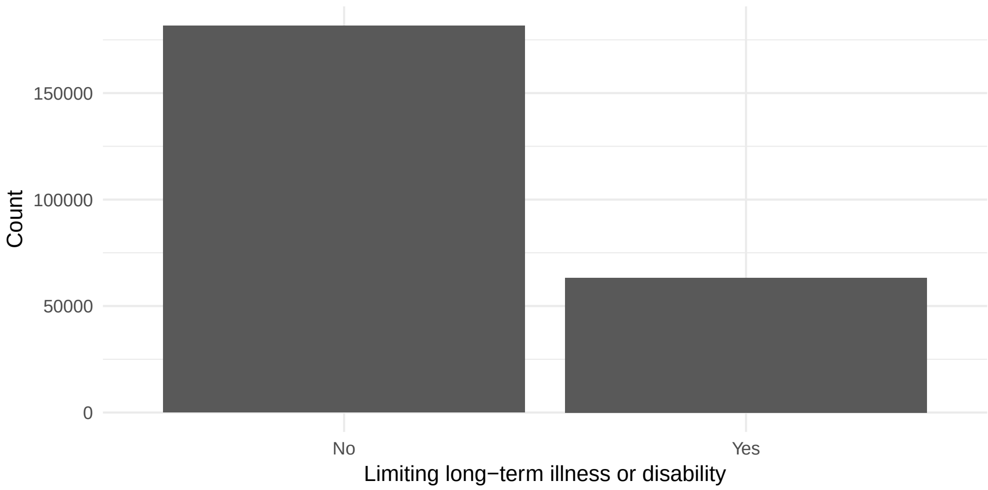

Figure 5.5: Limiting long-term illness or disability in Doncaster from 2011 census

2011 census possible responses are instead very good, good, fair, bad, or very bad. Matching responses between the two data sources is problematic. There is no clear way to match 'excellent' health or 'poor' health with the levels from the census. Similarly, fair health could be considered worse in *Understanding Society* than in the census, given their 'rank' in the ordering of the responses. For this reason I will use limiting–long term illness or disability for external validation, as matching the levels across the two data sets for this variable is less ambiguous.

5.5 Constraints

In Section 4.7 we saw that the configuration of contraints—the selection and ordering—may affect the simulation. In practice my choice of constraint was limited to variables that were available in both the census and *Understanding Society*, but this still offered a rich pool of information to choose from. To ensure the constraints I chose created an optimal model I selected them based on theory (see Section 2.7) and empirically tested these to ensure they correlated well with my target variable. This section describes how I made my initial selection of constraints, how I matched these in both the census and *Understanding Society*, and how I ensured the populations in the census matched across variables. In the next section I describe how I empirically tested their correlation with my target variable, then how I ordered these for entry into the simulation.

5.5.1 Initial choice

My choice of constraints included essential demographic information, such as age, ethnicity, and sex. I also chose to include highest education qualification as this has a known relationship with health outcomes. I also wanted to include an indicator for poverty or deprivation, as this is known to be a powerful predictor of health outcomes and inequality (see Section 2.7). There is no income variable in the census, so instead I used indirect measures as indicators of poverty, including car availabilty, housing quality, and home ownership (Senior, 2002). These are asked in both the census and in *Understanding Society* so can be used as constraint variables.

I am not suggesting that the lack of these amenities or non–ownership housing tenure should be enough to consider a household in poverty. Instead they may indicate reduced or limited financial resources which is linked to poorer health outcomes. Similar measures have been used as markers of deprivation, for example by Townsend et al. (1988).

My final choice of constraints for this pilot were: age; ethnicity; sex; highest educational qualification; car or van availability; and housing tenure.

5.5.2 Matching variable levels

In order to be used as constraints the variable levels must match *precisely* between the survey and the census data (Tanton and Edwards, 2013b). The definitions used in many demographic variables matched exactly, so no further recoding was required. In some cases it was necessary to recode some variables before performing the logistic regression tests. These are performed by the scripts `data-raw/0-prep-understanding-society.R` for survey variables and `data-raw/2-simulate-pilot.R` for census variables, and are described in brief here.

Ages were given as individual years in *Understanding Society*, so were recoded or 'cut' into bands that matched those already available in the census.

Ethnicity was only available with a reduced classification in the census with age information, which was necessary to remove children and young people aged under 16 from the constraints. To match the census, I therefore had to recode ethnicity in *Understanding Society* to match the reduced levels available in the census with age information by combining some ethnic groups.

Tenure was grouped differently in the census and in the survey. Those who owned their home with a mortgage or loan and those with shared ownership were grouped together in the census but separate in *Understanding Society* so were recoded in the survey to merge these two groups. In the census private renters and those living rent–free were grouped together, while social renters were a separate classification. In *Understanding Society* both private renters and social renters were grouped together. To address this it was necessary to amalgomate private renters, social renters, and those living rent–free in to one group in both the census and the survey. This was less than ideal, but the only classification the two data sources allowed.

Education was perhaps the most challenging variable to prepare for comparison between *Understanding Society* and the census. The census provides the highest qualification by level, from no qualifications, through Level 1 qualifications, Level 2 qualifications, apprenticeships, Level 3 qualifications, Level 4 qualifications and above, and finally other qualifications not captured with these categories. This allows regulated qualifications in

England and Wales from different frameworks to be compared (Gov.uk, 2016a).

Understanding Society, on the other hand, provides the highest education qualification the respondent reported, based largely on academic qualifications types, such as degree, higher degree, 'A'–level, GCSEs, other qualifications, and no qualifications. To compare the two classifications, I converted the qualifications given in *Understanding Society* to levels.

GCSEs were problematic because GCSEs grades D–G are considered level 1, while five GCSEs grades A*–C are considered level 2 (Gov.uk, 2016b). To solve this I coded all GCSEs as a 'Level 2 qualifications', in effect treating this as 'up to and including Level 2'. For the census data to match I simply added the number of individuals with Level 1 qualifications and Level 2 qualifications together, to create a comparable 'up to and including Level 2 qualifications' classification in the census.

AS and A level qualifications were coded as level 3 qualifications. Degree and other higher degree qualifications were coded as level 4. Apprenticeships in *Understanding Society* were coded as 'other qualification', so apprenticeships in the census were added to 'other qualifications' to match. Figure 5.6 shows the final coding and their respective proportions in *Understanding Society*.

5.5.3 Matching census populations

If the populations of the individual census constraints do not match the model will fail to constrain and not produce results. The number of respondents in each zone for each constraint must match exactly for the simulation to work. For example, the population of each zone in the age census table must match the population of the same zone in the ethnicity census table.

For most census tables the populations already matched so no preparation was necessary. In some cases the populations did not match and it was

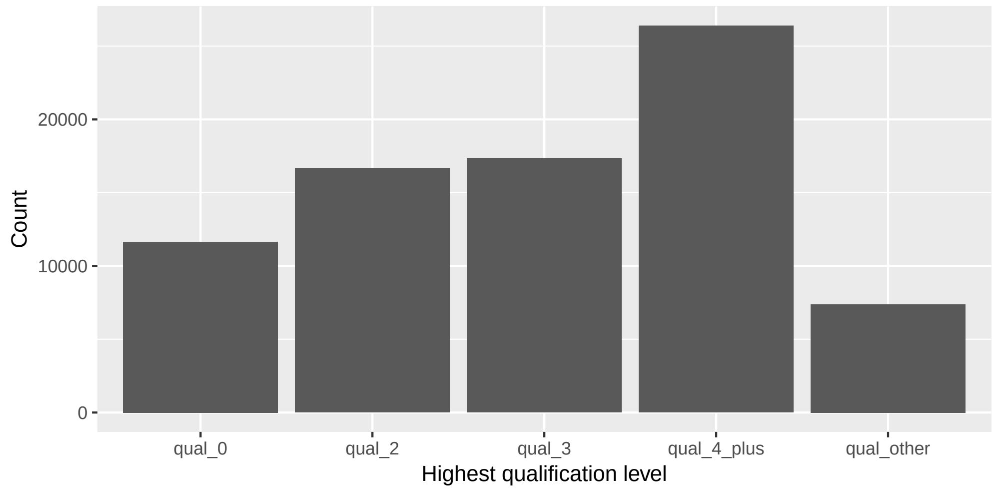

Figure 5.6: Highest qualification

necessary to re–weight these to match populations in other variables. It is possible that some form of statistical disclosure control, for example record swapping or small cell adjustment, has been applied to maintain the anonymity of indvididuals, particularly for the small areas geographies I am using, which could explain why the populations did not match exactly (Office for National Statistics, n.d.: 6–7). Some variables are collected about households or a household reference person, rather than individuals within the household. This too could cause the population of the zones to be different. In either case, the solution to infer populations is acceptable as many characteristics of the household, such as socio–economic position, are determined by the resources—such as cars—available to household overall.

There are three possible values for car ownership in the census, and these were recoded in the survey to match. These are the household has no cars, one car, or two or more cars. To impute 'correct' zone populations for the car ownership table I calculated the proportion of respondents that had no cars, one car, or two or more cars in each zone. This was then multiplied against the known population of the zone. The result was a fraction, which can be fatal for individuals, so this number was rounded to a whole number. Finally, where the imputed population now had one person too many or too few due to rounding, this was taken away from or added to the modal category to minimise the change in the proportions. This procedure is documented in `data-raw/2-simulate-pilot.R`.

5.6 Empirically test constraints

Because limiting long–term illness or disability is measured at the binary level—yes or no, the assumption of linearity between variables necessary for linear regression would be unmet, making linear regression unsuited to this task (Field et al., 2012: 314–315). Instead, logistic regression is the most appropriate statistical test to use. Logistic regression predicts the *probability* of an event occurring, rather than the *value* of the event, making it suitable for use with binary dependent variables. Equation (5.1) shows the logistic regression equation:

$$P(Y) = \frac{1}{1 + e^{-(b_0 + b_1 X_{1i} + b_2 X_{2i} + \ldots b_n X_{ni})}} \tag{5.1}$$

where $P(Y)$ is the probability of event Y occurring, b_0 is the intercept, and b_n is the regression coefficient of variable X_{ni}.

In `R` the model is set up using the `glm()`—generalised linear model—function with a binomial family. After removing missing cases the model was input as follows ($n = 58,740$):

```
m_llid <- glm(
  llid ~ age + eth + sex + qual + car + ten,
  data = us_llid,
  family = binomial())
```

The model overall is an improvement on the baseline model, as the AIC is lower than the baseline model 65062.43 (compared to 74985.91 for the baseline model). The Nagelkerke *pseudo*–R^2 of the model is 0.22 and the likelihood ratio = 9973.48 (*df* = 25, $\chi^2 \approx 0$). Most levels of age and tenure, and all levels of ethnicity, sex, highest qualification, and car ownership were statistically significant ($p << 0.05$; 95% confidence intervals of the odds ratio do not cross 1.0).

Most levels of age were statistically significant. Respondents aged 18–24 were not statistically significantly more likely to have a limiting long–term illness or disability compared to the reference group (ages 16–17). All other age groups were more likely to have a limiting long–term illness or disability, and the likelihood increases dramatically with age, as would be expected.

All ethnic groups had lower odds of having a limiting long–term illness or disability than the reference group (White British), and these differences were statistically significant. Black African and Black Caribbean respondents, for example, were only about half as likely to have a limiting long–term illness or disability as White British respondents. Males had lower odds of having a limiting long–term illness or disability than females.

Individuals with higher qualifications had lower odds of having a limiting long–term illness or disability than the reference group of no qualifications. The likelihood decreased with each additional level of qualification gained, so that individuals with level 4 and above qualifications were least likely to have a limiting long–term illness or disability. Respondents with 'other' qualifications were still less likely to have a limiting long–term illness or disability than the reference group.

Car ownership is associated with lower odds of limiting long–term illness or disability, and the odds continue to decrease for families with more than one car. Home ownership—either owning outright or owning with a mortgage—is associated with lower odds of limiting long–term illness or disability than renting, and the odds for both kinds of ownership are not statistically different so are similar.

With constraints selected it was then necessary to order them for entry in the spatial microsimulation algorithm. Some authors argue that constraints should be entered so that the 'weakest' variable is entered first

Table 5.6: LLID model summary

predictor	beta	p_value	sig	lower_ci	odds_ratio	upper_ci
(Intercept)	-1.39	0.00	**	NA	NA	NA
age_18_19	0.01	0.94		0.80	1.01	1.27
age_20_24	0.19	0.06		1.00	1.20	1.47
age_25_29	0.28	0.00	**	1.10	1.33	1.61
age_30_44	0.81	0.00	**	1.87	2.24	2.69
age_45_59	1.50	0.00	**	3.77	4.49	5.39
age_60_64	1.96	0.00	**	5.87	7.07	8.57
age_65_74	2.16	0.00	**	7.22	8.66	10.46
age_75_84	2.48	0.00	**	9.86	11.91	14.47
age_85_89	2.63	0.00	**	11.16	13.94	17.50
age_90_plus	2.63	0.00	**	10.78	13.82	17.82
eth_irish	-0.19	0.02	*	0.71	0.83	0.96
eth_other_white	-0.63	0.00	**	0.47	0.53	0.60
eth_mixed_multiple_ethnic	-0.17	0.02	*	0.73	0.84	0.98
eth_asian_asian_british	-0.53	0.00	**	0.55	0.59	0.63
eth_black_african_cbbean_british	-0.66	0.00	**	0.47	0.52	0.57
eth_other_ethnicity	-0.40	0.00	**	0.58	0.67	0.78
sex_male	-0.07	0.00	**	0.90	0.93	0.97
qual_2	-0.28	0.00	**	0.71	0.76	0.81
qual_3	-0.32	0.00	**	0.68	0.73	0.78
qual_4_plus	-0.45	0.00	**	0.60	0.64	0.68
qual_other	-0.12	0.00	**	0.82	0.89	0.95
car_1	-0.29	0.00	**	0.71	0.75	0.79
car_2_plus	-0.53	0.00	**	0.56	0.59	0.63
ten_owned_mortgage_shared	-0.03	0.34		0.92	0.97	1.03
ten_rented	0.34	0.00	**	1.32	1.40	1.48

and subsequent variables improve the model fit (see Section 4.7).

The order of the constraints can be informed by β coefficients, odds, or log odds. These do not identify the variables with the greatest effect on the predictive power of the model, but do suggest an order based on the effect of the variables on the outcome. Using log odds, the values furthest from 0.0 have the greatest effect on the outcome, so are entered last. The log odds of the model are listed in table 5.7.

While there is variation within the levels of each variable, broadly the order of the 'weakest' to the most powerful variable is: sex; highest qualification; housing tenure; ethnicity; car ownership; and finally age.

I tested to see if the constraint order affected the simulation results, as theoretically it should not. I did this by comparing simulation results produced using the constraint order outlined above with simulation results produced using a randomised constraint order. In all cases the difference between the results was negligible, so most likely other orderings of

Table 5.7: Model predictors with log odds

predictor	log_odds
(Intercept)	NA
age_18_19	0.01
age_20_24	0.19
age_25_29	0.28
age_30_44	0.81
age_45_59	1.50
age_60_64	1.96
age_65_74	2.16
age_75_84	2.48
age_85_89	2.63
age_90_plus	2.63
eth_irish	-0.19
eth_other_white	-0.63
eth_mixed_multiple_ethnic	-0.17
eth_asian_asian_british	-0.53
eth_black_african_cbbean_british	-0.66
eth_other_ethnicity	-0.40
sex_male	-0.07
qual_2	-0.28
qual_3	-0.32
qual_4_plus	-0.45
qual_other	-0.12
car_1	-0.29
car_2_plus	-0.53
ten_owned_mortgage_shared	-0.03
ten_rented	0.34

Table 5.8: Cases remaining by included variables

Variables	Cases
age	81,397
age, sex	81,396
age, sex, eth	63,455
age, sex, eth, qual	61,787
age, sex, eth, qual, car	58,906
age, sex, eth, qual, car, tenure	58,740

constraints will most likely have produced nearly identical results. See `data-raw/constraint-order-comparison.R` for code that made this comparison.

5.7 Number of cases

There are $81,540$ cases in *Understanding Society*, but after removing missing cases $58,740$ remain. Table 5.8 shows the number of cases remaining as additional variables are added to the model.

There is clearly a trade–off to be made between the number of variables included in the simulation and the number of cases lost through missing data. I decided to include ethnicity in the model because it is such a fundamental piece of demographic information that the richness of the simulation would be lost if it were excluded. Having decided to include ethnicity there were very few additional cases lost through missing data after adding qualifications, car ownership, and housing tenure, so these were including in the simulation.

5.8 Hardware

Spatial microsimulation is becoming more possible with the increasing capabilities of consumer computers, but is still a relatively computationally intensive task. A modern desktop computer with 16 gigabyes of memory was used to prepare and reweight the datasets. This was sufficient to work with the individual–level and aggregate–level datasets, and compute the reweighted matrices for each area.

5.9 Software

The spatial microsimulation was prepared and performed in `R` (R version 3.4.3 (2017-11-30)) (R Core Team, 2017). `R` was ideally suited to this task because it is both a statistical analysis programme as well as a programming language, so it was able to handle all of the tasks required of spatial microsimulation within one environment. These included: obtaining, loading (Wickham, Hester, et al., 2017) and 'tidying' (Wickham, 2014) datasets (Wickham and Henry, 2017; Wickham, Francois, et al., 2017); regression analysis for selecting and testing variables and results; for the spatial microsimulation (Jones et al., 2017) itself; for GIS (Bivand and Rundel, 2017; Bivand et al., 2017; Pebesma and Bivand, 2018); and for writing output and plotting charts (Tennekes, 2017; Wickham and

Chang, 2018).

`R` is free software (both in the sense of 'free of charge' and 'free speech'). `R` is also a scripted, interpreted, language. Both of these characteristics mean that it is easy for others to obtain the software and run the code used to perform the spatial microsimulation, validate the ouputs or use the code in their own projects. Code for this project can be obtained from `http://etheses.whiterose.ac.uk/id/eprint/19283` or from `https://github.com/philmikejones/thesis`.

There is a recent precedent of researchers using `R` for spatial microsimulation (Campbell, 2011; Lovelace, 2013; New Zealand Ministry of Business, Innovation and Employment, 2015). This body of good quality code, as well as policy and peer-reviewed outputs, attests to the robustness of `R` for spatial microsimulation.

I used 'Git' (Torvalds, 2015) and 'GitHub' (GitHub, Inc., 2015) for code version control, for backups, and to manage issues; Travis CI (Travis CI, GmbH, 2016) for continuous integration testing; and Codecov (Codecov LLC, 2016) for testing coverage. The simulation was run on a computer running a Linux operating system (Canonical, 2016).

5.9.1 rakeR

Building on code written by Lovelace and Dumont (2016) I wrote the `rakeR` package (Jones et al., 2017) to aid the spatial microsimulation. This removes the need for a lot of the manual data manipulation stages as these are now done automatically by the software. The advantages of this are the ability to create spatial microsimulations more quickly, as there are fewer steps involved, and the reduction of opportunities for human error to creep into the simulation.

`rakeR` accepts two data frames as arguments, one for the individual–level survey data and one for the combined aggregated census data. It then returns a single data frame of weights or integerised cases, as requested by the user, with the simulated microdata. Source code for `rakeR` can be obtained from `https://github.com/philmikejones/rakeR`; the CRAN (Comprehensive R Archive Network) package can be obtained from `https://cran.r-project.org/package=rakeR`; or installed in `R` with `install.packages("rakeR")`.

`rakeR` is designed to be robust and this is achieved through defensive programming practices. It will 'fail fast' (Wickham, 2015: 169) if it encounters any errors or ambiguities, for example if the inputs are not of the expected type. It will not, for example, attempt to infer an option or input data type if they are misspecified. This has the advantage that it will not produce spurious or incorrect answers, which is particularly dangerous if the user is unaware the answer provided are spurious. Component functions are unit tested to verify their results compared to known, good simulations. `rakeR` version 0.1.2 was used to simulate the pilot.

5.10 Weighting

Having chosen and prepared suitable constraints and decided on their order of entry into the model, I turned to performing the spatial microsimulated itself. The final model will be used to compare current circumstances and a small number of discrete 'snapshots' of proposed changes to factors that are hypothesised to affect resilience (Ballas et al., 2005: 8).

Fractional weights were created using the `weight()` function in the `rakeR` package which in turn uses the `ipfp` package (Blocker, 2013). This is used because it is written in `C` so is at least an order of magnitude quicker than base `R` (Lovelace and Dumont, 2016). The `weight()` function accepts the survey data frame and the census data as arguments and produces a table of fractional weights for each individual in each zone. Effectively the entire survey population is copied exactly into each zone then, in each zone in turn, the individuals in that zone are reweighted by the algorithm to match the appropriate constraint tables. The resulting data is structured like a cube, with each 'slice' through the cube representing a zone, and on each slice is a data frame containing the individuals in the survey with their appropriate weights.

Simulated data is produced for each zone in Doncaster that matches the real population in size and constraint characteristics as closely as possible. Alongside these simulated constraint variables the simulation contains the target variable, limiting long–term illness or disability, which I have used for external validation.

I simulated weights for output areas (OAs), lower–layer super output areas (LSOAs), and middle–layer super output areas (MSOAs) simultaneously. This allowed me to test if the size of the zone being simulated affects the accuracy of the model.

The fractional weights are difficult to work with on their own, so the reweighted results were 'extracted' and integerised using the `integerise()` function which aggregates the individuals in each zone and turns the fractional weights into integers. I used the 'truncate, replicate, sample' (TRS) method of integerisation because it provides more accurate results than other approaches (Lovelace and Ballas, 2013: 10).

Integerised weights are typically used if they are being entered into a dynamic or agent–based model. Although I am not using a dynamic model, integerised cases have the advantage of being more meaningful because each case represents a simulated individual. By simulating both I was able to compare the accuracy of the results of both methods. Integerised weights are also useful to provide 'case studies' that are illustrative of individuals affected by policy changes (Campbell and Ballas, 2013). I select a number of simulated individuals as case studies in Section 7.3.

5.11 Validation

The simulated population ($244,909$) matched the actual population exactly ($244,909$) indicating the simulation has constrained accurately overall. The correlation between the simulated and actual population in each zone can be used to assess the quality of the simulation. The correlation statistic is standardised, so a value of 1 is ideal. The correlation between the simulated and integerised zone populations and the actual zone populations for this simulation is 1.

A chart of the simulated and actual zone populations confirms the simulated populations are accurate (Figure 5.7). As the simulated constraint should equal the 'real' constraint, the 'ideal' line of $y = x$ is plotted on each chart rather than the actual line of best fit.

It is also possible to test the correlation between the simulated and actual populations by variable. Figures 5.8 and 5.9 show the model fit for age 30–44 and White British ethnicity respectively. They are illustrative of the accuracy of the fit of the model to the actual data. Plots for each level of each variable were created which fit the actual data very accurately. These are not shown because of their similarity to the displayed plots but can be found in the `figures/cache/` directory.

t–tests of the simulation indicated that all categories of all variables are from the same distribution because they are not statistically significant. These tests suggest the spatial microsimulation is performing well and that the results match 'real' data well. Table 5.9 summarises the results of the *t*–test for each category of variable.

The total absolute error overall is 0.00000000069, making the standardised absolute error overall 2.83×10^{-15}, both negligible amounts. These results are extremely encouraging; no area has an SAE of 0.1 or greater, which is suitable for simulating a relatively rare event (see Section 4.8). The mean SAE is 2.89×10^{-18} with a standard deviation of 2.43×10^{-18}, meaning the errors are well within the validation criteria and are effectively zero

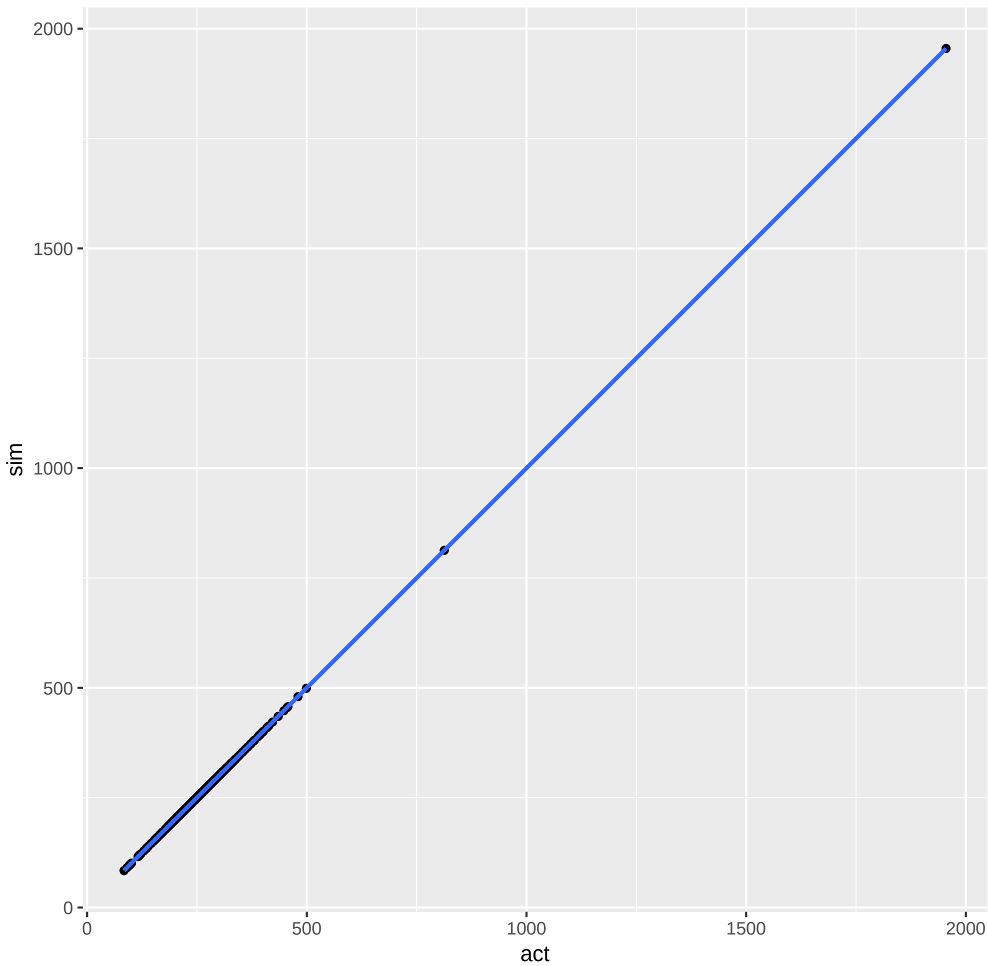

Figure 5.7: Actual population against simulated population at output area level

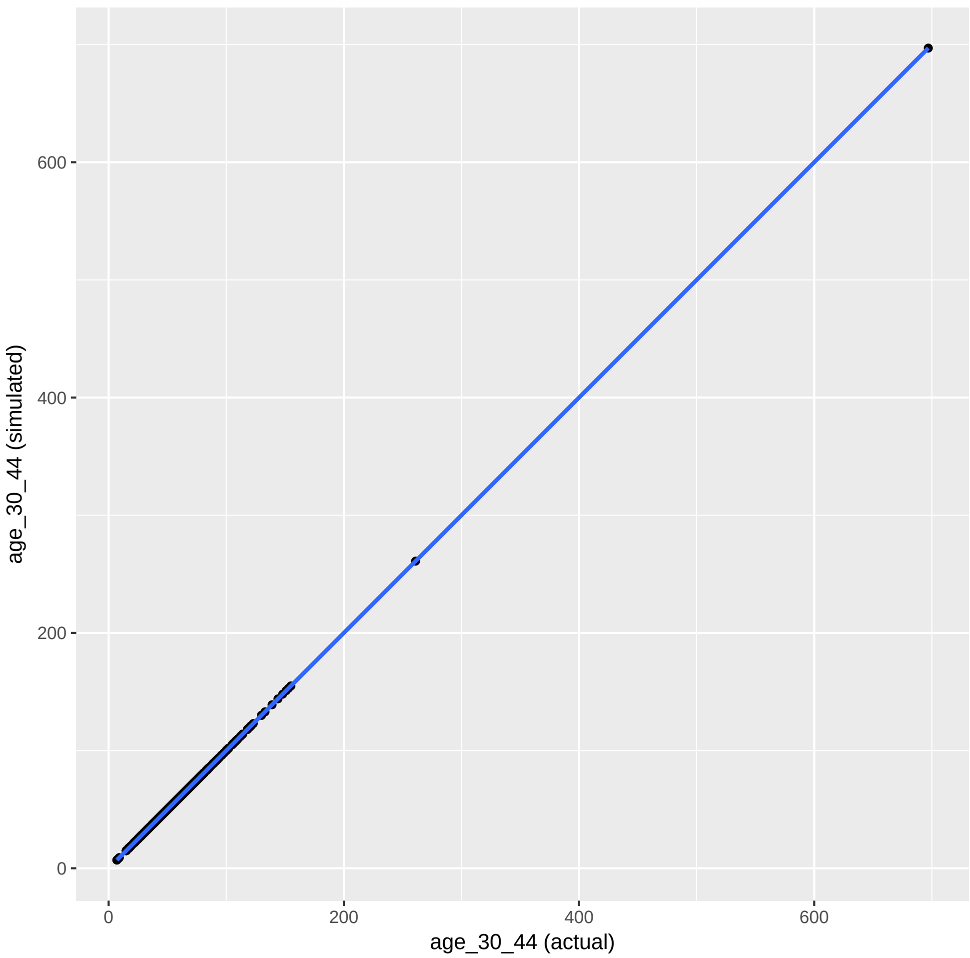

Figure 5.8: Internal validation of age

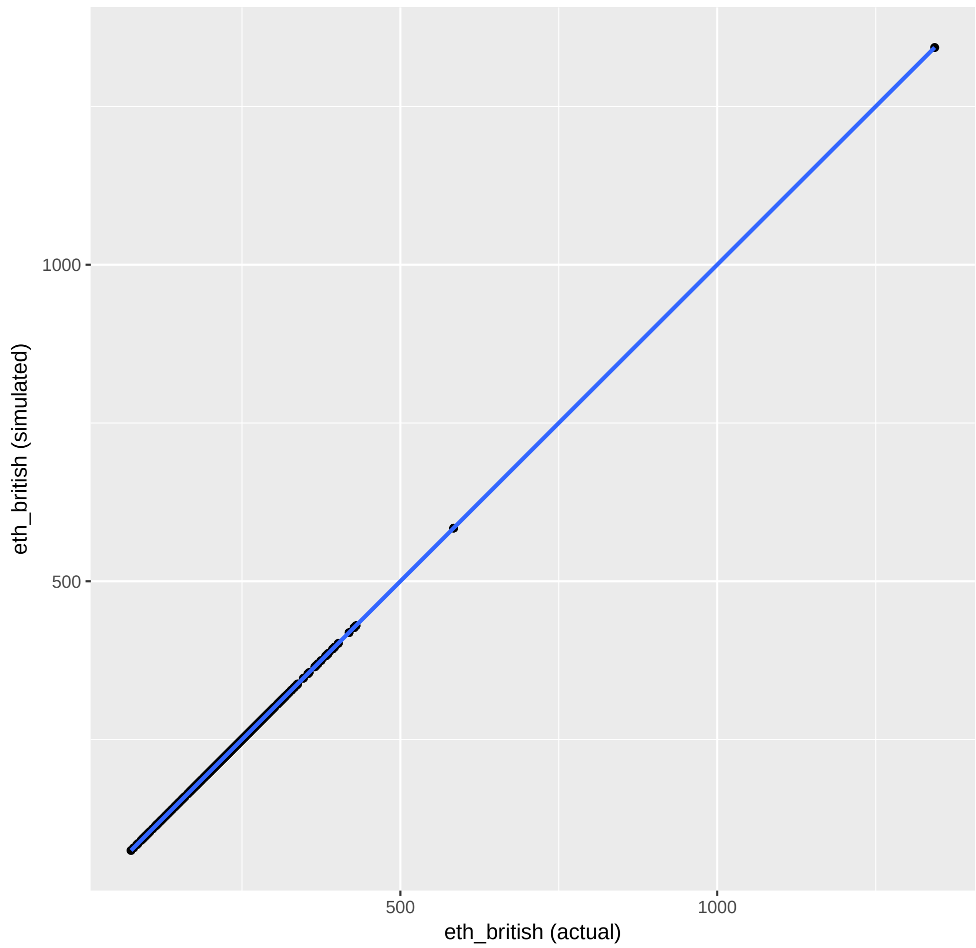

Figure 5.9: Internal validation of ethnicity

Table 5.9: Result of t-tests comparing simulated against actual data

variable	statistic	p_value
sex_female	0	1
sex_male	0	1
qual_0	0	1
qual_2	0	1
qual_3	0	1
qual_4_plus	0	1
qual_other	0	1
eth_british	0	1
eth_irish	0	1
eth_other_white	0	1
eth_mixed_multiple_ethnic	0	1
eth_asian_asian_british	0	1
eth_black_african_caribbean_british	0	1
eth_other_ethnicity	0	1
ten_owned_outright	0	1
ten_owned_mortgage_shared	0	1
ten_rented	0	1
car_0	0	1
car_1	0	1
car_2_plus	0	1
age_16_17	0	1
age_18_19	0	1
age_20_24	0	1
age_25_29	0	1
age_30_44	0	1
age_45_59	0	1
age_60_64	0	1
age_65_74	0	1
age_75_84	0	1
age_85_89	0	1
age_90_plus	0	1

for practical purposes. These are significantly better than thresholds suggested by Birkin and Clarke (1988) and Smith et al. (2009).

