Essays on the relationship between labor markets and health

by

Joe Spearing

(Under the Direction of Meghan Skira)

Abstract

This dissertation answers three questions about the relationship between labor markets and health. In the first chapter, I examine the welfare effects of work restrictions in disability benefit programs. Work restrictions can discincentivize people with high work capability from applying and therefore control fiscal costs. However, they also distort the labor supply of claimants. I study a disability benefit program in the UK which does not feature work restrictions or means tests, and show that it is well-targeted by income and health and that marginal increases in generosity can be welfare-improving. I estimate a structural lifecycle model with endogenous benefit application, work, and asset accumulation decisions. I estimate this model using UK data and use it quantify the effect of benefit generosity and introduction of work restrictions on benefit claims, labor supply, and welfare. A revenue-neutral doubling of benefit generosity generates a 24% increase in the benefit take up rate, while overall welfare is improved by the equivalent of 1.9% to 2.2% of lifetime consumption, depending on educational attainment. Introduction of work restrictions in my model reduces overall welfare by 1.3% to 1.6% of lifetime consumption, depending on educational attainment. The second chapter estimates the causal effect of work-related autonomy on mental health. Using Understanding Society data from the UK, I exploit within-person, within-occupation variation in autonomy and explore the robustness of my results to assumptions about the degree of confoundedness of unobservables. I find low work-related autonomy adversely impacts mental health. Finally, I investigate heterogeneity across occupational characteristics in the effect of retirement eligibility on mental health in the UK. I use K-means clustering to define three occupational clusters, differing across multiple dimensions. I estimate the effect of retirement eligibility using a Regression Discontinuity Design, allowing the effect to differ by cluster. The effects of retirement eligibility are beneficial, and greater in two clusters: one comprised of white-collar jobs in an office setting and another of blue-collar jobs with high physical demands and hazards. The cluster with smaller benefits mixes blue- and white-collar uncompetitive jobs with high levels of customer interaction.

Index words: [Health, Labor markets, Disability, Disability benefit, Retirement, Mental health, Working conditions, Health inequality]

Essays on the relationship between labor markets and health

by

Joe Spearing

B.A., University of Oxford, UK, 2015

A Dissertation Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Doctor of Philosophy

Athens, Georgia

2024

©2024
Joe Spearing
All Rights Reserved

Essays on the relationship between labor markets and health

by

Joe Spearing

Major Professor:	Meghan Skira
Committee:	Joshua Kinsler Matthew Knepper

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
May 2024

Acknowledgments

I am grateful to Meghan Skira, for her feedback, time, support, advice, and encouragement, Joshua Kinsler, Matthew Knepper, Roozbeh Hosseini, Carolina Caetano, Brad Curtis, and Abigail Cormier for their generous feedback and input, and participants in the University of Georgia student seminar series, University of Georgia Health Economics seminar series, University of Georgia Macro Reading Group, ASHEcon 2023, CINCH-dggö 2023, SEA 2023, and I85 2023 for their constructive comments. Conversations with Cormac O'Dea, Brent Davis, Krzysztof Karbownik, Victoria Barone, Nicole Maestas, Marisa Carlos, Davide Dragone, and Stefanie Schurer were also valuable and helped significantly improve this work. Any remaining errors are my own.

Contents

Acknowledgments **iv**

List of Figures **vii**

List of Tables **ix**

1 The welfare effects of work restrictions for disability benefit claimants: evidence from the UK **1**

1.1 Introduction . . . 1
1.2 Disability Benefits in the UK . . . 4
1.3 Data . . . 6
1.4 Stylized facts about DLA/PIP . . . 9
1.5 Modeling reform of UK disability benefits . . . 14
1.6 Counterfactual simulations . . . 26
1.7 Conclusion . . . 28

2 The causal effects of workplace autonomy on mental health **51**

2.1 Introduction . . . 51
2.2 Data . . . 54
2.3 The relationship between mental health and autonomy . . . 57
2.4 Assessing omitted variable bias . . . 60
2.5 Reverse causality . . . 63
2.6 Discussion and conclusion . . . 65

3 The effect of retirement eligibility on mental health in the United Kingdom: heterogeneous effects by occupation **86**

3.1 Introduction . . . 86
3.2 Data . . . 87
3.3 Estimation . . . 89
3.4 Results . . . 91
3.5 Discussion and conclusion . . . 93

Appendices **102**

A The welfare effects of work restrictions for disability benefit claimants: evidence from the UK **102**
A.1 Results with discount rate fixed across education types . . . 102
A.2 Model solution . . . 105
A.3 Health process parameters . . . 106
A.4 Simulated versus data moments . . . 109

B The causal effects of workplace autonomy on mental health **112**
B.1 Mental health variable questions and scoring . . . 112
B.2 Robustness to different summary measures of work-related autonomy . . . 115
B.3 Robustness to inclusion of respondents who are not consistently in work . . . 118
B.4 Trajectory of mental health about changes in work-related autonomy . . . 118

C The effect of retirement eligibility on mental health in the United Kingdom: heterogeneous effects by occupation **123**
C.1 GHQ12 questions and sub-indices . . . 123
C.2 Institutional details on women's retirement ages . . . 124
C.3 Sensitivity of all-sample estimates to changes in specification . . . 126
C.4 Sensitivity of estimates by occupation type to changes in specification . . . 127
C.5 Standard error estimates . . . 127
C.6 Analysis using four occupational clusters . . . 138
C.7 Robustness tests . . . 143

Bibliography **146**

List of Figures

1.1 PIP/DLA cases in payment 40
1.2 Sources of household income, claimants and non-claimants 41
1.3 Model fit of transitions between health states 42
1.4 Model fit of average wages by education and health status 43
1.5 Asset profile by education 44
1.6 Labor market participation across the life cycle 45
1.7 Life cycle of benefit applications 46
1.8 Life cycle of benefit claims 47
1.9 Health status over the life cycle 48
1.10 Asset accumulation by claimant status 49
1.11 The welfare effects of combining work restrictions and increases in benefits 50

2.1 Conditional associations between the GHQ12 caseness score and low work-related autonomy amongst men 77
2.2 Conditional associations between the SF12 index and low work-related autonomy amongst men 78
2.3 Conditional associations between the GHQ12 caseness score and low work-related autonomy across different dimensions amongst men 79
2.4 Conditional associations between the SF12 index and low work-related autonomy across different dimensions amongst men 80
2.5 Conditional associations between the GHQ12 caseness score and low work-related autonomy amongst women 81
2.6 Conditional associations between the SF12 index and low work-related autonomy amongst women 82
2.7 Conditional associations between the GHQ12 caseness score and low work-related autonomy across different dimensions amongst women 83
2.8 Conditional associations between the SF12 index and low work-related autonomy across different dimensions amongst women 84
2.9 Association between having low workplace autonomy and each value of the caseness score 85

3.1 Percent of people in each cluster scoring high minus percent scoring low 101

A.1 Asset accumulation by education in the model where the discount rate is fixed at 0.98 for all education types. 103
A.2 Welfare effects of reforming disability benefits in the model where the discount rate is fixed at 0.98 for all education types. 104
A.3 Estimated versus simulated health transitions, over 50s 109
A.4 Estimated versus simulated log wages persistence 110
A.5 Estimated versus simulated variance in log wages 111

B.1 The evolution of mental health around a transition between high and low autonomy work 122

C.1 Effect of retirement eligibility on GHQ12 caseness across specifications 128
C.2 Effect of retirement eligibility on Anxiety and Depression across specifications 129
C.3 Effect of retirement eligibility on Loss of Confidence across specifications 130
C.4 Effect of retirement eligibility on Social Dysfunction across specifications 131
C.5 Effect of retirement eligibility on GHQ12 caseness across occupational clusters across different specifications 132
C.6 Effect of retirement eligibility on Anxiety and Depression across occupational clusters across different specifications 133
C.7 Effect of retirement eligibility on Loss of Confidence across occupational clusters across different specifications 134
C.8 Effect of retirement eligibility on Social Dysfunction across occupational clusters across different specifications 135
C.9 Within-cluster sum of squares at different number of clusters 139
C.10 Percent of people in each cluster scoring high minus percent scoring low, four clusters . 140

List of Tables

1.1 Comparing Disability Living Allowance and Personal Independence Payments 30
1.2 Share of claimants by primary disability 31
1.3 Summary statistics by claimant status, men and women 32
1.4 Percentage of the population claiming and share of benefit spending by health status and income quintile 33
1.5 Associations between working and claiming benefits 34
1.6 Value of key paramters, sufficient statistics approach 34
1.7 The welfare effect of marginal increases in benefit generosity 35
1.8 Data moments used to identify coefficients using Simulated Method of Moments . . . 35
1.9 Structural parameters 36
1.9 Structural parameter estimates 37
1.9 Structural parameter estimates 38
1.10 Unmatched moments: benefit transitions and claimant labor force participation 38
1.11 Effects of policy reform 39

2.1 Summary statistics of key variables 68
2.2 Autonomy scores across different occupations 69
2.3 The effect of low autonomy on the mental health of men 70
2.4 The effect of low autonomy on the mental health of women 71
2.5 Causal effects of low autonomy on the GHQ12 caseness of men under different assumptions 72
2.6 Causal effects of low autonomy on men's SF12 under different assumptions 73
2.7 Causal effects of low autonomy on women's GHQ12 caseness under different assumptions 74
2.8 Causal effects of low autonomy on women's SF12 under different assumptions 75
2.9 The effect of low autonomy on the mental health of those who never change employers 76

3.1 O*NET measure of contact with people at work for CEOs, August 2020 94
3.2 O*NET indices and related variables 95
3.2 O*NET indices and related variables 96
3.3 Sample summary statistics 97
3.4 Share of occupational employment by cluster 98
3.5 Patterns of income and health amongst those working in different occupational clusters 98

3.6 Optimal bandwidths for each outcome 99
3.7 The effect of retirement eligibility by occupational cluster (using RDD) 100

A.1 Effects of policy reform where the discount rate is fixed across education 105
A.2 Health process parameter estimates 107
A.2 Health process parameter estimates 108

B.1 Scoring system for SF12 indices 114
B.2 Robustness of the Causal Effects of Low Overall Autonomy on the Mental Health of Men to Different Cutoffs 116
B.3 Robustness of the causal effects of low overall autonomy on the Mental Health of Women to Different Cutoffs 117
B.4 Robustness of the causal effect of low autonomy on the mental health of men using the full sample of workers 119
B.5 Robustness of the causal effect of low autonomy on the mental health of women using the full sample of workers 120

C.1 Women's retirement date by date of birth. 125
C.1 Women's retirement date by date of birth. 126
C.2 Standard errors using different calculations, for each dependent variable 137
C.3 Share of occupational employment by cluster 141
C.4 The effect of retirement by occupational cluster (four clusters) 142
C.5 The effect of retirement eligibility by occupational cluster, excluding wave fixed effects . 144
C.6 The effect of retirement eligibility by occupational cluster amongst people who work until the SPA 145

Chapter 1

The welfare effects of work restrictions for disability benefit claimants: evidence from the UK

1.1 Introduction

Poor health and disability are significant predictors of low income and low asset accumulation (De Nardi et al., 2022, Hosseini et al., 2021). Accordingly, many developed countries have large and growing disability benefits programs designed to insure against the economic risks associated with poor health. Most disability benefits programs impose work restrictions of some kind on claimants. For example, claimants in the US lose benefits if they earn income above the "substantial gainful activity" threshold.[1] Means-testing can also act as an effective work restriction if claimants lose benefits at relatively low levels of earnings, e.g. ,in Norway, a claimant loses all disability benefits when they earn more than 80% of their pre-disability earnings. A large literature establishes that disability benefit programs with work restrictions distort the labor supply of claimants (e.g., Maestas et al., 2013, French and Song, 2014, Kostol and Mogstad, 2014, Moore, 2015, Gelber et al., 2017, Markussen et al., 2018, Ahammer and Packham, 2022). This distortion could be especially costly for marginal applicants, who may be discouraged from applying if they are able to work or expect to be able to in the future. On the other hand, work restrictions may improve the targeting of disability benefit programs and reduce *ex post* moral hazard by discouraging those with high work capability from applying, and encouraging claimants whose health recovers to exit the program. The resulting reduction in fiscal cost may allow for a reduction in taxes. The effect of work restrictions in disability benefit programs on population labor supply is, therefore, theoretically ambiguous.

In this paper, I examine the effects of work restrictions in disability benefit programs on labor supply, benefit claims, and overall population welfare. One difficulty in credibly assessing the effects of work

[1] In 2023 this was USD 1,470 per month.

restrictions is that most studies focus on disability benefit programs which include work restrictions, and we therefore have little evidence about the behavior of agents in a program without work restrictions. I examine disability benefits in the UK— Disability Living Allowance (DLA) and Personal Independence Payments (PIP)— which do not impose work restrictions or means tests. Eligibility to claim these benefits depends only on the result of a health assessment.

I first provide a rigorous descriptive analysis of the functioning of these benefit programs. I use data from the British Household Panel Survey/Understanding Society (BHPS/US), administrative data on benefit applications, the English Longitudinal Survey of Ageing (ELSA), the Family Resources Survey (FRS), and the Living Costs and Food Survey (LCF). Overall, these data provide a rich picture of income sources, health status, consumption, asset accumulation, and application decisions for the UK population, broken down by claimant status, age, and sex. The descriptive analysis shows that despite the lack of work restrictions, the majority of public spending on these benefits goes to individuals in the bottom half of the household income distribution. I also provide evidence that although the labor supply of claimants is low, transitions on and off PIP or DLA are associated with small changes in labor supply relative to estimates of the effect of US disability insurance on labor supply. Finally, using a "sufficient statistics" approach as in Deshpande and Lockwood (2022), I estimate the marginal social value of increasing benefit generosity. Specifically, under assumptions about agents' utility functions, and the distorting effect of taxation on labor supply, I simulate a marginal increase in benefits and an income tax increase which leaves agents *ex ante* welfare-neutral, i.e., their expected utility before the realization of income and health shocks unchanged. Under all assumptions, the resulting change in government revenue is positive, suggesting that there is substantial insurance value from reducing the dispersion of consumption between claimants and non-claimants. This result demonstrates that a disability benefit program which does not have work restrictions or means tests can generate high insurance value relative to its distorting effect.

In order to formally evaluate the role and overall effect of work restrictions on welfare, I develop a structural life cycle model where agents make decisions about benefit applications, labor supply, and asset accumulation. My model includes several channels to generate income-health and asset-health gradients. Agents face exogenous wage offer and health risks, which differ by exogenous educational attainment. I allow discount rates to differ by education, reflecting heterogeneity in preferences which can be correlated with health. I also allow wages and the utility cost of work to differ by health status. Agents smooth consumption through labor supply, asset accumulation and benefit application decisions. Agents have preferences over consumption, and face a disutility from working and applying for benefits. I estimate the model using Simulated Method of Moments to target life cycle profiles of wages, work choices, asset accumulation, benefit applications and benefit claims by health status for UK men. The estimated model closely matches health transitions, wages and labor supply by health and education, and asset accumulation by education. The estimates reveal significant health penalties on wages. Agents face disutility from applying for benefits which is higher than the disutility of working, suggesting some combination of high stigma costs, hassle costs or aversion to the health assessment. My estimates also imply that health assessments are effective: the probability of successful application is 79% for those in bad health and less than 1.5% for those in good or fair health.

The descriptive analysis suggests that benefits are well-targeted in the UK even in the absence of work restrictions. However, the result might be driven by the relatively low monetary value of benefits. It is possible that as benefits become more generous, more applications will be induced and the lack of work restrictions will lead to poor targeting. I test this by using the structural model to simulate a counterfactual where the monetary value of benefits are doubled, along with a corresponding tax increase. I impose tax changes which vary by education so that the policy is revenue-neutral within each education group, which allows me to partially separate the redistributive effects of reform from efficiency changes. Benefit claims increase by only 24%, and there is an increase in average welfare for all education groups equivalent to between a 1.9% and 2.1% increase in lifetime consumption. The key mechanism which drives this result is the high utility cost of applying, which means that this policy change primarily increases payments to existing claimants. This fact suggests that health assessments with a high nuisance cost or stigma contribute to the targeting of the program.

I then simulate a counterfactual that imposes work restrictions by reducing benefit payments to zero for all claimants who work and removing them from benefits in subsequent periods.[2] Again I impose tax changes so that net revenue is unchanged within each education group. Imposing work restrictions leads to a reduction in the number of claimants by 47%, which is consistent with work restrictions reducing the value agents place on disability benefits. All education groups experience an average welfare loss from this change, equivalent to losing between 1.3% and 1.6% of lifetime consumption. Furthermore work restrictions decrease welfare even when coupled with higher levels of benefits. My results therefore indicate that, at least in the UK context, the benefit of reducing moral hazard and improving targeting is smaller than the cost of reducing the insurance value of disability benefits and distorting claimants' labor supply.

This paper contributes to two strands of literature. Firstly, my work contributes to the literature which investigates the value and optimal design of disability benefit programs. Much of this literature uses structural models to quantify the value of disability benefit and assess the effects of reforming disability benefit (Low and Pistaferri, 2015, Michaud and Wiczer, 2018, Autor et al., 2019, Hosseini et al., 2021, Kellogg, 2022). In addition, Haller, Staubli, and Zweimüller (2020) use a sufficient statistics approach to assess the effect of increasing leniency in applications and monetary payments. I contribute to this literature by considering the design of disability benefit across a different dimension: the inclusion of work restrictions. Research has investigated the effect of varying the effective tax rate on claimants' incomes in a structural setting (e.g., Hoynes and Moffitt, 1999, Benıtez-Silva et al., 2006, Dal Bianco, 2022) or exploiting existing reforms (e.g., Campolieti and Riddell, 2012, Krekó et al., 2022). My work differs from these papers by comparing disability benefit with and without work restrictions altogether, instead of varying the degree of restrictions within a small range.[3] This is an important contribution, because removal

[2] I follow Hosseini, Kopecky and Zhao (2021) in approximating work restrictions as a prohibition on all paid work.

[3] To my knowledge, the only research which studies disability benefit programs with no means-tests or work restrictions is Silver and Zhang (2022) who estimate the effect of receiving benefits from the VA's disability compensation program in the US on health outcomes and behaviors. They find effects on healthcare utilization but very small effects on health outcomes. My work differs in that I focus on the welfare effects of a similar program with national eligibility.

of work restrictions altogether likely has a much larger impact on the value of disability benefit to potential applicants.

I also contribute to the literature which studies the UK disability benefit system. There are multiple disability benefits in the UK, and most of the existing literature considers benefits such as Employment Support Allowance (ESA), which includes work restrictions (see Section 1.2 for further details). Ball and Low (2014) assess the relative role of self-insurance and ESA in protecting the consumption of disabled people. Low and Pistaferri (2020) quantify the tradeoff between the distortion to labor supply and insurance value of disability benefits which include work restrictions in a UK context. Dal Bianco (2022) models the behavior of older disabled people claiming ESA up to the year 2008 and assesses policy reforms which may improve labor market participation. I contribute by quantitatively studying DLA and PIP, disability benefits which do not impose work restrictions on their claimants. This contribution is important for two reasons: firstly, receipt of DLA and PIP has experienced a steep rise since 2010, and now represent a larger caseload than benefits such as ESA which do impose work restrictions; secondly, since DLA and PIP are very different to both ESA and disability benefit programs in other countries, findings about disability benefit programs in other countries are less likely to carry over to DLA and PIP.

This paper begins by explaining the key characteristics of DLA and PIP in section 1.2. Section 1.3 describes the data sources used in my analysis. Section 1.4 provides key descriptive evidence about DLA and PIP. Section 1.5 presents the structural model and estimation procedure. Section 1.6 describes the results of counterfactual simulations. Section 1.7 concludes.

1.2 Disability Benefits in the UK

The focus of this paper is on DLA and PIP. However, the UK welfare state includes a number of benefits for people who are disabled or otherwise in poor health.

1.2.1 Disability Living Allowance and Personal Independence Payments

Introduced in 1992, DLA provides financial support for those with financial needs owing to a health condition or disability. Those under 65 are eligible to claim based on their health state.

Applicants for benefits are assessed on their health across two dimensions: "mobility", which includes whether an applicant can walk, and "care", which includes aspects such as frequency of required supervision and ability to prepare food. Mobility scores are discretized into two levels and care scores are discretized into three. An applicant's eligibility and total payment conditional on successful application are based on their care and mobility scores. Assessors have the discretion to make indefinite awards (when a person is unlikely to recover) or definite awards, where reassessment is required at the end of the award term to continue claiming. Automatic qualification without an assessment is given to those with terminal or certain chronic conditions (such as blindness). Assessments can be made based on evidence from an applicant's physician or based on an in-person health assessment.

DLA does not impose any restrictions on a claimant's economic activity. Eligibility is unrelated to labor market status and benefits are not means-tested or asset-tested. Payments depend entirely on a claimant's care and mobility scores, rather than past or current earnings or tax payments.

Due to concerns about the overall level of public spending, the UK government introduced a series of reforms to the welfare state which included phasing out DLA for most people and replacing it with a new benefit, PIP, starting in April 2013.[4] PIP is similar to DLA but with tighter eligibility criteria and more frequent assessments. Table 1.1 compares the two benefits. Key differences are that the "care" component was streamlined into a "daily living" component, taking two possible levels instead of three, and chronic illnesses were removed as sources of automatic qualification. The Department for Work and Pensions (DWP) also made important discretionary changes about how applications are processed and awards made: namely fewer indefinite awards, and mostly requiring in-person assessments (Royston, 2017).

The key features of PIP and DLA are otherwise similar. Both benefits are paid regardless of income, dependent only on health status, with no means-testing or work restrictions. The monetary payments are also nearly identical (the only difference being that the lowest tier of "care" payments was phased out under PIP). For most of this paper, I treat DLA and PIP as being substantively the same. Similarly, studies often model US disability insurance over a period of time without explicitly incorporating reforms, such as changes in eligibility by different illnesses (e.g., Low and Pistaferri, 2015). PIP is a more stringent benefit and claimants therefore have worse health on average than DLA claimants. My estimates of the effect of work restrictions on welfare, labor supply, and claims therefore likely tend to understate the effect of work restrictions on DLA claimants and overstate the effect on PIP claimants.

Figure 1.1 plots the number of active claims each year for both DLA and PIP since May 2002.[5] The number of DLA claimants rises steadily from May 2002 to 2013, where it begins declining as new claimants must apply for PIP and DLA claimants are reassessed for PIP. Overall, the total number of claimants (DLA and PIP) increases as the new benefit is introduced. In August 2022, there were 4.4 million claimants overall, over 6% of the UK population. Another notable feature of the data is that while up to the introduction of PIP, men and women are roughly equally likely to claim benefits, women are more likely to claim PIP. The most likely explanation is that women were more likely to be reassessed for PIP, consistent with women being more likely to self-select for reassessment. Throughout the period, however, men make up between 49% and 51% of total claimants.

