Essays on the Economics of Health and Place

Jack Higgins

BSc Economics, MSc Economics (Economics of Health)

A thesis submitted to Lancaster University in fulfilment of the requirements for the degree of Doctor of Philosophy

January 2020

Lancaster University

Declaration of Authorship

I, Jack Higgins, hereby declare that the work contained in this thesis and the work presented in it is entirely my own and I have clearly documented all sources and materials used. I received supervisory advice and comments on each chapter, from Prof. Bruce Hollingsworth and Prof. Ian Walker, throughout the duration of writing this thesis.

This work has not been submitted in any form for the award of a degree at this university or any other institution, nor has this work been published.

Acknowledgments

This thesis considers the effects of moving home, living on the coast, and moving away to university. Throughout my time writing it, I experienced each of these first-hand. With $N = 1$, I can say that each of these can have a detrimental impact on one's mental and physical health; I couldn't have done this without the continuous support from family, friends and colleagues.

I thank my friends from my time at Lancaster: Rachael, Jamie, Cat, Luna, Annie, Maddie, Claire and Luke. I couldn't have done this without your patience and support. I have had the pleasure of working with many amazing academics over the past few years. In particular, I thank Katharina Janke for her support, understanding and for making me a better researcher. I also thank Colin Green and Céu Mateus for their time over the years. I of course thank my supervisors Bruce and Ian: thanks for supporting my research for so long... I also thank my friends and colleagues from Manchester, in particular: Luke, Phil, Al, Shing, Cheryl and Matt for your good humour and knowing support.

I thank my parents for your continuous love and support; particularly mum, without whom I would not have been able to handle the constant highs and lows of writing a PhD thesis. Thanks also to my sisters Louise and Pamela for always being there when I needed you!

A final special thanks goes to Helen. The usual spiel applies about love and support, but you went above and beyond what I could have asked for - especially getting me through the final stages of writing - I cannot thank you enough.

Notes

Various versions of Chapter 2 were presented at the 15th Portuguese National Conference on Health Economics in Coimbra (October, 2017); the North West Doctoral Training Centre (NWDTC) conference in Liverpool (May, 2018); at the Health Economists' Study Group (HESG) meeting in Bristol (June, 2018); and the European Health Economics Association (EuHEA) conference in Maastricht (July, 2018).

Chapter 3 was presented at the EuHEA Student-Supervisor conference in Barcelona (September, 2016); and the NWDTC conference in Manchester (May, 2017).

Chapter 4 was presented at the EuHEA Student-Supervisor conference in Porto (September, 2019).

Abstract

Title: Essays on the Economics of Health and Place
Author: Jack Higgins
Degree: PhD in Health Research (Doctor of Philosophy)
Submission Date: January 2020

This thesis considers the connection between an individual's health and place - that is, variation in where one lives and the process of moving itself. It comprises three self-contained empirical chapters that are unified under this broad theme. It begins by considering the impact of moving house more generally, before moving on to look at more specific topics: living on the UK coast and living at University.

The first empirical chapter asks the question: What is the effect of moving residence on health? It considers the empirical challenges in identifying causal effects for such phenomena. The main methodology for doing so, involves using spatial and inter-temporal variation in local school quality and house prices, as Instrumental Variables for changing address. These data (available publicly) are mapped to households that were interviewed as part of the British Household Panel Survey and Understanding Society providing a rich dataset containing information at the individual, household and local area level. This paper goes on to address the question of how a move affects short-term health using a regression discontinuity-type design, considering the differences in health between those who were interviewed just before and just after they changed address. It finds, in general, that local school quality is a strong instrument for moving residence, and doing so leads to worse self-assessed health.

The second empirical chapter considers a more specific question about the health of those who live on the UK coast. UK data shows that health amongst the working-age population (16-64 years) is worse on the coast than elsewhere. For example, there is a much greater prevalence of limiting long-term health conditions on the coast as opposed to the average for England and Wales. Despite this, there is a lack of literature that considers the potential reasons for these differences and how they can be identified; this paper addresses this gap. Using data on health and other characteristics from all five waves of Understanding Society, this chapter quantifies the differences in health and health-related outcomes on the coast compared to inland. Detailed geographic data are used to construct a distance to the coast measure, which is used as the main distinc-

tion between a coastal and non-coastal area. The analysis finds that most health-related outcomes are worse on the coast, including long-standing health conditions, disability benefit claimants and smoking and drinking prevalence being more likely.

The final empirical chapter considers the mobility of students attending university, and their life satisfaction in early adulthood. The move to university is often the first major independent change of residence that an individual faces. At the same time, a large proportion of students live at home while they study. Data on a cohort of university attendees from the Longitudinal Survey of Young People in England (LSYPE) and the follow-up study, Next Steps, is used to assess the impact of moving away from home on early-adult life satisfaction. A random sample of children, born in 1989/1990, were surveyed annually between the ages of 13 and 19 years old, and then again when aged 25 (Next Steps wave). Life satisfaction is modelled for graduates aged 25 years, using an ordered probit approach, controlling for individual characteristics at various points in the student's life, such as external locus of control and psychosocial health. I also partial out parental and household factors such as household income, parental education, parental occupation, and the number of siblings in the home. The analysis finds that life satisfaction of males who move away is much higher than those who do not; there is no effect for females. Instrumental variable ordered probit models address the endogeneity of moving away to university, and mediation analysis assess some potential mechanisms behind these differences.

Where an individual lives, and the process of moving, is a determinant of health. More research is needed to disentangle this complex and heterogeneous relationship, with a view to identify policies with which to facilitate internal migration, and improve health outcomes.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 General Background

There is a growing emphasis on health research in which *place*, or location matters, yet there remain many under-researched areas which have the potential to inform and improve health policy. For this thesis, the topic of health and place includes both location and relocation, or migration. How these phenomena affect the health of locals or those who move is the subject of this thesis.

The UK health budget is allocated to Health Authorities based on local need. Implicitly, variation in local characteristics affect the demand for health care in different areas. A greater understanding of how place-related features affect the health of local populations is vital for the adequate provision of resources.

Specifically, it could be that particular characteristics need to be accounted for in the resource allocation formula, or that some areas - such as the coast in the UK - have a greater need for health resources that is only revealed from a place-based, quantitative, approach. It may be possible to identify different patterns of need that arise through empirical consideration of spatial features this way.

Looking in detail at differences within groups, such as those who move and those who do not - whether through choice or not - may help identify vulnerable groups of society who are in a position potentially detrimental to their health and wellbeing. A greater understanding of these groups of people can help influence housing policy in a way that is both more efficient, and is closer to maximising social welfare.

More generally, health systems generate social value within local communities. It is important to further knowledge of how health is affected locally to maximise this value. The Health Foundation report highlights a need for the NHS to support “inclusive economies”, saying that there is a growing synergy between the “place-based” lens of the NHS and broader policy that emphasizes localism. It is crucial to support this growing synergy with robust empirical evidence of all aspects of place-based health.

Regional disparities in health have been highlighted between the North and South of England (see, for example, Bambra et al. (2018)). There could exist other such important, divisive, spatial domains. The coast in the UK is one such domain, and the left-behind economies that reside there may have a systematically different need for health and healthcare. Identifying such need may be far-reaching in terms of policies that are aimed at boosting these local areas, which may become even further left behind in the wake of Brexit.

The UK has seen a greater embrace of the devolution of power over the last decade. With more governing power in the hands of local authorities, it is critical to support local policy with robust quantitative evidence. How place-based factors affect health, and how migration means the ever-changing local populations place differing demands on the healthcare system are two core health policy issues that will benefit from further research in these areas.

There is also an important interplay between location and education. University attendance is a major factor in the migration of young populations within the UK, for example, and there is little known about the effects of doing so. School choice, due to catchment areas, is also inherently related to where a household resides, and plays a major role in determining when and where relocation may take place. As such, education plays a role in two out of the three empirical chapters, where school quality and moving to university play a role.

This thesis consists of three self-contained empirical chapters that are linked by the micro-econometric techniques they use, and that location and health are the central themes of interest. The first empirical chapter considers the effect of moving home on health, the second investigates differences in health and health-related outcome on the UK coast versus inland, and the final empirical chapter assesses how living away from home during university can affect early-adult outcomes.

The remainder of this introductory chapter provides some context and background for the empirical chapters and how they contribute to the relevant literatures relating to health and place. It begins with some definitions of what is meant by "place" in this thesis, with a brief overview of some related literature and descriptive statistics for the UK. Health, and how it is measured, is then outlined. The motivation and aims of the overall thesis are then explicitly stated, before providing an overview of the three empirical chapters to follow.

1.1.1 Place: definitions and its effect on health

The common theme of each empirical chapter is considering, generally, the effects of place on health. This section defines what is meant by place and, more generally, place in the context of this thesis. Specifically, each empirical chapter focuses on a variable of interest that either captures the effect of moving between places (migration), or the effects of a spatial characteristic - namely areas on the coast.

Migration

Migration - any change in residence between two places - can be divided into two broad themes: international migration and internal (or domestic) migration. Both have been in the remit of economists, in terms of both the determinants and consequences of migration, for some time. There has been a greater focus on international migration however, and there remains a dearth in the literature on internal migration - particularly in the UK. It is this that takes the focus of the thesis. The remainder of this section will give a brief overview of the international migration and health literature, before moving on to the internal migration and health evidence.

The economics literature has tended to focus on the employment and fiscal consequences of international migration on both receiving and donor countries (Borjas, 2015). Labour economists in particular have focused on the market implications of those who migrate, with the movement of people from countries of lower to higher labour productivity the main economic driver (Sachs, 2016). There is mixed evidence on wage effects (Dustmann et al., 2016). Some studies find, for native workers, positive wage effects (Ottaviano & Peri, 2012); some find minor effects (Card, 2009); and others negative effects (Borjas, 2003).

A subset of the international migration literature considers instead the health returns of migration. Most health economics and public health studies find a "healthy migrant effect" (Abraido-Lanza et al., 1999). A failure to control for self-selection by most early studies has led to a hypothesis that migrants are positively selected on health (Palloni & Arias, 2004; Farré, 2016). Some studies have adopted a pseudo-experimental estimation framework in an attempt to address this. Gibson et al. (2013) (2013), for example, exploit a migration lottery to estimate the impact of migration on measured blood pressure and hypertension. They find that migration increases both measures, and these results persist after adjusting for selective non-compliance within the natural experiment.

Internal migration is at the center of chapters 2 and 4. One of the earliest economic theories of internal migration was from Mincer's work on the household migration decision (Mincer, 1978). He recognised that positive net benefits at the *family* level drive relocation. If individuals moved purely based on their own costs and benefits of doing so, we would expect positive wellbeing returns to moving, and also positive health returns. With moving posited as a household-level decision, however, this leaves the option of gainers and losers within a family. Hence there can be positive health returns for some and negative for others.

The most compelling evidence of the impact of internal migration comes from literature which exploits the Moving to Opportunity (MTO) randomised experiment. The MTO programme enlisted households residing in public housing, based in poor areas from five US cities: Baltimore, Boston, Chicago, Los Angeles and New York (Chetty et al., 2016). The households were randomly assigned, via a lottery, to two treatment groups that received housing vouchers for low-poverty areas and regular housing vouchers; the control group received no assistance from MTO. The literature finds positive labour market and education effects for children who moved from high poverty to low poverty areas (Chetty et al., 2016; Sanbonmatsu et al., 2006) and large positive effects on physical health, mental health and subjective wellbeing for both adults and children (Katz et al., 2001; Clampet-Lundquist & Massey, 2008; Ludwig et al., 2013; Kling et al., 2007).

Much of the literature that considers migration and health does so using data from large countries, such as the US or China, and focuses on rural-urban migration. Johnson and Taylor (2018) for example, examine rural to urban migration in the United States throughout the early 20th century and its effect on long-term health and longevity. Using the location of railway lines as an instrument they find that, despite an increase in lifetime wealth, migrants are worse-off in later-life health. Chen (2011) explores rural to urban migration in China using a small household survey. They find evidence of a "healthy migrant phenomenon" on self-rated physical health. Using panel data from the Indonesian Family Life Survey, Lu (2008) finds that movers select based on past health status, but the magnitude of this depends on the type of move.

All of the empirical chapters in this thesis utilise data from the UK. There are several studies which have considered the effect of domestic mobility on health and/or wellbeing in the UK. Most studies use panel data, and are thus able to control for unobserved (time-fixed) individual heterogeneity in health outcomes. Moh'd and Ajefu (2017) find that individuals who move report higher health outcomes, but not mental health indicators. This is attributed to positive health selection, as apposed to any causal effect of moving on health. Other studies find negative health effects of moving (Tunstall et al., 2014; Morris et al., 2017), with some finding both negative and ambiguous effects in the short-run (Nowok et al., 2013; Whittaker, 2012).

In summary, the literature on migration and health is mixed, and there is a clear need for further research in this area. This is especially the case for research that considers internal migration, as the majority of papers tend to focus on international mobility (Moh'd & Ajefu, 2017). Alongside this general gap in the literature on domestic mobility and health, there is little evidence on how moving at different ages can influence health and wellbeing. Younger migrants tend to be healthier, and early-adulthood represents the peak age for migration (Norman et al., 2005). In migration studies more generally there is a lack of attention paid to the age-profiling of movers (Norman & Boyle, 2014), and this represents another gap in the literature that would significantly gain from more work in the area.

Local characteristics and the UK coast

Much of the literature that considers place in a health context focuses on a particular local characteristic, and how these influence the health of these populations. Some relevant examples include prevalence of green and blue space, commuting behaviour, neighbourhood effects and local community assets. This section continues by providing a contextual overview of some of these related literatures, before identifying a gap within them about the UK coast.

There are structural features, at the local level, that can influence health and quality of life. Community assets are one such way in which local features can influence the lives of local residents. Munford, Sidaway, et al. (2017) found, for example, that participation in community assets is associated with substantially higher health-related quality of life, and that there is a potentially substantial social value generated by developing these assets.

There is evidence to suggest that variation in local geographical factors can influence health. There is a large body of evidence on the influence of greenspace and health. De Vries et al. (2003) find, using Dutch data on the self-reported health of 10,000 people, that living in a green environment was associated with positive health outcomes. This effect is amplified for the elderly and those of a lower educational background. A systematic review of green spaces and mortality finds an inverse relationship between surrounding greenness and all-cause mortality (Rojas-Rueda et al., 2019). In a similar fashion to greenspace in the literature, "bluespace" is defined: "health-enabling places and spaces, where water is at the centre of a range of environments with identifiable potential for the promotion of human wellbeing" (Foley & Kistemann, 2015). Bluespace also matters: there is limited evidence to suggest that bluespace can affect mental health, but it is recognised that there is a strong need for further research in this area (Gascon et al., 2015). There is also the potential for bluespace to be framed as a health-enabling resource (De Vries et al., 2003; Foley & Kistemann, 2015), and used as a platform for physical exercise (Pasanen et al., 2019).

Another, related, strand of literature is that which considers health on the UK coast.

There is limited evidence which suggests that there are health differences in populations that reside on the coast. There has also been media interest highlighting potential problems with coastal health and the local economic climate of these areas more generally. The existing literature, which is outlined in more detail in chapter 3, does not consider in any detail the extent to which these differences in health are attributable to selection on unobservable characteristics that may affect living on the coast and an individual's health.

There has been a recent recognition that populations on the coast are generally worse off than those inland. Corfe (2017) analysed recent economic and social data at the local authority level, and highlights several problematic features. Five of the ten local authorities in Great Britain with the lowest average employee pay are in coastal communities, with average annual gross pay on the coast around £3,600 lower in coastal communities. These inequalities are also worsening: in 1997 economic output per capita was 23% lower in coastal communities, and by 2015 this gap had increased to 26%. Worse outcomes on the coast are not limited to economic indicators: two of the 20 local authorities with the highest proportion of poor health are coastal communities[1]. Despite these stark figures, there is relatively little empirical evidence that investigates health on the coast.

Chapter 3 makes a contribution to the small existing literature that considers health on the UK coast by bringing a robust econometric approach that specifically considers to what extend selection on unobservables may play a role in explaining the health differences. It also does this using a large panel data set which is new to the coastal literature.

1.2 Motivation and Aims

It is clear from this general background to health and place as a research topic that this is an important, policy relevant, research area with some underdeveloped areas of the literature. Specifically: how local characteristics and internal migration can affect health are two such literatures which can benefit from further research in these areas; this the-

[1]Specifically: Neath Port Talbot, Blackpool, Bridgend, Sunderland, Barrow-in-Furness, Carmarthenshire, East Lindsey, South Tyneside, County Durham and Hartlepool (Corfe, 2017).

sis aims do do so. The remainder of this section outlines the specific research questions and contributions that this thesis makes.

This thesis aims to extend the literature in three distinct, but related, areas. Firstly it considers how internal migration can affect health. This is a relevant question for policy-makers as there is relatively little known about the consequences of internal migration in the UK. Secondly, it investigates specifically the differences in health and health-related behaviour, and the potential mechanisms behind them. Again, given the pressures on the NHS in the UK, identifying areas of unmet need is of crucial importance to decision makers in order to improve the effectiveness and efficiency of health resource allocation. Finally, it look s at how internal migration can affect young adults - namely university students in the UK. It considers how moving away to university can differentially impact early-adult life satisfaction versus those who remain at home during study. Ultimately, this thesis provides new evidence on the role of place in an individual's health, at various stages in life, and helps to identify potential populations of need.

1.3 Thesis structure and Overview

In addition to this introductory chapter, and a final conclusive chapter, this thesis consists of three substantive empirical chapters. As this thesis is written in an alternative format, these empirical chapters are presented in the style of stand-alone journal publications. Each has its own literature review, introduction and conclusion. The intersection of each chapter, and the overall theme of the thesis however, is a contribution to the set of knowledge about how health and place are interrelated. The rest of this section outlines each of these empirical chapters, including information about the data and identification strategies used, and a brief overview of the results.

The first empirical chapter asks the question: What is the effect of moving residence on health? I consider at length the empirical challenges in identifying causal effects for such phenomena. The main methodology for doing so, involves using spatial and intertemporal variation in local school quality and house prices, as Instrumental Variables for changing address. These data (available publicly) are mapped to households that were interviewed as part of the British Household Panel Survey and Understanding Society providing a rich dataset containing information at the individual, household and local area level. In order to address the endogeneity of moving home, an Instrumental Variables identification strategy was used, with local school quality and the age of the youngest child comprising the main instrument set. The rationale behind this instrument is that the schooling decisions play a large role in a household's decision to move, and the potential health effects are either ignored or unknown to the household that moves. The main findings of the chapter are that moving house has a negative effect on an individual's self-assessed health outcomes, once instrumenting for moving home. However, the imprecision of these estimates once using an instrument mean that they must be treated with caution. The analysis went on to consider short-run effects in an RDD-type set up, which compared movers who were interviewed in the 12 months before, and 12 months after they moved home. Doing so revealed a negative anticipatory effect of moving in the few months prior to doing so followed by an offsetting positive effect in the three months post move.

The second empirical chapter considers a more specific question about the health of those who live on the UK coast. UK data shows that health amongst the working-age population (16-64 years) is worse on the coast than elsewhere. For example, there is a much greater prevalence of limiting long-term health conditions on the coast as opposed to the average for England and Wales. Despite this, there is a lack of literature that considers the potential reasons for these differences and how they can be identified; this paper addresses this gap. Using data on health and other characteristics from all five waves of Understanding Society, I quantify the differences in health and health-related outcomes on the coast compared to inland. Detailed geographic data are used to construct a distance to the coast measure, which I use as the main distinction between a coastal and non-coastal area. The analysis reveals that there are potential health-promoting features of the coast in terms of physical exercise - with those living on the coast participating in more frequent physical activity. However, this positive finding is offset by a higher prevalence of smoking, disability claimants and those with long-term health conditions. This suggests that theres a greater health need on the coast, compared with otherwise similar areas, and the resource allocation formula should take these differences into account.

The final empirical chapter considers the mobility of students attending university, and their health and labour outcomes. The move to university is often the first major independent change of residence that an individual faces. At the same time, a large proportion of students live at home while they study. I consider the differences in post-graduation outcomes for those who move away from, versus those who remain at, home. The data allow for the partialling out of pre-, during-, and post-university individual characteristics, as well as parental and household controls. Heterogeneous effects were found, with males reporting much higher life satisfaction if they lived away from home, whereas females who moved away showed no significant difference in early adult life satisfaction. It also considers the interactions between health and attending a Russell group university as a proxy for university quality, and Muslim students who are likely to face different financial constraints surrounding student finance and Islamic beliefs. The analysis shows that Male Muslim students are better off if the remain at home, whilst Males who move away are better off if they attend a Russell group University. The endogeneity of moving away is addressed by using the individual's grandparents' university attendance as an instrument in an IV ordered probit model, and mediation analysis was performed to assess the potential role of income and an individual's external locus of control in explaining the difference in effect for males and females. Neither income nor loci of control act as an indirect effect of moving on life satisfaction, and so the mechanism behind moving must either be direct or due to some other unobserved factor to the analysis.

CHAPTER 2

Home is where your health is: the impact of internal migration on health and wellbeing

Abstract

There exists a large body of evidence around the issue of internal migration, with most considering the labour market implications of relocation. Less attention has been paid, however, to the health implications of migration; specifically, how it can affect health and wellbeing. This paper contributes to that literature by considering the causal effect of internal migration and health, using rich data from the UK. Data is used from all waves of the British Household Panel Survey (BHPS), from 1991-2008, and five waves of Understanding Society (USoc), from 2010-2015, with access to local area-level data in each. Individuals are followed over this period, including if, when and where households move. There are around 6,700 individuals who are in both data, and our full sample size (N*T, for those in BHPS, USoc, or both), conditional on full information, is 107,736. For identification, age of the youngest child in the household interacted with local school quality, and house prices, are used as instrumental variables for internal migration. Measures of health and wellbeing, including the General Health Questionnaire (GHQ), self-assessed health, and whether or not the individual has a long-term health condition, are used as dependent variables. Once the endogeneity of migration is accounted for, the analysis shows that there are negative health consequences of moving. In particular, individuals who move are more likely to have a health condition, less likely to report very good or excellent self-assessed health and worse GHQ-12 scores. These results are insignificant when accounting for endogeneity with various instruments. This paper also finds that there are short-term negative mental health consequences of moving, but these return to a baseline level around 3 months after moving. These findings suggest that there may be unintended negative consequences of policies that, for example, remove barriers to internal migration. Future work should seek to find alternative sources of exogenous variation and to unpack the mechanisms behind this complex household decision.

2.1 Introduction

This chapter considers the health effects of internal migration, using data from two linked household panel surveys: the British Household Panel Survey (BHPS) and Understanding Society (USoc). Internal migration is defined, for the purposes of this analysis, as moving from one permanent address to another. This is opposed to movements between labour markets (typically government office regions in the UK), as is found in the labour economics literature, where the focus tends to be on wage returns. A threat to identification comes from the fact that individuals self-select into moving home. This study contributes to the literature by using local school quality and the age of the youngest child in a household as Instrumental Variables. The analysis also attempts to unpack the likely heterogeneous and timing effects (including anticipation effects) associated with internal migration.

Much of the literature on migration focuses on wages and wealth returns, and many of the papers that do consider health, do so from an international perspective. Gibson et al. (2013), for example, exploit a migration lottery to estimate the impact of migration on measured blood pressure and hypertension, and find that both are increased persistently after migration. As such, there exists a small subset of the migration literature that considers its effect on health. Most recently, Johnson and Taylor (2018) consider rural to urban migration in the US, in the early 20th century, and its effect on long-term health and longevity. They find, using the location of railway lines as an instrument for migration, that although lifetime wealth was increased, migrants paid a price in terms of later-life health. The authors suggest that the mechanism is through adopted risky health behaviour such as smoking and excess alcohol use. In a paper considering the health-based selection into migration in Indonesia, Lu (2008) finds that individuals do select based on past health status, but the magnitude of this depends on the type of move.

There is a body of literature that explores the effects of the Moving To Opportunity (MTO) experiment to estimate the causal effects of moving out of high-poverty neighbourhoods in the US. The MTO was a randomised housing mobility experiment that utilised a lottery system to distribute housing vouchers amongst families living in social housing (Sanbonmatsu et al., 2006). Chetty et al. (2016) convincingly shows that there were large benefits of MTO for younger children - who were more likely to attend college and earned on average $3,477 more than the control group. There were heterogeneous treatment effects however, with little to no impact of MTO on adult outcomes, and some small and negative effects for older children (aged 13-18 years). There were also positive effects of the MTO program on education outcomes (Sanbonmatsu et al., 2006). Many studies also considered the health effects of the MTO programme, generally finding that moving to lower poverty areas has large positive effects on individuals' physical health, mental health and subjective wellbeing (Katz et al., 2001; Clampet-Lundquist & Massey, 2008; Ludwig et al., 2013). Kling et al. (2007) find similar results, but alo show that positive health effects for female youth were offset by negative effects for the male youth.

In the UK literature, the direction of the effect of migration is inconclusive. Tunstall et al. (2014) consider the mental health of movers in the UK, using the BHPS. They focus on movements from more to less deprived areas (and vice versa) and find that movers are more likely to have mental health problems and have higher rates of poor health. Nowok et al. (2013) consider the post-move happiness of movers in the BHPS, and consider the duration effects after the migration event. They find that before a move takes place, individuals suffer from lower happiness, but then they return to previous levels once a move has taken place. Whittaker (2012) also uses the BHPS to compare pre- and post-move wellbeing scores, addressing endogeneity with a dynamic random effects probit model, with lagged health terms to account for selection. Their results suggest that the effects of migration are both positive and negative, and do not differ by motive but do differ by past health status.

2.1.1 Theoretical Underpinnings and contributions

Many of the typical reasons for moving (Farwick, 2009) are associated with positive life, wealth, and socioeconomic status gains: increased income from new job, more space or nicer environment, increased disposable income from downsizing, or better schooling. This would suggest a positive effect of moving on health.

Households move when the net gains from doing so are positive (Mincer, 1978), so a natural hypothesis for the effect of moving on health is that there are positive health returns. The net gains in consideration, however, are at the *household* level - with each household member having potentially differing weights. This means that there may be positive or negative net returns at the *individual* level, so the observed net effect on health - as explored in this chapter - is less clear cut.

Moving house is a significant life event, which can be extremely stressful to those experiencing it (Thoits, 2010; Lazarus, 1995; DeLongis et al., 1988). In the short-run, including just before a move takes place, it would be unsurprising to find negative effects for those who move, that could mask any short-run gains from doing so. Easterlin (2005) In contrast to the Easterlin paradox - whereby improved material circumstances do not bring about improved wellbeing as individuals adapt to their new level of living - life events in family and health domains can have lasting effects on health and wellbeing (Easterlin, 2005; Nowok et al., 2013). In the longer term then, we would expect a return to baseline levels of health and wellbeing, or reaching a higher level than before the move.

Many of these mechanisms are difficult to unpick, due to the endogeneity of moving with respect to health and wellbeing. This chapter makes a contribution to this literature by firstly using an IV approach in an attempt to overcome the endogeneity of migration. Secondly this paper provides a link between school choice, migration and health through the IV approach that it takes. Using geographical data on local school quality and distance to local schools to estimate the probability of attending the closest schools, provides a novel instrument to the literature that considers moving residence. Thirdly, the short-run effects of migration on health are investigated, considering the health and wellbeing of those just before and just after moving residence.

2.1.2 Roadmap

Section 2.2 begins the motivation for the instruments for migration, and provides a link between school choice, house prices and moving home; section 2.3 describes the datasets used; section 2.4 outlines the empirical methodology; section 2.5 presents the results; and section 2.6 concludes.

2.2 School choice and migration

Due to the fact that health itself is a determinant of migration, and the likelihood of the existence of unobservable factors that influence both, a focus is placed on the role of school choice and house prices in the search for exogenous variation. The results are presented for these two approaches separately, and the motivation for using each is outlined below.

2.2.1 School choice and quality

With respect to moving home, if a household contains a child of a relevant schooling age, then the decision to migrate is negatively associated with the expected utility (or expected quality) of their current local school choice set. If the local schools are low performing, and there is a young child in the household, then the probability of living at a different address in the next period is higher, *ceteris paribus*. It can therefore be viewed as a source of exogenous variation in the migration decision. The age of the youngest child, interacted with $E[U_{ha}]$, is used to capture this mechanism.

Following Weldon (2017), consider the utility function of household h, for school s:

$$U_{hs} = f(D_{hs}, Q_s) + \varepsilon_{hs}, \tag{2.1}$$

where D_{hs} and Q_s are the distance to, and quality of, school s; ε_{hs} represents unobserved household heterogeneity of preference to school s.

At the local level, the expected utility of schools in the local area a, for household h can be expressed as:

$$E[U_{ha}] = \sum_{s=1}^{S_a} p_{hs} U_{hs}, \tag{2.2}$$

where p_{hs} represents the probability of the children in household h attending school s, and S_a is the school choice set (i.e. the number of schools available in the local area).

For the purposes of this chapter, the probability of attending a school is computed purely as a function of distance:

$$p_{hs} = \frac{exp(\beta^D ln(D_{hs}))}{\sum_{i \in S_h} exp(\beta^D ln(D_{hi}))}, \tag{2.3}$$

The log-odds parameter β^D, represents a households' preference to be based closer to a school and is unobserved. For this chapter, β^D is set to -2.436, informed by a discrete-choice model of household school-choice preferences undertaken by Weldon (2017). As U_{hs} is not observed, a performance measure for school s is used as a proxy, and set S_a to be the five closest schools to each household h. In practice, the measure of $E[U_{ha}]$ can be viewed as expected school quality of the local area.[1]

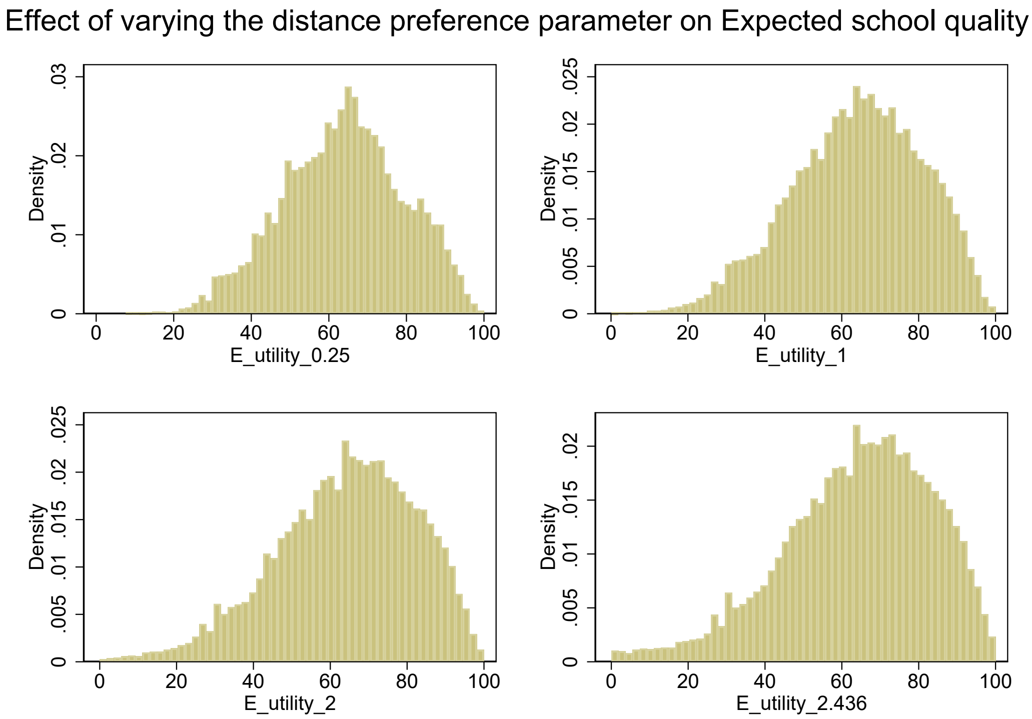
Choosing the parameter β^D represents an attempt to better represent a household's preferences using empirical evidence, versus simply choosing some arbitrary value, or ignoring it altogether. Choosing to base this value on evidence in Weldon (2017) may not be representative of the population at hand, as this was based off data from Lancashire alone. If this parameter varies between region - this may mask regional variation in the expected utility calculated. In practice, however, the choice of this parameter amongst a wide set of reasonable values does not affect the distribution, nor level of, expected utility, as shown in Figure 2.1 below.

2.2.2 House prices

Using expected school quality and age of the youngest child in the household as an instrument is a plausible source of exogenous variation in moving house. A caveat of this approach is apparent when considering the nature of the Local Average Treatment Effect (LATE) that it picks up. The estimates are only relevant for the "compliers" to the instrument. Loosely speaking these are households who, conditional on having a child of a relevant age, would be more likely to move house if their local school choice set was of a low standard. This limits the external validity of the analysis to those who are both willing and able to move on the basis of school quality. This represents a non-trivial

[1] As a robustness check, the measure of expected school quality is replaced with a simple unweighted average from the closest five schools in the area. The results (i.e. the instrument's predictive power of migration) do not differ greatly.

Figure 2.1: Choice parameter sensitivity



Notes: Graphical comparison of the distribution of the resulting values of $E[U]$, as β^D is allowed to vary. Going clokckwise, the parameter takes the value of -0.25, -1, -2.436, and -2, respectively.

group of households, and the question of how relocating affects their health outcomes remains an important one. However, this approach may under-represent many households who are not influenced by local school quality, including households which have chosen not to have children, and those who are older.

In order to try and capture the health effects of moving for a more general group of households, data on local house price shocks - relative to general UK prices - are incorporated from the previous period as an alternative instrument. A positive association between house prices and the decision to sell/buy is relatively well-established in the literature both theoretically (Stein, 1995) and empirically (Krainer et al., 2008).

A price shock in a previous period is also likely to affect renters in the current period. Hence it seems likely that there is a higher probability of moving home following a price shock, for both home-owners and renters. The use of local house prices, as opposed to rent paid, adds to the validity of the exclusion restriction from the health outcome equation. Implicitly assumed in using school choice and house prices as instruments is that in the decision to move, households have incomplete information about how a move will affect their health, and hence ignore this in their decision to move. These assumptions form the basis of the plausibility of the exclusion restrictions for each instrument set, respectively.

In the case of the school quality and age of youngest child instrument - a strong case for the excludability of the instrument from a structural health equation can be made based upon the complier group for which this instrument is relevant for. Economic theory states that household moves take place when the *household* net benefits are positive (Mincer, 1978). Parents of young children who face a decision about their child's schooling are likely to do so from a more altruistic perspective - placing a higher weight on their child's gains from moving than their own. This discounting of their own gains (including health gains) provides an argument for the instruments' excludability. In the case of house prices, the argument for the exclusion restriction is based upon price shocks. Conditional on wealth, price spikes that induce a move are unlikely to affect long-term health directly. This is unlikely to hold in the short-run, however - particularly in the case for mental health. As a result, the price spike instrument used in this chapter is to be viewed as secondary to the schooling instrument, and this in part motivates the need for the analysis of short-run associations between moving on health. These assumptions remain fundamentally untestable, and the reliability of the results presented depend upon how likely it is deemed that they hold, conditional on the set of controls and estimation methods.

2.3 Data

2.3.1 Understanding Society and the BHPS

This chapter makes use of data from two longitudinal surveys in the UK: the British Household Panel Survey (BHPS), and Understanding Society (USoc). The BHPS ran from 1991 to 2008, covering around 10,000 households. USoc began in 2009 and is ongoing, in its 7^{th} wave, covering around 40,000 households. The members of each household are revisited annually, with data collection for each wave taking place over a two-year period. Individuals in the household aged over 16 years have a face-to-face interview in addition to a short questionnaire. In the final wave of the BHPS, individuals were asked if the wished to participate in USoc, meaning some individuals are traceable through both datasets.

For the purposes of this analysis all available waves are used, restricting the sample to those of working age (16-65 years) and those residing in England and Wales. Local area level data linkage is not available for Scotland and Northern Ireland. A linkage between both datasets and 2011 census geographical data is also used; specifically, the household's lower-layer super output area (LSOA). The full sample of the working age population (N*T) consists of 342,488 observations. This is further restricted to households who are observed to have at least one child in at least one wave, and which are observed in at least two waves. School quality data is available from 2001 onwards. Once conditioning on having a complete set of outcome and control variables, this leaves N*T = 107,736, with N=31,216 and T ranging from 2 to 13 observed periods (from 2001 to 2014).

Outcome Variables

Several measures of health are used as the main outcome variables. To capture self-assessed health, individuals were asked "How would you rate your health in general?" to which they responded poor, fair, good, very good or excellent in USoc and very poor, poor, good, very good or excellent in BHPS. To work around this inconsistency, and to simplify estimation, this variable is set to equal to one if the response was "excellent", "very good" or "good", and equal to zero if equal to "fair", "poor" or "very poor". The variable is used, therefore, as a measure of self-assessed "good" health. Although a simple and subjective measure of health, it has been shown to have strong predictive power of future mortality (Idler & Benyamini, 1997).