As with internal validation, to externally validate the model the simulated values should exactly match the values from the census for each zone which, if plotted, produces a line of $y = x$. Figure 5.10 shows this comparison for this pilot simulation.

An equal variance, two-tailed t–test can again be used for the purpose of statistically comparing the actual with simulated counts. In this instance the result is statistically significant (25.74, $p \approx 4.64 \times 10^{-126}$). The residuals from the mean with no model were smaller than the residuals against the ideal line in this case because the simulation consitently over–estimated the number of people with a limiting long–term illness or disability.

As the over–estimation was consistent a commensurate, consistent, adjustment was made to the model data. A 'calibration' or 'alignment' such as this has been used in similar circumstances (Morrissey et al., 2013: 222). Each simulated value was reduced by the difference in means ($\overline{sim} - \overline{real}$) producing a more accurate simulation, illustrated in Figure 5.11. Applying a fixed correction to each value of the difference between the observed mean and the simulated mean improved the model fit, which is confirmed by a t–test (0, $p \approx 1$).

The 'Standard Error around Identity (SEI)' provides another measure of similarity and is a measure of residuals from the 'ideal' line of $y = x$. Ballas et al. (2007) stipulate the SEI is calculated as "[t]he square root of the average of the sum of the squared deviations about that line" (Ballas et al., 2007: 58–59), so is a measure of the 'average' deviation from the ideal line for all zones, with smaller numbers indicating smaller deviations. The SEI for the corrected vector is 14.39 against an average population size of 245.5, suggesting a good fit between simulated and actual data.

A final method of external validation is to calculate the percentage of simulated individuals in each area who are incorrectly classified, a method

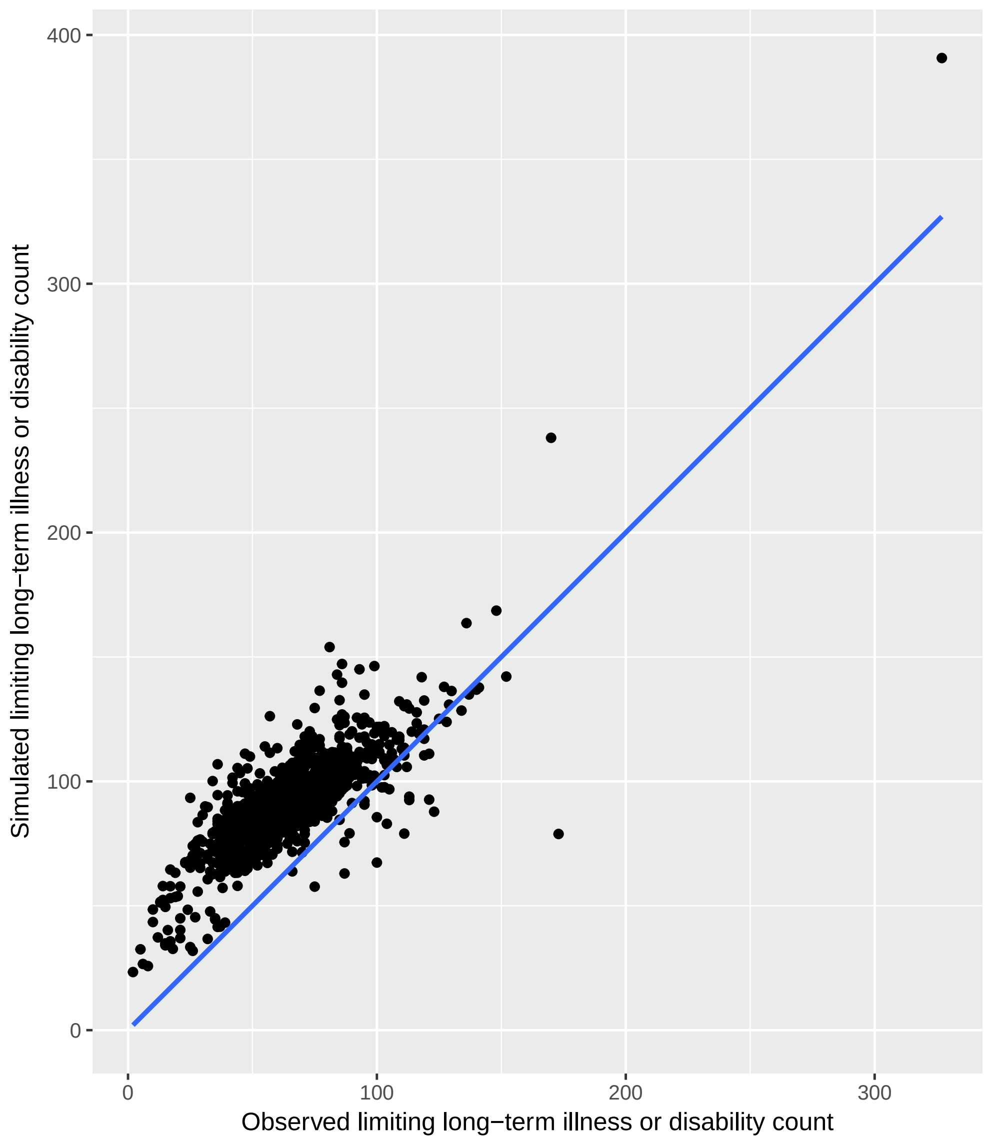

Figure 5.10: External validation of pilot simulation

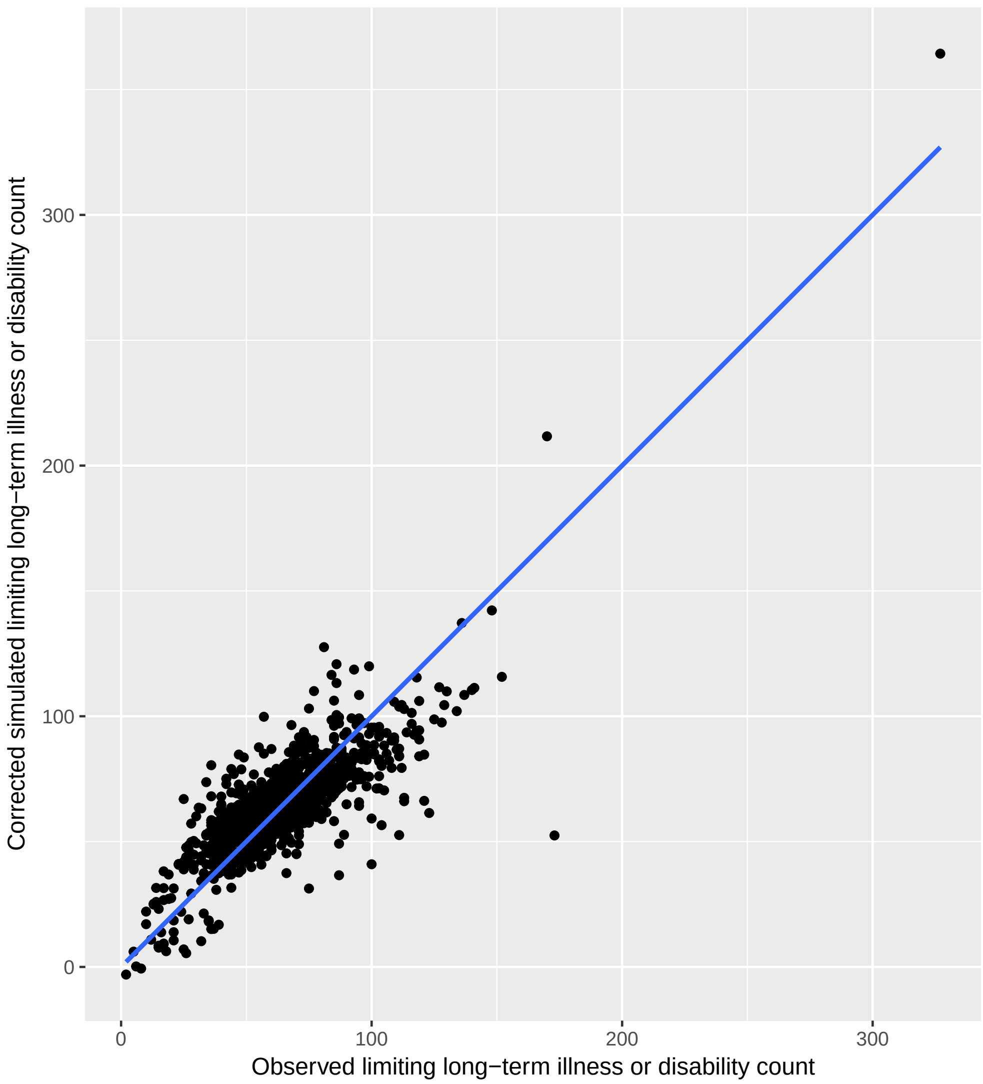

Figure 5.11: Actual against simulated population with a limiting long-term illness or disability, corrected

used by Smith et al. (2011), p. 621. The percentage error is calculated as the mean of the absolute differences between the number of simulated individuals and the number of 'real' individuals with a limiting long–term illness or disability, divided by the total individuals for each zone. The overall percentage error for the corrected simulation is therefore 4.34%. These results compared favourably with the results obtained by Smith et al. (2011).

5.12 Comparison of weights and integerised simulations

Figure 5.12 illustrates the similarity between the fractional and integerised weights. Red indicates a fractional weight, and blue indicates an integerised weight. A two–sided, equal variance t–test confirms the distributions are statistical similar (-0.25, $p = 0.8$). Based on this, either the integerised weights or fractional weights could be used. I use both the fractional and the integerised weights in different circumstances. For single variables I opted to use the fractional weights, but for comparison of multiple variables I used integerised weights instead because they are more intuitive. The integerised weights could be used for a future dynamic or agent–based model with confidence.

5.13 Comparison of geography zone sizes

All zone sizes simulate well when the correction is applied (Figure 5.13). The external validation shows incremental improvement as geographical zone size increases but the simulation of output areas is more detailed than LSOA or MSOA without much loss of accuracy. Smaller geographies provide less homogeneity and more detail, without much loss of accuracy, so a simulation at output area level is most appropriate.

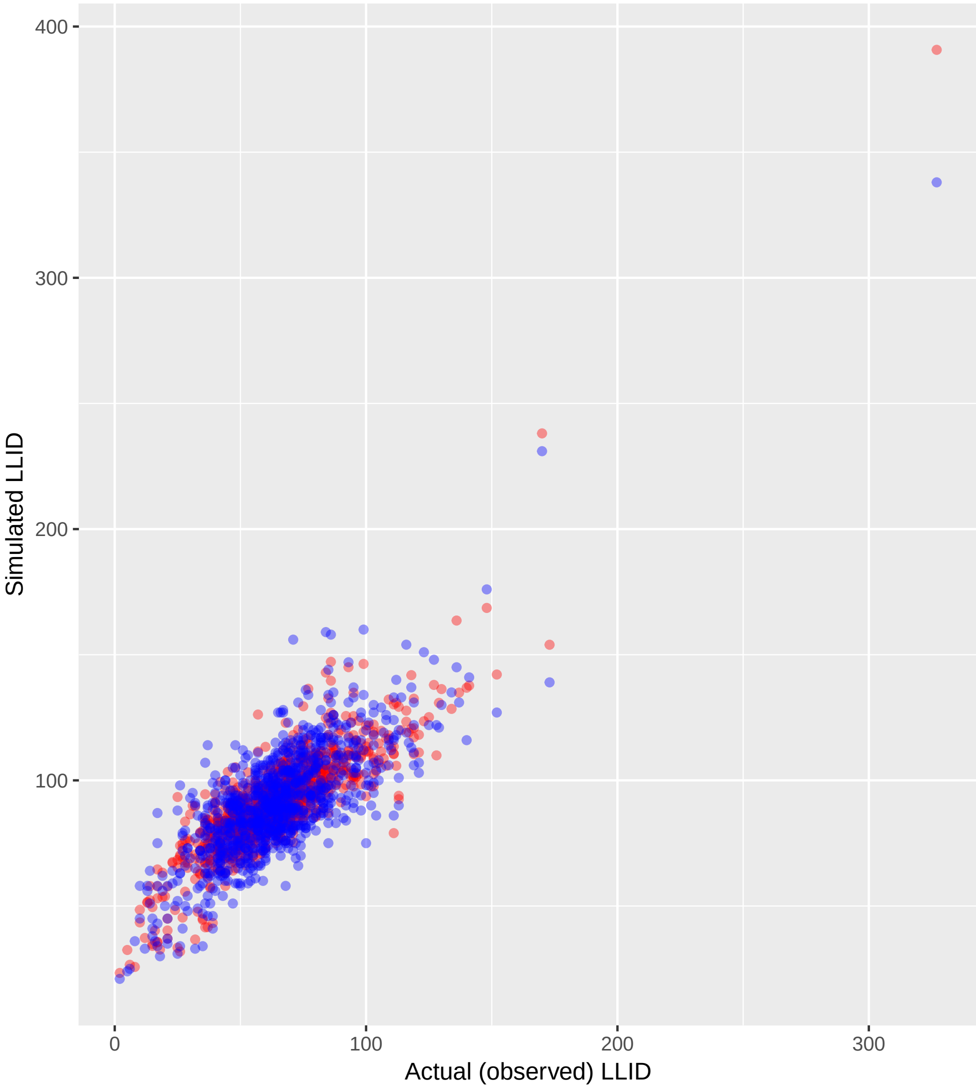

Figure 5.12: Comparison of fractional weight and integerised weight distributions (red = fractional weights; blue = integerised weights)

Table 5.10: Comparison of simulation at OA, LSOA, and MSOA geographies

zone_size	correlation	sae	sei	perc_error
OA	1	0	14	0.04
LSOA	1	0	45	0.03
MSOA	1	0	157	0.02

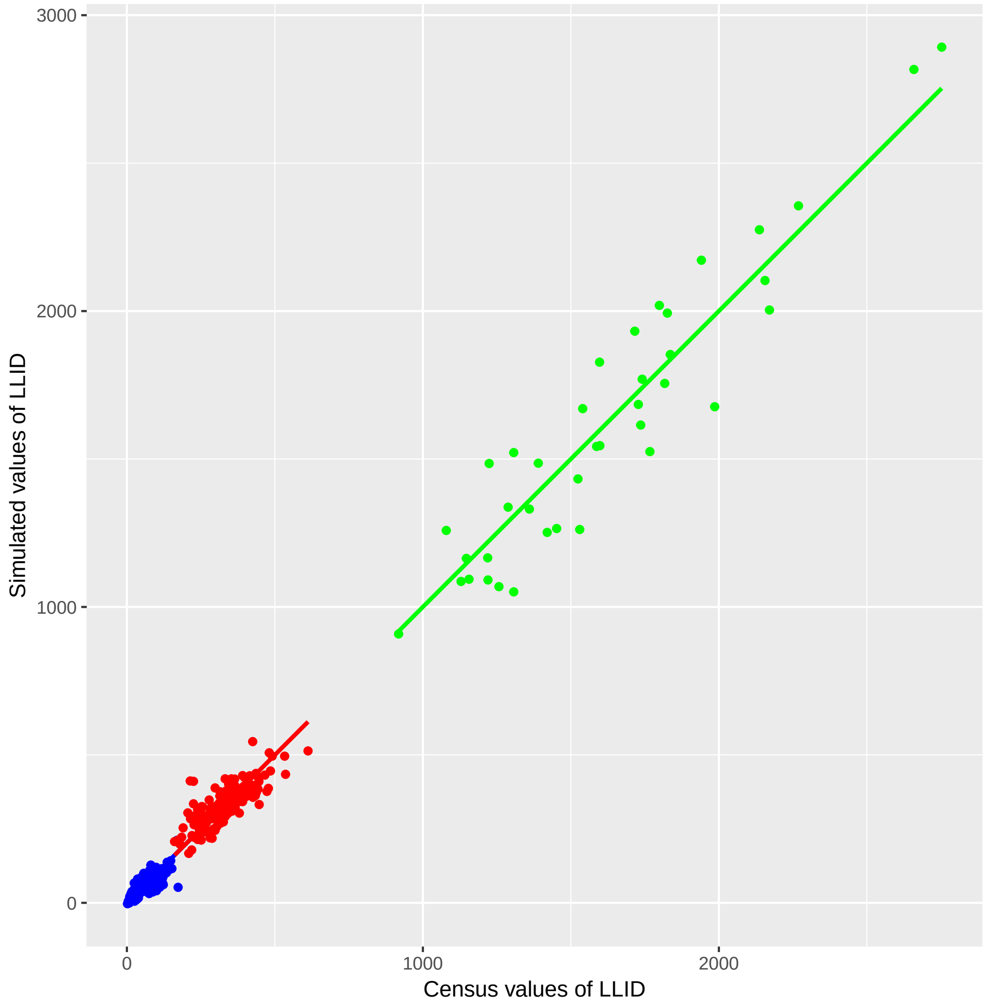

Figure 5.13: Comparison of OA (blue), LSOA (red), and MSOA (green) simulations

Table 5.10 compares summary statistics across the three geographical levels.

5.14 Results

5.14.1 Ouptut areas

By comparing the simulated and actual proportion of people with a limiting long–term illness or disability, it is clear how close the simulated values are to the known values. The mean simulated proportion is 0.26 with standard deviation of 0.06, compared to the actual median proportion of 0.26 with standard deviation of 0.09.

The biggest discrepancy between the simulated data and the actual values is that in a small number of cases the simulation does not assign enough people with a limiting long–term illness or disability: it undersimulates. The maximum proportion assigned by the simulation is 0.53 compared to 0.72 for the actual data. However, only eight output areas—or less than 1%—have an actual maximum proportion higher than 0.53, suggesting the simulation has estimated that majority of cases of limiting long–term illness or disability well.

This is further demonstrated by the similarity of the thematic maps produced. Figure 5.14 shows the simulated proportion of individuals in each output area with a limiting long–term illness or disability. Figure 5.15 shows the actual proportion of individuals in each output area with a limiting long–term illness or disability, based on the 2011 census.

The simulation picks up high proportions of limiting long–term illness or disability in areas across the borough, including Conisbrough and Mexborough to the west, Carcroft to the north, Thorne and Armthorpe to the east, and Rossington to the south.

The simulation has picked up one output area with a high proportion of

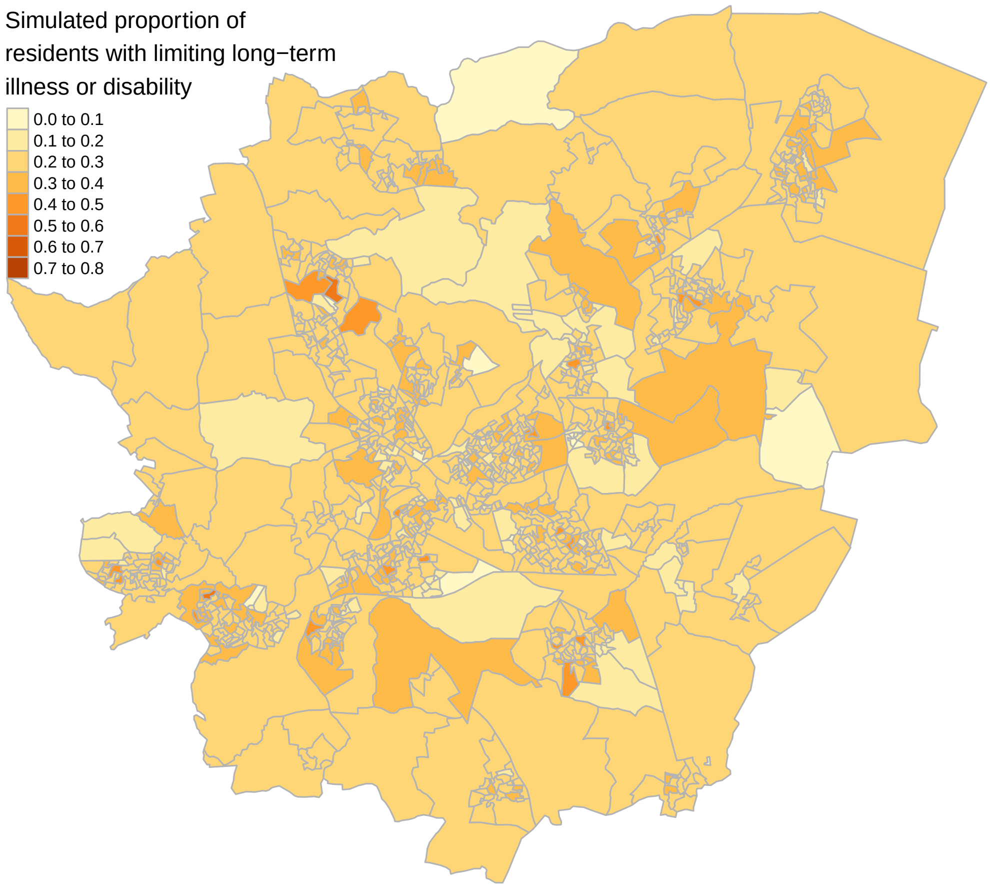

Figure 5.14: Limiting long-term illness or disability spatial microsimulation results for Doncaster output areas

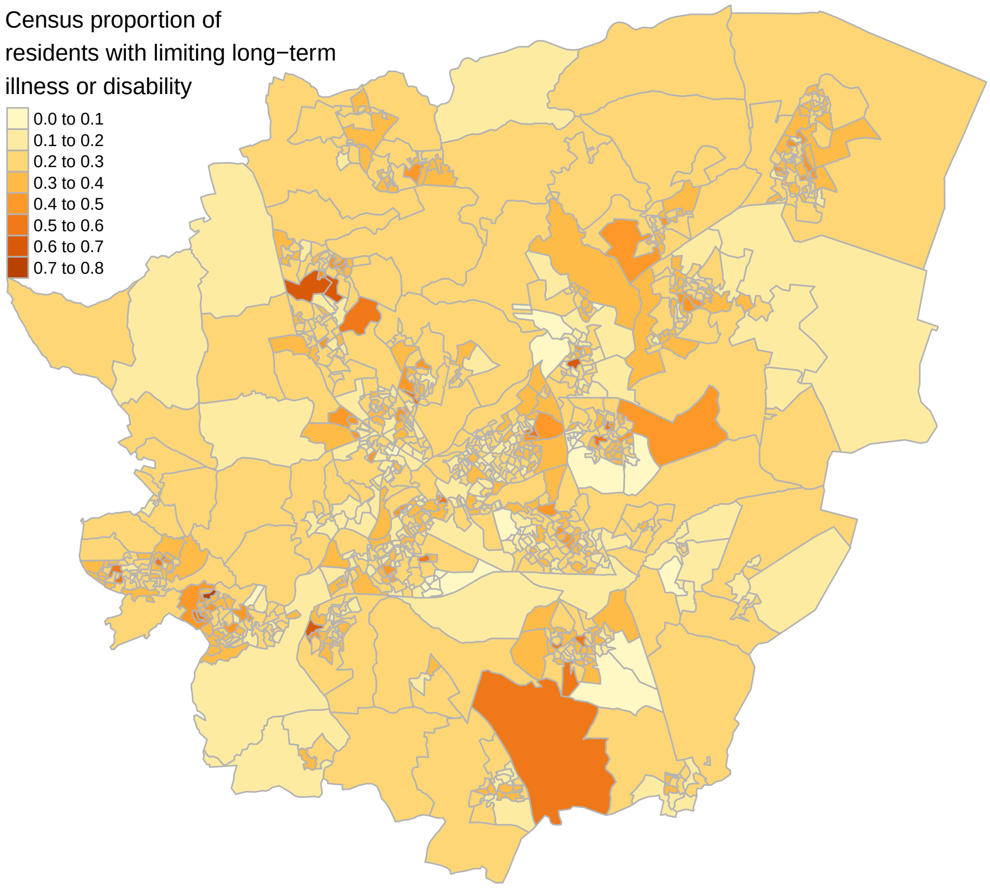

Figure 5.15: Limiting long-term illness or disability from census

limiting long–term illness or disability. This is the output area to the south of Rossington and east of the A1(M) at the council boundary. I believe this is because there is a high number of residents who have never worked or are long–term unemployed (NS–SEC 8) living in this area, 61 compared to a mean of just 16.05.

If the survey data set has a disproportionate number of people who have never worked or are long–term unemployed, but not on medical grounds, this may affect the simulation. For example, a bias in the survey sample design may have identified people who have never worked because their partner works instead which may not fit the demographic of this area.

5.15 Conclusion

This chapter is a proof–of–concept spatial microsimulation using Understanding Society (*Understanding Society*) respondents and 2011 census tables obtained from Nomisweb. It simulates if the population have or have had a health condition for each resident in Doncaster aged 16 and above. It uses a self–written package for R, `rakeR`, to perform the spatial microsimulation using the iterative proportional fitting method.

The final output of the spatial microsimulation model is a data table with one row per geographical zone and one column per variable, included the simulated variable. The results of the internal validation are encouraging and suggest this model forms a good basis to expand to include resilience and indicators of poverty, which I simulate in the next chapter.

Chapter 6

Health resilience spatial microsimulation

6.1 Introduction

After successfully simulating a pilot spatial micro dataset in Chapter 5 I moved on to simulate health resilience, which includes clinical depression and measures of deprivation, and indicators of poverty which I use to examine the likely effects of a number of local and national policy proposals in Chapter 7. This was again a simulation of Doncaster, my case study area, at output area level. To perform the simulation I used the same data sources as the pilot simulation, namely *Understanding Society* and the 2011 census tables. Where this simulation differed was in the increased number of target variables that I simulated to help identify health resilience, and in the increased number of constraint variables I used to improve the accuracy of the simulation.

For the target variables I compared two approaches to identify resilience. One approach was to simulate mental health outcomes, specifically prevalence of clinical depression, at the area level. I then combined these results with area–level deprivation measures to identify which area or areas could be considered resilient, if any. This is similar to the approach taken by much contemporary social science research into health resilience, such as that by Bartley (2006), Mitchell et al. (2009), or Cairns (2013). The other approach was to simulate variables that identify concepts thought to promote resilience, as outlined in Chapter 3. With this approach I was able to specify which areas might be resilient under certain assumptions. These two approaches are documented in Section 6.7. Finally I simulated various indicators of economic and social status, which I use to examine the possible effects of proposed national and local policy changes in Chapter 7.

For the constraints I wanted to test additional variables because more constraints can lead to a more accurate simulation, although some authors suggest the number of possible categories for each constraint is at least as important as the number of constraints themselves:

> . . . a model constrained by two variables, each containing 10 categories (20 constraint categories in total), will be better constrained than a model constrained by 5 binary variables such as male/female, young/old etc. (Lovelace and Dumont, 2016: 52).

Regardless of the efficacy of using multiple variables or multiple levels, by testing additional constraints I was able to satisfy both requirements, as many of the constraints have several response categories. Of course, the constraints are only as good as their ability to predict the target variable, so I empirically tested this relationship in Section 6.4.

6.2 Target variables

Each of the two approaches to identify resilience that I described in Section 6.1 required different target variables. The first approach identified areas as resilient if they have low prevalence of clinical depression but high area–level deprivation. I chose clinical depression as it is more closely associated with psychological resilience originated in early resilience literature, outlined in Chapter 2.1.

To calculate this I simulated the prevalence of clinical depression. In *Understanding Society* this was asked as 'Has a doctor or other health professional ever told you that you have any of these conditions?' (University of Essex et al., 2016). Respondents were asked if they had any one or more of 17 conditions, which included clinical depression. Self–reported depression has been shown to be adequately correlated with clinical records of depression (Sanchez-Villegas et al., 2008).

The second approach simulated characteristics thought to relate to higher levels of resilience, as identified by the systematic literature review described in Chapter 3. Table 3.2 outlines the characteristics identified in each paper thought to affect resilience. These included: social capital and social networks; a mentor or someone to provide support; place attachment; natural environment; being in or returning to employment, income, or social class; involvement in sports in childhood and youth; coping mechanisms; cognitive ability in childhood; behaviour change; sickness benefit provision; access to—especially primary—healthcare; demographics such as gender, age, ethnicity, and education level; congruity between individual circumstances and neighbourhood or area circumstances; absence of Adverse Childhood Experiences (ACE); parental and grandparental mental health; budgeting and money management skills; and bespoke resilience scales.

6.2.1 Social capital

Poortinga (2012) tested the role of bonding, bridging, and linking social capital in community resilience, and Cairns (2013) and Cairns–Nagi and Bambra (2013) also identify social capital as a source of resilience. Poortinga (2012) used nine variables from the 2007 and 2009 Citizenship Surveys in England to articulate social capital (Poortinga, 2012: 289–290).

The authors tested bonding social capital by asking about: the extent to which people in their respondent's neighbourhood pull together to improve the neighbourhood; how many people in the neighbourhood can be trusted; and how strongly the respondent feels they belong to their neighbourhood. In *Understanding Society* there is no exact analogy to the first question, but respondents are asked if they would be 'willing to work together with others on something to improve my neighbourhood' and if 'I think of myself as similar to the people that live in this neighbourhood'. I coded respondents who strongly agreed or agreed to both questions as a proxy for neighbourhood cohesion. Trust in people in the neighbourhood and feeling of belonging to the neighbourhood have more direct analogies in *Understanding Society*. Trust was asked as 'people in this neighbourhood can be trusted' in waves `f_` and `c_`, and as 'generally speaking would you say that most people can be trusted, or that you can't be too careful in dealing with people?' in wave `_a`. I coded neighbourhood trust as either strongly agree or agree to wave `f_` and `c_`, or 'most people can be trusted' to wave `a_`, taking the most recent response if respondents answered more than one wave. Belonging was asked as 'I feel like I belong to this neighbourhood'. I coded respondents who strongly agreed or agreed with this statement as feeling they belong to their neighbourhood.

Bridging social capital was asked by: if respondents think people from different backgrounds in their neighbourhood get on well together; if residents respect ethnic differences between people; what proportion of the respondent's friends have a similar income to them; and what proportion of the respondent's friends are of the same ethnic group as them. *Understanding Society* asks respondents to agree or disagree with the statement, 'People in this neighbourhood generally don't get along with each other'. This is reversed from the use in Poortinga (2012), but tests the same concept so I used this as a proxy. There is no direct analogy asking about respect for ethnic differences so I could not include this. Proportion of friends with a similar income and proportion of friends of the same ethnic group have direct analogies in *Understanding Society*. Poortinga (2012) suggested hetergeneous friendship groups were conducive to resilience, so I coded 'about half' and 'less than half' as hetergeneous in both cases.

Linking social capital asked: if respondents had contacted a political representative, such as a councillor or Member of Parliament, in the last twelve months; if the respondent had attended a public rally, meeting, demonstration, or protest, or signed a petition in the last twelve months; to what degree the respondent felt they could influence decisions affecting their local area; and how much they trust the local council, the police, and parliament. The first two questions ask about activities, except voting, that the respondent has participated in, for which there was no adequate analogy in *Understanding Society*, which forced me to exclude these questions from my analysis. The third question asks about the respondent's ability to influence local decisions, for which I used 'People like me don't have any say in what the government does' as a proxy. I coded respondents who strongly disagreed or disagreed as having political efficacy. Finally, levels of trust in the local council, police, and parliament were not asked so I could not use these.

6.2.2 Social networks

Reeves et al. (2014) reviewed the effectiveness of social networks for patients managing a long–term condition. They suggested network member characteristics, social network characteristics, and member change were important for effective social networks. They articulated these as: number of network members within five minutes; percentage of network members giving support within five minutes; number of network members in contact at least weekly; number of cohabitants; if the network includes a spouse or partner; number of different relationship 'types'; number of network members who know each other; amount of support given to other network members; score of social resources; extent of involvement in groups or organisations; a binary measure if any network members were lost in the previous twelve months; and number of members of the network lost in the previous twelve months.

In *Understanding Society* respondents are only asked details of up to three 'best friends' or network contacts, so it was not appropriate to use the number of contacts as this was capped. Instead, I used a binary yes or no if any one of the respondent's friends met the respective criteria. For network members within five minutes I used friends who live less than one mile away as a proxy. There was no suitable variable asking if friends provided support, so I was not able to include this. Respondents were asked how frequently they were in touch with friends, so I coded respondents as binary yes or no if they were in touch with at least one friend, at least weekly. I derived number of cohabitants by subtracting one—the respondent—from household size. Marital status was the most appropriate proxy for whether the network included a spouse or partner, but I could not include this because I included it as a constraint (see Section 6.3.3).