Table 1.2 shows the share of claimants categorized by major disabling condition in August 2022. The four most common primary disabilities represent 82% of claimants and are: psychiatric disorders, general musculoskeletal diseases, neurological diseases, and regional musculoskeletal diseases. These conditions

[4] PIP was phased in from April 2013, with new applicants over the age of 16 required to apply for PIP instead of DLA. Existing DLA claimants under the age of 65 began to be reassessed for benefits and either transitioned onto PIP or were removed from benefit rolls. The transition occurred first when DLA claimants had changes of circumstances or a definite award expired, and then long-term claimants were randomly selected to be reassessed for PIP. Throughout, claimants were entitled to self-select to be reassessed for the new benefit. Those over the age of 65 in April 2013 were allowed to continue claiming DLA.

[5] This is the full available history for active claims provided by the DWP.

also constitute a large caseload in the US, where musculoskeletal disorders represent 30% of disability insurance claims, and mental disorders 35%.[6]

1.2.2 Other disability benefits in the UK

In addition to PIP and DLA, there are a number of other benefits available for those in the UK who are disabled or otherwise in poor health. Those with a disability or health condition affecting their ability to work who have previously worked and paid sufficient taxes can claim Employment Support Allowance (ESA). ESA imposes strict conditions on claimants' right to do paid work. However, those who are assessed as able to prepare for work (for example, writing a CV) are required to prepare for work as a condition of receiving ESA. Additionally, low-income people in the UK can claim Universal Credit (UC). UC is an all-purpose means-tested benefit which pays households benefits based on their income. Those claimants who can demonstrate a health condition which prevents them from working see an increase in their benefits.[7] As with ESA, those judged able to prepare for work are required to do so as a condition of receiving their benefits. Both ESA and UC were introduced as replacements to similar benefits, in 2008 and 2013 respectively. Claimants may receive both DLA/PIP and ESA/UC. One important consideration is that although DLA and PIP impose no work restrictions on claimants, DLA and PIP claimants may claim other benefits which affect their legal right to work. 21.6% of DLA or PIP claimants also claim ESA, and 2.7% of DLA or PIP claimants also claim UC,[8] suggesting a minority of DLA and PIP claimants are affected in this way. Claiming cash benefits has no impact on access to healthcare via the National Health Service.

The relative importance of these benefits has changed over time as eligibility criteria have changed. Currently, there are 1.7 million UC claimants with eligibility for some health-related component of the benefit, and 1.6 million ESA claimants. By contrast, there are currently 4.4 million DLA and PIP claimants. Disability benefit programs which do not come with work restrictions therefore represent a larger caseload than disability benefit programs which do have work restrictions.

1.3 Data

I use several data sources which provide a detailed picture of the income, asset accumulation, consumption, and health of claimants and non-claimants. In this section, I describe these data sources and the key variables in my analysis.

[6] Author's calculations from the Annual Statistical Report on the Social Security Disability Insurance Program, 2021: https://www.ssa.gov/policy/docs/statcomps/di_asr/index.html.

[7] The maximum payment in 2022 was GBP 390.06 per month.

[8] Author's calculations from Understanding Society data set, using data from 2016 to 2021.

1.3.1 British Household Panel Survey/Understanding Society

The British Household Panel Survey/Understanding Society is a representative panel data set of UK households, spanning the years 1991 to 2021.[9] Surveyed individuals are interviewed annually. Data include detailed information about employment, income sources, health status, education, and demographic characteristics.

Monthly labor income is reported as the total income in the month before the survey is conducted. Monthly household income is also reported and decomposed into labor income, social benefit income, private benefit income,[10] investment income, pension income, and other income.

My primary measure of health is self-reported health, reported as "excellent", "very good", "good", "fair", or "poor". Self-reported health is commonly used as an overall summary measure of health status in the literature (e.g., Blundell et al., 2017, De Nardi et al., 2022, Borella et al., 2019). Further, self-reported health is an effective predictor of successful application for disability insurance in the US (Benítez-Silva et al., 2004). The survey additionally reports when respondents are deceased between waves.

The survey reports receipt of all state benefits by benefit and total income from each benefit. Unfortunately, in the BHPS/US data set, benefit applications are not observed.

Education is reported as the highest qualification ever achieved, from one of the following: degree or equivalent, other higher degree, A-level or equivalent, GCSE or equivalent,[11] other qualification, no qualification. Other qualifications are mostly non-academic qualifications.

I construct a sample from the BHPS/US data by taking all observations from all waves since 1991, excluding proxy responses. This means that my sample includes attritors and new respondents who were recruited to increase the sample size and maintain representativeness. This sample allows me to observe benefit claiming behavior across the life cycle and compare the labor market behavior of claimants and non-claimants.

1.3.2 Department for Work and Pensions Administrative data

I complement the survey data with administrative data.[12] The data report the number of DLA and PIP applicants by age and sex; the number of claims by age, sex, and main disabling condition; and the number of assessments and reassessments by age, sex and local authority. These data allow me to discern patterns of applications across the life cycle by sex and patterns of applications and application outcomes by local authority. On the other hand, I am unable to infer the self-reported health of applicants or the success of applications by self-reported health status from these data. I cannot construct a mapping of main

[9] University of Essex, Institute for Social and Economic Research. (2022). Understanding Society: Waves 1-12, 2009-2021 and Harmonised BHPS: Waves 1-18, 1991-2009. 17th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-18.

[10] Private benefit income includes sickness benefit and unemployment insurance payments paid by trade unions, friendly societies, and similar organizations.

[11] A-levels are national exams taken at the age of 18; GCSE (General Certificate of Secondary Education) exams are taken at the age of 16.

[12] DWP data can be accessed at https://stat-xplore.dwp.gov.uk/webapi/jsf/login.xhtml.

disabling condition to self-reported health for two reasons. Firstly, knowing a person's main disabling condition does not, in general, provide enough evidence to infer self-reported health. For example, Blundell, Borella, Commault and De Nardi (2023) require knowledge of all reported health conditions to link health conditions to self-reported health, because a person who has multiple health conditions is expected to have worse overall health. Secondly, the conditions in the DWP data do not map well onto conditions reported in BHPS/US, which reports self-reported health.

1.3.3 Other survey data sets

I use additional data sources for information not reliably avilable from BHPS/US such as assets, savings and consumption.

For consumption, I use the Living Costs and Food Survey,[13] which is a representative survey of UK households that gives detailed information about income and consumption. Income is broken down by source, allowing me to see which benefits a household claims. I am able to identify how consumption differs by whether there is a claimant in the household and age of head of household.

For savings, I use the Family Resources Survey,[14] which is a representative survey of UK households with a detailed breakdown of income from different sources, including benefits. I observe total savings per household, and I can break this down by claimant status, self-reported health status, and age of people within the household. I use savings by health status at the age of 26 in the Family Resources Survey to set initial conditions for capital accumulation in the structural model.

In later life, housing wealth is an important component of assets. Since the Family Resources Survey does not provide detailed information about the value of housing wealth, I use the English Longitudinal Survey of Ageing (ELSA)[15] to construct asset profiles for those over 50. ELSA is a representative survey of over 50s, which includes a valuation of household assets that can be broken down by age, self-reported health, and claimant status. One drawback of this data is that I cannot infer asset accumulation before the age of 50.

I report all financial variables in 2015 pounds using the CPI index,[16] express all household-level financial variables in equivalized terms,[17] and remove proxy respondents and observations which do not report health status. Key descriptive information is presented in Table 1.3.[18] DLA and PIP claimants are marginally older than non-claimants, significantly less likely to work, tend to be less educated, are in worse health, and tend to have lower income and assets. PIP and DLA claimants are broadly similar, although household income is higher for PIP claimants.

[13] Office for National Statistics. (2019). Living Costs and Food Survey. 3rd Release. UK Data Service. SN: 2000028

[14] Department for Work and Pensions, NatCen Social Research. (2021). Family Resources Survey. 4th Release. UK Data Service. SN: 200017

[15] NatCen Social Research, University College London, Institute for Fiscal Studies. (2023). English Longitudinal Study of Ageing. 6th Release. UK Data Service. SN: 200011

[16] e.g., Office for National Statistics (ONS), released 19 April 2023, ONS website, Statistical Bulletin.

[17] I use the OECD's equivalization index, which is equal to one if there is one adult in the household, and for each additional adult increases by 0.7. For each child, the index increases by 0.5. The index is intended to reflect economies of scale in household size and the different consumption demands of adults and children.

[18] Table 1.6 reports consumption data from the Living Costs and Food Survey.

1.4 Stylized facts about DLA/PIP

1.4.1 How redistributive are DLA and PIP?

One might be concerned that in the absence of work restrictions, PIP and DLA will be poorly targeted at those with low income. Table 1.4 investigates the pattern of claims by health and income quintile using data from BHPS/US. In Panel A, I report the probability that a person claims either PIP or DLA conditional on self-reported health status and real equivalized household income quintile. I draw three conclusions: firstly, conditional on household income, the probability of claiming increases as self-reported health worsens. This is not surprising, but it suggests that health assessments are effective in screening out healthy applicants.

Secondly, there is a notable income gradient in claimant status. Conditional on health status, the probability of claiming tends to decline in income. There are no features of either program which explicitly deliver this result. A high-income person in poor health is as eligible to claim benefits as a low-income person in poor health. One explanation for the gradient is that there is a cost to benefit applications—a stigma cost, a nuisance cost, a cost of gathering information, or a psychological cost of undergoing the health assessment—and this cost is more likely to outweigh the marginal utility of consumption for individuals in high-income households because of declining marginal utility of consumption. Alternatively, there may be health differences not captured by self-reported health between high-income and low-income respondents: high-income people who report being in poor health may be on average less likely to suffer from chronic conditions than low-income people who report being in poor health. This disparity could result from comparisons with different peer groups, i.e., high-income people may have a downward-biased estimate of the percentage of people with chronic conditions which might lead them to be more likely to report being in poor health when they are objectively relatively healthy compared to the population average.

Thirdly, conditional on being in poor health with high income, there is a high probability that a person claims. 24.5% of those in the top 20% of the household income distribution who are in poor health claim PIP or DLA. This fact gives some credence to the argument that a lack of means (or asset) testing in the program may cause outlays which are not well-targeted at those who need them most.

Panel B of Table 1.4 shows the share of program spending by real equivalized household income and self-reported health. The concern about large outlays going to high-income individuals appears less important: less than 1% of total program spending goes to people in the top 20% of household earnings with poor health; over half of program spending goes to people in the bottom 40% of the income distribution who are in poor or fair health. These patterns are explained by the health-income gradient. Although there is a high probability of claiming conditional on being in poor health and having high income, only a small percentage of the population is in poor health with high income.

To delve further into understanding the economic lives of DLA and PIP claimants, I break down the sources of real equivalized household income for households with DLA claimants, households with PIP claimants, and households with no claimants of either program. Results are shown in Figure 1.2. As

we would expect, overall household income is lower for benefit claimants, and the pattern of household income sources is similar across DLA claimants and PIP claimants. This fact underscores the similarity of the households who claim these two benefits. For households without a claimant, by far the largest source of income is labor income, followed by pension income.[19] Claimant households on average receive very little labor income from the claimant's work, but somewhat more income from other household members. Overall household labor income is much smaller than in non-claimant households, though. Claimant households also receive on average more income from social benefits other than PIP or DLA than non-claimant households.

A common finding in the literature on disability benefit programs is that program recipients tend to have low income (e.g., Deshpande and Li, 2019). Effective targeting of low-income people is a significant contributing factor to the insurance value of disability benefit programs (Deshpande and Lockwood, 2022). In addition to the health-income gradient, there are several features of disability benefit programs which may facilitate targeting low-income individuals: means-testing may ensure that high-income people are ineligible; and work-restrictions may make it incentive-incompatible for people who are able to earn high incomes to claim. Even though PIP and DLA do not have these features they are effective in targeting low-income households for two reasons: firstly, low-income households are disproportionately likely to have a member in poor health; and secondly, conditional on health status, low-income people are more likely to apply for benefits. The relative importance of these two effects is important for assessing the likely effects of policy changes: if self-selection on income is a significant driver of the redistributive nature of the programs, then making benefits more generous is more likely to attract higher-income applicants, thereby generating wasteful public spending.

1.4.2 Associations between working and claiming benefits

The evidence from the previous section suggests relatively low labor supply amongst claimants. This could be a result of their health condition or a result of claiming benefits. Although benefits are not tied to work or earnings, DLA and PIP may reduce labor supply due to the income effect of receiving benefits. If this effect exists, then there may be an additional fiscal cost of benefits due to lower tax receipts. To investigate associations between claiming behavior and labor supply, I estimate a series of linear probability models using data from the BHPS/US sample. Letting $work_{it}$ be a binary variable equal to 1 if a person works and zero otherwise, I estimate regressions of the following form:

$$work_{it} = \beta_0 + \beta_1 PIP_{it} + \beta_2 DLA_{it} + \gamma' X_{it} + \tau_t + \epsilon_{it} \tag{1.1}$$

PIP_{it} is a binary variable which is equal to 1 if a person claims Personal Independence Payments, and 0 otherwise. DLA_{it} is a binary variable which is equal to 1 if a person claims Disability Living Allowance, and 0 otherwise. X_{it} is a vector of control variables which vary across specifications, and τ_t is a wave fixed effect. In some specifications I also include person fixed effects.

[19] Although one cannot apply for benefits over the age of 65, over 65s who have an existing entitlement to benefits from an earlier application can continue to claim.

Table 1.5 reports estimates of β_1 and β_2 across a series of specifications. In specification (1), I control for wave fixed effects and age using a full set of dummies. In specification (2), I additionally include dummy variables for each value of self-reported health. Specification (3) additionally includes person fixed effects. Across every specification, there is a statistically significant negative association between claiming benefits and working. Adding controls reduces but does not eliminate the association. In particular, controlling for person fixed effects and health status attenuates the association to -0.047 (for DLA) and -0.058 (for PIP). This a much smaller association than causal effects estimated for US disability insurance. French and Song (2014) estimate that receiving benefits decreases the probability of working by 26 percentage points. Their result is consistent with work restrictions having a large effect on labor supply, which are not present in DLA and PIP.

These associations should not be interpreted as causal, since movement on or off benefits is likely correlated with unobservable factors which also affect the likelihood of working. People who claim disability benefits may have additional health deficits not captured by my self-reported health variable. These unobserved health-related differences also raise the cost of working on average, which will tend to bias my estimates away from zero relative to the underlying causal effect. We might think of these estimates as providing an upper bound of the reduction in labor force participation generated by receipt of DLA and PIP. The results in this section are therefore consistent with DLA and PIP being significantly less distortionary than programs with work restrictions or means-testing.

1.4.3 Marginal value of disability benefits

DLA and PIP effectively target people with low income, which is consistent with these benefits providing positive insurance value: the value to risk-averse agents of reducing the *ex post* dispersion of consumption (Low & Pistaferri, 2020). However we should compare this insurance value with the distortionary cost of the taxes required to pay for disability benefit. Since some spending on PIP and DLA goes towards high earners, the fiscal cost and corresponding distortionary effect of additional taxes may be higher than for a program with work restrictions. To assess whether benefits are overall welfare-enhancing, I consider the welfare effect of a marginal increase in benefit generosity. Following Deshpande and Lockwood (2022), I consider the following policy change: an increase in benefit generosity by 1 GBP and an increase in taxes such that overall, social welfare as defined by the expected utility of a consumer who does not know the realization of their stochastic income and health processes is unchanged. I then ask whether total net government revenue from this change is positive. If it is positive, then one may conclude that increasing the generosity of benefits is welfare-enhancing at the margin (since the government could theoretically rebate this revenue as additional consumption). The sufficient statistics welfare calculation requires setting key parameter values which can be calibrated directly from data or reduced-form estimates .

Define the set of people who claim B (and may or may not work) and those who do not claim and work W. For each individual i:

$$\Pr(i \in B) = \pi_B, \text{ and } \Pr(i \in W) = \pi_W \tag{1.2}$$

Each person has labor income I_i, and consumption c_i from which they derive flow utility $u(c_i)$.[20] I assume that any change in permanent income has the same effect on consumption. Define the change in income induced by an increase in benefit payments of 1 pound (GBP) $\frac{\partial I_i}{\partial DI}$. The elasticity of taxable income, the percentage change in taxable income induced by small changes in the marginal tax rate, is given by $\epsilon_{TI} = \frac{(1-\tau_i)}{I_i}\frac{\partial I_i}{\partial \tau_i}$. I consider the welfare effect of a marginal increase in benefit generosity: what is the effect on net government revenue of an expected utility-neutral increase in benefits and taxes?

There are three key determinants of the welfare calculation. The first is the rate at which agents are willing to transfer consumption from the state of the world where they do not claim to the state of the world where they do claim:

$$\frac{\pi_B}{\pi_W}\frac{\mathbb{E}[\lambda_i|i \in B]}{\mathbb{E}[\lambda_i|i \in W]} \tag{1.3}$$

where $\lambda_i = u'(c_i)$. This is equivalent to the slope of the indifference curve in consumption space in different states of the world. Note that the higher the probability of claiming benefits, and the higher the marginal utility of consumption when claiming benefits, the more the agent is willing to lose in the state of the world when they do not claim for an additional pound in the state of the world when they do claim.

However, non-claimant workers' income is reduced through distortionary taxation. The reduction in post-tax income is therefore the sum of two effects: firstly, income is reduced through the worker keeping a smaller share of their pre-tax earnings; and secondly, pre-tax earnings are reduced. Only a share of the reduction in income is therefore collected by the government as additional revenue, with the remainder being reduced labor income. The share going to the government is:

$$\frac{1}{1+\epsilon_{TI}} \tag{1.4}$$

where ϵ_{TI} is the elasticity of taxable income (ETI) with respect to the marginal tax rise.[21] The share of income reduction for the consumer which is rebated to the government is higher, the less elastic income is.

Finally, the cost to the government from increasing disability benefits is composed of the cost of paying the extra benefit plus the cost of reduced government revenue from the induced change to the labor income of benefit claimants:

[20] Since this analysis considers only marginal changes, the effect of income-induced changes in labor supply and future saving on utility is zero by the envelope theorem.

[21] To see why, note that where $I_i(1-\tau_i)$ is post-tax income, the differential of post tax income with respect to the marginal tax rate is given by:

$$\frac{d(I_i(1-\tau_i))}{d\tau_i} = \frac{dI_i}{d\tau_i}(1-\tau_i) - I_i = \frac{dI_i}{d\tau_i}\frac{1-\tau_i}{I_i}\frac{I_i}{1-\tau_i}(1-\tau_i) - I_i = -|\epsilon_{TI}I_i| - I_i$$

$-I_i$ is the change in post-tax income resulting from additional tax payments holding pre-tax income constant, and $-|\epsilon_{TI}|I_i$ is the change in post-tax income resulting from a reduction in pre-tax income. The share of the reduction in income which is additional government revenue is therefore $\frac{1}{1+\epsilon_{TI}}$.

$$1+\left|\frac{\partial I_i}{\partial DI}\right|\tau_i \tag{1.5}$$

τ_i is the marginal tax rate, and $\partial I_i/\partial DI$ is the marginal effect of additional benefit income on labor income. The cost to the government is greater the more claimants react to additional benefits by reducing their income.

The change in net revenue is the additional revenue from working non-claimants less the cost of increasing benefits to claimants, weighted by the respective shares of those groups within the population:

$$\Delta R=\pi_W\frac{1}{1+\epsilon_{TI}}\frac{\pi_B}{\pi_W}\frac{\mathbb{E}[\lambda_i|i\in B]}{\mathbb{E}[\lambda_i|i\in W]}-\pi_B\mathbb{E}\left[1+\left|\frac{\partial I_i}{\partial DI}\right|\tau_i|i\in B\right] \tag{1.6}$$

The first term is the welfare-neutral reduction in income for non-claimant workers, scaled by the share which goes to the government, weighted by the share of non-claimants workers within the population. This is the additional tax revenue for the government. The second term is the cost of increasing benefits to the government, weighted by the share of the population which claims benefits. The welfare gains from expanding generosity (the increase in government revenue per person holding total population welfare constant) are larger: the greater the marginal utility of consumption for claimants relative to non-claimants; the smaller the effect of benefit receipt on labor income; and the smaller the elasticity of taxable income.

Table 1.6 shows the chosen values of important parameters that feed into equation 1.6. Note that for risk aversion (which drives marginal utility), elasticity of taxable income, and the causal effect of benefit claiming on labor income, I use multiple values to reflect uncertainty and examine how results depend on the choice of parameters. The distribution of consumption by claimants and non-claimant workers comes from the Living Costs and Food Survey. There is a large disparity in consumption between claimants and non-claimants, and under the assumption of risk aversion, this will tend to drive the result that it is welfare improving to transfer income to claimants. This fact also strongly suggests that payments to high-income households are not quantitatively important. For the marginal utility of consumption, I assume a CRRA utility function: $u(c)=\frac{c^{1-\sigma}}{1-\sigma}$ and let the coefficient of risk aversion be equal to either 1 or 2. I calculate the share of claimants and non-claimant workers directly from the Understanding Society sample. I use a range of ETI estimates from the literature.[22] To my knowledge, no previous estimates of the effect of claiming PIP or DLA on total labor earnings exist. Again I use a range of assumptions, reflecting different reasonable bounds of the effect. Firstly, the effect is bounded by zero under the assumption that leisure is a normal good. Secondly, under the assumption that any time-varying confounders are positively correlated with claiming disability benefits and negatively correlated with earnings (e.g., unobserved changes in health status), a regression of labor income on total income from DLA and PIP provides a reasonable

[22]The ETI literature mostly uses difference-in-differences design exploiting changes in tax rates (e.g., Brewer et al., 2010, Browne and Phillips, 2017), or a bunching methodology which exploits mass points at "kinks" in the tax schedule (e.g., Britton and Gruber, 2020, Adam et al., 2021). There is a relatively wide range of estimates from close to zero to 0.5.

lower bound of the effect of benefit income on labor income.[23] The lower bound of the effect is given by -0.048, i.e., an extra pound of benefits leads to a 0.048 pound reduction in labor income.