The BHPS and USoc both contain data on the individuals' GHQ-12 score, which ranges from 0 to 36, and aggregates answers from 12 questions aimed at assessing their psychological well-being. Such questions include: "Have you recently lost much sleep over worry?"; "Have you recently felt capable of making decisions about things?"; "Have you recently been feeling unhappy or depressed?". The respondents answer each of these questions on a likert scale: "not at all" (0), "no more than usual" (1), "rather more than usual" (2), "much more than usual" (3).

The final health measure captures whether or not the individual has a self-reported health problem at the time of interview. In both the BHPS and Understanding Society, respondents were shown a list of health conditions and asked if they currently had any, and to indicate which ones if so. Amongst the conditions were: "Coronary Heart disease"; "High Blood Pressure"; "Diabetes"; "Epilepsy" etc. This is included as a variable equal to one if a respondent identified any of the conditions they were shown and zero otherwise. There was a slight difference in the categories that the respondents were shown between the BHPS and USoc. However as variation across these categories is not exploited in the analysis, an individuals' response as measured by the binary variable is highly likely to be the same regardless of the cards shown.

Migration and area-level variables

Much of the previous literature that considers internal migration as a variable of interest, has generally relied on a large move - typically from one labour market to another. For instance in a UK setting, Rabe and Taylor (2012) use the BHPS to consider moves from one local authority region to another as migration. In this chapter, however, as the focus is on moves of any type - whether related to the labour market or not - internal migration is defined as a household's change of address from time $t-1$ to t. Moving home - regardless of the distance moved - is a significant life-event and as such small moves should not be ignored. For this reason, any move picked up in the data is used, which is a departure from most studies that look at internal migration. As mentioned earlier, this necessitates an individual to be observed in at least two *consecutive* waves; those who aren't are dropped from the analysis.

In order to allow the inclusion of area-level covariates, including the Index of Multiple Deprivation (IMD) and its sub-indexes, and local schooling quality (see below), special license data on each household's Lower-layer Super Output Area (LSOA) is incorpo-

Table 2.1: Descriptive Statistics - differences by move status

	Move = 0 [N = 91,899]	**Move = 1** [N = 9,495]	**Difference**	**P-Value**
Outcomes:				
Self-Assessed Health	0.80	0.77	-0.03	0.00
GHQ Score	11.31	11.47	0.16	0.01
Long-term condition	0.03	0.03	-0.01	0.01
Covariates:				
Household net Income	2950.57	2654.73	-295.83	0.00
Male	0.41	0.39	-0.02	0.00
Age	36.91	32.69	-4.21	0.00
Employed	0.72	0.63	-0.09	0.00
Married	0.67	0.52	-0.14	0.00
GHQ Score	11.24	11.32	0.07	0.07
Highest Qualification:				
No Qualifications	0.10	0.09	-0.01	0.04
Other qualification	0.04	0.04	-0.00	0.04
GCSE or equiv.	0.27	0.28	0.02	0.00
A-level or equiv	0.18	0.18	0.00	0.40
Other Higher Qual.	0.22	0.21	-0.01	0.05
Degree	0.20	0.20	-0.00	0.74
Housing Tenure:				
Owned/mortgage	0.47	0.34	-0.12	0.00
Shared Ownership	0.27	0.13	-0.13	0.00
Rent Private	0.08	0.26	0.19	0.00
Rent Public	0.19	0.26	0.07	0.00
Instruments:				
Age of youngest child (years)	6.72	4.40	-2.32	0.00
E[School Quality]$_{t-1}$ (units = %A-C)	63.38	60.50	-3.88	0.00
House price spike$_{t-1}$ (LSOA)	0.027	0.065	0.038	0.00

rated to each dataset.[2]

Instrumental Variables

The main specification uses the age of the youngest child in the household (in categories), and local school quality (continuous), as an instrumental variable for migration. The motivation behind the instrument is that parents of young children of a schooling age, coupled with the quality of local schools, that make them more or less likely to move. As such, these two variables are included in the instrument set, as well as the interaction between them Age, in years, of the youngest child in the household is available in each wave of both the BHPS and USoc, and takes a value for each schooling category: 0-4 years; 5-10 years; and 11-15 years. In this way, the ages enter as dummy variables, with 0-4 years omitted as the base category. This variable is lagged, as it is expected that the age of the youngest child in the household at time $t-1$, will affect whether there is a change of address between time $t-1$ and t.

Data for school quality was obtained from the Department for Education[3], who provide a range of performance measures obtained from exam boards and the school census, for all schools in the UK. For the purposes of this chapter, the percentage attainment of 5 A*-C grades at GCSE-level is taken as a proxy measure of school quality. To incorporate distance to these schools at the LSOA level, each school is mapped (based on postcode) to the population-weighted centroid of each LSOA, and the Euclidean distance between each is subsequently calculated[4]. The approach outlined in section 2.2 is applied, taking the performance of the 5 closest schools to each LSOA, and used as the measure of school quality available to the household. The lagged age of youngest child variable is then interacted with this distance-weighted school quality measure, to construct the instrumental variable for moving home. Different lags and continuous age are also included as robustness checks.

House price spikes are also used as an alternative source of exogenous variation that

[2] An LSOA comprises of between 1,000-3,000 individuals and there are approximately 32,000 LSOAs in England.

[3] `https://www.compare-school-performance.service.gov.uk/`

[4] This distance is calculated using the easting and northing coordinates, which is available for both LSOAs and school postcodes.

influences moving home. Price paid data was accessed from the UK Land Registry[5] at the postcode level, from 1995 onwards. This was subsequently aggregated to both the LSOA and MSOA level, allowing the use of variation in local house prices. A local house price spike is defined as: equal to one if the percentage change in (LSOA-level) price between $t-2$ and $t-1$ is greater than two standard deviations away from the expected one-period price change, and zero otherwise.

Independent Variables

In terms of the included covariates of health and migration, the following are included: age and age^2, in years, of the respondents; their gender, which takes a value of one if the individual is male and zero otherwise; and whether they are married or in a civil partnership, equal to one if so and zero otherwise.

Education is also included. Respondents were asked about their highest educational qualification to date choosing one from “No qualifications; Degree; Other higher qualification; A-Level or equivalent; GCSE or equivalent; Other”, from which dummy variables for each category were created. Information on income was collected from all adult respondents and was used to construct a net income variable, which was trimmed of it’s 1^{st} and 99^{th} percentiles, and subsequently log-transformed[6]. Employment status is included as a dummy variable, equal to one if the individual is in paid employment and zero otherwise. Housing tenure comprised of a categorical variable for each of four types of residency: Ownership/Mortgage; Shared Ownership; Private Rental; and Public Rental.

[5] `http://landregistry.data.gov.uk/`

[6] The variable is trimmed, as $ln(0)$ is undefined. In earlier specifications the cube-root of income was used instead (as $0^{1/3}=0$), to check for differences between the two. There were no differences, suggesting that zero income is not a problem here.

2.4 Methods

2.4.1 Instrumental Variables methods

Consider the following linear regression of health on migration:

$$Y_{it} = \alpha_i + X_{it}'\beta + M_{it}'\tau + \varepsilon_{it}, \tag{2.4}$$

where Y is a measure of health, X a vector of determinants of health, and M a dummy variable equal to one if the individual has migrated at time t and zero otherwise. We are interested in τ, the marginal effect of migration on health, controlling for other health determinants. Using this approach ensures an unbiased estimate of τ, under the Gauss-Markov assumptions including, crucially, that migration is exogenous with respect to health ($cov(M, \varepsilon) = 0$). Including individual fixed effects, α_i, in the specification allows for time-fixed unobservables to be correlated with $\varepsilon_i t$.

This assumption is likely to be violated. It is plausible to suggest there exists some time-varying, common cause of health and migration that we cannot observe in the data, nor find a suitable proxy variable in its place. This model also suffers from reverse causality: migration affects health, but health also affects the decision to migrate. Both of these cases seem reasonable, as is often the case when the mechanism of interest takes the form of a "choice" variable. In other words, individuals self-select into the migration decision, so estimation via OLS yields a biased estimate of τ. In the absence of a natural experiment, an Instrumental Variable (IV) approach is used.

The following represent the first and second stage equations of the IV approach:

$$Y_{it} = \alpha_i + X_{it}'\beta^Y + \widehat{M}_{it}\tau + \varepsilon_{it1} \tag{2.5}$$

$$M_{it} = \delta_i + X_{it}'\beta^M + Z_{it}'\gamma + \varepsilon_{it2}, \tag{2.6}$$

where Y,X and M are defined as above. The new term, Z, which appears in the migration equation but is excluded from the health production function, represents the vector of instruments for migration. Note that α_i represent the individual-fixed effects, and thus estimation relies on the two-stage least-squares "within" estimator, which sweeps away the observable and unobserved, time-fixed, individual heterogeneity.

Identification relies on the assumption that the instrument(s) Z_{it} do not directly influence the outcome Y_{it}, other than through M_{it} (i.e. they are excludable from the health equation). Monotonicity of the instrument is also assumed (for a LATE interpretation), as is the relevance of the IVs with respect to migration. The latter of which is directly testable from the first stage, in the form of an F-test of joint significance of the instruments. Standard errors are clustered at the household level.

2.4.2 Alternative function form - recursive bivariate probit

Assuming that both health and moving, measured as binary variables, follow a bivariate normal distribution yields the recursive bivariate probit model. This approach explicitly allows for the two error terms to be correlated ($\rho = corr(\varepsilon_1, \varepsilon_2) \neq 0$), which is useful given the concerns with omitted variables bias in this model. Identification, in this particular system, still relies on the exclusion restriction placed on the health equation, however. This model is estimated identically to a standard bivariate probit model. We do not need to pay special attention to its recursive nature upon estimation, as in obtaining the density of Y, we already condition on M (Wooldridge, 2010; W. H. Greene, 2012).

Owing to the fact that both the observed outcomes and treatment variables are close to either zero or one in the data, this may be a preferable approach (Chiburis et al., 2012). However doing so does not exploit the panel nature of the data, and places a much stronger parametric assumption on the two error terms than linear 2SLS. The cost of estimating both stages linearly is that we can only recover the LATE, and not the ATE. Owing to the fact that these two estimates are likely to be close in nature anyway (local information is all that is provided in the data (Angrist & Pischke, 2008)), the trade-off of small gains in terms of the point estimates at the cost of imposing strong parametric assumptions and not exploiting the panel nature of the data, does not seem to be net positive. As a result, this model is left as a robustness check of the functional form in Appendix A, table A.1.

Average Partial (Treatment) Effects

We are interested in the average partial treatment effect of M on Y:

$$\frac{\partial E[Y|X,\varepsilon]}{\partial M}, \tag{2.7}$$

averaged over the population distribution of the unobserved heterogeneity, ε. In the discrete case, this is equivalent to calculating the following difference:

$$G(Y=1|X,M=1) - G(Y=1|X,M=0), \tag{2.8}$$

where $G(.)$ is the relevant cumulative distribution function, and averaging this over the population. This is the average treatment effect (ATE). In most cases, calculation of such effects and their standard errors are straightforward, using Stata's - `margins, dydx()` - command in a probit or logit approach. In the bivariate probit case however, this command does not produce the ATE[7]. A bootstrap procedure is used to calculate the standard errors for the ATE calculated as follows:

$$\Phi(X_i\hat{\beta^Y} + \hat{\tau}) - \Phi(X_i\hat{\beta^Y}), \tag{2.9}$$

averaged over the sample, where Φ is the cumulative standard normal distribution. 500 replications are used for the bootstrap procedure, and the Stata code for the above is presented in Appendix A.

2.4.3 Short-run effects

The above methods overlook the contemporaneous nature of the relationship between moving residence and health outcomes. Due to the panel nature of both data, and the way in which households were interviewed throughout the year, the date of moving can be exploited to address issues around the timing of a move. It is reasonable to suggest that many of the health effects arising as a result of a move occur in the short-run: around the time of the move itself. The GHQ-12 score of mental health is also a contemporaneous measure; it prefaces its questions with "over the past few weeks...". To

[7] -`margins`- in this case, focuses on joint probabilities - not required for calculation of the ATE, which uses marginal probabilities

address this the sample is restricted to those who are observed moving at least once during the sample period, and outcomes of those interviewed just before a move are compared to those just after.

This section uses a regression discontinuity-type set up. A typical regression discontinuity (RDD) approach involves analysing the effect of a treatment using observations close to the cut-off, assuming that such individuals are otherwise identical, and the running variable is effectively random. This section's approach compares individuals who were interviewed just before and just after they moved home. The interpretation of this model relies upon the assumptions made about how interview dates get determined, and whether or not a household has any input into this. If a household has no input, the dates are essentially random with respect to moving, and this model estimates the short-run causal effect of moving on health. On the other hand, as is more likely the case, some correspondence will take place prior to the interview, meaning the household will have some input as to the interview date. In this case, this approach does not offer a sharp identification of the effects of moving, as there are unobserved factors that could affect when an interview took place, relative to when the household moved. Therefore chapter takes this conservative, more probable, stance and will refer to this analysis as an RDD-style approach, to make this distinction clear.

Considering only moves that take place within the time frame of the data (1991-2015), variables which indicate both the date of the next, and the last, move at time t are created. Using the date of last move, the number of months (from the interview date) since the last move took place is created; the same is done for the amount of months before the next move. The sample is restricted further, using the date of interview and date of last move, to those who move within 12 months of the interview date. This allows for comparisons just before and just after a move has taken place, and to investigate anticipatory and temporary effects of moving. The following is estimated:

$$Y_{it} = \alpha_i + X_{it}'\beta + m_{it}\delta + M_{it}\tau + \varepsilon_{it}, \tag{2.10}$$

where m_{it} is the time, in months, before or since the individual moved house and $M_{it} =$

$1[m_{it} > 0]$. This equation is estimated using the `-rdrobust-` command in Stata (Calonico et al., 2017). This package allows for the implementation of bias-correction in the form of robust, nonparametric confidence intervals. This corrects for the fact that typical optimal bandwidth selectors tend to choose a large bandwidth, which can lead to biased confidence intervals (Calonico et al., 2014). In practice, the adapted confidence interval estimator is run using a higher-order polynomial than is used for point-estimation, which in this case is order one (i.e. local linear regression). The point-estimate bandwidth is restricted to four months either side of the cutoff[8], and present these point-estimates with the bias-corrected t-statistics and p-values.

This approach departs from the explicit Instrumental Variables estimation used previously. Identification relies on the assumption that whether an individual was interviewed just before or just after they moved was essentially random with respect to the outcome variables. This seems unlikely to hold in this context, so this assumption is not invoked, nor is there a casual interpretation of these results.

This mechanism is investigated further by fitting a piecewise linear spline function, with knots at various numbers of months after moving, m_{it}. Each knot is spaced by 3 months, and ranges from -12 to 12 months since moving location. This allows us to consider, at least descriptively, whether there are more extreme short-run effects, just before (anticipation period effects) and just after ("honeymoon" period effects).

2.4.4 Attrition

As is often the case when working with panel data, we may be concerned about attrition of individuals and households from subsequent waves. If the propensity to drop out of the survey in the next wave is unrelated to the variables of interest, individuals are said to be "missing at random" and this will not affect the estimates. However, in the health literature, health-related attrition has been a large concern when using longitudinal data on health. For this chapter, these concerns are worsened by the fact that the focus is on internal migration, which is another common cause of individuals dropping out. Simply put, considering the health effects of movers implicitly brings about a

[8]The aim here is to capture short run effects. This was as tight a bandwidth as possible without inducing an invertibility/cell-size problem.

sample selection: those who have worse health, and those who move, are more likely to drop out of the sample. The IV approach used earlier does not address this concern.

A simple diagnostic test for attrition, proposed by Verbeek and Nijman (1992), is performed on the data. Much of the literature that concerns attrition in panel data treats it as an absorbing state (i.e. those who drop out, stay out), whilst general attrition (with re-entrants) can be complicated to deal with (Wooldridge, 2010). Therefore the data are set up as though attrition is "absorbing"[9], and are first checked if there are any observable differences from the main sample. The variable addition test outlined originally in Verbeek and Nijman (1992), and used in a health setting with the BHPS in Jones et al. (2006), is subsequently performed.

For this test, define $s_{it} = 1$ if the data, (x_{it}, y_{it}) are observed for that period and 0 otherwise. The absorbed panel described above is used, so that the vector of indicators, $(s_{i1}, ..., s_{iT})'$ is Tx1, where $t = 1$ indicates the first wave in which the individual is observed (not necessarily the first wave), and $t = T$ indicates the last observed wave, with $T_i = \Sigma_{t=1}^{T} s_{it}$.

The intuition behind tests of this type is that, under the null of no attrition bias, we would not expect the outcomes to be correlated with the pattern of survey response, so a simple t-test of this response pattern, included in the outcome regression, will be informative of non-random attrition in the sample. To capture the pattern of response Verbeek and Nijman (1992) suggest including T_i as an additional regressor, an approach followed by Jones et al. (2006). The number of observed waves is, of course, constant within individuals and therefore cannot be used in a fixed-effects setup. Wooldridge (2010) extends this approach to fixed effects estimation by suggesting that s_{it+1} be included, allowing for within variation[10]. Alternatively r_{t+1}, a count of the number of waves after t that individual i is observed in the sample, can be used. Both of these approaches are used, including r_{t+1} and s_{it+1}, separately, in models with fixed effects. Fully robust standard errors, clustered at the individual level, are used in both cases.

[9] In other words, once an individual drops out of the household, they are not allowed to re-enter in later waves.

[10] s_{it-1} can be used equivalently, as suggested by Verbeek and Nijman (1992). However, doing so necessitates the loss of the first wave of the sample for all individuals, hence the use of a lead indicator.

2.5 Results

2.5.1 First-Stage results

Table 2.2 presents the results from the first-stage regressions, for both instrument sets and all outcomes. The upper panel shows the school choice instrument set, comprising of the interaction between the measure of expected utility from local schools, the age of the youngest child in the household, and their constitutive terms. With M binary, these represent linear probability models and ($\beta * 100$) are percentage-point movements in the probability of moving between the last period and the current period.

Though the coefficients from these models are not directly of interest per se, it is reassuring that the age of the youngest child in the household affects the probability of moving house in the direction we would expect. Households with a child aged 5-10 years and 11-15 years are 7.2 and 9.1 percentage points less likely ($p<0.01$ in both cases) to move house. In other words, those with a child of a relevant primary or secondary schooling age are less likely to move. These results hold as controls, individual fixed-effects and interview wave fixed-effects, are incorporated. The measure of expected utility from local schooling quality is statistically significant, but too granular to provide a meaningful coefficient. As expected, there is a negative sign on the coefficient, indicating that living in an area with low school quality is positively correlated with moving to another area. This is true for those with children of a relevant schooling age, as shown by the interaction term's statistical significance. As fixed-effects are added, however, this statistical significance drops away.

The main interest in the first-stage specification is whether or not the instrument set has relevance for, i.e. is a statistical predictor of, moving residence. A typical indicator of this power is the F-statistic from the first stage. For this analysis, due to the clustering of the standard errors at the individual level, Kleibergen-Paap-Wald F-statistics are reported. This instrument set shows a strong statistical power for predicting whether or not a household moves location. The usual rule-of-thumb criteria of $F > 10$ is only relevant for specifications that are just-identified; nevertheless these F-stats are in a com-

Table 2.2: First stage estimates of both instrument sets on the probability of moving

Panel A: School Choice Instrument				
Age of youngest Child$_{t-1}$:				
Child Aged 0-4	0.006	-0.035***	-0.073***	-0.074***
	(0.004)	(0.004)	(0.007)	(0.007)
Child aged 5-10	-0.072***	-0.075***	-0.092***	-0.093***
	(0.006)	(0.006)	(0.009)	(0.009)
Child aged 11-15	-0.091***	-0.081***	-0.076***	-0.078***
	(0.008)	(0.008)	(0.012)	(0.012)
Child aged 16+	0.064***	0.036***	0.022	0.022
	(0.014)	(0.014)	(0.019)	(0.019)
$E[Utility]_{t-1}$	-0.283***	-0.022	0.030	0.017
	(0.070)	(0.065)	(0.109)	(0.121)
$E[Utility]_{t-1}$*Age Youngest$_{t-1}$	0.162***	0.119***	0.087	0.095
	(0.042)	(0.040)	(0.061)	(0.061)
N	107736	107736	100406	100406
Kleibergen-Paap F stat	181.660	135.299	81.244	82.189
Panel B: House Price Instrument				
Local House Price Spike$_{t-1}$	0.117***	0.086***	-0.010	-0.010
	(0.007)	(0.007)	(0.008)	(0.008)
N	104152	104152	97768	97768
Kleibergen-Paap F stat	467.945	263.767	3.951	3.815
Controls		✓	✓	✓
Fixed Effects			✓	✓
Wave Dummies		✓		✓

Notes: Coefficients from the first-stage regressions of the instruments on the probability of moving home. Panel A shows the results from using the school choice instrument, where locally distance-weighted average school quality, interacted with the age of the youngest household, in the previous period, is the instrument set. Panel B shows the same results, but using instead the house price spike instrument, where living in an area which experienced a change in its averaged house price, of more than two standard deviations from the contemporaneous UK mean, is flagged as a "spike". The Kleibergen-Paap F statistics are shown at the bottom of each panel, indicating the relative strength of the instrument sets in their predictive power of moving home.

fortably large range, even in specifications that rely only on the within variation.

The bottom panel indicates the same information for the second instrument set: whether there was a house price spike in the previous period. We can see, as expected, a previous-period spike in local house prices means a household is more likely to move in the next period (8.6-11.7 percentage-points, $p<0.01$). Likewise, we can see remarkable power in this instrument as shown by the large F-statistics. However, the majority of this is

Table 2.3: OLS and IV second stage estimates of moving house on health outcomes: results using school quality and child age as an instrument

	OLS			2SLS IV: School Choice			
Outcomes:							
>Good SAH	0.003	-0.004	-0.010***	-0.263***	-0.366***	-0.088	-0.082
	(0.004)	(0.004)	(0.004)	(0.054)	(0.067)	(0.062)	(0.061)
GHQ-36 Score	0.145**	0.076	-0.004	-0.844	2.645***	1.352	1.300
	(0.062)	(0.061)	(0.056)	(0.725)	(0.915)	(0.872)	(0.868)
Health Problem	-0.001	-0.001	-0.002	0.469***	0.373***	0.248***	0.193***
	(0.005)	(0.004)	(0.004)	(0.073)	(0.071)	(0.078)	(0.074)
Controls		✓	✓		✓	✓	✓
Fixed Effects			✓			✓	✓
Wave Dummies		✓	✓		✓		✓
N	107736	107736	107736	107736	107736	100406	100406
First-Stage F				181.660	135.299	81.244	82.189

Notes: Second-stage coefficients of moving home, as instrumented by school choice, on various health outcomes. Each row represents a different outcome model, as indicated by the leftmost column. The school choice instrument consists of locally distance-weighted average school quality, interacted with the age of the youngest household, in the previous period. The Kleibergen-Paap F statistics from the relevant first-stage regressions are shown at the bottom of the table, indicating the relative strength of the instrument set in its predictive power of moving home.

driven by time-fixed factors, as this power dissipates with the inclusion of individual fixed-effects.

These results indicate that the proposed mechanism through which these instruments operate in the model, holds. Households with a child of a relevant school age, in areas with low school quality, are more likely to move. Likewise, those who live in areas which experience a local housing price spike are more likely to move. The tests of the instruments' statistical power suggests that the school quality IVs are in better stead than the house price instrument, when including individual fixed effects. These second-stage results for the latter, must therefore be treated with caution.

2.5.2 Second-stage results

Table 2.3 shows the main results from the schooling instrument set. The first three columns show the single-equation OLS results; the last four show the second-stage IV estimates. The outcome variable for each is indicated in the leftmost column.

For self-assessed health, there is no statistically significant effect observed until both individual fixed-effects and wave dummies are included in the model. Here, individuals who have moved are 1 percentage-point ($p<0.01$) less likely to report good, very good or excellent health. In the models which include instruments for moving, the coefficients become implausibly large in magnitude without the inclusion of fixed-effects. Taken literally, these imply that moving results in a 36 percentage point movement in the probability of reporting "good" health. It makes sense, given the impact of including them on the OLS results, to focus on the fixed effects models. Here, the point estimates fall in a large, but plausible range; movers are around 8 percentage points less likely to report at least good health. As is common when using a two-stage approach however, the standard errors are also much larger. The null of a zero effect cannot be rejected, nor can a significant difference from the OLS results at any reasonable confidence level.

The signs of the coefficients for the other two outcomes provide a consistent story: movers are more likely to have a health condition and report a higher (worse) GHQ score. There is no significant effect present in the fixed effects models however, and the statistically significant effect found in the IV specifications for health problems is perhaps implausibly large.

Table 2.4 shows the results from the house price instrument. Note the different sample and hence OLS estimates from the school quality instrument results. This is due to the fact that the price spike is constructed from $t-2$ and $t-1$ movements in local prices. This is offset (in terms of N) by the fact that data from 1995 onwards is used as opposed to 2000 onwards for the school quality IV.

In general, these estimates show the same qualitative interpretation as for the previous instrument set: moving is associated with worse physical and psychosocial health. The preferred specifications (the far right column) show similar point estimates for self assessed health, much larger for GHQ score, and much smaller for having health problem. The standard errors, on the other hand, are much larger. This is most likely a result of the weak instrument problem, which these fixed-effects models suffer from. This is illustrated by comparing the IV fixed effects standard errors to those from the standard

Table 2.4: OLS and IV Second Stage estimates of moving house on health outcomes

	OLS			2SLS IV: House Price Spike			
Outcomes:							
>Good SAH	-0.003	-0.007	-0.007*	-0.084	0.034	0.008	-0.075
	(0.005)	(0.004)	(0.004)	(0.064)	(0.101)	(0.470)	(0.463)
GHQ-36 Score	0.178**	0.185***	-0.012	-0.444	-0.972	6.754	5.975
	(0.071)	(0.068)	(0.062)	(0.883)	(1.424)	(7.383)	(7.072)
Health Problem	0.012**	0.002	-0.001	0.635***	0.056	-0.221	0.055
	(0.006)	(0.005)	(0.004)	(0.084)	(0.107)	(0.482)	(0.451)
Controls		✓	✓		✓	✓	✓
Fixed Effects			✓			✓	✓
Wave Dummies		✓	✓		✓		✓
N	104152	104152	104152	104152	104152	97768	97768
First-Stage F				467.945	186.581	3.951	3.815

Notes: Second-stage coefficients of moving home, as instrumented by house price spikes, on various health outcomes. Each row represents a different outcome model, as indicated by the leftmost column. The house price instrument is a dummy variable, where living in an area which experienced a change in its averaged house price, of more than two standard deviations from the contemporaneous UK mean, is flagged as a "spike". The Kleibergen-Paap F statistics from the relevant first-stage regressions are shown at the bottom of the table, indicating the relative strength of the instrument set in its predictive power of moving home.

2SLS estimates; they are roughly seven times as large.

Including the two instrument sets together in the same first stage is avoided in the main results, as the interpretation of the LATE - specifically, who other compliers to treatment are, becomes trickier to interpret, and less generalisable. This represents a trade of between the internal validity of the estimated LATE itself versus how externally valid this is to the population of interest. Nevertheless, given the relative poor performance of the price instruments in the fixed effects first stage, the two instrument sets are joined together and second stage results shown in table 2.5 below. The results do not differ substantively using this approach, despite the marked improvement in the first stage performance. For the purposes of the interpretation of the complier groups for each instrument set, the separate instrument sets are carried through for the rest of the chapter.

Number of observed moves

Of the 26,000 individuals in the estimation sample, 32% were observed to move at some point across all waves. Of these, 58% moved only once during the observation period, 22% moved twice, and the remaining 10% moved 3 or more times across all waves of the

Table 2.5: OLS and IV Second Stage estimates of moving house on health outcomes

	OLS			2SLS IV: School Qual & House Price Spike			
Outcomes:							
>Good SAH	0.002	-0.003	-0.011**	-0.216***	-0.819***	-0.075	-0.052
	(0.005)	(0.005)	(0.005)	(0.053)	(0.105)	(0.061)	(0.061)
GHQ-36 Score	0.205***	0.221***	0.102	-0.708	8.354***	1.129	0.980
	(0.076)	(0.075)	(0.073)	(0.725)	(1.431)	(0.849)	(0.847)
Health Problem	0.006	0.006	0.001	0.515***	0.554***	0.188***	0.076
	(0.006)	(0.005)	(0.005)	(0.069)	(0.104)	(0.069)	(0.066)
Controls		✓	✓		✓	✓	✓
Fixed Effects			✓			✓	✓
Wave Dummies		✓	✓		✓		✓
N	91879	91879	91879	91879	91879	85193	85193
First-Stage F				167.544	67.628	90.583	90.439

Notes: Second-stage coefficients of moving home, as instrumented by house price spikes and school quality instruments, on various health outcomes. Each row represents a different outcome model, as indicated by the leftmost column. The house price instrument is a dummy variable, where living in an area which experienced a change in its averaged house price, of more than two standard deviations from the contemporaneous UK mean, is flagged as a "spike". The Kleibergen-Paap F statistics from the relevant first-stage regressions are shown at the bottom of the table, indicating the relative strength of the instrument set in its predictive power of moving home.

BHPS and USoc. Ignoring the number of observed moves may induce omitted variable bias, as it affects the probability of a subsequent move, and may be correlated with health outcomes, given the hypothesised relationship in this chapter. As a robustness check, I include the cumulative total of observed moves for each individual, at each wave. Using a cumulative total allows for its use in fixed effects models. The results are shown in Appendix A, table A.2. Controlling for the number of moves does not substantively change the results, except for the probability of having a health problem which is now insignificant at the 10% level.

2.5.3 Heterogeneous treatment effects: Age

In order to try and pick up any heterogeneous effects in the results presented this far, the results are stratified by various age bands. Table 2.6 displays the fixed effects IV estimates of moving on health outcomes, by these different strata. Table 2.7 shows the same stratification, but using prices as an IV. The first stage F-stats from table 2.6 show some variation in the predictive power of the school quality instrument. This does not change too much across age bands, but suggests that younger individuals are more likely to be influenced to move house by local schooling quality. The results for self-assessed health

Table 2.6: Fixed effects IV estimates stratified by various (overlapping) age bands, using school quality instrument

	18-30 years	20-35 years	25-40 years	30-45 years	35-50 years
Outcomes:					
>Good SAH	0.043	0.035	0.071	0.018	0.036
	(0.078)	(0.099)	(0.118)	(0.118)	(0.128)
GHQ-36 Score	0.087	-0.755	-1.349	-1.455	0.726
	(1.125)	(1.293)	(1.474)	(1.561)	(1.944)
Health Problem	0.302^{***}	0.304^{***}	0.250^{*}	0.129	0.147
	(0.091)	(0.115)	(0.129)	(0.131)	(0.152)
Controls	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓
Wave Dummies	✓	✓	✓	✓	✓
N	20045	25838	31632	36869	38381
First-Stage F	26.666	18.749	17.834	22.731	21.017

Notes: Age-stratified second-stage fixed effects IV estimates of moving home, as instrumented by school choice, on various health outcomes. Each row represents a different outcome model, as indicated by the leftmost column. The school choice instrument consists of locally distance-weighted average school quality, interacted with the age of the youngest household, in the previous period. The Kleibergen-Paap F statistics from the relevant first-stage regressions are shown at the bottom of the table, indicating the relative strength of the instrument set in its predictive power of moving home.

Table 2.7: Fixed effects IV estimates stratified by various (overlapping) age bands, using price instrument

	18-30 years	20-35 years	25-40 years	30-45 years	35-50 years
Outcomes:					
>Good SAH	0.846	0.015	-0.018	-1.223	0.189
	(0.637)	(0.292)	(0.260)	(1.702)	(0.601)
GHQ-36 Score	7.523	-1.375	-1.454	5.342	-0.177
	(7.313)	(4.165)	(4.038)	(15.527)	(6.584)
Health Problem	-0.694	-0.091	0.826^{**}	0.679	-0.390
	(0.568)	(0.277)	(0.352)	(1.257)	(0.728)
Controls	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓
Wave Dummies	✓	✓	✓	✓	✓
N	11481	20782	26395	22027	12646
First-Stage F	5.839	12.288	30.118	2.828	12.086

Notes: Age-stratified second-stage fixed effects IV estimates of moving home, as instrumented by house price spikes, on various health outcomes. Each row represents a different outcome model, as indicated by the leftmost column. The house price instrument is a dummy variable, where living in an area which experienced a change in its averaged house price, of more than two standard deviations from the contemporaneous UK mean, is flagged as a "spike". The Kleibergen-Paap F statistics from the relevant first-stage regressions are shown at the bottom of the table, indicating the relative strength of the instrument set in its predictive power of moving home.

remain fairly consistent, as does the imprecision of these estimates, however. The results for having a health problem become implausibly large when broken down by age bands, whilst the GHQ score shows worse mental health for those who move in the youngest and oldest categories, and a positive effect (negative sign) for the age-bands in between. The results in Table 2.7 show some improvement in the first stage for those aged 25 to 40 years, but again these results remain statistically insignificant and do not offer any additional interpretation to the main results presented earlier.

2.5.4 Heterogeneous treatment effects: Housing tenure changes

It is clear that there are many different reasons and mechanisms behind why a move takes place. Given the instrument that we use, the analysis is limited to a local average treatment effect (LATE), which is only relevant for the subgroup of compliers who are in a position to move house, based on their child's schooling. Even within this local effect, however, there are many different types of move. This is perhaps best captured by changes to housing tenure: moving from one owned house to another is likely to have a different effect to moving from a private rental to ownership. The LATE estimated comprises of a weighted average of the effects of all combinations of housing tenure changes. Table 2.8 shows the housing tenure transition probabilities for the sample: the probability of moving to housing tenure at time t (columns), given residence in housing tenure at time t-1 (rows), conditional on a move having taken place.

Table 2.8: Housing tenure transition probabilities for full sample

Time t:	Owned/ Mort.	Shared Own.	Private Rent	Public Rent
Time t-1:				
Owned/ Mortgage	95.89	0.75	2.69	0.67
Shared Ownership	3.54	93.77	2.25	0.43
Private Rent	10.50	4.13	75.00	10.37
Public Rent	2.60	0.63	7.49	89.28

Notes: Transition probabilities are calculated as the frequency of individuals who reside in a given tenure type in wave $t-1$ and a given housing tenure type in wave t, for all waves.

These heterogeneities are illustrated by interacting relevant changes to housing tenure

Table 2.9: Housing tenure transition OLS coefficients of moving house, for full sample

Time t:	Owned/ Mort.	Shared Own.	Private Rent	Public Rent
Time t-1:				
Owned/ Mortgage	0.050***	0.065***	0.006	-0.035
Shared Ownership	0.029	0.038***	0.027*	-0.040
Private Rent	-0.024*	0.018	-0.008	-0.033*
Public Rent	-0.012	-0.003	-0.028*	-0.090***

Notes: Migration coefficients from OLS models of self assessed health, regressed on each housing tenure change (conditional on migration); Base category for all models is all other housing tenure change; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; N=24,338

with migration. If using the instrumental variable approach, this would involve estimating separate first-stages for each interaction term, each instrumented by their interaction with age of the youngest child and school quality instruments. As illustrated by Table 2.8, there are 16 possible combinations of housing tenure movements. Using these interactions for all 16 combinations of move would push the data too far, so for this secondary analysis, OLS is used to simplify the exposition (and thus no causal claims are made at this point).

Table 2.9 shows the analogue to Table 2.8, but reports OLS coefficients from the migration coefficient for each type of housing tenure change. All of these models are conditional on a move having taken place, and as such, the base category is all other types of move. Some clear heterogeneities in the effect of migration on health, by housing tenure, appear. The effect of moving from private rent to home ownership (row 3, column 1) is negative in terms of self-assessed health, whereas simply moving house from one that is owned, to another (row 1, column 1) shows a positive effect of migration. These represent two different types of move, with the former often being an individual's first house purchase, involving securing a mortgage and may place a different type of pressure on a household than one of a more experienced move, such as that captured by the latter. Of particular note from this table are the negative signs on each of the moves to, or from, public rental tenure. Considering the difference in probabilities of remaining in the same type of rental property between private and public rent (see Table 2.8), this brings to light a potential problem with public renters in the U.K.: it is likely that once an individual rents publicly, they will continue to do so in the next period (around 90%), and doing so is associated with much lower self-assessed health.

2.5.5 Short run effects (RDD) results

In this section the regression discontinuity results are presented, comparing the health outcomes of those who were interviewed around the time they moved residence. The running variable, m, is defined as the number of months after the last move that the interview took place. The cutoff is where $m = 0$. For these individuals, who were interviewed in the same month as they moved, it is impossible to determine whether the interview took place before or after move itself, so they are dropped from the analysis.