There were no suitable measures for number of different relationship 'types' or number of network members who know each other, so I could not include these. The paper used a count of up to seven types of support given to others by the participant in the last month, but it is not known what these seven types of support were. *Understanding Society* asks if the respondent cares for others either inside or outside of the household, but I was not able to use these responses as they might not capture all of the types of support used by Reeves et al. (2014). Social resources were assessed using the Resource Generator–UK (RG–UK) instrument (Webber and Huxley, 2007). This asked 27 items about the help available to the respondent across four domains, such as if the respondent had a friend who could help with jobs around the home or who had a professional occupation (Webber and Huxley, 2007: 486). *Understanding Society* did not ask comparable questions about the nature and extent of support provided by friends so I could not include these measures.

Extent of involvement in groups or organisations was asked as the number attended from a list of 14 different types. They did not specify what the 14 types are, but in *Understanding Society* respondents are asked if they participate in any of 16 organisations or activities. While there is no guarantee the 16 items in *Understanding Society* map to the 14 in Reeves et al. (2014), they do cover a broad range of organisations and groups and respondents are asked if they participate in any other groups not captured. I coded respondents as being involved if they participated in at least one group or organisation. *Understanding Society* does not ask if any friendships or network 'nodes' have been lost in the preceeding twelve months or about work done by lost 'nodes' in that period. I was therefore unable to include these concepts in my analysis.

6.2.3 Peer support

Matthews and Sykes (2012) found that respondents who self–reported that they had "...someone to support, push or encourage them" were more likely to look after their health and seek treatment when necessary (Matthews and Sykes, 2012: 404). *Understanding Society* asks about social networks, but not if the respondent feels they receive support from members of their network. Similarly only respondents completing the youth questionnaire—those aged 16–21—were asked if they feel they receive support from their family. For this reason I was not able to include this measure, but other measures of the quality and quantity of the respondent's social network are inlcuded based on measures in Section 6.2.1.

Robinson et al. (2015) identified peer support as a protective factor against poor health in men. As discussed above there were no suitable measures in *Understanding Society* for this concept so I was not able to include it.

6.2.4 Place attachment

Cairns (2013) and Cairns–Nagi and Bambra (2013) are based on the same doctoral research so repeat the same measures. They identify self–reported place attachment, social capital, and the quality of the natural environment as potential protective mechanisms. Place attachment was defined by the authors as “the emotional attachment acquired by individuals to their environmental surroundings which enables them to develop a strong sense of belonging, which is important for personal identity and emotional well–being” (Cairns–Nagi and Bambra, 2013: 232). *Understanding Society* asks if the respondent feels like they belong to their neighbourhood, which I already coded in Section 6.2.1.

6.2.5 Natural environment

Cairns (2013) and Cairns–Nagi and Bambra (2013) identify the quality of the natural environment as a potential protective mechanism. Bambra et al. (2015) hypothesised a reduced or limited proximity to ‘brownfield’ sites—sites that are categorised as previously developed land (PDL)—and low environmental deprivation are potential sources of health resilience. *Understanding Society* does not ask about the local environment so I was not able to include these concepts.

6.2.6 Employment status and occupational capital

Cameron (2013) found that self–reported employment status, financial situation, social isolation, ‘occupational capital’, and social support affected health outcomes. Employment status is already used as a constraint so I had to exclude it. Respondents in *Understanding Society* are asked about their current subjective financial status, so I included this as a proxy for financial situation. I coded respondents who reported they were living comfortably or doing alright as a ‘good’ financial situation and potential source of resilience. The number of close friends (which can also include family members) is asked in *Understanding Society*, so I coded respondents with one or more close friends as not socially isolated. Occupational capital is defined by the author as “accessible external opportunities” (Cameron, 2013: 197), which I take to mean as the availability or number of jobs which the candidate could reasonably perform and be appointed to within a reasonable distance. This is only applicable to individuals who are currently seeking work, mostly those who are unemployed, so is not applicable to the general population. I could not combine this in any way with employment status, either, as I used this as a constraint (see Section 6.3.1) For these reasons I excluded this from my analysis. I was not able to include social support as I discussed in Section 6.2.3.

6.2.7 Sports participation

Haycock and Smith (2014) determined that sports participation in youth had a strong association with sports participation, and therefore improved health, in adult life. In *Understanding Society* sports participation is asked, but only for the youth panel or if there is a child in the home, so it was not possible to include this measure.

6.2.8 Coping mechanisms

Lai and Oei (2014) provide a systematic review of coping mechanisms employed to mitigate stress and challenges from caregiving. As this is a review of other literature, multiple instruments were identified to measure coping ability and strategy including Coping Health Inventory for Parents (CHIP), Ways of Coping Scale (WCS), and the Multidimensional Coping Inventory (MCI), as well as qualitative and self–reported measures. *Understanding Society* does not capture this breadth of information about coping, and likely should not as many of these instruments are not designed to be self–completed. It does, however, ask the General Health

Questionnaire (GHQ) which includes items on the respondent's ability to overcome difficulties and to face problems. I used these as a proxy for 'coping' overall, although these will not articulate the nuances of *how* respondents cope. I coded 'not at all' or 'no more than usual' to problems overcoming difficulties and 'more so than usual' or 'same as usual' to ability to face problems as potentially sources of resilience.

Erskine et al. (2016) looked at the protection provided by repressive coping in old age. I cannot include detailed information about coping styles because these are not asked in *Understanding Society*. I have included the GHQ which asks about coping overall, but not about *how* the respondent copes.

6.2.9 Cognitive ability

Mõttus et al. (2012) tested the efficacy of cognitive ability, measured with the Moray House Test no. 12 (Mõttus et al., 2012: 1370), as a protective mechanism for health. I had to exclude this because there was no suitable comparable measure in *Understanding Society*.

6.2.10 Behaviour change

Mackenbach et al. (2015) describe the relationship between education and cause–specific mortality in Europe, from which mortality deviated from the 'expected' level in some circumstances. They determined that much of the deviation, particularly for preventable diseases, is due to behaviour change, medical intervention, and injury prevention (Mackenbach et al., 2015: 59). Medical intervention and injury prevention, although clearly important, are not of interest to this study because they focus on the prevention and treatment of a specific pathology or event, not on psychological or physiological improvement overall. Behaviours they identified as protective included not smoking and low alcohol consumption.

Smoking is recorded in *Understanding Society* as the usual number of cigarettes smoked per day, which I coded as either no cigarettes for non–smokers or one or more cigarettes per day for smokers. Alcohol consumption is not directly asked in *Understanding Society* but the amount of money the household spent on alcohol in the preceeding four weeks is. By dividing this figure by the average unit cost of alcohol (Institute of Alcohol Studies, 2014) I estimated the household alcohol consumption in units. Dividing this figure by four gave the weekly household alcohol consumption. I further divided this by the number of individuals living in the household aged 16 and over to arrive at an estimated consumption of alcohol per person in units. Consumption of more than 14 units per week is considered risky (Department of Health et al., 2016: 4) so I have coded respondents as low or high risk based on this threshold. This should be treated as highly indicative only as it is based on a number of assumptions, not least that all individuals within the household drink the same amount of alcohol. Parental attitudes and behavious towards alcohol consumption demonstrably influence child alcohol consumption (Nash et al., 2005; Yu, 2003) but clearly there will be variation within the household to a greater or lesser degree. There are no analogies for diet and exercise in *Understanding Society* so I have had to exclude these.

6.2.11 Sickness benefit arrangements

Wel et al. (2015) compare sickness benefit arrangements across Europe and their effect on health inequalities. Sickness benefit is an important safety net, potentially applicable to any and all employed individuals. *Understanding Society* asks if the respondent is usually employed but on sick leave in the last week, but does not include details of any amounts paid because of sick leave. Further, sickness benefit will only apply to respondents who are employed which accounts for only about 46% of the sample. Employment status is a constraint, so I was not able to combine this with sickness benefit provision to create a measure for the whole population. For these reasons I was not able to include sickness benefit in my analysis of resilience.

6.2.12 Accessing health care

Mastrocola et al. (2015) identified barriers women involved in street–based prostitution face in accessing health care, especially primary care, and suggest that improved access would be a protective factor for these women. Respondents in *Understanding Society* are asked if they experienced any difficulties accessing local services, but this is grouped together as one question which includes healthcare, food shops, and learning facilities. I was therefore not able to include this measure as there was no way to differentiate between access to health care services and all other services.

6.2.13 Personal and area demographics

Glonti et al. (2015) is a systematic review of health resilience during economic crises across ten countries. Extracting just the UK–based papers, the sources of resilience were, variously, gender, age, education level, employment, financial constraints, and low area–level deprivation. I could not include gender, age, education level, and employment because they are already included in the simulation as constraints. *Understanding Society* asks about subjective financial situation which I used as an indicator for financial constraints, as I coded in Section 6.2.6. Area–based methods of deprivation, such as IMD score, are not recorded in *Understanding Society* but I attached these to the aggregated simulation.

6.2.14 Neighbourhood congruity

Albor et al. (2014) tested to see if sharing a similar socio–economic status to other residents in the neighbourhood—neighbourhood congruity—can be a source of health resilience. Individual socio–economic status was derived from household occupational class and educational achievement, and neighbourhood socio–economic status was based on census occupational status and educational status. *Understanding Society* asks if respondents agree or disagree that they are similar to others in their neighbourhood, which is what I based neighbourhood congruity on. I was not able to include occupational status or educational status as they are both constraints.

6.2.15 Adverse Childhood Experiences (ACEs)

Bellis et al. (2014) explored the association between adverse childhood experiences (ACEs) and health–harming behaviours, specifically if an absence of ACEs can lead to resilience. Respondents were asked about ACEs using the Centers for Disease Control and Prevention short ACE tool which covered: physical, verbal, and sexual abuse; parental separation; exposure to domestic violence; or growing up in a household with mental illness, alcohol abuse, drug abuse, or incarceration (Bellis et al., 2014: 3). I was not able to include ACEs as *Understanding Society* does not ask respondents about household conditions during childhood or adolescence, but I was able to code household alcohol consumption (Section 6.2.10).

6.2.16 Familial mental health

Johnston et al. (2013) used the 1970 British Cohort Study to test if parental or grandparental mental health affected the mental health of the grandchild. Childhood or adolescent household conditions were not asked of respondents in *Understanding Society* so I was therefore unable to include parental or grandparental mental health. I was able to include an indicator for the respondent's mental health using the General Health Questionnaire (GHQ).

6.2.17 Financial and budgeting skills

Fenge et al. (2012) used semi–structured interviews to explore older peoples' resilience to the effects of economic recession, specifically if budgeting and money management skills enabled them to maintain their well–being and quality of life. *Understanding Society* asks respondents if they save any money, which is a binary response, and about the respondent's subjective financial situation, which I have already coded in Section 6.2.6.

6.2.18 Resilience scale (RS–25)

Sull et al. (2015) used the Resilience Scale (RS–25) to measure resilience among NHS workers which tests concepts of "a purposeful life, perserverance, equanimity, self–reliance and existential aloneness" (Sull et al., 2015: 3). The RS–25 is a proprietary measure of resilience marketed as the 'True Resilience Scale' which can be licensed for use from The Resilience Centre (The Resilience Centre, 2017). I contacted The Resilience Centre by email in April 2017 asking to see the items on the RS–25, explaining the nature of this research and that I did not intend to use the resilience scale in a clinical or organisational setting. After repeated emails (Wagnild, 2017) The Resilience Centre did not provide the items, so I could not include them. The RS–25 instrument might be valid but is of limited use for policy or research if it cannot be reviewed by other researchers.

Table 6.1 summarises the concepts and variables I used to operationalise these.

6.3 Constraints

In selecting constraints I began with those I used in the pilot simulation (see Section 5.5). These constraints simulated limiting long–term illness

Table 6.1: Operationalisation of resilience sources

Paper	Original measure	Understanding Society Variable
13	Neighbourhood cohesion	scopngbhe; scopngbhg
	Neighbourhood trust	nbrcoh3; sctrust (wave a_ only)
	Neighbourhood belonging	scopngbha
	Civic participation	orga
	Social cohesion	nbrcoh4 (reversed)
	Mutual respect	No suitable measure
	Heterogeneous relationships	siminc
		simrace
	Political participation	No suitable measure
	Political activism	No suitable measure
	Political efficacy	poleff4 (reversed)
	Political trust	No suitable measure
16	Cognitive ability	No suitable measure
23	Support/encouragement	No suitable measure
32	Place attachment	As neighbourhood belonging
	Social capital	As paper 13
	Natural environment	No suitable measure
37	Employment	Excluded - constraint
	Finances/income	finnow
	Social isolation	closenum (>0)
	Occupational capital	No suitable measure
	Social support	No suitable measure
46	Place attachment	As paper 32
	Social capital	As paper 32
	Natural environment	No suitable measure
67	Sport involvement in youth	No suitable measure
78	Coping strategy	GHQ
90	Smoking	ncigs (0)
	Alcohol consumption	xpaltob_g3 (household measure)
	Diet	No suitable measure
	Exercise	No suitable measure
96	Sickness benefit provision	Excluded - not generally applicable
98	Peer support	No suitable measure
173	Repressive coping	No suitable measure
195	Access to healthcare	Excluded - not generally applicable
204	Greater distance to brownfield	No suitable measure
	Low environmental deprivation	No suitable measure
206	Resilience Scale (RS-25)	Items not provided
208	Gender	Excluded - constraint
	Age	Excluded - constraint
	Education level	Excluded - constraint
	Employment	Excluded - constraint
	Financial problems in last year	As paper 37
	Area-level deprivation	No suitable measure
241	Number of 'nodes' within 5 minutes	netlv (< 1 mile)
	Support from network	No suitable measure
	Frequent contacts (>1/week)	netph (at least weekly)
	Number of cohabitants	hhsize
	Binary: network include spouse/partner	Excluded - constraint
	Number of different relationship 'types'	No suitable measure
	Number of network pairs who know each other	No suitable measure
	Support given to others	Unknown types of support
	Social resources	No suitable measures
	Involvement in groups or organisations	As paper 13
	Binary: network member lost in last 12 months	No suitable measure
	Total network members lost in 12 months	No suitable measure
242	Similarity with area status	scopngbhg
250	Adverse Childhood Experiences (ACEs)	No suitable measures
272	Parental and grandparental mental health	No suitable measures
307	Budgeting/money management skills	finnow; save

or disability well because they correlated well with this variable, and my aim here was to simulate similar health–related variables. The constraints I used were sex, highest qualification, ethnicity, housing tenure, car ownership, and age.

In addition to these I wanted to test an increased number of constraints, now I had a working model; as in the pilot simulation (Chapter 5) I was limited by the variables that are available in both the census and the survey data, which in practice usually means the census was the limiting factor. Nevertheless the census contained additional variables that I tested for inclusion in the simulation. These were: economic activity; overcrowding (greater than 1.0 person per room, as described by Townsend et al. (1988)); marital status; and social class.

6.3.1 Economic activity

The first additional variable I tried was economic activity, as this is a powerful predictor of many health outcomes (Bartley et al., 2003, 2006). Most levels matched across both the survey and the census data, but a few required recoding or re–aggregating.

Economic activity data in the census covered only individuals aged 16–74 whereas *Understanding Society* covered all individuals aged 16 and above. To solve this issue in setting up the census I added all individuals aged 75 and above from the census to the 'retired' category. This was the most pragmatic choice as, even though some individuals aged 75 and above may still be working, especially in part–time or informal capacities, the majority will have left the primary employment or career which influenced their social class.

An option for maternity leave was present in the survey data but not in the census data so I needed to choose the most suitable group to combine this with. Similarly apprenticeships, government training schemes, and 'unpaid worker in family business' were options in the survey data but not in the census. I ultimately decided that because apprenticeships and government training schemes were conceptually similar I would combine these into ‘other’ in both the census and survey levels.

Combining government training scheme and apprenticeship with unpaid worker in a family business was not ideal as they are conceptually different forms of economic activity. However, only a small number of respondents in *Understanding Society* were unpaid workers in a family business ($n = 48$) so the effect was negligible, so the ‘other’ group could be thought of as mostly comprising individuals on training schemes designed to enhance their skills and improve their careers.

Because of this, it did not seem appropriate to include people on maternity leave in the ‘other’ group, as women on maternity leave can choose to return to their previous role and economic activity. I considered grouping maternity leave and long–term sick and disabled together in the survey, as both groups have ‘paused’ their previous economic activity. However, maternity leave comes with an expectation that the individual returns to their previous economic activity within a defined period, usually twelve months. Individuals who are long–term sick or disabled and receiving a personal independence payment (PIP) must have a condition expected to last at least nine months, but in practice there is no maximum length of time people can claim for before returning to their previous economic activity as they are ‘regularly reassessed’ (Gov.UK, 2017).

I ultimately decided to group individuals on maternity leave with individuals looking after family or home. This has the same issue that those on maternity leave are likely to return to their ‘previous’ economic activity while those looking after the family or home or those who are long–term sick or disabled are more likely to remain so. It has the advantage, though, of the two being conceptually similar involving care for family members. In addition, *by definition*, people with a long–term illness or disability will necessarily have a health issue, while both those on maternity leave and those looking after family or home may or may not have a health issue: a health issue is not *a priori* known for these individuals. Finally, it preserves the distinction between individuals who are fundamentally performing a caring role to those who are receiving formal training.

In the census students are split between those who are economically active and those who are economically inactive, which is usually students who are studying full–time. In *Understanding Society* students are not distinguished in this way, so it was necessary to group economically active and economically inactive students in the census. Even though economically active students may not be full–time students, or may participate in the labour market in other ways, their primary economic activity is arguably studying to improve their skills so the two are conceptually similar.

The census splits self–employed groups by part–time and full–time, and those with employees and those without employees. These had to be aggregated to match the survey, which had a single category for self–employed. Similarly full–time and part–time employed individuals in the census were aggregated—to simply 'employed'—to match the survey. *Understanding Society* does not explicitly state the 'unemployed' group is the same as 'economically active unemployed' from the census. To be 'economically active unemployed' requires the individual to be "actively looking for work" or "waiting to start a new job" (Nomis, 2013b), while *Understanding Society* instead asks respondents to choose the economic activity that 'best' describes their current circumstances. Again, I do not believe this will affect the simulation significantly as they fundamentally measure the same concept; an individual looking to return to some other form of economic activity, be that employment, self–employment, or studying.

The final levels for economic activity in the census and the survey I used are: employed; looking after home or family; long–term sick or disabled; retired; self–employed; student; unemployed; and other. These are coded in `data-raw/0-prep-understanding-society.R` in the thesis source code.

6.3.2 Overcrowding

The concept of 'overcrowding' is based on the definition used by Townsend et al. (1988: 36–37) in their construction of a deprivation index. A private household is considered overcrowded if there is more than one person per room in the household. The definition of room excludes bathrooms, toilets, halls or landings, rooms that can only be used for storage, or any rooms shared between different households. All other rooms, including kitchens and utility rooms, are included. If two rooms have been converted in to one room they are counted as one room (Nomis, 2014).

Unfortunately it proved impossible to use overcrowding as a constraint variable. The data is available in the census for households or individuals, but crucially only for the whole population: it is not possible to obtain persons per room with an associated age breakdown. This makes it impossible to subset the data and remove individuals aged less than 16 from the census tables so there are approximately 50,000 'extra' individuals.

Arguably I could reweight the overcrowding population using the respective proportions to that of the known population that is 16 and above, as I did for car ownership (Section 5.5.3). The discrepancy for car ownership was approximately $5,000$ individuals, or approximately 2.1%, so the reweighting had a much smaller effect on the data than reweighting $50,000$ individuals would. This is additionally problematic because children are not randomly distributed among households that are overcrowded and those that are not. A hypothesis test using logistic regression with data from *Understanding Society* indicates that the number of children in the household and overcrowding are correlated (Nagelkerke pseudo–$R^2 = 0.33$, model χ^2 p–value ≈ 0). This would not be the case if families with more children had access to larger houses, but clearly something—perhaps income or availability of suitable housing stock—is preventing many families with children from moving into suitably–sized accommodation.

For these reasons I decided recalculating the populations was not appropriate and chose not to include overcrowding, or persons per room, as a constraint. This is unlikely to pose an issue for the simulation, however, because other constraints capture different dimensions of reduced material or economic circumstances or deprivation, which overcrowding is associated with.

6.3.3 Marital status

Evidence suggests marital status is associated with health outcomes (Hosseinpour et al., 2012; Robards et al., 2012), although not conclusively (Sacker et al., 2009), and not always equally across social class (Choi and Marks, 2013).

For the most part, levels recorded in *Understanding Society* closely matched those in the census. There were levels for married, in a civil partnership, single, separated, divorced, or widowed, and these required no additional matching. For respondents in *Understanding Society* there were additional levels for separated from a civil partnership, divorced from a civil partnership, or a surviving partner in a civil partnership. I simply combined these with separated, divorced, or widowed, respectively and there were relatively small number of respondents in a civil partnership so this did not affect the simulation.

6.3.4 Social class

Socio–economic position or social class is another powerful determinant of health. Social class is usually measured using the National Statistics Socio–economic Classification (NS–SEC) (Office for National Statistics, 2015).

There were a large number of missing cases for social class in *Understanding Society* (missing $n = 30,979$). To help in deciding whether to remove or include social class I ran a logistic regression test to see if NS–SEC is useful in predicting limiting long–term illness or disability, as a proxy for a health outcome. The model was statistically significant ($p \approx 0.01$) but the predictive power was negligible (Nagelkerke pseudo–$R^2 \approx 0$), the difference in deviances was small (19.63), and none of the levels of the variable were statistically significant. The poor predictive power of social class and the fact that there were so many missing data points led me to exclude this variable from the simulation. I did not consider this a significant problem as I was able to include education in the model which is arguably a more robust measure. Because highest level of education is generally 'fixed' there is no problem of 'reverse causality', making it clearer if poor health in old age affects socio–economic position, or if socio–economic position negatively affects health.

6.3.5 Final constraint choice

After excluding social class and overcrowding, the final list of constraints I tested were: age; sex; ethnicity; marital status; highest qualification; economic activity; car ownership; and housing tenure.

6.4 Empirically test constraints

In this section I tested the constraints to see if they correlated with clinical depression. Respondents in *Understanding Society* are asked if they have a broad range of health conditions, including clinical depression, and responses are coded as 'yes' or 'no'. Of the $81,540$ respondents in *Understanding Society*, $3,532$ reported having clinical depression.

As with the pilot microsimulation the dependent variable is binary, so logistic regression is the most appropriate technique to establish correlation between the constraints and depression. I set up an initial model using age, sex, ethnicity, marital status, highest qualification, car ownership, housing tenure, economic activity, and limiting long–term illness or

Table 6.2: Overall results of depression model

diff_deviance	diff_df	chisq_prob	cox_snell	nagelkerke	hosmer
3206	38	0	0.06	0.17	0.14

disability as independent variables. Clinical depression, with 'no clinical depression' coded as the base category, was the dependent variable. The overall results of this model are displayed in table 6.2.

The AIC of the model (19155.95) is less than the AIC of the baseline (22285.66) so the model overall predicts depression (difference in deviances = 3205.71, Nagelkerke pseudo–R^2 = 0.17, $p \approx 0$). The breakdown of individual results are provided in table 6.3.

The odds ratios suggest all age groups except age 85–89 are statistically significantly more likely to have clinical depresseion than respondents aged 90 and over. The odds of having clinical depression increase from age 16–17 to their peak between ages 25–44, then decline again with age to their lowest at age 85 and above. The increase in odds to age 44 might be a result of cumulative exposure to evironments and events that contribute to clinical depression. After this age the decreasing likelihood of clinical depression may be a genuine change so that older people 'recover' from or are otherwise resistant to clinical depression. It may also be a cohort effect such that older generations are less likely to report or seek diagnoses for mental illness.

Sex is statistically significant, with males less likely than females to have a diagnosis of clinical depression. Most levels of ethnicity were statistically significant compared to the reference group of White British; only the Irish ethnic group was not statistically significant. White British respondents are the most likely to have clinical depression, with all other ethnic groups having lower odds. Black African or Black Caribbean British respondents were less than half as likely to have clinical depression that White British respondents. These are consistent with the findings of the limiting long–term illness or disability model in Section 5.6.

Table 6.3: Individual results of depression model

predictor	beta	p_value	sig	lower_ci	odds_ratio	upper_ci
(Intercept)	-5.05	0.00	**	NA	NA	NA
age_16_17	0.98	0.02	*	1.20	2.67	5.97
age_18_19	1.64	0.00	**	2.67	5.14	10.43
age_20_24	1.84	0.00	**	3.48	6.28	12.12
age_25_29	2.08	0.00	**	4.47	7.99	15.28
age_30_44	2.04	0.00	**	4.38	7.69	14.51
age_45_59	1.77	0.00	**	3.37	5.88	11.02
age_60_64	1.51	0.00	**	2.62	4.54	8.48
age_65_74	1.04	0.00	**	1.66	2.82	5.17
age_75_84	0.58	0.05	*	1.04	1.78	3.29
age_85_89	0.36	0.28		0.76	1.43	2.83
sex_male	-0.47	0.00	**	0.58	0.63	0.68
eth_irish	-0.15	0.36		0.61	0.86	1.18
eth_other_white	-0.46	0.00	**	0.48	0.63	0.82
eth_mixed_multiple_ethnic	-0.31	0.03	*	0.54	0.73	0.96
eth_asian_asian_british	-0.72	0.00	**	0.41	0.49	0.58
eth_black_african_caribbean_british	-0.97	0.00	**	0.30	0.38	0.47
eth_other_ethnicity	-0.67	0.00	**	0.35	0.51	0.72
mar_civil_part	0.50	0.20		0.72	1.64	3.28
mar_divorced	0.16	0.03	*	1.02	1.18	1.36
mar_married	-0.18	0.00	**	0.74	0.84	0.94
mar_separated	0.33	0.00	**	1.13	1.39	1.71
mar_widowed	0.06	0.60		0.84	1.06	1.33
qual_2	0.21	0.00	**	1.08	1.24	1.42
qual_3	0.30	0.00	**	1.16	1.35	1.56
qual_4_plus	0.26	0.00	**	1.13	1.29	1.49
qual_other	0.22	0.01	**	1.06	1.24	1.46
car_1	-0.13	0.02	*	0.79	0.88	0.98
car_2_plus	-0.27	0.00	**	0.67	0.76	0.86
ten_owned_outright	-0.43	0.00	**	0.57	0.65	0.73
ten_owned_mortgage_shared	-0.36	0.00	**	0.62	0.70	0.78
eca_homefam	0.41	0.00	**	1.29	1.50	1.75
eca_ltsick	1.23	0.00	**	2.98	3.43	3.95
eca_other	0.18	0.40		0.77	1.20	1.79
eca_retired	0.29	0.01	**	1.08	1.33	1.64
eca_selfemp	0.18	0.05		1.00	1.19	1.42
eca_student	-0.20	0.09		0.64	0.82	1.03
eca_unemp	0.67	0.00	**	1.68	1.95	2.25
llid_yes	1.47	0.00	**	3.95	4.34	4.76

Respondents who are married were less likely to have clinical depression compared to those who were single and never married. Respondents who were divorced or separated were more likely to have clinical depression than those who were single and never married. Respondents in a civil partnership and who were widowed were not not statistically significantly different to the reference group (single), suggesting similar levels of clinical depression. The confidence intervals for the odds for civil partnership are wide, perhaps because of the small number of respondents in a civil partnership ($n = 147$).

Interestingly, respondents with any level of qualification were *more* likely to have clinical depression than those with no qualifications. This could be because individuals with qualifications may be more likely to know of services available or more willing to obtain an appropriate diagnosis in order to obtain support.

Individuals from households with at least one car were less likely to have clinical depression than the reference group (no car), with decreasing odds ratios for individuals from households with more cars. Home owners, either those who owned their home outright or with a mortgage, were less likely to have depression than individuals who rent their homes (the reference group).

These suggest that increased financial means are associated with lower risks of clinical depression. This is supported by the fact that employed respondents are least likely to have clinical depression compared to other statistically significant levels of economic activity. Respondents looking after the home or family, who are long–term sick, retired, or unemployed are all more likely to have clinical depression than employed respondents. Respondents who are self–employed or who are students have similar levels of clinical depression to employed respondents.

Limiting long–term illness or disability is also correlated with clinical depression. The correlation is not high (pseudo–$R^2 = 0.08$), but it does suggest that: either some people have depression severe enough for them to consider it ‘limiting’; or that some people have a different limiting condition with clinical depression as a co–morbidity; or both.

Overall these variables correlated meaningfully with clinical depression, so I was able to use them as constraints for the spatial microsimulation model.

6.4.1 Constraint order

As seen in Section 5.6 the order the constraints were entered into the model made negligible differences to the outcome. I used the absolute β values to guide the order I entered the constraints into the model, although a number of random orders converged on the same result. The final order of entry I used was: car ownership, housing tenure, highest qualification, marital status, economic activity, sex, ethnicity, and age.

6.5 Weight

Weighting was performed with the `rakeR` package. I ordered the constraints as specified in Section 6.4.1 in both the census and survey and then checked for compatibility using `rakeR::check_constraint()`. I produced the fractional weights using the iterative proportional fitting algorithm (Section 5.10), as was the case for the pilot simulation. For this I used the `rakeR::weight()` function. I then ‘extracted’ the weights to produce aggregate results for each variable in each zone with `rakeR::extract()`. I integerised the weights to use as case studies in Section 7.3, but I used the extracted weights in most of my analysis because I do not need cases to use in a subsequent agent–based or dynamic model. As demonstrated in Section 5.12 the fractional weights are also slightly more accurate than the integerised weights. Figure 6.1 shows the initial results of simulated clinical depression by output area in Doncaster. Output areas with significant prison populations have been removed as

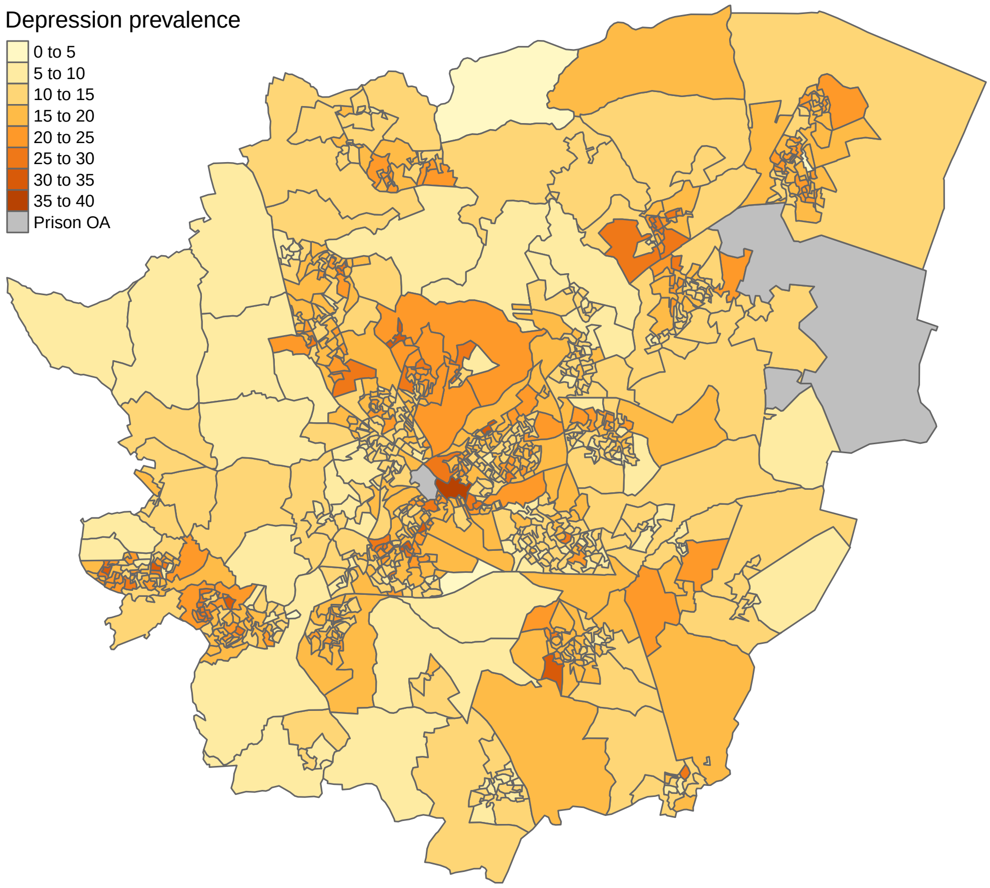

Figure 6.1: Simulated clinical depression prevalence in Doncaster

discussed in Section 5.3.2, and are displayed in grey.