Table 1.7 reports the value of the change in revenue per person estimated using equation 1.6 for different values of parameters of interest. The additional government revenue from an *ex ante* welfare-neutral marginal increase in benefit generosity is positive under all assumptions. Its lowest value is 0.038, i.e., a welfare-neutral increase in benefit generosity of GBP 1 and tax increase would generate additional net revenue of GBP 0.038 per person. At higher levels of assumed risk aversion, low assumed effect of taxation on taxable income, and the upper bound estimate for the effect of benefit generosity on the labor income of claimants, the additional net revenue can be as large as GBP 0.42 per person. The largest determinant of the benefit of making disability benefits more generous is the coefficient of relative risk aversion. If we assume a coefficient of relative risk aversion of 2, the effects of making benefits more generous are strongly positive under any reasonable assumptions about the degree of distortion.

These results strongly suggest that even in the absence of work restrictions, a disability benefit program can be welfare-improving. This finding is in line with the literature which shows the beneficial welfare effect of disability benefit outweighs the distortionary cost in programs with work restrictions (e.g., Low and Pistaferri, 2015, Hosseini et al., 2021, Deshpande and Lockwood, 2022). The key determinant of this result is the stark difference in consumption between claimants and non-claimants, which indicates that the lack of work restrictions does not lead to people in high-income households claiming disability benefit.

However, my results should not be interpreted as a definitive claim about the value of increasing benefits, as there are likely effects of changes in disability benefits not captured by this exercise. Firstly, I assume that the share of the population claiming benefits π_B and working π_W are unchanged by the reform. In practice, more generous benefits could induce an increase in applications, potentially diluting the targeted nature of benefits. Secondly, this calculation does not incorporate changes in self-insurance (i.e., asset accumulation) which could dampen the beneficial effects of expanding benefits. Such effects are likely larger for large increases in benefit generosity. This consideration is important in my context because one may suspect that the effective targeting of the disability benefit program results from the low monetary value of benefits; only those with very low consumption choose to apply. Finally, this approach only allows me to assess the effects of marginal changes in benefit generosity. A sufficient statistics approach does not permit me to assess the overall welfare effects of large changes in benefit generosity or work restrictions. These limitations support the structural approach taken in the remainder of the paper.

1.5 Modeling reform of UK disability benefits

The stylized facts established in the previous section inform the construction of a structural model to assess non-marginal changes in the program, including the introduction of work restrictions. Disability benefits are broadly redistributive, going disproportionately to those who are in poor health and low income. In particular, the sufficient statistics welfare analysis shows that the disparity in consumption between claimants and non-claimants is a key driver of claims about the welfare effects of disability benefits. These

[23] I also include person fixed effects, age and self-reported health as controls.

findings underscore the importance of capturing the degree of asset and income inequality in the economy. Accordingly, my model includes several sources of income and asset inequality, including heterogeneity in wages, health risks, and discount rates by educational attainment, and persistent wage risks.

Additionally, although there is little evidence that claiming disability benefits creates large distortions in the labor market, claimants mostly do not work. This result is consistent with health-related work costs being the main driver of low labor supply amongst claimants. I therefore construct a model in which wages and preferences over labor supply differ by health status.

In my model, agents are born aged 26, retire at 65, and may die at any age from 65 to 85. If they are still alive at the end of their 85th year, they die. Agents, indexed by i, differ by education e_i which can take three values:

$$e_i = \begin{cases} 3 & \text{if a person's highest qualification is a degree or equivalent} \\ 2 & \text{if their highest qualification is a GCSE, A-level or equivalent} \\ 1 & \text{if their highest qualification is another (mostly non-academic) qualification} \\ & \text{or they have no qualifications} \end{cases}$$

Education is exogenous for each person and does not change across periods. It is the key source of predetermined heterogeneity in my model, and I allow health and wage offer processes to differ by education. This modeling decision captures the empirical fact that people with different levels of education have markedly different health and labor outcomes, but I do not take a stand on what portion of this association, if any, is causal. This is a common feature of similar models (e.g., Keane et al., 2020, Hosseini et al., 2021). It reflects the fact that the association between education and health is well-documented, but the causal link is unclear (Galama et al., 2018).

In the years up to 65, agents choose whether to work and whether to apply for benefits. Conditional on these choices, they choose which portion of their total resources to consume and save the remainder. They receive a return on assets of r. After the age of 65, agents may not work or apply for benefits, which reflects the fact that over-65s are not eligible to begin claiming PIP or DLA. Over-65s make decisions about what portion of their resources to consume each period.

Agents face risks over health, which affects their non-pecuniary cost of working and wage offers, and direct wage shocks, which affects the wages they earn conditional on working. They manage these risks through precautionary saving and government transfers including disability benefits.

1.5.1 Health process

Health h_{it} for agent i aged t takes one of three values:

$$h_{it} = \begin{cases} 1 & \text{if a person's self-reported health is "excellent", "very good" or "good"} \\ 2 & \text{if their health is "fair"} \\ 3 & \text{if their health is "poor"} \end{cases}$$

Agents' health follows a stochastic process over the life cycle. Following De Nardi, Pashchenko and Porapakkarm (2022), I allow the distribution over health states for individual i aged t to depend on more than just i's health state at age $t-1$. I define person i's health as being "persistent" if their current health state is the same as their health state last period i.e., $p_{it} = 1$ iff $h_{it} = h_{it-1}$. The distribution over health states in the next period depends upon the persistence of an agent's current health state. Given an agent's health this period h_{it}, their education e_i, health persistence p_{it}, and age, t, the probability of their health taking the value g' next period is a multinomial logit function. Specifically:

$$\Pr(h_{it+1} = g' | h_{it}, p_{it}, t, e_i) = \frac{\exp[(\gamma_{0eh}^{g'} + \gamma_{1eh}^{g'} t + \gamma_{2eh}^{g'} \mathbb{1}\{p_{it} = 1\})]}{\Sigma_{g=1}^{3} \exp[(\gamma_{0eh}^{g} + \gamma_{1eh}^{g} t + \gamma_{2eh}^{g} \mathbb{1}\{p_{it} = 1\})]} \tag{1.7}$$

The transitions are a function of age and health persistence, and vary by education and current health status. γ_{0eh}^g, γ_{1eh}^g, and γ_{2eh}^g are coefficients which vary by the health status a person may transition to, current health status and health persistence. g indexes health status in the next period, e indexes education and h indexes current health status. The process therefore involves the following 81 parameters: $\{\gamma_{0eh}^g, \gamma_{1eh}^g, \gamma_{2eh}^g\}_{g=1,2,3,h=1,2,3,e=1,2,3}$.

At the end of each period, over 65s die with probability:

$$1 - \text{p_surv}(h_{it}, t) = \frac{\exp(s[t, h_{it}])}{1 + \exp(s[t, h_{it}])} \tag{1.8}$$

where:

$$s[t, h_{it}] = \kappa_0 + \kappa_1 t + \kappa_2 t^2 + \mathbb{1}(h_{it} = 2)[\kappa_3 + \kappa_4 t + \kappa_5 t^2] + \mathbb{1}(h_{it} = 3)[\kappa_6 + \kappa_7 t + \kappa_8 t^2] \tag{1.9}$$

Note that the probability of transitioning between health states does not depend on work choices, income, or healthcare utilization. In this respect my model is in line with similar papers which study disability benefit (e.g., Low and Pistaferri, 2015, Hosseini et al., 2021, Kellogg, 2022). There are two potential concerns with this approach. One is that since I do not account for the income effect on health, I may understate the value of disability benefit. Some studies have estimated a beneficial effect of disability benefit on health outcomes (e.g., Silver and Zhang, 2022, Gelber et al., 2022), but these effects are typically relatively small. Additionally, in the UK context where healthcare is mostly provided by the National Health Service to all citizens at nearly zero out-of-pocket cost, the effect of income on health is likely smaller than in the US. Secondly, I may potentially overstate the beneficial effects of disability benefit by failing to account for *ex ante* moral hazard. If disability benefit blunts the incentive to invest in health, it has an additional distorting effect. My model does not account for this. However, Cole, Kim, and Krueger (2019) investigate optimal insurance against health risks in a model with endogenous health and *ex ante* moral hazard and find that it is still optimal to insure 80% of health-related income shocks. In an economy with stark health-related income risks, therefore, *ex ante* moral hazard is likely a secondary concern.

1.5.2 Wage processes and other income

The wage process is exogenous from the perspective of agents. Each period, agents receive an annual wage offer, w_{it}, defined by:

$$\log(w_{it}) = \delta_0 + \delta_1 t + \delta_2 t^2 + \delta_3 \mathbb{1}\{h_{it} = 2\} + \delta_4 \mathbb{1}\{h_{it} = 3\} + \delta_5 \mathbb{1}\{e_i = 2\} + \delta_6 \mathbb{1}\{e_i = 3\} + \epsilon_{it} \quad (1.10)$$

Log wage offers include a polynomial in age, with shifters reflecting different productivity by health and education.[24] Log wage shocks, ϵ_{it}, are persistent, and follow an AR(1) process:

$$\epsilon_{it} = \rho \epsilon_{it-1} + \zeta_{it} \quad (1.11)$$

where ρ is the persistence of wage shocks and $\zeta_{it} \sim N(0, V_\zeta)$ is a normally distributed white noise term. Given this wage offer, an agent may choose to work. Their labor supply is denoted $\ell_{it} \in \{0, 1\}$, which is equal to 1 if the agent works.

Other sources of income for agents are spousal income and other state benefits. Since I do not explicitly track marital status as a state variable, spousal income is a deterministic function of own education, age, and age-squared:

$$\text{SPOU}_{it} = \phi_0 + \phi_1 t + \phi_2 t^2 + \phi_3 \mathbb{1}\{e_i = 2\} + \phi_4 \mathbb{1}\{e_i = 3\}$$

Spousal income reflects the fact that an important source of income for both claimants and non-claimants is other labor income in the household. Note that spousal income is not a function of current health status, consistent with empirical findings that spousal income does not react significantly to non-fatal health shocks (Fadlon & Nielsen, 2021).

Other state benefits differ by labor income, health state, and age:

$$\text{BEN}(t, h_{it}, \ell_{it} w_{it}) = \max(0, Y^B(t, h_{it}, \ell_{it} w_{it}))$$

$$Y^B(t, h_{it}, \ell_{it} w_{it}) = \begin{cases} \mu_0 + \mu_1 \mathbb{1}\{h_{it} = 2\} + \mu_2 \mathbb{1}\{h_{it} = 3\} & \text{if } t \geq 65 \\ \mu_3 + \mu_4 \ell_{it} w_{it} + \mathbb{1}\{h_{it} = 2\}(\mu_5 + \mu_6 \ell_{it} w_{it}) & \text{otherwise} \\ + \mathbb{1}\{h_{it} = 3\}(\mu_7 + \mu_8 \ell_{it} w_{it}) & \end{cases}$$

Other benefits are intended to reflect means-tested benefits, such as UC, and include a health-related component to reflect the existence of the health-related component of UC and ESA. They therefore build in the average disincentive effect to working from the wider UK benefits system. Since I do not allow labor supply to vary on the intensive margin, I match average benefit payments by income and health, rather than focusing on precisely matching marginal effective taxes at each level of income.

[24] Note that I do not allow wages to differ between agents' health persistence. A regression of log wages on health status allowing for a difference between those in bad health and persistent bad health produces a statistically insignificant difference of 0.04 between those in bad health and persistent bad health.

Agents pay progressive taxes $T(\ell_{it}w_{it})$. $T(\ell_{it}w_{it})$ is a convex function reflecting the UK tax system which includes progressive income taxes and National Insurance contributions, which is similarly progressive.

1.5.3 PIP/DLA Benefits

Each period, an agent under 65 who does not currently claim PIP/DLA may apply for PIP/DLA benefits ($\text{app}_{it} = 1$). An applicant has disability benefits next period $b_{it+1} = 1$ with probability:

$$\Pr(b_{it+1} = 1|\text{app}_{it} = 1, h_{it}) = \frac{\exp[\alpha_0 + \alpha_1 \mathbb{1}(h_{it} = 2) + \alpha_2 \mathbb{1}(h_{it} = 3) + \alpha_3 \mathbb{1}(h_{it} = 3)\mathbb{1}(p_{it} = 1)]}{1 + \exp[\alpha_0 + \alpha_1 \mathbb{1}(h_{it} = 2) + \alpha_2 \mathbb{1}(h_{it} = 3) + \alpha_3 \mathbb{1}(h_{it} = 3)\mathbb{1}(p_{it} = 1)]} \tag{1.12}$$

The probability of receiving benefits next period conditional on applying depends on one's current health state. This feature reflects the fact that applicants are required to undergo a health assessment, which is less likely to result in an award for those in better health. I allow the probability of successful application to depend on whether a person's bad health is "persistent", that is, an agent who has been in bad health for two or more periods has a higher chance of receiving benefits. This feature reflects the fact that chronic conditions are more likely to require the type of care which qualifies an applicant for benefits and the empirical fact that those who have been in bad health for multiple periods are more likely to claim. This specification allows for type 1 and type 2 errors— that is, people in bad health being denied benefits and people in good health being awarded benefits.

An agent claiming benefits b_{it} receives income y^{PIP} each period, which is independent of health.[25] An agent currently claiming benefits loses benefits going into the next period with probability:[26]

$$\Pr(b_{it+1} = 0|b_{it} = 1, h_{it}, p_{it}) = \frac{\exp(\lambda(h_{it}, p_{it}))}{1 + \exp(\lambda(h_{it}, p_{it}))} \tag{1.13}$$

where

$$\lambda(h_{it}, p_{it}) = \lambda_0 + \lambda_1 \mathbb{1}(h_{it} = 2) + \lambda_2 \mathbb{1}(h_{it} = 3) + \mathbb{1}(p_{it} = 1)[\lambda_3 + \lambda_4 \mathbb{1}(h_{it} = 2) + \lambda_5 \mathbb{1}(h_{it} = 3)] \tag{1.14}$$

This function reflects the risk of reassessment, which is more likely to lead to loss of benefits if an agent is in good health and less likely if the agent is in poor health. I let health persistence affect the probability of benefit removal.

[25] In reality, DLA and PIP benefit payments vary based on the severity of a claimant's condition. I do not incorporate this feature, firstly, because my health categories are insufficiently fine-grained to reflect differences between different chronic conditions, and secondly because this would necessitate including an additional state variable: health state when benefits were initially awarded.

[26] Note that in the model agents cannot choose to leave the program.

1.5.4 Preferences

Agents have preferences over consumption, and derive dis-utility from working, $\ell_{it} = 1$, and applying for benefits, $\text{app}_{it} = 1$. The cost of applying for benefits reflects a combination of nuisance cost, stigma, and aversion to the health assessment. Agents' discrete choices can be expressed as the ordered pair $d_{it} = [\ell_{it}, \text{app}_{it}]$. Conditional on their discrete choice, flow utility is given by:

$$u(c_{it}, d_{it}, h_{it}, p_{it}) = \omega_c \left[\frac{c_{it}^{1-\sigma}}{1-\sigma}\right] + f(h_{it}, p_{it}, t)\mathbb{1}(\ell_{it} = 1) + \omega_{\text{app}}\mathbb{1}(\text{app}_{it} = 1) + \nu_{itd} \tag{1.15}$$

where c_{it} is consumption and ν_{itd} is a type 1 extreme value preference shock specific to the agent i, their age t, and each discrete choice d.[27] The marginal utility of consumption is declining in consumption for $\sigma \geq 1$. The non-pecuniary cost of working $f(h_{it}, p_{it}, t)$ depends on an agent's health state and age. This feature reflects the empirical fact that labor force participation declines with poor health and age, as a result of physical and cognitive deterioration, social barriers to working, and changing preferences. I allow the cost to differ by the persistence of bad health, again reflecting the difference between chronic and acute conditions:

$$\begin{aligned} f(h_{it}, p_{it}, t) = \omega_{h0} + \omega_{h=2}\mathbb{1}(h_{it} = 2) + \omega_{h=3}\mathbb{1}(h_{it} = 3) + \omega_{h=3,p=1}\mathbb{1}(h_{it} = 3)\mathbb{1}(p_{it} = 1) + \\ t[\omega_t + \omega_{t,h=2}\mathbb{1}(h_{it} = 2) + \omega_{t,h=3}\mathbb{1}(h_{it} = 3) + \omega_{t,h=3,p=1}\mathbb{1}(h_{it} = 3)\mathbb{1}(p_{it} = 1)] \end{aligned} \tag{1.16}$$

In order to better match the dispersion of wealth and wealth profile in retirement, I include a bequest motive as in De Nardi (2004), and De Nardi, French, and Jones (2016). Agents' utility from dying with wealth k is given by:

$$u_B(k) = \psi_1(\psi_2 + k)^{1-\sigma} \tag{1.17}$$

The parameter ψ_1 determines the strength of the bequest motive, and ψ_2 controls the extent to which a bequest motive is a luxury.

Agents discount the future at the rate $\beta_i \in \{\beta_{e=1}, \beta_{e=2}, \beta_{e=3}\}$. Following Capatina, Keane, and Maruyama (2020), I allow the discount rate to differ by education. I do this to match empirical differences in asset accumulation by education group, and to reflect the empirical correlation between education

[27] The variance of the preference shock is fixed at the variance of the type 1 extreme value distribution. However, this is without loss of generality because the variance of the preference shock relative to the other terms in the equation is therefore controlled by the scale of the other parameters. Multiplying parameters ω_c, ω_{h0}, $\omega_{h=2}$, $\omega_{h=3}$, $\omega_{h=3,p=1}$, ω_t, $\omega_{t,h=2}$, $\omega_{t,h=3}$, $\omega_{t,h=3,p=1}$, and ω_{app} by the same multiple greater than one decreases the relative weight of preference shocks in the agent's decision-making i.e., to the extent that observably similar agents make the same discrete choices, the weight on these terms increases. Multiplying these same parameters by a number between 0 and 1 increases the relative weight of preference shocks in the agent's decision-making.

and patience (Castillo et al., 2019). Appendix A.1 reports results when I fix the the discount rate across education levels.

1.5.5 State variables and model timing

For people above the age of 65, state variables are given by:

$$\Omega_{it}^{R} = \{t, k_{it}, h_{it}, p_{it}, b_{it}, e_i\} \tag{1.18}$$

Working-age people also have their current wage shock and preference shocks as state variables:

$$\Omega_{it}^{W} = \{t, k_{it}, h_{it}, p_{it}, \epsilon_{it}, b_{it}, e_i, \nu_{itd}\} \tag{1.19}$$

In each period, the timing is:

1. The agent learns the realization of state variables, Ω_{it}^{R} or Ω_{it}^{W}.
2. Agents receive returns on their assets from the last period.
3. For agents who are working age, they make their discrete choice: $d_{it} = [\ell_{it}, \text{app}_{it}]$.
4. Given their discrete choice, agents decide how much of their resources to consume.
5. At the end of the period, agents who currently claim benefits discover the realization of the shock to whether they will have benefits next period; agents who applied for benefits discover the realization of the shock to whether they have benefits next period; and retired people discover whether they will survive to the next period.

I solve the model using backward recursion under the assumption of rational expectations. Details are provided in Appendix A.2, including the optimization problem and how I discretize continuous variables.

1.5.6 Estimation

My model as specified is a better fit for the behavior of men than women, since I do not allow for part-time work or for fertility choices.[28] Therefore, I estimate the model on English men.[29] All monetary variables are expressed in thousands of 2015 pounds. I estimate the model in two stages.

In the first stage, parameters which can be directly estimated from data are estimated outside of the model. Additionally, some preference parameters have been widely estimated and have well-established values. I set a subset of parameters using estimates from the literature.

I estimate the remaining parameters using Simulated Method of Moments: I solve for values of parameters which minimize the distance between selected model-generated and observed moments.

[28] A large majority of UK men who work do so full time (Meghir & Phillips, 2008).

[29] I am restricted to not considering men in Scotland, Wales or Northern Ireland by the ELSA data set.

Calibrated Parameters

Some parameters can be directly estimated from data. The multinomial logit health process is estimated from the Understanding Society data using maximum likelihood. As with any multinomial logit, there must be an "omitted" outcome. Therefore, without loss of generality, I set $\gamma^1_{0eh} = \gamma^1_{1eh} = \gamma^1_{2eh} = 0$ and solve:

$$\max_{\{\gamma^g_{0,e,h},\gamma^g_{1,e,h},\gamma^g_{2,e,h}\}_{g=2,3,h=1,2,3,e=1,2,3}} \Sigma_i \Sigma_t \Sigma_{g'} \mathbb{1}(h_{it+1} = g') \log \left(\frac{\exp[\gamma^{g'}_{0eh} + \gamma^{g'}_{1eh} t + \gamma^{g'}_{2eh} \mathbb{1}\{p_{it} = 1\}]}{\Sigma^3_{g=1} \exp[\gamma^{g'}_{0eh} + \gamma^{g'}_{1eh} t + \gamma^{g'}_{2eh} \mathbb{1}\{p_{it} = 1\}]} \right) \tag{1.20}$$

Analogously, I estimate the logit survival probability parameters $\{\kappa_0, \kappa_1, \kappa_2, \kappa_3, \kappa_4, \kappa_5, \kappa_6, \kappa_7, \kappa_8\}$ via maximum likelihood, using information on whether a person is reported as deceased in the Understanding Society data.

Spousal income is estimated directly from the Understanding Society data. I run a regression of real spousal labor income on an age polynomial and education dummies. Because the growth in spousal income over the life cycle is correlated with time, I control for wave dummies when estimating the coefficients on age and education.[30]

Social benefit income is likewise estimated as a regression of social benefit income on health interacted with labor income, separately for those younger and older than 65, from the Understanding Society data.[31] In the model, I impose that values of social benefit income predicted to be negative based on estimates from the linear model are set to zero.

Progressive income taxes are income tax and national insurance. Income tax is paid at a rate of 20% on income above GBP 12,570, plus an additional 20% above GBP 50,270. National insurance is paid at a rate of 12% on income above GBP 12,570, and 2% above GBP 50,270. These numbers correspond to the UK tax code.

The probability of being removed from benefits at the end of each period is estimated from the Understanding Society data from the logit transition probabilities for those who currently claim benefits.

There are some parameters I calibrate using estimates provided by other researchers. I set the interest rate to r =1.55%, the Holston-Laubach-Williams (Holston et al., 2017) estimate for the UK. σ is the coefficient of relative risk aversion. Following Keane, Capatina, and Maruyama (2020), amongst others, I set σ=2.

Initial conditions

Agents enter the model aged 26. Education is distributed according to the unconditional distribution in the Understanding Society data. I let the distribution of health and persistence at the age of 26 differ by education, estimated from the Understanding Society data amongst those aged 26. I let the distribution

[30] I do not include the estimated wave coefficients or track calendar time in the model. Spousal income therefore corresponds to the fitted values of spousal income conditional on age and education in the last wave of Understanding Society. I do not exclude zero values from the data when fitting this model.