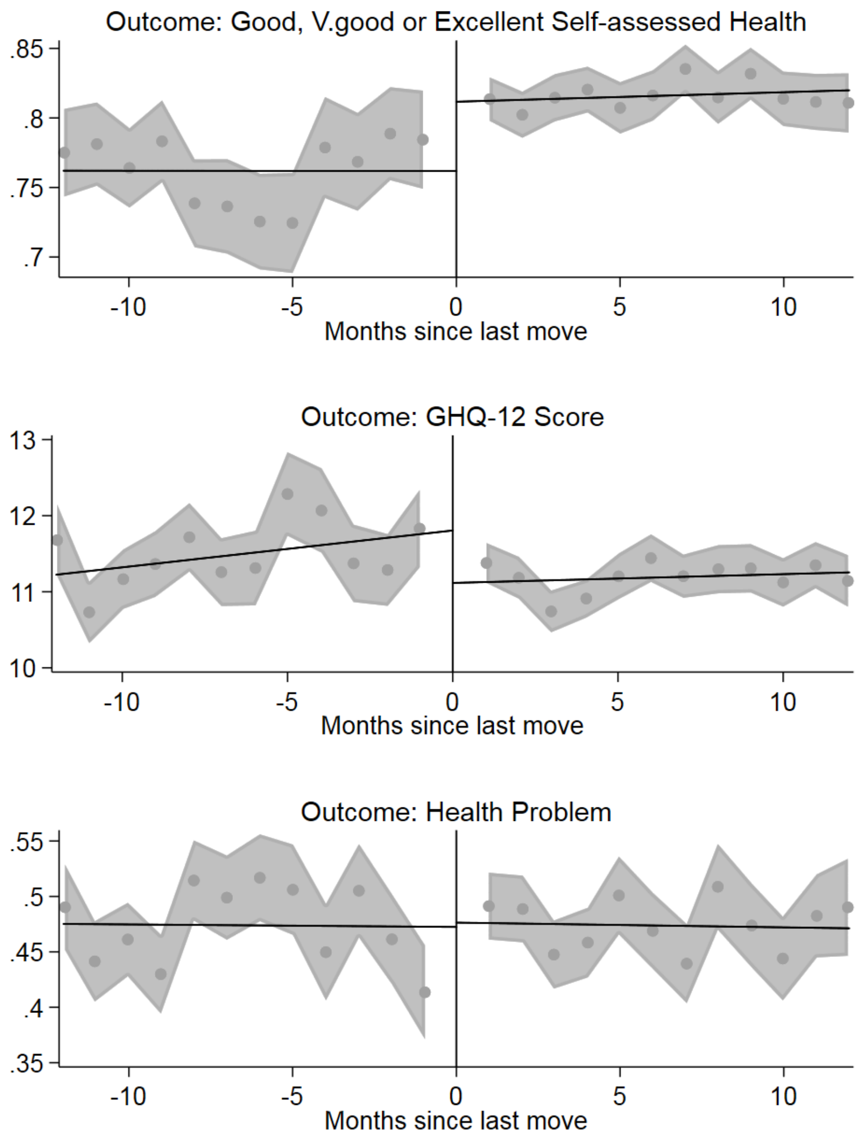
Figure 2.2 illustrates a simplified version of this set up, and shows the outcomes over m. We can see small differences for the GHQ-12 and self-assessed health measures. Another consideration for this approach is whether the covariates vary over the cutoff. This is often seen as a placebo test, providing credit to the identification assumption if they vary smoothly. Figure 2.3 shows this for several variables. For each of these a discontinuity can be observed, and thus are controlled for in the main specification.

Table 2.10 shows the regression discontinuity results. As expected, the null of no effect for the physical health measures (self-assessed health and whether or not the individual has a health problem) cannot be rejected. It can be seen, however, that there is a statistically significant positive effect of moving residence on psychosocial health, as measured by the GHQ-12 score.

Anticipatory and temporary effects of moving

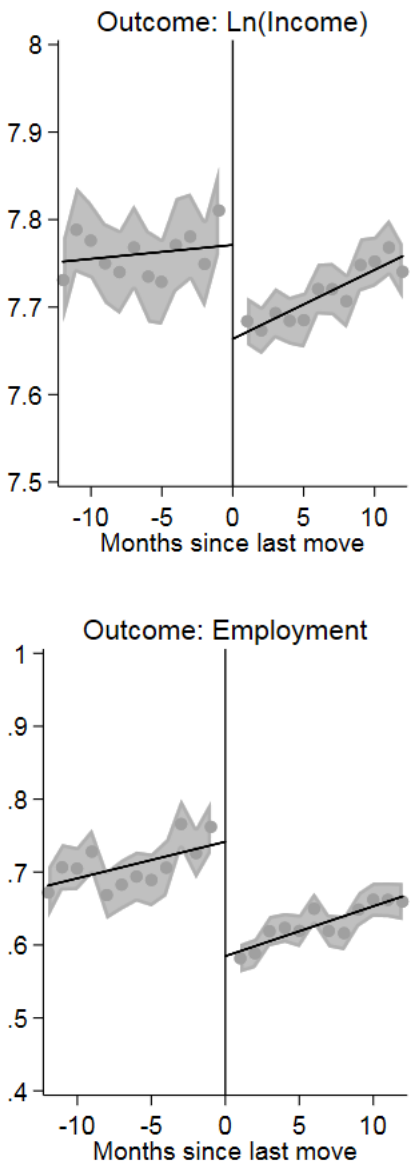
This final piece of descriptive analysis investigates whether the (lack of) short-run effects are the result of two offsetting mechanisms. The hypothesis is that, especially in the case of psychosocial health (and therefore the GHQ-12 score), most of the stress associated with a move occurs before the event itself. Likewise, this heightened stress may well dissipate quickly in the months following a move, as individuals adjust to new surroundings. To capture this, a piecewise-linear spline function of the months prior to a move is fit against each health outcome. The focus is on those who are interviewed within 24 months of moving, and nine knots in three-month intervals from 12 months

Figure 2.2: Health outcomes before and after moving residence



Notes: Graphical comparison of individuals' outcomes who are interviewed up to 12 months before a move, with those who are interviewed up to 12 months after a move. Months since last move = 0 represents the month in which a move took place: if the interview fell in this month, the individuals were dropped as it is impossible to determine from the data whether they were interviewed before or after a move.

Figure 2.3: Various covariates of health before and after moving residence



Notes: Graphical comparison of individuals' covariates who are interviewed up to 12 months before a move, with those who are interviewed up to 12 months after a move. Months since last move = 0 represents the month in which a move took place: if the interview fell in this month, the individuals were dropped as it is impossible to determine from the data whether they were interviewed before or after a move.

Table 2.10: Short-run effects: RDD results around the cutoff of moving residence

	Better than good SAH	GHQ-12 Score	Health Problem
RD Estimate	0.006	-0.514	0.131[a]
	[-0.112, 0.190]	[-3.705, 0.626]	[-0.054, 0.314]
Controls	-	-	-
Bandwidth (both sides of cutoff)	4	4	4
Effective N (left,right)	(1,171, 5,595)	(1,146, 4,781)	(1,170, 2,655)
RD Estimate	0.029	-0.711*	0.113[a]
	[-0.057, 0.252]	[-4.471, -0.233]	[-0.135, 0.235]
Controls	✓	✓	✓
Bandwidth (both sides of cutoff)	4	4	4
Effective N (left,right)	(1,066, 4,906)	(1,105, 4,495)	(1,096, 2,403)

Robust 90% confidence intervals in brackets; * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
[a] Significant at the 5% level when regular standard error calculations are used.

prior, to 12 months post a move, are included. This approach is illustrated in Figure 2.4. It is clear from this plot, that six months prior to moving, the average GHQ-12 score is elevated. This rapidly moves in the opposite direction once a move has taken place, and then returns to a baseline level after nine to twelve months.

These results are in line with the notion that any short run effects picked up in the previous analysis are an average over these two periods: an anticipatory effect of worse psychosocial health, offset by a "honeymoon" period of better psychosocial health.

Figure 2.4: Illustration of the spline function

Notes: A linear piece-wise spline function, fitted to the sample of those who are interviewed within two years before or after moving. Knots are fitted in three-month intervals, ranging from 12 months prior to a move and 12 months post.

2.5.6 Attrition

Table 2.11 shows the tests of health-related attrition that were described earlier. The top row shows the coefficients from the fixed effects models that do not include the test variables. The remaining two "mover" rows show these same coefficients when the models are added. Adding the test variables does not address the attrition in any way, but they are useful for comparative purposes. There is little difference between these estimates. The row for r_{t+1}, shows the test using a count of the number of waves remaining after t. The row for s_{t+1}, shows the test using an indicator of whether the individual is observed in the following wave. In general, it seems that patterns of attrition out of the dataset are not associated with health outcomes, with the null of no association between attrition and health outcomes failing to be rejected in all but one case. This case suggests that the number of future sample periods remaining at time t is associated with having a health problem at time t. This estimated association, however, is extremely small in magnitude (0.2 percentage points). It can be concluded therefore, that health-related attrition does

Table 2.11: Coefficients and t-stats from tests of health-related attrition

	SAH	GHQ-12	Health Prob.
Mover	-0.010**	-0.004	-0.002
	(0.004)	(0.056)	(0.004)
Mover with r_{t+1}	-0.009*	-0.005	-0.002
	(0.004)	(0.056)	(0.004)
r_{t+1}	-0.011	0.087	0.002**
	(-1.64)	(0.87)	(3.10)
N	107736	107736	107736
Mover with s_{t+1}	-0.012**	-0.017	-0.003
	(0.004)	(0.064)	(0.004)
s_{t+1}	0.008	-0.100	-0.001
	(1.54)	(-1.46)	(-0.23)
N	90389	90389	90389

Parentheses: t stats for test coefficients, SEs for Mover coefficients; Mover represents the coefficient on moving residence from these fixed effects models, with and without the added attrition test variables; r_{t+1} represents number of observed waves remaining; s_{t+1} is an indicator of whether observed in next period; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

not seem to be a major concern for this analysis, after using this approach that is the norm in the health literature.

2.6 Discussion

There is a shortage of literature that considers the role of moving from one place of residence to another, on health outcomes. This chapter has sought to address this gap in the literature using an instrument that has not been utilised as a predictor of moving house in the economics literature before.

The results from the IV analysis concludes that moving house is associated with worse physical and psychosocial health. Any causal conclusions must be treated with caution, however. The fact that the IV estimates are larger than the OLS estimates may, in part, be due to the fact these estimates represent a Local Average Treatment Effect (LATE); they represent the effect only for the compliers with the instrument. In the case of the school choice, for example, the point estimates represent the effect of moving house, *for those who are both willing and able to move location for their child's schooling quality*. The other possibility cannot be ignored: that the large coefficients could be due to a failure of the exclusion restriction. This seems less likely for the fixed effects models that use the local variation in school quality in the instrument set; but this assumption remains fundamentally untestable.

There is a trade off between consistency and variance amongst the choice of estimators. We cannot ignore the endogeneity that renders OLS inconsistent. However, turning to a seemingly plausible and carefully thought-out instrument results in a large variance (the standard errors almost always contain the OLS estimates in this case), meaning that the results do not necessarily provide more confidence in the conclusions made. This remains a common and important issue in papers that rely on an IV identification strategy, but it is one that often gets overlooked in the discussion of the results[11]. Nevertheless, the coefficients of the full specifications with fixed effects lie in a plausible range for the self-assessed measures and represent the take-away estimates from this analysis: Individuals who move are 8 percentage points less likely to report very good or excellent health and report 1.3 points higher on the GHQ score.

The secondary analysis of this chapter reveals a short-run effect of moving house that occurs before and after the move takes place. Individuals who are interviewed within six months prior to moving house show elevated GHQ-12 scores, reflecting worse psychosocial health. In addition to these anticipatory effects, a period of better psychosocial health is observed in the three months post-move, which returns to the baseline level in the long-run. This suggests that whilst moving can be a period of intense stress for a household, individuals quickly adjust once the move has taken place. These results, along with the regression discontinuity approach do not identify causal effects of moving, however. Nevertheless, the lack of clean identification of causal effects should not deter research on an important question that remains unanswered in the literature (Ruhm, 2018).

Also apparent from the analysis is the likelihood of heterogeneous treatment effects. This is driven by the fact that there are many different reasons for moving residence, and

[11] See Young (2017)

many different types of move. Any effect estimated, even with a perfect instrument, will comprise of a weighted-average of other, more local, average treatment effects depending on the type of move and the motivations for doing so. The sixteen combinations (see Table 2.8) of moves between housing tenure provides a framework for literature looking at internal migration. For example, a recent paper by Munford, Fichera, et al. (2017) considers the effect of home ownership on health, using exogenous variation from the Right to Buy policy in the 1980s. As this policy affected those who resided in council houses and encouraged home ownership, this paper estimates the causal effect of moving from private rent to home ownership (analogous to row 4, column 1 of Table 2.8). This paper is an important contribution to the literature that explicitly considers a specific mechanism of home ownership, and this framework should help guide future work that considers different mechanisms of internal migration.

There are other potential features of migration that warrant further research. The effect of a move for tied movers versus non-tied movers, for example, is likely to have a very different effect on health outcomes. The effects of moving house at older ages presents another avenue for research, as do changes to housing tenure. Future work involves teasing out these mechanisms further in a sensible way, carefully using different instruments to construct other LATEs of migration on health. These estimates will help unpick complex, yet commonplace behaviour.

CHAPTER 3

Are Health Outcomes Worse on the English Coast? Selection on Observables vs Unobservables

Abstract

UK data shows that health amongst the working-age population (16-64 years) is worse on the coast than elsewhere. For example, there is a much greater prevalence of limiting long-term health conditions on the coast as opposed to the average for England and Wales. Despite this, there is a lack of literature that considers the potential reasons for these differences and how they can be identified; this paper addresses this gap. Using data on health and other characteristics from all five waves of Understanding Society, a UK household panel dataset, we quantify the differences in health and health-related outcomes on the coast compared to inland. Detailed geographic data are used to construct a distance to the coast measure, which we use as the main distinction between a coastal and non-coastal area. We find that most health-related outcomes are worse on the coast, including long-standing health conditions, disability benefit claimants and smoking prevalence being more likely. A notable exception is frequent physical exercise, which is more prevalent on the coast (2.8 %pts more likely, $p<0.001$). It seems living on the coast may well be related to worse health, but further research into the potential mechanisms and identification issue is needed.

3.1 Introduction

This chapter considers the difference in health and health-related outcomes between individuals who live on the coast versus those who live inland. A recent study by the Office for National Statistics (ONS, 2014), which uses 2011 census data to provide an overview of coastal towns in the United Kingdom (UK), shows that health is worse on the coast than inland. Indicators other than health, notably education and employment, also paint the UK coast in an adverse light. For example, a recent report conducted by the Future Leaders Trust (2015) concluded that schools on the coast are among the worst performing in the UK, attributing this to geographical isolation and industrial decline.

This chapter addresses two main research questions: Are health outcomes worse on the English coast, and how likely is it that these differences are driven by selection on unobservable factors? It does so using Understanding Society with lower-layer super output area (LSOA) linkage. Distance to the coast is measured from the population- weighted centroid of each household's LSOA, to the nearest coast. This is then used to create a binary "coastal" variable, which is included in regression analysis of the outcomes, amongst other covariates. The effects of interest are then estimated using a range of different methods, and show the results do not differ significantly based on functional form.

There are likely unobservable, systematic, differences between individuals living on the coast versus those who do not due to self-selection. An absence of plausibly exogenous variation in living on the coast rules out identification through an instrumental variable or natural experiment. This chapter therefore uses methods first proposed by Altonji et al. (2005) and recently developed by Krauth (2016) and Oster (2019) to assess how large the selection on unobservables would have to be, relative to selection on observables, for these results to be statistically insignificant. Subsequent analysis also looks more closely at the role of house prices, access to health care providers, movers and non-movers to disentangle the mechanisms at play.

Our main results show that individuals residing on the coast are 1.9 percentage points more likely to claim disability benefits ($p<0.01$), 3.2 percentage points more likely to have a long-standing health condition ($p<0.01$), 8 percentage points more likely to have smoked ($p< 0.01$) and 3.1 percentage points more likely to have drank frequently in the past week. Contrastingly, individuals are more likely to have participated in frequent moderately intense exercise ($p<0.01$). There were no differences in self-assessed physical and psychosocial health - a result contradictory to previous literature. These are important new findings, and raise new questions for policy, surrounding unmet need on the coast.

The chapter continues as follows: Section 3.2 provides some background to the features of the coast that motivate the study; Section 3.3 outlines the data and relevant variables used, including details on how the coast variable is constructed; Section 3.4 outlines the methods used; Sections 3.5 and 3.6 report the findings and Section 3.7 concludes with a discussion of the results and limitations.

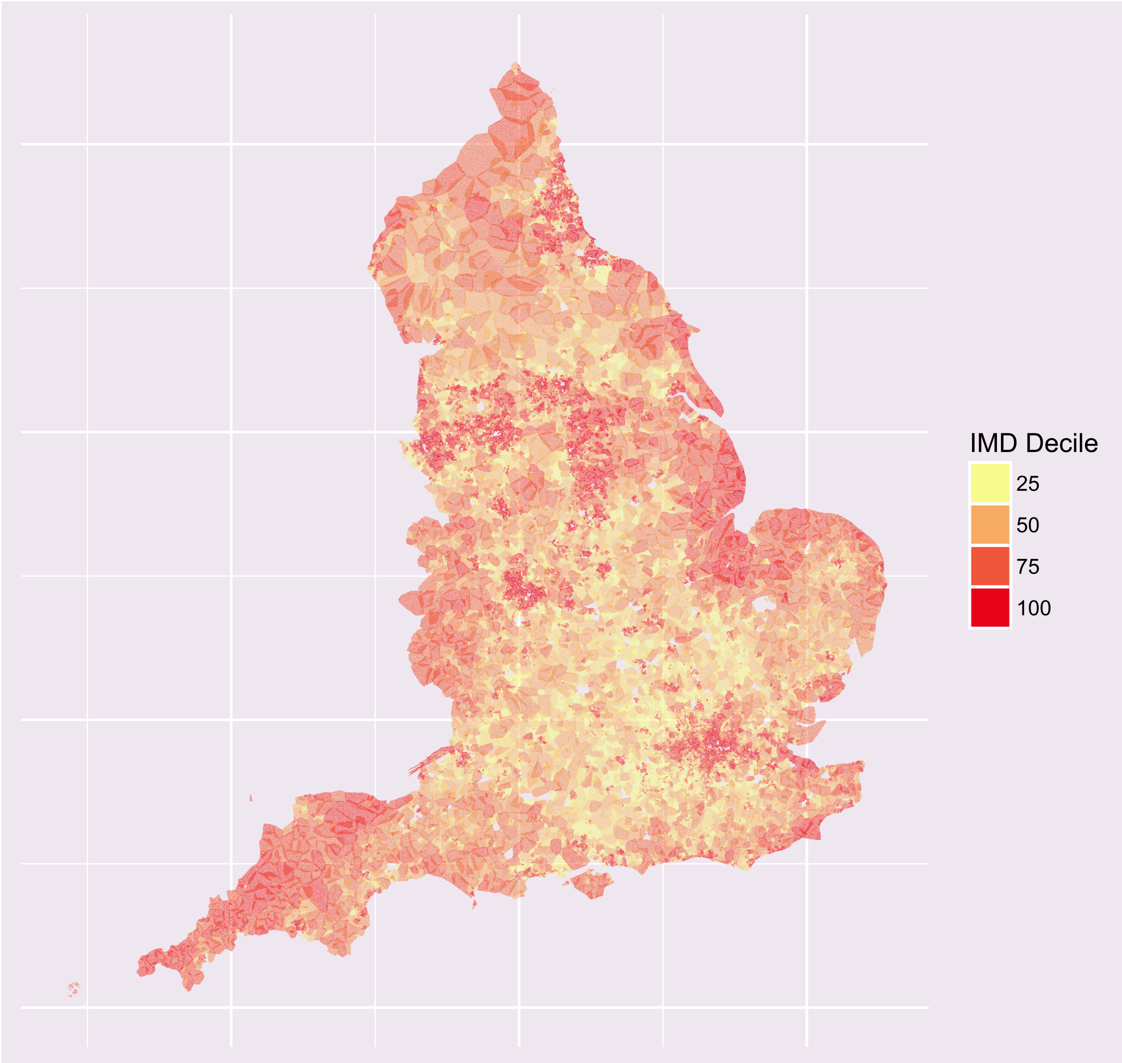
3.2 Background

There is observational evidence to show that the English coast is associated with worse outcomes for a host of health-influencing factors. Towns on the English coast are associated with higher deprivation, increased drug use, poorer education, and worse employment outcomes. Figure 3.1 plots area-level Index of Multiple Deprivation (IMD) rankings on a map of England. Once we ignore the main cities (i.e London, Manchester, Birmingham, Leeds and Sheffield), the most deprived areas are clearly concentrated along the coast. A 2018 ONS report showed that 6 of the 10 local authority districts in England and Wales with the highest rates of heroin and/or morphine misuse-related deaths were on the coast[1]. According to a report by the Social Market Foundation (2020), the two local authorities with the smallest proportion of those aged 16+ holding a level 4 or above qualification, and 5 of the 10 local authorities in Great Britain with the lowest average pay, were coastal communities.

Yet, there is little empirical evidence on whether health-related outcomes themselves are indeed worse. This motivates the primary question this paper seeks to address:

[1] https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/halfofheroinmorphinemisusedeathhotspotsinenglandandwalesareseasidelocations/2018-04-04

Figure 3.1: Deprivation in England: by LSOA-level IMD Decile



are health- related outcomes worse on the English coast, after conditioning on health-influencing factors?

3.2.1 Why are coastal regions more deprived?

The prevalent issues in coastal areas in the UK which contribute to worse deprivation can be summarised as: a decrease in popularity for living and holidaying in these locations, poor infrastructure, the socioeconomic status of residents, and the reinforcing of these problems by the "brain drain" from these regions.

As an island, the UK has historically been heavily reliant on its maritime trade. This became of even greater importance during the Industrial Revolution; increased demand for ship building and repair was brought about by increases in coastal, foreign and slave trades (British Museum, 2008). This, along with the already strong agricultural and fishing industries, led to a huge surge in jobs and wealth on the coast, drawing in populations from the inland cities. The coast has also prospered from being a holiday destination for many, throughout the 1900's, boosting the economy of the areas not built upon the aforementioned industries. However, these industries are no longer what they once were, having faced a steady decline over the past 30-40 years, leaving the populations who still live on the coast worse off as a result. Plant closures have resulted in heavy job losses and cheap flights mean the prospect of going to the English coast for a holiday has been replaced by a Spanish coast. These factors can offer some potential explanation for the deprivation that affects certain parts of the English coast today.

Coastal towns are generally characterised by seasonal business, with high numbers of small or medium sized businesses with poor digital skills, which impedes growth through a strong digital infrastructure. Seaside towns also exhibit a high proportion of the elderly population, as popular retirement locations, and low opportunities for entry level jobs, especially for yount people (Parliament, 2019). This results in outward migration of younger individuals and skilled labour ("brain drain") resulting in an ageing population. This has consequences on natural reproduction of the population, and therefore a decrease in the human capital necessary for further development, further reinforcing the socio-economic issues (Farwick, 2009).

3.2.2 Theoretical Underpinnings

The factors outlined in section 3.2.1 outline the distinct mix of problems in coastal towns which contribute to worse social deprivation. Socio-economic status is a known contributor to health status, particularly at old ages (Salas, 2002; Siciliani & Verzulli, 2009). This study is concerned with whether living on the coast is associated with health and wellbeing, holding these factors constant. Therefore, I hypothesise two, conflicting, potential mechanisms for why living on the coast should affect health and wellbeing, all else equal:

1. Living on the coast has a direct, positive, effect on health and well-being: the coast as

"bluespace".

Bluespace is defined in a similar fashion to greenspace in the literature: "health-enabling places and spaces, where water is at the centre of a range of environments with identifiable potential for the promotion of human wellbeing" (Foley & Kistemann, 2015). The coast is a subset of bluespace, and this approach suggests a possible hypothesis that living by the coast is good for health. In particular, a positive effect through increased physical activity is a salient feature in the coast-health mechanism.

2. Living on the coast has a negative impact on health through insufficient provision of healthcare services. In coastal regions, this unique set of issues is not properly reflected in resource allocation formulas, therefore need is not properly accounted for, and the supply of healthcare services does not properly meet the demand for healthcare in coastal regions, therefore health worsens.

A recent House of Lords select committee report on the future of seaside towns (REF) provides anecdotal evidence that though the national funding formula reflects an area of deprivation score, it does not adequately account for the complex set of circumstances faced by healthcare professionals in these areas. Examples include inability to notify the population of public health interventions such as screening programmes through conventional methods of communication by practices (letters), as many patients live in unstable accommodation with high turnover. The high proportion of multimorbid patients in these areas, and the higher utilisation of services they require and extra work this entails for GPs is also not reflected in the funding formula for GPS, resulting in low numbers compared with the demand.

3.2.3 Literature Overview

There is a small body of literature which considers the physical or mental health effects of living near to the coast. Mostly, the international literature approaches the problem from a "bluespace" perspective. A recent systematic review considering, amongst other phenomena, the mental health benefits of bluespace (Gascon et al., 2015) identified only three studies (White et al., 2013; De Vries et al., 2003; Triguero-Mas et al., 2015) which fit this criteria, concluding that there is limited evidence of casual effects. Völker and Kistemann (2011) consider a wide range of studies, in a systematic review of those which consider the health effects of inland surface waters. Although this specifically excludes the coast, the authors identify a need for more quantitative evidence in such areas.

There are several studies which consider coastal health effects, rather than bluespace in general, in a UK setting. Using 2011 census data from the UK, Wheeler et al. (2012) consider whether living by the coast improves health and well-being. They employ multivariate linear regressions, with measures for self-assessed health and well-being as the outcomes of interest. They find that good health is more prevalent the closer an individual lives to the coast. In a more detailed approach, White et al. (2013) consider the same research question, but employ panel data methods. They find in a fixed-effects specification, that physical and mental health is better the closer one lives to the coast, but life satisfaction is no different. However, they can only identify effects for those who move to or from the coast, and these may be imprecise, due to small individual or household-level variation. In a paper which considers many different types of environment (including the coast) and health, Wheeler et al. (2015) find some statistically significant and positive associations between living in a coastal environment and good self assessed health. However this study, as do all of those considering health on the UK coast, fails to address the endogeneity of living on the coast with respect to health. The most recent study to consider health and the UK coast explores the relationship between childhood obesity and proximity to the coast (Wood et al., 2016). Regressing obesity prevalence (BMI $\geq 95^{th}$ percentile) on coastal proximity and area-level confounders, the authors find that living nearer to the coast is associated with lower obesity, though this relationship depends on the type of area (rural or urban areas). Specifically, the coastal gradient in obesity was not found for large urban areas.

The common theme within all of the literature reviewed here is that, ex ante, bluespace is thought to be *beneficial* for health, and these hypotheses are carried forward into the literature which considers the effects of living in a coastal environment. Another feature of this literature is that the health measures used tend to be self-assessed, and assessed in absolute terms; using objective measures of health and assessing distributional effects could add to the existing evidence. Finally, an important omission of the current literature is that the endogeneity of the coast with respect to health is not explicitly considered, nor are the implications discussed. This paper aims to offer some insight into the role of non-random selection into living on the coast.

3.2.4 Identification Issues

There are several identification issues faced when considering the health-related consequences of living on the English coast. These can best be summarised as the likely systematic difference between individuals living on the coast versus those who do not. This means that a simple comparison of averages will yield results that are likely biased.

If these differences are along observable dimensions, such as age, income and education, then simply controlling for the variation in these characteristics using Ordinary Least Squares (OLS) regression will provided consistent and unbiased estimates of the coastal differences. If, however, these differences are unobservable (as is likely the case when using observational data) then we cannot give any estimate a causal interpretation, in the absence of a quasi-experiment or other plausibly exogenous source of variation in living on the coast (i.e. an Instrumental Variable (IV)).

In the absence of such an instrument, there are few approaches one can take in the search for a causal interpretation. In light of this, methods first proposed by Altonji et al. (2005), who assess the role of selection on observed and unobserved variables in establishing the effectiveness of Catholic schools, are utilised. These methods were subsequently formalised by Krauth (2016) and Oster (2019) as a way to bound a linear causal effect by relaxing the relative correlation restrictions placed on the model. This allows an assessment of how large the correlation between the coast and the unobservables (i.e the selective migration) would have to be, relative between its correlation with the observables, for the estimate of the coastal effect to tend to zero.

Aside from the observable determinants of health, there are several potential sources of unobserved heterogeneity in individual responses, that relate to living on the coast.

One of the main features of a coastal environment in the UK is geographical isolation. This can have potentially far-reaching consequences for health on the coast, through limited access to the same level of health care as one would otherwise have in, say, a city. In supplementary analysis to assess this, proxies for the level of health care access at the local area level included in order to soak up the variation geographical isolation that affects health. Additionally, local house prices are added to the analysis to pick up heterogeneity in the coast and it's effect on health and health related outcomes.

Another potential source of systematic differences in health on the coast is selective migration. This can arguably take the form of two different types of selection: "unhealthy" inward migration and "healthy" outward migration. The former is driven by retirement, and captures the older (and sicker) populations moving towards the coast in their retirement, whilst the latter relates to younger populations moving inland to cities to improve labour market outcomes. In supplementary analysis the main models are stratified by these age groups as a basic test to see if the data supports this hypothesis. Analysis is carried out separately on only on those who move, and then on only those who do not move, between the coast to get a hold of these potential mechanisms.

The approaches taken in this chapter contribute to the literature on coastal health by using new data, new methods that test the sensitivity of point estimates to selection on unobservables, and by testing several mechanisms that may confound the main effect.

3.3 Data

This chapter uses data from Understanding Society (USoc), a longitudinal survey of approximately 40,000 households in the United Kingdom. The members of each household are revisited annually, with data collection for each wave taking place over a two-year period. Individuals in the household aged over 16 years have a face-to-face interview in addition to a short questionnaire. The survey began in 2009, and is currently in its 5th wave, which corresponds to 2013/14. For the purposes of this analysis all 5 waves[2] are used, restricting the sample to those residing in England[3], as the area-level

[2]With the exception of models which consider smoking, drinking and physical activity; these variables are only available in certain waves.

[3]Those living on the Isle of White are also dropped, as doing so perfectly predicts treatment (living within a close proximity to the coast).

data for Wales and Scotland was insufficient for the proposed coastal variable. For this variable, a linkage between USoc and 2011 census geographical data is used; specifically the household's lower-layer super output area (LSOA). Knowing whether an individual *lives* at the address which is deemed to be on the coast or not is important for the analysis and the sampling design of USoc ensures that the individuals interviewed are permanent residents of the address (Lynn et al., 2009), so individuals are not mistakenly flagged as being on the coast. There is no information on whether or not a holiday home is owned, so those who may experience the benefits of the coast through this mechanism cannot be disentangled from those who do not.

3.3.1 Variable Definitions

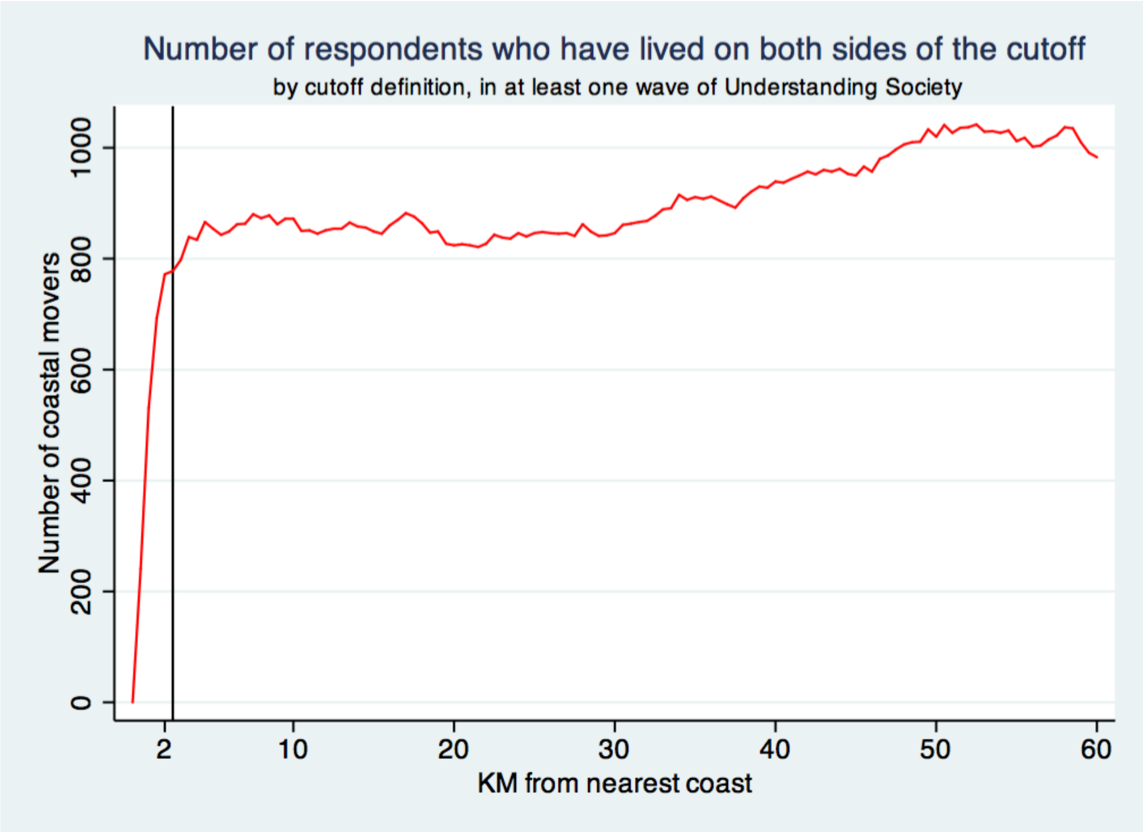
USoc contains data on individual characteristics, and variables which measure both the determinants of and levels of health and health-related outcomes. These health-related measures take the form of our main outcome variables, while the former were used as controls to capture observed individual heterogeneity.

Defining the Coast

Crucial to the analysis of this paper is a variable which captures whether or not an individual lives on (or near to) the coast. Following (Wheeler et al., 2012) and (White et al., 2013), data on the household's LSOA has been incorporated to measure this. LSOAs are small areas of roughly the same population size. Each LSOA covers between 1,000 and 3,000 people. The population-weighted centroid coordinates were obtained from the (Office for National Statistics, 2012) and merged in to each LSOA. Separately, a shapefile was obtained from (OpenStreetMapData, 2016) which contains over 400,000 longitude and latitude coordinates which trace the UK coast. Finally, the distance to the nearest coast was calculated for each LSOA, in kilometres[4]. This distance to the coast variable is used to split the sample into those who live on the coast and those who do not, and the main variable of interest. While these distances are accurate (they are only limited by the geographical size of the LSOA, which are small areas), deciding the bounds of this coastal variable is more subjective. In the bluespace literature, the norm is to use 0-5 km

[4] Using the Stata command `-geonear-` (Picard, 2010) to do this. Rather than calculate all pairwise combinations and take the minimum distance for each LSOA (population weighted centroid), this command determines the nearest neighbour using a "divide and conquer" algorithm which greatly reduces the number of calculations, without a loss of accuracy.

Figure 3.2: Number of "movers" vs. distance to the coast



Notes: The data for this figure are from the sample of those who have moved at least once, either to or from the coast. I.e: they have lived on both sides of the coastal cutoff, which is allowed to vary on the x-axis.

as the coastal category (Wheeler et al., 2012; White et al., 2013). For the purposes of this chapter, however, there is a clear trade off between accurately capturing those who live on the coast, and the sample size - especially when considering only movers between the coast and inland.

This sample of movers is created by keeping only those who have lived on one side of the coastal cut-off in at least one wave, and on the other side of the cut-off in at least one other wave. Figure 3.2 shows how this sample varies across different definitions of the coast. A distance of 2.5km seems to provide a balance between accurately capturing the coast, and providing a large enough sample of movers and is thus used as the definition of the coast throughout this chapter, for both the full sample and the sample of movers. Other distances are used in robustness checks, to check the sensitivity of the results to this definition. I also employ a "donut" in the distance to the coast, where data between 2.5km and 15km from the coast is dropped. This imposes a clear distinction between the coast and inland (which is otherwise somewhat arbitrary), and ensures that the control group remain fixed when testing different definitions of the coast.

Outcome Variables

To capture self-assessed health, individuals were asked "How would you rate your health in general?" to which they responded Poor, Fair, Good, Very Good or Excellent. In these analyses, this was treated as continuous categories or dichotomised to equal one if the response was "Excellent" or "Very Good" and zero otherwise. In another question about their physical health, respondents were asked whether they had any of a number of long-term limiting illnesses. The responses from these questions were used to create a binary variable equal to one if the respondent had *any* long-term limiting health condition and zero otherwise. USoc contains data on the individuals' GHQ-12 score, which ranges from 0 to 36, and aggregates answers from a range of questions aimed at assessing their psychological well-being. This variable was reversed so that a higher score indicates "better" mental health than a lower one, for ease of interpretation. An outcome variable capturing disability claimants, where individuals identify as claiming one or more types of welfare, is constructed. Individuals were asked if they were currently receiving any health-related benefits: incapacity benefit, employment and support allowance, severe disablement allowance, carer's allowance, disability living allowance, return to work credit, attendance allowance, industrial injury disablement benefit, war disablement benefit, sickness and accident insurance, and any other disability related benefit or payment. The variable is equal to one if the individual is claiming any of these benefits, and zero otherwise.