6.6 Validate

As with the pilot simulation, it is possible to statistically compare the simulated constraints with the actual, known constraints to internally validate the accuracy of the model. This will involve an assessment of: correlation; a two–sided, equal variance t–test; total absolute error and standardised absolute error of the model overall; and standardised absolute error for each zone.

6.6.1 Correlation

The simulated population $(244,909)$ matched the actual population $(244,909)$ exactly, indicating the simulation constrained accurately overall.

This was further confirmed by the correlation statistic, which is a standardised statistic so a value of 1.0 is ideal. The correlation statistic was 1, indicating the population simulated in each area accurately matched the respective known population.

Figure 6.2 compares the simulated population against the actual, known population for each output area. The simulated populations were a perfect match with their known counterparts, indicating that each individual area simulated accurately.

In addition to the overall plot for each area shown in figure 6.2, I created a plot for each level of each variable for inspection. These all demonstrated the same high level of fit as the overall area plot, further indicating the model simulation was accurate. These figures are not displayed here to avoid repetition, as they all show essentially the same relationship, but can be found in the `figures/cache/` directory of the thesis source code if required.

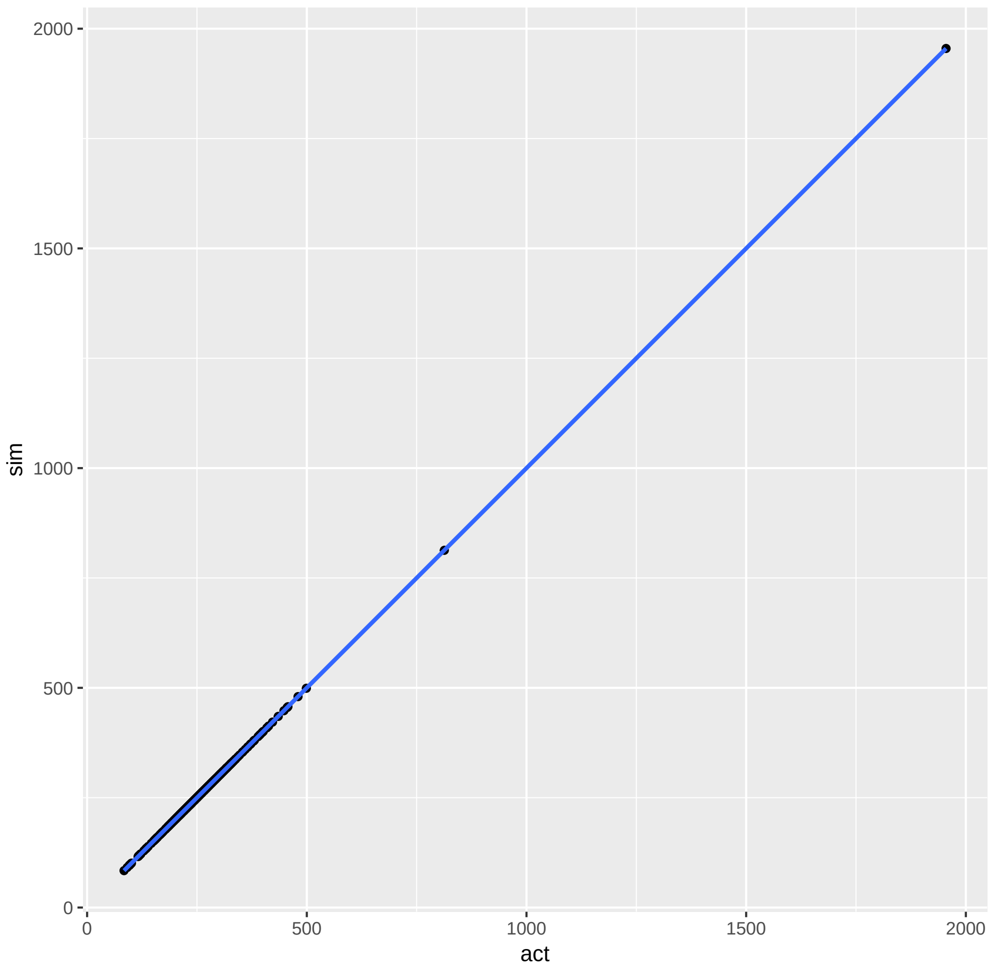

Figure 6.2: Actual population against simulated population by output area

6.6.2 t–test

Table 6.4 shows the results of the equal variance, two–sided t–test for each constraint. This statistically compares the simulated value with the actual, known value from the census and tests the null hypothesis that the two distributions are not different. In all cases the result of the t–test was not statistically significant so we accept the null hypothesis that the two distributions are not statistically different. This indicates the simulation was a good fit with the census data.

6.6.3 Total absolute error

The total absolute error and the standardised absolute error were both overall ≈ 0. Together, these indicate the model overall simulated very well as the differences between the simulated and the observed data are negligible, and certainly well within the thresholds suggested by Smith et al. (2009: 1256) discussed in Section 4.8.

6.6.4 External validation

By aggregating the simulated values for clinical depression I was able to determine the total simulated prevalence for the Doncaster local authority area. I then compared this aggregated value against a known value to provide reassurance that the simulation was realistic and plausible. These values were unlikely to match precisely because of differences in the populations and because I had to exclude output areas whose population was predominantly prisoners.

The population of the simulation was individuals aged 16 and above as this is based on the sample of individuals in *Understanding Society*. The measures from Public Health England (PHE) only include those aged 18 and over. I also had to exclude three output areas had a population consisting predominantly of prisoners. The prison population was $2,522$

Table 6.4: Result of t-tests comparing simulated against actual data

variable	statistic	p_value
car_0	-0.01	1.00
car_1	0.00	1.00
car_2_plus	0.00	1.00
ten_owned_outright	-0.02	0.99
ten_owned_mortgage_shared	0.02	0.99
ten_rented	0.00	1.00
qual_0	-0.01	0.99
qual_2	0.00	1.00
qual_3	0.02	0.99
qual_4_plus	0.00	1.00
qual_other	0.00	1.00
mar_civil_part	0.00	1.00
mar_divorced	0.00	1.00
mar_married	0.00	1.00
mar_separated	0.01	0.99
mar_single	0.01	0.99
mar_widowed	-0.03	0.97
eca_emp	-0.12	0.90
eca_homefam	-0.08	0.94
eca_ltsick	-0.05	0.96
eca_other	-0.01	1.00
eca_retired	0.30	0.76
eca_selfemp	-0.04	0.97
eca_student	-0.21	0.84
eca_unemp	-0.08	0.94
sex_female	0.03	0.97
sex_male	-0.01	0.99
eth_british	0.01	0.99
eth_irish	0.00	1.00
eth_other_white	-0.02	0.99
eth_mixed_multiple_ethnic	-0.02	0.99
eth_asian_asian_british	-0.02	0.99
eth_black_african_caribbean_british	-0.01	0.99
eth_other_ethnicity	-0.01	0.99
age_16_17	0.00	1.00
age_18_19	0.00	1.00
age_20_24	0.00	1.00
age_25_29	0.00	1.00
age_30_44	0.00	1.00
age_45_59	0.00	1.00
age_60_64	0.00	1.00
age_65_74	0.00	1.00
age_75_84	0.00	1.00
age_85_89	0.00	1.00
age_90_plus	0.00	1.00

in 2011, and it is likely a substantial proportion of these individuals will have clinical depression.

Data from Public Health England (2016) provides the prevalence of depression in the Doncaster clinical commissioning group (CCG) area for patients registered with a GP aged 18 and over, for the years 2011–12 to 2015–16. The clinical commissioning group area is coterminous with the local authority boundaries in Doncaster, so the two could be compared directly.

Based on the results of my simulation, the number of people in Doncaster with clinical depression was $15,288$, or approximately 6.24% of the overall population aged 16 and above.

The 'known' prevalence of clinical depression was 12.8% in 2011–12 for Doncaster CCG. I used the 2011–12 prevalence because the simulation was constrained by census data from this year. The population aged 18 and above in Doncaster was $237,002$ in 2011, so the prevalence of depression was approximately $30,336$ individuals.

On face value this indicated the model only simulated about half the cases of clinical depression. A more careful examination of the PHE data suggested the 2011–12 data point was problematic and the simulation was more accurate than initial inspection suggested. I believe the 'known' prevalence provided by Public Health England (2016) for 2011–12 is inconsistent with the data from the surrounding time points, suggesting this data point could be spurious.

Figure 6.3 depicts the trend in clinical depression prevalence in Doncaster and the Yorkshire and The Humber region between 2009–10 and 2015–16 (Public Health England, 2016). This trend data indicates that the prevalence of clinical depression in Doncaster in 2012–13 was only 6.1%, less than half that of the 2011–12 figure. This figure is more congruous with subsequent years, for which the prevalence of clinical depression increased to 8.2% by 2015–16. The 2011–12 prevalence figure therefore seems at odds with later data points.

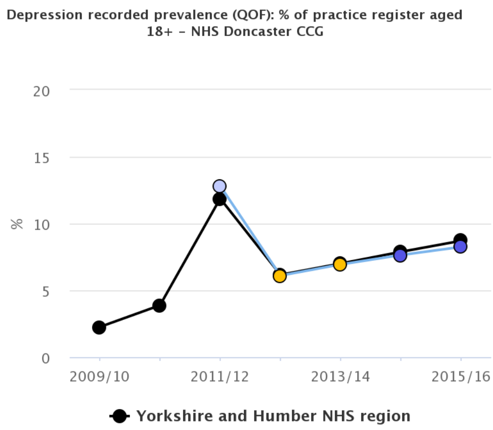

Figure 6.3: Prevalence of clinical depression in Doncaster (blue) and the Yorkshire and The Humber region (black), source: Public Health England (2016)

Data before 2011–12 for Doncaster is not provided, but data for the Yorkshire and The Humber region suggest the prevalence of clinical depression prior to 2011–12 was less than 5.0%. This is congruous with 2012–13 and later data, further suggesting the 2011–12 figure is anomalous.

One possible explanation for this discrepancy is the Quality and Outcomes Framework (QOF), "...the annual reward and incentive programme detailing GP practice achievement results" (NHS Digital, 2016), changed between 2010–11 and 2011–12. Indicators for clinical depression—DEP2/DEP4 and DEP3/DEP5—were changed to be worth fewer 'points', potentially affecting the measurement and reporting of this diagnosis (NHS Employers, 2011–2012: 3).

For this reason I believe it is likely that the prevalence of clinical depression is closer to 5–6% than the chart initially suggests. This would be the approximately prevalence if the 2011–12 data point was removed and the trend used instead. This places my simulated results in line with the surrounding data, suggesting they are plausible and certainly more likely to be valid than initial comparison to 'known' data suggested.

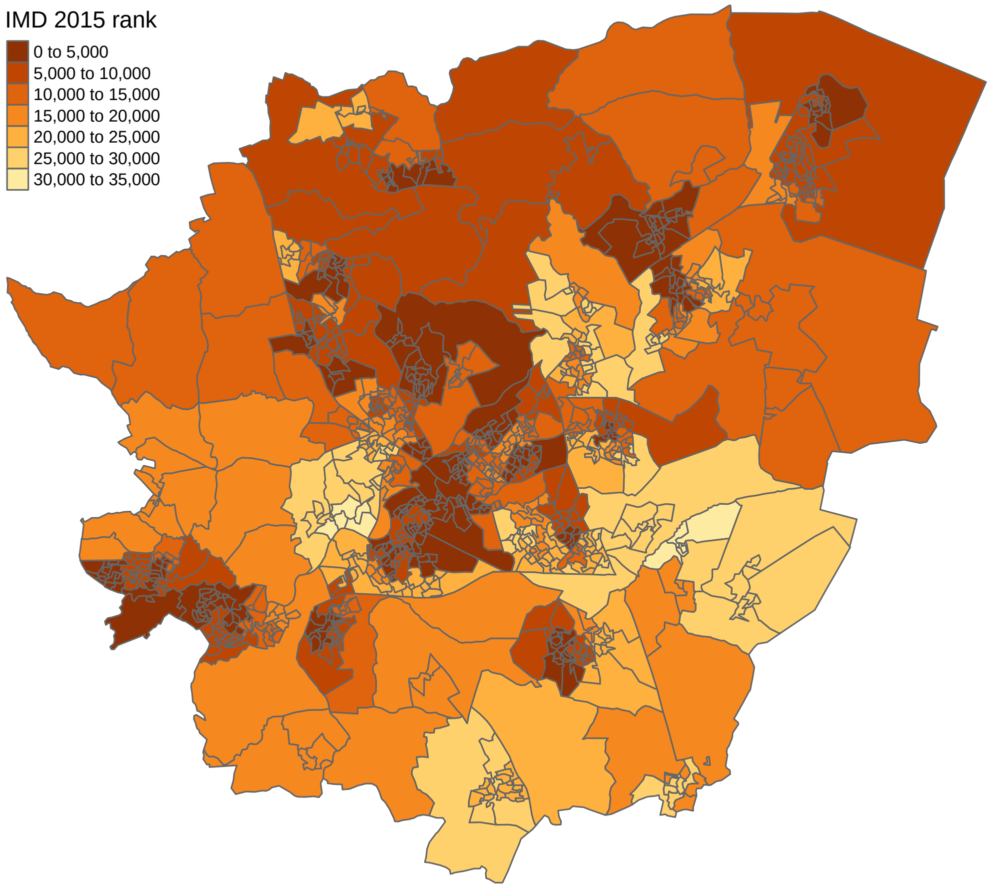

Figure 6.4: Doncaster IMD 2015 rank (lower rank is more deprived)

6.7 Results

6.7.1 Resilience

Having simulated and validated prevalence of clinical depression I compared this with various indicators of area–based socio–economic deprivation. These were: unemployment; long–term unemployment; low–grade employment (routine employment, NS–SEC 7); index of multiple deprivation (IMD) score; and output area classification supergroup 'hard–pressed living'.

Deprivation based on unemployment, long–term unemployment, and low–grade employment were calculated by summing the number of individuals in each output area matching these criteria and selecting the areas with the highest number of these individuals.

The 2015 Index of Multiple Deprivation (IMD) is provided for lower layer super output areas (LSOAs), but not output areas directly. An official tool to lookup the IMD score for individual postcodes is provided by Swirrl IT Ltd. and Department for Communities and Local Government (2017), so it is possible to use indices of multiple deprivation scores at geographies smaller than the LSOAs provided. For each LSOA I applied the overall LSOA score to each of its constituent output areas, then selected the lowest ranks as the most deprived areas of Doncaster. Figure 6.4 shows the IMD score for each output area in Doncaster, with lower scores representing higher deprivation.

Areas classified as being in the 'hard–pressed living' supergroup are used to identify high deprivation areas using the output area classification system. These areas are indicative of higher rates of social renting, lower rates of higher–level qualifications, and unemployment rates above the national average (Office for National Statistics, 2015d: 19). Figure 1.4 shows the output area classification supergroup of Doncaster output areas.

I considered output areas as 'resilient' if they had both high deprivation, using the indicators described above, and low prevalence of clinical depression. To determine what to classify as 'low' and 'high' I tested a number of thresholds from 20% to 40% of respondents being both clinically depressed and being in the highest deprivation classification. Table 6.5 summarises the results of these tests.

Selecting a threshold will always include an element of subjective choice and is arguably more an art than a science. There are two properties that I used to help guide my decision in selecting a threshold, however. First, resilience is, by definition, an outlying phenomenon so a threshold should mark a relatively small number of areas as resilient. Second, I suggest it is desirable if a threshold does not treat too many cases as 'high' deprivation or 'low' health, as it is important for these to remain

Table 6.5: Number of resilient areas by area-based deprivation measure

Area-based deprivation measure	Threshold (%)	Number of resilient areas
High unemployment	20	0
High long-term unemployment	20	0
High low-grade employment	20	0
'Hard-pressed living'	20	0
IMD score	20	2
High unemployment	25	0
High long-term unemployment	25	1
High low-grade employment	25	0
'Hard-pressed living'	25	0
IMD score	25	5
High unemployment	30	2
High long-term unemployment	30	1
High low-grade employment	30	0
'Hard-pressed living'	30	0
IMD score	30	8
High unemployment	33	4
High long-term unemployment	33	1
High low-grade employment	33	4
'Hard-pressed living'	33	0
IMD score	33	13
High unemployment	40	19
High long-term unemployment	40	15
High low-grade employment	40	11
'Hard-pressed living'	40	0
IMD score	40	35
total	NA	4
depr	NA	0

differentiated from ‘background’ cases.

After testing, thresholds of 20%, 25%, and 30% resulted in very few ‘resilient’ areas, sometimes none at all. Conversely, a threshold of 40% arguably resulted in too many resilient areas being identified. Using 40% also felt unsatisfactory as this resulted in similar numbers of areas being classified as ‘high’ deprivation and ‘low’ clinical depression as not.

A threshold of $\frac{1}{3}$ (specifically 33%) resulted in approximately 1% of output areas being classified as resilient. I selected this threshold because I believe it offered the most satisfactory balance between identifying suitable resilient areas and maintaining separation of ‘high’ and ‘low’ areas. Of course, this decision is my own and could be argued to be arbitrary, but I will progress on this basis because any reasonable threshold can be used to provide useful insight, and other thresholds can be selected and tested by subsequent researchers using the code in this repository.

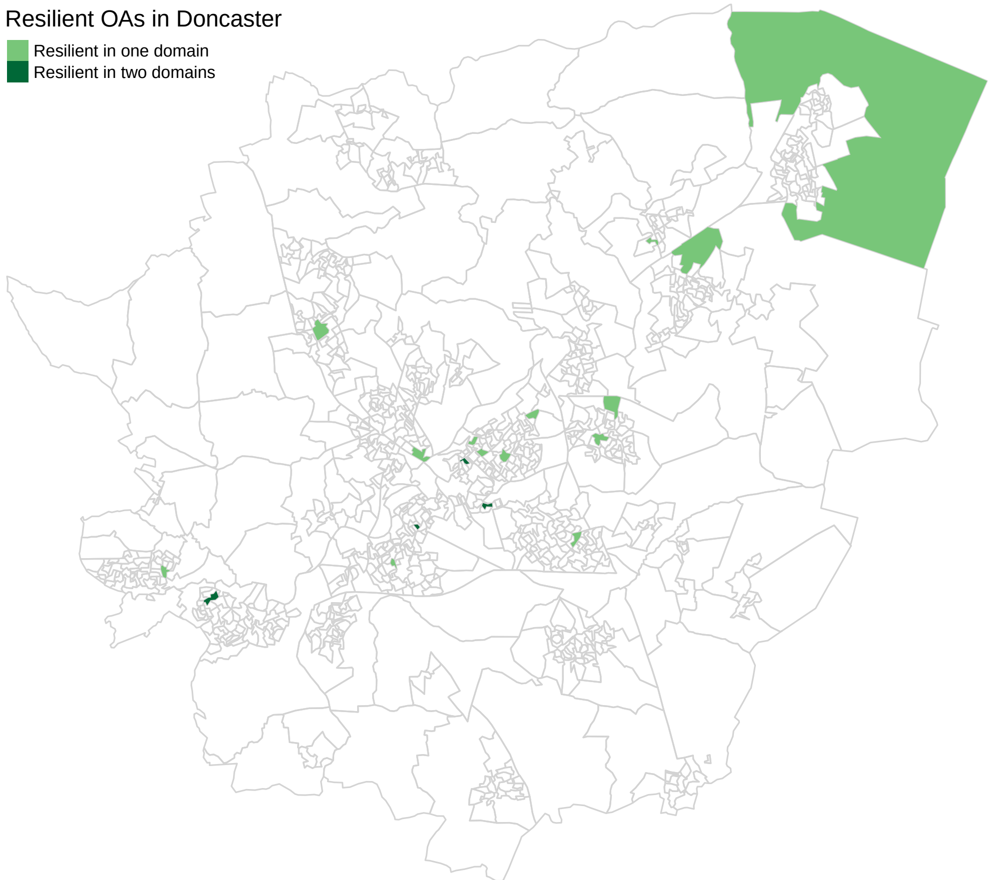

Figure 6.5: Resilient output areas in Doncaster

Having selected an appropriate threshold, I plotted the output areas that the various models identified as resilient. The simulation identifies 18 output areas as resilient in total based on the five deprivation criteria, of which 4 are identified as resilient by two or more measures of area–based deprivation.

One area, to the north east near Thorne, is rural but the majority of resilient areas were in urban or suburban centres. These include output areas in: Adwick le Street to the north; Stainforth to the north east; Armthorpe to the east; New Edlington to the south; Conisborough, Mexborough, and Denaby Main to the west; as well as Doncaster town itself.

6.7.2 Resilient characteristics

In addition to simulating resilient areas based on low clinical depression, I also simulated a comprehensive range of characteristics that I identified in my systematic literature review. Chapter 3 outlines the process I used to conduct the review, while Section 6.2 and table 6.1 summarise the measures and variables I used to operationalise these characteristics. I then simulated individuals with these characteristics into each area and calculated the prevalence of these characteristics at the small–area level.

Figure 6.6, for example, shows the proportion of residents in each area who state they have a 'good' financial situation. This figure illustrates a pattern that is fairly typical of many of the resilient characteristics, with residents in the central urban area of Doncaster and the urban areas of Conisbrough, Mexborough, Carcroft, Askern, and Thorne reporting more constrained financial means than those in the wealthier rural areas around these urban centres. Similar patterns are seen throughout many of the GHQ items, for example, high confidence, good decision making, high ability to face problems, low unhappiness or depressed scores, high feeling useful scores, low feeling worthless scores, low social isolation scores, low 'problems overcoming difficulty' scores, and in areas with high neigh-

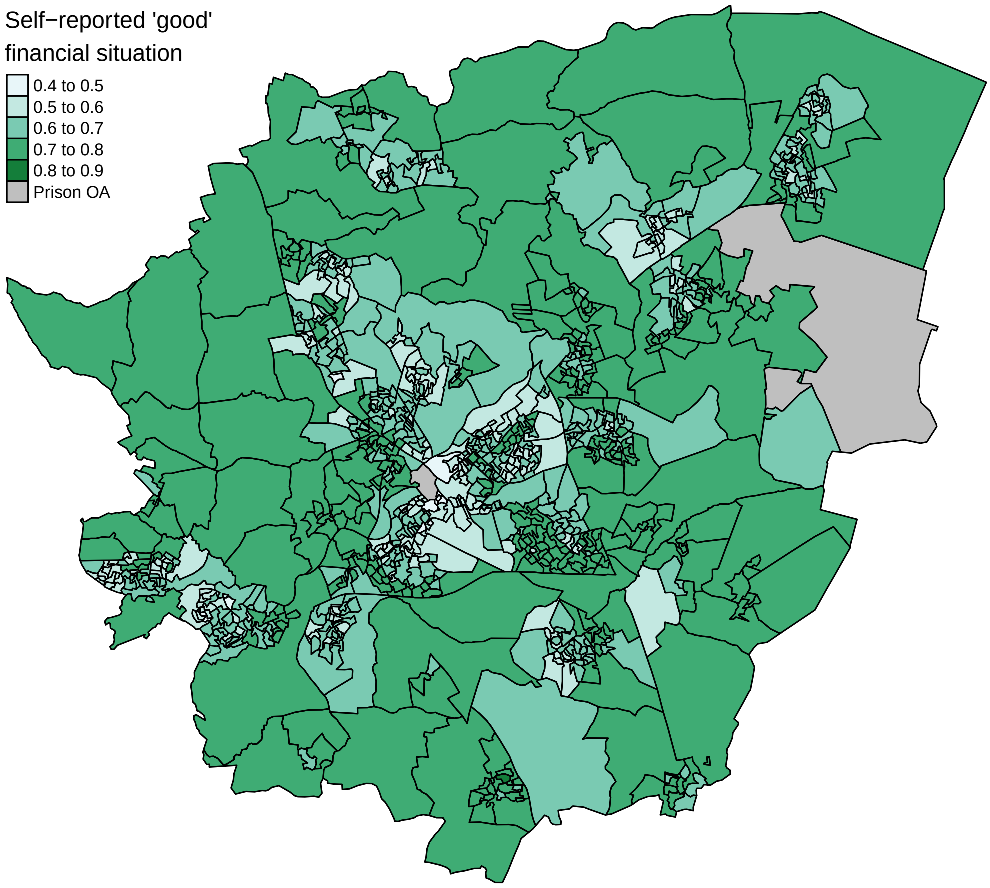

Figure 6.6: Areas with good subjective financial situation

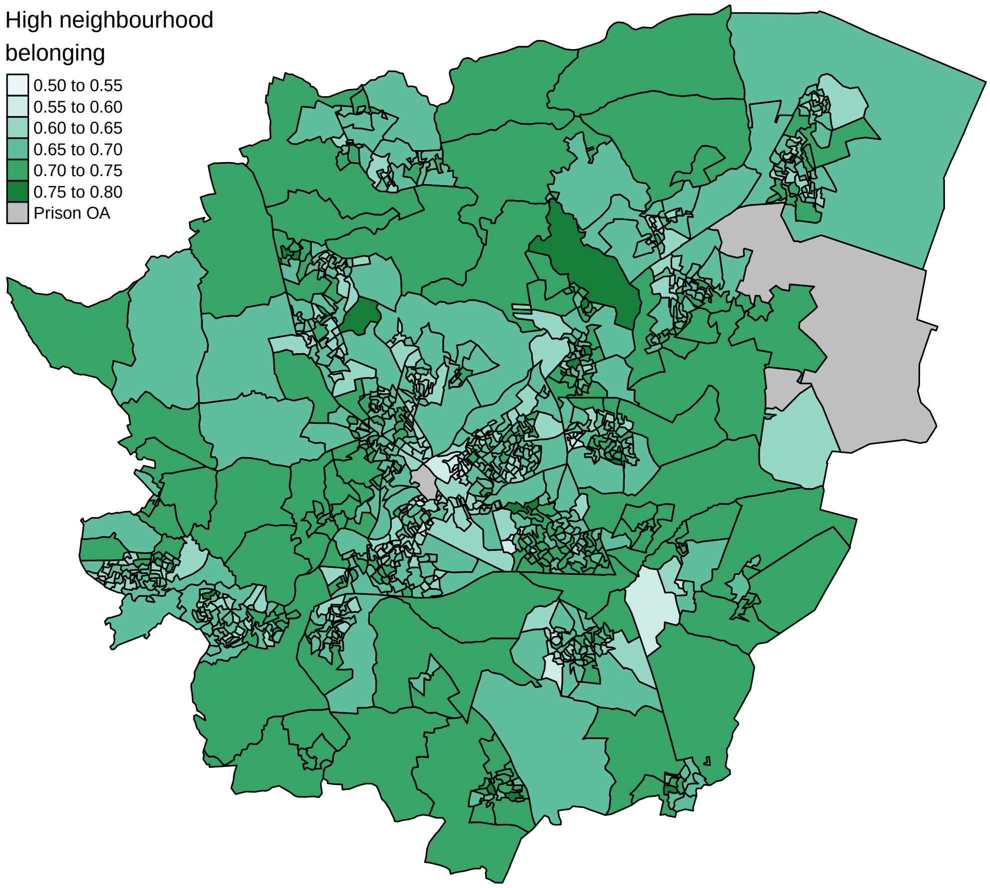

Figure 6.7: Areas with high neighbourhood belonging

bourhood cohesion. These figures can be found in the `figures/cache/` directory with filenames beginning `res_char_`.

This suggests many resilient characteristics in the individual are associated with subjective financial circumstances. Neighbourhood cohesion also seems tied to subjective financial circumstances of the individual, so that as fewer individuals report having financial pressures the perceived characteristic of the area also improves. This is a useful example of how the spatial microsimulation can help to illustrate the relationship between individual–level and area–level characteristics at the small–area level.

Areas with high neighbourhood belonging (Figure 6.7), high GHQ concentration scores, low difficulty sleeping scores, low 'constantly under strain' scores, high happiness scores, and high neighbourhood trust show

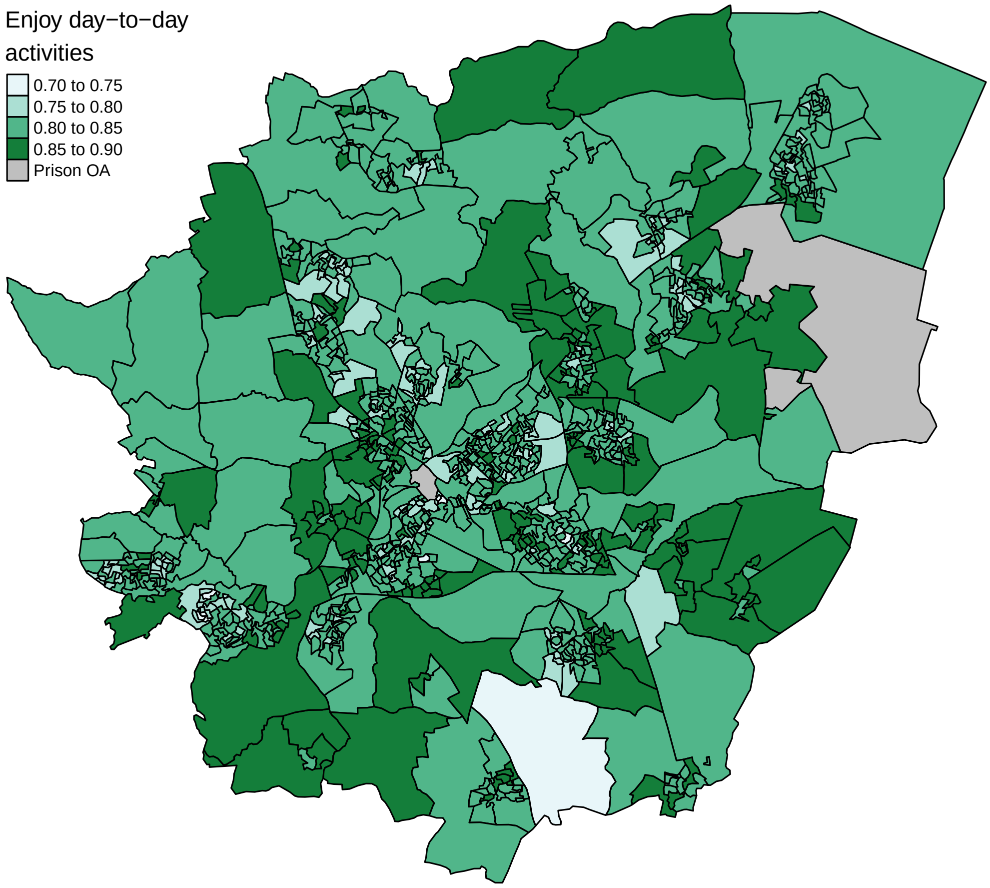

Figure 6.8: Areas with high 'enjoy day-to-day activity' scores

a similar pattern but with key differences in a small number of rural areas. The rural areas to the north east, north central, and north west of the map have lower proportions of residents with these characteristics than might be expected given the previous pattern. These may indicate differences in both individual and area–level characteristics that are not as strongly associated with subjective financial situation. As these are mainly rural areas individuals in these areas may experience additional pressures that are not offset by perceived financial resources.