[31] Again, I maintain zero values to avoid overestimating overall average social benefit income.

of initial claiming depend on initial health and persistence. Agents draw an initial wage shock from the stationary distribution. I set initial assets to average savings by health status at age 26 in the Family Resources Survey.[32]

Simulated Method of Moments

The remaining structural parameters are estimated using Simulated Method of Moments (Gourinchas & Parker, 2002). Wage offer parameters must be estimated within the model because offers are not observed for non-workers. The average wage received by a worker conditional on age, health and education is different from the average wage that "would be" received by a non-worker. Likewise, wage persistence and conditional variance are only observed conditional on labor supply choice. The model accounts explicitly for this selection.

The probability of successful application conditional on health status is not observed in my data set: administrative data from the DWP give me the share of people of each age and sex who *apply* for benefits but not the probability of success conditional on health status. Further, applicants are a selected group: the share of applicants in good health who are successful is likely not indicative of the probability of a non-applicant in good health being successful if they applied.

I simulate an economy of 10,000 agents and solve for those parameters which minimize the squared distance between key moments simulated from my model and from various data sources. Formally, let the vector of structural parameters to be estimated be:

$$\theta = \{\delta_0, \delta_1, \delta_2, \delta_3, \delta_4, \delta_5, \delta_6, \rho, V_\zeta, \alpha_0, \alpha_1, \alpha_2, \alpha_3, \omega_c, \omega_{\text{app}}, \omega_{h0}, \omega_{h=2},$$
$$\omega_{h=3}, \omega_{h=3,p=1}, \omega_t, \omega_{t,h=2}, \omega_{t,h=3}, \omega_{t,h=3,p=1}, \psi_1, \psi_2\, \beta_{e=1}, \beta_{e=2}, \beta_{e=3}\}$$

.

Define the set of moments estimated from a simulation using my model $\hat{g}(\theta)$, and let the vector of corresponding data moments be M. Then I numerically solve:

$$\min_\theta (M - \hat{g}(\theta))' W (M - \hat{g}(\theta)) \tag{1.21}$$

W is a diagonal weighting matrix with the inverse of the variance of the data moments as its elements. This is equivalent to minimizing the standard deviation-weighted squared distance between simulated and data moments. For the most part, I estimate the variance of data moments using standard errors.

Table 1.8 lists the moments I use to identify each parameter in θ. The utility parameters which pertain to consumption and labor supply are estimated by matching life-cycle work probabilities by age and health status up to the age of 64. Life-cycle probabilities of working are estimated using regressions of a dummy variable for working on a full set of age and health dummies, controlling for wave fixed effects so as not to conflate age trends in labor market participation with broader macroeconomic trends. The life cycle

[32] The measure of assets in the Family Resources Survey does not include housing wealth. Implicitly I assume that agents are not homeowners at the age of 26.

profile used in the data moments M is then the fitted values from this regression for each age and health status in the last wave of Understanding Society.[33] I estimate the disutility of applying for benefits by matching the probability of applying at each age. I calculate these probabilities directly from DWP data.[34] Conditional on my model correctly reproducing the share of applicants, the percentage of applications of each health status which are successful is identified by the percentage of people with each health status who claim PIP/DLA. The profile of benefit claims by health status and age is estimated using a regression on a full set of age and health status dummies, controlling for wave fixed effects. The discount rates are identified from asset accumulation profiles by education. Bequest motive preferences are identified by the profile of asset decumulation in retirement. Asset accumulation profiles are estimated by education level from a regression of asset ownership on age dummies controlling for wave fixed effects.

For the deterministic component of wages, I match average log wages by age, health status and education. I estimate these profiles from the data by regressing log wages on fully-interacted age, health status and education dummies, controlling for wave fixed effects. I then take the residuals from these regressions to identify parameters associated with the stochastic component of the wage process. I match the 1st to 10th autocorrelation of the log wage residuals to the analogous moments calculated from my simulated data (i.e., I run the same regression in my simulated data and estimate the 1st to 10th autocorrelation of residuals). This pins down the persistence of the stochastic process.[35] To identify the variance, I match the variance of the residual in my simulated and observed data at each age.

My model does not include population growth or trends in education decisions over time. As a consequence the distribution over age and education differs between the data and the simulations from my model. I therefore construct model moments by sampling from the simulated data which is weighted to match observed distribution of education and age. For each moment, I draw 1000 observations from my simulated data which are stratified by education and age, if applicable,[36] where the distribution of education and age are computed from the Understanding Society data set. I follow the same procedure for all model moments presented in the remainder of the paper.

1.5.7 Model fit

In this section, I report the estimation results, the extent to which my model matches the targeted moments, and the extent to which it matches non-targeted moments.

Appendix A.3 reports estimated health process parameters. Other estimated and calibrated parameters are reported in Table 1.9. Of note, agents in fair health receive wage offers on average 0.451 log points lower than agents in good health, and agents in bad health receive wage offers 0.896 log points lower than agents

[33] I use this procedure whenever the moments are estimated using panel data.

[34] To calculate the variance of the estimated share of people applying for benefits, I exploit the fact that the number of applicants is a binomial-distributed variable with known variance, i.e., if the percentage of people aged a who apply is p_{app}, and there are N_a people in the UK population aged a, then the variance of the estimator $\hat{p}_{\text{app}}$ is $\frac{\hat{p}_{\text{app}}(1-\hat{p}_{\text{app}})}{N_a-1}$.

[35] I aim to match the full autocorrelation profile of wages over a ten year look-back because differences in autocorrelations over different periods can affect conclusions about optimal insurance (De Nardi et al., 2021).

[36] This is not necessary for moments which are specific to a given age or education, e.g., average wages for 35 year olds with a college degree.

in good health. Shocks to wage offers are highly persistent. In terms of the probability of successful application, agents in good and fair health have around a 1.2% and 1.4% chance of successful application, respectively. Conditional on applying, agents in bad health and persistent bad health are successful with 79% probability. My results are therefore consistent with low levels of type II errors (failing to reject healthy applicants) and moderate levels of type I errors (rejecting unhealthy applicants). Regarding preference parameters, it is notable that there is a larger negative utility cost to applying for benefits than to working in any health state. Since working is likely more costly than applying for benefits in terms of forgone leisure, this likely reflects stigma costs or aversion to the health assessment. Discount factors range from 0.988 to 0.948, within the range of common estimates. Agents with more education are more patient. Finally, the bequest motive is small relative to estimates from US data (e.g., De Nardi, French and Jones 2016), when one accounts for the ratio of the bequest motive to the weight on consumption (the weight on consumption is assumed to be 1 in models without preference shocks).

Figure 1.3 reports the simulated and estimated transitions between health states for those under the age of 51. I separate the sample into under 51s and over 50s in order to check that the average fit of transition probabilities does not obscure disparities at different points of the life cycle. Appendix A.4 reports the transitions for over 50s. Overall, my model effectively matches the transition probabilities. There is a low probability of transition from good health to fair health, and a lower probability of transition to bad health. Persistence of health states plays an important role, particularly in transitioning from a bad health state. Those who have been in bad health for more than one period have over a 60% chance of continuing to be in bad health next period and nearly zero probability of transitioning to good health. On the other hand, those who have been in bad health for only one period have a higher chance of transitioning to fair health than remaining in bad health.

Turning to labor market and financial outcomes, in Figure 1.4, I plot average wages by health and education in my simulated model and the data. Again, I match patterns relatively closely. Agents in bad health have lower wages conditional on working and there is an additional education gradient in wages. Figure 1.5 plots life cycle asset accumulation by education in my model and in the data. In order to smooth out fluctuations due to sampling error when plotting variables over the life cycle, I pool these data into 5-year age bins. Note that I can only observe assets from the age of 50 in ELSA. There is a steep gradient in asset accumulation by education. Agents with a degree or equivalent have around three times the assets of agents with no or other qualifications at the age of 60. Finally, in Figure 1.6, I plot labor force participation by current health status over the life cycle. There is a significant difference in labor force participation by health status. Additionally, those in persistent bad health are much less likely to work than those in non-persistent bad health. This fact underscores my decision to impose different non-pecuniary costs of working on these groups.

Finally, considering the profile of benefit claims and applications, Figure 1.7 shows how I match the share of people applying for benefits each period. I capture the magnitude and trend over the life cycle well. Figure 1.8 shows the life cycle of benefit claims by current health status. I am mostly effective in matching the probability of claiming, except that the model predicts a lower share of people in non-persistent bad health claiming than is observed in the data. The most likely explanation is that while in the model I

impose a period between applying for benefits and starting to claim, in reality, agents may transition to poor health and apply for benefits between survey interviews. However, I am able to match the stylized fact that those in persistent bad health are significantly more likely to claim.

Appendix A.4 reports additional matched moments, including those moments which identify the persistence and variance of wage shocks.

Turning to unmatched moments, there are a number of key moments which affect the credibility of my model's conclusions. Firstly, Figure 1.9 plots the share of people at each age in each health status. Although the health process matches period-to-period transitions rather than health status at each age, I am successful in capturing the trend in health status within the population and its trend over the life cycle. In general, I capture aggregate health risk well.

I am able to match the stylized fact, though not the exact magnitude, that claimants tend to have lower levels of assets than non-claimants (see Figure 1.10). As I show in Section 1.4.3, the lower consumption of claimants is a key driver of welfare effects. To the extent that my model overestimates the asset accumulation of claimants, I will tend to underestimate the benefits of more generous disability benefit and underestimate the cost of work restrictions for claimants.

Regarding behavioral responses, my model endogenously returns the effect of taxation on male labor supply because men make decisions about whether to work taking into account the monetary value of wages compared to the disutility of work. Agents can adjust labor supply on the extensive margin only, which corresponds to the empirical evidence that intensive margin adjustments to wage rates are very small for UK men (see e.g., Meghir and Phillips, 2008). In order to test whether the distorting effect of taxation is reasonable in this model, I simulate a 10% increase in after-tax wages (i.e., not spousal income), and derive a participation elasticity of 0.16, around the mid-point of estimates for UK men (Meghir & Phillips, 2008). My model therefore produces behavioral responses to taxation which are in line with observed behavior.

Finally, I consider the behavior of claimants and applicants. While in the DWP data, roughly 50% of applications are successful, in my model, 40% of applications are successful. Table 1.10 reports simulated and actual data moments around the transition onto benefits and the probability of working conditional on claiming. My model overestimates the probability of agents in bad health transitioning onto benefits and underestimates the probability of agents in good and fair health transitioning onto benefits. Most agents in good or fair health who claim benefits in my model transitioned to benefits during a period when they were in bad health. I also overestimate the probability that a claimant works, most significantly for claimants in good health. The most likely explanation is that within each health category, agents who claim benefits have some health condition which may make it harder for them to work. This effect is especially pronounced for those in good health who nevertheless claim benefits. Some people who report being in good health are in fact disabled in a meaningful way. This fact implies that my model suggests more moral hazard than is observed in the data. Therefore, my model will likely overstate the fiscal benefits of work restrictions.

1.6 Counterfactual simulations

In this section, I report the simulated effects of counterfactual policy reforms using the estimated model. All reforms accompany tax changes which make them revenue-neutral. Agents' labor supply reacts endogenously to changes in taxation. For spouses, I impose a change in spousal labor supply to taxation using an elasticity of taxable income of 0.3.[37] In each policy reform, I separately balance the budget for each education category. I introduce flat taxes or subsidies on each education type until the "deficit", that is, the difference between spending on benefits and tax revenue, is the same as before the increase in benefits. This is a similar procedure to the one used by Hosseini, Kopecky and Zhao (2021), and it has the advantage of separating the efficiency gains due to the policy change from the effect of redistributing between education types.

The first reform I simulate is a doubling of the monetary value of disability benefits, with a corresponding increase in taxation. The purpose of this exercise is to investigate whether a program without work restrictions can effectively redistribute to disabled people without large distortions of labor supply when the monetary value of benefits increases. It is possible that a program without work restrictions only remains well-targeted when monetary benefits are small. Table 1.11 reports how key moments change in response to policy reforms. In the column labeled "Doubling benefits", I report the effect of a revenue-neutral doubling of benefit generosity. The overall spending on benefits per person increases by more than 100% because in addition to greater payments conditional on receipt, more agents apply for benefits. There is a 24% increase in the number of claimants, implying an elasticity of claiming with respect to the monetary generosity of benefits of 0.24.[38] This muted effect is likely the result of the large non-pecuniary cost of applying for benefits. One key insight therefore is that the disutility of applying for benefits is an important factor keeping benefits well-targeted. The overall effect of the reform on labor earnings is small even with the tax distortion. This is because there are two offsetting effects. Firstly, spousal income unambiguously declines with the additional taxes, and male labor supply is additionally distorted. Secondly, there is an offsetting effect on male labor supply which is the income effect of men having lower spousal income. There is a larger effect on asset accumulation. Even though the reform redistributes income to people in poor health, their asset accumulation at the age of 65 increases by less than GBP 4000. The asset accumulation of those in good health and fair health declines. One implication is that state insurance may crowd out self-insurance through asset markets.

I measure welfare as the percentage increase in consumption in each period of life in my baseline model, holding fixed all other decision-making, such that overall expected utility at age 26 is increased by the same amount as the policy reform. I do this separately for those with different levels of education. This approach can be understood as a "veil of ignorance" approach (Rawls, 2020, Harsanyi, 1955), that is, I ask how expected welfare changes when summing over the realization of all possible states of the world. I distinguish between different education categories for two reasons: firstly, assessing policy reforms by education groups allows me to account for heterogeneous welfare effects and to partially distinguish be-

[37] See Section 1.4.3 for the range of ETI estimates in the reduced-form literature.

[38] This is at the lower end of the range of estimates in the literature on the US (see Low and Pistaferri, 2015)

tween redistributive effects of policy and efficiency gains (e.g., Hosseini et al., 2021). Secondly, since I allow the discount rate to differ by education, it is not meaningful to compare changes in lifetime utility across education groups. For example, in the population as a whole, any policy which redistributes income from agents with less education to agents with more education will appear welfare-improving. In Appendix A.1, I repeat this analysis for a model where the patience parameter is fixed across education levels. I find qualitatively consistent results.

I find that all education categories see welfare gains from this policy change. Those with a degree or equivalent are indifferent between a revenue-neutral doubling of benefits and a 2.14% increase in lifetime consumption. Those with A-levels or GCSEs are indifferent between the policy change and a 2.2% increase in lifetime consumption. Those with no qualifications or other qualifications are indifferent between the policy change and a 1.87% increase in lifetime consumption. The welfare benefits are surprisingly smallest for those with the least education, despite this group facing the highest health risks. One explanation is that the tax rates required to balance the budget for those with no or other qualifications are around four times that of those with a degree or equivalent. For those with a degree or equivalent, an additional tax rate of 0.5% is required for the policy to be revenue-neutral. For those with A-levels or GCSEs, the tax rate is 1%, and for those with no qualifications or other qualifications, it is 2.2%. Another explanation may be that since those with no qualifications or other qualifications discount the future at a higher rate, they place less value on increases in benefits which are more likely to be paid later in life.

Secondly, I simulate the effect of adding work restrictions to the benefits program. From a theoretical perspective, work restrictions have ambiguous effects on welfare. On the one hand, work restrictions may reduce the moral hazard associated with disability benefit programs by making it incentive-incompatible for those with high wage offers to claim disability benefit. The fiscal cost of the program and the taxation needed to fund it is therefore reduced. On the other hand, the insurance value of benefits is reduced when one cannot be insured against poor health if one can work (or expects to in the near future). There may also be a greater distortion of work decisions amongst those who claim.

In this policy simulation, a person claiming benefits who chooses to work will experience two differences with respect to the baseline model. Firstly, he receives no benefits payments this period. Secondly, he loses benefits next period with probability 1.

Again, the responses of key variables are reported in Table 1.11. I implement tax changes to ensure that the change is revenue neutral for the government in each education category. There is a decline of around 50% in benefit receipt as a result of the policy change. This occurs both because benefit claimants who receive a high wage offer lose their benefits as a result of taking the offer, and because receiving benefits become less attractive to applicants who may expect to work and be required to stop claiming benefits in the future. Applications for benefits fall by 28%. Falling benefit claims are consistent with a reduction in moral hazard caused mostly by claimants who recover work capacity, either through improved health or favourable wage shocks, choosing to exit the program in order to work. Taxes are reduced, but again, the overall labor supply effect is muted, due to offsetting effects of 1) greater labor supply distortion amongst claimants, 2) smaller labor supply distortions for spouses, 3) the income effect of higher spousal earnings for men, and 4) smaller distortions of male labor supply. Work restrictions overall cause a decline in

welfare for all education categories. Consumption-equivalent welfare falls by 1.3% for those with a degree or equivalent, 1.4% for those with A-levels or GCSEs, and 1.6% for those with no or other qualifications. For work restrictions, the welfare loss is highest for those with less education, consistent with those with less education having higher health risks.

I also simulate combinations of these two policies. One hypotheses may be that the value of work restrictions changes with the generosity of benefits: as benefits become more generous, the risk of moral hazard increases since healthier and wealthier people will have a greater incentive to apply for benefits. Work restrictions may dampen this effect. In order to explore this, I simulate a series of policy experiments combining increased benefits with work restrictions. Figure 1.11 reports the welfare effects. I estimate the consumption-equivalent welfare changes for an increase of 100%, 200%, and 300%, both with and without introducing work restrictions. The circular points report the welfare changes for each level of education when work restrictions are not introduced, and the triangular points report the welfare changes when work restrictions are introduced. For larger increases in benefit generosity, increases in taxation can be large and vary across education (e.g., when I increase benefits by 300% and impose work restrictions, the additional flat tax for those with no qualifications or other qualifications is 6.8%, but is only 1.3% for those with a degree or equivalent). For all levels of education, increasing benefits always generates improvements in welfare. Work restrictions always cause a decline in welfare for agents with the least education, but cause an increase in welfare for those with degrees or A-levels/GCSEs when combined with tripling of benefit payments or greater. However, agents would still prefer a tax-neutral increase in benefits without work restrictions. Overall, my model implies that increases in disability benefit payments are welfare improving across a reasonable range of increases, and work restrictions always decrease welfare relative to having no work restrictions.

1.7 Conclusion

Improving the design of disability benefit programs has potentially large welfare benefits due to their large fiscal cost and the high cost of poor health and disability. In this paper, I contribute to the literature on the optimal design of disability benefits by assessing the role of work restrictions. Specifically, I analyze an understudied disability benefit program which does not include work restrictions or means-testing. I ask whether benefits are well-targeted in the absence of these features, whether increases in benefit generosity are overall welfare-improving, and assess the effects of introducing work restrictions.

I find that the majority of disability spending is on low-income people with poor health, despite the lack of work restrictions and means-testing. The insurance value of the program is high relative to its distortion. Turning to reforms, I find that work restrictions always reduce welfare relative to not imposing work restrictions and across a reasonable range of benefit increases, increases in benefits are welfare-improving. The key mechanisms which drive these results are the large utility cost of applying for benefits limits the increase in claims when benefits become more generous, and that benefits become significantly less attractive to agents as they lose the right to work while they claim.

My research contributes to our understanding of the optimal design of disability benefit in two ways. Firstly, I show that a disability benefit program without work restrictions and means-testing can still be well-targeted and have high insurance value compared to its distortionary effect. Secondly, within my estimated model, introducing work restrictions makes the program overall *less* effective. This result suggests that, at least in a UK context, work restrictions are not an optimal feature of a disability benefit program. Theoretically, this implies that the reduction in moral hazard caused by introducing work restrictions is smaller than the loss of insurance value. Further research could apply a similar structural framework to the US and investigate whether the removal of work restrictions from US disability insurance is welfare-improving.

The major limitations of my research relate to model fit. There are key out-of-sample moments which my model cannot match. Most notably, my model overpredicts the propensity of claimants to work. As a result, I likely overestimate the fiscal benefits of work restrictions. Further research will refine the model, including part-time work, incorporating a more active spousal labor supply process, allowing for agents to choose to exit them program, and incorporating human capital accumulation.

Table 1.1: Comparing Disability Living Allowance and Personal Independence Payments

	DLA	PIP
Award length	71% indefinite	83% are $\leq$ 3.5 years
Conditions that automatically qualify	chronic & terminal	terminal only
In-person assessment	Mostly not	Almost always
Payments begin	Claim made	Decision made
Components	Care and mobility	Daily living and mobility
Means-tested	N	N
Taxable	N	N
Qualifies carer for CA	Y	Y
Max (weekly) Allowance	£142.15	£142.15

Notes: The table describes key stylized differences between Disability Living Allowance (DLA) and Personal Independence Payments (PIP). Indefinite awards are given until claimants report a change in circumstances, whereas definite awards require a reassessment at the end of the award period in order for the claimant to continue to receive benefits. The "components" of benefits refer to the dimensions on which an applicant is scored and can qualify for benefits on.

Table 1.2: Share of claimants by primary disability

Primary Disability	Share of claimants
Psychiatric disorders	0.37
Musculoskeletal disease (general)	0.20
Neurological disease	0.13
Musculoskeletal disease (regional)	0.12
Respiratory disease	0.042
Malignant disease	0.031
Cardiovascular disease	0.026
Visual disease	0.018
Endocrine disease	0.013
Hearing disorders	0.011
Gastrointestinal disease	<0.01
Genitourinary disease	<0.01
Skin disease	<0.01
Autoimmune disease (connective tissue disorders)	<0.01
Diseases of the liver, gallbladder, biliary tract	<0.01
Infectious disease	<0.01
Unknown or missing	<0.01
Haematological Disease	<0.01
Metabolic disease	<0.01
Multisystem and extremes of age	<0.01
Diseases of the immune system	<0.01

Notes: The share of claimants by main disabling conditions in August 2022. Data are from the Department for Work and Pensions (DWP).