Waves 2 and 5 of USoc contain information on health behaviours, including smoking and physical activity. Individuals were asked if they had ever smoked, which is included as a dummy variable equal to one if they have and zero otherwise, in our analysis. For physical activity, the respondents were asked to choose which sports and activities they have undertaken. They are subsequently asked how often (from a list of pre-defined "moderate-intensity sports") they have done this in the last 12 months. Individuals responded: no sports were done in the past 12 months; once in the past 12 months; twice in the past 12 months; less often than once a month but at least three or four times a year; less often than once a week but at least once a month; at least once a week but less than three times; three or more times a week. This was coded as a binary variable equal to one if the individual had participated at least once a week and zero otherwise.

Waves 3 and 4 of USoc contain information on drinking behaviour. Participants were asked: "Thinking back over the last week, on how many days (if any) did you have a drink? (A 'drink' is one pint/bottle/can of beer or cider, 2 alcopops, one small glass of wine, a single measure of spirits)". For our purposes this was coded to equal one if they responded 3 days or more and zero otherwise.

Control Variables

In addition, USoc provides information on the respondents age, gender, ethnicity, and employment status (including job classification). Individuals were asked to select the most appropriate marital status from "Single, never married or in a civil partnership; married; civil partner (legal); separated; legally married; divorced; widowed; separated from civil partner; a former civil partner; surviving civil partner". This was dichotomised to equal one if the respondent was either married or in a civil partnership and zero otherwise. Information on income was collected from all of the adult respondents in the household and used to construct a household net income variable, which was subsequently log-transformed. Respondents were also asked about their highest educational qualification to date, choosing one from "No qualifications; Degree; Other higher qualification; A-Level or equivalent; GCSE or equivalent; Other", from which dummy variables for each category were created. The LSOA data also allows us to control for area level characteristics. These comprise of their Index of Multiple Deprivation scores, and time (in minutes) to the nearest hospital and GP practice, if one were to travel by car (Department for Transport, 2015).

3.3.2 Descriptive Statistics

Table 3.1 reports the descriptive statistics for the full sample from waves 1-5 of USoc. Column one shows the descriptives for those living on the coast, column two for those living outside of the coastal cut-off and the third and fourth columns report the difference between the two and the corresponding p-value for the t-test of the null that the differences are zero. The upper panel shows the differences between the control variables on the coast versus inland. With the exception of A-level prevalence and Gender, those who live on the coast differ statistically significantly from those who do not across all variables. The magnitude of these differences is not, in general, economically significant. Of note, is that those on the coast tend to be several years older, less likely to hold a degree, earn (net) around £315 a month less (per household) and live just less than 1 minute further away from a hospital, by car. That households earn £3,780âĂň (£315*12) is in line with the figures from Corfe (2017). The households on the coast have a lower IMD score (i.e are more deprived) on average, a result better illustrated in Figure 3.1.

The lower panel of Table 1 shows the descriptive statistics for various outcome measures. Again, there are statistical differences between each (with the exception of GHQ score and frequent physical activity), with some more meaningful than others. Respondents living on the coast are 6 percentage-points more likely to have a long term health condition than those who do not, 3 percentage points more likely to claim disability benefits, and 9 percentage points more likely to smoke, than those who live further from the coast. In addition to this, Figures 3.3-3.5 show distributional comparisons of the outcomes, on the coast vs. inland, and Figure 3.6 shows the same for various covariates. From these, it seems that the differences between the coast and inland are in terms of mean levels, with little difference in the spread of each variable. One exception to this is the distribution of age on the coast vs inland, with a higher concentration of older population on the coast, and relatively fewer aged between 30 and 45 years, unlike the peak in these ages inland.

3.4 Methods

3.4.1 Simple Linear Model

The analysis begins with a simple linear model for a given health outcome (long-term health condition; self-assessed health; GHQ-12 score), y_i, as follows:

$$y_i = \alpha + \beta Coast_i + X_i'\gamma + \varepsilon_i, \tag{3.1}$$

Table 3.1: Descriptive Statistics

	Coast = 1	**Coast = 0**	**Difference**	**P-Value**
Observations	16,578	143,859		
Covariates:				
Age	48.59	45.93	2.66	0.00
Married	0.50	0.52	-0.03	0.00
White	0.96	0.75	0.21	0.00
(Full-time) Employment Status & Job Classification				
Not Employed (not retired)	0.21	0.24	-0.03	0.00
Retired	0.27	0.22	0.05	0.00
Large employers & higher management	0.02	0.03	-0.01	0.00
Higher professional	0.03	0.05	-0.01	0.00
Lower management & professional	0.15	0.16	-0.01	0.00
Intermediate	0.07	0.08	-0.01	0.01
Small employers & own account	0.06	0.05	0.01	0.00
Lower supervisory & technical	0.05	0.04	0.01	0.00
Semi-routine	0.10	0.09	0.01	0.00
Routine	0.05	0.05	-0.00	0.40
Highest Qualification:				
Degree	0.19	0.24	-0.05	0.00
Other Higher Qual.	0.11	0.11	0.01	0.01
A-level or equivalent	0.21	0.20	0.01	0.09
GCSE or equivalent	0.24	0.21	0.03	0.00
Other Qualification	0.11	0.10	0.01	0.00
No Qualifications	0.14	0.14	-0.01	0.03
Net HH Monthly Income	£2685.65	£3001.31	-£315.66	0.00
Male	0.47	0.46	0.00	0.40
Time to GP (mins, by car)	7.79	7.87	-0.08	0.00
Time to Hospital (mins, by car)	18.15	17.30	0.85	0.00
IMD Score	24.87	22.43	2.43	0.00
Outcomes:				
Self-Assessed Health (SAH)	2.65	2.58	0.06	0.00
GHQ Score	24.86	24.91	-0.05	0.30
Long-term Health condition	0.39	0.33	0.06	0.00
SAH = V. Good or Excellent	0.49	0.51	-0.02	0.00
Ever Smoked	0.62	0.52	0.09	0.00
Drank frequently last wk.	0.34	0.31	0.02	0.00
Frequent physical activity	0.35	0.34	0.01	0.27
Disability Benefits claimants	0.12	0.09	0.03	0.00

Figure 3.3: Distribution of various health outcomes - coast vs. inland

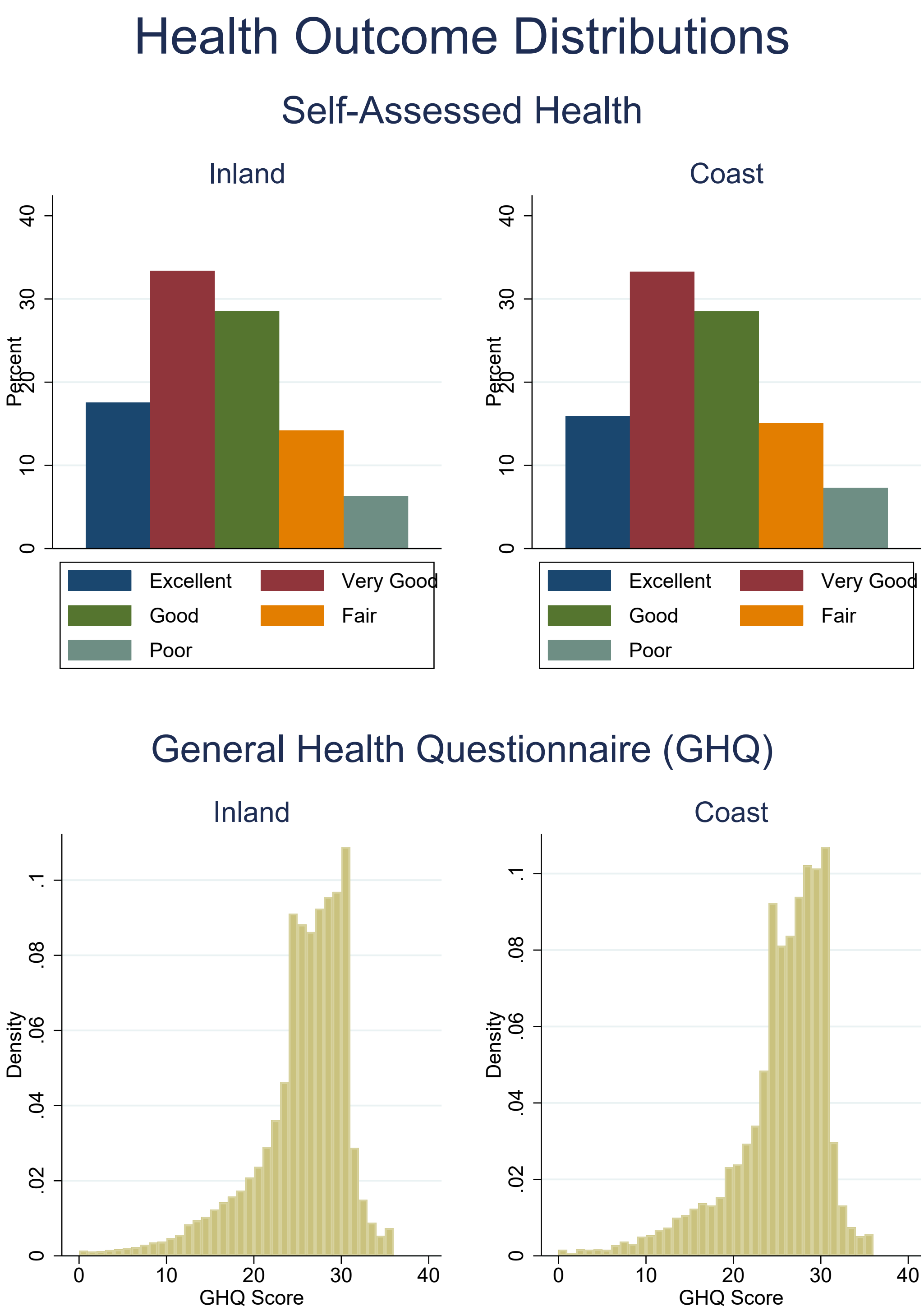


Figure 3.4: Distribution of various health outcomes continued - coast vs. inland

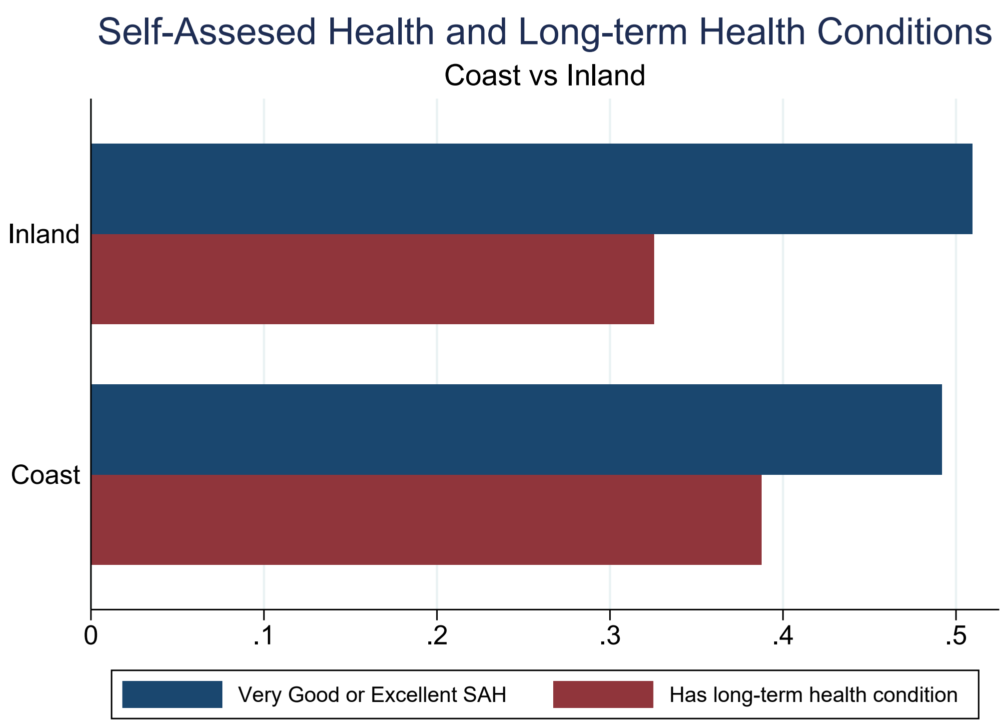


Figure 3.5: Distribution of risky health behaviours and disability claimants - coast vs. inland

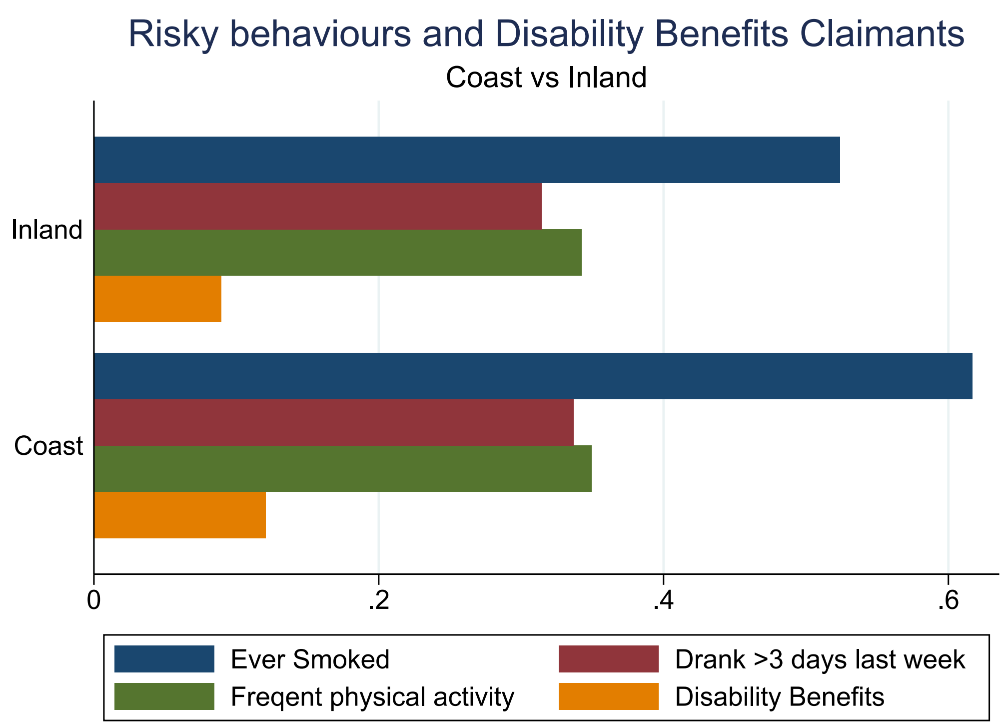
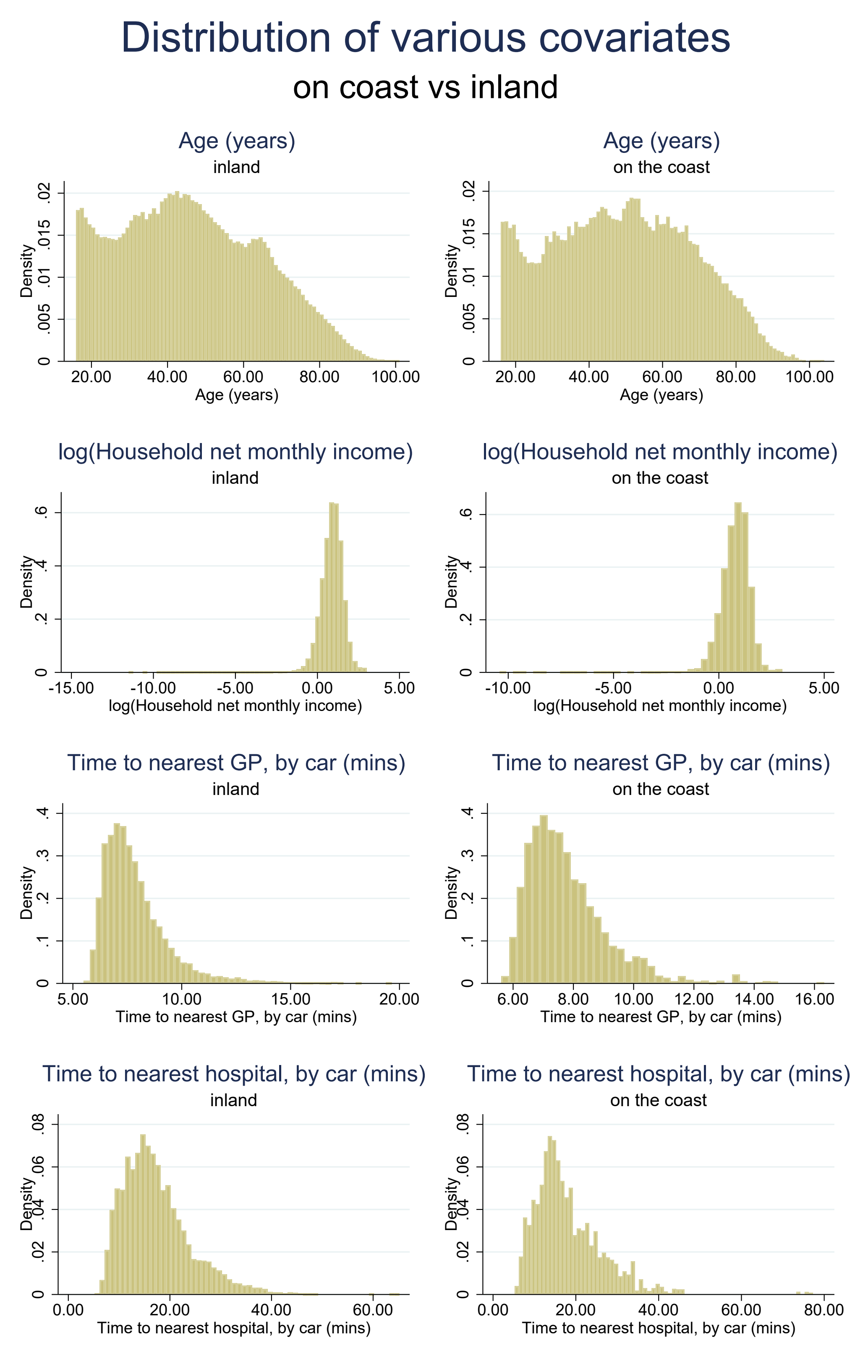


Figure 3.6: Distribution of various covariates - coast vs. inland



where $Coast_i$ is equal to one if the individual resides on the coast (within 2.5km) and zero if they live further than 15km away. In this respect, most specifications use a "donut" definition of the coast, where the mid range is not included. This allows for a clear distinction between those on the coast versus those inland, and for robustness checks, allows moving this coastal cutoff without altering the control group[5]. ε_i denoted unobserved determinants of y_i and X_i is a vector of individual, household and LSOA-level characteristics which are believed to be correlated with both y_i and $Coast_i$.

As noted in the previous section, Usoc provides a wealth of individual, household and local area (LSOA) level characteristics. The inclusion of these as controls allow the reduction of, but not elimination of, the bias of other confounding factors. Specifically, a causal interpretation of β relies on a "selection on observables" assumption: conditional on the observable characteristics contained in X_i', individuals who live on the coast do not differ systematically in terms of unobservable factors. As was mentioned in earlier sections, it is unlikely that this assumption will hold, when considering the English coast.

The data in Understanding Society are longitudinal in nature: individuals are followed through subsequent waves. For the main analysis, equation 3.1 is estimated via pooled OLS where participants are allowed to contribute variation to the analysis more than once. To account for serial correlation in the variance-covariance structure, all standard errors are clustered at the individual and household level. A natural extension of this analysis is, using only within-individual variation, including fixed effects. However, as the coast does not vary over time, this would only make use of variation generated by individuals moving between the coast across two or more waves. There are two reasons this analysis does not feature as part of the main results of this chapter. Firstly, considering the effect of moving to or from the coast on health is a different research question to the one at hand. Secondly, there are not many coastal movers relative to the rest of the sample. Due to combination of these two factors, I reserve this analysis as an extension to the main question in section 3.6.1.

[5]Robustness to this defintion is shown in section 3.6.4, and to including this data as an additional category in Appendix B, tables B.4 and B.5

Finally, to address any concern from treating each wave as a repeated cross-section I include supplementary estimation of 3.1 where I use only one response per individual (using the latest response), and where I exploit the variance structure of the full panel in a Random Effects (GLS) framework (Wooldridge, 2010). These can be found in Appendix B, tables B.6 and B.7. The results from these analyses are not explicitly discussed further as they qualitatively the same, the standard errors less conservative than those of the main analysis, and the Random Effects setup is more restrictive in its assumptions for consistency (Wooldridge, 2010).

3.4.2 Propensity Score Methods

Let the propensity score of living by the coast be defined as:

$$p(x) = Pr(Coast_i = 1|X_i). \tag{3.2}$$

Several methods make use of propensity scores in order to estimate average treatment effects under two assumptions: the aforementioned selection on observables assumption, and the "overlap" assumption which rules out the propensity score of living on the coast ever equaling zero or one (i.e that: $0 < Pr(Coast_i = 1|X_i) < 1$).

Estimating Propensity Scores

Perhaps the most commonly used method in the economics literature is specifying a parametric form and estimating the propensity scores via a logit or probit model:

$$Coast_i = 1(f(X_i')\gamma + v_i > 0), \tag{3.3}$$

where the choice of X_i is governed primarily by economic intuition, theory and previous research, and $f(.)$ is usually chosen such that the specification is linear in predictors. However, there is little guidance in the literature on how is best to specify this parametric form. Imbens (2015) and Imbens and Rubin (2015) have attempted to provide some guidance to practical applications using propensity scores suggesting that misspecification of the propensity score can be problematic, especially when estimating treatment effects. Specifically, they say that using a linear specification (in predictors) is generally not advisable; the choice of polynomial terms poses another decision for the researcher, however. A similar algorithm proposed by the aforementioned authors is used, in order to select the predictors, X_i that allows the researcher to ensure variables of theoretical importance can be included, as well as data-driven choices (see Imbens (2015) and Imbens and Rubin (2015) for a complete description of the method).

The aim of the algorithm is to choose X_c, X_l and X_q, from the set of all potential predictors X. X_l are the linear predictors, and X_q are second-order terms (i.e including squared and interaction terms), as chosen by the algorithm. X_c are the linear terms chosen by the researcher, and in this case include only income and age[6], whilst X_l and X_q are empty sets upon the first iteration. A stepwise approach is then taken, adding in a predictor, x_l from the set $[X - X_c - X_l]$ and estimating a logit model:

$$Coast_i = F(X_c'\gamma + X_l'\beta + x_l\phi + v_i) \tag{3.4}$$

and calculating the likelihood ratio (LR) statistic against the restricted model that only includes X_c and X_l (which in the first iteration is empty). This process is repeated for each covariate individually, and if the largest of these LR statistics is less than one[7], then the algorithm moves on to the quadratic selection phase. If the LR statistic is greater than or equal to one, then the corresponding predictor, x_l is added to X_l, and this step is repeated again for the ever-smaller set $[X - X_c - X_l]$, until either all of the covariates have been included, or the LR statistic is less than one.

The next phase of the algorithm is performed after X_c and X_l have been selected, and chooses a subset of the second-order terms to be included in X_q. Only second-order terms for covariates which have been included linearly are considered. The process is identical to the linear phase, including one second-order term at a time, calculating the LR statistic after each iteration. The largest LR statistic is considered at the end of each instance of the phase, including the second-order term in X_q and repeating the step while the largest LR statistic after each iteration is greater than or equal to 2.71 [8].

[6]There is little to no theory on what characteristics should predict living on the coast; an advantage of this algorithm is that X_c can be chosen to be empty, and allowing the data to choose all of the predictors.

[7]The choice of this cut-off must be made by the researcher; the value of one is taken from Imbens (2015)

[8]Again, Imbens (2015) is followed in choosing the value for this cut-off.

After the algorithm has chosen the first and second-order terms to include, the propensity scores are then estimated via maximum likelihood estimation of the corresponding logit model.

Propensity Score Matching

Using the previously estimated propensity scores, those living close to the coast are matched to those living further away, if they have similar propensity scores. Specifically, a nearest-neighbour approach is used, with a caliper of 0.1[9].

The Average Treatment Effects (ATE) for each outcome are calculated as the sum of the differences in outcomes between the matched individuals on the coast and not on the coast.

Inverse Probability Weighted Regression Adjustment (IPWRA)

The second method that makes use of the estimated propensity scores, $p(x)$, is IPWRA. This approach has the attractive property of being "doubly robust" (Wooldridge, 2010), which has the implication that only one of the outcome model or the treatment model has to be correctly specified, not both. In this approach, the full OLS specification is taken as the conditional outcome model, and a weighted version of the following is estimated, using the reciprocal of the estimated propensity scores as weights:

$$(1/p(x))y_i = (1/p(x))(\alpha + \beta Coast_i + X_i'\gamma + \varepsilon_i) \tag{3.5}$$

A nuanced difference (other than the weights) between this approach and OLS, is that regression adjustment calculates the ATE as the sum of the matched differences, as opposed to the difference in average outcomes.

3.4.3 Coarsened Exact Matching (CEM)

For the sake of robustness to the estimation method, comparisons are made with a method that does not rely on propensity score estimation, and is much closer to exact matching. CEM (as outlined in (Iacus et al., 2012)) is an exact matching algorithm ap-

[9]This is implemented in the Stata package `-teffects psmatch-`, and the standard errors are generated under the assumption that the matched data are independently and identically distributed.

plied to a set of data matched on "coarsened" variables chosen by the researcher. If these variables are continuous, they are divided into discrete "coarsened" intervals, while if they are categorical, they are regrouped into fewer "coarsened" categories. Individuals are matched with those who have identical values of the coarsened variables and the treatment effect is calculated as per exact matching.

3.4.4 Relative Correlation Restrictions

In order to assess the potential influence of selection on unobservables, the methods first used by Altonji et al. (2005) and recently developed by Krauth (2016) and Oster (2019) are applied to the data. Referring to Equation (3.1), note that we are only interested in a potential causal interpretation of β; we are not interested in the causal effects of the variables in X, only that the individual variation is controlled for. In other words, they are treated as exogenous:

$$corr(X, \varepsilon) = 0. \tag{3.6}$$

However, as mentioned earlier, living in close proximity to the coast is endogenous:

$$corr(Coast, \varepsilon) \neq 0. \tag{3.7}$$

Thus, in the absence of an instrument for the coast, one must either assume it is exogenous and estimate its effects using the methods above, or accept that the effect of interest is not identified. The methods in Krauth (2016) and Oster (2019) offer a "middle-ground" between these two options, by defining a relative correlation parameter, λ, that satisfies:

$$corr(Coast, \varepsilon) = \lambda corr(Coast, X'\gamma), \tag{3.8}$$

therefore providing a description between the correlation of the coast and unobservables, and the correlation between the coast and the observables, weighted by their statistical relationship with the outcome variable[10].

It is useful to note that $\lambda = 0$ implies the assumption that the coast is exogenous with respect to the outcome variable, and thus sufficient for point estimation and consistent

[10]The choice of using these weights follows from Krauth (2016), owing to several useful properties. A different index would simply mean a different value of λ, and is only a matter of convenience.

estimation by OLS. We can consider a weaker relative correlation restriction:

$$\lambda^l \leq \lambda \leq \lambda^h, \tag{3.9}$$

for some specified λ^l and λ^h. This allows the estimation of bounds around β from (3.1) and the construction of confidence intervals. Additionally, the λ can be used as a sensitivity parameter that enables us to find its smallest value that pushes either the bounds or confidence intervals to include zero. In other words this involves estimating the size of the correlation between living on the coast and the unobservables, relative to the correlation between the coast and observables, that implies no effect (that zero lies within the bounds).

3.5 Results

3.5.1 Ordinary Least Squares

The main results of the effect of living on the coast are now presented, firstly from OLS models and followed by the various matching methods. Living on the coast is described in section 3.3.1 above, and all models are run twice: once using all available data, and the other using a "donut" where data between 2.5km and 15km is dropped (as opposed to an additional category). Doing so represents a more conservative approach, allows the definition of the coast to be altered without affecting the control group (for robustness checks), and doesn't affect the results by much (see tables B.4 and B.5). For these reasons, only the results from these models are presented. In what follows the results are displayed as two distinct sets: one for direct health outcomes, and one for risky health behaviours and disability benefit claims.

Health Outcomes

Table 3.2 shows the coefficient on the coast variable, for each health outcome variable, for each OLS specification employed. The full specifications for each of these models can be found in the Appendix, in Tables B.1 and B.2. For the most part, residing within the coastal threshold is correlated with worse health-related outcomes. The unadjusted difference in means (leftmost column of Table 3.2) shows that living on the coast is associated with reporting worse self-assessed health: 0.063 scale points higher ($p<0.01$)

Table 3.2: Coefficients on the coastal variable: Health Outcomes

Self-Assessed Health (SAH)	0.063***	-0.007	-0.007
	(0.016)	(0.016)	(0.016)
N	109666	109666	109666
GHQ Score	-0.051	0.112	0.113
	(0.077)	(0.080)	(0.080)
N	109666	109666	109666
L/Term Health Problem	0.062***	0.016**	0.016**
	(0.007)	(0.007)	(0.007)
N	109666	109666	109666
SAH: V. Good or Excellent	-0.018***	0.005	0.005
	(0.006)	(0.007)	(0.007)
N	109666	109666	109666
Controls		✓	✓
Wave Dummies			✓

Notes: Point estimates of the coefficients from the Coast variable from OLS models are reported. Standard Errors are in parentheses. The coast is defined using a donut, with households which reside within 2.5km of the coastline (as measured from the population-weighted centroid of the LSOA) =1, whilst those greater than 15km =0. Statistical significance is denoted by: * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

(higher is worse) and 1.8 percentage-points less likely to report very good or excellent health ($p<0.01$). Likewise, individuals are 6.2 percentage-points more likely to have a long-term health condition if they live on the coast ($p<0.01$). The GHQ score also shows a negative effect, but is not significantly different from zero. As controls and wave dummies are added, however there is a sign switch for all outcomes except for long term health conditions, which are persistently worse on the coast albeit to a lesser extent (1.6 percentage points, $p<0.01$). All outcomes which change sign are no longer significant at the 10 percent level.

Risky health behaviours and disability benefit claimants

Table 3.3 shows the coast coefficients from models with dependent variables that capture risky health behaviours, and disability claimants.

In terms of health behaviours the results are similar, with those living on the coast more

Table 3.3: Coefficients on the coastal variable: Health Behaviours and Benefits

Ever Smoked	0.093***	0.026***	0.026***
	(0.008)	(0.009)	(0.009)
N	38821	38821	38821
Drank≥3 days last week	0.022***	-0.001	-0.001
	(0.008)	(0.009)	(0.009)
N	38821	38821	38821
Freqent physical activity	0.007	0.028***	0.028***
	(0.007)	(0.008)	(0.008)
N	38821	38821	38821
Disability Benefits	0.031***	0.012**	0.012**
	(0.005)	(0.005)	(0.005)
N 109666	109666	109666	
Controls		✓	✓
Wave Dummies			✓

Notes: Point estimates of the coefficients from the Coast variable from OLS models are reported. Standard Errors are in parentheses. The coast is defined using a donut, with households which reside within 2.5km of the coastline (as measured from the population-weighted centroid of the LSOA) =1, whilst those greater than 15km =0. Statistical significance is denoted by: * p<0.1; ** p<0.05; *** p<0.01.

likely to have ever smoked by 2.6 percentage points (p<0.01) and are associated with being more likely to claim disability benefits by 1.2 percentage points (p<0.05). However, those in a coastal proximity are also 2.8 percentage points more likely to participate in moderate-intensity sports and activities (p<0.01). These results, with the exception of physical activity, are in-line with the health outcomes results and generally suggest that health and health-related outcomes are worse on the English coast than inland.

It is worth noting that, so far, the sample is allowed to vary for each specification. This is due to the sample size for some outcomes being much lower - mostly due to the questions not being asked in certain waves. Tables B.3, in the appendix, reports all of these coefficients from models which are conditional on non-missingness across all covariates and outcomes. The resulting estimation sample size is 32,520 and the results are very similar, with the exception of some loss of statistical power for the dichotomised self-assessed health variable and the alcohol consumption variable.

Overall, for the health and unhealthy behaviour outcomes here, there are several takeaways. The main result is that those who live on the coast are more likely to have a long-term health condition than those who do not. This suggests that there is greater need for care in these coastal areas that has not been shown previously in the literature. Policy-wise, this has implications for the estimation of the resource allocation formula, and suggests that coastal areas may warrant higher endowments than otherwise similar areas inland. Another result from this analysis is that the self-reported measures do not reflect this higher prevalence of ill-health. This is suggestive that relative, self-assessed measures cannot always be relied upon to uncover health differences. Alternatively, it may reflect that these are more contemporaneous measures of health (i.e. relative to the last 12 months), whereas chronic health condition prevalence is, by definition, long-term. Finally, the result that individuals on the coast take part in more frequent physical activity is a finding that is in line with the previous literature on the coast, and suggests that the coast itself may serve a health-promoting asset. This is likely offset by the prevalence of smoking behaviour however. A targeted smoking cessation policy would likely improve coastal public health, and may be more successful than in other areas if the physical activity result signals a higher propensity of health investment, amongst coastal populations.

3.5.2 Robustness of results to choice of functional form

The main results are restricted to a parametric estimation that relies on normality of the error term. Table 3.4 shows the robustness of the OLS results to different estimation techniques, including both semi- and non-parametric methods. All specifications of the outcome model, where appropriate, use the same covariates as the OLS models. Essentially, this table shows that once appropriate covariates are included, the results produced by linear regression are robust to different estimation methods. In short: results which are insignificant to begin with, stay so, while statistically significant results stay of a similar magnitude and significance level.

Table 3.4: Coast coefficients, by outcome model and estimation method

Outcome: Variable	SAH	GHQ	Health	Good Health	Smoking	Drinking	Physical Activity	Disability Benefits
OLS	-0.007	0.113	0.016**	0.005	0.026***	-0.001	0.028***	0.012**
	(0.016)	(0.080)	(0.007)	(0.007)	(0.009)	(0.009)	(0.008)	(0.005)
PSMatch	-0.005	-0.011	0.028***	0.002	0.056***	0.012	0.042***	0.014***
	(0.013)	(0.066)	(0.006)	(0.006)	(0.009)	(0.009)	(0.008)	(0.004)
CEM	-0.028*	-0.001	0.026***	0.017***	0.076***	0.006	0.034***	0.016***
	(0.015)	(0.077)	(0.006)	(0.006)	(0.008)	(0.007)	(0.007)	(0.004)
IPWRA	-0.008	0.012	0.029***	0.006	0.049***	0.001	0.033***	0.012***
	(0.010)	(0.052)	(0.005)	(0.004)	(0.009)	(0.008)	(0.007)	(0.004)

Notes: For outcome models, the same covariates are included as in the standard OLS models. For the propensity score specifications, the same linear terms are used as the OLS models, with the following non-linear and interaction terms: ln(income)^2; ln(income):- $*$Education, $*$Wave, $*$Age, $*$IMD Rank; Age^2; Age:- $*$Education, $*$Wave; IMD Rank:- $*$Education, $*$Wave. These covariates were chosen as per the algorithm outlined in the methods section. Standard Errors are in parentheses. The coast is defined using a donut, with households which reside within 2.5km of the coastline (as measured from the population-weighted centroid of the LSOA) =1, whilst those greater than 15km =0. Statistical significance is denoted by: * p<0.1; ** p<0.05; *** p<0.01.

3.5.3 Relative Correlation Restrictions

Table 3.5 reports the OLS point estimates again, with 95% confidence intervals in brackets. The table also shows the RCR bounds and 95% confidence intervals for different ranges of the relative correlation parameter, λ. Results for outcome measures that are not statistically significantly different from zero are not reported, as no additional information is gained.

Column 1 shows the results for long-standing health conditions. The results suggest that the positive association between living on the coast and having a long-term health condition are robust to mild correlation between living on the coast and unobserved factors, relative to the correlation between living on the coast and observable characteristics. If the correlation between the coast variable and unobservables is up to 10% of the correlation between the coast and control variables, the the bounds are [0.026, 0.032], compared to the OLS point-estimate of 0.032; these bounds are significant at the 0.1% level. Once the relative correlation is allowed to be larger, however, the bounds around the point-estimate contain zero. Specifically, if the correlation between coastal proximity and unobservables is 57% as large as the correlation with observables then the bounds contain zero.

There is a similar pattern of results for the association between living on the coast and claiming disability benefits. The OLS point-estimate is robust to moderate correlation with the unobservables relative to observables; the bounds include zero when λ takes a value of 0.679. The results for health behaviours, however, are much more robust to relaxing the relative correlation restriction. The association between living on the coast and having ever smoked is robust to the correlation between living in a coastal proximity and unobservables being up to 109% as large as the correlation with observables. This is 113%(in absolute value) for our measure of alcohol consumption and 45% for physical exercise.