Areas with high scores for 'enjoy day–to–day activities' GHQ item again show a similar pattern, but a larger number still of rural areas score lower on this resilient characteristic. Larger areas to the north east across to the north west, as well as areas south of Doncaster town centre itself have

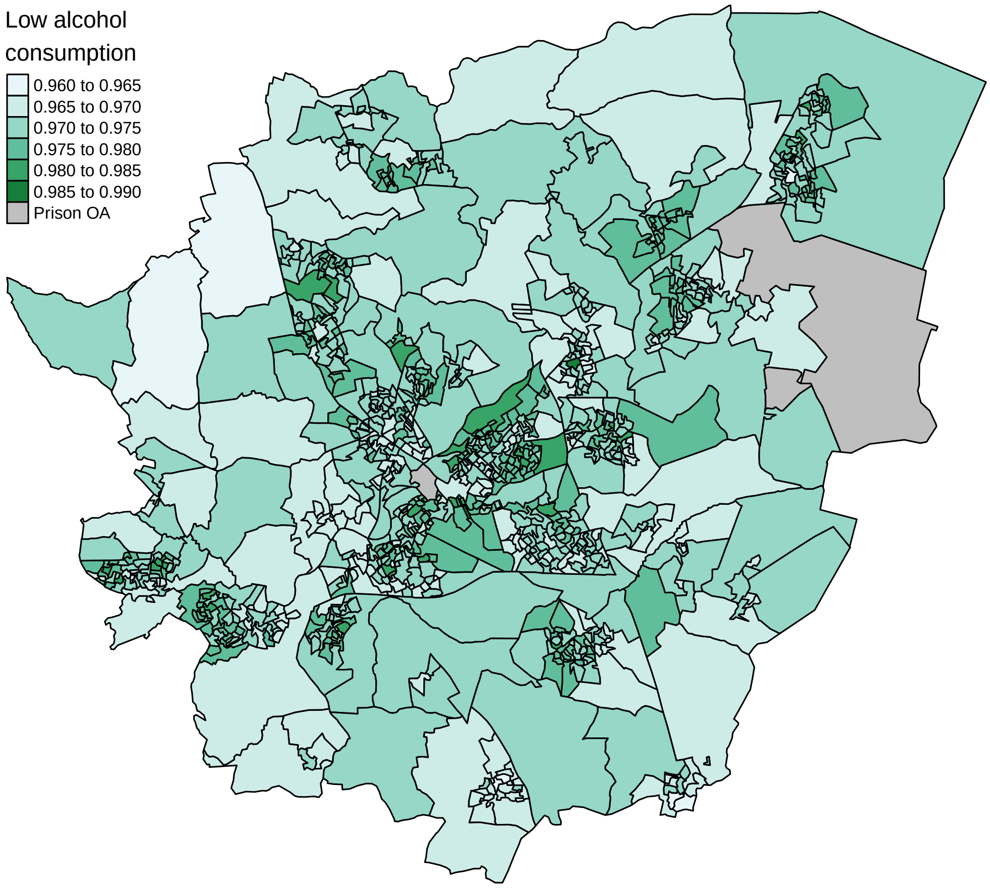

Figure 6.9: Areas with low alcohol consumption

lower scores on this GHQ item than would perhaps be expected if it were tied to financial circumstances. This again suggests that there may be additional pressures on residents of rural areas that are not offset by good perceived financial circumstances. This could be because location and proximity, as well as simply financial resources, play an important role in quality of life. I explore aspects of this in more detail in Section 7.2.

Finally, low alcohol consumption seems to show an inverse relationship with perceived financial circumstances with residents in the poorer urban areas reporting consuming less alcohol than their rural and wealthier neighbours (Figure 6.9). Assuming this is not a reporting inaccuracy (Monk et al., 2015) this suggests that low alcohol consumption is associated with low financial means and could indicate a protective factor against depression for residents in poor financial circumstances.

Comparing these areas with IMD 2015 rank (figure 6.4) suggested that many of the resilient characteristics are associated with affluence, but that this was not always the case, and was indeed the opposite for alcohol consumption.

6.8 Conclusion

In this chapter I have outlined how I produced the full resilience simulation. This simulation built on the pilot simulation outlined in Chapter 5 by adding target variables to study resilience and by increasing the number of constraint variables used and target variables simulated.

The principal variables I used to operationalise resilience were clinical depression and measures of deprivation. I also simulated a range of characteristics thought to promote resilience which were informed by the systematic literature review I documented in Chapter 3. I operationalised as many of these as possible, using variables available in *Understanding Society*. I outlined this process in Section 6.2 of this chapter.

These characteristics are summarised in table 3.2, and include social capital, social networks, cognitive ability, peer support, place attachment, the natural environment, employment status and occupational capital, sports participation, coping mechanisms and coping strategy, behavioural change, sickness benefit, accessible health care, personal and area demographics, neighbourhood congruity, adverse childhood experiences, familial mental health, and financial and budgeting skills. I was able to include measures of neighbourhood cohesion, neighbourhood trust, confidence, abilities, financial coping, health behaviours, general coping, and general happiness.

Many of these characteristics were associated with perceived and actual financial resources, but this relationship did not always hold, especially for rural areas, and indeed was the opposite for alcohol consumption. This suggests there may be strategies that can be employed in less affluent or rural areas to improve resilience.

Alongside these I simulated a range of economic and social status indicators which I use in Chapter 7 to explore the likely effects of proposed national and local policy changes. These include benefit receipt, proportion of income spent on rent, in–work poverty, and neighbourhood safety.

I expanded the constraints from those used in the pilot study to include marital status and economic activity. I was not able to include social class and overcrowding but I do not believe this adversely affected the model because I was able to include other measures of relative social rank—such as education—and deprivation—such as economic activity. I included additional constraints, and constraints with more levels, to help ensure the simulation was as accurate as possible (Section 6.3).

I performed the actual simulation using the iterative proportional fitting algorithm, as I did with the pilot simulation (Chapter 5). To do this I used the `rakeR` package with data from the 2011 census and *Understanding Society*.

When validating the simulation I found the internal consistency of the model to be excellent (Section 6.6). The external validation was less clear–cut. At face value the simulated aggregated prevalence of clinical depression was about half of the 'known' value from Public Health England. Nevertheless the value of the 2011–2012 data point was, at least, problematic, and the overall trend suggested the 2011 value to be closer to the value produced in the simulation. I believe the discrepancy between the given point value and the overall trend can be explained by the changes in measurement of clinical depression prevalence around this time. In any case the simulated prevalence of clinical depression is very similar to the prevalence given by the overall trend, suggesting the model is a better fit than a cursory inspection of the data would suggest.

With the simulated data I illustrated areas with high deprivation but low clinical depression prevalence. These could broadly be thought of as 'resilient' areas. I used a variety of measures to operationalise deprivation including: unemployment and long–term unemployment, low–grade employment (routine employment), index of multiple deprivation (IMD) score, and output area classification supergroup 'hard–pressed living'. I tested a number of thresholds from 10% to 40%, and opted to use 33% this was optimal for this data.

In the next chapter I return to discuss some of the areas identified by the simulation as resilient, both based on the prevalence of clinical depression and the resilient characteristics. I also discuss the likely effects of proposed national and local policy changes to individuals, households, and areas.

Chapter 7

Policy analysis

7.1 Introduction

In the previous chapter I simulated and validated health resilience at the output area level in Doncaster. The principal variables I used to articulate health outcomes and health resilience were clinical depression and measures of deprivation. After simulating clinical depression at the individual level I was able to calculate the prevalence in each output area of Doncaster. I then compared this with a number of area–based measures of deprivation such as the Index of Multiple Deprivation 2015 (IMD) and unemployment.

Alongside clinical depression I also simulated a number of variables that operationalised 'resilient' characteristics. These characteristics were identified in the systematic literature review in Chapter 6 and were hypothesised as promoting health resilience and good health outcomes. Using these two strategies I identified a number of output areas in Doncaster that could be considered resilient (Figure 6.5).

In this chapter I review the findings of these two strategies and compare them to other demographic, socio–economic, and area characteristics. This allowed me to hypothesise which characteristics at the individual–

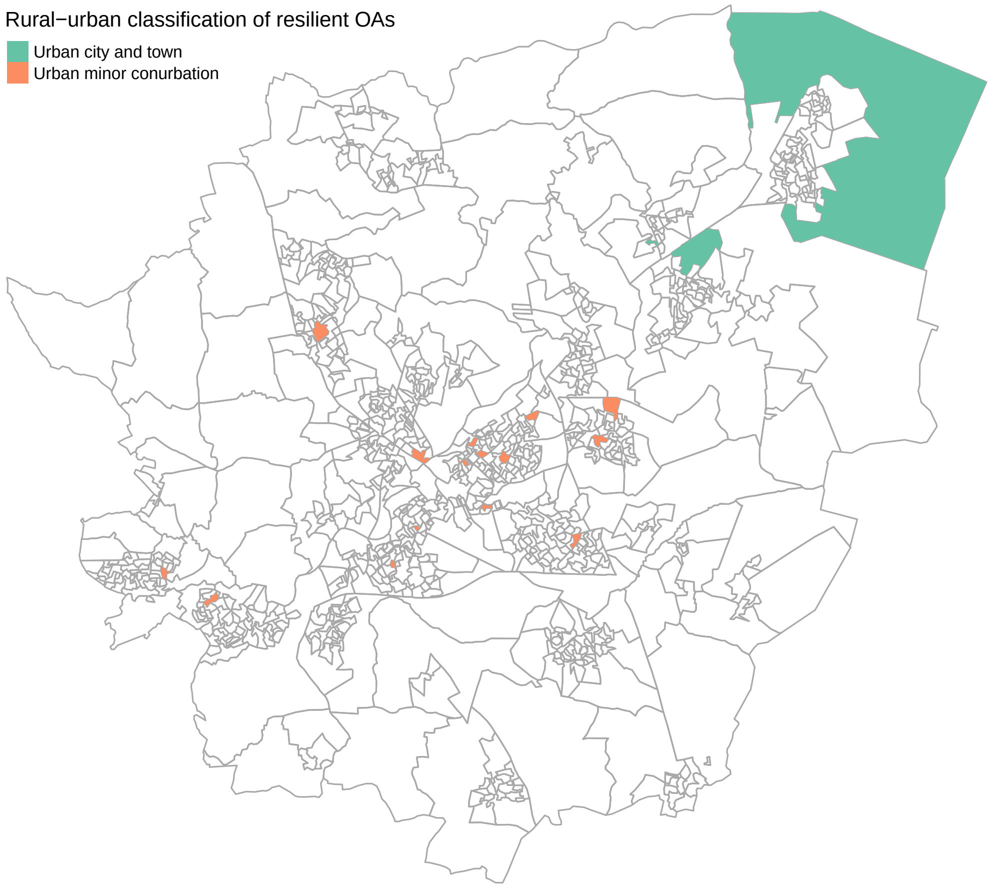

Figure 7.1: Rural-urban classification of resilient areas

level affect resilience at an area level. I provide case studies of four areas in Doncaster that my simulation identified as resilient. I also review a number of local and national policies that potentially have a differential effect due to differences in population or area characteristics to determine which individuals and areas might be affected by these changes. These analyses were made possible by simulating a range of economic and social status variables alongside resilience which I described in Chapter 4.

7.2 Local area

The majority of the resilient areas are high density urban areas that are part of the South Yorkshire minor conurbation. A smaller number of output areas to the north east of the borough are urban, but not part of the South Yorkshire conurbation, so are not as densely populated.

A settlement is considered urban if it has a population of at least 10,000, and an output area is considered urban if "the majority of the population of a particular OA live in such a settlement..." (Bibby and Brindley, 2013a, para 1.5). Areas are further considered a conurbation if they have a high population density per hectare cell (Bibby and Brindley, 2013b, para 4.11), so are the urban areas with the highest population densities.

As many of the resilient areas are in high density conurbations it may be that residents in more densely populated areas are more likely to be resilient. If this is the case, this could be because of greater proximity to health–conducive amenities, for example a general practice, leisure facilities, or green space.

I explored this by comparing the mean distance to a doctor's surgery, leisure centre, and green space from the resilient output areas against the mean distance from the non–resilient output areas. These were post–hoc tests, so further analyses with new data or in other areas are necessary to make any conclusive statements. A similar approach has been used by Campbell and Ballas (2013) to examine if distance affects a variable of interest (2013: 281–283). I used the centroids of the output areas as the origin which provides an 'average' distance from the output area to the nearest amenity. This approach does not take account of population density within the output area—as the majority of the population may not live near the centroid of the output area—or how the road or footpath network affects the journey. Nevertheless, as output areas are small geographies this should not have affected the analysis significantly. Doctor's surgeries and leisure centres were recorded

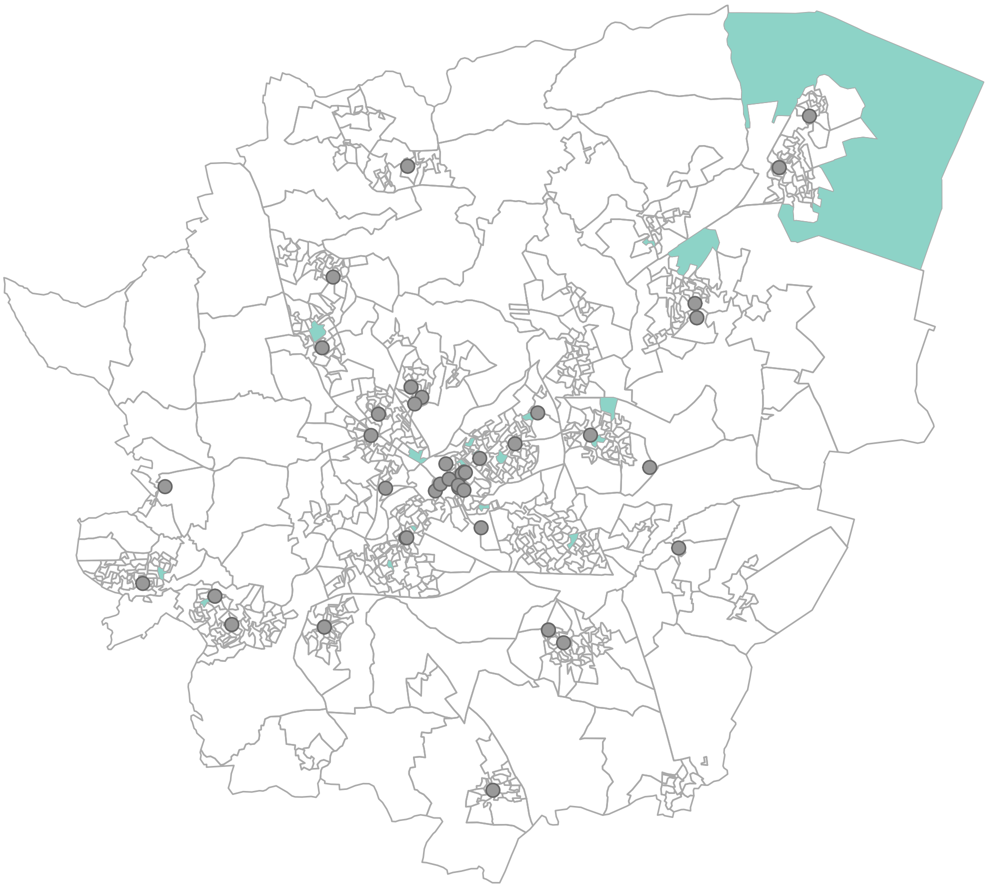

Figure 7.2: Doctors surgeries in Doncaster with resilient areas shaded

as spatial points so the distances could be calculated directly. Green spaces were recorded as polygons so I again used their centroids when calculating distances.

I obtained the postcodes for all GP surgeries in Doncaster from the Care Quality Commission (Care Quality Commission, 2017) and geocoded these into coordinates using the Google Maps Geocoding API (Google, 2017a). I obtained coordinates for leisure centres in Doncaster directly from the Google Maps Places API (Google, 2017b). Finally I obtained polygons for green spaces from the Ordnance Survey Open Greenspace data file (Ordnance Survey, 2017) and obtained centroids from the polygon data.

Figure 7.2 shows the location of doctor's surgeries in Doncaster with resilient output areas highlighted. The mean distance to a doctor's surgery for resilient areas is 935 metres (sd 812m), and for non–resilient areas is 1,219 metres (sd 1,041m). Ease of access to a doctor's surgery could therefore be an important protective factor for clinical depression, and additional practices, especially in areas outside the main urban centres, might be beneficial for resilience and clinical depression.

Figure 7.3 shows the location of leisure centres in Doncaster, with resilient output areas highlighted. The mean distance to a leisure centre for resilient areas is 2,216 metres (sd 1,379m), while the mean distance for non–resilient areas is 2,835 metres (sd 1,826m). As with a GP surgery, the proximity of a leisure centre could confer some protective benefits to nearby residents and careful planning of future sites could be beneficial to currently under–served populations.

Figure 7.4 shows the location of green spaces in Doncaster, taken from Ordnance Survey's newly–released Open Greenspace data set (Ordnance Survey, 2017). I created a matrix of distances between green spaces and output areas, using the centroids of each. I then filtered these, leaving the minimum distance as an indication of the distance to the closest green space. The mean distance to green space from a resilient output area is 510 metres (sd 297m), while for non–resilient areas the mean distance is

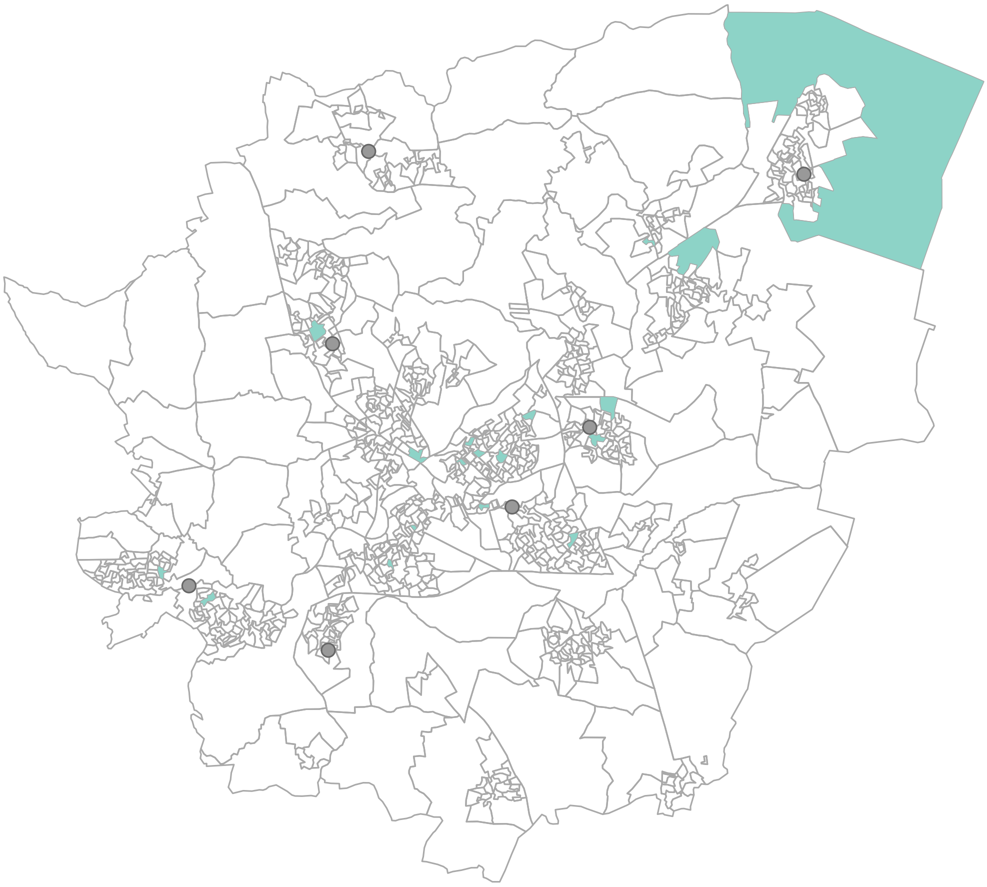

Figure 7.3: Location of leisure centres in Doncaster with resilient areas shaded

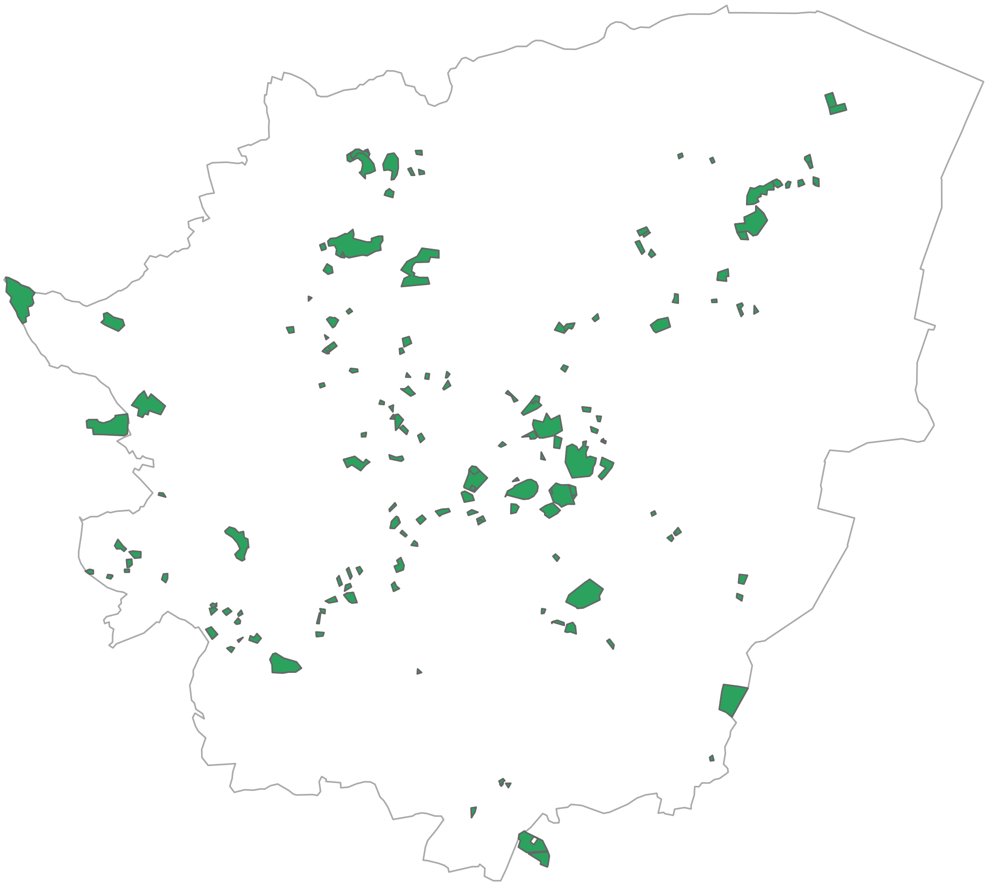

Figure 7.4: Doncaster green space

608 metres (sd 479m).

The spatial microsimulation model is not coded to be aware of, or constrain based on, the distance to a GP surgery, leisure centre, or green space. Nevertheless the distance to these facilities for resilient areas is consistently shorter than for non–resilient areas based on clinical depression outcomes. This supports the hypothesis that the local area has an important role to play in health and resilience, and that improving access to facilities could be an important step in maintaining and improving health.

7.3 Case studies

As part of my research I visited four areas in Doncaster that were identified by the simulation as resilient for clinical depression. Using the definition of low clinical depression I chose two areas that were resilient despite high unemployment, and two areas that were resilient despite high proportions of routine and manual (NS–SEC 7) employment. I chose these criteria to represent those in poverty or living in a deprived area but who are relatively well–off in terms of outcomes for clinical depression. Unemployment is associated with poor economic circumstances, but increasingly families in employment are being affected (Tinson et al., 2016), so it was appropriate to include those in the lowest grades of employment who are most likely to be at risk of poverty.

I selected these areas to be as representative as possible of similar areas in Doncaster, although I do not claim they are representative in any statistical or formal sense. Nevertheless they are illustrative of the broad range of communities in the Doncaster borough. From these areas I sampled a small number of simulated individuals who represent the modal or most common characteristics of residents in these areas. This is a technique that has successfully been used before with spatially microsimulated data (Campbell and Ballas, 2013: 283).

Two areas are in the main Doncaster town itself, a short distance from

Figure 7.5: Case study areas. Motorways (blue), primary roads (green), and urban areas (light grey) are shown for context

the town centre. One area is a suburb to the north east of the town, in Armthorpe. Finally, one area is in Denaby Main, one of the most deprived areas in the borough.

7.3.1 Wheatley

Wheatley is an area to the north of Doncaster town centre part of the South Yorkshire minor conurbation. It is characterised as an 'urban professionals and families' area under the Output Area Classification system. Accommodation is typically in the form of semi–detached housing (Figure 7.6). The distance to a GP, leisure centre, and green space is

Figure 7.6: Case study area 1: Wheatley

599.73 metres, 2515.25 metres, and 936.45 metres respectively.

Based on the 2011 census the percentage of the over–16 population aged 65 and over is 15.2%, compared to 21% for the borough overall. Similarly the percentage of residents of Wheatley with a self–reported limiting long–term illness or disability is 20%, compared to 26% in Doncaster overall. The percentage of the Wheatley population with clinical depression, based on the microsimulated data, is 5%.

Selecting individuals with the modal constraint characteristics left twelve simulated individuals. Individuals were selected who owned one car, owned their home with a mortgage or with shared ownership, held level two qualifications, were either married or single, were either employed or retired, and were White British. The selected individuals were aged 25–29, 30–44, or 45–59.

None of these twelve individuals reported having clinical depression, in keeping with the resilient nature of the individuals in the area. Individual–level GHQ item responses were consistently in line with those expected of resilient characteristics. Most respondents reported having good concentration, sleeping well, managing difficulties, managing or coping with strain. All twelve respondents reported feeling useful, being able to make decisions, enjoying day–to–day activities, having confidence, being happy, and not being socially isolated. These suggest many of the individual–level characteristics measured by the GHQ can contribute to resilience. Responses to neighbourhood cohesion, neighbourhood trust, neigbourhood belonging, and subjective financial situation were more mixed, however.

Two–thirds of the individuals did not save any money, which may contribute to financial pressures later in life, which may in turn be linked to mental health problems. Most consumed low levels of alcohol. The median household income was 2764.97, but inter–quartile range did range from £1,245 to £4,285 so a number of these households were below the poverty line yet still exhibited resilient characteristics.

7.3.2 North Armthorpe

North Armthorpe is an area on the northern edge of Armthorpe, itself to the north east of Doncaster town. It is also part of the South Yorkshire minor conurbation.

It is characterised as 'semi–detached suburbia' under the Output Area Classification system. As with Wheatley, accommodation is typically in the form of semi–detached housing (Figure 7.7). The distance to a GP, leisure centre, and green space is 1128.75 metres, 955 metres, and 571.75 metres respectively. Unusually the distance to a GP is greater than to a leisure centre.

Based on the 2011 census the percentage of the over–16 population aged 65 and over is 26.13%, compared to 21% for the borough overall. Similarly the percentage of residents of North Armthorpe with a self–reported limiting long–term illness or disability is 24%, compared to 26% in Doncaster overall. Based on the spatial microsimulated data, the percentage of the

Figure 7.7: Case study area 2: North Armthorpe

North Armthorpe population with clinical depression is also 5%.

I used the same procedure to select individuals with modal characteristics that I used for Wheatley. This time I selected individuals who owned one or more cars, who owned their home outright or with a mortgage or shared ownership, who had no qualifications or up to level two qualifications only, were married, were employed or retired, were White British, and were aged either 30–44, 45–59, or 65–74. This resulted in 40 individuals sharing these characteristics.

Of the 40, only two had clinical depression, again in keeping with the resilience of the area. Approximately two–thirds reported having high neighbourhood cohesion, and a greater proportion again reported trust in the neighbourhood and a feeling of belonging to the neighbourhood. Based on responses to the GHQ items most respondents reported being able to concentrate, did not have difficulty sleeping, felt useful, being able to make decisions, were confident, and felt happy. Only one individual reported feeling socially isolated.

Figure 7.8: Case study area 3: ‘Five Streets’

About half of the respondents did not save, and only a small proportion consumed high levels of alcohol. The inter–quartile range of household incomes was £1,138 to £5,883 so again a number of households were in relative poverty but still exhibited resilient characteristics. The range of incomes but relative consistency across resilient characteristics suggests there may be an area–level factor or factors that support individual–level resilience in North Armthorpe.

7.3.3 Five Streets

Five Streets is located to the west of Doncaster town centre, near to the train station. It is part of the South Yorkshire minor conurbation.

It is characterised as ‘Challenged Asian Terraces’ under the Output Area Classification system, reflecting the constrained circumstances residents of the area face. Since the 2011 census many of the streets in the area have been redeveloped and rebuilt, although keeping the terraced nature of the original housing (Figure 7.8). The distance to a GP, leisure centre, and green space is 798.88 metres, 1780.77 metres, and 428.17 metres respectively.

Based on the 2011 census the percentage of the over–16 population aged 65 and over is 8.02%, compared to 21% for the borough overall. Similarly the percentage of residents of Five Streets with a self–reported limiting long–term illness or disability is 20%, compared to 26% in Doncaster overall. Based on the spatial microsimulated data, the percentage of the Five Streets population with clinical depression is 7%.

Individuals with modal characteristics in Five Streets: owned either no cars or only one car; rented their home; had no qualifications, or up to level two qualifications; were single; were employed; were White British or Asian or Asian British; and were aged 30–44. The population in Five Streets was therefore younger than the populations of Wheatley or North Armthorpe.

Four simulated individuals matched these criteria. None had depression, again in line with the expected resilience of the area. Neighbourhood cohesion, neighbourhood trust, and neighbourhood belonging were mixed or low in this area, especially compared to North Armthorpe. Sample individuals had mixed financial situations, too. However, the GHQ items still overall suggested the individuals in this area had resilient characteristics, such as concentration, good quality sleep, decision making, overcoming difficulties, confidence, and did not report social isolation.

None of the respondents drank high levels of alcohol, although most did not save money regularly. Household income ranged from £1,103 to £2,733, so some households were in relative poverty but exhibited resilient characteristics.

Figure 7.9: Case study area 4: Denaby Main

7.3.4 Denaby Main

Denaby Main is located to the west of Doncaster town centre at the edge of the district, near Conisborough and Mexbrough. It is part of the South Yorkshire minor conurbation.

It is characterised as 'Industrious Communities' under the Output Area Classification system. The resilient output area of Denaby Main features a community centre, a number of shops and a cafe, a number of sheltered accommodation buildings (Figure 7.8), and local green space. These are reflected in the short distances to a GP, leisure centre, and green space. These are 234.1 metres, 769.82 metres, and 257.9 metres respectively. These short distances could be helping to maintain and protect the health outcomes of the area, despite its deprivation.

Based on the 2011 census the percentage of the over–16 population aged 65 and over is 20.37%, compared to 21% for the borough overall. Similarly the percentage of residents of Denaby Main with a self–reported limiting long–term illness or disability is 30%, compared to 26% in Doncaster overall. Based on the spatial microsimulated data, the percentage of the Denaby Main population with clinical depression is 6%.

In Denaby Main the individuals with the modal characteristics: owned one car; owned their home outright or with a mortgage or shared ownership; had no qualifications; were married; were employed; were White British; and were aged 30–44 or 45–59.

Three simulated individuals matched these characteristics, and as with the other case study areas, none had clinical depression. The respondents report that neighbourhood cohesion and trust are good in their area. Their individual–level characteristics measured by the GHQ items are consistently good, including concentration, sleep quality, decision making, managing strain, overcoming difficulties, feeling useful, enjoying day–to–day activities, did not feel socially isolated, and consumed a low level of alcohol.

Across the four case study areas very few sample simulated individuals had clinical depression, which was expected given that these areas were identified as being resilient. This was despite many of the households having a household income that would place them below the poverty line.

Individual–level characteristics, mainly measured by the GHQ items, were consistently good even among the individuals and households in poverty, suggesting these may provide protection against the effects of poor mental health and clinical depression.

Area–level characteristics were more mixed, but most simulated individuals in these areas still reported good neighbourhood cohesion, trust, and a sense of belonging, as well as access to facilities such as leisure centres, green space, and GP practices. This suggests the nature of the local area and the facilites and amenities available to the residents is important. The proximity and availability of amenities and facilities could be important to residents' health outcomes, and further research could test this hypothesis.

In the next sections I move on to review a number of local and national policies and their likely effect on the resources available to residents, and therefore their likely effect on their mental health outcomes and health resilience.

7.4 Local and national policy effects

Spatially microsimulated datasets are powerful tools for identifying individuals, households, or areas who are likely to be affected by policy changes. In cases where no appropriate data exists for examining these changes, simulated data makes it possible to at least examine *likely* changes. Spatial microsimulated data has been used in this way in numerous cases, as outlined in Chapter 4.

After creating a spatially microsimulated dataset, this data can be filtered to identify observations that meet certain criteria. For example, Campbell and Ballas (2013) were interested in several proposed policy changes including the suggestion of raising the bottom income tax threshold from £7,475 to £10,000. Using their simulated data they were able to identify 'low earners' and map the proportion of these individuals in each small area (2013: 269–272). A number of other spatial microsimulation used in the health domain are outlined in Chapter 4.

I use a similar technique in this section to assess the effects of local and national policy on Doncaster residents, for better or for worse. In each case I outline the policy, identify what group or groups of individuals are most likely to be affected, and what these effects are likely to be. Using these criteria I highlight the number of people in each output area in Doncaster who are likely to be affected by the policy, based on the additional economic and social variables simulated in Chapter 6. This provides valuable small–area level evidence that could be used to address health inequality and improve quality of life for residents of Doncaster by ensuring that the implementation of local policy is used to mitigate or reduce the effects of these inequalities.

Doncaster Metropolitan Borough Council (DMBC) is currently prioritising four key policy areas: “Doncaster Learning”; “Doncaster Working”; “Doncaster Caring”; and “Doncaster Living” (Tillman et al., 2017). These will be the policy focus of DMBC to 2021.

7.4.1 Doncaster Learning

The Doncaster Learning policy is designed to offer development and learning for all ages, but will start with early years. People of all ages could therefore benefit, but it will be young people in compulsory education who will be most exposed to this policy and benefit most from it, if it is successful.