Table 1.3: Summary statistics by claimant status, men and women

	Does not claim	Claims DLA	Claims PIP
	BHPS/US		
Age	49.887	55.870	50.576
% working	0.578	0.143	0.156
Income conditional on work	2,112.36	1,403.63	1,304.44
% highest qualification is higher degree	0.267	0.089	0.120
% highest qualification is a degree	0.123	0.091	0.103
% highest qualification is A-level	0.212	0.148	0.188
% highest qualification is GCSE	0.198	0.194	0.238
% no qualifications	0.089	0.146	0.140
% highest qualification is other qualification	0.111	0.332	0.211
Excellent health	0.127	0.029	0.013
Very good health	0.340	0.083	0.044
Good health	0.329	0.170	0.133
Fair health	0.160	0.330	0.319
Poor health	0.044	0.387	0.491
Equivalized household income	2,206.98	1,279.24	1,589.90
Equivalized social benefit income	390.88	582.43	665.42
	Family Resources Survey		
Age	49.068	54.842	54.953
Very good health	0.328	0.044	0.023
Good health	0.405	0.137	0.101
Fair health	0.208	0.316	0.274
Bad health	0.049	0.351	0.407
Very Bad health	0.010	0.152	0.196
Assets (except housing), 1000s	3.589	1.054	1.001
	ELSA		
Age	57.085	54.622	62.608
Excellent health	0.159	0.039	0.012
Very good health	0.330	0.135	0.096
Good health	0.311	0.269	0.208
Fair health	0.151	0.313	0.322
Poor health	0.049	0.244	0.362
Assets, 1000s	398.048	164.056	163.045

Notes: Income and wealth variables are expressed in 2015 pounds. The BHPS/US and Family Resources Survey are representative data sets. ELSA is representative of the over 50s population. Equivalized variables are adjusted by household size and composition. Flow variables such as income are expressed per month.

Table 1.4: Percentage of the population claiming and share of benefit spending by health status and income quintile

	Household income quintile				
	5th (Low)	4th	3rd	2nd	1st (High)
	Share claiming				
Poor health	0.44	0.48	0.39	0.32	0.24
Fair health	0.19	0.15	0.10	0.07	0.06
Good health	0.06	0.04	0.03	0.02	0.01
Very good health	0.03	0.02	0.01	0.01	0.01
Excellent health	0.03	0.01	0.01	0.01	0.00
	% of spending				
Poor health	0.18	0.14	0.06	0.03	0.01
Fair health	0.13	0.10	0.05	0.02	0.01
Good health	0.06	0.05	0.03	0.02	0.01
Very good health	0.03	0.02	0.02	0.01	0.01
Excellent health	0.01	0.01	0.01	0.01	0.00

Notes: Panel A shows the probability that a person claims either PIP or DLA conditional on self-reported health and household income quntile. Household income is equivalized and reported in terms of 2015 GBP. Panel B shows the share of program spending by each combination of health status and household income. Estimates are based on calculations from the BHPS/US data set.

Table 1.5: Associations between working and claiming benefits

	Dependent variable: indicator for working		
	(1)	(2)	(3)
DLA	-0.355^{***}	-0.249^{***}	-0.047^{***}
	(0.003)	(0.003)	(0.004)
PIP	-0.464^{***}	-0.327^{***}	-0.058^{***}
	(0.006)	(0.006)	(0.008)
R^2	0.345	0.363	0.738
Observations	370,125	370,125	370,125
Wave controls	Y	Y	Y
Age	Y	Y	Y
Health Status	N	Y	Y
Person FE	N	N	Y

Notes: Each column presents coefficients from a regression of whether a person works on an indicator for whether they claim benefits, with various controls (see equation 1.1). Data come from the British Household Panel Survey/Understanding Society data set. In specification (1) I control for a full set of age dummies and wave dummies. Specification (2) additionally controls for self-reported health using a full set of dummies. In Specification (3) I control for person fixed effects. Standard errors are clustered at the individual level. *p<0.1; **p<0.05; ***p<0.01

Table 1.6: Value of key paramters, sufficient statistics approach

Parameter	Interpretation	Values
ϵ_{TI}	Elasticity of taxable income	0.2, 0.35, 0.5
σ	Coefficient of relative risk aversion	1,2
c_W	Weekly equivalized consumption, non-claiming workers	534.13
c_B	Weekly equivalized consumption, claimants	211.49
$\frac{\partial I_i}{\partial DI}$	Income effect of benefits on labor income	$0, -0.048$
π_B	Share of the population claiming benefits	0.057
π_W	Share of the population working and not claiming benefits	0.54

Notes: Values of key parameters used in equation 1.6. For some parameters, I use multiple values to assess sensitivity to different assumptions. Values for the elasticity of taxable income come from various estimates in the reduced-form literature. The coefficient of risk aversion is set to 2, which is a common value in the literature (Keane et al., 2020), and a conservative value of 1. The income effect of benefits on labor income is based on a theoretical bound of 0, and an (absolute) upper bound from a regression of the probability of working on a binary variable for claiming PIP or DLA with person fixed effects and controls for age, health and survey wave.

Table 1.7: The welfare effect of marginal increases in benefit generosity

Coefficient of relative risk aversion	Taxable income elasticity	Income effect of marginal benefit expansion	Marginal social benefit of 1 GBP expansion
1	0.200	0 (theoretical upper bound)	0.063
1	0.200	−0.048 (estimated lower bound)	0.062
1	0.350	0	0.050
1	0.350	−0.048	0.049
1	0.500	0	0.039
1	0.500	−0.048	0.038
2	0.200	0	0.420
2	0.200	−0.048	0.419
2	0.350	0	0.367
2	0.350	−0.048	0.366
2	0.500	0	0.325
2	0.500	−0.048	0.324

Notes: The marginal social benefit of making benefits more generous by GBP 1, using different sets of assumptions. I use equation 1.6 to estimate the additional net revenue per person the government would derive from an *ex ante* welfare-neutral marginal increase in benefits and tax rise. The income effect of marginal benefit expansion refers to the reduction in annual income resulting from a GBP 1 increase in benefits.

Table 1.8: Data moments used to identify coefficients using Simulated Method of Moments

Parameters	Moments	# moments	Data source
ω_c,ω_{h0},$\omega_{h=2}$,$\omega_{h=3}$ $\omega_{h=3,p=1}$,ω_t,$\omega_{t,h=2}$ $\omega_{t,h=3}$,$\omega_{t,h=3,p=1}$	Probability of working by age and health status	156	US
ω_{app}	Probability of applying for benefits by age	39	DWP
α_0,α_1,α_2,α_3	Probability of claiming benefits by health status, age	156	US
δ_0,δ_1,δ_2,δ_3,δ_4,δ_5,δ_6	Average log wage by age, health status, education	351	US
ρ	1st-10th autocorrelation log wage residual	10	US
V_ζ	Log wage residual by age	39	US
β_1,β_2,β_3,ψ_1, ψ_2	Asset accumulation by education	108	ELSA

Notes: I use the Simulated Method of Moments to identify parameters which cannot be estimated directly from data or calibrated from widely-used values. I choose parameters to minimize the variance-weighted squared difference between data moments and moments simulated from my model (as in equation 1.21).

Table 1.9: Structural parameters

	Parameter	Estimate
	Death/survival logit parameters	
κ_0	Intercept	-8.51
κ_1	Age coefficient	-0.00
κ_2	Age-squared coefficient	0.0008
κ_3	$h_{it} = 2$ coefficient	2.53
κ_4	$h_{it} = 2\times$ age coefficient	0.13
κ_5	$h_{it} = 2\times$ age-squared coefficient	-0.0011
κ_6	$h_{it} = 3$ coefficient	1.95
κ_7	$h_{it} = 3\times$ age coefficient	0.04
κ_8	$h_{it} = 3\times$ age-squared coefficient	-0.0053
	Spousal income parameters	
ϕ_0	Intercept	11.82
ϕ_1	Age coefficient	0.54
ϕ_2	Age squared coefficient	-0.006
ϕ_3	Education = 2 coefficient	-3.86
ϕ_4	Education = 3 coefficient	-7.07
	Other benefit income parameters	
μ_0	Intercept, if over 65	11.99
μ_1	$h_{it} = 2$, if over 65	1.42
μ_2	$h_{it} = 3$, if over 65	3.69
μ_3	Intercept, if under 65	8.09
μ_4	Coefficient on labor income, if under 65	-0.12
μ_5	Coefficient, $h_{it} = 2$, if under 65	3.13
μ_6	Coefficient, $h_{it} = 2\times$ labor income, if under 65	-0.08
μ_7	Coefficient, $h_{it} = 3$, if under 65	5.60
μ_8	Coefficient, $h_{it} = 3\times$ labor income, if under 65	-0.15
	Loss of DLA/PIP logit parameters	
λ_0	Intercept	-1.09
λ_1	Coefficient, $h_{it} = 2$	-0.30
λ_1	Coefficient, $h_{it} = 3$	-0.31
λ_1	Coefficient, persistent health	0.45
λ_1	Coefficient, $h_{it} = 2\times$ persistent health	-0.38
λ_1	Coefficient, $h_{it} = 3\times$ persistent health	-1.01

Notes: Death/survival parameters are estimated using a logit model. Spousal income, other benefit income and loss of DLA/PIP parameters are estimated using regressions. Wage parameters, conditional application success parameters and preference parameters are solved using the Simulated Method of Moments, i.e., minimizing the distance between observed and simulated moments listed in Table 1.8. This is equivalent to numerically solving the problem in equation 1.21. Simulated moments are weighted by age and education to be comparable with representative survey data.

Table 1.9: Structural parameter estimates

	Parameter	Estimate
	Other parameters calibrated outside the model	
y^{PIP}	DLA/PIP benefit payments (1000s)	6.72
r	Interest rate	0.0155
σ	Relative risk aversion	2
	Wage parameters	
δ_0	Intercept	1.986
δ_1	Age	0.08
δ_2	Age squared	-0.001
δ_3	Fair health	-0.451
δ_4	Bad health	-0.896
δ_5	A-level/GCSE	-0.436
δ_6	No/other qual	-0.546
ρ	Shock persistence	0.99
V_ζ	Shock variance	0.378
	P(success | application) logit parameters	
α_0	Intercept	-4.377
α_1	Fair health	0.17
α_2	Bad health	5.705
α_3	Persistent bad health	0.005
	Preference parameters	
ω_c	Consumption weight	354.61
ω_{h0}	Work disutility	-0.649
$\omega_{h=2}$	Work disutility, fair health	-1.82
$\omega_{h=3}$	Work disutility, bad health	-1.85
$\omega_{h=3,p=1}$	Work disutility, persistent bad health	-2.491
ω_t	Work disutility, age trend	-0.050
$\omega_{t,h=2}$	Work disutility, age trend $\times$ fair health	-0.010
$\omega_{t,h=3}$	Work disutility, age trend $\times$ bad health	-0.010
$\omega_{t,h=3,p=1}$	Work disutility, age trend $\times$ persistent bad health	-0.012
ω_{app}	Application disutility	-5.703

Notes: Death/survival parameters are estimated using a logit model. Spousal income, other benefit income and loss of DLA/PIP parameters are estimated using regressions. Wage parameters, conditional application success parameters and preference parameters are solved using the Simulated Method of Moments, i.e., minimizing the distance between observed and simulated moments listed in Table 1.8. This is equivalent to numerically solving the problem in equation 1.21. Simulated moments are weighted by age and education to be comparable with representative survey data.

Table 1.9: Structural parameter estimates

	Parameter	Estimate
	Patience parameters	
$\beta_{e=1}$	Discount rate, degree or equiv	0.988
$\beta_{e=2}$	Discount rate, A-level/GCSEs	0.970
$\beta_{e=3}$	Discount rate, no/other qual	0.948
	Bequest motive	
ψ_0	Bequest motive weight	−23,844.860
ψ_1	Bequest motive shifter	8.666

Notes: Death/survival parameters are estimated using a logit model. Spousal income, other benefit income and loss of DLA/PIP parameters are estimated using regressions. Wage parameters, conditional application success parameters and preference parameters are solved using the Simulated Method of Moments, i.e., minimizing the distance between observed and simulated moments listed in Table 1.8. This is equivalent to numerically solving the problem in equation 1.21. Simulated moments are weighted by age and education to be comparable with representative survey data.

Table 1.10: Unmatched moments: benefit transitions and claimant labor force participation

	model	data
$\Pr(b_{it} = 1 \vert b_{it-1} = 0)$		
good health	0.00	0.02
fair health	0.00	0.03
non-persistent bad health	0.24	0.08
persistent bad health	0.28	0.16
$\Pr(\ell_{it} = 1 \vert b_{it} = 1)$		
good health	0.640	0.245
fair health	0.505	0.134
non-persistent bad health	0.263	0.118
persistent bad health	0.209	0.095

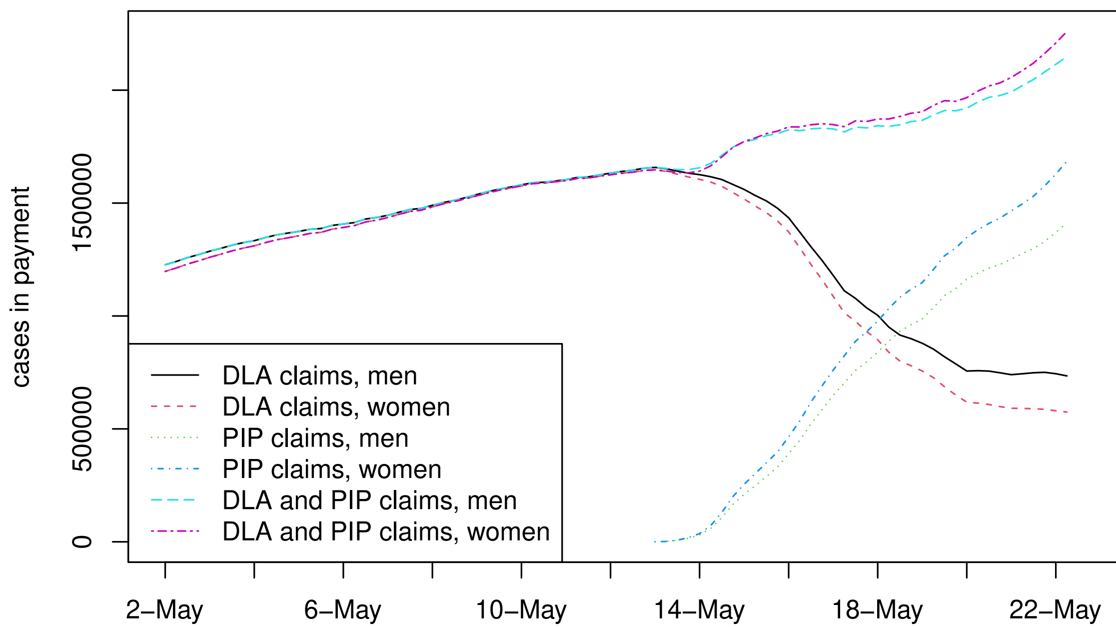
Notes: The probability of transitions onto benefits and the share of people working conditional on claiming in the simulated and observed data. Simulated moments are weighted by age and education to be comparable to the representative survey data.

Table 1.11: Effects of policy reform

	Baseline	Doubling benefits	Work restrictions
Income sources (average per period)			
DLA/PIP payment per person (GBP)	230	572	123
Average labor income (GBP 1000s)	23.074	23.042	23.037
Average assets aged 65, (GBP 1000s)			
Good health	535.279	530.324	534.939
Fair health	529.964	526.937	528.777
Bad health	518.682	522.081	516.081
Welfare equivalent % change in (lifetime) consumption			
Degree or equivalent		2.144	−1.303
A-level/GCSE		2.203	−1.428
No/other qualification		1.868	−1.647

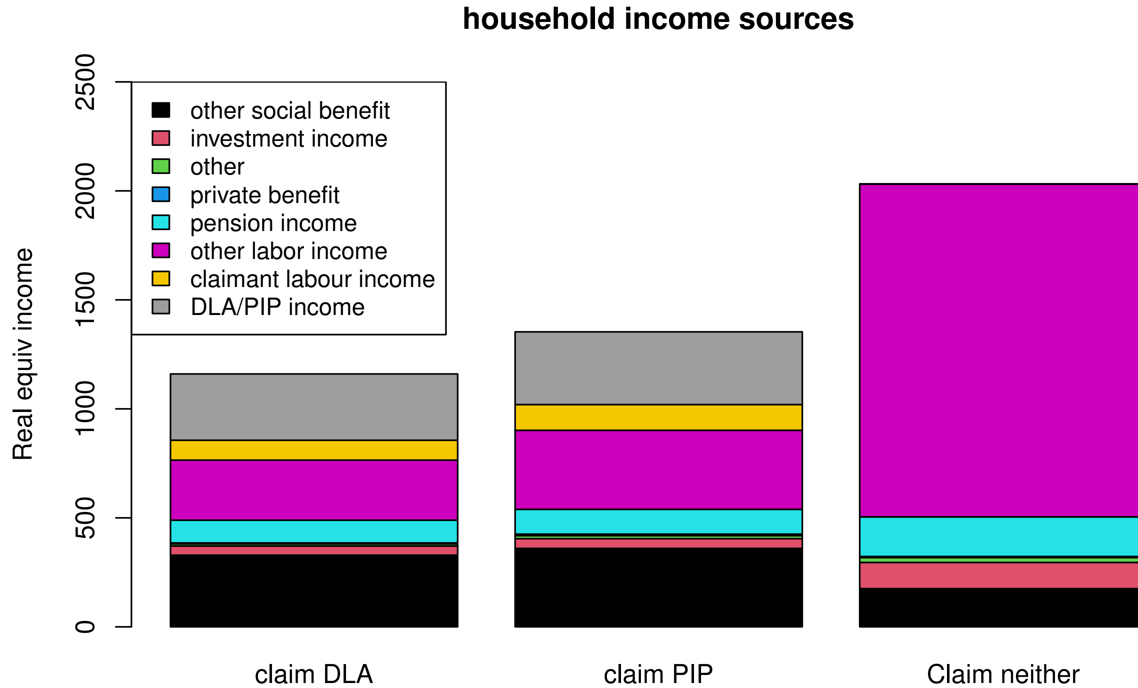
Notes: Simulated effects of various policy reforms on average annual DLA/PIP payments per person (including non-claimants), average household labor income, average household assets when the man is aged 65, and welfare equivalent changes in consumption. Simulated moments are weighted by age and education to be comparable to the representative survey data. All policy changes are revenue neutral conditional on education, i.e., I impose education-specific increases or decreases in taxes such that overall spending less tax receipts are unchanged for each education group. Male labor supply changes endogenously within the model. Spousal labor supply is simulated using an assumed elasticity of taxable income of 0.3. Doubling benefits means increasing benefit monetary generosity by 100% and imposing additional taxes until the government's deficit is overall unchanged. Imposing work restrictions means that if a claimant chooses to work, he forgoes benefits this period and does not have the right to claim next period. I calculate the welfare equivalent change in lifetime consumption by calculating the percentage change in consumption each period, holding all other decision-making fixed, which would be required to effect the same change in expected lifetime utility as the policy change.

Figure 1.1: PIP/DLA cases in payment



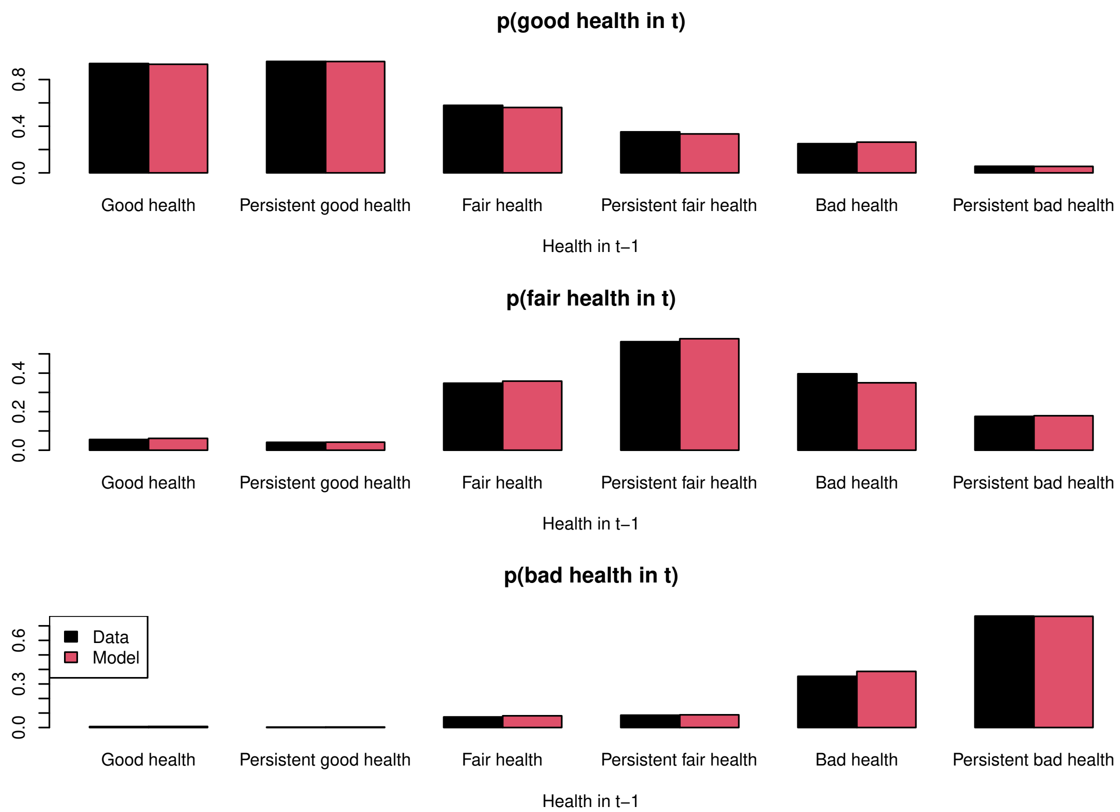
Notes: The number of claimants of Disability Living Allowance (DLA) and Personal Independence Payments (PIP) ("cases in payment") by sex over time. Data from the Department for Work and Pensions (DWP).