The last two outcomes have a negative minimum lambda value. The interpretation of this is the same in absolute terms, but means that the correlation between the coast and unobservables must have a different sign to the correlation between the coast and observables. Finally, the implications of this additional analysis means the OLS point-estimates for physical health, disability benefit claimants and smoking behaviour can be considered upper-bounds, while the coefficients for drinking and physical activity are lower-bounds around a causal effect of living on the coast.

Table 3.5: Comparison of OLS estimates with and without relative correlation restrictions

	LT Health	Disability Benefits	Ever Smoked	Drink $\geq$3 days	Frequent Sport
OLS Point Est.	0.032***	0.019***	0.080***	0.031***	0.032***
(95% Conf. Int.)	(0.018, 0.045)	(0.008, 0.026)	(0.069, 0.105)	(0.012, 0.046)	(0.013, 0.045)
Bounds, $0 \leq \lambda \leq 0.1$	[0.026, 0.032]***	[0.015, 0.019]***	[0.079, 0.080]***	[0.031, 0.033]***	[0.032, 0.036]***
(95% Conf. Int.)	(0.014, 0.043)	(0.007, 0.025)	(0.063, 0.102)	(0.013, 0.047)	(0.015, 0.050)
Bounds, $0 \leq \lambda \leq 0.5$	[0.004, 0.032]	[0.005, 0.019]	[0.048, 0.080]***	[0.031, 0.042]***	[0.032, 0.063]***
(95% Conf. Int.)	(−0.009, 0.042)	(−0.004, 0.025)	(0.030, 0.102)	(0.014, 0.059)	(0.016, 0.078)
Bounds, $0 \leq \lambda \leq 1$	[−0.024, 0.032]	[−0.008, 0.019]	[0.007, 0.080]	[0.031, 0.056]***	[0.032, 0.098]**
(95% Conf. Int.)	(−0.042, 0.042)	(−0.019, 0.025)	(−0.018, 0.102)	(0.015, 0.077)	(0.016, 0.119)
Minimum λ where $0 \in$ bounds	0.569	0.679	1.089	−1.126	−0.447

Notes: Intervals in square brackets are the bounds themselves, while (individual) cluster-robust 95% confidence intervals appear in rounded brackets; * $p<0.05$; ** $p<0.01$; *** $p<0.001$; standard errors calculated as in Imbens and Manski (2004); calculation of bounds follows Krauth (2016) and Oster (2016).

These results suggest that the conditional differences in health behaviours on the coast are not entirely driven by selection on unobservables. The selection on unobservables is found to have to be proportionally larger than the selection on observables, to push these effects to zero. It is unlikely that the correlation between the coast and these unobservables is larger than it's correlation with the control variables, given the rich set of individual, household and area level controls included in this analysis. Ultimately, this analysis supports the claim that the main results found in 3.5.1 do not reflect serious omitted variables bias.

3.6 Further robustness checks and potential mechanisms

The following section assesses several different mechanisms around living on the coast and how this may affect health and health-related outcomes. Movements to and from the coast are considered first, before moving on to investigate the role of house prices, access to care providers and regional heterogeneity in the effects of living on the coast. The section concludes with analysis that alters the definition of the coast to test the sensitivity of the coefficients.

3.6.1 Moving to and from the coast

This section focuses on the role of movers across the coastal threshold. One explanation for the main results - and a contributor to the selection on unobservables, is selective migration. Those who are capable and have the means to move are likely to be in better health. Understanding Society follows individuals over time and so permits the use of fixed-effects models to sweep away unobserved individual heterogeneity in health outcomes; doing so would help shed light on this issue. Doing so, however, changes the nature of the research question analysed: only individuals who have been observed on both sides of the coastal cutoff contribute variation to the model. For this reason, and due to the fact that this is a separate yet relevant question, these results are presented distinctly from the earlier analysis[11]. This section documents these effects, starting with fixed effects estimation of the original OLS specifications before moving on to those who move to the coast and those who move away. Finally it considers those who do not move,

[11]Using a fixed effects specification is also the primary identification strategy of the previous literature that considers health on the coast.

resulting in comparisons of those who reside on the coast to those who live off the coast in all the waves in which they are observed.

Table 3.6 displays the results of the fixed effects analysis in column 1. There are some sign changes: disability benefits claimants, having a long-term health problem and reporting very good or excellent health. However, all estimates from these models are not statistically different from zero with the exception of drinking on three or more days in the last week which is 12 percentage-points more likely for movers who live on the coast ($p<0.1$). The standard errors from this model are all much larger than the OLS estimates: a consequence of relying on a small amount of variation in living on the coast over time.

Columns 2 and 3 of Table 3.6 shows the effects of those who move to the coast. This variable takes a value of one if, compared to the previous period, the individual moved across the 2.5km threshold towards the coast. Whilst taking a value of zero if the individual remains off the coast in both periods. In the samples with the coastal donut in place, there are large statistically significant effects of moving to the coast on self-assessed health (-0.199 scale points, $p<0.1$) and GHQ score (1.43 points, $p<0.1$), whilst the other estimates are either insignificant or drop out of the models altogether. This is due to the sample restrictions: firstly, using fixed effects forces out anyone who is observed in one wave only; secondly, it only features those on the coast who moved there from the previous period; lastly, the coastal donut means that on top of a move having taken place it has to have been greater than 12.5km, and in the right direction, to be considered in the treatment group. Table 3.7 shows why this analysis that relies on having moved across the threshold is potentially problematic - only a tiny proportion of the sample in a given wave move to or from the coast as it is defined here. For this reason, the donut is removed in column 3 to allow more variation in the coast variable. This results in more precise estimation and the estimates no longer drop out. The magnitudes of the self-assessed coefficients halves in size, but the interpretation is the same: moving to the coast vs remaining off it results in higher self-assessed and psychosocial health. Even with this inclusion of those residing between 2.5 and 15km from the coastline, however, there is not enough variation in those who move for these results to

Table 3.6: Investigating the role of movers: fixed effects estimates, moving to and from the coast, and those who do not move

	Fixed Effects	Moving to Coast		Moving away from Coast		No Movers
Self-Assessed Health (SAH)	-0.039	-0.199*	-0.108	-0.127	-0.007	-0.003
	(0.059)	(0.112)	(0.068)	(0.124)	(0.066)	(0.018)
N	85849	50114	91312	6078	9177	86045
GHQ Score	0.069	1.433*	0.555	3.563***	0.090	0.027
	(0.459)	(0.803)	(0.443)	(0.946)	(0.481)	(0.092)
N	75808	44850	80794	5661	8468	76000
L/Term Health Problem	-0.015	0.004	0.015	0.010	0.024	0.018**
	(0.033)	(0.059)	(0.035)	(0.075)	(0.039)	(0.008)
N	85826	50100	91297	6074	9172	86022
SAH: V. Good or Excellent	-0.038	0.034	0.008	0.090	0.037	0.003
	(0.035)	(0.078)	(0.041)	(0.066)	(0.032)	(0.008)
N	85849	50114	91312	6078	9177	86045
Ever Smoked	-0.003	-	-0.008	-	0.002	0.026***
	(0.037)	-	(0.038)	-	(0.058)	(0.009)
N	44327	16117	47182	1963	4699	44427
Drank$\geq$3 days last week	0.120*	-	0.068	-	-0.125	-0.006
	(0.071)	-	(0.077)	-	(0.099)	(0.009)
N	35766	16116	38080	1963	3856	35857
Freqent physical activity	0.011	-	-0.087	-	-0.043	0.029***
	(0.055)	-	(0.087)	-	(0.075)	(0.008)
N	44138	16063	46992	1961	4685	44238
Disability Benefits	-0.023	0.056	0.027	0.047	0.025	0.014**
	(0.021)	(0.035)	(0.017)	(0.050)	(0.019)	(0.005)
N	85548	49922	91006	6055	9140	85744
Fixed Effects?	✓	✓	✓	✓	✓	-
Donut?	✓	✓	-	✓	-	✓

Notes: Col 1: Point estimates of the coefficients from the Coast variable from Fixed Effects models are reported, using the same specification as the OLS models (time-fixed variables drop out by definition namely gender). *Cols 2-4:* show the point estimates for moving to the coast (vs remaining off-coast) and moving from the coast (vs staying on). These are shown from samples both with and without the donut. *Col 5:* Estimates of living on the coast, using OLS on only the sample of households who do not move across the coastal threshold. Standard Errors are in parentheses. The coast is defined using a donut, with households which reside within 2.5km of the coastline (as measured from the population-weighted centroid of the LSOA) =1, whilst those greater than 15km =0. Statistical significance is denoted by: * p<0.1; ** p<0.05; *** p<0.01.

Table 3.7: The problem with analysing movers: Coastal transition probabilities for full sample

Panel A: Transition probabilities using the coast donut		
Time t:	Non-Coast	Coast
Time t-1:		
Non-Coast	99.86	0.14
Coast	0.83	99.17
Panel B: Transition probabilities without using the coast donut		
Time t:	Non-Coast	Coast
Time t-1:		
Non-Coast	99.75	0.25
Coast	2.44	97.56

be reliable. Columns 4 and 5 show similarly insignificant results - only the increase in probability of reporting very good or excellent health is statistically different from zero at the 10% level.

The final column of Table 3.6 shows the results from a different approach to assessing the importance of those who move as opposed to those who do not: running the same OLS specification as earlier, on a sample of those who do not move across the coastal threshold at all. The results are extremely similar to those in column 3 of Tables 3.2 and 3.3 - both in terms of the magnitude and the pattern of statistical significance across all outcomes. This suggests that movers do not contribute much to these main results. Again, this is unsurprising given the lack of variation (see Table 3.7) in living on the coast over time. However these results highlight that, at least in the short-run, selective migration is not a major concern for the original OLS estimates.

3.6.2 House Prices and access to healthcare

This section investigates two potential mechanisms that may drive health differences on the coast: house prices and access to health care providers. It does so making use of the LSOA-level linkage that the special license of Understanding Society provides access to. The UK coastline is highly heterogeneous - and where a household is based, conditional on being on the coast, is likely to play a role. Experiencing the benefits of a coastal view likely demands a premium on the housing market, for example. LSOA-level average house prices are included in the analysis to evaluate the effect of living on the coast, conditional on the environment the household is based in, as proxied by prices.[12] Access to healthcare providers, namely GPs and Hospitals, may be limited on the coast due to its isolated nature. Data is available on the time by car, in minutes, to the nearest GP and the nearest hospital at the LSOA level. This analysis exploits this information to partial out the effect of provider isolation from the effect of the coast on the set of health-related outcomes.

Table 3.8 shows the coefficients for the coast, log house prices, time to nearest GP and time to nearest hospital in minutes. For all of the health outcomes the coast coefficients do not change in a meaningful way, after these controls are added, decreasing slightly for having a long-term health condition.

Living in an area with higher house prices is positively associated with better health outcomes in each specification, and statistically significant at the 1% confidence level. Access to health care leaves more of a mixed picture: GP access is associated with better self-assessed health (for the binary outcome), but worse psychosocial and physical health. Access to a hospital is associated with worse self-assessed health in the categorical outcome, better SAH in the binary model, and being more likely to have a health problem. The coefficients are persistently smaller for access to a hospital versus access to a GP. For example, living 10 minutes away from a hospital[13] is associated with being 1% more likely to have a health condition, as opposed to being 3% more likely if living 10 minutes away from a GP. This is likely explained by the distance to a GP being a better proxy for geographical isolation, as they are much more densely distributed than hospitals.

Table 3.9 shows the same coefficients for the other outcomes. Including house prices and access to health care variables has little effect on the coast coefficients for these outcomes. Areas with higher house prices are associated with being less likely to claim

[12]Data was collected from the Land Registry. Every year was used, resulting in a postcode-level house price dataset, based on sales, from 1995-2018. This was subsequently aggregated up to the LSOA level, and merged in to Understanding Society based on the LSOA and interview year.

[13]The average response time target for an ambulance is 7 minutes for an urgent call-out in the UK.

Table 3.8: Estimates of the effect of house prices and access to health care providers

Panel A: Dependent variable: Self-Assessed Health (SAH)					
Coast	-0.007	-0.013	-0.008	-0.006	-0.013
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
ln(House Price)		-0.135***			-0.134***
		(0.010)			(0.010)
Time to GP (mins)			-0.008**		-0.004
			(0.003)		(0.004)
Time to Hosp (mins)				-0.002**	-0.001
				(0.001)	(0.001)
N	127479	127479	127479	127479	127479
Panel A: Dependent variable: GHQ Score					
Coast	0.115	0.123	0.114	0.112	0.119
	(0.081)	(0.081)	(0.081)	(0.081)	(0.081)
ln(House Price)		0.228***			0.230***
		(0.055)			(0.055)
Time to GP (mins)			-0.006		-0.016
			(0.017)		(0.017)
Time to Hosp (mins)				0.004	0.004
				(0.004)	(0.004)
N	109498	109498	109498	109498	109498
Panel A: Dependent variable: L/Term Health Problem					
Coast	0.016**	0.014**	0.017**	0.016**	0.014**
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
ln(House Price)		-0.043***			-0.044***
		(0.004)			(0.004)
Time to GP (mins)			0.003**		0.003**
			(0.001)		(0.001)
Time to Hosp (mins)				0.001*	0.000
				(0.000)	(0.000)
N	127442	127442	127442	127442	127442
Panel A: Dependent variable: SAH: V. Good or Excellent					
Coast	0.005	0.008	0.006	0.005	0.007
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
ln(House Price)		0.047***			0.046***
		(0.004)			(0.004)
Time to GP (mins)			0.003**		0.002
			(0.001)		(0.002)
Time to Hosp (mins)				0.001*	0.000
				(0.000)	(0.000)
N	127479	127479	127479	127479	127479

Notes: Point estimates of the coefficients from the Coast, house price and access variables from OLS models are reported. Standard Errors are in parentheses. The coast is defined using a donut, with households which reside within 2.5km of the coastline (as measured from the population-weighted centroid of the LSOA) =1, whilst those greater than 15km =0. Statistical significance is denoted by: * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

Table 3.9: The effect of house prices and access to health care providers continued

Panel A: Dependent variable: Ever Smoked					
Coast	0.026***	0.028***	0.026***	0.026***	0.027***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
ln(House Price)		0.031***			0.031***
		(0.006)			(0.006)
Time to GP (mins)			0.000		-0.002
			(0.002)		(0.002)
Time to Hosp (mins)				0.001**	0.001**
				(0.000)	(0.000)
N	48043	48043	48043	48043	48043
Panel A: Dependent variable: Drank≥3 days last week					
Coast	-0.001	0.001	0.000	-0.002	0.001
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
ln(House Price)		0.032***			0.031***
		(0.005)			(0.005)
Time to GP (mins)			0.009***		0.007***
			(0.002)		(0.002)
Time to Hosp (mins)				0.002***	0.001***
				(0.000)	(0.000)
N	38790	38790	38790	38790	38790
Panel A: Dependent variable: Freqent physical activity					
Coast	0.028***	0.030***	0.028***	0.028***	0.029***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
ln(House Price)		0.031***			0.032***
		(0.005)			(0.005)
Time to GP (mins)			0.000		-0.001
			(0.002)		(0.002)
Time to Hosp (mins)				0.001*	0.001*
				(0.000)	(0.000)
N	47843	47843	47843	47843	47843
Panel A: Dependent variable: Disability Benefits					
Coast	0.012**	0.010**	0.012**	0.012**	0.010**
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
ln(House Price)		-0.030***			-0.030***
		(0.003)			(0.003)
Time to GP (mins)			-0.001		-0.001
			(0.001)		(0.001)
Time to Hosp (mins)				0.000	0.000
				(0.000)	(0.000)
N	126080	126080	126080	126080	126080

Notes: Point estimates of the coefficients from the Coast, house price and access variables from OLS models are reported. Standard Errors are in parentheses. The coast is defined using a donut, with households which reside within 2.5km of the coastline (as measured from the population-weighted centroid of the LSOA) =1, whilst those greater than 15km =0. Statistical significance is denoted by: * p<0.1; ** p<0.05; *** p<0.01.

disability benefits, and more likely to drink, smoke and participate in frequent physical activity ($p<0.01$).

In summary, including house prices and health care access in the health outcome models do not alter the coefficients much. This suggests that, although somewhat predictive of these health related outcomes, living on the coast is associated with systematically worse health-related outcomes through some other mechanism not observed here.

3.6.3 Results by working age and retirement age

Age is likely an important feature of how living on the coast can influence health outcomes. In particular, the UK coastline is anecdotally a popular place for individuals to move to upon retirement. If this is the case then, due to the association between age and worse health outcomes, this "unhealthy migration" effect may explain some of the findings from the main analysis. Likewise, there could be a "healthy" outward migration away from the coast of younger individuals, who may move inland to the city in the hope of improving labour market outcomes. As a test for these phenomena, the main analysis is replicated and the coast dummy interacted with an indicatoir for those who are of a working age (16-64 years) versus those who are of retirement age (65+ years). Robustness of these results to using a *retirement* flag, as opposed to age, is included in Appendix B, where the results are the same, if slightly smaller in magnitude (see table B.8).

Table B.8 presents the coast coefficients for these stratified models with panel A showing those for the health outcome models. Individuals on the coast, regardless of being of retirement or working age, are more likely to have a long term health condition. For self-assessed health, however, those who live on the coast in retirement age are much more likely to report very good or excellent health (4.5pp, $p<0.01$), and more likely to report a better category[14] of health (-0.097 scale points, $p<0.01$). This result is suggestive of a different impact of living on the coast for retirees. This could be due to a slackening of a retirees time constraint, allowing more leisure time to be spent enjoying coastal features. Likewise, it may be that the self-assessed health question used in the survey is a proxy

[14] A higher value represents worse self-assessed health

Table 3.10: OLS results with working/retirement age-coast interactions

Panel A: Health Outcomes				
	SAH	GHQ Score	Health Problem	V. Good Health
Coast	-0.082**	-0.005	0.021	0.040***
	(0.037)	(0.162)	(0.015)	(0.015)
Working Age	0.178***	-1.335***	0.082***	-0.052***
	(0.022)	(0.100)	(0.009)	(0.009)
Interaction	0.097**	0.140	-0.005	-0.045***
	(0.040)	(0.182)	(0.016)	(0.016)
N	127682	109664	127645	127682
Panel B: Risky health behaviours and disability benefits				
	Smoking	Drinking	Physical Activity	Disability Benefits
Coast	0.022	0.032*	0.033**	0.020
	(0.018)	(0.018)	(0.015)	(0.013)
Working Age	0.013	0.113***	-0.020*	0.077***
	(0.012)	(0.012)	(0.010)	(0.007)
Interaction	0.005	-0.043**	-0.006	-0.010
	(0.020)	(0.020)	(0.018)	(0.013)
N	48098	38821	47898	126279

Notes: Point estimates of the coefficients from the Coast variable from OLS models, and its interaction with the retirement indicator, are reported. Working age is defined as all individuals aged 16-64, whilst retirement age is defined as 65 years and older. Cluster-robust standard errors are in parentheses. The coast is defined using a donut, with households which reside within 2.5km of the coastline (as measured from the population-weighted centroid of the LSOA) =1, whilst those greater than 15km =0. Statistical significance is denoted by: * p<0.1; ** p<0.05; *** p<0.01.

for general wellbeing, and living on the coast in retirement is positively associated with wellbeing regardless of the negative association with physical health.

Panel B of Table B.8 shows the same stratified results for the health behaviour and disability benefit claimants. Living on the coast is linked with higher smoking and physical activity regardless of being of retirement age, but those on the coast aged over 65 are more likely to drink frequently (4.3pp, p<0.05). Those who live on the coast in retirement age are no more or less likely to claim disability benefits, with the effect found in the original analysis being driven by those of a working age.

The results from this analysis hint at a slightly different mechanism for retirees who live on the coast, particularly around self-assessed health measures. Further work is needed in this area which would help to shed light on who gains and who is worse-off from living on the coast. This is particularly important for decision makers who are involved with minimising unmet need in terms of the provision of health services.

3.6.4 Robustness to the definition of the coast

This section reports the sensitivity of the original OLS results to the definition of the coastal variable. As the donut was used to remove all data between the distances of 2.5km and 15km of the coastline, changing the definition (i.e moving the 2.5km cutoff) does not affect the definition of the control group. Plots of the coefficients from the fully specified models, where coast= $1[D < Z]$ and zero if $D >$15km are found in figures 3.7 and 3.8 below. The coastal cutoff, Z, is allowed to vary from 0.5km to 15km.

A similar pattern occurs for most of these coefficient plots: defining the coastal variable as equal to one if the household resides less than or equal to 1km from the coastline does not provide enough statistical power to determine the effect of living there. Above this threshold, the coefficients jump to being statistically different from zero (the plot shows 95% confidence intervals) for models which showed statistical significance from the main analysis. Ultimately, the main results reported from previous sections are robust to the definition of the main variable of interest.

Figure 3.7: Robustness of results to definition of the coast - Health Outcomes

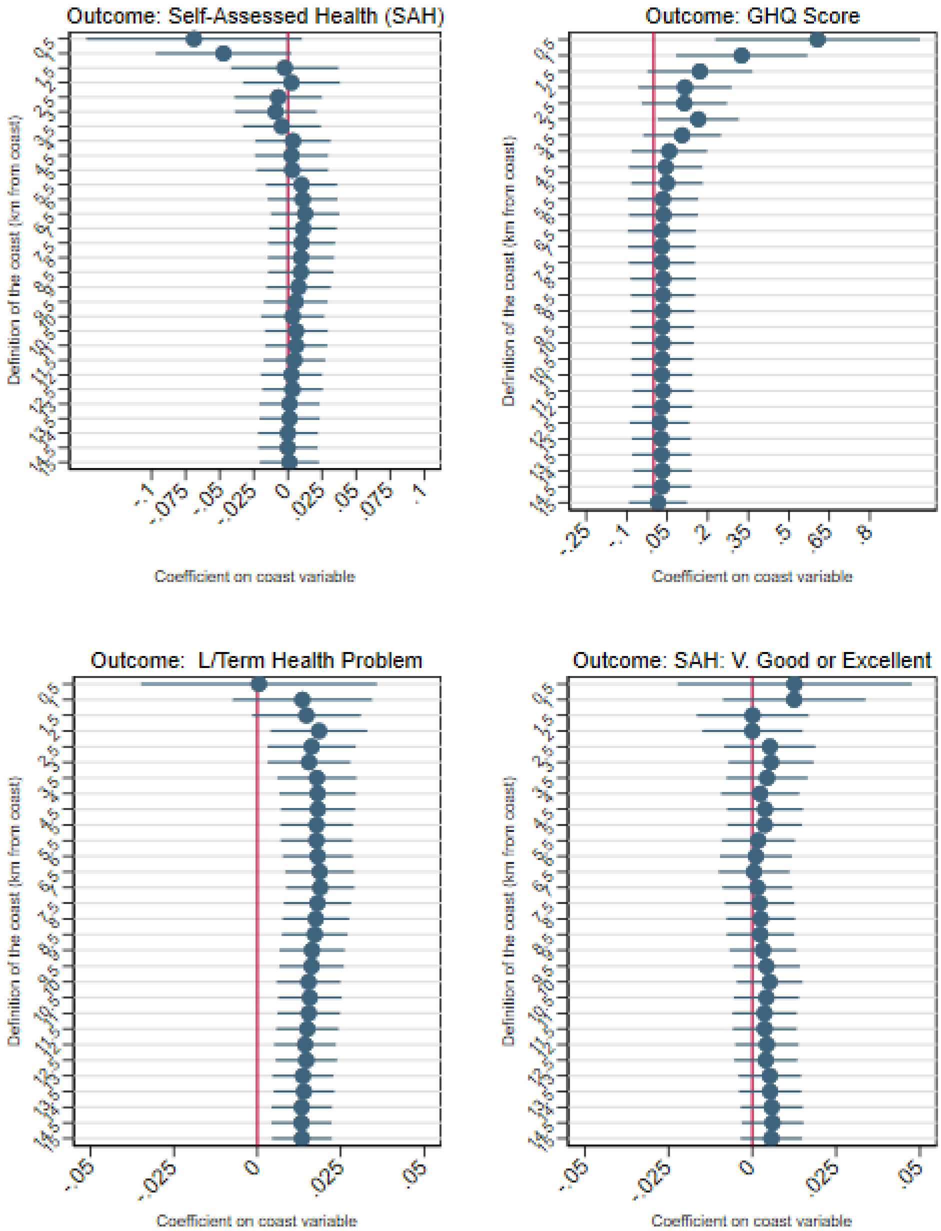
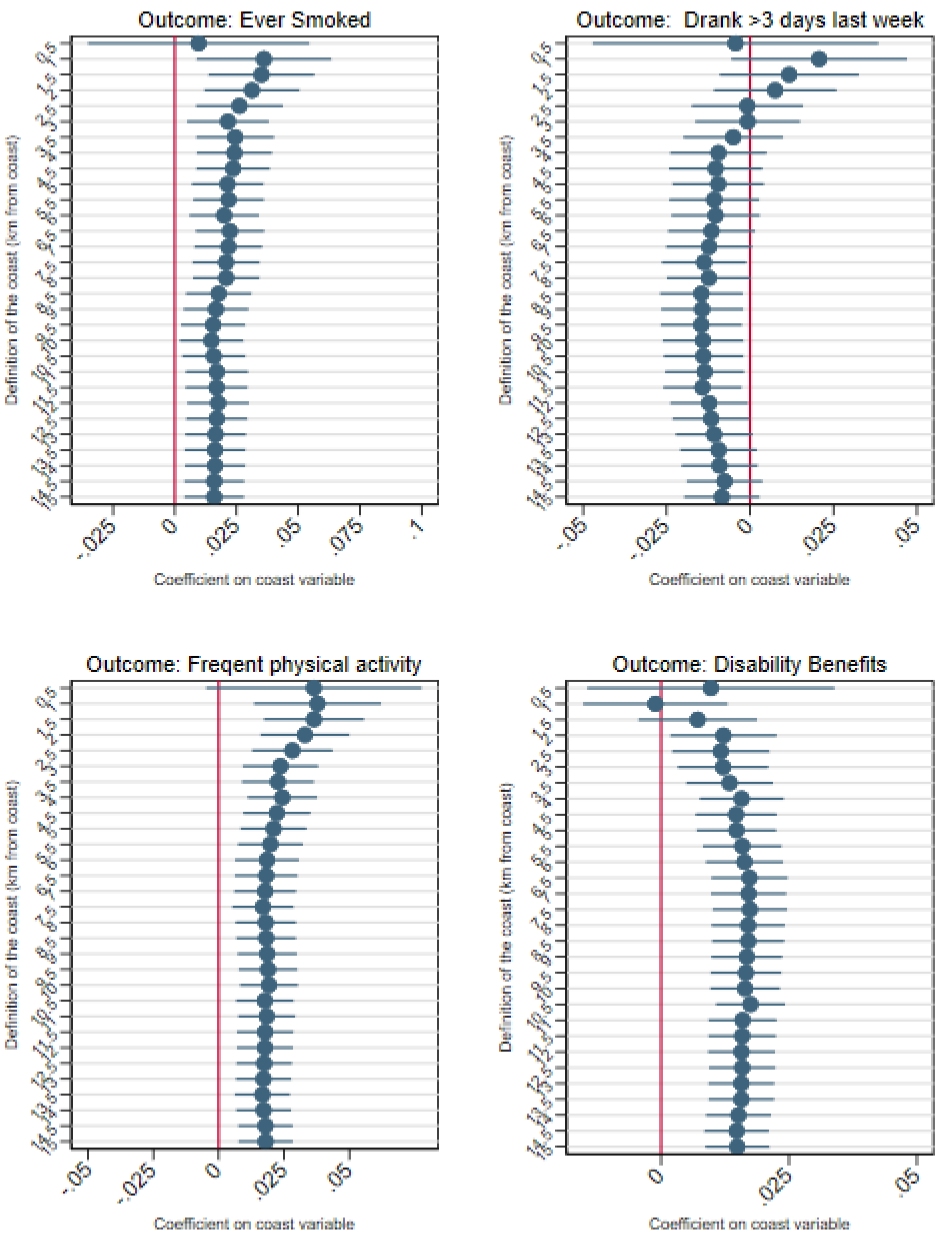


Figure 3.8: Robustness of results to definition of the coast - Health Behaviours and Disability Benefits



3.7 Discussion

This chapter has considered the differences in health and health-related outcomes between individuals living on the English coast and those living inland. The results show that many of the unconditional mean differences persist, once controlling for observable correlates of health at the individual and area level. Those differences that are explained away by the inclusion of covariates, are exclusive to health measures that are self-assessed. For the others, such as the number of disability claimants, those with a long-standing health condition and smokers and drinkers, this is an important result, as these outcomes can be considered much more objective measures of health. That health is worse on the coast is in-line with the census figures from the ONS (ONS, 2014), and suggests that the differences in health may well be driven by determinants other than socio-economic status, education and demographic factors.

However, the one notable exception here is that of frequent sport activity, which is much more prevalent on the coast than otherwise. This is an important result to highlight as much of the previous literature, which focuses on the positive effect of living on the coast on health and wellbeing, uses physical exercise as a primary reason why the coast (and more generally, bluespace) is good for health (Wheeler et al., 2012, 2015; White et al., 2013). These results support this aspect of the literature, and goes further to show that this result seems robust to selection on unobservables.

This chapter's analysis finds, in general, no significant effects of living on the coast on self-assessed physical and mental health. This is contradictory to the results in the literature considering health on the coast, which finds positive, statistically significant effects. It seems plausible that the previous results found in these studies can be considered upper bounds, owing to the fact that this work has controlled for some coastal features that have not been previously. Namely, house prices and access to health care, as proxied by the average time to nearest hospital and GP by car, are included. The careful treatment of these potential mechanisms are another contribution to the literature of this work. Another possible explanation for the null effect found in this chapter is that these are self-assessed measures. Health measures of this type may well be self-assessed, relative to those around them. The fact that there are persistent effects in terms of having a health condition, smoking and drinking support this claim.

Arguably the main contribution of this chapter - over and above using new data and outcomes to an under-researched question - is the analysis of the relative size of this selection on unobservables to selection on observables. This analysis showed that for physical health, and disability claimants, only a mild to moderate selection on unobservables relative to observables was needed to push these results to zero. It seems fairly likely that differences in physical health on the coast are driven by selective migration, or some other unobserved mechanism. The secondary analysis which stratifies by working age casts doubt on the selective migration argument, as does the analysis which considers movers and non-movers. It is important to note that this fixed effects analysis represents the main methods used in the previous literature. They do not offer comparisons as this chapter has done, including robustness to functional form, sensitivity to definitions of the coast, nor a careful inspection of the role of selection on unobservables. This is the chapter's fundamental contribution to the literature.

In terms of disability-related welfare claimants, a report published by the Communities and Local Government Comittee (2007) documents this finding (they report a difference in means) and suggests that it is *inward* migration of benefit claimants could be the main contributing factor. Thus, while we can reason the direction of healthy and unhealthy migration, further work considering these migration patterns in more detail would greatly contribute to this evidence.

Finally, the results suggest that the conditional differences in health behaviours on the coast are not solely driven by selection on unobservables. The selection on unobservables is found to have to be proportionally larger than the selection on observables, to push these effects to zero. While it is likely that these outcomes are influenced in part by selection bias, it is unlikely that the correlation between the coast and these unobservables is larger than it's correlation with the control variables. While the outcome models presented here can be considered reduced-form, there is scope for further research that considers health on the coast via structural modelling. In these structural models of health, it seems plausible that the coast can affect health through differences in health behaviours. Smoking, drinking and physical activity could, at the least, be included as covariates with the health outcomes, and perhaps used in a joint estimation of health behaviours and health outcomes.

3.7.1 Limitations

The main limitation of this study revolves around the trade off between having enough variation in the variables of interest and making use of panel data to sweep away unobserved heterogeneity. Whilst this chapter has attempted to unpick this problem in section 3.6.1, the analysis relies on cell sizes that are too small: those who moved away from and those who moved to the coast. This demonstrates the unreliability with using a fixed effects estimation strategy: there is simply too little variation in each wave to estimate reliable coefficients. A further problem is that, even with sufficient statistical power, the question (of looking solely at movers) is complicated by the fact there are two treatments relative to one baseline. Future work which has access to sufficiently data to investigate this could make further use of the IPWRA methods used in this chapter, allowing for multiple treatment effects.

Living on the coast is endogenous with respect to health, and there exists no natural experiment which would provide sharp identification of a causal effect. Other than using a richer source of data, and estimating the relative correlation restriction bounds as this chapter has done, future work could identify an instrumental variable for living on the coast that does not affect health other than through the coast variable itself. There is no guarantee that such a variable exists however, so a focus on estimating lower bounds is perhaps the most reliable strategy in terms of informing policy.

3.7.2 Conclusion

The results found in this paper, with the exception of physical activity, differ from those found in the other literature that considers coastal health. This paper has considered a wider range of outcome measures and provides a careful treatment and investigation into selection on unobservables. It is the first to consider non-random selection to the coast and has demonstrated the flexibility and applicability of the methods in Altonji et al. (2005), Krauth (2016) and Oster (2019) to a health setting, where there are often identification issues.

The results suggest that policies based on the existing literature may well overestimate the health benefits of living by the coast. It seems that an Instrumental Variables specification is the way forward in this literature, although finding a plausibly exogenous source of variation in living on the coast with respect to health, could be problematic. This is a complex issue that deserves much more quantitative-based research. Furthermore, this topic would benefit from a study that considers internal migration patterns with respect to health; there are likely policy relevant conclusions to be had from this kind of analysis.

There is considerable scope for further research in this area, and the lack of causal evidence certainly does not mean there is a lack of an effect. Descriptively, it seems that any differences in health on the coast are mitigated by other factors which affect both the probability of living on the coast and health. This does not seem to be the case when considering long-term health conditions and smoking, however, and it is important in terms of policy to investigate these differences further. Finding an exogenous source of variation in coastal living would greatly contribute to the literature, and provide potentially far-reaching policy implications.

CHAPTER 4

Does moving away from home to university affect life satisfaction? Evidence from Next Steps

Abstract

Geographical mobility, and its effect on socioeconomic status, health and wellbeing, has become an increasingly important topic for policy and research. Moving to University is often the first opportunity for such movement in the life of a young person. A recent report by the Sutton Trust showed that roughly half of students moved more than 55 miles from their home address; those that do so are from socially, ethnically and geographically distinct groups. There have been calls in the media for schools to encourage pupils to move away for university. Despite this, there is little to no empirical evidence of the impact of moving away from home to university on post graduation outcomes. This paper addresses that gap, using data on a cohort of university attendees from the Longitudinal Survey of Young People in England (LSYPE) and the follow-up study, Next Steps, to assess the impact of moving away from home on early-adult life satisfaction. A random sample of children, born in 1989/1990, were surveyed annually between the ages of 13 and 19 years old, and then again when aged 25 (Next Steps wave). Life satisfaction is modelled for graduates aged 25 years, using an ordered probit approach, controlling for individual characteristics at various points in the student's life, such as external locus of control and psychosocial health. I also partial out parental and household factors such as household income, parental education, parental occupation, and the number of siblings in the home. Preliminary results show that moving away from home to university increases life satisfaction. However, a striking gender difference persists through specifications: moving away greatly increases life satisfaction for males (between 5 and 7 percentage points more likely to report "very satisfied" $p<0.01$; between 1 and 4 percentage points less likely to report "fairly" or "very dissatisfied" $p<0.01$), but has no effect for females. The paper considers several potential mechanisms for this result, including the mediating effect of wage premiums from moving away. These results suggest that some caution should be taken by policy makers aiming to influence pupils' decision to move away from home, as there appears to be substantial heterogeneous effects, particularly by gender.

4.1 Introduction

Moving home is a significant life-event which has become an increasingly important topic for health economics and health research more generally. Moving home is directly related to age, with younger people representing the most mobile section of society. The decision to move from home to university is a significant aspect of many young people's lives and, for most, represents the first time they have moved away from their parents. At the same time a non-trivial amount of students[1] do not move away, instead choosing to remain living at home. Student mental health has become a bigger policy issue over the last decade, with the number of university students with a serious mental illness having risen significantly (Storrie et al., 2010). These concerns can be generalised to potential differences in wellbeing attributable to the student experience and related outcomes. The decision to move splits university students into two broad groups who are likely to face a completely different experience from university life (Holton, 2015). These contrasting experiences could lead to significant differences in early-adult outcomes, such as labour outcomes and wellbeing. It is the latter that takes the focus of this study.

This paper is the first to consider differences in outcomes attributable to moving away from home to university. The main focus in this thesis has been on, broadly, health outcomes. The WHO defines health as a "state of complete physical, mental and social well-being and not merely the absence of disease or infirmity" (Organization et al., 2017). Much of the focus, both in this thesis and in the health economics literature more generally, has been on physical health. This chapter turns to a different facet of health, as defined by the WHO, and focuses on a measure of mental/social wellbeing as measured by life satisfaction. A distinction is made in this chapter between wellbeing and mental health: with the latter taken as a subcomponent of a broader measure of wellbeing. An individuals' life satisfaction is taken to be a proxy for wellbeing, once mental health is partialled out. The primary research question is: are there differences in early-adult life satisfaction for those who move away to university versus those who live at home whilst studying? This chapter uses data from *Next Steps*, formerly known as the *Longitudinal Survey of Young People in England*, which follows a sample of 15,000 children from the

[1] Roughly 33% according to the data used in this paper; 55.8% remain local ($<$ 50 miles) according to Donnelly and Gamsu (2018).