The policy should benefit all people in Doncaster who want to learn, and especially young people in compulsory education. To ensure the Doncaster Learning policy benefits all young people equitably, it may be useful to understand the financial resources available to the families of young people, especially those with reduced circumstances.

Figure 7.10 illustrates the number of children in households in a poor subjective financial situation. These are households that report they are “finding it quite difficult” or “finding it very difficult” to manage financially, a characteristic that has been suggested to be important to resilience (see Chapter 3).

Areas in Conisborough and Mexbrough, Carcroft and Bentley, New Rossington, Doncaster town itself, and also Thorne, have relatively high numbers of children living in families that are finding it quite difficult or very difficult to manage financially. It is therefore these areas that any goals should be mindful of when planning Doncaster Learning services. For example, with the possible exception of those living in Doncaster town itself, free or low–cost, frequent, and reliable public transport should be available to those areas with high numbers of children and young people

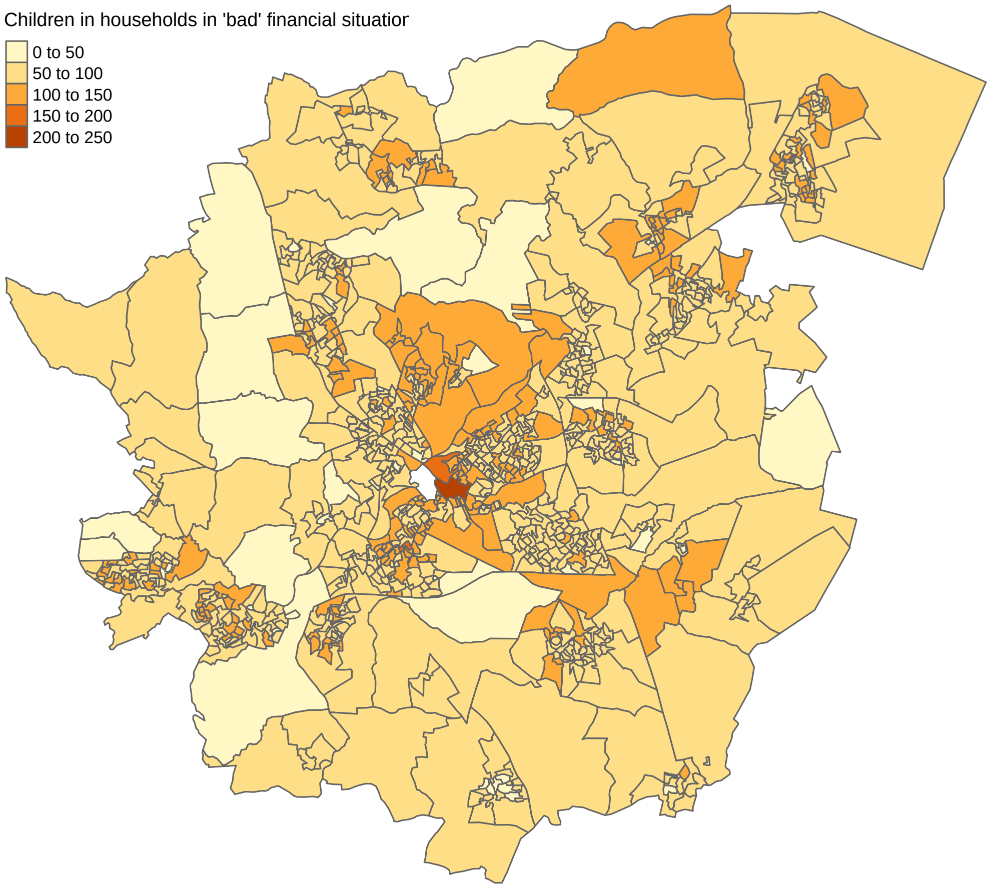

Figure 7.10: Number of children living in households in 'bad' financial situation

in poverty. This would allow them to access education, particularly if they do not live in a family with access to a car.

Provision for free school meals are likely to be higher in primary and secondary schools in these areas. Community centres, youth groups, and after–school clubs could be supported in these areas to encourage children and young people to make the most of their education and to provide educational activities to support their studies. It may also be necessary to ensure health provision for children and young people in these areas—including mental health and sexual health services—is sufficient to meet demand, which is likely to be higher in these areas. Planning and provision of supplementary services that enable these children and young people to get the most out of the education are vital to ensure they flourish and are able to “pursue fulfilling jobs, careers and lives” (Tillman et al., 2017).

Figure 7.11 shows the number of people in each output area who do not save money, again illustrating the nature of material deprivation in Doncaster. While it is more common for people not to save, even in relatively economically wealthy areas, there is still a pattern of low savings in the areas already identified lending further weight to the need for transport, health, and educational support services to help young people from these areas access education.

7.4.2 Doncaster Working

The purpose of the ‘Doncaster Working’ policy is to encourage and create high–quality jobs for local people, to improve their quality of life. Improving conditions of lower grade and lower quality employment, often typified by routine and manual employment, is most likely to improve feelings of fulfilment and job opportunity.

Figure 7.12 shows the number of people in each Doncaster output area who are of working age (16–65) and in employment, but who live in

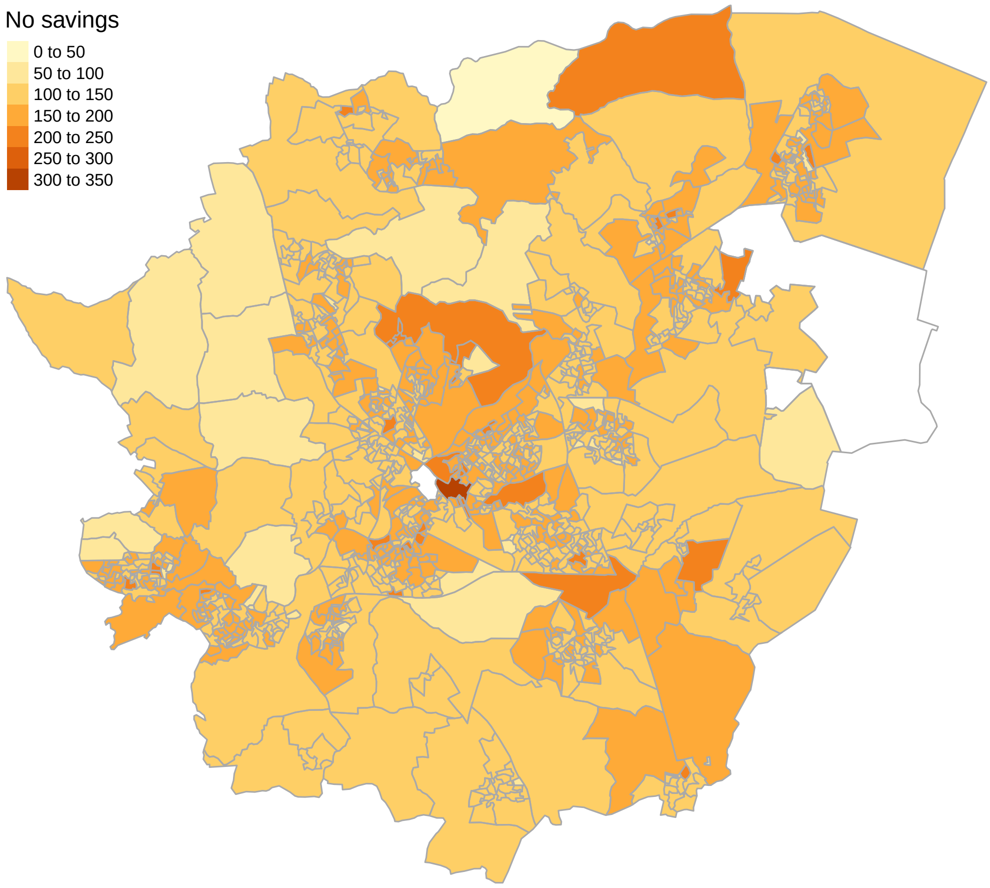

Figure 7.11: Number of individuals who do not regularly save

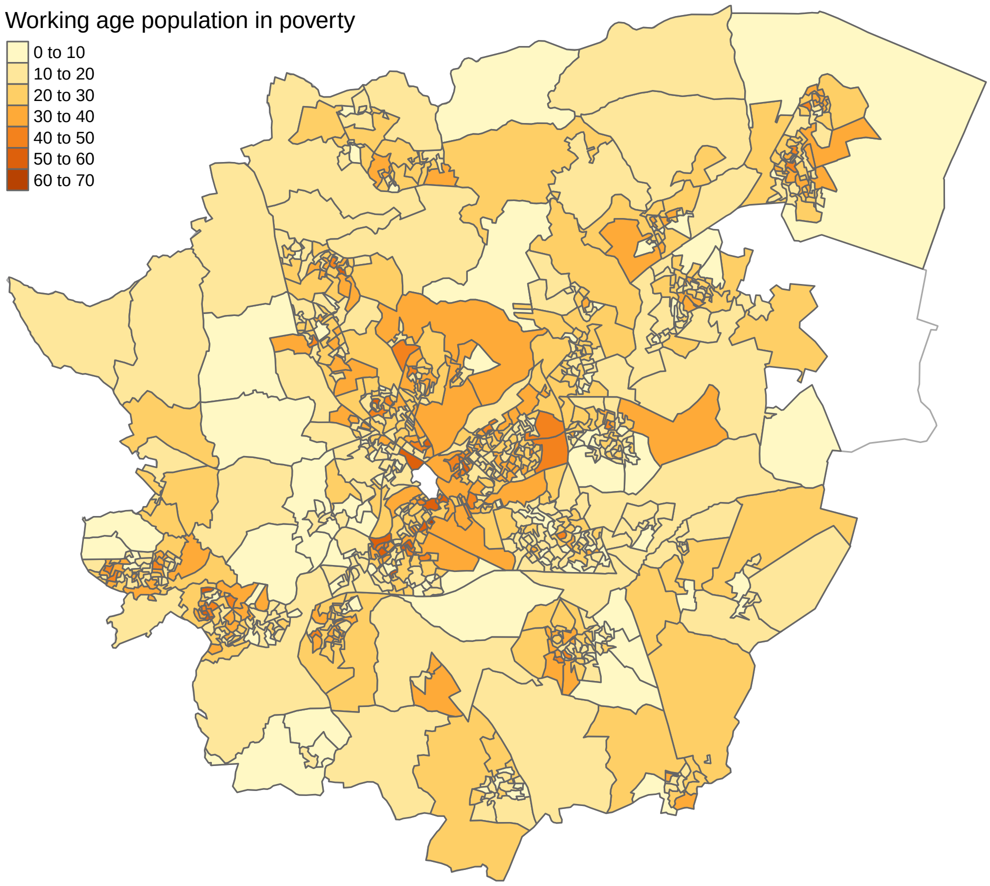

Figure 7.12: Number of individuals of working age (16-65) in employment but living in poverty

poverty. Areas with higher numbers of people in working poverty include: Doncaster town itself; Conisborough and Mexborough to the west; Carcroft and Bentley to the north; and Stainforth and Thorne to the north east.

This gives an indication of the number of the number of individuals working in low quality employment, for example in routine and manual occupations. They suggest people are working, but that the amount they are paid does not allow them to make ends meet. This could be for a variety of reasons, including: employment in casual labour ('zero–hour') contracts that do not guarantee enough income when averaged over a week or month; having to take part–time work that does not pay sufficiently; or simply a low–paying job, for example because it is unskilled.

These areas should receive focus from the Doncaster Working policy. Policies implemented as part of this programme should aim to mitigate the effects of poverty and deprivation for these individuals in these areas. Taking the examples above, this could be by: reducing or legislating against 'zero–hour' contracts; helping those in part–time work, for example by assisting with caring responsibilities or 'topping up' incomes of part–time workers; or improving the job base in Doncaster and moving towards a higher–skilled labour force.

7.4.3 Doncaster Caring

Figure 7.13 shows the number of people in each output area who report experiencing or feeling socially isolated. Social isolation and loneliness are demonstrably associated with poorer physical and mental health (see section 2.7.9) so should be addressed by the Doncaster Caring policy.

The spatial distribution of people in social isolation differs somewhat from the pattern of poverty or deprivation. For example, as well as the higher numbers that are typically seen in Conisborough, Mexborough, Doncaster, Bentley, Sprotborough, and Thorne for measures of deprivation, there are

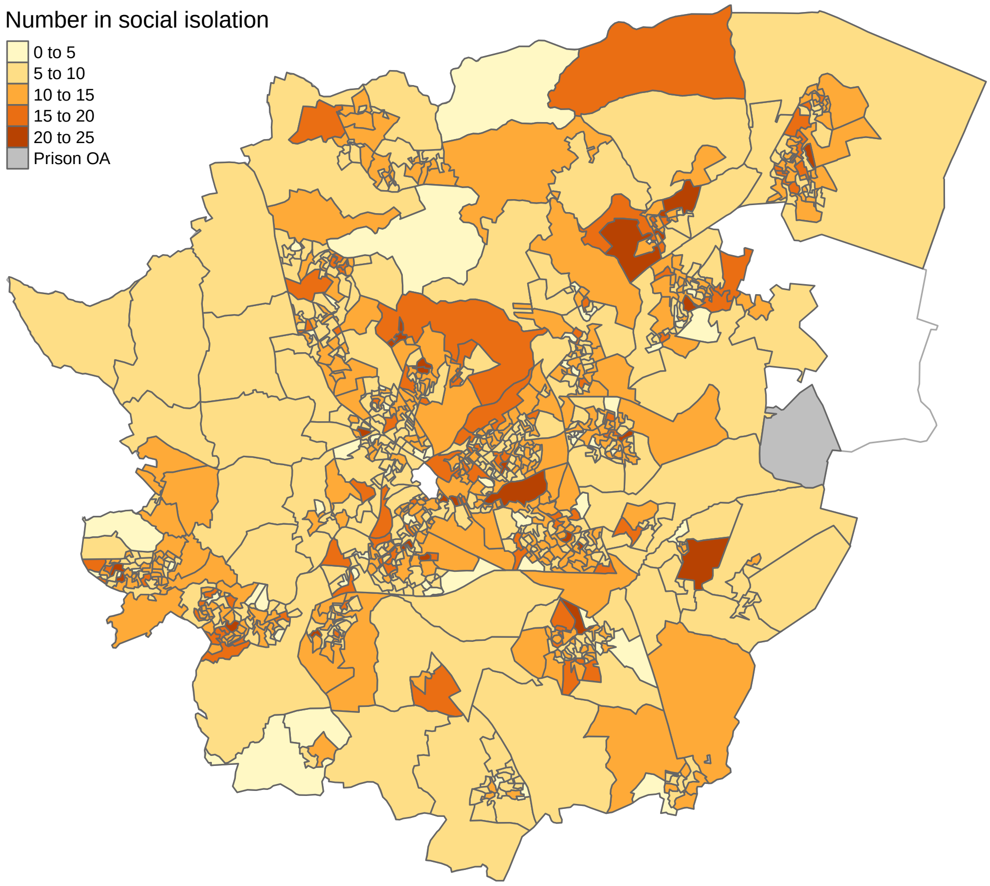

Figure 7.13: Number of people in social isolation

also areas of Sykehouse to the north (west of Thorne) and Bawtry to the south that have a relatively high number of people experiencing social isolation.

As part of the Doncaster Caring policy it should be determined what demographic factors, or combination of factors, are associated with social isolation and loneliness and attempt to tackle this. For example, it may be that older people or certain ethnic groups—or a combination of the two—are more likely to experience social isolation despite the relatively low deprivation experienced by these areas.

The proportion of the population aged 65 and over in the output area near Bawtry to the south of the borough ('E00038534') is 0.31, compared to just 0.21 for the borough overall. The proportion of people aged 65 and over is closer to this figure for the other two output areas. Despite this the *number* of older people aged 65 and over is higher in each of these output areas (97, 77, and 95, compared to a median of 49.5 per output area for the borough overall), and in two of the three the number is nearly double.

The number of individuals in these three output areas who are White British is higher than the median number per output area for the borough overall (401.99, 336.99, and 286.98, compared to the median of 228 overall). Further exploration of the factors that affect social isolation in these areas should be undertaken, but these initial descriptions suggest that age, ethnicity, or an interaction of the two, could affect social isolation in Doncaster.

As a majority urban area, Doncaster residents are at greater risk of the effects of air pollution compared to individuals living in less urban areas (Pascal et al., 2013). Air pollution—especially particulate matter—affects the respiratory and cardiovascular health of individuals, and recent evidence from a study in the United States suggests there may be an association between air pollution and anti–depressant use in women (Kioumourtzoglou et al., 2017). In Europe the greatest number of deaths

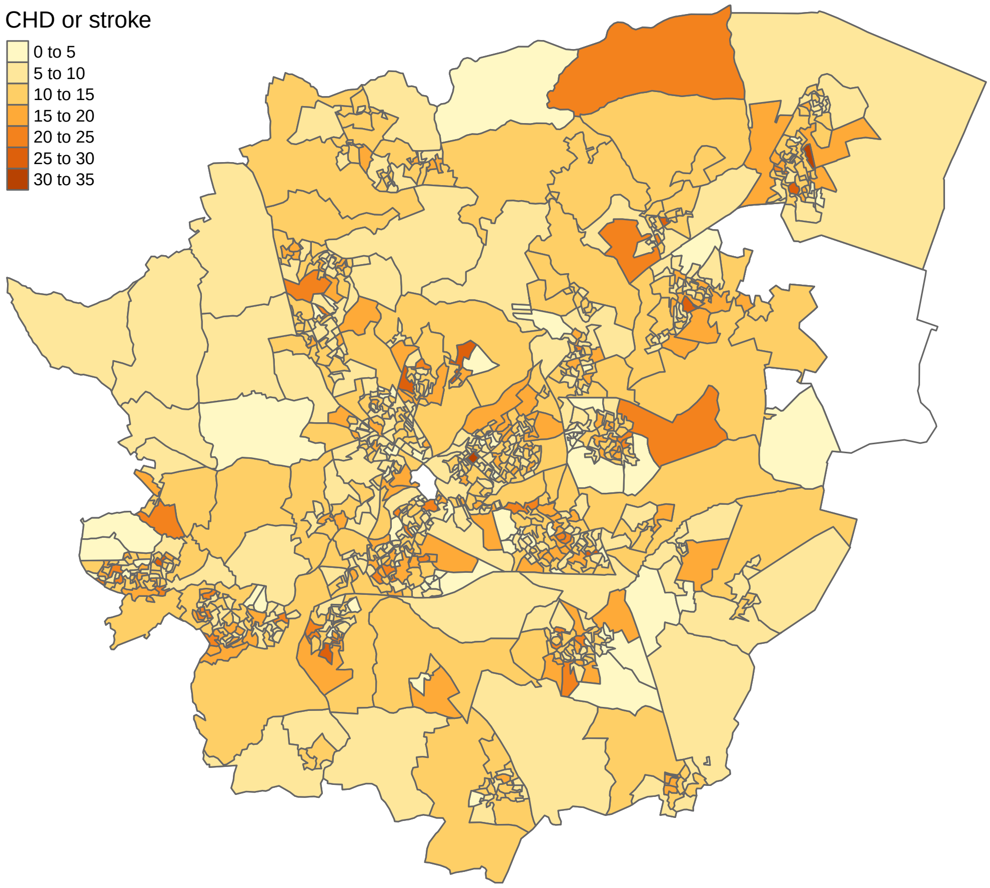

Figure 7.14: Number of residents with CHD or stroke

attributable to air pollution are caused by coronary heart disease (CHD), including angina, heart attack or myocardial infarction, and stroke (World Health Organization, 2016: 41, Figure 14).

Figure 7.14 shows the number of residents with a diagnosis of CHD or stroke in each output area. The area of Sykehouse again features prominently on this map, although this may be because of the age of residents in this output area. The risk of developing CHD increases with age and is most common in the over–50 age group (NHS Choices, 2016).

Figure 7.15 shows the proportion of the population aged 45 and over with a diagnosis of CHD. I have had to include those aged 45–50 because of how the data is constructed to match the census tables, but should still represent the 'at–risk' population for CHD.

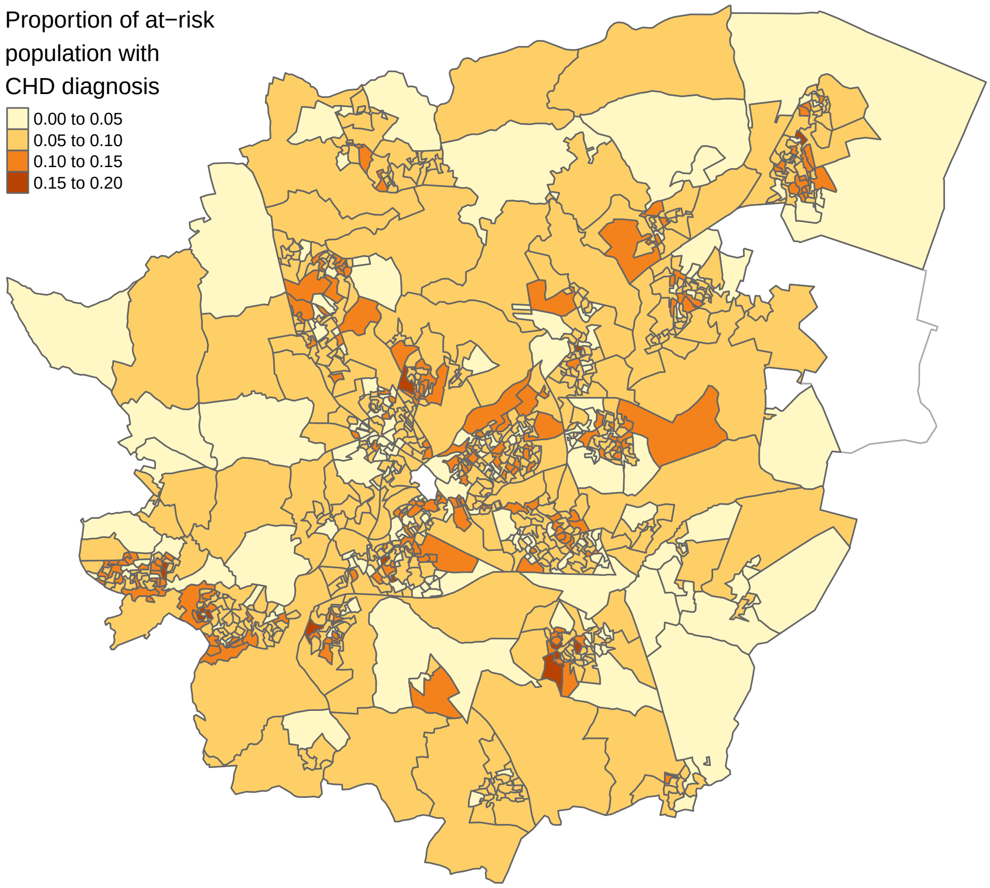

Figure 7.15: Proportion of over-45 population with a diagnosis of coronary heart disease (CHD)

With this figure it is easier to see how the proportion of the at–risk population with CHD is associated with the proximity to an urban area, with greater prevalence near town centres. Doncaster Caring could examine the level of air pollution in these areas—or traffic density as a proxy—and identify areas where individuals are at increased risk of mortality as a result of poor air quality.

7.4.4 Doncaster Living

Doncaster Living aims to improve the living environment for Doncaster residents:

> [Doncaster Living will provide...] access to a full range of housing options, offering homes for life, a top class culture and leisure offer; quality, thriving town centres; an increasingly safe and secure setting. (Tillman et al., 2017)

Examining housing options first, Figure 7.16 shows the median proportion of household income spend on housing costs (rent or mortgage) for each output area. This shows the areas where residents are spending a higher proportion of their household income on housing costs, leaving them less money for other necessities and limiting their disposable income. These are the areas that could most benefit from increased affordable housing, or support with housing costs for residents in these areas, or support with finding better quality employment.

Figure 7.3 shows the location of leisure centres in Doncaster. Other maps of leisure amenities could be produced using geocoded data, and goals for Doncaster Living may wish to take these in to account in devising policy goals to maximise the leisure facilities on offer in Doncaster.

Figure 7.17 shows the number of people in each area who state they do not feel safe out alone after dark in their area (‘a bit unsafe’, ‘very unsafe’, and ‘do not go out alone after dark’). While this could include some individuals who do not go out after dark for other reasons, this overall

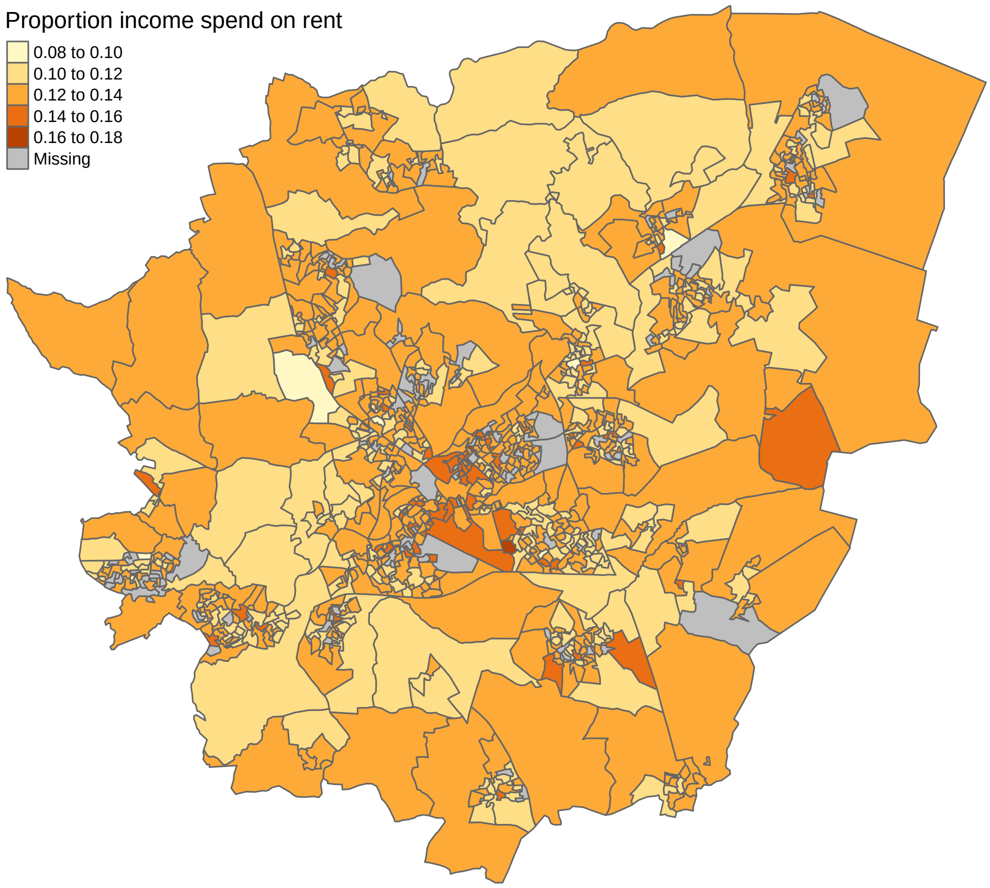

Figure 7.16: Median proportion of household income spend on rent or mortgage, £

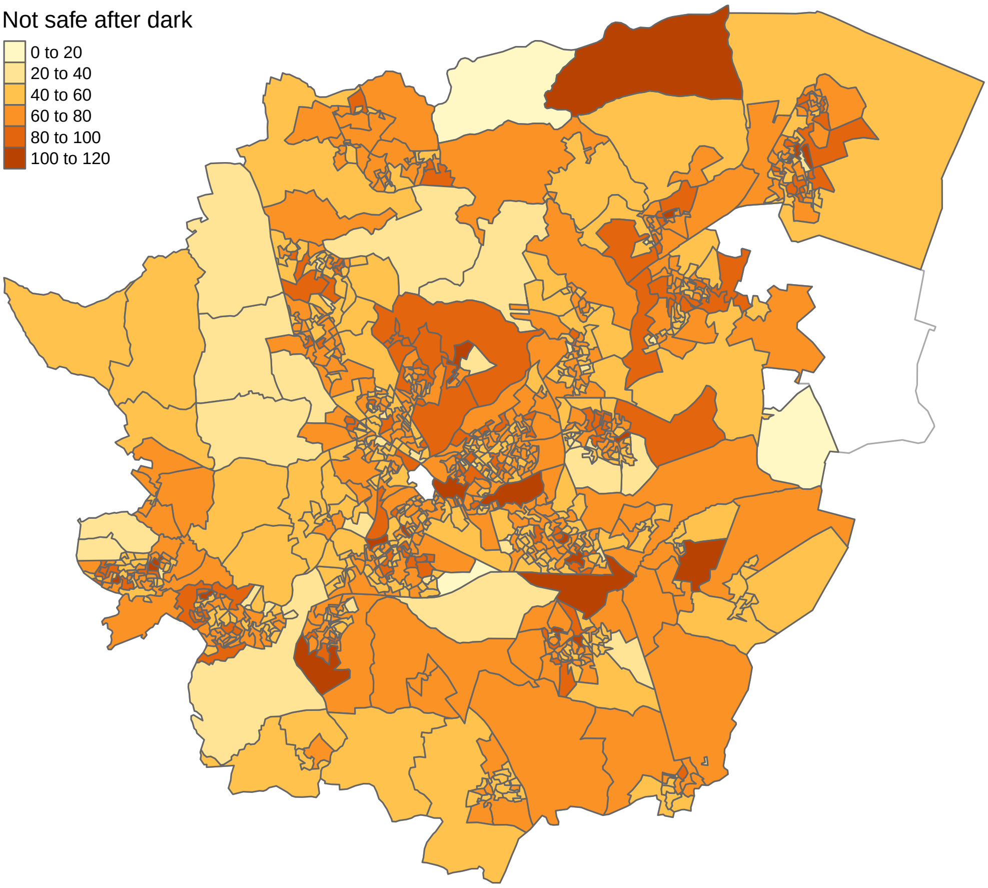

Figure 7.17: Number of people who do not feel safe alone after dark in their area

Table 7.1: Standard allowance amounts for Universal Credit claimants. Source: Gov.uk

Circumstances	Monthly standard allowance
Single and under 25	£251.77
Single and 25 or over	£317.82
In a couple and both under 25	£395.20
In a couple and either partner 25 or over	£498.89

captures the areas of low subjective community safety. The subjective feeling might be as important, or more important, than objective levels of crime or anti–social behaviour, so should be considered as part of the Doncaster Caring alongside any counts of crimes.

7.4.5 Universal credit

As well as local policy initiatives discussed above, there are a number of national policy initiatives implemented by central government that will affect the residents of Doncaster. For example Universal Credit is set to replace a number of existing benefits, including: income–based jobseeker's allowance (JSA); housing benefit; working tax credit; child tax credit; income–related employment and support allowance (ESA); and income support (Gov.uk, 2017b). The transition to Universal Credit in its current form has been heavily criticised, as it is driving some people into debt and poverty (Butler and Asthana, 2017; Mason, 2017).

Those who are on a low income or out of work are eligible for Universal Credit (Gov.uk, 2017c). Table 7.1 illustrates the circumstances under which individuals and couples can claim Universal Credit.

There are three issues that are likely to require policy intervention from DMBC and other service providers in relation to Universal Credit. One is that some people may be affected by a short period when their existing benefits stop and before they receive payment of Universal Credit (Gov.uk, 2017b). Such a period for someone near the poverty line may temporarily cause them to 'dip' into poverty, and they are likely to need additional support during this transition.

Two, the number of people eligible for at least some portion of Universal Credit is difficult to calculate because the eligibility rules are broad, merely specifying 'low income' or 'out of work' (Gov.uk, 2017c). Using data on household income, benefits, and cohabiting status in *Understanding Society*, I have been able to estimate the number of individuals in each area of Doncaster eligible for at least some Universal Credit.

Three, even those individuals who earn more than the work allowance amounts are not necessarily earning a sufficient income to keep them out of relative poverty. Other approaches should be considered to help these people out of poverty in the absence of state benefits.

For each of these three cases I highlight the areas where individuals are most likely to be affected and discuss effects and possible mitigation below.