Figure 1.2: Sources of household income, claimants and non-claimants



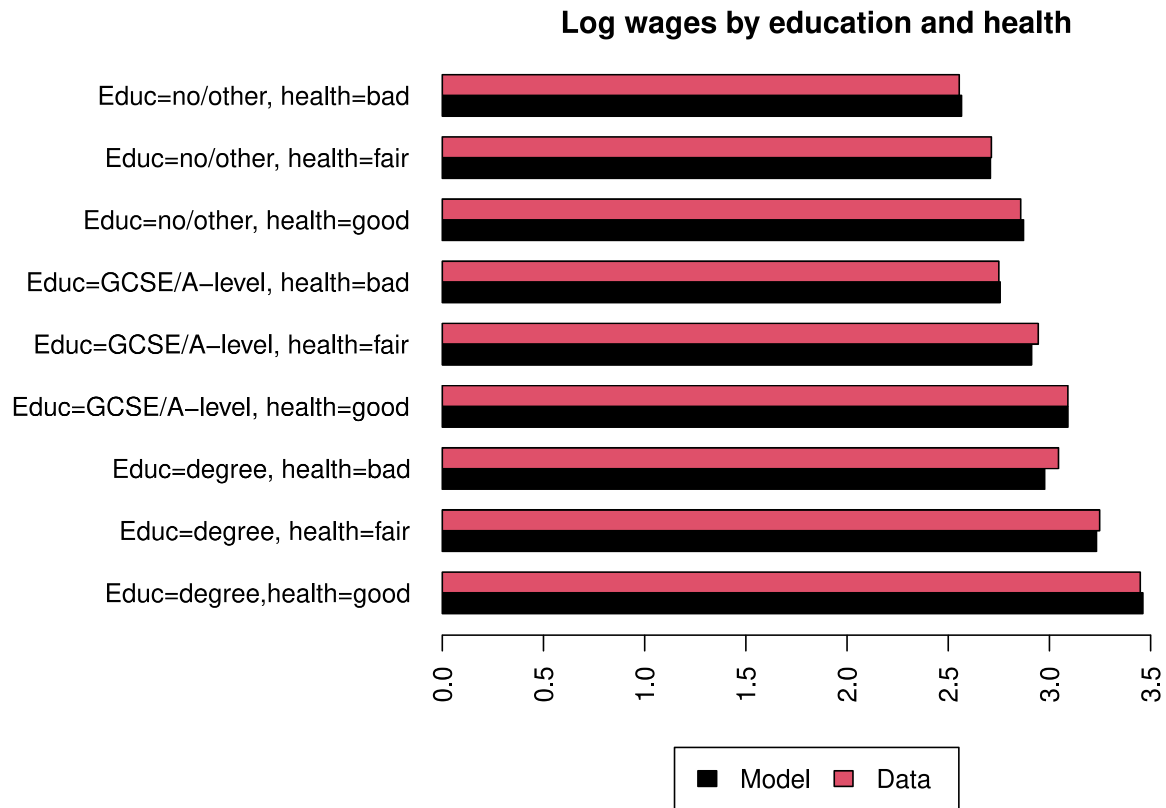
Notes: Sources of household income by claimant status. Data are from the British Household Panel Survey/Understanding Society. Claimant labor income is labor income earned by someone who claims either DLA or PIP. Other labor income is labor income from other members of the household. Other social benefits refers to government benefit payments other than DLA or PIP. Private benefits are payments from Trade Unions, Employers or Friendly Societies. Pension income refers to income from private pension plans. All values are equivalized by household size and expressed in 2015 pounds.

Figure 1.3: Model fit of transitions between health states



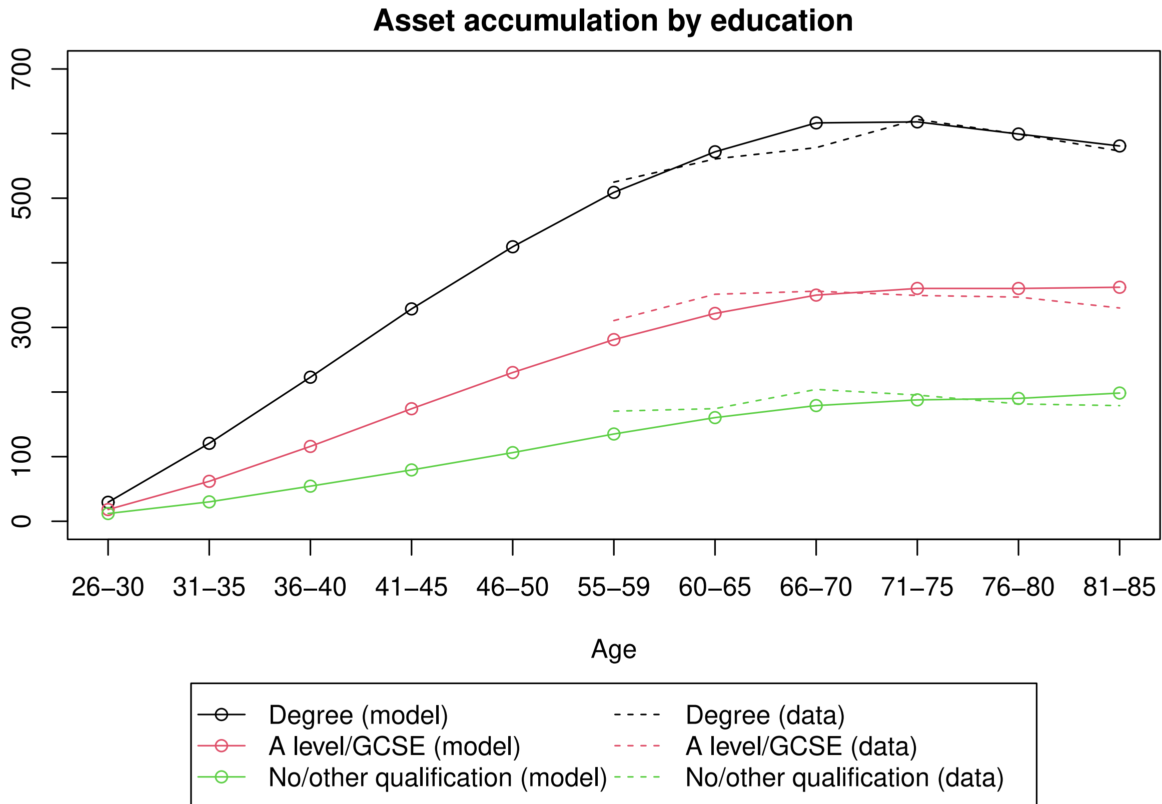
Notes: Transitions between health states for under 51s, data and simulations. Persistent health states are health states an agent has been in for more than one year. The health process is a multinomial logit (see equation 1.20). Simulated moments are weighted by education and age to be comparable with survey data.

Figure 1.4: Model fit of average wages by education and health status



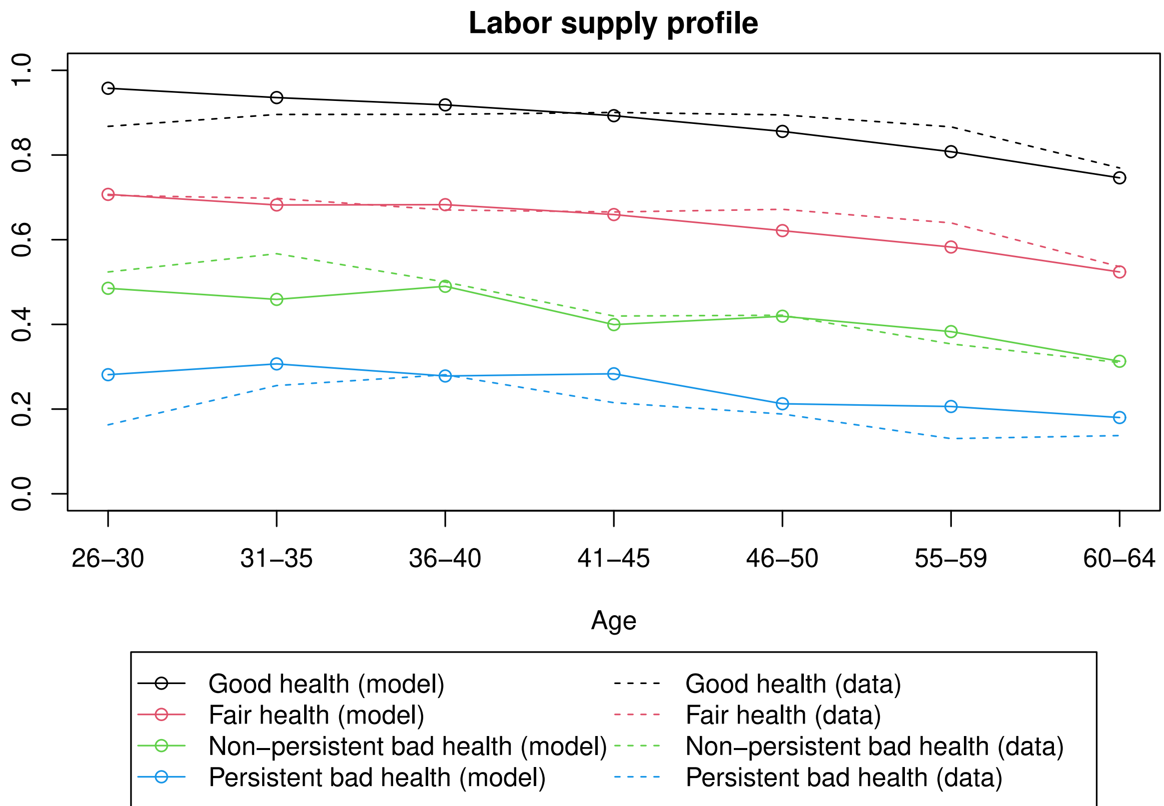
Notes: Average wages by health and education in the data and simulations. Education is measured by the highest qualification achieved, and can be either a degree or equivalent, A-levels or GCSEs, or no or other (non-academic) qualifications. Simulated moments are weighted by education and age to be comparable with survey data.

Figure 1.5: Asset profile by education



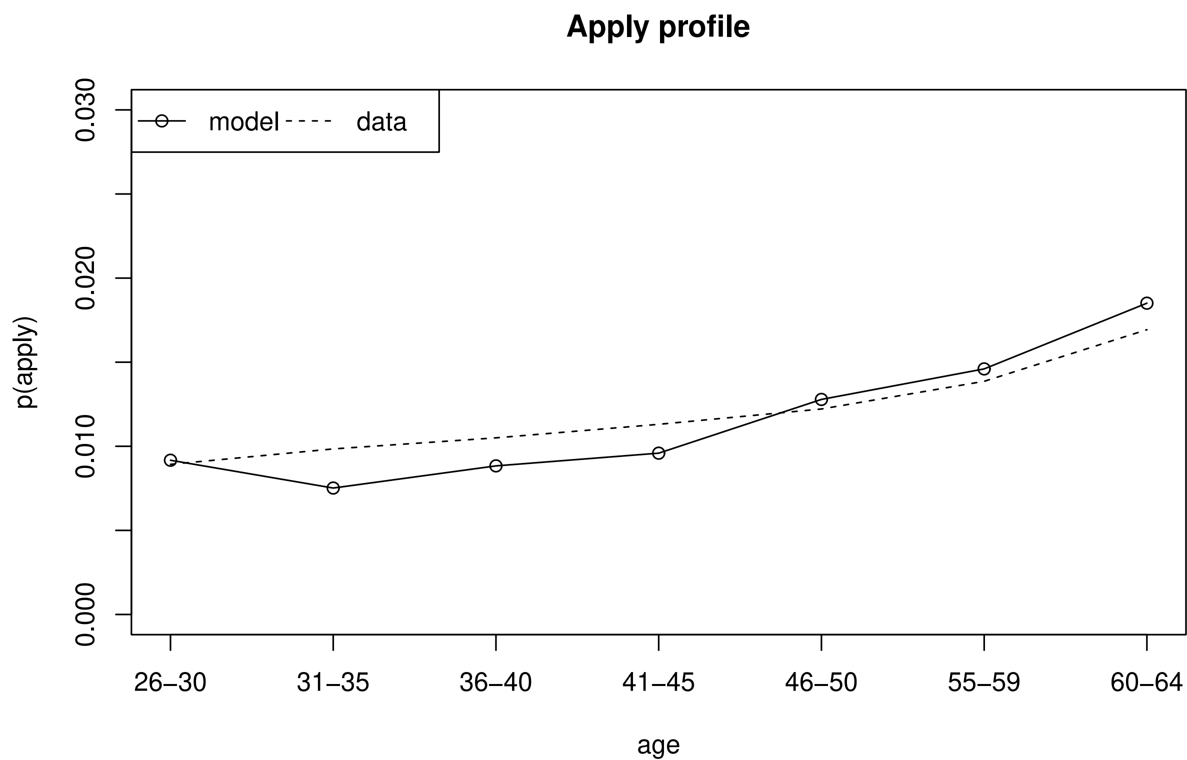
Notes: Asset ownership by 5-year age bin and education, simulated and observed. Education categories are defined in the note to Table 1.4. Simulated moments are weighted by education and age to be comparable with survey data. Age profiles are constructed from the data by regressing asset ownership on age dummies, controlling for wave fixed effects to control for macroeconomic trends, separately for each education category.

Figure 1.6: Labor market participation across the life cycle



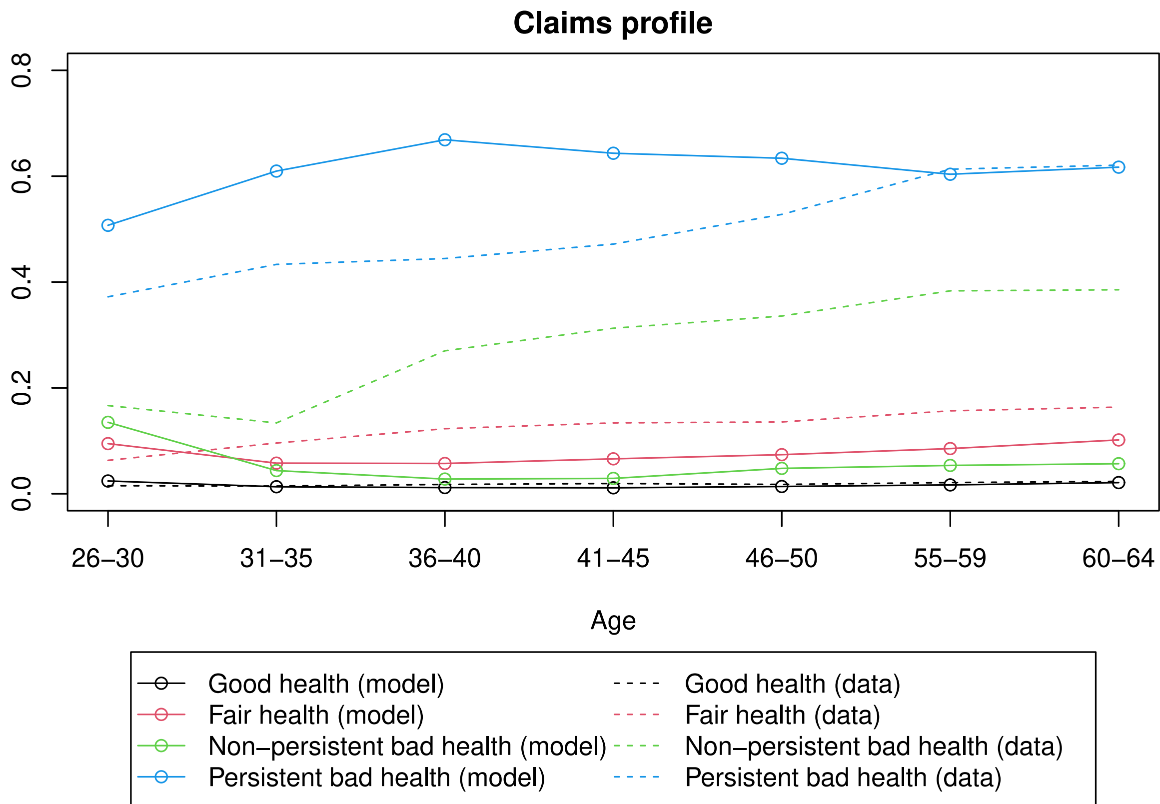
Notes: Labor force participation by health status 5-year age bin. Simulated moments are weighted by education and age to be comparable with survey data. Bad health is persistent if a person is in bad health for more than one period.

Figure 1.7: Life cycle of benefit applications



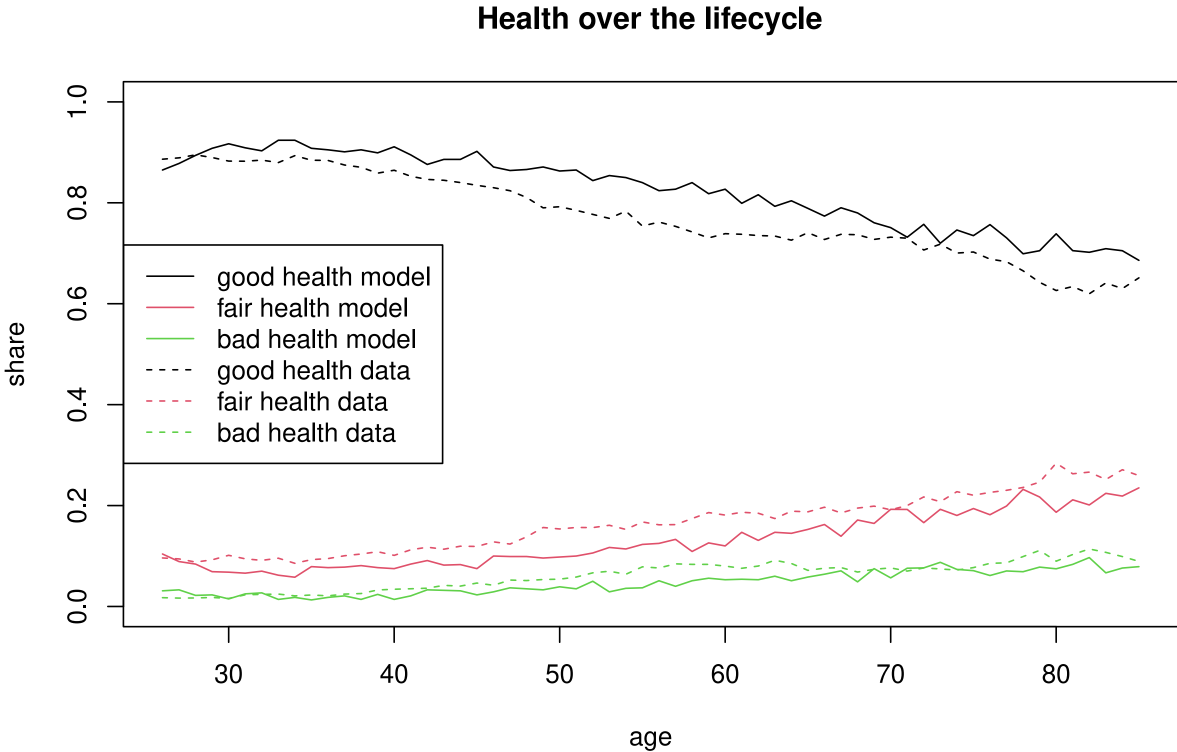
Notes: Share of people applying for benefits by 5-year age bin. Simulated moments are weighted by education and age to be comparable with survey data.

Figure 1.8: Life cycle of benefit claims



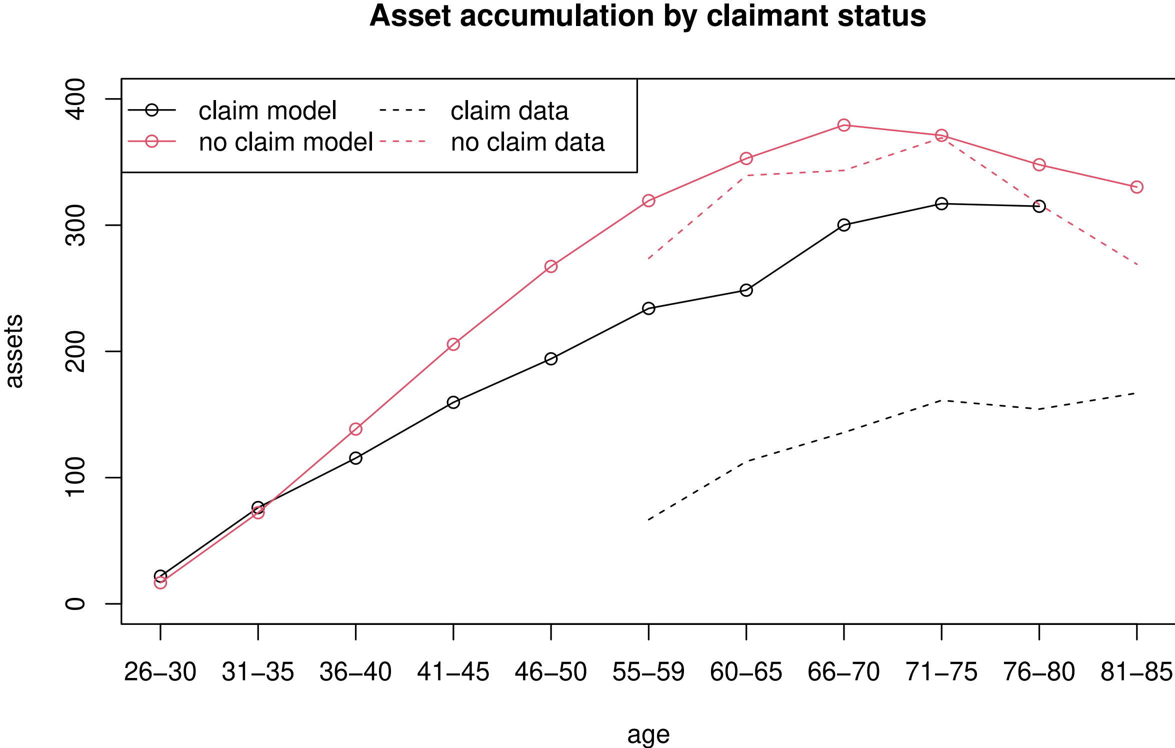
Notes: Probability of claiming by 5-year age bin and health status, simulated and observed. Simulated moments are weighted by education and age to be comparable with survey data. Bad health is persistent if a person is in bad health for more than one period.