1989/90 birth cohort in the UK. The individuals are first interviewed at the age of 13, every year until aged 19, and finally again at age 25. Controlling for pre- and post- move characteristics of the students, their parents and the area they live in, early-adult life satisfaction (at age 25) of students who attended university is considered, comparing those who left home to those who stayed. Life satisfaction at age 25 is modelled using linear, probit and ordered probit functional forms.

This paper also considers heterogeneous effects of specific groups, namely those of Islamic faith, those who attend Russell group universities and those who are first-generation university attendees. Subsequent analysis looks into the potential transmission mechanism through which moving away can influence early-adult life satisfaction. This mediation analysis decomposes the total effect of moving away into direct and indirect effects. The latter being through wages at age 25 and through the individual's external locus of control.

In recent years there has been an emerging literature concerning the mobility of students, their determinants, and the difference in experience between local and non-local students. Holton (2015), for example, consider differences in how local and no-local students accrue capital during their time at university, and how this can smooth the transition into adult life. Based on qualitative analysis of 31 students from the University of Portsmouth. They identify a "transformative potential" of University, but also that there is a great deal of heterogeneity in how student's experience this. Gamsu et al. (2018) use administrative data on all 412,000 students attending university in 2014-2015 combined with spatial census data. They look at how diversity in the area in which the children grew up affects where they attend university. This is one of the first studies to use quantitative methods to look at student trajectories in university attendance, and the authors conclude that students' ethnicity and University choice are key to determining whether they move to more or less diverse areas.

The closest related study to this one is a paper by Sage et al. (2013). In it, they investigate *graduate* migration of a cohort of students who left the University of Southampton between 2001 and 2007, 5 years post graduation. They find that there is a "parental safety net" offered by moving back in with parents during the 5 year post-graduation period. They make reference to the fact that stable trajectories of living at home include both those who remain at home during term time, and those who moved away and back again upon graduating. The authors do not consider the role of moving away from home to university any further, however.

This paper contributes to the literature in several ways. Firstly, it poses a novel research question. Most of the current literature focuses on determinants of moving and where students move to, and does so using qualitative methods. There remains a dearth of quantitative evidence surrounding the move to university and its consequences for the students who do (or do not). This paper represents the first to consider the longer term effects of having moved away from home, and thus opens up a new avenue for research surrounding student mobility. Secondly, this paper uses cohort data, randomly sampled from the UK. This is in contrast to the literature which has either focused on a particular university, or large but limited administrative data on university attendees. Thirdly, once establishing that differences in early-adult outcomes seem to exist, it goes on to consider and test for potential mechanisms through which these differences occur.

4.1.1 Theoretical Underpinnings

There are several mechanisms through which moving away from home to university can affect early adult wellbeing, all else equal. The direction of the net effect, however, of all of these things is not clear: there is no empirical evidence on this question. Many of the potential mechanisms are through social elements. Factors such as forming new social support networks, gaining an exposure to different social groups and structures, and living independently with peers for the first time can have potentially long-lasting positive effects over and above other features associated with going to university. Peer support networks are crucial during adolescent years (Roach, 2018); therefore a move away from home could have a positive or negative effect on wellbeing, depending on the strength of these connections.

There are some clear negative mechanisms: family ties are important for mental health

(Woodman & McArthur, 2018) - youth who are more connected to parents are less likely to report depression (Foster et al., 2017), so moving away from home may disrupt this and have a long term detrimental impact on wellbeing. Another clear mechanism is through debt: moving away from home places an additional financial burden on students - particularly those from a disadvantaged background. There are low-barrier credit markets aimed at student to combat this, but the taking on of debt can have long term consequences for mental health (Fitch et al., 2011).

There are also likely to be gender differences in these effects, particularly in the case of social ties. Males and females process and experience the costs and benefits of peer relationships during adolescence differently, with social support shown to be more strongly correlated with lower depression and higher self-esteem for males than for females (Moran & Eckenrode, 1991). It is plausible that the strong social networks that are formed by moving to university have a more beneficial effect for males than they do for females - perhaps because they may not have been formed otherwise. Men have also been shown to display higher rates of depression than women amongst those with low emotional support networks - suggesting that, conditional on peer groups being formed, they are more beneficial for wellbeing, for men (Sonnenberg et al., 2013).

4.1.2 Roadmap

The chapter proceeds as follows. Section 4.2 provides an overview of the broad literature that it contributes to. Owing to the fact that there is extremely limited evidence on student mobility and its effect on early-adult outcomes, links are made to a broader literature, and the value added over and above the novel research question. Section 4.3 describes the data used, going into detail about the construction of each variable, the nature of attrition across waves, and the survey weights used in light of this. Section 4.4 outlines the baseline analysis. This is referred to as the "Total Effect" of moving away, in order to use the language of mediation analysis which is later conducted. The results of this analysis follow in section 4.5, followed by subgroup analysis examining heterogeneous effects in section 4.6. Section 4.8 outlines the methods used in the mediation analysis, in order to try and disentangle the indirect and direct effects of moving away. The mediation results follow in section 4.9. The analysis concludes in section 4.7, in which the endogeneity problem is addressed by using the individuals' grandparents' university attendance as an instrumental variable for moving away. Section 4.10 discusses the results, limitations of the study, and makes some concluding remarks.

4.2 Literature

This section provides an overview of relevant literature. As there is no research that attempts to answer the same question as this paper, this section begins by briefly discussing the broader literature which considers the interplay between internal migration and health, concluding that there is a need to consider different age profiles in migration studies. This invokes student mobility in particular as the focus of this study. The limited literature that considers the determinants of, and differences in experience of, those who choose to move away and stay at home whilst at university is then reviewed.

4.2.1 Internal Migration and Health Selection

There has been a recent increase in studies considering the effects of moving home - internal migration - and it's association with health. Much of this literature, based on data from large countries such as the US and China, focus on rural to urban migration. Johnson and Taylor (2018) for example, examine rural to urban migration in the United States throughout the early 20th century and its effect on long-term health and longevity. They find that despite an increase in lifetime wealth, migrants are worse-off in later-life health. Chen (2011) explore rural to urban migration in China using a small household survey. They find evidence of a "healthy migrant phenomenon" on self-rated physical health.

There are several UK-based studies looking at internal migration and health which do so using the British Household Panel Survey. There is inconclusive evidence about the direction of the effect of internal migration on health in the UK. Some studies find negative health effects when focusing on moves to less deprived areas (Tunstall et al., 2014), while others find short-run temporary negative effects and effects in both directions (Nowok et al., 2013; Whittaker, 2012).

In contrast to the healthy migrant selection argument, some research identifies a “salmon bias” or selective return migration (Abraido-Lanza et al., 1999). This hypothesis posits that unhealthy migrants have a greater tendency to return home than healthier migrants. Lu and Qin (2014) tests these two hypotheses with Chinese data, finding support for both.

With respect to these two hypotheses about health selectivity and migration, they can be in part be explained by the age-profile of migrants. Younger migrants tend to be healthier and early-adulthood also represents the peak age for migration, with people relocating for education and employment opportunities (Norman et al., 2005). Despite this, there is a lack of attention paid to this age profiling in migration studies (Norman & Boyle, 2014), which in part motivates this study in focusing on student mobility.

4.2.2 Student Mobility

There is a paucity of literature related to student mobility, though there is a recent surge of interest in the topic. The most recent relevant work stems from a report on student mobility published by the Sutton Trust (Donnelly & Gamsu, 2018). The authors explore how staying at home and studying locally is strongly differentiated by socioeconomic status and ethnic background using data from the Higher Education Statistics Agency on all UK-based students (international students were omitted) attending university in 2009/10 and 2014/15. They find that the decision to stay at home or move away is a strong determinant of inequality in higher education choice and experience, and that more disadvantaged students are more likely to study from home. They recommend, amongst other suggestions, that Halal Student Loans are instated, to enable Muslim students to borrow money in accordance with their religious beliefs. The authors identify that students of Islamic faith do not face the same opportunities for mobility as their student peers.

Much of the literature surrounding student mobility is qualitative in nature. Early work on this topic was conducted by Christie (2007). The author investigates student mobility decisions and specifically the experiences of those who chose to live at home whilst attending university, along with the reasoning for doing so. For all of the students interviewed (N = 12), staying at home was an economically pragmatic decision. However, there were also more complex emotional ties with home that played a part in their decision. Many of these students faced a significant commute, which was identified as a disadvantage versus those who live at the university itself. Another study interviewed 31 students attending the University of Portsmouth in 2012 (Holton, 2015). They conclude that the university experience is transformative generally, but there are more subtle effects for those living in non-university accommodation. Ultimately, though, these students tend to form similarly influential peer groups, albeit through different mechanisms.

There is also some quantitative work considering the decision to move to university. Gamsu et al. (2018) use administrative data on all 412,000 students attending university in 2014-2015 combined with spatial census data. They look at how diversity in the area in which the children grew up affects where they attend university. This is one of the first studies to use quantitative methods to look at student trajectories in university attendance, and the authors conclude that students' ethnicity and university choice are key to determining whether they move to more or less diverse areas. Holton (2018) collect data from students attending a "post-1992" university in the South-East of England, from which they obtain 1,147 valid responses. Again, the author finds that for those students choosing to remain at home, the decision was a pragmatic one. Additionally to this, many of these students cited that living close to home was the main reason they chose their university. They conclude that there is a clear distinction in the way students choose to experience university, between traditional (including those who live at university) and non-traditional students (including those who live with their parents). Adding to the distinction between groups of students who live at home and away at university, a study on the determinants of student loan take-up (using the same data as this paper) finds that living at home whilst studying is a significant debt avoidance mechanism (de Gayardon et al., 2019). They conclude that this is problematic as it limits choice of university, and ultimately which labour market they end up in. They also draw attention to the role of religion and loan take-up, with Muslim students hugely less likely to take out student loans. As mentioned in the previous section, the closest in nature to this paper is work by Sage et al. (2013) who consider the migration decisions of graduates from the University of Southampton between 2001 and 2007. Living at home during term time is a determinant of living at home post graduation, but the authors do not consider how this can affect early-adult outcomes.

This paper contributes to these two literatures. It adds to the evidence of the wellbeing-related effects of moving home, focusing on young adults who attended university, for whom there is little evidence for. It also adds to the recent surge of literature considering student mobility. The main contribution is the novel question of how the difference in general student experience, brought about by living at university versus at home, can affect early-adult life satisfaction.

4.3 Data

The following section gives a general overview of the dataset used, including some information about attrition in the data, followed by a more detailed exposition of the variables constructed and used in the analysis itself.

4.3.1 LSYPE

The analysis is undertaken using all seven waves of Next Steps, formerly known as the Longitudinal Survey of Young People in England (LSYPE)[2]. The survey follows a cohort of 15,500 children based in English secondary schools, born in 1989/90. Starting in 2004 at the age of 14 years, the respondents are interviewed each year for six years, until the age of 20 in 2010. The follow-up wave took place when the individuals were 25 years, in 2015. The timeline of LSYPE means that data is available pre-university (waves 1-5), during university (the first two years, conditional on attendance, in waves 6 and 7), and post-university (wave 8). Following a nationally representative cohort this way allows the analysis of university-related themes, conditional on pre-university conditions, on

[2]For brevity, and to avoid confusion, throughout the remainder of this chapter these data are referred to as LSYPE.

early-adult outcomes.

The first five waves (or sweeps) of LSYPE comprised of face-to-face interviews at the individual-level, including separate interviews with the both the children and their guardian(s) living in the household at the time. This allows the models to control for variation in parental background as well as household characteristics.

LSYPE suffers from an attrition problem over the course of its eight sweeps. The initial sample size of around 15,500 is reduced to just 8,000 by the eighth wave. Due to the fact that the analysis specifically considers early-adult outcomes - those available in wave 8 - the sample is immediately restricted to those who responded in 2015. Furthermore, the analysis explicitly considers variation in wellbeing for those who attended university, again restricting the sample further to 3,155 (1,322 males and 1,833 females). Conditional on non-missing responses to the included covariates in each model, an estimation sample of around 2,283 (985 males and 1298 females) remains. LSYPE also uses a stratified sampling approach, using schools as the primary sampling unit. The study further stratified by deprivation, resulting in an oversampling of more deprived schools and students from minority ethnic groups. On top of this, there is an oversampling of those who enter into higher education, as noted by Hosein (2019); Anders et al. (2012). As a result, longitudinal sample weights from the final wave are used, which address both the potential non-response bias and LSYPE's sample design. Not doing so can result in highly-misleading point-estimates (Anders et al., 2012).[3]

4.3.2 Outcome Variables

LSYPE provides information on early-adult outcomes, at age 25, in it's eighth and final wave. As part of the self-completion questionnaire, individuals were asked "How dissatisfied or satisfied are you about the way your life has turned out so far?" and chose from the following responses: Very Satisfied; Fairly Satisfied; Neither Satisfied nor Dissatisfied; Fairly Dissatisfied; Very Dissatisfied. The categorical nature of this variable means that it takes the value of one to five for each of the above responses, respectively. This is recoded so that the variable takes a value of zero for responses of "Very Dissat-

[3] This technical report by Anders et al. (2012) contains a detailed exposition of the sampling design and the sample weights for LSYPE.

isfied" and four for responses of "Very Satisfied".

Some specifications rely on a dichotomisation of this variable so it becomes binary in nature. To do this, responses of "Very Satisfied" are set to equal one, and all other responses to equal zero. To this end, the binary version of this variable captures how likely an individual is to report being very satisfied, versus any other category of life satisfaction.

Table 4.1: Life Satisfaction at age 25 - Movers vs. Remainers

	Move=0	Move=1	Difference	P-Value
Very dissatisfied	0.02	0.01	-0.01	0.01
Fairly dissatisfied	0.06	0.07	0.01	0.13
Neither satisfied nor dissatisfied	0.17	0.13	-0.03	0.01
Fairly satisfied	0.55	0.54	-0.00	0.92
Very satisfied	0.21	0.24	0.03	0.03

4.3.3 University Movers

In waves six and seven, LSYPE contains information on whether or not the individual went to university in that wave. Those who are in higher education during these waves are asked "Do you live at home with your parents or guardians during term time?". It is this variable, coded as one if the individual responds "yes" and zero otherwise, that constitutes the main variable of interest in this chapter. Of the 3,243 individuals who attended university, 62% (N = 2,038) moved away from home and 38% (N = 1,205) remained at home during study.

4.3.4 Control Variables

In what follows in this subsection, all covariates are included in all models, unless otherwise specified in the table of results. LSYPE contains a wealth of individual, parental and area-level characteristics, from different time-points in the respondents life (i.e. from different waves). The nature of the outcomes at age 25 means that they are only observed in this final wave. As such, the analyses are cross-sectional in nature, but utilise data from all waves.

LSYPE contains information on gender. All of the analysis in this chapter is broken down by the individual's gender, in addition to both genders pooled, for two main reasons. Firstly, there are possible systematic differences in the way that men and women process information when considering their self-assessed health and wellbeing (Benyamini et al., 2000). Splitting the analysis this way (as opposed to simply controlling for gender) may shed light on these differences. Secondly the nature of moving away from home, and the university experience in general, is likely to be heterogeneous for men and women. These potential heterogeneous experiences, if present, are important to bear in mind when implementing policies that influence the nature of university attendance. Gender differences in life-satisfaction as a result of moving away from home to university pose an important feature of this analysis.

Another individual-level characteristic which potentially confounds the relationship between moving away from home to university and life-satisfaction is religious faith. In particular, individuals of Islamic faith are particularly less likely to move away from home due to financial constraints and barriers to access credit markets (see Section 4.6). Respondents in LSYPE are asked "What is your religion?", which is subsequently harmonised into the 7 2001 census categories. From this, a dummy variable is created equal to one if the individual identifies as Muslim, and zero otherwise. The latest possible version of this variable is used, from wave 8. Missing values are replaced with the latest non-missing wave available[4]. This is included as a control variable in all analyses, and also for stratification in subgroup analysis.

LSYPE contains information on the respondent's ethnicity as part of the "Identity" section of the questionnaire. They are asked "What is your ethnic group?", to which there are 18 categories to choose from. This is subsequently condensed using the ONS 6 category census classification: "White"; "Mixed"; "Indian"; "Pakistani and Bangladeshi"; "Black or Black British"; "Other Ethnic Group". Section C.1.1 provides detail of how this is done. This is included as a categorical variable in the analysis, and again using

[4]This is done only for unexplained missingness. Some cases are explained by a refusal to answer or other reasons which may reflect a change in religiosity itself. This occurs in very few cases.

the latest non-missing wave for each individual.

A major correlate of life satisfaction which is also likely a determinant of moving away from home to university is the individual's mental health. As a measure of the individual's psychosocial health, LYSPE asks respondents to complete the General Health Questionnaire (GHQ). The GHQ aggregates answers from 12 questions aimed at assessing the respondents' psychological well-being. Such questions include: "Have you recently lost much sleep over worry?"; "Have you recently felt capable of making decisions about things?"; "Have you recently been feeling unhappy or depressed?" (see Appendix section C.1.2 for the full list of questions). The respondents answer each of these questions on a likert scale: "not at all" (0), "no more than usual" (1), "rather more than usual" (2), "much more than usual" (3). Aggregating these answers from all 12 questions results in the GHQ-score, which ranges from 0 to 36. A higher score on this scale indicates a greater likelihood of mental ill health. Including the GHQ score in each analysis is appropriate, as omitting it would likely confound the effect of moving away. However, changes to an individual's mental health could be considered an outcome of moving away from home to university, and hence post-university GHQ would be a "bad control" (Angrist & Pischke, 2008). For this reason, *pre*-university GHQ, which as such constitutes a baseline level of psychosocial health during the individual's formative years are included.

An important feature of the mediation analysis which follows the main section is locus of control. Following (Lefcourt, 1991), LSYPE contains information on an individual's (internal or external) loci of control through asking respondents to what extent they agree or disagree with the following statements:A) "If someone is not a success in life, it is usually their own fault"; B) "I can pretty much decide what will happen in my life"; C) "How well you get on in this world is mostly a matter of luck"and D) "If you work hard at something you'll usually succeed". They answer by choosing one of: "(1) Strongly agree";"(2) Agree";"(3) Disagree" and"(4) Strongly disagree". The response to question C is reversed to reflect the direction of the question, and the results are summed to give a score between 4 and 16. A higher score on this scale indicates a more external locus of control. See section 4.8 for more detail about the mechanism behind this variable, moving away from home and life satisfaction. This variable is available in many waves: locus of control from waves 2 and 7 are included in all analysis, and use wave 8's iteration as a mediating variable.

University and course choice

LSYPE contains data on, conditional on attendance at university, satisfaction with student's choices. Individuals were asked "Was [university] your first choice of institution?", and also "Was [subject] your first choice of subject?", to which they responded "yes", "no" or "don't know". These two factors: whether or not a student attends their desired university or studies their preferred subject, may have a significant impact on the university experience, performance, and ultimately the outcome of their degree. In turn, this is likely to affect life satisfaction as a result, so these two binary variables are included in all analyses.

Ideally, in all analyses, university fixed effects would be included to sweep up within-university unobserved heterogeneity. However this information is not available without special license access to LSYPE. Some of the potential heterogeneity across university type is captured by controlling for whether the individual's university was a member of the Russell group or not. LSYPE asks respondents in wave 6 and 7 whether their university is a member of the Russell group, to which they respond "yes" or "no". As well as controlling for this in the main analysis, results are stratified by Russell group status in the secondary analysis. Another factor is university "quality", broadly speaking. Student satisfaction and experience is a major factor of both University life and the University quality rankings. Universities compete both in terms of student outcomes, but also in large part on the student experience they offer. Thus, attendance of a higher quality university should, on average, be positively correlated with early adult outcomes. Again, the university attended is not observed, so this chapter makes use of the Russell group indicator to pick up the effects of attending a university of a higher quality. Although this is a crude measure and not necessarily related to an insitutions' quality or ranking, on average we can expect differences in quality, experience and life satisfaction to show up through Russell group status.

Parental and Household characteristics

Alongside the individual questionnaires LSYPE collected information from their parent(s) including questions at the household level. In each analysis, the following controls for parental health, work conditions and marital status, are included, alongside the total number of siblings in the household.

Each parent in the household was asked "Do you have any longstanding illness, disability, or infirmity? By longstanding I mean anything that has troubled you over a period of at least 12 months or that is likely to affect you over a period of at least 12 months?" to which they responded Yes or No. This is included as a binary variable equal to one if the parent responds yes.

Each parent is asked about their employment status. This is recoded as a variable that indicates whether they are employed full or part-time, or whether they are currently non-employed.This employment dummy is included to capture correlation between parental employment status and moving away from home to university.

LSYPE also derives marital status of the parents in the household, listing the responses: "Single, that is, never married"; "Married and living with husband or wife"; "Living with a partner"; "Married and separated from husband or wife"; "Divorced"; "Widowed"; "Other". A variable equal to one is created if the response indicates that the parent is married and living with husband or wife, and zero otherwise. It is expected that stability in the home co-varies with the probability of moving away from home to university. Therefore included are martial status of the parents as a proxy variable for this.

In wave 2, information about the number of siblings of the young person was collected. This is subsequently updated in wave 4 to account for the boosted sample, giving the total number of siblings (including natural, step, adoptive or foster) of the young person. Again this is expected to influence the probability of moving away, hence motivating its inclusion in the analysis.

Information on household income is derived in wave 4 of LSYPE for both parents (where applicable). This gives total monthly household income, banded into 12 bins ranging from: "Up to £2,599", "£2,600 up to £5,199", then bands incrementing by £5,199 up to "£52,000 or more". Finance for students plays a large role in whether or not they move away from home, and household income plays several roles: either through directly facilitating a move at higher incomes, or receiving a higher means-tested grant at lower incomes. Household income is a likely predictor of moving away therefore, and omitting it from the analyses would invoke bias in the estimates. The above mentioned categories of household income at wave 4 (pre-university) are included as separate dummies for each band.

Area-level controls

There are likely regional disparities in the propensity to move away from home to university as households based in different areas face different costs of moving. The fact that some areas have a higher concentration of universities means that the distance of a potential move is lower and perhaps less likely. To capture these differences, region fixed-effects appear in all of the analysis, to sweep away within-region variation in the probability of moving away from home to university.

LSYPE includes the Income Deprivation Affecting Children Index (IDACI) score for each young person in the survey, collected at wave 2. The IDACI is an indicator measuring the percentage of children living in low income households based upon their postcode. This is included in each analysis to control for variation in local deprivation, which may affect both the probability of moving away, and life satisfaction at later life.

Income

A key variable for the mediation analysis in section 4.8 is the respondents income at age 25. Income was collected in sweep 8 of Next Steps using five separate banded questions. The first question gives respondents a choice between four bands, and the four remaining questions subdivide each band into four finer bands. In total the scale consists of 16 bands. Income was missing for 9.4% of the 7,707 respondents in wave 8. LSYPE imputes continuous and missing income using interval regression, using the log-upper and log-lower bands as the dependent variables (intervals). They use a host of predictors, which are listed in the appendix C.1.3.

Table 4.2: Descriptive Statistics - split by migration status

	Move=0	Move=1	Difference	P-Value
Male	0.40	0.43	0.03	0.16
Muslim	0.31	0.05	-0.26	0.00
IDACI score	0.26	0.14	-0.12	0.00
GHQ-12 Score (wave 2)	1.63	1.89	0.26	0.01
Parent has L/t health cond. (wave4)	0.16	0.13	-0.03	0.02
No. of Siblings (wave 4)	2.26	1.69	-0.56	0.00
Parents Married & living together (wave 4)	0.78	0.79	0.00	0.80
Attended preferred University	0.76	0.80	0.05	0.00
Read preferred course at Uni.	0.88	0.92	0.05	0.00
Locus of Control (wave 2)	7.45	7.56	0.11	0.11
Locus of Control (wave 7)	8.51	8.62	0.12	0.07
Locus of Control (wave 8)	9.03	9.06	0.03	0.62
Parent employed (wave 4)	0.66	0.82	0.17	0.00
Government Region				
North Eest	0.06	0.03	-0.03	0.00
North West	0.15	0.12	-0.03	0.01
Yorkshire and the Humber	0.11	0.09	-0.02	0.07
East Midlands	0.06	0.08	0.03	0.01
West Midlands	0.13	0.09	-0.04	0.00
East of England	0.07	0.11	0.03	0.00
London	0.30	0.26	-0.04	0.02
South East	0.07	0.15	0.08	0.00
South West	0.04	0.07	0.02	0.01
Ethnicity				
White	0.45	0.74	0.29	0.00
Mixed	0.03	0.05	0.02	0.00
Indian	0.16	0.08	-0.08	0.00
Pakistani	0.12	0.02	-0.10	0.00
Bangladeshi	0.12	0.01	-0.11	0.00
Black Caribbean	0.04	0.03	-0.01	0.07
Black African	0.05	0.05	0.00	0.53
Other	0.05	0.03	-0.02	0.00

4.4 The (total) effect of moving away to university on life satisfaction

4.4.1 Basic model

Basic setup:

$$Y_i = \alpha + X_i'\beta + Moved_i\tau + \varepsilon_i,$$

where Y_i is a dummy variable, equal to one if the individual reports their life satisfaction as 'very satisfied', and zero otherwise, X_i is a vector of control variables from different stages in the individual's life, including parental education, occupation and health, as well as individual-level measures. $Moved_i$ is an binary variable, indicating whether the individual moved away from home to university, or whether they stayed at home whilst studying; τ is the main parameter of interest. This model is estimated with and without controls, using both Linear Probability and Probit models.

4.4.2 Ordered Probit model

So far the basic approach has been to dichotomise individual-level life satisfaction to 'Very Satisfied' vs. not. This approach does not make use of variation between other response categories, which form an ordinal set. Therefore the next approach is to capture this variation across categories, via an ordered Probit model. The following is adapted from Wooldridge (2010) and W. H. Greene (2012), to which the reader is referred for a comprehensive overview of ordered response models.

Assume life satisfaction is measured by a latent variable, Y^*, determined by

$$Y^* = X'\beta + Moved\tau + \varepsilon, \qquad \varepsilon|X \sim Normal(0,1)$$

where X and $Moved$ are defined as before, for the population of interest. Given this latent specification, we can define the observed life satisfaction categories (Y) as:
Very Dissatisfied:

$$Y = 0, \qquad if \quad Y^* \leq \lambda_1$$

Fairly Dissatisfied:

$$Y = 1, \qquad if \quad \lambda_1 < Y^* \leq \lambda_2$$

Neither Satisfied nor Dissatisfied:

$$Y = 2, \qquad if \quad \lambda_2 < Y^* \leq \lambda_3$$

Fairly Satisfied:

$$Y = 3, \qquad if \quad \lambda_3 < Y^* \leq \lambda_4$$

Very Satisfied:

$$Y = 4, \qquad if \quad Y^* > \lambda_4$$

The response parameters, λ, are to be estimated along with β. Given the assumption that ε follows a standard normal distribution, the response probabilities are calculated as follows:

$$P(Y = 0|X, Moved) = P(Y^* \leq \lambda_1|X, Moved) = P(X'\beta + Moved\tau + \varepsilon \leq \lambda_1|X, Moved)$$

$$= \Phi(\lambda_1 - X'\beta - Moved\tau)$$

$$P(Y = j|X, Moved) = P(\lambda_{j-1} < Y^* \leq \lambda_j|X, Moved)$$

$$= P(\lambda_{j-1} < X'\beta + Moved\tau + \varepsilon \leq \lambda_j|X, Moved)$$

$$= \Phi(\lambda_j - X'\beta - Moved\tau) - \Phi(\lambda_{j-1} - X'\beta - Moved\tau),$$

for $1 < j < 4$.

$$P(Y = 4|X, Moved) = P(Y^* \leq \lambda_4|X, Moved) = P(X'\beta + Moved\tau + \varepsilon \leq \lambda_4|X, Moved)$$

$$= 1 - \Phi(\lambda_4 - X'\beta - Moved\tau)$$

Finally, as we are interested in the marginal effect of moving away from home to university on life satisfaction at 25 years old, we have:

$\frac{\delta p_0(X,Moved)}{\delta Moved} = -\tau\phi(\lambda_1 - X'\beta - Moved\tau)$

$\frac{\delta p_j(X,Moved)}{\delta Moved} = \tau[\phi(\lambda_{j-1} - X'\beta - Moved\tau) - \phi(\lambda_j - X'\beta - Moved\tau)]$, for $1 < j < 4$

$\frac{\delta p_4(X,Moved)}{\delta Moved} = \tau\phi(\lambda_4 - X'\beta - Moved\tau)$

Noting that the nonlinearity of this model means the partial effects are themselves functions of the covariates, we must either calculate these partial effects at the averages of the covariates, or calculate the average partial effects over each observation.

4.5 Results - Total Effect of moving away

Table 4.3 shows the basic estimation results from Linear Probability and Probit models where Y_i takes the value of one if the individual responds that they are Satisfied with their life, and zero otherwise. For males, moving away from home to university has a large and statistically significant ($p<0.1$) effect on reported life satisfaction at age 25. The reverse is true for females: those who move away from home to university report worse life satisfaction than those who remain at home, albeit statistically insignificant at the 10% level. These results remain persistent as controls are added, and the choice of linear or non-linear estimation method makes little difference.

Table 4.4 shows the ordered probit estimates for life satisfaction. Again, male movers enjoy a higher life satisfaction at age 25, whereas women do not. These estimates provide a little more detail. Male movers are less likely to report being fairly dissatisfied and neither satisfied nor dissatisfied by one and two percentage points, respectively ($p<0.1$); and more likely to report being satisfied by four percentage points ($p<0.1$). Females on the other hand are less likely to report being satisfied, and more likely to report being in all categories below (although again, not statistically significant at the 10 % level).

The take-away result from Tables 4.3 and 4.4 is that moving away from home to University is positively associated with life satisfaction at age 25, but only for Males.

Table 4.3: Average Marginal Effects of moving away to university on Life Satisfaction at age 25 years

	Males			Females		
LPM	0.06**	0.05*	0.05*	0.02	-0.02	-0.02
	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Probit	0.06**	0.05*	0.05*	0.02	-0.02	-0.02
	(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
N	1322	985	985	1833	1298	1298
Controls (excl Wages)	-	✓	✓	-	✓	✓
Wages (at age 25)	-	-	✓	-	-	✓

Robust standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.4: Estimated marginal effect of moving away to University on Life Satisfaction probabilities

	Males			Females		
Life Satisfaction:						
Very Dissatisfied	-0.01***	-0.00*	-0.00	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Fairly Dissatisfied	-0.02***	-0.02**	-0.01*	-0.00	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Neither Satisfied	-0.04***	-0.03**	-0.02*	-0.00	0.01	0.01
nor Dissatisfied	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Fairly Satisfied	0.01**	0.00	0.00	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Satisfied	0.06***	0.04**	0.04*	0.01	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
N	1322	985	985	1833	1298	1298
Controls (excl Wages)	-	✓	✓	-	✓	✓
Wages (at age 25)	-	-	✓	-	-	✓

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.6 Results - Subgroup analysis

In the context of moving away from home to university, there are several subgroups of students for whom it is reasonable to expect heterogeneous effects of moving. Three cases are examined here: those of Islamic faith, those attending Russell Group universities, and first-generation university attendees.

For many students, the decision to move away hinges on access to credit markets. The UK Government offers undergraduate students a Tuition Fee and Maintenance Loan as part of its Student Finance funding system. These loans incur a rate of interest which, since 2012, are set at commercial rates. In the case of Muslim students, many may feel unable to take on such interest-bearing loans for religious reasons; namely a lack of

Sharia-compliance. This issue has been recognised by the Government, but there is no time frame for plans to introduce an alternative (Department for Business & Skills, 2016). As such, it seems likely that Muslim students who do move away are likely to be systematically different from those who remain at home.

Universities compete both in terms of student outcomes, but also in large part on the student experience they offer. Thus, attendance of a higher quality university should, on average, be positively correlated with early adult outcomes. Taking Russell group status as a proxy for quality, those who attend Russell group universities are likely to face a different university experience than those who do not. This section therefore investigates whether there is a differential impact of moving away to a Russell group University, over and above moving away to a non-Russell group University. Likewise, a different experience at University is likely to be faced by those who are first-generation university attendees. Those who have a familial history of university attendance are likely to be, *ceteris paribus*, better prepared for undergraduate life. Whether there is a differential impact of moving away for first-time attendees, above those who are not, is analysed.

The possibility of heterogeneous effects are investigated by interacting each of the above mentioned subgroups with the variable indicating whether a student moved away from home. LSYPE contains information on faith so Muslim students are easily identifiable, as are Russell group attendees.A (crude) first-generation attendee variable is created by exploiting the parental questionnaires in the first wave. Both parents in the household were asked if they attended university, and they were also asked if their parents attended. Information on siblings' attendance at university is not available. A variable equal to one is created if at least one parent or grandparent attended university and zero otherwise, to proxy first-generation attendance. The original Linear Probability Models are used to estimate these interactions, both for simplicity in their estimation and to avoid difficulties with interaction terms in non-linear models.[5]

Table 4.5 shows the original LPM coefficients (left-most column), followed by the coeffi-

[5]See Ai and Norton (2003); W. Greene (2010) for a more detailed exposition of the issues with interaction terms in non-linear models.

Table 4.5: Estimated marginal effect of moving away to University on Life Satisfaction probabilities

	Males			
University Mover	0.06**	0.09***	0.01	0.06
	(0.03)	(0.03)	(0.03)	(0.05)
Muslim		0.19***		
		(0.07)		
Mover x Muslim		-0.15*		
		(0.08)		
Russell Group			-0.13	
			(0.09)	
Mover x Russell Group			0.23**	
			(0.09)	
1st-Gen Uni attendee				-0.04
				(0.05)
Mover x 1st-Gen Uni attendee				-0.01
				(0.06)
N	1134	1130	1082	1095
	Females			
University Mover	-0.02	-0.01	-0.03	0.03
	(0.02)	(0.03)	(0.03)	(0.05)
Muslim		0.05		
		(0.07)		
Mover x Muslim		-0.13		
		(0.08)		
Russell Group			0.12*	
			(0.07)	
Mover x Russell Group			-0.04	
			(0.08)	
1st-Gen Uni attendee				0.05
				(0.05)
Mover x 1st-Gen Uni attendee				-0.07
				(0.05)
N	1537	1533	1459	1467

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

cients from the interacted models. For Muslim males who move away from home, they are 15 percentage points worse-off than non-Muslim movers (p$<$ 0.1): those who do move are 6 percentage points less likely to report being very satisfied with their life at age 25. Female Muslims who move are also worse-off, but these results are not statistically significant at the 10 % level. Males who move away from home to a Russell group University are significantly better off: they are 23 percentage points more likely to report being very satisfied with their health (p$<$ 0.05), an effect size equal to roughly 55%

of a standard deviation. Female movers to Russell group universities are differentially worse-off, but not statistically significantly so. There are no statistically significant differences found for first-generation attendees. A result that could be due to measurement error in familial history of attendance, as it seems likely that an individuals' siblings' university experience would influence their own.

4.7 Endogenous movers - Instrumental Variables approach

The decision to move away from home to university is not random; those who do so, may self-select into moving away and there are unobservable factors that are likely to influence the decision to move and life satisfaction. In an attempt to address these endogeneity concerns, an instrumental variable for moving away is used that, conditional on X, is arguably excludable from the Life Satisfaction equation.

LSYPE has information on whether or not the individual's grandmother or grandfather attended university. A binary variable equal to one if at least one grandparent attended university is used as an instrument for moving away. The validity of this approach relies on the instrument's excludability - that an individual's grandparents' university attendance doesn't affect their life satisfaction other than through their own university attendance. One major factor this assumption overlooks is the influence of a grandparents' university attendance on social mobility. There are theories in the social mobility literature which suggest that a future generation is independent of its past generations, conditional on the present generation (Zeng & Xie, 2014; Mare, 2011). Fortunately, LSYPE provides sufficient information to control for parental confounds. This means that the exclusion restriction holds, if the generational theory holds true. There is some contention to this theory in the literature, however, with some studies finding there can be direct effects, conditional on parental socioeconomics status (Cherlin & Furstenberg Jr, 1992; McLanahan & Percheski, 2008). In terms of the instrument's relevance to a child's university attendance, linear first-stage regressions show that the F-stats pass the Stock and Yogo (2002) test for weak instruments: having a grandparent who attended university is a positive, statistically significant predictor of whether a student moves away from home to university.

For the two-stage estimation, the analysis follows Cullinan and Gillespie (2016) and uses a Probit first-stage, and an Ordered Probit second stage. This is operationalised in Stata 16 with the `-cmp-` suite of commands (Roodman, 2011)[6]. Table 4.6 shows the first stage probit results and Table 4.7 shows the second-stage Probit and Ordered Probit IV estimates for Life Satsfaction at age 25, by gender.