7.4.6 Transition to Universal Credit

Understanding Society contains rich information about benefits received by respondents, including the benefits Universal Credit is replacing (see section 7.4.5). Of the benefits that are being replaced with Universal Credit, only income–related employment and support allowance (ESA) does not have complete details. Claimaints can apply for contribution–based ESA, which is paid if their national insurance contributions are sufficient, or a means–tested income–based payment. Current government literature states only the income–based ESA is being replaced with Universal Credit, but it is not possible to separate contribution–based claimants from income–based claimants of ESA in *Understanding Society.* Therefore the following estimates are likely to be slightly higher as I cannot remove contribution–based ESA claimaints from my model. Nevertheless, the model contains information on income support, job seeker's allowance, working and child tax credit, and housing benefit, so provides useful

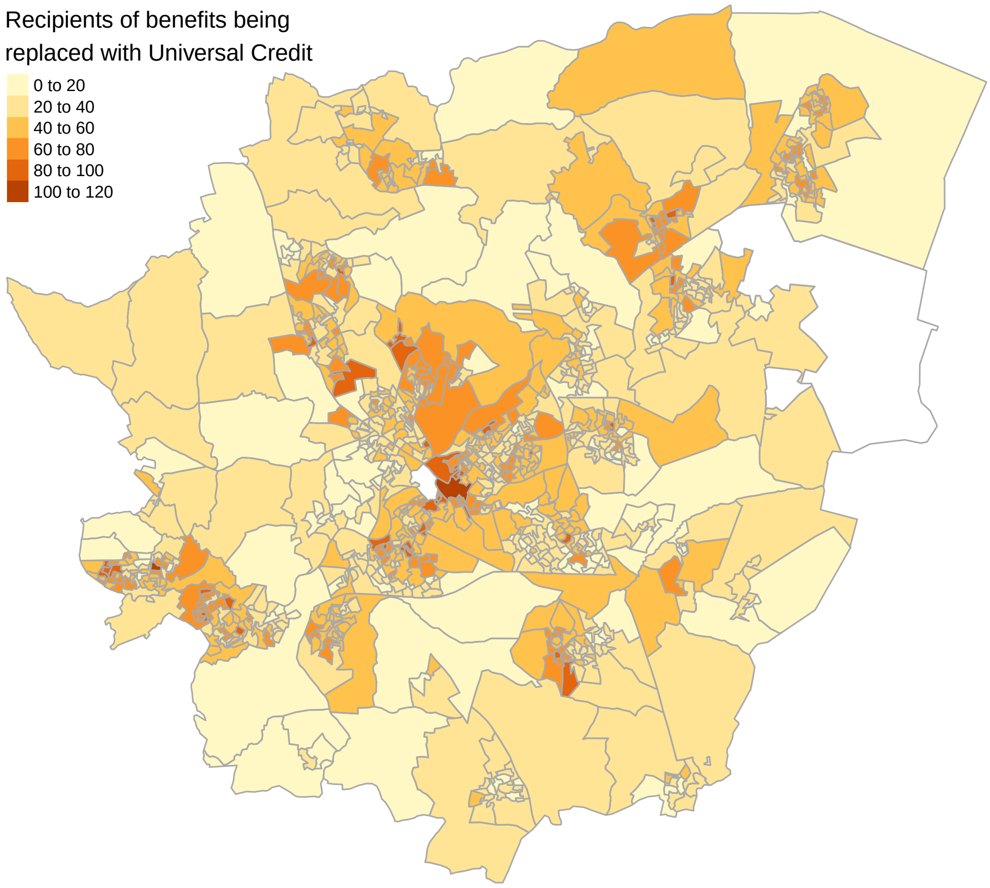

Figure 7.18: Number of individuals receiving benefits that are to be replaced by Universal Credit

evidence about the number of people in each area who are likely to be affected by the transition to Universal Credit. The transition can take several weeks which, although short–term, can cause long–term issues as claimants try to recover from poverty (Presser, 2016).

Figure 7.18 shows the number of claimants in each area receiving a state benefit that is to be replaced by Universal Credit, so are likely to have a transition period where they may experience additional financial strain and may 'dip' into temporary poverty.

DMBC should use this information to plan ways to mitigate any harm caused by this transition and support claimants during this time. For example, claims for discretionary housing payments (Department for

Table 7.2: Maximum monthly earnings before being ineligible for Universal Credit. Source: gov.uk; own calculation

Circumstances	Maximum earnings
Single under 25	£796.63
Single 25 or over	£901.48
Couple (both under 25)	£1024.30
Couple (either partner over 25)	£1188.88

Work and Pensions, 2017; Shelter, 2016) could help to cover housing costs for claimants during this time.

7.4.7 Universal Credit eligibility

Calculating a precise number of people who are eligible for Universal Credit is challenging because the criteria are broad without much specificity (Gov.uk, 2017c). I have estimated the number of people eligible for Universal Credit in each area of Doncaster using information in Understanding Society on benefits, household income, and cohabiting status.

Individuals and couples claiming Universal Credits have a 'work allowance' up to which they can still receive their full Universal Credit payment. For every £1 earned above this work allowance threshold, the payment of Universal Credit is reduced by 63 pence, often referred to as the 'taper rate'. The work allowance is £192 per month for those who get help with housing costs, and £397 per month for those who do not get help with housing costs (Gov.uk, 2017c).

Taking the greater figure I have calculated the amounts each individual can earn before losing all Universal Credit, based on their circumstances. This has the effect of implicitly assuming that Universal Credit covers the claimant's housing costs which will not necessarily always be the case, especially when claimants lose some of their housing support because of under–occupancy, sometimes referred to as the 'bedroom tax'. These amounts are presented in Table 7.2.

Figure 7.19 shows the number of people in each area who are eligible to

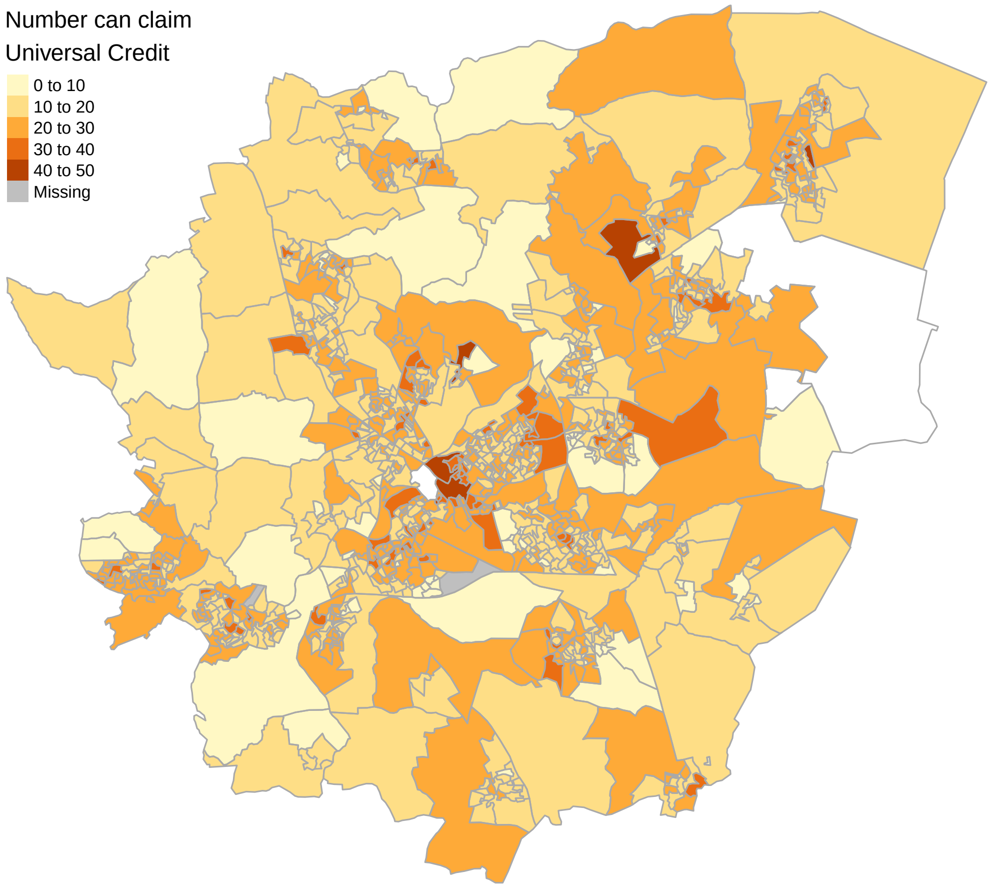

Figure 7.19: Number of residents who are eligible for Universal Credit

Table 7.3: Extra payment amount eligibility for Universal Credit claimaints. Source: Gov.uk. Notes: *based on child born after 6 April 2017

Circumstances	Extra monthly amount
For first child*	£277.08
For second child	£231.67
Help with childcare costs	85% of costs
Disability, or severely disabled child	£649.38 (maximum)
Health condition prevents working	£318.76
Care for a disabled person	£151.89

Table 7.4: Benefit cap amounts for claimants living outside Greater London aged 16–64. Source: gov.uk

Circumstances	Limit per week
Couple (with or without children)	£384.62
Single person with children	£384.62
Single person without children	£257.69

claim Universal Credit using these criteria and assumptions. Areas in Doncaster town centre, Conisborough, Mexborough, Bentley, Carcroft, Stainforth, Thorne, Rossington, and Bawtry have relatively high numbers of individuals who are eligible to claim Universal Credit. Doncaster should support these individuals who are on the lowest incomes and most vulnerable to poverty.

One such approach could be to help them maximise their incomes. Some people eligible for Universal Credit are also able to claim extra payments on top of their standard allowance (Gov.uk, 2017c). Table 7.3 summarises these additional payments and their eligibility requirements:

Support from DMBC and other public agencies, as well as voluntary and third sector support services, could help these individuals to claim any extra entitlement. It should be noted that social welfare payments are capped (the 'benefits cap') at the following amounts for those aged 16–64 living outside the Greater London area (Gov.uk, 2017a). Therefore Universal Credit, including any extra payments, might not offer a complete solution to move people out of poverty. Table 7.4 shows the cap amounts.

This cap might have detrimental effects on claimants if their needs are complex, for example if they are a carer or single parent, and the benefits cap prevents them from obtaining enough financial support to cover their living costs.

7.4.8 Moving out of poverty

Some individuals who are earning more than the work allowance, and therefore are not eligible for Universal Credit, are not necessarily earning a sufficient income to keep them out of relative poverty. Figure 7.20 shows the number of individuals in each area in Doncaster who are in relative poverty—that is they earn less than 60% of median income—but who earn more than their relevant work allowance.

For these people Universal Credit, and benefits, are not able to help them move out of poverty and thus need to be supported in other ways, such as by helping them to transition to more stable or higher–skilled employment. Ideally the Universal Credit thresholds would be raised to help people like this move out of relative poverty, but in the absence of such an initiative from central government other approaches should be considered. Not all individuals are able to work, and therefore move into more stable or higher–skilled employment, and any policy implemented should consider this to help all individuals out of poverty.

7.5 Conclusion

In this chapter I have used spatial microsimulation outputs to illustrate some of the issues facing residents of Doncaster. These have included a geographical analysis of the provision of health amenities and facilities, case studies of four resilient areas, and policy analysis of local and national policies aimed at supporting those on low incomes or in deprivation.

The geographical analysis highlighted that resilient areas, whether urban

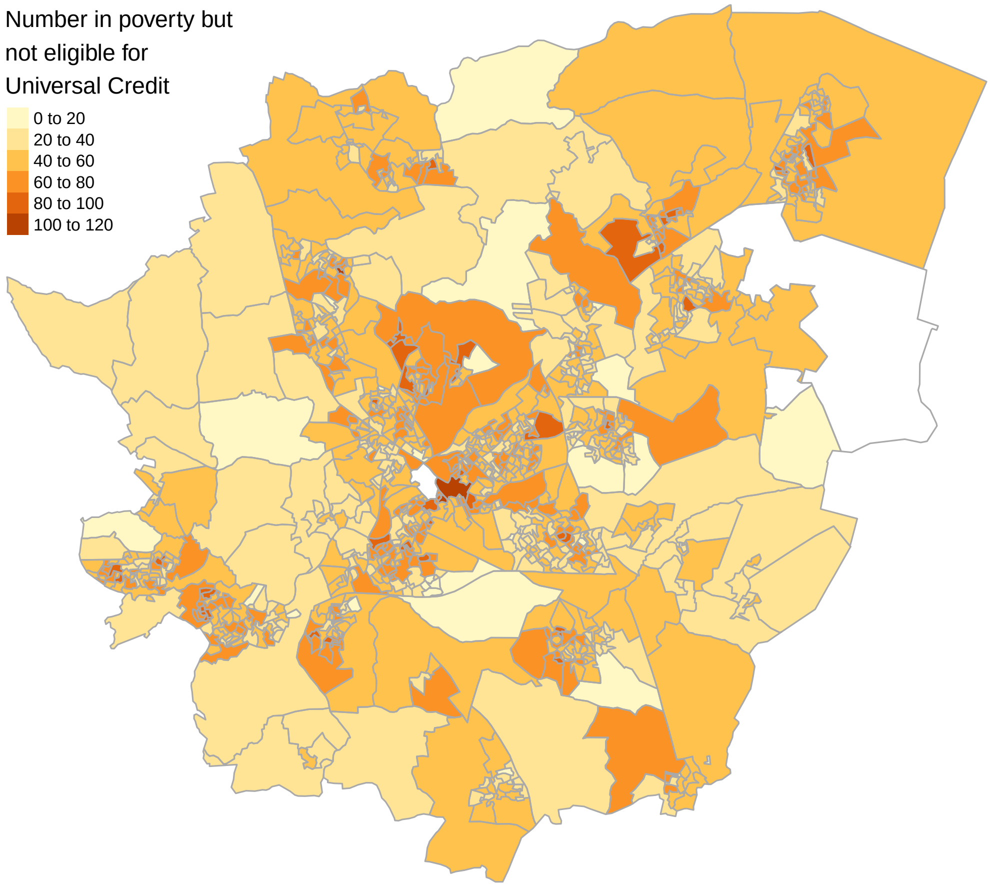

Figure 7.20: Number of individuals in relative poverty but not eligible for Universal Credit

or not, tended to have better access to health amenities and facilities, such as a GP practice, leisure centre, or green space. DMBC and other agencies should consider the location and proximity of any future facilities at the planning stage to maximise the benefits to health and resilience for as many individuals as possible. Doncaster could also use this information to review and rationalise their current provision.

The case studies illustrated the characteristics and nature of four resilient areas, and the people who live in them. Despite many of the individuals facing poverty and deprivation, none of the sampled individuals reported having clinical depression, in keeping with the resilient nature of the area. Many of the individuals reported good mental health characteristics, reported through the use of the General Health Questionnaire instrument, so there is evidence that good mental health could confer resilience or protective benefits.

In two of the areas sampled residents were also highly diverse in their backgrounds and demographics, which may suggest the diversity of the area's population may confer additional protective or resilient benefits for residents. This is at odds with the other two areas, where a sense of shared identity may be considered positive. Further research might explore why some areas benefit from diversity, while others benefit from homogeneity.

Local policy—Doncaster Learning, Doncaster Working, Doncaster Caring, and Doncaster Living—are designed to help people obtain the best education they can, move into better quality work, and access high quality housing and leisure opportunities. Doncaster Learning should prioritise children in poverty or who do not live near good quality public transport links to ensure they are able to access educational sites. Other issues of access to education are important, such as the quality of education provision, but without addressing children in poverty they are likely to get left behind in terms of achievement and ultimately employment.

This is linked with Doncaster Working which aims to move residents out of low quality, low pay employment into better quality jobs. With the spatially microsimulated data I have highlighted the areas where people are in work but in relative poverty, indicating where such policies should be targeted and which would not be possible without the spatial microsimulation. Moving these individuals out of low quality work may also help improve their mental health, and therefore resilience, independent of any financial improvement in their circumstances.

Doncaster Caring seeks to maximise the provision of care and health care to those most in need. Physical health is clearly important, but interventions and funding should seek to improve mental health, and therefore resilient, too. Social isolation, for example, has been associated with detrimental physical and mental health, so with the spatial microsimulation I have been able to highlight areas with the greatest numbers experiencing social isolation. There are clearly other concerns and criteria, but this should illustrate some of the areas that require support and funding under Doncaster Caring.

Doncaster Living aims to improve housing quality and community safety, as well as leisure provision. Having already demonstrated the proximity of leisure centres and green space is associated with resilience, I also considered housing affordability and *perceived* community safety. First, I was able to highlight where individuals spend the greatest proportion of their household incomes on housing. This suggests people in these areas have low incomes and may need support or better opportunities to move to better quality housing. For example, discretionary housing payments could be used to support these individuals with the financial cost of moving to a new, better quality, home. It may also be necessary to see where these individuals overlap with those in low quality employment—Doncaster Working—to support them into higher quality jobs with higher incomes. I was also able to highlight where individuals perceive their area as unsafe after dark. In this case perception may be as important as any 'objective' measure of community safety when considering the quality of life of residents. Improving housing quality and community safety are laudible goals, but it is important to not miss the areas most in need, such as those identified here.

National welfare policy is to move claimants to Universal Credit. I cannot link the move to Universal Credit directly to health resilience empirically, but there are two key effects of Universal Credit that are likely to affect claimant's mental health. The first is when transitioning to Universal Credit, claimants stop receiving their current benefit and then begin receiving their new benefit. This can take several weeks, and can drive some people into poverty which becomes a long–term issue as they try to recover from it (Presser, 2016). Whether short– or long–term, periods spent in poverty are detrimental to mental health which, as demonstrated above, is likely to have tangible effects on people's resilience and health.

The second issue with Universal Credit is that even those who earn more than the work allowance do not necessarily have an income that is sufficient. Without being able to claim social welfare support, these individuals are then stuck in relative poverty, with the detrimental effect on health and resilience that this is associated with. As they cannot claim social welfare benefits to help them out of poverty, other approaches should be considered by Doncaster. Policies such as Doncaster Working—moving individuals to higher–skilled and better paying positions—should help many move into better quality and more stable employment. Not everyone is able to work, however, for example because of caring responsibilities or disabilities, so any policy implemented should not just assume that better quality work can move everyone out of poverty.

These case studies, figures, and geographical analyses demonstrate the power of the spatial microsimulation technique and this model. Identifying the people and areas most affected by poor health and resilience, low incomes, and deprivation would not be possible with only data available from the census or *Understanding Society* when analysed separately and without spatial microsimulation modelling. With the spatial microsimulation it has been possible to identify specific individuals and areas that are most at risk of the effects of social welfare policies and health inequalities, which could allow the targeted intervention of local policy to mitigate or reduce the detrimental effects of these experienced by Doncaster residents.

Chapter 8

Conclusions

In this thesis I have examined the relationship between socio–economic risk and positive mental health at the small–area level in Doncaster. More generally this framing of exposure to a risk and a positive outcome is articulated as resilience. I outline these ideas and the emergence of resilience, especially of early psychological resilience, in Chapter 2.

Because of the ambiguity of measures used to operationalise risk and positive outcomes, in Chapter 3 I conducted a systematic scoping literature review of measures in contemporary resilience literature to inform my analyses. For this review I use the PRISMA method of reporting. I synthesise the results in a narrative analysis, as it was not necessary or appropriate to carry out a meta–analysis of the range of factors that affect resilience.

Because of the strong association between resilience and the psyche in this literature, I chose clinical depression as the positive mental health outcome to examine at the small–area level. This study, to the best of my knowledge, is the only resilience research to use individual clinical depression diagnoses as a health outcome at the small–area level in the United Kingdom. Previous studies have been able to include morbidity and mortality data at the small–area level, but these have generally been constrained by information available in the census. They have also used area–based measures, such as area–level premature mortality, to examine the resilience of the area. This is the first study to be able to use rich individual–level data available through a representative and comprehensive survey, but at the small geographical level.

I achieve this through the use of spatial microsimulation. Spatial microsimulation is a technique to statistically estimate the spatial distribution of individual–level data, and has been successfully used for decades in the social sciences for policy analysis. I outline the spatial microsimulation method and a number of previous applications of the technique in the health domain in Chapter 4.

For the analyses of Doncaster resilience I created a static, rather than dynamic, model which was most appropriate for this application. In the first instance I constructed a pilot model, using limiting long–term illness or disability as a target or dependent variable. This allowed me the opportunity to develop and test code for the construction of a spatial microsimulation model, specifically using the iterative proportional fitting (IPF) technique. From this code I developed the `rakeR` package for `R`. This package vastly simplifies the process of performing IPF spatial microsimulation and is available free–of–charge for use by other researchers wishing to use IPF.

A second advantage of conducting a pilot simulation was that this helped me to externally validate my final model. External validation—validation of the model results with known data—is challenging because the data that is necessary to perform the validation is typically the same data that we wish to simulate. One approach to external validation is to test how accurately the model simulates a correlated variable. Limiting long–term illness and disability is available in the census, so I was able to test the accuracy of the simulation against this known data. Using this technique I was able to determine that the model was over–simulating limiting long–term illness or disability, which I was both able to correct and use to inform my assessment of the final simulation model. Using this pilot simulation I also compare the accuracy of the resulting data at different geographical output levels, and compare the accuracy of fractional weights and integerised weights from the simulation output. I describe this process in Chapter 5.

In Chapter 6 I extended the pilot model to simulate health resilience. I made use of the `rakeR` package I developed in the previous chapter to perform the simulation process. This model incorporated additional constraints to further improve the model fit and used clinical depression as the target variable. With the additional constraints the model still converged, so a more precise fit may have been achieved leading to a more nuanced fit of individuals at the small–area level. Internal validation of the simulation was excellent, with negligible absolute errors and non–significant *t*–tests suggesting the simulated data matched the known constraints very well. To externally validate the model I aggregated clinical depression prevalence to the Doncaster local authority overall and compared this with known data from Public Health England. The published point data for 2011 was higher than the simulated prevalence of depression, but this point data did not fit with the surrounding trend data and may have been an anomaly in the data collection process. As a result it is not unambiguous that the simulated prevalence of clinical depression matched known values, but the simulated value is in line with the trend data for this period so I argue the simulation is a reasonable estimate.

At this stage I also simulated a range of resilient characteristics that I identified in the systematic scoping literature review in Chapter 3. These allowed me to explore some of the characteristics that may form protective factors for resilient individuals, and complemented my analysis of clinical depression. These showed a number of distinct geographical distributions and indicated differences between urban and rural areas, suggesting key differences between the two. High alcohol consumption seemed to be related to affluence, but high alcohol consumption was based on a number of assumptions so this would need to be explored further.

After running the simulation, I identified a number of small areas in Doncaster that had low clinical depression prevalence but high socio–economic risk (Figure 6.5). These risks included unemployment and low–grade employment (NS–SEC 7). A number of output areas were resilient for both measures of socio–economic risk, suggesting the model is identifying genuinely resilient areas.

Many of these resilient areas—which therefore had high deprivation but low clinical depression prevalence—exhibited a number of resilient characteristics. These included both neighbourhood characteristics, such as high neighbourhood cohesion and social capital, as well as individual characteristics such as positive GHQ scores.

Having identified resilient areas of Doncaster under these assumptions, it was possible to analyse a number of factors that may be conducive to resilience. Using geocoded locations of GP surgeries, leisure centres, and green space, I was able to determine that the resilient areas were consistently closer to these amenities than non–resilient areas. This suggests that proximity to amenities such as these could confer protective benefits to individuals that allows them to maintain positive health outcomes in spite of exposure to socio–economic risk that cause others to succumb to depression.

The case studies I illustrated in Chapter 7 provide rich details about a selection of individuals who live in four resilient areas. I chose the four areas to demonstrate a cross–section of resilient areas in Doncaster that I identified with my simulation. The (simulated) individuals were selected from the most common characteristics of the area. The vast majority did not have clinical depression, in keeping with the resilient nature of the selected area. This was despite many of the individuals living below or near the poverty line.

Many of the individuals had positive individual mental health characteristics, reported through responses to the General Health Questionnaire instrument, which is itself identified as a useful instrument to articulate resilience through the systematic literature review chapter. Responses to area–level questions were more mixed, but still primarily suggested that residents of these areas enjoyed neighbourhood cohesion, trust, and a sense of belonging. As such, area–level characteristics may support or enhance individual resilient characteristics.

One of the most powerful uses of spatially simulated micro–data is the ability to estimate the spatial effects of policy change at the small–area level. Doncaster Metropolitan Borough Council (DMBC) have proposed four policies to provide focus to its activities to 2021, as well as being subject to national policy change. I examined the likely effect of these policies on individuals and areas by including additional economic and social status variables in the simulation.

Local policy—Doncaster Learning, Doncaster Working, Doncaster Caring, and Doncaster Living—are designed to help people obtain the best education they can, move into better quality work, receive the care they need in their community, and access high quality housing and leisure opportunities.

By filtering the spatially microsimulated data set I was able to identify areas that DMBC could focus on to maximise the benefit of their policy initiatives. Children and young people in low income households, for example, stand to lose most if they cannot access education so any policy implemented should consider these individuals.

Similarly with the spatially microsimulated data it was possible to highlight areas of 'in–work poverty' where people are working but still do not have a sufficient income, often because of precarious, low paid, or low skilled employment. DMBC has a real opportunity to help these people out of poverty by supporting the creation of training and education opportunities and of higher skilled jobs in the longer–term. In the short–term it is at least possible to help mitigate as many of the effects of poverty in these areas. I was also able to identify areas where people spend a high proportion of their income on housing. DMBC can again use this information to support such individuals, for example in the short–term with discretionary housing payments (DHPs) or in the long–term by identifying sites for development or redevelopment of affordable housing.

Based on the results of the simulation, there are areas of Doncaster with high levels of social isolation, which demonstrably have negative effects on physical and mental health. By identifying these areas with higher numbers of individuals experiencing social isolation DMBC have the opporunity to mitigate this.

One of the main changes in contemporary welfare policy is the transition to Universal Credit. Universal Credit is highly controversial and can affect claimants in a number of ways. For example, when transitioning to Universal Credit from benefits that are being phased out, claimants can experience delays of several weeks. This can cause these individuals to drop into poverty, which can take a significant period of time to recover from. Having identified areas with high numbers of people who are likely to transition to Universal Credit, DMBC could use this information to intervene at this crucial period and provide short–term support to individuals at risk of entering poverty due to delays in payments. This short–term, low–level, intervention could stop people at risk dropping into poverty and could have long–term benefits if they do not spend a long time recovering.

Universal Credit is also not always sufficient to support claimants out of poverty. Individuals can earn a certain amount—their work allowance—before their payment of Universal Credit is gradually reduced—the taper amount. These are too low for people who earn just enough to have their Universal Credit payments reduced completely. Using the simulation I identified individuals who earned an income high enough so that they were no longer eligible for Universal Credit, but who still did not earn enough to move them out of relative poverty. For these individuals moving out of poverty will be difficult, as they cannot claim social welfare payments to top–up their income to above poverty levels, and cannot easily move out of low–paid employment. In the absence of national government policy to increase the provision of Universal Credit, DMBC could consider supporting such individuals in the short–term as well as plan for the transition to higher quality employment outlined in the 'Doncaster Working' policy.

These analyses represent a unique aspect of this research that would not be possible with traditional data sources separately, so provide valuable evidence about the likely spatial characteristics of individuals in small areas in Doncaster.

8.1 Opportunities for further study

A longitudinal understanding of resilience could allow for the construction of a dynamic spatial microsimulation model. A dynamic model could be used to evaluate the effect of life events on resilience and health outcomes for individuals and their friends and family members.

The accuracy of the spatial microsimulation model could be improved further by using a *k*–means clustering technique (Smith et al., 2009: 1259). Rather than using one configuration of constraints for every output area, cluster analysis could identify statistical groups of output areas that would perform better using adjusted constraint configurations, thus improving the model fit further. This approach could be tested to see if it could improve the accuracy of this model further. However, the accuracy of the model is already well within conventional tolerances and further attempts at improvements have diminishing returns in accuracy. Further, *Understanding Society* has a far greater sample size than many surveys so has greater diversity from which to sample individuals in each small area.

The analysis of straight–line distance between resilient areas and health amenities has the advantage of being simple and quick to set up and obtain results from, and as such the technique is fairly common (Smith et al., 2006: 913). The accuracy of this model could be further enhanced by using network analysis (Morrissey et al., 2010: 17) or a spatial interaction model (Morrissey et al., 2010: 14–15; Smith et al., 2006: 914). These have the advantage of being able to include additional parameters that might affect the level of access to the amenity, such as the availability of public transport or the road network more generally, as well as behavioural factors such as the 'attractiveness' of the facility to service users.

8.2 Final reflections on resilience

> We would not wish a resilience perspective to become an excuse for blaming those who succumb to the effects of poverty or adversity of any kind just because it may be possible to identify some people and places who do, to an extent, 'beat the odds' (Mitchell et al., 2009: 22).

For a social scientist interested in mitigating health and social inequalities, health resilience is a problematic topic of study. On the one hand, in an ideal world resilience would be unnecessary if policies and interventions were taken to mitigate and reduce the inequalities in health. On the other hand austerity, dramatic reductions in public spending, and poverty are a reality for millions of people in the UK and Europe. In the absence of the ideal, health resilience could be an important way local authorities protect the health of their residents as much as possible by creating an environment that supports positive health and mental health outcomes. With this research I have identified a number of characteristics that could create healthly environments and areas where practical interventions could make the most of diminishing resources.

These are not replacements for proper investment in public services, a universal health care system, and an adequate welfare state.

> . . . we may once have hoped that increasing resilience was a way of saving money by proofing people against the difficulties they face, there is little reason to think that health benefits

> bought this way are any cheaper than those which would come from reducing the underlying disadvantage itself (Wilkinson, 2006).

Health inequalities may be mitigated by resilience, but this is akin to treating the symptoms of a disease. Health resilience may mitigate some of the worst symptoms but it is, by definition, only necessary if the health inequalities remain that it is trying to eliminate.

Bibliography

Age UK (2017) Conditions and illnesses. Available from: http://www.ageuk.org.uk/health-wellbeing/conditions-illnesses/ (accessed July 2017).

Albor C, Uphoff EP, Stafford M, et al. (2014) The effects of socioeconomic incongruity in the neighbourhood on social support, self-esteem and mental health in England. *Social Science & Medicine* 111: 1–9.

Anderson B (2007) Creating small–area income estimates: Spatial microsimulation modelling. Department for Communities and Local Government.

Anderson B (2013) Estimating small-area income deprivation: An iterative proportional fitting approach. In: Tanton R and Edwards KL (eds), *Spatial Microsimulation: A reference guide for users*, Dordrecht: Springer.

Aranda MP and Lincoln KD (2011) Financial strain, negative interaction, coping styles, and mental health among low–income Latinos. *Race and Social Problems* 3: 280–297.

Ballas D and Dorling D (2011) Human–scaled visualizations and society. In: Nyerges T, Couclelis H, and McMaster R (eds), *The Sage handbook of GIS and society*, London: Sage, pp. 177–201.

Ballas D, Rossiter D, Thomas B, et al. (2005) Geography matters: Simulating the local impacts of national social policies. Joseph Rowntree Foundation. Available from: https://www.jrf.org.uk/report/

geography-matters-simulating-local-impacts-national-social-policies.

Ballas D, Clarke G, Dorling D, et al. (2007) Using SimBritain to model the geographical impact of national government policies. *Geographical Analysis* 39: 44–77.

Ballas D, Clarke G, Hynes S, et al. (2013) A review of microsimulation for policy analysis. In: O'Donoghue C, Ballas D, Clarke G, et al. (eds), *Spatial microsimulation for rural policy analysis*, Heidelberg; New York; Dordrecht; London: Springer.

Ballas D, Clarke G, Dorling D, et al. (2005b) SimBritain: A spatial microsimulation approach to population dynamics. *Population, Space and Place* 11: 13–34.

Bambra C, Cairns J, Kasim A, et al. (2015) This divided land: An examination of regional inequalities in exposure to brownfield land and the association with morbidity and mortality in england. *Health & Place* 34.

Barford A and Dorling D (2006) A new approach to mapping the world: Visualizing facets of international health. *National Medical Journal of India* 4(19): 234–235.

Bartley M, Blane D, Brunner E, et al. (2003) Social determinants of health: The solid facts. 2nd ed. Wilkinson R and Marmot M (eds), World Health Organization. Available from: http://www.euro.who.int/en/publications/abstracts/social-determinants-of-health.-the-solid-facts (accessed January 2015).

Bartley M, Ferrie J and Montgomery SM (2006) Health and labour market disadvantage: Unemployment, non–employment, and job insecurity. 2nd ed. In: Marmot M and Wilkinson RG (eds), *Social determinants of health*, Oxford: Oxford University Press.

Bartley M (ed.) (2006) *Capability and Resilience: Beating the Odds*. UCL Department of Epidemiology; Public Health. Available from: http://www.ucl.ac.uk/capabilityandresilience/beatingtheoddsbook.pdf (accessed

June 2015).

Bellis M, Hughes K, Leckenby N, et al. (2014) National household survey of adverse childhood experiences and their relationship with resilience to health–harming behaviors in England. *BMC Medicine* 12.

Bibby P and Brindley P (2013a) The 2011 rural–urban classification for small area geographies: A user guide and frequently asked questions (v1.0). Department for Environment Food & Rural Affairs; Department for Communities; Local Government; Office for National Statistics; Welsh Government. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/239478/RUC11user_guide_28_Aug.pdf (accessed May 2017).