Figure 1.9: Health status over the life cycle



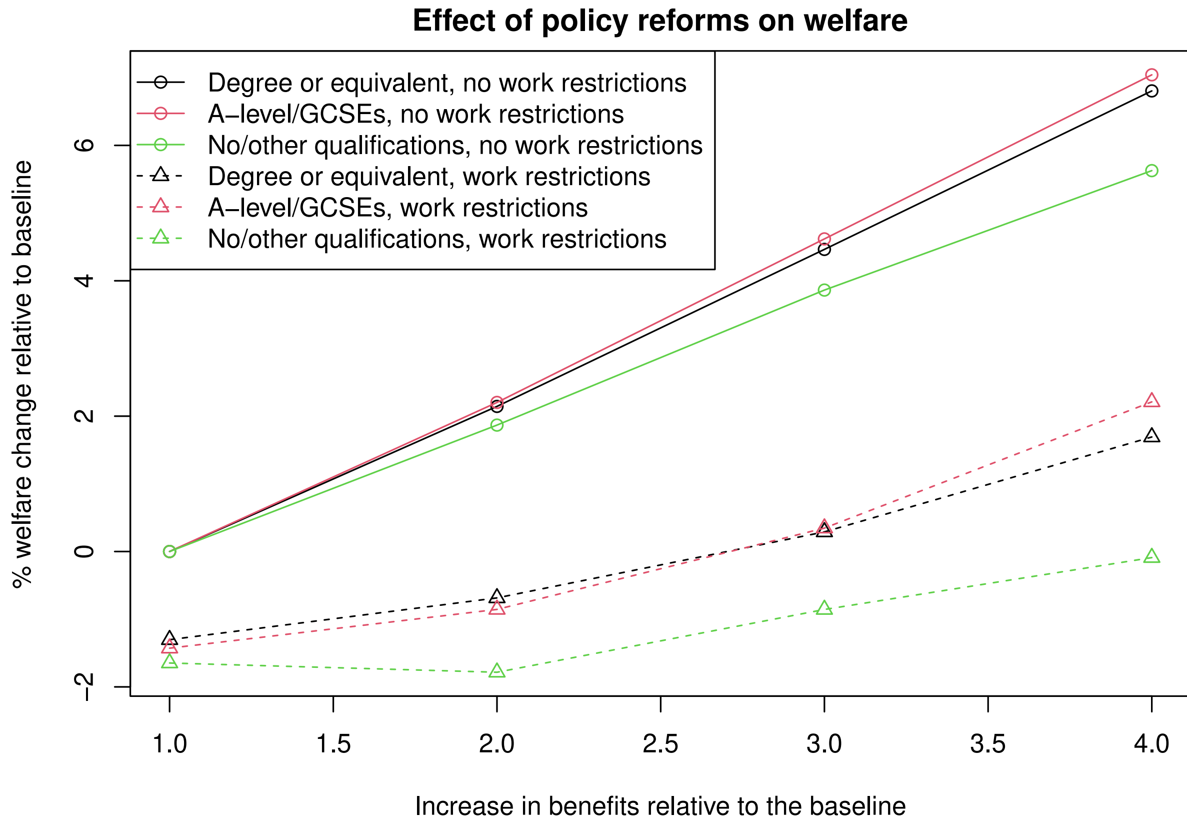
Notes: Probability of being in each health status by age, simulated and observed. Simulated moments are weighted by education and age to be comparable with survey data.

Figure 1.10: Asset accumulation by claimant status



Notes: Asset ownership by claimant status, simulated and observed. Simulated moments are weighted by education and age to be comparable with survey data. Age profiles are constructed from the data by regressing asset ownership on age dummies, controlling for wave fixed effects to control for macroeconomic trends, separately by claimant status.

Figure 1.11: The welfare effects of combining work restrictions and increases in benefits



Notes: Simulated welfare effects of policy changes. Circular points show the effect of multiplying benefits relative to the baseline, and triangular points show the effect of the combination of this policy and work restrictions. Work restrictions are imposed by removing all people who work from benefits until they re-apply. Welfare effects are calculated separately by education, and are calculated as the percentage change in lifetime consumption which has the equivalent effect on average expected lifetime utility at age 26.

Chapter 2

The causal effects of workplace autonomy on mental health

2.1 Introduction

Poor mental health is a significant source of global disease. In the United Kingdom (UK) nearly half the population has a diagnosable psychiatric disease in their lifetime, with many patients attributing poor mental health to the workplace environment.[1] Autonomy at work is a key aspect of the workplace environment. Work-related autonomy refers to the level of control a worker has over their working conditions and the nature of their work. There is evidence that workers value work-related autonomy (Dube et al., 2022), and psychologists have argued that it might be important in explaining mental health outcomes (e.g., Stansfeld et al., 1999).

I estimate the causal effect of different aspects of work-related autonomy on mental health. I consider two clinical screening measures of mental health: the GHQ12 caseness score and the SF12 mental health index. I exploit within-person, within-occupation variation in work-related autonomy and include a rich set of controls including recent mental health history, marital status, and age. I show that low levels of work-related autonomy are consistently associated with poor mental health, and the inclusion of observable controls only moderately decreases the association despite increasing the R-squared substantially. This result could indicate the existence of a causal effect, but one may be concerned about remaining confounders. Therefore, using Oster's (2019) method of selection on unobservables, I calculate the causal effect of work-related autonomy on mental health under various assumptions about the degree of selection on unobservables, and the percentage of variation in the dependent variable explained by the treatment and its confounders. I find that low work-related autonomy has an adverse effect on mental health under most assumptions. The selection bias attributable to unobserved confounders would need to be larger than the selection bias attributable to occupation and person fixed effects to conclude that there is no causal effect. For individual components of autonomy, my results are generally less robust to different assumptions about the degree of confounding. I find an adverse effect of low autonomy over work tasks

[1] https://www.hse.gov.uk/statistics/causdis/stress.pdf

on mental health under most assumptions, though these results are not always statistically significant. I find a statistically significant adverse effect of low autonomy over work hours on women's mental health under all assumptions.

I investigate the potential for reverse causality, that is, whether a deterioration in mental health, holding occupation and person fixed effects constant, can cause decreases in work-related autonomy. Reverse causality could happen if workers who experience adverse mental health shocks choose to switch to low-autonomy jobs, or if firms impose lower levels of autonomy on workers who experience negative mental health shocks. On the first point, I restrict my sample to those who do not change employer, and show that the estimated associations are similar. Thus, it is unlikely that my results are driven by workers switching employer to change their work-related autonomy as a result of changes in mental health. One interpretation of this result is that workers with poor mental health do not necessarily desire lower levels of autonomy. To address the second concern, I consider the possibility that employer adjustments of working conditions for those with mental illness drive my results. I argue that it is unlikely that employers systematically reduce the autonomy of workers with poor mental health for two reasons: firstly, the association between mental health and autonomy is present even at low and moderate levels of symptoms (that are unlikely to be detectable by employers). Secondly, UK employment law requires changes in working conditions to accommodate mental illness which are as likely to *increase* work-related autonomy as decrease it.

My results suggest that short- and medium-run trends in working conditions in the UK could have important implications for mental health. Firstly, in the public sector, currently composed of 5.8 million employees,[2] work practices have tended towards tighter supervision and greater use of targets (Le Grand, 2003). Secondly, a shift in the industrial composition of the UK economy has increased the surveillance of employees. For example, there has been a large increase in employment in call centers, where very high levels of monitoring are used to increase productivity (Bain and Taylor, 2000, Burgess and Connell, 2004). My results suggest that both of these trends could have adverse population-level effects on mental health outcomes.

On the other hand, more recently, the COVID-19 pandemic led to a large increase in remote and hybrid working, with almost half of UK workers working remotely at the height of the pandemic, and a majority of them continuing to work either hybrid or remotely.[3] Remote working increases one's work-related autonomy. One has more control over the exact environment in which one works, and may have increased flexibility over when one works. On the other hand, adaptations to working practices since 2020 have seen employers claw back some control with additional surveillance (Aloisi & De Stefano, 2022). To the extent that the shift to remote working engenders a permanently higher level of work-related autonomy, this development could have positive consequences for overall population health.

Finally, the rise of the "gig economy" (Woodcock & Graham, 2019) could increase or decrease average employee autonomy. On the one hand, there is an increase in a worker's autonomy about exactly when

[2] https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/publicsectorpersonnel/bulletins/publicsectoremployment/september2022#public-sector-employment-data

[3] https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/ishybridworkingheretostay/2022-05-23. Hansen, Lambert, Bloom, Davis, Sadun, and Taska (2023) also document an uptick in rates of employment opportunities offering remote or hybrid work across English-speaking countries since the pandemic.

to do work. On the other hand, conditional on accepting a job, the specific online platform may dictate the manner and pace of work. The contrast illustrates the importance of differentiating between various dimensions of autonomy. My results could have implications for the optimal regulatory response to the gig economy.

My research contributes to a growing literature which relates working conditions to health outcomes (Ravesteijn et al., 2018, Michaud and Wiczer, 2018), including mental health (Jolivet & Postel-Vinay, 2020). A common finding is that greater levels of work-related autonomy cause higher overall life satisfaction (Coad & Binder, 2014), better mood (Johannsen & Zak, 2020), and better clinical mental health outcomes (Stansfeld et al., 1997, Stansfeld et al., 1999, Bentley et al., 2015, Carrino et al., 2020, Belloni et al., 2022). Work-related autonomy may be valuable because it delivers a wider "locus of control" (Rotter & Mulry, 1965), which is theorized to cause better mental health (Bandura, 1997, Carr, 2013, Frenkel et al., 1995, Churchill et al., 2020). Alternatively, higher work-related autonomy may be a marker of status or "work dignity" (Dube et al., 2022). I contribute to this literature in three ways. Firstly, my research estimates the effect of work-related autonomy on mental health using a representative sample of the working-age population. By contrast, previous literature has estimated the effect of work-related autonomy on outcomes other than mental health such as life satisfaction (Coad and Binder, 2014), and mood (Johannsen and Zak, 2020), estimated the effect within particular occupations such as civil servants (Stansfeld et al., 1997, Stansfeld et al., 1999), estimated the effect of "skills and discretion", a variable which includes aspects of autonomy but also features such as task complexity (Belloni et al., 2022), and investigated the effect of working in a job with low control over decisions amongst women nearing retirement age (Carrino et al., 2020). The closest paper to mine is Bently et al. (2015) who estimate associations between work control and mental health outcomes in a representative sample of the Australian population. My paper builds on theirs by using Oster's (2019) methodology, which allows me to assess the robustness of my results to potential unobserved confounders. Secondly, where most previous literature has used a single overall summary measure of workplace autonomy, my data allow me to disaggregate overall work-related autonomy into 5 areas and show which are most important. Thus, I provide a more nuanced analysis of the relationship between work-related autonomy and mental health. Thirdly, I demonstrate that there is significant within-occupation variation in work-related autonomy and exploit variation within occupations. The majority of previous research exploits variation across occupations (e.g., Carrino et al., 2020), occupation-specific trends (e.g., Belloni et al., 2022), or variation within one specific occupation which may not generalize to the wider population (e.g., Stansfeld et al., 1997, Stansfeld et al., 1999). My data allow me to control for occupation fixed effects which avoids confounding with other occupation-specific characteristics.

This paper also relates to a literature estimating the value that workers place on flexibility about when they work. Employees are often found to place high value on flexibility over working hours (Mas and Pallais, 2017, He et al., 2021) and it is associated with higher workplace satisfaction (Lyness et al., 2012). Greater flexibility over when one works has also been linked to greater productivity in some contexts (Chen et al., 2019), which may enhance work satisfaction. I contribute to this literature by showing that greater autonomy over work hours causes improvements in overall mental health amongst women. This

finding suggests that women may value hours flexibility because of its beneficial effect on their mental health. This result is particularly interesting when contrasted against the finding that selection into jobs with greater flexibility over hours contributes substantially to the gender wage gap (Goldin, 2014).

This paper proceeds as follows: section 2.2 describes key features of the data, and sample construction. Section 2.3 exploits within-person, within-occupation variation in autonomy to estimate the effect of autonomy on mental health. Section 2.4 discusses Oster's (2019) method and presents the results of applying it to my estimates. Section 2.5 discusses the potential for reverse causality, and argues it is unlikely to be a major driver of results. Section 2.6 discusses implications and concludes.

2.2 Data

I use Understanding Society data, a large, representative, panel data set of the UK population. The data cover the years 2009-2021. Respondents are surveyed annually. The data provide information about labor market outcomes and history, health outcomes (including mental health), demographic characteristics including race, gender, marital status, and education. The data include very detailed information about a person's work-related autonomy, but respondents are only asked about work-related autonomy in even waves. There is roughly 5% sample attrition between waves, and new respondents are recruited in order to maintain the representativeness of the survey.

Mental health variables

I consider two summary measures of mental health: the GHQ12 caseness and the SF12 mental health summary.

The GHQ12 caseness score is constructed from the general health questionnaire. Respondents are asked 12 questions about their current mental health symptoms, which are scored from 1 (best) to 4 (worst). The caseness score is the number of symptoms on which a person scores 3 or higher (McCabe et al., 1996). It ranges from 0 to 12, with higher numbers indicating worse mental health symptoms. The GHQ12 caseness score has been used in various contexts to screen for mental illness (e.g., Gureje and Obikoya, 1990, Anjara et al., 2020) and to measure mental health in economics research (e.g., Gathergood, 2013, Belloni et al., 2022).

The SF12 mental health index is a summary of mental health status derived from the SF12 questionnaire. The SF12 involves a more comprehensive series of questions about health status. The scoring system applied to the answers gives a total mental health score which ranges between 0 (the worst possible mental health) and 100 (the best possible mental health) (Jenkinson and Layte, 1997). The SF12, like the GHQ12 caseness score, is used as a screening measure for mental health problems (Kontodimopoulos et al., 2007, Tibubos and Kröger, 2020), and has been used previously by economists studying mental health (e.g., Davalos and French, 2011, Wallace et al., 2016, Jolivet and Postel-Vinay, 2020).

The individual questions relating to each mental health measure and scoring system for the SF12 mental health measure are shown in Appendix B.1. Both measures are intended to capture mental health,

broadly defined. There is a a -0.66 correlation between the two measures, suggesting significant overlap. However, the GHQ12 caseness covers a larger range of symptoms than the SF12, such as the effect of mental health on sleep and ability to concentrate.

Work-related autonomy

I construct a set of binary variables which describe the autonomy a person has at work across several dimensions. The autonomy variables must be binary to maintain a causal interpretation of the OLS estimator if controls are required to purge omitted variable bias (Goldsmith-Pinkham, Hull and Kolesár 2021).[4]

The Understanding Society data set has five questions about work-related autonomy, asked in waves 2, 4, 6, 8, and 10. The questions are as follows:

1. In your current job, how much influence do you have over what tasks you do in your job?
2. In your current job, how much influence do you have over the pace at which you work?
3. In your current job, how much influence do you have over how you do your work?
4. In your current job, how much influence do you have over the order in which you carry out tasks?
5. In your current job, how much influence do you have over the time you start or finish your working day?

The respondent is asked to choose between one of four answers: "a lot", "some", "a little", or "none".

I construct 6 autonomy variables as follows: for each autonomy question, I score a person-observation as 1 if they report having "a little" or "none" of this kind of autonomy, and 0 otherwise. This procedure yields 5 autonomy variables which relate to different aspects of autonomy that one might have at work. I then construct an overall autonomy variable, which acts as a summary of the amount of work-related autonomy a person has. This variable is 1 if a person answers "a little" or "none" to three or more of the autonomy questions, and 0 otherwise.[5]

[4]Goldsmith-Pinkham, Hull and Kolesár (2021) show that if a treatment takes more than two values (i.e., more than "treated" and "untreated") and is randomly assigned conditional on controls, then regressing the dependent variable of interest on the treatment with controls does not necessarily uncover the causal effect of the treatment on the dependent variable. If, however, treatments can be measured in a binary fashion, one can recover estimates with a causal interpretation.

[5]I assess whether results are robust to using cutoffs of one, two, four or five in appendix B.2 (Table B.21). My results are broadly consistent with my main results (Tables 2.3 and 2.4). Across all estimates, having low autonomy is associated with worse mental health, however low autonomy is defined. Furthermore, adding in new controls does not radically change any of the estimates, which is also consistent with the results in the text.

Control variables

I use a number of control variables to assess the degree of confoundedness of the relationship between mental health and work-related autonomy. I control for age at last birthday. This variable is important since both mental health and job experience can vary over the life cycle. The survey reports education as the highest qualification a person has received, and marital status as married, never married, divorced, or in a civil partnership. Both education and marital status are potentially correlated with both working conditions and mental health. I control for both of these variables with a full set of dummies for each value they can take. The occupation of a person's current main job is reported at the ISCO 3-digit level. Controlling for occupation means holding fixed all characteristics of jobs which vary only at the occupation level (these characteristics might include, for example, the level of physical risk or regularity of interaction with other people). Wave 3 of the Understanding Society data reports a person's score on the "big five" personality indices.[6] Personality variables may be correlated with both job choice and mental health. I assume that personality is unchanged over time and assign a person's wave 3 scores on the personality indices to all other waves in which they appear.[7] For those in work, labor income is reported. Controlling for labor income addresses the concern that high or low autonomy jobs may simply be higher or lower paying. I use log labor income expressed in 2015 GBP. The survey also reports how many biological children a respondent has living with them. I use a dummy variable for whether a person has biological children living with them, since parental status can affect both mental health and the choice of job.

Sample construction and statistics

I limit the sample to those who are continuously in work (that is, report being employed in every survey wave in which they are interviewed). I make this decision for two reasons. Firstly, including those who move in and out of work could create a selection problems: if employment and mental health are jointly determined, and those entering the labor force disproportionately work in jobs with, say, low levels of autonomy, then selection into employment by mental health status could generate a (non-causal) relationship between mental health and autonomy. Secondly, amongst those who experience periods out of work, unemployment may confound work-related autonomy if transitions from high autonomy to low autonomy are correlated with an intervening period of unemployment.[8]

I drop all observations which do not report values for any of the controls, so that any change in the point estimates and R-squared when controls are added entirely reflect the effect of additional controls and not changes in the sample. Since the autonomy variables only appear in even waves, I only use data from even waves of the survey. My main sample consists of observations from even waves of people who

[6] The big five personality indices are openness, conscientiousness, agreeableness, neuroticism, and extraversion.

[7] One concern may be that personality may be malleable and correlated with mental health. If so, personality variables may not be useful controls. However, my preferred specification includes person-specific fixed effects, which absorb the personality measures (imputed under the assumption they are time-invariant).

[8] Appendix B.3 shows the results when including all observations of workers in my sample.

continuously work which report a value for all control variables. I have 33,563 person-wave observations of 9,744 unique individuals.[9]

Table 3.3 presents summary statistics for variables of interest. The GHQ12 caseness varies between 0 and 12. Its mean is relatively low, with the average person having around 1.5 poor mental health symptoms. The SF12 ranges between 0 and 75.5 in my sample (though the theoretical maximum is 100). For most of the autonomy variables, over half of people score 0 (i.e., have high autonomy). The exception is autonomy over work hours, where over half of people score 1 (i.e., low autonomy over work hours). 20% of people score 1 on the overall autonomy index, that is, they have low autonomy on three or more dimensions. Finally, since my identification strategy relies heavily on within-person variation in autonomy, I calculate the within-person standard deviation for each variable. For all variables, there remains substantial variation after person fixed effects are taken out.

Table 2.2 shows how average autonomy differs across the 15 most common occupations in my data (where occupation is defined at the 3-digit ISCO code level). These 15 occupations account for 58% of employment in my sample. While levels of autonomy differ by occupation, there is clear evidence of within-occupation variation. For example, drivers have the lowest average overall autonomy. However, over 60% of them do not have low autonomy. Additionally, disaggregating autonomy over aspects of one's job reveals a more nuanced picture of work-related autonomy. While some occupations tend to have high overall autonomy, they can have low autonomy over certain aspects of work. For example, high-school teachers have relatively high overall autonomy, but have the lowest autonomy over work pace. This observation underscores the importance of assessing different aspects of work-related autonomy.

2.3 The relationship between mental health and autonomy

In this section, I estimate the relationship between work-related autonomy and mental health. Specifically, I run a series of regressions of the following type:

$$\text{mental_health}_{it} = \text{low_autonomy}'_{it}\beta + X'_{it}\gamma + \tau_t + \mu_i + \epsilon_{it} \tag{2.1}$$

where i indexes individual and t indexes survey wave. Here mental_health$_{it}$ is either the GHQ12 caseness or the SF12 index for person i at time t; autonomy$_{it}$ is either my summary index measure of having low work-related autonomy, or a vector of measures of work-related autonomy for person i at time t (i.e., a dummy for whether a person has low autonomy over job tasks, a dummy for whether she has low autonomy over work pace, etc.). X_{it} is a vector of controls for person i at time t. τ_t are wave fixed effects and included in every specification. μ_i are person fixed effects, which are ommitted in some specifications. ϵ_{it} is the OLS residual. β, the coefficients of interest, give the conditional association between mental health outcomes and work-related autonomy.

[9]Additionally, since lagged mental health is a control variable I exclude all observations which do not report the lagged dependent variable in each regression of that dependent variable. The sample therefore differs across dependent variables, but not across specifications for a given dependent variable.

Tables 2.3 and 2.4 show the results of these exercises, separately for men and women. In each specification, I control for wave fixed effects. The first and fourth columns show the association between a measure of low work-related autonomy and a measure of mental health, conditional on wave fixed effects. In the second and fifth columns I additionally control for person fixed effects, leveraging within-person variation in autonomy and mental health. In the third and sixth columns I add controls for all individual-level time-varying confounders. I control for age, marital status, education, and occupation using dummy variables for each possible value the variable can take. I include a dummy variable for whether a person has biological children living in the household. Lagged mental health is the person's score on the dependent variable in the previous wave. Standard errors are clustered at the person level.

Before discussing the results in Tables 3 and 4, it is informative how point estimates change when controls are added. For example, if point estimates are notably changed by the addition of a control, it could indicate that this control is an important confounding variable. Therefore, to complement Tables 2.3 and 2.4, Figures 2.1 to 2.8 show in detail how point estimates on the low work-related autonomy variables and the R-squared change as additional variables are included as controls. Some specifications exclude person fixed effects but control for variables which vary at the individual level, such as personality. The results from these specifications indicate which individual characteristics are important in explaining mental health. I control for personality variables using dummy variables for each possible value the variable can take. I control for labor income using the log of real labor income. Once education, marital status and age are controlled for, including labor income does not lead to a large increase in the R-squared in any specification. Person fixed effects absorb large percentages of the variation in the dependent variable, absorbing notably larger percentages of variation than variables such as personality, occupation and lagged mental health.

Lower overall autonomy is associated with worse mental health in all specifications. For men, moving from a high autonomy job to a low autonomy job is associated with an increase in the GHQ12 caseness score of around 0.44, and a decrease in the SF12 index of around 1.8. For women, the corresponding results are around a 0.51 increase in the caseness score and a 1.3 reduction in the SF12 index. Including controls generates a large increase in the R-squared, but comparatively small changes in the coefficients. My preferred specification is in column 3 of Tables 2.3 and 2.4 and includes person fixed effects, wave fixed effects and controls for education, marital status, occupation, having children and lagged mental health.[10] In this specification, I leverage within-person, within-occupation variation in work-related autonomy. The estimated association remains statistically significant, consistent with a causal effect. The corresponding estimated effects are an increase in the caseness score of 0.41 and 0.37 for men and women, respectively, and a reduction in the SF12 index of 1.3 for men and 0.88 for women. These results reflect a deterioration in mental health of between 10% and 15% of a standard deviation.