For the binary response model, the coefficients are close in magnitude to those of the LPM and Probit models, but the increase in standard errors owing to the uncertainty in the first stage mean these are not statistically different from zero. Unpacking these in the ordered probit IV model: for both males and females, the signs of the effect of moving away on each category of Life Satisfaction at age 25 are the same as the original specification (see Table 4.4). For males, the coefficients are of a similar magnitude to that of the ordered probit model, but are too imprecise to be deemed statistically different from zero. The coefficient on satisfied, for example, increases by around three percentage points (1.75 times as large), whilst the standard errors become around seven times as large. For females, on the other hand, the standard errors again increase (as expected when using instrumental variables), but the coefficients balloon to up to ten times as large.

Assuming excludability and monotonicity of the instrument, this analysis suggests a large, positive, causal effect of moving away from home to university on life satisfaction for males who are "compliers" with the instrument. In other words, for those who are influenced to move away from home by their grandparents' university attendance, who wouldn't have done so otherwise, there are positive wellbeing-returns to doing so. However, there is no guarantee that an individual's grandparents' university attendance affects their life satisfaction only through influencing them to move away from home to university. This, if true, would lead to the estimator being inconsistent, and the coefficients biased.

Another issue with this analysis is the first-stage. Grandparents' university attendance

[6]The `-cmp-` command essentially fits SUR models, and is extremely flexible in terms of functional form of each estimating equation. The accompanying article to the command by Roodman (2011), explores the many different combinations of estimation methods, including ordered probit IV.

Table 4.6: First-Stage Probit estimates of moving away to university

	Males	Females	All
G.Parent attended University	0.077**	0.115***	0.092***
	(0.033)	(0.031)	(0.023)
Muslim	-0.180***	-0.211***	-0.229***
	(0.064)	(0.062)	(0.054)
Mixed	0.169**	0.272***	0.199***
	(0.071)	(0.067)	(0.033)
Indian	-0.026	-0.033	-0.027
	(0.048)	(0.048)	(0.036)
Pakistani	0.092	-0.028	0.031
	(0.084)	(0.080)	(0.056)
Bangladeshi	-0.086	-0.007	-0.037
	(0.096)	(0.091)	(0.070)
Black Caribbean	-0.050	0.156**	0.082
	(0.094)	(0.074)	(0.052)
Black African	0.161**	0.236***	0.176***
	(0.078)	(0.077)	(0.039)
Other	0.143	0.040	0.077*
	(0.089)	(0.065)	(0.047)
IDACI score	-0.491***	-0.440***	-0.466***
	(0.086)	(0.079)	(0.058)
GHQ-12 Score	-0.009	0.006	0.002
	(0.006)	(0.004)	(0.003)
Parent has L/t health cond	0.030	0.019	0.023
	(0.036)	(0.034)	(0.025)
Parent Working FT	-0.076**	-0.035	-0.053**
	(0.037)	(0.033)	(0.023)
Parent Working PT	-0.074*	-0.031	-0.049**
	(0.040)	(0.035)	(0.025)
No. of Siblings	0.012	0.016	0.013*
	(0.010)	(0.010)	(0.007)
Parents Married & living together	-0.073**	-0.077**	-0.073***
	(0.035)	(0.030)	(0.021)
Attended preferred University	-0.036	-0.009	-0.020
	(0.032)	(0.029)	(0.021)
Read preferred course at Uni	0.073*	0.073*	0.076**
	(0.043)	(0.038)	(0.031)
Locus of Control (wave 2)	0.009	0.001	0.005
	(0.007)	(0.006)	(0.005)
Locus of Control (wave 7)	0.001	0.017***	0.010**
	(0.008)	(0.006)	(0.005)
Locus of Control (wave 8)	0.002	-0.005	-0.002
	(0.007)	(0.007)	(0.005)
ln(income)	0.508***	0.491***	0.500***
	(0.096)	(0.091)	(0.066)
Male			-0.068***
			(0.017)
N	1127	1541	2674
F - Statistic	11.371	13.949	23.100
Region Fixed Effects	✓	✓	✓

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.7: IV Ordered Probit estimates of moving away to university on life satisfaction

	Males	Females	All
Binary Life Satisfaction:			
V. Satisfied	0.062	-0.018	0.013
	-0.279	-0.286	-0.247
Ordered Life Satisfaction:			
Very Dissatisfied	-0.009	0.023	0.011
	(0.019)	(0.019)	(0.012)
Fairly Dissatisfied	-0.029	0.076*	0.038
	(0.055)	(0.040)	(0.035)
Neither Satisfied nor Dissatisfied	-0.043	0.082**	0.050
	(0.080)	(0.032)	(0.042)
Fairly Satisfied	0.004	0.030***	0.008
	(0.007)	(0.011)	(0.007)
Satisfied	0.077	-0.211**	-0.107
	(0.147)	(0.097)	(0.095)
N	1133	1542	2675
First-Stage F-stat	11.371	13.949	23.100

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

only has borderline predictive power for moving away from home to university. The uncertainty brought about in this stage might mean there is too much imprecision in the second-stage to reach meaningful conclusions about the coefficients.

4.8 Mediation Analysis: the indirect effects of moving away

The analysis in the previous section presents two main results: that there is a negative association between moving away from home to university and life satisfaction in early adulthood; and that this effect is only apparent for males. This section aims to shed some light on the potential transmission mechanisms at play behind these results. With a view to provide insight on the gender difference, focus is placed on the indirect (mediating) role of an individual's external locus of control, and their wages.

Locus of control refers to individual beliefs about whether life events are mostly internally or externally determined (Rotter, 1966). People with an external locus of control believe that what happens in life is largely determined by events beyond their control, whereas individuals with internal locus of control generally perceive a sense of personal control; their own decisions and behaviours are the main determinants of their outcomes. There exists a link between an individual's locus of control and their life satisfaction. For example, those who believe they have little or no control over their life are more distressed, and are likely to have a lower life satisfaction as a result (Mirowsky, 2017). This has led to an individual's locus of control being examined as a mediator on the causal path to life satisfaction (see, for e.g. Fiori et al. (2006)) There is some evidence to suggest that the home environment plays a role in a young person's locus of control (Bansal et al., 2006). In the context of moving away from home to university, this is often the first time an individual may have lived away from home for an extended period. It is hypothesized that this may have an effect on an individual's perception of self-control, which may not otherwise be experienced at this early stage in an adult life.

Another factor influencing early-adulthood life satisfaction is wages, which has been considered as a mediator in an education context in previous literature (del Mar Salinas-Jiménez et al., 2013). Being restricted to living at home whilst studying at University greatly reduces the choice set that would otherwise be available. Prospective students in the UK list five choices when they apply, meaning the probability of attending a higher-quality institution is lower with a choice set restricted to within a commuting distance of home, *ceteris paribus*, and thus are likely to earn less on average than those who move away. Given the well-documented gender gap in pay, this could provide some explanation as to why moving affects male life satisfaction only.

4.8.1 Methods

In order to disentangle the direct effect of moving away on early-adult life satisfaction, from the indirect effects of loci of control and wages, the approach of Han et al. (2011) is followed. Adapting the basic model from the earlier analysis gives the Life Satisfaction equation as:

$$Y_i = \alpha + X_i'\beta + Moved_i\tau + Locus_i(Moved_i)\gamma + Wages_i(Moved_i)\eta + \mu_i,$$

where $Locus_i$ is the individual's external locus of control, which is a function of whether they moved away from home to university. Likewise, $Wages_i$ is the level of wages faced by the individual at 25 years of age, also a function of whether the individual moved.

Taking the total differential of life satisfaction with respect to moving away to university gives:

$$\frac{dY_i}{dMoved_i} = \tau + (\frac{\delta Y_i}{\delta Locus_i}\frac{\delta Locus_i}{\delta Moved_i}) + (\frac{\delta Y_i}{\delta Wages_i}\frac{\delta Wages_i}{\delta Moved_i}),$$

where the direct effect of moving away from home on life satisfaction (holding fixed locus of control and wages, as well as X_i) is given by τ. The total effect of moving away is further decomposed into two indirect effects which appear as the partial derivative terms in the two sets of parentheses.

For estimation purposes, the basic linear probability specification for life satisfaction are used. There is little informational (nor statistical significance) difference between the ordered probit and LPM/probit specifications, and doing so makes computation of the direct and indirect effects more straightforward. The first partial derivative terms in parentheses, $\frac{\delta Y_i}{\delta Locus_i}$ and $\frac{\delta Y_i}{\delta Wages_i}$, are taken from this model. The second two partial derivative terms are estimated via the following:

$$Locus_i = {\gamma_0}^L + Moved_i{\gamma_1}^L + X_i'{\gamma_2}^L + {\mu_i}^L$$

$$Wages_i = {\eta_0}^W + Moved_i{\eta_1}^W + X_i'{\eta_2}^W + {\mu_i}^W$$

and the coefficients ${\gamma_1}^L$ and ${\eta_1}^W$ are taken as estimates of $\frac{\delta Locus_i}{\delta Moved_i}$ and $\frac{\delta Wages_i}{\delta Moved_i}$. This gives a value for each of the indirect effects.This entire procedure is then bootstrapped to obtain estimated standard errors for the direct, indirect and total effects of moving away from home to university on early-adult life satisfaction. Using the above parameters, the equation for decomposing the total effect becomes:

$$\frac{dY_i}{dMoved_i} = \tau + (\gamma \times \gamma_1^L) + (\eta \times \eta_1^W),$$

4.9 Results - Direct and Indirect effects of moving away

Table 4.8 shows the estimated direct, indirect and total effects of moving away to university on life satisfaction at age 25. Unsurprisingly given earlier results, only significant effects are found for males. The mediation analysis shows that the positive total effect of moving away from home to university on life satisfaction is not explained by either mechanism relating to locus of control or wages. Of the 6.5 percentage points (p<0.05) that an individual who moves is more likely to report being very satisfied with their life, 6.4 percentage points (p<0.05) are attributed to a *direct* effect of moving.

Table 4.8: Estimated marginal effect of moving away to University on Life Satisfaction probabilities

	Full Sample	Males	Females
Indirect Effect: Locus of control	0.002	-0.001	0.003
$(\gamma \times \gamma_1^L)$	(0.003)	(0.005)	(0.005)
Indirect Effect: Wages	0.001	0.001	0.002
$(\eta \times \eta_1^W)$	(0.003)	(0.004)	(0.003)
Total Indirect Effect	0.003	0.000	0.005
$(\gamma \times \gamma_1^L + \eta \times \eta_1^W)$	(0.004)	(0.006)	(0.006)
Total Effect	0.016	0.065**	-0.015
$(dY_i/dMoved_i)$	(0.020)	(0.028)	(0.024)
Total Indirect/Total Effect	0.212	0.044	-0.142
	(1.252)	(0.199)	(2.720)
Direct Effect	0.013	0.064**	-0.020
(τ)	(0.019)	(0.028)	(0.023)

Standard errors in parentheses
$^{*}\ p < 0.1$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

For females, ignoring the imprecision of the estimates, the results appear to show a case of "inconsistent mediation" (MacKinnon et al., 2012), whereby the indirect effects have a different sign to the direct effects. The total effect estimate is smaller (closer to zero) than the direct effect, which suggests the presence of a suppressing variable. The overall effect of moving away from home is negative for females, but the particular mediating paths through Locus of Control and Wages are positive ones. This suggests that there are other, larger, factors associated with moving away that offset these positive effects.

The caveat of this analysis is that its plausibility relies on the identification of the initial, total, effects (i.e the results from the initial analysis). Clearly, owing to the potential endogeneity of moving away from home vs remaining whilst studying, causal claims can not be based on the results shown here. A further issue with the mediation analysis, is that even if moving were exogenous, the two indirect mechanisms analysed are potentially endogenous too. This, in some sense, “doubles” the endogeneity problem and so in order to address causality one would need three instruments: one for each of moving away from home, wages, and locus of control. This is deemed to be beyond the scope of this chapter, and therefore leave this line of analysis to further research.

4.10 Discussion, Limitations and Conclusions

This chapter has considered the impact of moving away from home to university on life satisfaction in early-adulthood. OLS, Probit and Ordered Probit methods were applied to a cohort of university attendees from LSYPE, some of whom stayed at home during study, whilst others moved away. In addition to this baseline analysis it explored whether there exist heterogeneous effects of moving away, and potential transmission mechanisms using mediation analysis. In particular the latter considers how external locus of control and wages may act as channels through which moving away from home affects life satisfaction.

The main result is that moving away has a differential effect for men and women. Conditional on both pre- and post-move characteristics at the individual, parental and local level, males report much higher life satisfaction at age 25 if they moved away; females show no difference. The move to university is often the first time a young person moves away from home, and thus represents a significant life event that occurs at age 18. There is perhaps an argument for a gender difference in developmental and maturity levels at this age, and thus it seems plausible that having a vastly different university experience (through moving away) may impact men and women differently. Subsequent analysis of heterogeneous effects within gender, and of potential mediators of the direct effect of moving away, shed some light on this baseline difference in early-adulthood life satisfaction.

Secondary analysis considered heterogeneous effects for Muslim respondents, those who attend Russell group universities, and those who are a “first-generation” university attendee. The analysis for females shows no statistically significant interaction effects. For males however, those who identify as being of Islamic faith are much worse off if they move away from home, than movers who do not. Due to the issues around access to credit markets for those of Islamic faith, this result may arise because of the financial strain placed upon those who move away from home. Perhaps taking on a job alongside study or the financial stress more generally are drivers of this effect. Further research is needed in this area, particularly for policy surrounding access to sharia-compliant student finance.

The effect of moving away from home to university is amplified if that university is a member of the Russell group. This is perhaps unsurprising, as these "elite" universities recruit nationally from more affluent families and are more isolated from their local communities (Donnelly & Gamsu, 2018). These universities are likely more geared towards facilitating this type of student, thus maximizing the benefit of them moving away from home to attend.

As previously mentioned, there is a feasible argument for a gender difference in emotional maturity at age 18. A study of undergraduates from Delhi University, for example, found females to be much more capable of emotional adjustment to life at university in their first year (Mahanta & Kannan, 2015). It could be the case that, for males, moving away from home triggers a significant emotional development throughout their time living away. Whereas females are further along in their emotional development and thus don't face the same benefits from moving away. The psychology literature has made links between an individual's external locus of control, maturity, and psychosocial development (Brackney & Westman, 1992). LSYPE contains information on the individual's locus of control, and so this provides an opportunity to test a loose version of this hypothesis. The mediation analysis conducted in this paper suggests that moving away from home to university does not impact life satisfaction indirectly through influencing an individuals' external locus of control. The effects must therefore either be driven by the direct mechanism of moving away, or by some other confounder which is not accounted for in the analysis.

Another possible channel though which we might explain the gender difference and university experience is through wages. However, the mediation analysis present here suggests that differences in wages, which come about as a result of moving away, do not impact life satisfaction. Future research should seek to investigate this finding further: whether the effect of moving away is purely direct, as this chapter shows suggests. The question of which potential mechanisms are at play remains unanswered.

Individuals self-select into the decision to move away from home, and there are common causes of moving away and early-adult life satisfaction. The analysis here goes some way to controlling for pre-move characteristics at the individual, parental, household and area level. Whilst this alleviates some of the bias from joint confounders, we cannot make causal claims from the data due to the self-selection issue. To address this problem, the individual's grandparents' university attendance is used as an instrumental variable for moving away to university. This analysis is suggestive of a large, causal effect of moving away on life satisfaction for males, and the reverse for females. The reliability of these estimates are still limited, however, as the IV only has weak predictive power of moving away in the first-stage. This lack of predictive power magnifies the standard errors to the extent that meaningful inference is difficult. Future work on this paper will focus on addressing the endogeneity of moving away, through looking at whether the individual is attending their first-choice university and first-choice course as an IV.

4.10.1 Limitations

There are several main limitations of this study, many of which revolve around the nature of LSYPE. There are two main threats that arise from using LSYPE: non-random attrition of students from LSYPE over the 8 waves, and the inability to use panel data methods when looking at the outcomes presented here.

There is a large amount of attrition from LSYPE: around 15,000 students are present in wave 1, and around 8,000 remain in wave 8. To account for this, the survey weights included in LSYPE are used, from wave 8. However, these were designed to be used for attrition in general within LSYPE, and not for the specific case of university attendees who drop out of the survey. Conditional on attendance at university in wave 6, there were 3,412 students who attended; this drops by around a third to 2,349 of these students who remain at age 25. The attrition of these students seems to be randomly distributed between those who remain at home and those who move away. Of those at university in wave 6, 63.3% moved away from home. Of the university attendees who remain in wave 8, 64% moved away. This suggests that the attrition of students who move away is not a problem over and above the issue of attrition from LSYPE in general, and therefore justifies the use of the general survey weights. Further work will look into this issue more carefully, considering the characteristics of those who attend university and subsequently drop out of the survey.

LSYPE offers the advantage of following the same cohort of children over time. The nature of the data naturally lends itself to the use of panel data methods, which can help attenuate bias that occurs through unobserved time-fixed heterogeneity at the individual level. However there are inconsistencies in the surveys across waves, meaning that not all variables are available in all waves. In the most extreme case, a variable is available only in one wave. This is the case with the main outcome of interest in this paper: life satisfaction. For this reason the analysis is constrained to cross-sectional methods, and thus the assumption that the unobserved effects are uncorrelated with moving away from home to university. This assumption is both untestable and unlikely to hold, considering that at the very least, individuals self-select into moving away. Another important consideration is that this cohort, born in 1989 and 1990, attended university during the financial crisis of 2008, graduating and competing on the job market during its aftermath. This will have affected these students and their success on the labour market - likely to the detriment of their life satisfaction in early-adulthood. Consequentially, this threatens the external validity of this chapter's findings, and one must be careful in extrapolating to other university cohort years who attended much earlier or later.

Another limitation of the study is the endogeneity of the decision to move away from home to university. This issue is exacerbated in the mediation analysis in which the two mediators used, wages and locus of control, are themselves endogenous with respect to life satisfaction later in life. Whilst there do exist methods involving instruments for each endogenous variable in this setting, the literature is underdeveloped, and addressing this issue is beyond the scope of this paper. Whilst an arguably excludable instrumental variable is used to address the baseline endogeneity issue, there is clearly room for more work in this area to overcome the identification issue.

Student mobility is only considered as binary - this has been criticised by some, such as (Finn, 2017), who identify social mobility as a broader issue, including flows, stops and starts that emerge throughout higher education. This paper is constrained by the data it uses, and so does not capture this more nuanced view of student mobility. Ideally, panel data on students would be available, containing data on their outcomes over time, as well as where they are living contemporaneously. This would allow for this issue to be looked at in more detail. Nevertheless, moving away to university is likely to be binary in nature for the vast majority of students. Those who opt to change from living at university to living back with their parents are a small group, and are unlikely to affect the results shown here.

4.10.2 Conclusions

This paper has contributed to the internal migration and student mobility literature by asking a novel question about how a young person's wellbeing in early-adulthood can be affected by living away form home at university. The results suggest that males are better off, whilst females are not. It has also considered potential mechanisms empirically, using mediation analysis to disentangle the indirect from direct effects of moving. Finally, this paper has addressed the endogeneity issue surrounding self-selection into moving away by using instrumental variables. It opens up a new strand of research in the student mobility literature, considering early-adult consequences of moving away, and provides a baseline for further research in this area to compare to.

There have been conflicting calls surrounding how governments and universities should influence student mobility. Damian Hinds, the UK education secretary, suggested that universities should offer "commuter courses", where students stay at home to cut costs (*Grammar school expansion and faith school reforms: Damian Hinds sets out his stall*, 2018). On the other hand there have been recent calls, influenced through a recent report by Donnelly and Gamsu (2018), for schools to encourage students to move away from home.

The main results from this analysis would support such claims. However, policies aimed at influencing attendance at university and moving location to do so must be careful to ensure they don't leave vulnerable groups of students behind. As part of a suite of policies designed to increase moving away from home, decisions makers should ensure that Muslim students in particular are facilitated in their move if they do so. It would seem that moving away from home for Islamic students is part of a different mechanism, and policies which may influence access to credit markets and thus facilitate a move away from home need to be based on further work which specifically looks into this problem. The baseline results here suggest that facilitating such moves would be detrimental to Muslim students' early adult life satisfaction. Without further research into disentangling this issue, policy-makers should tread with caution on what is a complex issue.

More generally decision-makers, including those within universities, should ensure that appropriate support is offered both to those who have moved, and in particular to those who are living at home during study. There are vulnerable parts of society who are perhaps constrained to living at home, which not only greatly limits university choice and thus quality, but also leaves (especially for males) them worse-off in early-adulthood than if they were to move away.

CHAPTER 5

Conclusion

5.1 Aims and objectives & how they were met

This thesis has considered three separate but related research questions centered around how where an individual lives can influence their health and health related outcomes. There are many under-researched areas within health and place, which have the potential to improve health policy. This thesis aimed to understand how local characteristics and internal migration can affect health and wellbeing. Three specific objectives within this thesis were (1) to understand how moving house can affect health and wellbeing; (2) to explore the differences in health outcomes on the UK coast; and (3) to explore the role of moving away from home to university on early adult life satisfaction. This thesis used large longitudinal datasets to add to the evidence surrounding these topics. Within each specific objective, various recently developed methods were employed to try and address the self-selection and endogeneity issues in answering these questions. Furthermore, a range of health and wellbeing measures were used in the analyses which capture the multifaceted nature of health [1].

5.2 Empirical chapter summary

Chapter Two considered how moving house can affect health and wellbeing. It did so by making use of two household panel surveys in the UK: BHPS and USoc, and a special license linkage with 2011 census geographical data. This allowed for the analysis of 31,216 individuals over 2-13 observed periods. In order to address the endogeneity of moving home, an Instrumental Variables identification strategy was used, with local school quality and the age of the youngest child comprising the main instrument set. The rationale behind this instrument is that the schooling decisions play a large role in a household's decision to move, and the potential health effects are either ignored or unknown to the household that moves. The main findings of the chapter are that moving house has a negative effect on an individual's self-assessed health outcomes, once instrumenting for moving home. However, the imprecision of these estimates once using an instrument mean that they must be treated with caution. The analysis went on to consider short-run effects in an RDD-type set up, which compared movers who were interviewed in the 12 months before, and 12 months after they moved home. Doing so revealed, descriptively, a negative anticipatory effect of moving in the few months prior

[1] `https://www.who.int/about/who-we-are/constitution`

to doing so, followed by an offsetting positive effect in the three months post move.

Chapter Three investigated the difference in health outcomes on the UK coast, making use of Understanding Society and a special-license access to 2011 census geographical data meant that households could be located with respect to the UK coastline. This distance to the coast measure was then used to define a treatment group which resided within 2.5km of the coast, and a control group that lived further than 15km from the coast (the data in-between were dropped to create a sharp distinction between coast and inland). The results paint a stark picture of the coast: contrary to the previous literature, health outcomes are worse on the coast, conditional on individual, household and local area characteristics. As are health behaviours, with smoking and drinking more prevalent on the coast, and coastal residents being more likely to claim disability benefits. These results were scrutinised by making use of coefficient and R^2 changes when including covariates to assess the coefficient stability and selection on unobservables. Doing so showed that the unhealthy behaviour prevalence on the coast holds up well, even with an extreme selection on unobservables issue. These results are a strong contribution to a limited literature, which has so far only considered self-assessed health benefits of those on the UK coast: a result that is not found from the data used in this chapter.

Chapter Four focused on a younger age group and considered the role of moving away to university on early adult life satisfaction. It did so by using LSYPE, a cohort study that followed children from the age of 14 years every year until aged 19 years, and then once again in early adulthood at age 25. The life satisfaction at age 25 was compared between those who lived away from home at university and those who remained living with their parents, in an ordered probit model. The data allow for the partialling out of pre-, during-, and post-university individual characteristics, as well as parental and household controls. Heterogeneous effects were found, with males reporting much higher life satisfaction if they lived away from home, whereas females who moved away showed no significant difference in early adult life satisfaction. The endogeneity of moving away is addressed by using the individual's grandparents' university attendance as an instrument in an IV ordered probit model. The first stage induced too much uncertainty in the second stage estimates for these results to be more reliable than the standard analysis. Finally, a mediation analysis was performed to assess the potential role of income and an individual's external locus of control in explaining the difference in effect for males and females. Neither income nor loci of control act as an indirect effect of moving on life satisfaction, and so the mechanism behind moving must either be direct or due to some other unobserved factor to the analysis.

5.3 Contribution to knowledge

Overall, this thesis has contributed to several different literatures in several different ways. Novel research questions were explored within health and place. Only several papers have previously considered health on the UK coast as is addressed in chapter 3. Chapter 4 is the first study to consider the wellbeing effects of moving away from home to university, and this opens a new branch of literature entirely. Chapter 2 extended the literature by considering any move as internal migration, as opposed to the typically larger moves found in the labour literature.

A new dataset was applied to the literature in chapter 3 (Understanding Society with special-license access to 2011 census geographical data) It also used outcomes that have not previously been used- namely health behaviours, disability benefit claimants, and chronic health conditions. These outcomes shed new light on the problems faced by residents on the coast.

Finally, new methods were used in order to address the endogeneity issues that arise in exploring the relationship between health and place. Chapter 2 applied an instrument set that has not been used in the literature before. Using local distance-weighted school quality in this way- through LSOA linkage- shows potentially new avenues for future research to address this question.

5.4 Policy Implications

There are many implications for policy and decision makers from the results presented in this thesis. Primarily, location and relocation matter for individual-level health. Each of the chapters has shown a different mechanism or feature through which this is the case, has identified populations at greater health risk. This in the UK, especially in the face of greater strain on the NHS and a worsening health care crisis more generally, is of particular importance to decision makers whose aim is to ensure the efficient allocation of resources. A failure to account for the health mechanisms highlighted in this research - especially for resource allocation and unmet need - will lead to ineffective or potentially detrimental policies. Chapter 2 provides evidence on the fluctuating effects of moving house on health and wellbeing, and policymakers should be aware of the differing demands this may place on local healthcare systems. Results of chapter 3 suggest that the optimal allocation of resources may not be being achieved due to the disparity in health and wellbeing outcomes between coastal and non-coastal regions. Chapter 4 addresses important policy concerns over whether adequate support is given to both those who have moved, and those who are living at home during study, and the potential negative implications for vulnerable parts of society who are constrained to living at home.

The results of this thesis provide evidence on particularly important issues, given the economic issues as a result of the COVID-19 pandemic; namely: What are the implications on health and well-being of the influx of people moving house? What does the increase in demand for UK holiday locations mean for coastal towns and the health and well-being of residents of these areas? Is this change in behaviour sustainable, and what are the future implications if not? What are the potential effects on well-being of the rise in students working remotely from home?

5.5 Limitations and future research

In each of the empirical chapters in this thesis, the variable of interest - whether location or relocation based - is endogenous with respect to health outcomes. Individuals, for the most part, choose where to live and when to move which makes the estimation of causal effects extremely difficult. Given the data constraints, this thesis has attempted to use methods that either address the endogeneity directly, through the use of instrumental variables, or to estimate bounds around a potentially biased effect. Though each of the chapters has addressed this endogeneity problem to some extent, there is clearly scope for future work that adequately accounts for the unobservable mechanisms which pose a threat to identification.

An example of an extension to Chapter 4 could be to exploit changing patterns in students not moving away from home to university due to COVID-19 - and the between-university variation in these policies - to address the self-selection issue in this area of research.

APPENDIX A

Appendix to Chapter 2

A.1 Calculation of APEs and their standard errors in a bivariate probit model

```
capture program drop bivpro_boot

program bivpro_boot, rclass

    biprobit (‘1’ = ‘2’ ‘3’) (‘2’ = ‘3’ ‘4’), cluster(pid)
    predict xb_hat, xb1
    gen xb_hat0 = xb_hat - _b[‘2’]*‘2’
    gen xb_hat1 = xb_hat - _b[‘2’]*‘2’ + _b[‘2’]
    gen pe1 = normal(xb_hat1) - normal(xb_hat0)
    summarize pe1
    return scalar ape1_‘1’ = r(mean)

    drop xb_hat xb_hat1 xb_hat0 pe1

end

bootstrap r(ape1_‘yvar’), cluster(pidj) reps(500) seed(280191): ///
 bivpro_boot ‘Y’ ‘M’ ‘X’ ‘Z’
```

Table A.1: Bivariate Probit Estimates: Average Marginal Effects

	Good SAH	GHQ-12	Health Problem
OLS	-0.011**	0.102	0.001
	(0.005)	(0.073)	(0.005)
Bivariate Probit (no IV)	-0.003**	0.003	0.001
	(0.001)	(0.002)	(0.000)
ρ	[0.049]	[-0.066]	[-0.043]
Bivariate Probit (w/ IV)	-0.004*	0.001	0.001
	(0.002)	(0.002)	(0.001)
ρ	[0.096]	[-0.030]	[-0.132]
N	107736	107736	107736

Notes: Bivariate probit estimates of average marginal effects of moving home, including those instrumented by school choice, on various health outcomes. Each row represents a different estimation method, as indicated by the leftmost column. The school choice instrument consists of locally distance-weighted average school quality, interacted with the age of the youngest household, in the previous period. ρ indicates the degree of correlation between the two error terms. Standard errors estimated through 500 bootstrapped replications. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

A.2 Robustness Checks

Table A.2: OLS and IV second stage estimates of moving house on health outcomes: including number of cumulative moves

	OLS			2SLS IV: School Choice			
Outcomes:							
>Good SAH	0.003	-0.003	-0.012***	-0.263***	-0.479***	-0.060	-0.048
	(0.004)	(0.004)	(0.004)	(0.054)	(0.079)	(0.052)	(0.051)
GHQ-36 Score	0.145**	0.083	0.083	-0.844	4.823***	1.135	1.107
	(0.062)	(0.063)	(0.061)	(0.725)	(1.072)	(0.728)	(0.721)
Health Problem	-0.001	0.002	-0.001	0.456***	0.427***	0.167***	0.092
	(0.005)	(0.004)	(0.004)	(0.073)	(0.081)	(0.065)	(0.061)
Controls		✓	✓		✓	✓	✓
Fixed Effects			✓			✓	✓
Wave Dummies		✓	✓		✓		✓
N	107736	107736	107736	107736	107736	100406	100406
First-Stage F				181.669	116.378	126.302	128.798

Notes: Second-stage coefficients of moving home, as instrumented by school choice, on various health outcomes. Each row represents a different outcome model, as indicated by the leftmost column. The school choice instrument consists of locally distance-weighted average school quality, interacted with the age of the youngest household, in the previous period. The Kleibergen-Paap F statistics from the relevant first-stage regressions are shown at the bottom of the table, indicating the relative strength of the instrument set in its predictive power of moving home.

APPENDIX B

Appendix to Chapter 3

Table B.1: Full set of coefficients from health outcome models

	SAH	GHQ	Health Cond.	> Good Health
Coast	-0.007	0.113	0.016**	0.005
	(0.016)	(0.080)	(0.007)	(0.007)
Age	0.018***	-0.032***	0.009***	-0.007***
	(0.000)	(0.002)	(0.000)	(0.000)
Married	-0.075***	0.392***	-0.044***	0.023***
	(0.010)	(0.053)	(0.004)	(0.004)
Retired	-0.394***	3.318***	-0.108***	0.112***
	(0.021)	(0.111)	(0.008)	(0.008)
Large employers & higher management	-0.523***	1.795***	-0.169***	0.176***
	(0.028)	(0.135)	(0.012)	(0.013)
Higher professional	-0.493***	1.855***	-0.150***	0.161***
	(0.023)	(0.115)	(0.010)	(0.010)
Lower management & professional	-0.412***	1.698***	-0.145***	0.124***
	(0.016)	(0.087)	(0.006)	(0.007)
Intermediate	-0.379***	1.712***	-0.149***	0.110***
	(0.018)	(0.097)	(0.007)	(0.008)
Small employers & own account	-0.438***	2.037***	-0.181***	0.124***
	(0.021)	(0.110)	(0.009)	(0.010)
Lower supervisory & technical	-0.336***	2.108***	-0.156***	0.091***
	(0.023)	(0.114)	(0.009)	(0.011)
Semi-routine	-0.331***	1.826***	-0.149***	0.082***
	(0.017)	(0.092)	(0.007)	(0.007)
Routine	-0.367***	2.190***	-0.176***	0.085***
	(0.020)	(0.111)	(0.008)	(0.009)
Other higher qualification	0.125***	-0.172**	0.019***	-0.055***
	(0.017)	(0.085)	(0.007)	(0.008)
A level etc	0.148***	-0.083	0.012**	-0.067***
	(0.015)	(0.074)	(0.006)	(0.007)
GCSE etc	0.196***	-0.028	0.009	-0.092***
	(0.015)	(0.076)	(0.006)	(0.007)
Other qualifications	0.293***	-0.313***	0.038***	-0.122***
	(0.020)	(0.102)	(0.008)	(0.008)
No qualifications	0.442***	-0.334***	0.050***	-0.157***
	(0.019)	(0.098)	(0.008)	(0.008)
ln(Income)	-0.057***	0.361***	-0.007**	0.027***
	(0.007)	(0.036)	(0.003)	(0.003)
Male	0.017*	0.895***	0.005	-0.005
	(0.010)	(0.050)	(0.004)	(0.004)
IMD Score	0.008***	-0.022***	0.001***	-0.003***
	(0.000)	(0.002)	(0.000)	(0.000)
White (vs "non-white")	-0.014	-0.451***	0.092***	0.026***
	(0.012)	(0.070)	(0.005)	(0.005)
Wave 2	0.021***	-0.193***	-0.010***	-0.007**
	(0.006)	(0.038)	(0.003)	(0.003)
Wave 3	0.007	-0.192***	-0.013***	0.004
	(0.007)	(0.042)	(0.003)	(0.003)
Wave 4	0.031***	-0.082*	-0.010***	-0.013***
	(0.007)	(0.044)	(0.003)	(0.004)
Wave 5	0.025***	-0.321***	-0.022***	-0.010***
	(0.008)	(0.046)	(0.003)	(0.004)
Constant	1.766***	24.862***	-0.063***	0.806***
	(0.024)	(0.126)	(0.009)	(0.010)
N	109666	109666	109666	109666

Notes: Point estimates of the coefficients from OLS models are reported. Standard Errors are in parentheses. The coast is defined using a donut, with households which reside within 5km of the coastline (as measured from the population-weighted centroid of the LSOA) =1, whilst those greater than 15km =0. The base category for the education variable is having a Degree. Statistical significance is denoted by: * $p<0.1$; ** $p<0.05$; *** $p<0.01$.

Table B.2: Full set of coefficients from risky health behaviour and benefits models

	Smoked	Alcohol	Physical Activity	Disability Benefits
Coast	0.026***	-0.001	0.028***	0.012**
	(0.009)	(0.009)	(0.008)	(0.005)
Age	0.002***	0.003***	-0.006***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
Married	-0.051***	0.015***	-0.023***	-0.032***
	(0.006)	(0.005)	(0.005)	(0.003)
Retired	-0.015	0.001	0.121***	-0.204***
	(0.011)	(0.010)	(0.009)	(0.007)
Large employers & higher management	0.027	0.062***	0.076***	-0.199***
	(0.018)	(0.017)	(0.017)	(0.005)
Higher professional	0.011	0.039***	0.071***	-0.189***
	(0.014)	(0.014)	(0.013)	(0.005)
Lower management & professional	0.038***	0.045***	0.064***	-0.190***
	(0.009)	(0.009)	(0.008)	(0.004)
Intermediate	0.018	0.015	0.051***	-0.179***
	(0.011)	(0.010)	(0.010)	(0.005)
Small employers & own account	0.040***	0.052***	0.033***	-0.212***
	(0.013)	(0.012)	(0.012)	(0.005)
Lower supervisory & technical	0.062***	-0.020	0.009	-0.201***
	(0.014)	(0.013)	(0.013)	(0.005)
Semi-routine	0.012	-0.024***	-0.002	-0.188***
	(0.009)	(0.008)	(0.008)	(0.004)
Routine	0.058***	-0.020*	-0.006	-0.205***
	(0.012)	(0.011)	(0.011)	(0.005)
Other higher qualification	0.040***	-0.031***	-0.029***	0.018***
	(0.010)	(0.009)	(0.009)	(0.004)
A level etc	0.019**	-0.048***	-0.052***	0.008**
	(0.009)	(0.008)	(0.008)	(0.003)
GCSE etc	0.057***	-0.072***	-0.087***	0.012***
	(0.009)	(0.008)	(0.008)	(0.004)
Other qualifications	0.058***	-0.093***	-0.107***	0.050***
	(0.011)	(0.010)	(0.009)	(0.006)
No qualifications	0.048***	-0.166***	-0.162***	0.085***
	(0.010)	(0.009)	(0.009)	(0.006)
ln(Income)	-0.026***	0.023***	0.036***	0.015***
	(0.004)	(0.004)	(0.004)	(0.002)
Male	0.118***	0.083***	0.072***	0.014***
	(0.006)	(0.005)	(0.005)	(0.003)
IMD Score	0.001***	-0.002***	-0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
White (vs "non-white")	0.301***	0.201***	0.026***	0.044***
	(0.007)	(0.006)	(0.006)	(0.003)
Wave 5	-0.031***	-0.119***	0.011***	-0.002
	(0.003)	(0.004)	(0.004)	(0.002)
Wave 2				0.002
				(0.002)
Wave 3				0.005***
				(0.002)
Wave 4				0.001
				(0.002)
Constant	0.137***	0.068***	0.597***	-0.060***
	(0.014)	(0.013)	(0.012)	(0.006)
N	38821	38821	38821	109666

Notes: Point estimates of the coefficients from OLS models are reported. Standard Errors are in parentheses. The coast is defined using a donut, with households which reside within 5km of the coastline (as measured from the population-weighted centroid of the LSOA) =1, whilst those greater than 15km =0. The base category for the education variable is having a Degree. Statistical significance is denoted by: * p<0.1; ** p<0.05; *** p<0.01.