Bibby P and Brindley P (2013b) Urban and rural area definitions for policy purposes in England and Wales: Methodology (v1.0). Department for Environment Food & Rural Affairs; Department for Communities; Local Government; Office for National Statistics; Welsh Government. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/239477/RUC11methodologypaperaug_28_Aug.pdf (accessed May 2017).

Biogeo (2016) Global Administrative Boundaries (GADM). Available from: http://data.biogeo.ucdavis.edu/ (accessed May 2017).

Birkin M and Clarke M (1988) SYNTHESIS—a synthetic spatial information system for urban and regional analysis: Methods and examples. *Environment and Planning A* 20: 1645–1671.

Birkin M and Clarke M (1989) The generation of individual and household incomes at the small area level using Synthesis. *Regional Studies* 23(6): 535–548.

Bivand R and Rundel C (2017) *Rgeos: Interface to geometry engine - open source ('geos')*. Available from: https://CRAN.R-project.org/package=

rgeos.

Bivand R, Keitt T and Rowlingson B (2017) *Rgdal: Bindings for the 'geospatial' data abstraction library.* Available from: https://CRAN.R-project.org/package=rgdal.

Blickem C, Blakeman T, Kennedy A, et al. (2013) The clinical and cost-effectiveness of the BRinging Information and Guided Help Together (BRIGHT) intervention for the self-management support of people with stage 3 chronic kidney disease in primary care: Study protocol for a randomized controlled trial. *Trials* 14(28).

Blocker AW (2013) Ipfp: Fast implementation of the iterative proportional fitting procedure in c. Available from: http://CRAN.R-project.org/package=ipfp.

Booth A, Papaioannou D and Sutton A (2012) *Systematic approaches to a successful literature review.* London: Sage.

Brunner E and Marmot M (2006) Social organization, stress, and health. 2nd ed. In: Marmot M and Wilkinson RG (eds), *Social determinants of health*, Oxford: Oxford University Press.

Buck N and McFall S (2012) Understanding society: Design overview. *Longitudinal and Life Course Studies* 3(1): 5–17.

Butler P and Asthana A (2017) Universal credit flaws pushing claimants towards debt and eviction: Recipients falling into rent arrears because of payment delays, forcing them to turn to food banks, guardian investigation reveals. Available from: https://goo.gl/wWTPzz (accessed July 2017).

Cairns JM (2013) Defying the odds?: Identifying and understanding the relationship between health and resilience. PhD thesis, University of Durham. Available from: http://etheses.dur.ac.uk/127/1/FINAL_REVISED_JMC_THESIS_2013.pdf (accessed April 2017).

Cairns JM, Curtis SE and Bambra C (2012) Defying deprivation: A cross-sectional analysis of area level health. *Health & Place* 18: 928–933.

Cairns–Nagi JM and Bambra C (2013) Defying the odds: A mixed-methods study of health resilience in deprived areas of England. *Social Science and Medicine* 91: 229–237.

Cameron J (2013) Work–related needs and experiences of people recovering from mental health problems. PhD thesis, University of Brighton. Available from: http://eprints.brighton.ac.uk/12263/ (accessed April 2017).

Campbell M and Ballas D (2013) A spatial microsimulation approach to economic policy analysis in Scotland. *Regional Science Policy & Practice* 5(3).

Campbell M and Ballas D (2016) SimAlba: A spatial microsimulation approach to the analysis of health inequalities. *Frontiers in Public Health* 4.

Campbell MH (2011) Exploring the social and spatial inequalities of ill-health in Scotland: A spatial microsimulation approach. PhD thesis, University of Sheffield. Available from: http://etheses.whiterose.ac.uk/1942/.

Canonical (2016) *Ubuntu*. Available from: https://www.ubuntu.com/desktop/developers.

Care Quality Commission (2016) 619 results from our latest checks on all services in Doncaster local authority. Available from: https://goo.gl/NhRi48 (accessed October 2016).

Care Quality Commission (2017) CQC care directory - zip format. Available from: https://goo.gl/bNrhfg (accessed June 2017).

Cassells R, Miranti R and Harding A (2013) Building a static spatial microsimulation model: Data preparation. In: Tanton R and Edwards KL (eds), *Spatial Microsimulation: A reference guide for users*, Dordrecht:

Springer.

Choi H and Marks N (2013) Marital quality, socioeconomic status, and physical health. *Journal of Marriage and Family* 75(4): 903–919.

Cleal B, Panton UH, Willaing I, et al. (2017) Diabetes and depression in Denmark 1996–2010: National data stratified by occupational status and annual income. *Diabetic Medicine* 34(1): 108–114.

Codecov LLC (2016) *Codecov: Code coverage done right*. Available from: https://codecov.io/.

Corman H, Curtis MA, Noonan K, et al. (2016) Maternal depression as a risk factor for children's inadequate housing conditions. *Social Science & Medicine* 149: 76–83.

Courtin E and Knapp M (2017) Social isolation, loneliness and health in old age: A scoping review. *Health & Social Care in the Community* 25(3): 799–812.

Department for Communities and Local Government (2015) English indices of deprivation 2015. Available from: https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015 (accessed January 2017).

Department for Work and Pensions (2017) Use of Discretionary Housing Payments GB: Analysis of returns from local authorities, april 2016 – march 2017. Gov.uk. Available from: https://goo.gl/Bzkzqq (accessed July 2017).

Department of Health, Welsh Government, Department of Health (NI), et al. (2016) UK Chief Medical Officers' low risk drinking guidelines. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/545937/UK_CMOs___report.pdf (accessed May 2017).

Deurzen I van, Ingen E van and Oorschot WJH van (2015) Income inequality and depression: The role of social comparisons and coping resources. *European Sociological Review* 31(4): 477–489.

Doran T, Drever F and Whitehead M (2006) Health underachievement and overachievement in English local authorities. *Journal of Epidemiology and Community Health* 60: 686–693.

Edwards KL and Clarke GP (2009) The design and validation of a spatial microsimulation model of obesogenic environments for children in Leeds, UK: SimObesity. *Social Science and Medicine* 69: 1127–1134.

Edwards KL and Tanton R (2013) Validation of spatial microsimulation models. In: Tanton R and Edwards KL (eds), *Spatial Microsimulation: A reference guide for users*, Dordrecht: Springer.

Edwards KL, Clarke GP, Ransley JK, et al. (2010) The neighbourhood matters: Studying exposures relevant to childhood obesity and the policy implications in Leeds, UK. *Journal of Epidemiology and Community Health* 64(3). Available from: http://jech.bmj.com/content/64/3/194.

Edwards KL, Clarke GP, Thomas J, et al. (2011) Internal and external validation of spatial microsimulation models: Small area estimates of adult obesity. *Applied Spatial Analysis & Policy* 4: 281–300.

Erskine J, Kvavilashvili L, Myers L, et al. (2016) A longitudinal investigation of repressive coping and ageing. *Aging & Mental Health* 20(10).

Fazel S and Baillargeon J (2011) The health of prisoners. *Lancet* 377(9769): 956–965.

Fenge L-A, Hean S, Worswick L, et al. (2012) The impact of the economic recession on well-being and quality of life of older people. *Health and Social Care in the Community* 20(6): 617–624.

Field A, Miles J and Field Z (2012) *Discovering statistics using R*. London: Sage.

Fienberg SE (1970) An iterative procedure for estimation in contingency tables. *The Annals of Mathematical Statistics* 41(3): 907–917. Available from: http://goo.gl/J9SvNL.

Garmezy N and Masten AS (1986) Stress, competence, and resilience: Common frontiers for therapist and psychopathologist. *Behavior Therapy* 17: 500–521.

Garmezy N and Streitman S (1974) Children at risk: The search for the antecedents of schizophrenia. part 1: Conceptual models and research methods. *Schizophrenia Bulletin* 1(8): 14–90. Available from: http://goo.gl/kVgmbq.

Garmezy N, Masten AS and Tellegen A (1984) The study of stress and competence in children: A building block for developmental psychopathology. *Child Development* 55(1): 97–111. Available from: http://www.jstor.org/stable/1129837.

Gastner MT and Newman MEJ (2004) Diffusion-based method for producing density equalizing maps. *Proceedings of the National Academy of Sciences in the United States of America* 101(20): 7499–7504.

GitHub, Inc. (2015) GitHub. Available from: https://github.com (accessed September 2017).

Glonti K, Gordeev V, Goryakin Y, et al. (2015) A systematic review on health resilience to economic crises. *Plos One* 10(4).

Google (2017a) Google Maps Geocoding API. Available from: https://developers.google.com/maps/documentation/geocoding/start (accessed June 2017).

Google (2017b) Google Maps Places API. Available from: https://developers.google.com/places/ (accessed July 2017).

Gov.UK (2017) Personal Independence Payment (PIP). Available from: https://www.gov.uk/pip/overview (accessed March 2017).

Gov.uk (2016a) What qualification levels mean. Available from: https://www.gov.uk/what-different-qualification-levels-mean/overview (accessed

February 2017).

Gov.uk (2016b) What qualification levels mean: List of qualification levels. Available from: https://www.gov.uk/what-different-qualification-levels-mean/list-of-qualification-levels (accessed February 2017).

Gov.uk (2017a) Benefits cap: Benefit cap amounts. Available from: https://www.gov.uk/benefit-cap/benefit-cap-amounts (accessed July 2017).

Gov.uk (2017b) Universal Credit: If you're claiming other benefits. Available from: https://www.gov.uk/universal-credit/claiming-other-benefits (accessed July 2017).

Gov.uk (2017c) Universal Credit: What you'll get. Available from: https://www.gov.uk/universal-credit/what-youll-get (accessed July 2017).

Haycock D and Smith A (2014) Sports participation and health during periods of educational transition: A study of 30–35 year–olds in north–west England. *Sport, Education and Society* 19(2): 168–185.

Hiilamo H (2014) Is income inequality 'toxic for mental health'? An ecological study on municipal level risk factors for depression. *PLoS ONE* 9(3).

Hosseinpour AR, Stewart Williams J, Amin A, et al. (2012) Social determinants of self–reported health in women and men: Understanding the role of gender in population health. *PLoS ONE* 7(4).

Institute of Alcohol Studies (2014) Alcohol pricing factsheet. Available from: http://www.ias.org.uk/uploads/pdf/Factsheets/Alcohol%20pricing%20factsheet%20April%202014.pdf (accessed May 2017).

Jefferis BJ, Nazareth I, Marston L, et al. (2010) 024 prospective cohort study of unemployment and clinical depression in Europe and Chile: The Predict study. *Journal of Epidemiology & Community Health*, BMJ Publishing Group Ltd 64(Suppl 1): A9–A10.

Johnston D, Schurer S and Shields M (2013) Exploring the intergenerational persistence of mental health: Evidence from three generations.

Journal Of Health Economics 32(6).

Johnston R and Pattie C (1993) Entropy-maximizing and the Iterative Proportional Fitting procedure. *The Professional Geographer* 45(3): 317–322. Available from: http://dx.doi.org/10.1111/j.0033-0124.1993.00317.x.

Jones PM, Lovelace R and Dumont M (2017) *RakeR: Easy spatial microsimulation (raking) in r.* Available from: https://CRAN.R-project.org/package=rakeR.

Kaminska O (2015) Understanding Society user support: Support #437. Available from: https://goo.gl/Mv62Zx (accessed October 2015).

Khlat M, Sermet C and Le Pape A (2004) Increased prevalence of depression, smoking, heavy drinking and use of psycho–active drugs among unemployed men in France. *European Journal of Epidemiology*, Springer 19(5): 445–451.

Kioumourtzoglou M, Power M, Hart JE, et al. (2017) The association between air pollution and onset of depression among middle–aged and older women. *American Journal of Epidemiology* 185(9): 801–809.

Knies G (ed.) (2015) *Understanding Society—UK household longitudinal study: Waves 1–5, 2009–2014, user manual.* Colchester: University of Essex, Institute for Social; Economic Research. Available from: https://goo.gl/vfXNpZ.

Knott KE (2013) Alcohol consumption and adherence to self–care behaviours in Type 2 diabetes: The inclusion of brief interventions for alcohol in diabetes care. PhD thesis, University of Leicester. Available from: https://lra.le.ac.uk/handle/2381/28265 (accessed April 2017).

Korhonen K, Remes H and Martikainen P (2017) Education as a social pathway from parental socioeconomic position to depression in late adolescence and early adulthood: A Finnish population–based register study. *Social Psychiatry and Psychiatric Epidemiology* 52(1): 105–116.

Lai WW and Oei TPS (2014) Coping in parents and caregivers of children with Autism Spectrum Disorders (ASD): A review. *Review Journal of Autism and Developmental Disorders* 1(3): 207–224.

Livingston M and Lee D (2014) ‘The Glasgow effect?’—The result of the geographical patterning of deprived areas? *Health & Place* 29: 1–9.

Lovelace R (2013) The energy costs of commuting: A spatial microsimulation approach. PhD thesis, University of Sheffield. Available from: http://etheses.whiterose.ac.uk/id/eprint/5027.

Lovelace R and Ballas D (2013) ‘Truncate, replicate, sample’: A method for creating integer weights for spatial microsimulation. *Computers, Environment and Urban Systems* 41: 1–11. Available from: http://www.sciencedirect.com/science/article/pii/S0198971513000240.

Lovelace R and Dumont M (2016) *Spatial microsimulation with R*. Boca Raton, FL: CRC Press; Taylor & Francis.

Lovelace R, Birkin M, Ballas D, et al. (2015) Evaluating the performance of Iterative Proportional Fitting for spatial microsimulation: New tests for an established technique. *Journal of Artificial Societies and Social Simulation* 18(2). Available from: http://jasss.soc.surrey.ac.uk/18/2/21.html.

Lumley T (2010) *Complex surveys: A guide to analysis using R*. Hoboken, NJ: Wiley.

Luthar S, Cicchetti D and Becker B (2000) The construct of resilience: A critical evaluation and guidelines for future work. *Child Development* 71(3): 543–562.

Mackenbach JP, Kulhanova I, Bopp M, et al. (2015) Variations in the relation between education and cause-specific mortality in 19 European populations: A test of the ’fundamental causes’ theory of social inequalities in health. *Social Science & Medicine* 127: 51–62.

Maheswaran R, Pearson T, Strong M, et al. (2014) Assessing the impact of selective migration and care homes on geographical inequalities in health: A total population cohort study in sheffield. *Spatial and Spatio-temporal Epidemiology* 10: 85–97. Available from: http://dx.doi.org/10.1016/j.sste.2014.06.001.

Marmot MG, Shipley MJ and Rose G (1984) Inequalities in death—specific explanations of a general pattern? *The Lancet* 1(8384): 1003–1006.

Mason R (2017) Universal credit rollout should be delayed as it is 'failing too many people': Citizens Advice produces fresh evidence that new benefit is causing debt and financial insecurity among recipients. Available from: https://www.theguardian.com/society/2017/jul/06/universal-credit-rollout-should-be-delayed-as-it-is-failing-too-many-people (accessed July 2017).

Mastrocola E, Taylor A and Chew–G. C (2015) Access to healthcare for long-term conditions in women involved in street-based prostitution: A qualitative study. *BMC Family Practice* 16.

Matthews S and Sykes S (2012) Exploring health priorities for young people leaving care. *Child Care in Practice* 18(4): 393–407.

Matthews T, Danese A, Wertz J, et al. (2015) Social isolation and mental health at primary and secondary school entry: A longitudinal cohort study. *Journal of the American Academy of Child and Adolescent Psychiatry* 54(3): 225–232.

McLaren ME, Szymkowicz SM, Kirton JW, et al. (2015) Impact of education on memory deficits in subclinical depression. *Archives of Clinical Neuropsychology* 30(5): 387–393.

McMunn A, Breeze E, Goodman A, et al. (2006) Social determinants of health in older age. 2nd ed. In: Marmot M and Wilkinson RG (eds), *Social determinants of health*, Oxford: Oxford University Press.

Mertz J (1991) Microsimulation: A survey of principles, developments and applications. *International Journal of Forecasting* 7: 77–104.

Ministry of Justice (2015) Prisons in england and wales. Available from: http://www.justice.gov.uk/contacts/prison-finder (accessed October 2015).

Mitchell R, Gibbs J, Tunstall H, et al. (2009) Factors which nurture geographical resilience in Britain: A mixed methods study. *Journal of Epidemiology and Community Health* 63: 18–23.

Moher D, Liberati A, Tetzlaff J, et al. (2009) Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ* 339.

Monk RL, Heim D, Qureshi A, et al. (2015) 'I have no clue what i drunk last night': Using smartphone technology to compare in–Vivo and retrospective self–Reports of alcohol consumption. *PLoS ONE* 10(5).

Morrissey K, Hynes S, Clarke G, et al. (2010) Examining the factors associated with depression at the small area level in Ireland using spatial microsimulation techniques. *Irish Geography* 45(1): 1–22.

Morrissey K, Ballas D, Clarke G, et al. (2013) Spatial access to health services. In: O'Donoghue C, Ballas D, Clarke G, et al. (eds), *Spatial microsimulation for rural policy analysis*, Heidelberg; New York; Dordrecht; London: Springer.

Mossakowski K (2009) The influence of past unemployment duration on symptoms of depression among young women and men in the United States. *American Journal of Public Health* 99(10): 1826–1832.

Mõttus R, Gale CR, Starr JM, et al. (2012) 'On the street where you live': Neighbourhood deprivation and quality of life among community–dwelling older people in Edinburgh, Scotland. *Social Science & Medicine* 74(9): 1368–1374.

Mukherjee S, Trepka MJ, Pierre–Victor D, et al. (2016) Racial/ethnic disparities in antenatal depression in the United States: A systematic review. *Matern Child Health J* 20: 1780–1797.

Nash SG, McQueen A and Bray JH (2005) Pathways to adolescent alcohol use: Family environment, peer influence, and parental expectations. *Journal of Adolescent Health* 37: 19–28.

New Zealand Ministry of Business, Innovation and Employment (2015) Modelled territorial authority gross domestic product. Available from: https://goo.gl/XuQqjy (accessed December 2015).

NHS Choices (2016) Cardiovascular disease. Available from: http://www.nhs.uk/Conditions/cardiovascular-disease/Pages/Introduction.aspx (accessed July 2017).

NHS Digital (2016) Quality and Outcomes Framework. Available from: http://content.digital.nhs.uk/qof (accessed April 2017).

NHS Employers (2011–2012) Summary of 2011/12 QOF indicator changes, points and thresholds. Available from: https://goo.gl/L1HQnj (accessed April 2017).

Nomis (2013a) 2011 census data for England and Wales on Nomis. Available from: http://www.nomisweb.co.uk/census/2011 (accessed June 2014).

Nomis (2013b) QS601EW - economic activity. Available from: https://www.nomisweb.co.uk/census/2011/qs601ew.pdf (accessed March 2017).

Nomis (2014) QS409UK - persons per room - households. Available from: https://www.nomisweb.co.uk/census/2011/qs409uk.pdf (accessed March 2017).

Nomis and Office for National Statistics (2017) Labour market profile - Doncaster. Available from: https://www.nomisweb.co.uk/reports/lmp/la/1946157121/report.aspx?town=doncaster (accessed February 2017).

Office for National Statistics (2011) 2011 Census: Aggregate data (England and Wales). Available from: http://infuse.mimas.ac.uk (accessed October 2011).

Office for National Statistics (2014a) 2011 Census glossary of terms. Available from: https://www.ons.gov.uk/ons/guide-method/census/2011/

census-data/2011-census-data/2011-first-release/2011-census-definitions/2011-census-glossary.pdf (accessed January 2017).

Office for National Statistics (2014b) 2011 Census variable and classification information: Part 4 derived variables. Available from: http://goo.gl/BN0LDp (accessed October 2015).

Office for National Statistics (2015a) 2011 area classifications. Available from: http://goo.gl/AFmfkx (accessed October 2015).

Office for National Statistics (2015b) 2011 Census microdata individual safeguarded sample (local authority): England and Wales. Available from: http://dx.doi.org/10.5255/UKDA-SN-7682-1 (accessed December 2015).

Office for National Statistics (2015c) Life expectancy at birth and age 65 by local areas in England and Wales. Available from: https://goo.gl/fkdpBp (accessed October 2016).

Office for National Statistics (2015d) Pen portraits for the 2011 area classification for output areas. Available from: https://goo.gl/SBozTV (accessed April 2017).

Office for National Statistics (2015) SOC2010 volume 3: National Statistics socio-economic classification (NS-SEC rebased on the SOC2010). Available from: http://goo.gl/Y8C6ML.

Office for National Statistics (n.d.) Statistical disclosure control for 2011 census. Available from: https://goo.gl/5DjiYF (accessed February 2017).

Office for National Statistics and National Archives (2016) Census 2011: Questionnaires, delivery, completion and return. Available from: http://goo.gl/LqiRTF (accessed March 2016).

Office for National Statistics, National Records of Scotland and Northern Ireland Statistics and Research Agency (2017) 2011 census aggregate data.

Orcutt GH (1960) Simulation of economic systems. *The Americal Eco-*

nomic Review 50(5).

Orcutt GH (1962) Microanalytic models of the United States economy: Need and development. *The Americal Economic Review* 52(2): 229–240. Available from: http://www.jstor.org/stable/1910887.

Ordnance Survey (2017) OS Open Greenspace. Available from: https://www.ordnancesurvey.co.uk/business-and-government/products/os-open-greenspace.html (accessed July 2017).

O'Donoghue C, Ballas D, Clarke G, et al. (2013) *Introduction*. O'Donoghue C, Ballas D, Clarke G, et al. (eds), Heidelberg; New York; Dordrecht; London: Springer.

O'Dougherty M and Wright FS (1990) Children born at medical risk: Factors affecting vulnerability and resilience. In: Rolf J, Masten AS, Cicchetti D, et al. (eds), *Risk and protective factors in the development of psychopathology*, Cambridge: Cambridge University Press.

Pantzar A, Atti RA, Fratiglioni L, et al. (2017) Cognitive performance in unipolar old–age depression: A longitudinal study. *International Journal of Geriatric Psychiatry* 32: 675–684.

Pascal M, Corso M, Chanel O, et al. (2013) Assessing the public health impacts of urban air pollution in 25 European cities: Results of the Aphekom project. *Science of The Total Environment* 449: 390–400.

Pebesma E and Bivand R (2018) *Sp: Classes and methods for spatial data*. Available from: https://CRAN.R-project.org/package=sp.

Pickett K and Wilkinson RG (2015) Income inequality and health: A causal review. *Social Science & Medicine* 128: 316–326.

Poetz A, Eyles JD, Elliott S, et al. (2007) Path analysis of income, coping and health at the local level in a Canadian context. *Health and Social Care in the Community* 15(6): 542–552.

Poortinga W (2012) Community resilience and health: The role of bonding, bridging, and linking aspects of social capital. *Health & Place* 18: 286–295.

Poynter B, Hon MS, Diaz–Granados N, et al. (2009) Sex differences in the prevalence of post–stroke depression: A systematic review. *Psychosomatics* 50(6): 563–569.

Presser L (2016) Changeover to universal credit is driving people into long–term debt. Guardian. Available from: https://www.theguardian.com/society/2016/jan/27/changeover-to-universal-credit-driving-people-into-debt (accessed July 2017).

PRISMA (2013) Transparent reporting of systematic reviews and meta-analyses. Available from: http://www.prisma-statement.org (accessed January 2015).

Procter KL, Clarke GP, Ransley JK, et al. (2008) Micro–level analysis of childhood obesity, diet, physical activity, residential and social capital variables: Where are the obesogenic environments in Leeds? *Area* 40(3): 323–340.

Pruckner N and Holthoff–Detto V (2017) Antidepressant pharmacotherapy in old–age depression: A review and clinical approach. *European Journal of Clinical Pharmacology* 73: 661–667.

Public Health England (2015) Mortality rankings. Available from: http://healthierlives.phe.org.uk/topic/mortality/comparisons#are/ /par/E92000001/ati/101/pat/101 (accessed January 2017).

Public Health England (2016) Common mental health disorders. Available from: http://fingertips.phe.org.uk/profile-group/mental-health/profile/common-mental-disorders (accessed April 2017).

R Core Team (2017) *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.R-project.org/.

Reeves A, McKee M, Clair A, et al. (2016) Reductions in housing benefit increases symptoms of depression in low–income UK households. *European*

Journal of Public Health 26.

Reeves D, Blickem C, Vassilev I, et al. (2014) The contribution of social networks to the health and self-management of patients with long–Term conditions: A longitudinal study. *Plos One* 9(6).

Robards J, Evandrou M, Falkingham J, et al. (2012) Marital status, health and mortality. *Maturitas* 73(4): 295–299. Available from: http://www.sciencedirect.com/science/article/pii/S0378512212002691.

Robinson M, Raine G, Robertson S, et al. (2015) Peer support as a resilience building practice with men. *Journal of Public Mental Health* 14(4): 196–204.

Rose GA (1992) *The strategy of preventive medicine*. Oxford: Oxford University Press.

Rutter M (1987) Psychosocial resilience and protective mechanisms. *Amer. J. Orthopsychiat.* 57(3).

Sacker A, Worts D and McDonough P (2009) Social influences on trajectories of self-rated health: Evidence from Britain, Germany, Denmark and the USA. *Journal of Epidemiology and Community Health* 65(2): 130–136.

Sameroff AJ and Seifer R (1983) Familial risk and child competence. *Child Development* 54: 1254–1268.

Sameroff AJ and Seifer R (1990) Early contributors to developmental risk. In: Rolf J, Masten AS, Cicchetti D, et al. (eds), *Risk and protective factors in the development of psychopathology*, Cambridge: Cambridge University Press.

Sanchez-Villegas A, Schlatter J, Ortuno F, et al. (2008) Validity of a self-reported diagnosis of depression among participants in a cohort study using the structured clinical interview for DSM-IV (SCID-I). *BMC Psychiatry* 8(1): 43. Available from: https://doi.org/10.1186/1471-244X-8-43.

Schoon I (2006) *Risk and resilience: Adaptations in changing times.*

Cambridge: Cambridge University Press.

Senior M (2002) Deprivation indicators. In: Rees P, Martin D, and Williamson P (eds), *The census data system*, Chichester: John Wiley & Sons.

Shelter (2016) Discretionary housing payments (DHP). Available from: https://england.shelter.org.uk/housing_advice/housing_benefit/discretionary_housing_payments_dhp (accessed July 2017).

Smith DM, Clarke GP, Ransley J, et al. (2006) Food access and health: A microsimulation framework for analysis. *Studies in Regional Science* 35(4): 909–927. Available from: https://www.jstage.jst.go.jp/article/srs/35/4/35_4_909/_article.

Smith DM, Clarke GP and Harland K (2009) Improving the synthetic data generation process in spatial microsimulation models. *Environment and Planning A* 41: 1251–1268.

Smith DM, Pearce JR and Harland K (2011) Can a deterministic spatial microsimulation model provide reliable small-area estimates of health behaviours? An example of smoking prevalence in New Zealand. *Health & Place* 17(2): 618–624. Available from: http://www.sciencedirect.com/science/article/pii/S1353829211000037.

Smith T, Noble M, Noble S, et al. (2015) The English indices of multiple deprivation 2015: Technical report. Department for Communities; Local Government. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/464485/English_Indices_of_Deprivation_2015_-_Technical-Report.pdf (accessed January 2017).

Stankunas M, Kalediene R, Starkuviene S, et al. (2006) Duration of unemployment and depression: A cross-sectional survey in Lithuania. *BMC Public Health* 6(174).

Sull A, Harland N and Moore A (2015) Resilience of health-care workers in the uk and a cross-sectional survey. *Journal Of Occupational Medicine*

And Toxicology 10.

Swirrl IT Ltd. and Department for Communities and Local Government (2017) English indices of deprivation 2015: Postcode lookup. Available from: http://imd-by-postcode.opendatacommunities.org/ (accessed March 2017).

Tanton R and Edwards K (2013a) Introduction to spatial microsimulation: History, methods and applications. In: Tanton R and Edwards K (eds), *Spatial Microsimulation: A reference guide for users*, Dordrecht: Springer.

Tanton R and Edwards KL (2013b) Limits of static spatial microsimulation models. In: Tanton R and Edwards KL (eds), *Spatial microsimulation: A reference guide for users*, Dordrecht: Springer.

Tarricone I, Stivanello E, Poggi F, et al. (2012) Ethnic variation in the prevalence of depression and anxiety in primary care: A systematic review and meta–analysis. *Psychiatry Research* 195(3): 91–106.

Tennekes M (2017) *Tmap: Thematic maps*. Available from: https://github.com/mtennekes/tmap.

The Marmot Review Team (2010) Fair society, healthy lives: The Marmot review. University College London Institute of Health Equity. Available from: http://www.instituteofhealthequity.org/projects/fair-society-healthy-lives-the-marmot-review (accessed October 2013).

The Resilience Centre (2017) True Resilience Scale. Available from: http://www.resiliencecenter.com/assessments/true-resilience-scale/ (accessed April 2017).

Tillman L, Simmonds M and Marsh C (2017) Borough strategy progress: Update for team doncaster. Team Doncaster.

Tinson A, Ayrton C, Barker K, et al. (2016) Monitoring poverty and social exclusion 2016. Joseph Rowntree Foundation. Available from: https://www.jrf.org.uk/report/monitoring-poverty-and-social-exclusion-2016 (accessed June 2017).

Tomintz MN, Clarke GP and Rigby JE (2008) The geography of smoking in Leeds: Estimating individual smoking rates and the implications for the location of stop smoking services. *Area* 30(3): 341–353.

Torvalds L (2015) Git: Distributed version control system. Available from: https://git-scm.com/ (accessed October 2015).

Townsend P, Phillimore P and Beattie A (1988) *Health and deprivation: Iinequality and the north.* Beckenham: Croom Helm.

Travis CI, GmbH (2016) *Travis CI.* Available from: https://travis-ci.com/.

Tunstall H, Mitchell R, Gibbs J, et al. (2007) Is economic adversity always a killer? Disadvantaged areas with relatively low mortality rates. *Journal of Epidemiology and Community Health* 61: 337–343.

UK Data Service (2016) UK Data Service data catalogue record for Understanding Society: Waves 1–6, 2009–2015. Available from: https://discover.ukdataservice.ac.uk/catalogue/?sn=6614&type=Data%20catalogue (accessed February 2017).

University of Essex, Institute for Social and Economic Research, NatCen Social Research, et al. (2016) *Understanding society: Waves 1–6, 2009-2015 [data collection; serial number: 6614].* 8th ed. Colchester, Essex: UK Data Service [distributor]. Available from: http://dx.doi.org/10.5255/UKDA-SN-6614-6 (accessed February 2017).

Wagnild G (2017) Personal communication.

Webber MP and Huxley PJ (2007) Measuring access to social capital: The validity and reliability of the Resource Generator–UK and its association with common mental disorder. *Social Science & Medicine* 65(3): 481–492. Available from: http://www.sciencedirect.com/science/article/pii/S0277953607001578.

Wel KA, Bambra C, Dragano N, et al. (2015) Risk and resilience: Health inequalities, working conditions and sickness benefit arrangements: An analysis of the 2010 European Working Conditions survey. *Sociology of Health & Illness* 37(8): 1157–1172.

Werner EE and Smith RS (1977) *Kauai's children come of age*. Honolulu: The University Press of Hawaii.

Werner EE and Smith RS (1992) *Overcoming the odds: High risk children from birth to adulthood*. Ithaca: Cornell University Press.

Wickham H (2014) Tidy data. *The Journal of Statistical Software* 59(10). Available from: http://www.jstatsoft.org/v59/i10/.

Wickham H (2015) *Advanced R*. Boca Raton, FL: CRC Press; Taylor & Francis Group.

Wickham H and Chang W (2018) *Ggplot2: Create elegant data visualisations using the grammar of graphics*.

Wickham H and Henry L (2017) *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. Available from: https://CRAN.R-project.org/package=tidyr.

Wickham H, Francois R, Henry L, et al. (2017) *Dplyr: A grammar of data manipulation*. Available from: https://CRAN.R-project.org/package=dplyr.

Wickham H, Hester J and Francois R (2017) *Readr: Read rectangular text data*. Available from: https://CRAN.R-project.org/package=readr.

Wilkinson R (2006) Foreward. In: Bartley M (ed.), *Capability and resilience: Beating the odds*, UCL Department of Epidemiology; Public Health. Available from: http://www.ucl.ac.uk/capabilityandresilience/beatingtheoddsbook.pdf.

Williamson P, Birkin M and Rees PH (1998) The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning A* 30: 785–816.

World Health Organization (2015) Ageing and health. Available from:

http://www.who.int/mediacentre/factsheets/fs404/en/ (accessed July 2017).

World Health Organization (2016) Ambient air pollution: A global assessment of exposure and burden of disease. Available from: http://www.who.int/phe/publications/air-pollution-global-assessment/en/ (accessed July 2017).

Yu J (2003) The association between parental alcohol–related behaviors and children's drinking. *Drug and Alcohol Dependence* 69: 253–262.