Table B.3: Coefficients on the coastal variable: All Outcomes, fixed sample

Self-Assessed Health (SAH)	0.093***	0.016	0.016
	(0.022)	(0.020)	(0.020)
N	32520	32520	32520
GHQ Score	-0.083	0.072	0.071
	(0.108)	(0.105)	(0.105)
N	32520	32520	32520
L/Term Health Problem	0.058***	0.017*	0.017*
	(0.010)	(0.009)	(0.009)
N	32520	32520	32520
SAH: V. Good or Excellent	-0.031***	-0.005	-0.005
	(0.010)	(0.010)	(0.010)
N	32520	32520	32520
Ever Smoked	0.062***	0.015	0.015
	(0.010)	(0.010)	(0.010)
N	32520	32520	32520
Drank≥3 days last week	0.009	-0.011	-0.012
	(0.010)	(0.009)	(0.009)
N	32520	32520	32520
Freqent physical activity	-0.002	0.025***	0.025***
	(0.009)	(0.009)	(0.009)
N	32520	32520	32520
Disability Benefits	0.024***	0.006	0.006
	(0.006)	(0.005)	(0.005)
N	32520	32520	32520
Controls		✓	✓
Wave Dummies			✓

Notes: Point estimates of the coefficients from the Coast variable from OLS models are reported. Standard Errors are in parentheses. The coast is defined using a donut, with households which reside within 5km of the coastline (as measured from the population-weighted centroid of the LSOA) =1, whilst those greater than 15km =0. Statistical significance is denoted by: * p<0.1; ** p<0.05; *** p<0.01. The estimation sample is conditional on non-missingness for all outcomes and covariates.

Table B.4: Robustness of results to including the donut observations as an additional category

	SAH	GHQ	Health Cond.	> Good Health
Coast	-0.008	0.111	0.016^{**}	0.006
	(0.016)	(0.080)	(0.007)	(0.007)
off-coast (2.5-15km)	0.007	-0.043	0.012^{**}	0.006
	(0.013)	(0.066)	(0.006)	(0.006)
N	128704	128704	128704	128704

Coefficients from OLS models including those residing between 2.5km and 15km as an additional category in the coast variable. Cluster-robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.5: Robustness of results to including the donut observations as an additional category

	Smoked	Alcohol	Physical Activity	Disability Benefits
Coast	0.025^{***}	0.000	0.029^{***}	0.011^{**}
	(0.009)	(0.009)	(0.008)	(0.005)
off-coast (2.5-15km)	0.010	-0.014^{**}	0.011^{*}	0.017^{***}
	(0.007)	(0.007)	(0.006)	(0.004)
N	45384	45384	45384	147222

Coefficients from OLS models including those residing between 2.5km and 15km as an additional category in the coast variable. Cluster-robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.6: Estimated coefficients from a sample of latest responses, and Random Effects specifications

	SAH	GHQ	Health Cond.	> Good Health
Panel A: Coefficients from latest observed response (one observation per individual)				
Coast	-0.009	0.171*	0.013*	0.006
	(0.018)	(0.100)	(0.008)	(0.008)
N	37540	31456	37528	37540
Panel B: Coefficients Random Effects models				
Coast	-0.002	0.132*	0.017***	0.002
	(0.015)	(0.078)	(0.006)	(0.006)
N	127686	109666	127649	127686

Coefficients from OLS and Random Effects models * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.7: Estimated coefficients from a sample of latest responses, and Random Effects specifications

	Smoked	Alcohol	Physical Activity	Disability Benefits
Panel A: Coefficients from latest observed response (one observation per individual)				
Coast	0.006	0.004	0.030**	0.013**
	(0.014)	(0.014)	(0.014)	(0.005)
N	12990	10435	12944	37142
Panel B: Coefficients Random Effects models				
Coast	0.023***	-0.000	0.026***	0.008*
	(0.009)	(0.008)	(0.008)	(0.004)
N	48098	38821	47898	126283

Coefficients from OLS and Random Effects models * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table B.8: OLS results with working/retirement-coast interactions

Panel A: Health Outcomes				
	SAH	GHQ Score	Health Problem	V. Good Health
Coast	0.012	0.176*	0.014*	-0.005
	(0.018)	(0.091)	(0.008)	(0.008)
Retired	-0.385***	3.349***	-0.109***	0.108***
	(0.021)	(0.113)	(0.008)	(0.008)
Interaction	-0.073*	-0.245	0.007	0.037**
	(0.038)	(0.180)	(0.015)	(0.015)
N	127680	109662	127643	127680
Panel B: Risky health behaviours and disability benefits				
	Smoking	Drinking	Physical Activity	Disability Benefits
Coast	0.025**	-0.011	0.029***	0.012**
	(0.010)	(0.010)	(0.009)	(0.005)
Retired	-0.016	-0.003	0.122***	-0.204***
	(0.011)	(0.011)	(0.009)	(0.007)
Interaction	0.006	0.036*	-0.002	-0.001
	(0.019)	(0.019)	(0.017)	(0.012)
N	48094	38818	47894	126277

Notes: Point estimates of the coefficients from the Coast variable from OLS models, and its interaction with the retirement indicator, are reported. Cluster-robust standard Errors are in parentheses. The coast is defined using a donut, with households which reside within 2.5km of the coastline (as measured from the population-weighted centroid of the LSOA) =1, whilst those greater than 15km =0. Statistical significance is denoted by: * p<0.1; ** p<0.05; *** p<0.01.

APPENDIX C

Appendix to Chapter 4

C.1 Variable Definitions

C.1.1 Ethnicity

The 18 categories, subsequently collapsed to the 6 2001 ons census classifications, are as follows:

Table C.1: Ethnicity variable definition

LSYPE 18 category response	Census Classification
1. White - English/Welsh/Scottish/Northern Irish/British	1-4 : 1. White
2. White - Irish	
3. White - Gypsy or Irish Traveller	
4. Any other White background	
5. Mixed/multiple ethnic groups - White and Black Caribbean	5-8 : 2. Mixed
6. Mixed/multiple ethnic groups - White and Black African	
7. Mixed/multiple ethnic groups - White and Asian	
8. Any other mixed/multiple ethnic background	
9. Asian/Asian British - Indian	9 : 3. Indian
10. Asian/Asian British - Pakistani	10-11 : 4. Pakistani and Bangladeshi
11. Asian/Asian British - Bangladeshi	
14. Black/African/Caribbean/Black British - African	14-16 : 5. Black or Black British
15. Black/African/Caribbean/Black British - Caribbean	
16. Any other Black/African/Caribbean background	
12. Asian/Asian British - Chinese	12-13,17-18 : 6. Other Ethnic Group
13. Any other Asian background	
17. Other Ethnic group - Arab	
18. Any other ethnic group	

C.1.2 General Health Questionnaire

The respondents are asked the following questions:

1. Have you recently been able to concentrate on what you're doing?
2. Have you recently lost much sleep over worry?
3. Have you recently felt that you are playing a useful part in things?
4. Have you recently felt capable of making decisions about things?
5. Have you recently felt constantly under strain?

6. Have you recently felt you couldn't overcome your difficulties?
7. Have you recently been able to enjoy your normal day to day activities?
8. Have you recently been able to face up to your problems?
9. Have you recently been feeling unhappy or depressed?
10. Have you recently been losing confidence in yourself?
11. Have you recently been thinking of yourself as a worthless person?
12. Have you recently been feeling reasonably happy, all things considered?

To which they may respond:

0. Not at all
1. No more than usual
2. Rather more than usual
3. Much more than usual

C.1.3 Income in LSYPE

Income was imputed using interval regression (Stewart 1983). This method allowed us to impute a continuous value within a band, rather than assuming that all cases in a band had the same midpoint income. This was achieved using Stata's INTREG command (StataCorp 2007; Conroy 2005). INTREG fits a model of y = [dependent variable 1, dependent variable 2] on independent variables where the dependent variable 1 was the log lower income band and dependent variable 2 was log upper income band. The INTREG procedure also allowed us to impute all missing values on the income questions.
Note that the left-hand-side bound for the lowest band is Âč0 per week and the right-hand- side bound for the top band was fixed at Âč1,700 per week. The predictors are shown below.
5.4.3 Predictors of income in Sweep 8
Family circumstances and cohort members' (CM) characteristics in sweep 1:

1. Highest qualification held by main parent (sweep 1)
2. Employment status of main parent (sweep 1)
3. Social status NS-SEC of the family (sweep 1)
4. Marital status of main parent (sweep 1)
5. Cohort member's gender (sweep 1)
6. Cohort member's ethnic group (sweep 1)
7. Whether cohort member ever identified as having special educational needs sweep
1)
8. Government office region (GOR sweep 1)

Cohort

member's circumstances in sweep 7:

1. Housing tenure (sweep 7)
2. Current activity including education and employment (sweep 7)
3. Whether cohort member ever tried cannabis (sweep 7)
4. Month of interview (sweep 7)
5. Interview mode (sweep 7)

C.2 Full Tables of coefficients

Table C.2: Average Marginal Effects for all covariates from Life Satisfaction LPM models

	Males		Females	
Government Region				
North West	0.14*	0.14*	-0.07	-0.07
	(0.08)	(0.08)	(0.06)	(0.06)
Yorkshire and the Humber	0.14*	0.14*	-0.06	-0.06
	(0.08)	(0.08)	(0.06)	(0.06)
East Midlands	0.04	0.05	0.03	0.03
	(0.08)	(0.08)	(0.07)	(0.07)
West Midlands	0.13	0.13	-0.11*	-0.11*
	(0.08)	(0.08)	(0.06)	(0.07)
East of England	0.09	0.10	-0.07	-0.07
	(0.08)	(0.08)	(0.06)	(0.07)
London	0.16**	0.16**	-0.06	-0.06
	(0.07)	(0.07)	(0.06)	(0.06)
South East	0.03	0.03	-0.06	-0.06
	(0.08)	(0.08)	(0.06)	(0.06)
South West	0.05	0.05	-0.09	-0.08
	(0.08)	(0.08)	(0.07)	(0.07)
HH Income: £28,000-46,000	-0.00	-0.00	-0.01	-0.01
	(0.04)	(0.04)	(0.04)	(0.04)
HH Income: £46,000+	0.11***	0.11***	-0.00	-0.00
	(0.04)	(0.04)	(0.04)	(0.04)
Ethnicity				
Mixed	-0.02	-0.03	0.05	0.05
	(0.07)	(0.07)	(0.06)	(0.06)
Indian	0.01	-0.00	-0.07	-0.08
	(0.05)	(0.05)	(0.05)	(0.06)
Pakistani	-0.02	-0.03	-0.06	-0.06
	(0.06)	(0.07)	(0.06)	(0.07)
Bangladeshi	-0.06	-0.07	-0.16**	-0.17**
	(0.09)	(0.10)	(0.07)	(0.08)
Black Caribbean	-0.20**	-0.20**	0.01	0.01
	(0.10)	(0.10)	(0.08)	(0.08)
Black African	-0.15**	-0.16**	-0.15*	-0.15*
	(0.08)	(0.08)	(0.08)	(0.08)
Other	0.12	0.11	-0.17***	-0.18**
	(0.09)	(0.10)	(0.06)	(0.07)
GHQ-12 Score (wave 2)	0.00	0.00	-0.01**	-0.01**
	(0.01)	(0.01)	(0.00)	(0.00)
Parent has L/t health cond. (wave4)	-0.01	-0.01	-0.03	-0.03
	(0.04)	(0.04)	(0.04)	(0.04)
Parent employed (wave 4)	0.00	0.01	0.07*	0.07*

	(0.04)	(0.04)	(0.03)	(0.04)
No. of Siblings (wave 4)	-0.00	-0.00	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Parents Married & living together (wave 4)	0.03	0.03	0.04	0.04
	(0.04)	(0.04)	(0.03)	(0.03)
Attended preferred University	0.03	0.03	-0.01	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)
Read preferred course at Uni.	-0.02	-0.02	0.01	0.01
	(0.05)	(0.05)	(0.04)	(0.04)
Locus of Control (wave 2)	-0.00	-0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Locus of Control (wave 7)	-0.00	-0.00	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Locus of Control (wave 8)	-0.04***	-0.04***	-0.06***	-0.06***
	(0.01)	(0.01)	(0.01)	(0.01)
log(Income) at age 25		-0.03		-0.01
		(0.11)		(0.10)
N	985	985	1298	1298

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.3: Average Marginal Effects for all covariates from Life Satisfaction Probit models

	Males		Females	
Government Region				
North West	0.16*	0.16*	-0.06	-0.06
	(0.09)	(0.09)	(0.06)	(0.06)
Yorkshire and the Humber	0.16*	0.17*	-0.05	-0.05
	(0.09)	(0.09)	(0.06)	(0.06)
East Midlands	0.05	0.06	0.02	0.03
	(0.09)	(0.10)	(0.07)	(0.07)
West Midlands	0.14	0.15	-0.11*	-0.11*
	(0.09)	(0.09)	(0.06)	(0.06)
East of England	0.11	0.12	-0.06	-0.06
	(0.09)	(0.09)	(0.06)	(0.06)
London	0.18**	0.18**	-0.05	-0.05
	(0.08)	(0.08)	(0.06)	(0.06)
South East	0.05	0.05	-0.06	-0.05
	(0.09)	(0.09)	(0.06)	(0.06)
South West	0.07	0.08	-0.06	-0.06
	(0.09)	(0.09)	(0.07)	(0.07)
HH Income: £28,000-46,000	-0.00	0.00	-0.01	-0.01
	(0.04)	(0.04)	(0.03)	(0.04)
HH Income: £46,000+	0.11***	0.11***	0.00	0.00
	(0.04)	(0.04)	(0.04)	(0.04)
Ethnicity				
Mixed	-0.04	-0.05	0.06	0.05
	(0.07)	(0.08)	(0.05)	(0.06)
Indian	0.01	-0.00	-0.07	-0.07
	(0.04)	(0.05)	(0.05)	(0.06)
Pakistani	-0.02	-0.03	-0.05	-0.06
	(0.06)	(0.07)	(0.06)	(0.07)
Bangladeshi	-0.08	-0.09	-0.18**	-0.19**
	(0.10)	(0.11)	(0.08)	(0.09)
Black Caribbean	-0.26*	-0.27*	0.00	0.00
	(0.14)	(0.14)	(0.08)	(0.09)
Black African	-0.19**	-0.20**	-0.14*	-0.15
	(0.09)	(0.10)	(0.08)	(0.09)
Other	0.11	0.10	-0.19**	-0.20**
	(0.08)	(0.09)	(0.08)	(0.08)
GHQ-12 Score (wave 2)	0.00	0.00	-0.01***	-0.01***
	(0.01)	(0.01)	(0.00)	(0.00)
Parent has L/t health cond. (wave4)	-0.01	-0.01	-0.03	-0.03
	(0.04)	(0.04)	(0.04)	(0.04)
Parent employed (wave 4)	0.00	0.01	0.07*	0.07*
	(0.04)	(0.04)	(0.04)	(0.04)
No. of Siblings (wave 4)	-0.00	-0.00	0.01	0.01

	(0.01)	(0.01)	(0.01)	(0.01)
Parents Married & living together (wave 4)	0.03	0.04	0.04	0.05
	(0.04)	(0.04)	(0.03)	(0.03)
Attended preferred University	0.03	0.03	-0.01	-0.01
	(0.03)	(0.03)	(0.03)	(0.03)
Read preferred course at Uni.	-0.02	-0.02	0.01	0.01
	(0.05)	(0.05)	(0.04)	(0.04)
Locus of Control (wave 2)	-0.01	-0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Locus of Control (wave 7)	-0.00	-0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Locus of Control (wave 8)	-0.04***	-0.04***	-0.06***	-0.06***
	(0.01)	(0.01)	(0.01)	(0.01)
log(Income) at age 25		-0.04		-0.01
		(0.11)		(0.10)
N	985	985	1298	1298

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

APPENDIX D

Ethics Approval

Applicant: Jack Higgins
Supervisor: Bruce Hollingsworth
Department: Health Research
FHMREC Reference: FHMREC18025

22 October 2018

Dear Jack

Re: Essays on the Economics of Health and Place

Thank you for submitting your research ethics amendment application for the above project for review by the **Faculty of Health and Medicine Research Ethics Committee (FHMREC)**. The application was recommended for approval by FHMREC, and on behalf of the Chair of the Committee, I can confirm that approval has been granted for the amendment to this research project.

As principal investigator your responsibilities include:

- ensuring that (where applicable) all the necessary legal and regulatory requirements in order to conduct the research are met, and the necessary licenses and approvals have been obtained;
- reporting any ethics-related issues that occur during the course of the research or arising from the research to the Research Ethics Officer at the email address below (e.g. unforeseen ethical issues, complaints about the conduct of the research, adverse reactions such as extreme distress);
- submitting details of proposed substantive amendments to the protocol to the Research Ethics Officer for approval.

Please contact me if you have any queries or require further information.

Tel:- 01542 593987
Email:- fhmresearchsupport@lancaster.ac.uk

Yours sincerely,

R.E. Case

Becky Case
Research Ethics Officer, Secretary to FHMREC.

References

Abraido-Lanza, A. F., Dohrenwend, B. P., Ng-Mak, D. S., & Turner, J. B. (1999). The latino mortality paradox: a test of the" salmon bias" and healthy migrant hypotheses. *American journal of public health*, *89*(10), 1543–1548.

Ai, C., & Norton, E. C. (2003). Interaction terms in logit and probit models. *Economics letters*, *80*(1), 123–129.

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools. *Journal of political Economy*, *113*(1), 151–184.

Anders, J., et al. (2012). *Using the longitudinal study of young people in england for research into higher education access* (Tech. Rep.). Department of Quantitative Social Science-UCL Institute of Education.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Bambra, C., Munford, L., & Brown, H. (2018). *Health for wealth: Building a healthier northern powerhouse for uk productivity*. Newcastle.

Bansal, S., Thind, S., & Jaswal, S. (2006). Relationship between quality of home environment, locus of control and achievement motivation among high achiever urban female adolescents. *Journal of Human Ecology*, *19*(4), 253–257.

Benyamini, Y., Leventhal, E. A., & Leventhal, H. (2000). Gender differences in processing information for making self-assessments of health. *Psychosomatic Medicine*, *62*(3), 354–364.

Borjas, G. J. (2003). The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market. *The quarterly journal of economics*, *118*(4), 1335–1374.

Borjas, G. J. (2015). Immigration and globalization: A review essay. *Journal of Economic*

Literature, *53*(4), 961–74.

Brackney, B. E., & Westman, A. S. (1992). Relationships among hope, psychosocial development, and locus of control. *Psychological reports*, *70*(3), 864–866.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression discontinuity designs. *Stata Journal*, *17*(2), 372–404.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295–2326.

Card, D. (2009). Immigration and inequality. *American Economic Review*, *99*(2), 1–21.

Chen, J. (2011). Internal migration and health: re-examining the healthy migrant phenomenon in china. *Social Science & Medicine*, *72*(8), 1294–1301.

Cherlin, A., & Furstenberg Jr, F. F. (1992). *The new american grandparent*. Harvard University Press.

Chetty, R., Hendren, N., & Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, *106*(4), 855–902.

Chiburis, R. C., Das, J., & Lokshin, M. (2012). A practical comparison of the bivariate probit and linear iv estimators. *Economics Letters*, *117*(3), 762–766.

Christie, H. (2007). Higher education and spatial (im) mobility: nontraditional students and living at home. *Environment and Planning A*, *39*(10), 2445–2463.

Clampet-Lundquist, S., & Massey, D. S. (2008). Neighborhood effects on economic self-sufficiency: A reconsideration of the moving to opportunity experiment. *American Journal of Sociology*, *114*(1), 107–143.

Communities and Local Government Comittee. (2007). *Coastal towns*. HC 351.

Corfe, S. (2017). Living on the edge: Britain's coastal communities. *London: Social Market Foudnation*.

Corfe, S. (2020). *Britain's coastal communities amongst worst performing areas for earnings, jobs, health and education.* Social Market Foundation. Retrieved from `https://www.smf.co.uk/britains-coastal-communities-amongst-worst-performing-areas-earnings-jobs-health-education/`

Cullinan, J., & Gillespie, P. (2016). Does overweight and obesity impact on self-rated health? evidence using instrumental variables ordered probit models. *Health economics*, *25*(10), 1341–1348.

De Vries, S., Verheij, R. A., Groenewegen, P. P., & Spreeuwenberg, P. (2003). Natural environments-healthy environments? an exploratory analysis of the relationship between greenspace and health. *Environment and planning A*, *35*(10), 1717–1731.

de Gayardon, A., Callender, C., & Green, F. (2019). The determinants of student loan take-up in england. *Higher Education*, 1–19.

del Mar Salinas-Jiménez, M., Artés, J., & Salinas-Jiménez, J. (2013). How do educational attainment and occupational and wage-earner statuses affect life satisfaction? a gender perspective study. *Journal of Happiness Studies*, *14*(2), 367–388.

DeLongis, A., Folkman, S., & Lazarus, R. S. (1988). The impact of daily stress on health and mood: psychological and social resources as mediators. *Journal of personality and social psychology*, *54*(3), 486.

Department for Transport. (2015). *Journey times to key services by lower super output area.* (`https://www.gov.uk/government/statistical-data-sets/journey-times-to-key-services-by-lower-super-output-area-jts05`)

Department for Business, I., & Skills. (2016). Success as a knowledge economy: Teaching excellence, social mobility and student choice. *Higher Education White Paper*.

Donnelly, M., & Gamsu, S. (2018). Home and away: social, ethnic and spatial inequalities in student mobility.

Dustmann, C., Schönberg, U., & Stuhler, J. (2016). The impact of immigration: Why do studies reach such different results? *Journal of Economic Perspectives*, *30*(4), 31–56.

Easterlin, R. A. (2005). Building a better theory of well-being. *Economics and happiness: Framing the analysis*, 29–64.

Farré, L. (2016). New evidence on the healthy immigrant effect. *Journal of Population Economics*, *29*(2), 365–394.

Farwick, A. (2009). Internal migration-challenges and perspectives for the research infrastructure. *Available at SSRN 1452667*.

Finn, K. (2017). Multiple, relational and emotional mobilities: Understanding student mobilities in higher education as more than âĂŸstaying localâĂŹand âĂŸgoing awayâĂŹ. *British Educational Research Journal*, *43*(4), 743–758.

Fiori, K. L., Brown, E. E., Cortina, K. S., & Antonucci, T. C. (2006). Locus of control as a mediator of the relationship between religiosity and life satisfaction: Age, race, and gender differences. *Mental Health, Religion & Culture*, *9*(3), 239-263. doi:

10.1080/13694670600615482

Fitch, C., Hamilton, S., Bassett, P., & Davey, R. (2011). The relationship between personal debt and mental health: a systematic review. *Mental Health Review Journal*.

Foley, R., & Kistemann, T. (2015). Blue space geographies: Enabling health in place. *Health & place*, *35*, 157–165.

Foster, C. E., Horwitz, A., Thomas, A., Opperman, K., Gipson, P., Burnside, A., ... King, C. A. (2017). Connectedness to family, school, peers, and community in socially vulnerable adolescents. *Children and youth services review*, *81*, 321–331.

Future Leaders Trust. (2015). *Combatting isolation: Why coastal schools are failing and how headteachers are turning them around.* (http://www.future-leaders.org.uk/documents/187/Combatting_Coastal_Isolation_-_Future_Leaders_Trust.pdf)

Gamsu, S., Donnelly, M., & Harris, R. (2018). The spatial dynamics of race in the transition to university: Diverse cities and white campuses in uk higher education. *Population, Space and Place*, e2222.

Gascon, M., Triguero-Mas, M., Martínez, D., Dadvand, P., Forns, J., Plasència, A., & Nieuwenhuijsen, M. J. (2015). Mental health benefits of long-term exposure to residential green and blue spaces: A systematic review. *International journal of environmental research and public health*, *12*(4), 4354–4379.

Gibson, J., Stillman, S., McKenzie, D., & Rohorua, H. (2013). Natural experiment evidence on the effect of migration on blood pressure and hypertension. *Health Economics*, *22*(6), 655–672.

Grammar school expansion and faith school reforms: Damian hinds sets out his stall. (2018, May). Retrieved from https://www.tes.com/news/grammar-school-expansion-and-faith-school-reforms-damian-hinds-sets-out-his-stall

Greene, W. (2010). Testing hypotheses about interaction terms in nonlinear models. *Economics Letters*, *107*(2), 291–296.

Greene, W. H. (2012). *Econometric analysis*. Pearson Education India.

Han, E., Norton, E. C., & Powell, L. M. (2011). Direct and indirect effects of body weight on adult wages. *Economics & Human Biology*, *9*(4), 381–392.

Holton, M. (2015). Learning the rules of the âĂŸstudent gameâĂŹ: transforming the âĂŸstudent habitusâĂŹ through [im] mobility. *Environment and Planning A*, *47*(11), 2373–2388.

Holton, M. (2018). Traditional or non-traditional students?: incorporating uk studentsâĂŹ living arrangements into decisions about going to university. *Journal of Further and Higher Education*, *42*(4), 556–569.

Hosein, A. (2019). Girls' video gaming behaviour and undergraduate degree selection: A secondary data analysis approach. *Computers in Human Behavior*, *91*, 226–235.

Iacus, S. M., King, G., Porro, G., & Katz, J. N. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 1–24.

Idler, E. L., & Benyamini, Y. (1997). Self-rated health and mortality: a review of twenty-seven community studies. *Journal of health and social behavior*, 21–37.

Imbens, G. W. (2015). Matching methods in practice: Three examples. *Journal of Human Resources*, *50*(2), 373–419.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Johnson, E. J., & Taylor, E. J. (2018). *The long run health consequences of rural-urban migration.* Manuscript under review, University of Minnesota. Retrieved from `https://sites.google.com/umn.edu/jannaj/research`

Jones, A. M., Koolman, X., & Rice, N. (2006). Health-related non-response in the british household panel survey and european community household panel: using inverse-probability-weighted estimators in non-linear models. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *169*(3), 543–569.

Katz, L. F., Kling, J. R., & Liebman, J. B. (2001). Moving to opportunity in boston: Early results of a randomized mobility experiment. *The Quarterly Journal of Economics*, *116*(2), 607–654.

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica*, *75*(1), 83–119.

Krainer, J., et al. (2008). Falling house prices and rising time on the market. *FRBSF Economic Letter*.

Krauth, B. (2016). Bounding a linear causal effect using relative correlation restrictions. *Journal of Econometric Methods*, *5*(1), 117–141.

Lazarus, R. (1995). *Psychological stress in the workplace, occupational stress: a handbook.* London: Taylor & Francis.

Lefcourt, H. M. (1991). *Locus of control.* Academic Press.

Lu, Y. (2008). Test of the âĂŸhealthy migrant hypothesisâĂŹ: a longitudinal analysis of health selectivity of internal migration in indonesia. *Social science & medicine*, *67*(8), 1331–1339.

Lu, Y., & Qin, L. (2014). Healthy migrant and salmon bias hypotheses: a study of health and internal migration in china. *Social science & medicine*, *102*, 41–48.

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2013). Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American economic review*, *103*(3), 226–31.

Lynn, P., et al. (2009). Sample design for understanding society. *Underst. Soc. Work. Pap. Ser*, *2009*.

MacKinnon, D. P., Cheong, J., & Pirlott, A. G. (2012). Statistical mediation analysis.

Mahanta, D., & Kannan, V. (2015). Emotional maturity and adjustment in first year undergraduates of delhi university: An empirical study. *Indian Journal of Psychological Science*, *5*(2), 84–90.

Mare, R. D. (2011). A multigenerational view of inequality. *Demography*, *48*(1), 1–23.

McLanahan, S., & Percheski, C. (2008). Family structure and the reproduction of inequalities. *Annu. Rev. Sociol*, *34*, 257–276.

Mincer, J. (1978). Family migration decisions. *Journal of political Economy*, *86*(5), 749–773.

Mirowsky, J. (2017). *Education, social status, and health*. Routledge.

Moh'd, R. I., & Ajefu, J. B. (2017). Understanding the relationship between health and internal migration in the united kingdom. *International Journal of Migration, Health and Social Care*, *13*(4), 432–448.

Moran, P. B., & Eckenrode, J. (1991). Gender differences in the costs and benefits of peer relationships during adolescence. *Journal of Adolescent Research*, *6*(4), 396–409.

Morris, T., Manley, D., Northstone, K., & Sabel, C. E. (2017). How do moving and other major life events impact mental health? a longitudinal analysis of uk children. *Health & place*, *46*, 257–266.

Munford, L. A., Fichera, E., Sutton, M., et al. (2017). *Is owning your home good for your health? evidence from exogenous variations in subsidies in england* (Tech. Rep.). HEDG, c/o Department of Economics, University of York.

Munford, L. A., Sidaway, M., Blakemore, A., Sutton, M., & Bower, P. (2017). Associations of participation in community assets with health-related quality of life and healthcare usage: a cross-sectional study of older people in the community. *BMJ open*, *7*(2), e012374.

Norman, P., & Boyle, P. (2014). Are health inequalities between differently deprived areas evident at different ages? a longitudinal study of census records in england and wales, 1991–2001. *Health & place*, *26*, 88–93.

Norman, P., Boyle, P., & Rees, P. (2005). Selective migration, health and deprivation: a longitudinal analysis. *Social science & medicine*, *60*(12), 2755–2771.

Nowok, B., Van Ham, M., Findlay, A. M., & Gayle, V. (2013). Does migration make you happy? a longitudinal study of internal migration and subjective well-being. *Environment and Planning A*, *45*(4), 986–1002.

Office for National Statistics. (2012). *Population-weighted centroids.* (`http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/geography/products/census/spatial/centroids/index.html`)

ONS. (2014). *2011 census: Coastal communities.* (`http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/rel/census/2011-census-analysis/coastal-communities/index.html`)

OpenStreetMapData. (2016). *Coastlines | Data | OpenStreetMapData.* (`http://openstreetmapdata.com/data/coastlines`)

Organization, W. H., et al. (2017). Constitution of the world health organization: Principles. *Diperoleh pada*, *20*.

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, *37*(2), 187–204.

Ottaviano, G. I., & Peri, G. (2012). Rethinking the effect of immigration on wages. *Journal of the European economic association*, *10*(1), 152–197.

Palloni, A., & Arias, E. (2004). Paradox lost: explaining the hispanic adult mortality advantage. *Demography*, *41*(3), 385–415.

Parliament, H. o. (2019). *The future of seaside towns - hl paper 320*. Select Committee on Regenerating Seaside Towns and Communities.

Pasanen, T. P., White, M. P., Wheeler, B. W., Garrett, J. K., & Elliott, L. R. (2019). Neighbourhood blue space, health and wellbeing: the mediating role of different types of physical activity. *Environment international*, *131*, 105016.

Picard, R. (2010). *GEONEAR: Stata module to find nearest neighbors using geodetic distances.* Statistical Software Components, Boston College Department of Economics. (https://ideas.repec.org/c/boc/bocode/s457146.html,)

Rabe, B., & Taylor, M. P. (2012). Differences in opportunities? wage, employment and house-price effects on migration. *Oxford Bulletin of Economics and Statistics*, *74*(6), 831–855.

Roach, A. (2018). Supportive peer relationships and mental health in adolescence: An integrative review. *Issues in mental health nursing*, *39*(9), 723–737.

Rojas-Rueda, D., Nieuwenhuijsen, M. J., Gascon, M., Perez-Leon, D., & Mudu, P. (2019). Green spaces and mortality: a systematic review and meta-analysis of cohort studies. *The Lancet Planetary Health*, *3*(11), e469–e477.

Roodman, D. (2011). Fitting fully observed recursive mixed-process models with cmp. *The Stata Journal*, *11*(2), 159–206.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied*, *80*(1), 1.

Ruhm, C. J. (2018). *Shackling the identification police?* (Tech. Rep.). National Bureau of Economic Research.

Sachs, J. D. (2016). Toward an international migration regime. *American Economic Review*, *106*(5), 451–55.

Sage, J., Evandrou, M., & Falkingham, J. (2013). Onwards or homewards? complex graduate migration pathways, well-being, and the âĂŸparental safety netâĂŹ. *Population, Space and Place*, *19*(6), 738–755.

Salas, C. (2002). On the empirical association between poor health and low socioeconomic status at old age. *Health Economics*, *11*(3), 207–220.

Sanbonmatsu, L., Kling, J. R., Duncan, G. J., & Brooks-Gunn, J. (2006). Neighborhoods and academic achievement results from the moving to opportunity experiment. *Journal of Human resources*, *41*(4), 649–691.

Siciliani, L., & Verzulli, R. (2009). Waiting times and socioeconomic status among elderly europeans: evidence from share. *Health Economics*, *18*(11), 1295–1306.

Sonnenberg, C., Deeg, D., Van Tilburg, T., Vink, D., Stek, M., & Beekman, A. (2013). Gender differences in the relation between depression and social support in later life. *International psychogeriatrics*, *25*(1), 61–70.

Stein, J. C. (1995). Prices and trading volume in the housing market: A model with down-payment effects. *The Quarterly Journal of Economics*, *110*(2), 379–406.

Stock, J. H., & Yogo, M. (2002). *Testing for weak instruments in linear iv regression.* National Bureau of Economic Research Cambridge, Mass., USA.

Storrie, K., Ahern, K., & Tuckett, A. (2010). A systematic review: students with mental health problems-a growing problem. *International journal of nursing practice*, *16*(1), 1–6.

Thoits, P. A. (2010). Stress and health: Major findings and policy implications. *Journal of health and social behavior*, *51*(1_suppl), S41–S53.

Triguero-Mas, M., Dadvand, P., Cirach, M., Martínez, D., Medina, A., Mompart, A., ... Nieuwenhuijsen, M. J. (2015). Natural outdoor environments and mental and physical health: relationships and mechanisms. *Environment international*, *77*, 35–41.

Tunstall, H., Mitchell, R., Pearce, J., & Shortt, N. (2014). The general and mental health of movers to more-and less-disadvantaged socio-economic and physical environments within the uk. *Social science & medicine*, *118*, 97–107.

Verbeek, M., & Nijman, T. (1992). Testing for selectivity bias in panel data models. *International Economic Review*, 681–703.

Völker, S., & Kistemann, T. (2011). The impact of blue space on human health and well-being–salutogenetic health effects of inland surface waters: A review. *International journal of hygiene and environmental health*, *214*(6), 449–460.

Weldon, M. (2017). *School choice, competition and ethnic segregation in lancashire: evidence from structural models of two-sided matching* (Unpublished doctoral dissertation). Lancaster University.

Wheeler, B. W., Lovell, R., Higgins, S. L., White, M. P., Alcock, I., Osborne, N. J., ... Depledge, M. H. (2015). Beyond greenspace: an ecological study of population general health and indicators of natural environment type and quality. *International journal of health geographics*, *14*(1), 1.

Wheeler, B. W., White, M., Stahl-Timmins, W., & Depledge, M. H. (2012). Does living by the coast improve health and wellbeing? *Health & place*, *18*(5), 1198–1201.

White, M. P., Alcock, I., Wheeler, B. W., & Depledge, M. H. (2013). Coastal proximity, health and well-being: Results from a longitudinal panel survey. *Health & place*, *23*, 97–103.

Whittaker, W. (2012). The pursuit of happiness? subjective wellbeing and internal migration in great britain. Retrieved from `http://dx.doi.org/10.2139/ssrn.2234981`

Wood, S. L., Demougin, P. R., Higgins, S., Husk, K., Wheeler, B. W., & White, M. (2016). Exploring the relationship between childhood obesity and proximity to the coast: A rural/urban perspective. *Health & place*, *40*, 129–136.

Woodman, E., & McArthur, M. (2018). Young peopleâĂŹs experiences of family connectedness: Supporting social work practice with families and young people. *British Journal of Social Work*, *48*(3), 693–713.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Young, A. (2017). Consistency without inference: Instrumental variables in practical application. *Unpublished manuscript, London: London School of Economics and Political Science. Retrieved from: http://personal. lse. ac. uk/YoungA*.

Zeng, Z., & Xie, Y. (2014). The effects of grandparents on childrenâĂŹs schooling: Evidence from rural china. *Demography*, *51*(2), 599–617.