[10] I do not include labor income in my preferred specification, because one may argue that it is a post-determined variable, i.e., labor income may be affected both by one's mental health status and the level of work-related autonomy. Controlling for post-determined variables, or "colliders", can bias estimates (Elwert & Winship, 2014). Including labor income as a control in my preferred specification, I estimate coefficients on the low autonomy variable of 0.411 (0.00) for men using the GHQ12 caseness, 0.371 (0.00) for women using the GHQ12 caseness, −1.23 (0.00) for men using the SF12 index, and −0.884 (0.00) for women using the SF12 index (p-values are in parentheses). Including labor income would therefore not affect my conclusions significantly.

Turning to specifications which examine different aspects of autonomy, low autonomy over job tasks associates significantly with worse mental health in most specifications, with the exception of the SF12 index for men. This result suggests an effect on symptoms of mental illness present in the GHQ12 caseness measure but not measured by the SF12 index. This result is robust to various control sets. Low autonomy over work pace is consistently associated with worse mental health for men, but is significantly attenuated by controls and is only significant at the 10% level in my preferred specification for women. In my preferred specification, men benefit from autonomy over task order, and women benefit from autonomy over work hours. However, the adverse effect of low autonomy over work hours for women's mental health is only statistically significant when both person fixed effects and time-varying controls are included. This result could indicate some degree of selection, where, for example, women with better overall mental health tend to choose jobs with lower autonomy over hours. For low autonomy over work manner, point estimates always imply an adverse effect, but the effect is not generally statistically significant in my preferred specification.

Appendix B.3 reports results of the same analysis but on a larger sample, including all observations of current workers instead of those who continually work. I derive similar point estimates for regressions which include the full set of controls, but controls play a larger role in reducing the association. This suggests that entry and exit from different types of work by workers with different amounts of mental health symptoms generates greater associations between mental health and work-related autonomy, but a greater portion of this variation can be explained by controls.

My results confirm that there is a robust relationship between work-related autonomy and mental health. Specifically, exploiting within-person, within-occupation variation in work-related autonomy with controls for lagged mental health, marital status, education, having biological children, and age, lower work-related autonomy is associated with worse mental health. It is notable that in many cases the addition of controls does not significantly attenuate the relationship, despite a large increase in the R-squared.

There are multiple possible selection effects, many of which I am able to control for. Work-related autonomy is a non-pecuniary amenity which may be offered only to higher productivity workers, or workers may otherwise select into high autonomy jobs based on personality traits or preferences. Both of these mechanisms are addressed by the use of person fixed effects. Secondly, reductions in work-related autonomy may be correlated with reductions in economic security or downward mobility, i.e., periods of unemployment or lower status jobs. I partially address this concern by estimating effects in a sample of people who continuously work, and controlling for occupation fixed effects which may capture a subset of internal promotions and demotions. Additionally, in section 2.5, I restrict my sample to only those who never change employer, and find similar results. Thirdly, changes in family composition may be closely correlated with job choice and mental health. I control for this explicitly with variables for having children and marital status. A final source of bias I can address is occupation-level changes which might be correlated with autonomy. A particular industry may experience regulation, technology and practices which drive both autonomy and other working conditions. My use of occupation fixed effects allows me to account for this concern.

Any remaining confounders therefore have the following properties: they differ within individuals across time; they are not closely correlated with changes in family composition; they do not involve changes in employment status or employer; and they are idiosyncratic to broader occupational characteristics. Candidate examples might include the profitability of a company, which might cause employers to change working conditions and also may be correlated with worsening mental health of employees as they fear losing their jobs. There may be other time-varying characteristics of individuals such as substance use, or time-varying preferences, which are correlated with both mental health and the type of job a person does.

We may also be concerned about the possibility of reverse-causality, i.e., that mental health might have a direct effect on autonomy. The next two sections consider additional confounders and reverse causality in turn.

2.4 Assessing omitted variable bias

The strong and robust associations between mental health outcomes and autonomy are consistent with low work-related autonomy causing worse mental health. However, it is possible that there are unobserved confounders which explain the association. In order to place bounds on the causal effects, I use Oster's (2019) method of deriving the causal effect of a treatment under assumptions about the behavior of confounders. Oster shows that under assumptions about the explanatory power of remaining (unobserved) confounders, and the correlation of those confounders with the treatment variable of interest, one can uncover the omitted variable bias in the OLS estimator with only observed confounders.

Specifically, let A be the treatment of interest (in this case, a measure of work-related autonomy), W_1 be a projection of mental health onto the observable controls, and W_2 be a projection of mental health onto unobserved confounders. Then given:

1. An assumption about the R-squared which would obtain in a regression with controls for all confounders (the R^2_{MAX}), and

2. A parameter, δ, which describes how similar the relationship between the treatment and unobserved confounders is to the relationship between the treatment and existing controls:

$$\delta = \frac{COV(A, W_2)/VAR(W_2)}{COV(A, W_1)/VAR(W_1)} \tag{2.2}$$

Define:

$$\hat{\beta} - \nu(\delta, R^2_{\text{MAX}}) \rightarrow \beta \tag{2.3}$$

where $\hat{\beta}$ is the OLS estimator with observable controls, β is the true treatment effect, and ν is the bias due to ommitted variables.

ν is a function of δ and R^2_{MAX}. Conditional on δ and R^2_{MAX}, ν is greater the more the inclusion of observable controls attenuates the coefficient on the treatment, and smaller the more those same observable controls increase the R-squared. Intuitively, if observable controls move the OLS coefficient by a small

amount despite a large increase in the R-squared, additional controls which behave similarly to observable controls would lead to a small additional change in the OLS coefficient for a given R-squared increase.[11] I use the $\hat{\beta}$ estimated using the richest OLS regression, including controls for lagged mental health, person fixed effects, occupation dummies, education and marital status dummies, and wave fixed effects.

This method provides a mapping from our assumptions about how unobservable confounders behave to the causal effect of interest. It is a formalization of the idea that if there are few additional confounders, and if they are confounding to the same degree as, or less than, the existing controls, then the causal effect is close to the OLS estimator.

With regard to assumptions about δ, $-1 \leq \delta \leq 1$ corresponds to the assumption that unobservable confounders are not more important in explaining the treatment than variables I have already controlled for. This rationale is frequently given for setting δ to be smaller than 1 in absolute terms (e.g., Altonji et al., 2005). In my setting, this is a reasonable assumption given that I have already controlled for all of the most obvious candidate confounders, including person fixed effects, occupation and lagged mental health; it is unlikely that subsequent controls would be as confounding.[12] Setting a positive δ corresponds to the assumption that unobserved confounders push the coefficient in the same direction as the observed confounders, while setting δ to be negative corresponds to the assumption that unobserved confounders push the coefficient in the opposite direction. The majority of research which uses this method (e.g., Aizer et al., 2022, Egan et al., 2022) assumes $\delta > 0$, i.e., selection on unobservables is in the same direction as selection on observables. However, since in section 2.3, I find evidence of positive and negative selection on observables, I allow for both positive and negative values of δ. In this respect, my assumptions are weaker. I explore how the causal effect of low autonomy changes over a range of assumptions. In particular, I choose values of -1.3, -1, $+1$ and $+1.3$ for δ.

The R^2_{MAX} corresponds to the maximum percentage of the variation in the dependent variable which can be explained by a combination of the treatment and all of its confounders. By definition the most conservative assumption is $R^2_{\text{MAX}} = 1$. This assumption corresponds to the claim that mental health is fully explained by the treatment and confounders. It is therefore a very extreme assumption. Instead, Oster (2019) recommends setting R^2_{MAX} equal to the R-squared of the regression with all observable controls scaled by a factor greater than 1. Specifically, she suggests a ratio of 1.3.[13] I vary the R-squared

[11] ν is the root of a cubic function of R^2_{MAX}, δ, $\hat{\beta}$, the R-squared of a "short regression" of the dependent variable on the treatment, the coefficient on the treatment in this short regression, the variance of the dependent variable, the share of the variance of the treatment which is explained by existing controls, and the variance of the treatment. If there is more than one real root of the equation, Oster proposes selecting solutions based on the sign of $-\nu$, i.e., if one assumes that additional confounders have a correlation with the treatment which has the same sign as the correlation between the treatment and the observable controls, $-\nu$ has the same sign as the difference between $\hat{\beta}$ and the OLS estimator of the short regression. If there are multiple real roots which meet this criterion, she suggests selecting the smallest root in absolute terms.

[12] As verification of the reasonable range of values for δ, I calculate the value of δ implied by the introduction of lagged mental health as a control in my main regressions. I derive estimates between 0.24 and 0.37 in absolute value. Thus, 1 and -1 appear to be extreme values of δ in my setting.

[13] A ratio of 1.3 is relatively extreme. Oster (2019) shows that using the data from randomized controlled trials published in top 5 journals, applying this procedure with a maximum R-squared equal to 1.3 times the R-squared of a regression of the outcome variable on all observables and a δ coefficient of 1 would be sufficient to conclude that the causal effect is confounded

ratios between 1.1 and 1.3, in order to explore how important this assumption is for conclusions about the causal effect.

I present estimated causal effects of low autonomy on the GHQ12 caseness and the SF12 index under different assumptions about selection on unobservables (δ) and the maximum R-squared for men and women in Tables 2.5 to 2.8. My assumptions about the maximum R-squared are multiples of the R-squared of my preferred specification (including wave fixed effects, education and marital status fixed effects, occupation fixed effects, person fixed effects, having biological children in the household, and lagged mental health). Therefore, because the R-squared of the models which use the SF12 as a dependent variable are larger, the max R-squared assumption for the causal effect of autonomy on the SF12 measure is larger. Standard errors are calculated through block bootstrapping.[14]

This exercise has two purposes. Firstly, given assumptions about the reasonable range of the behavior of confounders, we can form a belief about which causal effects exist. Secondly, given some range of assumptions about the confoundedness of the conditional association, we can now form beliefs about the magnitude of the causal effect.

I have defended the position that $-1 \leq \delta \leq 1$ is a reasonable assumption. Under this assumption, there is a causal effect of low overall work-related autonomy on the GHQ12 caseness and the SF12 which is statistically different from zero for both men and women. In all cases, low overall autonomy leads to worse mental health. Having a job which scores low on autonomy on three or more dimensions causes between 0.37 and 0.43 additional mental health symptoms (the GHQ12 index) and a decline of between 0.61 and 1.52 on one's SF12 mental health score for men; for women, it causes between 0.23 and 0.40 additional mental health symptoms and a decline in their SF12 score of between 0.44 and 1.1. If one takes an extreme view that the unobserved confounders could be as much as 30% more correlated with the treatment than the observed confounders, the only result which survives is that low overall autonomy has an adverse effect on men's mental health as measured by the GHQ12 caseness (between 0.34 and 0.43).

For individual components of the autonomy index, I again assess the existence of causal effects under the assumption that $-1 \leq \delta \leq 1$. Overall, results are significantly less robust for individual components of autonomy. For men, low autonomy over job tasks, work manner, and task order causes a worsening of mental health under all assumptions, but these effects are not generally statistically significant under the assumption of an R-squared of 30% higher than in my preferred regression and a δ coefficient of 1. The effect of low autonomy over work hours on men's GHQ12 caseness score has an inconsistent sign and is not statistically significant under the assumption of a negative δ. Results for women are similarly inconsistent. Low autonomy over job tasks adversely affects women's mental health under most assumptions, but is not always statistically significant. There is an adverse causal effect of low autonomy over work hours on women's mental health as measured by the GHQ12 caseness and the SF12 index. The effect is an increase

in 10% of cases. Several studies use 1.3 as the ratio of the maximum R-squared to observed R-squared (e.g., Bursztyn et al., 2020, Campos-Mercade et al., 2021, Aizer et al., 2022, Egan et al., 2022).

[14] I use a block bootstrap which samples unique individuals. This necessarily means that the sample size varies with a range of around 2.4% of the initial sample size across draws. I use 1000 draws. This difference in sample size occurs because of sample attrition, and the introduction of new recruited respondents to maintain the representativeness of the sample. The sample is therefore not balanced in calendar time.

of between 0.11 and 0.41 symptoms of mental illness and a decline of between 0.28 and 1.5 on the SF12 index. This result is robust to loosening the bounds on δ to 1.3 and -1.3, but the range of causal effects increases to between a 0.10 and 0.65 increase in symptoms of mental illness and a decline of between 0.23 and 2.5 on the SF12 index.

To give a sense of the relative magnitude of my results, unemployment has been estimated to increase the caseness by 0.87 (Gathergood, 2013). Rose (2020) finds that the short-run effect of retirement on mental health is a decrease of around 0.06 in the caseness score. This suggests that the effect of low overall autonomy could be as large as half the effect of unemployment on mental health, and somewhat larger than the effect of retirement. Wallace, Nazroo and Bécares (2016) estimate an effect of racial discrimination of mental health of a 1.96 reduction in the SF12 index, and an effect of multiple exposures to discrimination of a 8.26 decline. My estimated effects are much smaller. Turning to other studies investigating the effects of working conditions, Belloni, Carrino, and Meschi (2022) find that the effect of a one-standard deviation increase in the "skills and discretion" index on the caseness is a 0.078 decrease for women. My estimates imply a slightly larger effect of 0.09 for women.[15] They find an insignificant effect of skills and discretion on men's mental health. The skills and discretion index is the closest variable to work-related autonomy in their paper, but we should not expect their estimates to map precisely to mine: skills and discretion include features of work other than autonomy, such as the opportunity to carry out complex tasks. Additionally, Belloni, Carrino and Meschi (2022) exploit occupation-level changes in skills and discretion amongst those who never change occupation, so it is likely that a local treatment effect in a differently selected population will lead to a different estimate.

Overall my results suggest an adverse causal effect of low work-related autonomy on mental health which is robust to reasonable assumptions about the degree of confoundedness of unobservables. Autonomy over job tasks and, for women, work hours appear to be the most important dimensions of autonomy for mental health.

2.5 Reverse causality

The previous section indicates that under reasonable assumptions about the extent of confoundedness of unobservables, a causal effect exists which links levels of autonomy at work to symptoms of mental illness. A remaining threat to my claim that my estimates reflect a causal effect of work-related autonomy on mental health is reverse causality, i.e., it could be that an increase in mental health symptoms causes low levels of autonomy at work. There are two plausible pathways for reverse causality: those in poor mental health might sort into jobs which offer low levels of workplace autonomy. This sorting would happen if the value of workplace autonomy was lower or even negative for people in poor mental health. Secondly, managers may respond to symptoms of mental health which are noticed or reported by reducing workers' autonomy. Reducing work-related autonomy might occur because employers believe that people in poor mental health are more productive when they have less autonomy. If either of these two dynamics are at

[15] I calculate this effect by multiplying my lowest estimated treatment effect by the standard deviation of the autonomy index 0.40.

play, they could drive within-person variation in work-related autonomy which is correlated with mental health.

I address the first concern by narrowing my sample to only those who never change their employer during the sample period, and re-estimating the association between mental health and work-related autonomy with a full set of controls (my preferred specification). In this specification, variation in work-related autonomy comes from changes to autonomy made by the employer, rather than workers switching employer in order to optimize their autonomy. (Workers' long-term proclivity to choose jobs with a particular level of autonomy is captured by person fixed effects.) Since in this specification, workers' scope for choosing their level of work-related autonomy is reduced, we would expect the estimated association to be attenuated if workers in poor mental health choose low autonomy jobs. Table 2.9 shows the results. Point estimates are of the same sign and broad magnitude as my preferred specification in the main sample. It is therefore unlikely that reverse causality where workers in poor mental health move to jobs with low levels of work-related autonomy is a significant driver of my main results.

Turning to the second concern, there are two arguments against this relationship being primarily driven by an effect of mental health on autonomy. Firstly, this narrative, in which an employer changes a person's working conditions as their mental health deteriorates, requires mental health to deteriorate to noticeable or notable levels. It is unlikely that workplace conditions would change in response to only one or two mental health symptoms (such as sleeping poorly or not feeling reasonably happy). Reverse causality via this channel is most likely to happen in response to changes in mental health when a person has a large number of negative mental health symptoms.

In Figure 2.9, I assess whether there is a non-linear relationship between the caseness and the probability of working in a low-autonomy job. Specifically, I estimate linear probability models by regressing a dummy for whether a person has a low autonomy job on dummy variables for having each value of the GHQ12 caseness score, controlling for person fixed effects, wave dummies, occupation dummies, education, marital status, and whether the person has biological children in the household. In these regressions, I use the same sample as in my main results (respondents continuously work). I plot the increased probability of having low work-related autonomy at each level of the caseness relative to the probability of having low work-related autonomy for someone who has a caseness score of 0. If the cause of the relationship between low autonomy and poor mental health was that employers imposed lower autonomy on employees who reported or were observed to have mental illnesses, we might expect the probability of having a low autonomy job to increase rapidly as respondents have clinically significant levels of caseness, such as 8 (e.g., Anjara et al. 2020). In fact, the relationship is relatively linear, with increases in the probability of being in a low autonomy job beginning before clinically relevant levels of caseness are observed. To formally test for non-linearity, I regress the probability of having a low autonomy job on the caseness and the caseness squared with the same controls, and plot the resulting estimated relationship. The relationship is well-approximated as linear, and the coefficient on the caseness squared is insignificant and close to zero.

A second reason why this story seems unlikely is that adjustments employers might be required to make in response to an employee experiencing mental illness are as likely to increase as decrease work-

place autonomy. UK disability law (e.g., the 2010 Equality Act) requires employers to make "reasonable adjustments" for disabled employees, which includes employees with mental illnesses (Lockwood et al., 2012). Most legal guidance on what these adjustments might entail suggests adjustments such as more flexible work structures, changes in working environments which might include working from home, and a greater emphasis on specific goal-setting and feedback (Bell, 2015). Except for the latter, these adjustments mostly *increase* work-related autonomy. It is unlikely that overall, employer adjustments for those with mental illness would systematically drive work-related autonomy downward.

Appendix B.4 studies the path of mental health around transitions between high and low autonomy work. If changes in mental health proceed changes in work-related autonomy then this may be evidence that changes in mental health has an effect on work-related autonomy (Boyce and Oswald, 2012). I find that overall most of the change in mental health when a person's work-related autonomy changes occurs in the same period as the change in working conditions. We should be cautious in drawing conclusions about reverse causality from this result, because the precise ordering of changes in autonomy and mental health are obscured by the two-year period between each observation.

A final consideration about reverse causality is that if there is an effect of mental health on work-related autonomy, then controlling for lagged mental health should attenuate estimated regression coefficients unless the effect of mental health on working conditions always occurs within a year. However, I do not find evidence of a large change in regression coefficients when I control for lagged mental health (see Figures 2.1 to 2.8).

2.6 Discussion and conclusion

My research studies the relationship between work-related autonomy and mental health outcomes in the UK. I use the Understanding Society data set, which has two different summary measures of mental health and measures of autonomy which disaggregate work-related autonomy into five components.

First, I estimate the effect of autonomy on mental health using OLS, controlling for an increasingly large set of variables. I find that there is a robust adverse relationship between mental health and low autonomy. Lower levels of overall work-related autonomy are associated with worse mental health, as measured by both the GHQ12 caseness and the SF12 mental health index, consistent with a causal relationship. Amongst individual components of the autonomy index, the most robust relationships are between low autonomy over job tasks and mental health.

Secondly, I use Oster's (2019) method to investigate the causal effect of autonomy at work on mental health under different assumptions about 1) the extent of selection on unobservables relative to selection on observables and 2) the percentage of variation in the dependent variable (mental health) explained by the conjunction of the treatment, observable confounders and unobservable confounders. I argue that overall, having low overall autonomy at work likely causes an increase of between 0.37 and 0.42 negative mental health symptoms, or between a 0.61 and 1.52 reduction in the SF12 measure of mental health for men. For women, having low autonomy across 3 or more dimensions causes an increase of between 0.23 and 0.40 symptoms of mental illness and a decline in the SF12 index of between 0.44 and 1.1. Turning to

specific aspects of autonomy, my most robust result is that autonomy over work hours is important for women's mental health.

Finally, I discuss the potential for reverse causality. One way that reverse causality could occur is if employees in poor health are more likely to choose jobs where they have low work-related autonomy. I test for this kind of sorting by restricting my sample to only those workers who never change employers. I find results of a comparable magnitude to my main results, making it unlikely that my main results are driven by workers choosing low-autonomy jobs with new employers when they are in poor mental health. I argue that it is unlikely that employers systematically restrict the autonomy of employees in poor mental health for two reasons. Firstly the association between lower autonomy at work and more symptoms of mental illness is present even amongst those with few negative mental health symptoms, where the employee is less likely to report poor mental health and the employer less likely to make adjustments. Secondly, the adjustments employers are required to make under UK law for those with poor mental health are as likely to involve granting employees more work-related autonomy as they are to require restrictions.

My research has numerous implications for the relationship between mental health and labor markets. Firstly, where workers value non-pecuniary characteristics of their work, inequality can be mis-stated by simply observing income or wealth (Maestas et al., 2023). In my data set, both men and women who work in low-autonomy jobs have labor income that is around 27% lower than their counterparts in high-autonomy jobs. Accounting for inequality in the mental health effects of employment therefore increases estimates of overall inequality. Furthermore, inequality in workplace flexibility could contribute to mental health inequality.

Secondly, given the presence of a primarily state-funded healthcare system in the UK, low levels of workplace autonomy constitute a negative externality: neither workers nor firms have an incentive to account for the monetary cost of mental illness arising from working conditions. The cost of medical services for people in the UK with depression and anxiety is forecast to rise to GBP 5 billion by 2026 (McCrone et al., 2008), over 2% of total NHS spending. To the extent that working conditions are a contributing factor to poor mental health, the potential public cost is large.

Thirdly, there are policy implications of particular working conditions having adverse effects on mental health. Economic evaluations typically find that workplace interventions to improve mental health can deliver net benefits (including to employers) (Knapp et al., 2011). My research suggests two things: firstly, employers who offer their workers lower levels of workplace autonomy may benefit especially from interventions such as offering cognitive behavioral therapy to staff; and secondly, where possible, employers may find that offering greater levels of autonomy, especially over aspects of work such as tasks and flexible hours, may improve their workers' mental health.

There are two major limitations of this study. The first is that exact estimates of the causal effect of work-related autonomy on mental health depend on assumptions about the extent of selection on unobservables, though, under most reasonable assumptions, the adverse effect of low autonomy on mental health survives and the range of effects is fairly tight. Secondly, my results do not allow me to adjudicate the mechanism by which work-related autonomy affects mental health. As discussed in the introduction, better work-related autonomy may improve mental health by 1) expanding a person's locus of control 2)

or symbolizing status or "work dignity". Disentangling the extent to which each mechanism contributes to the effect is an interesting avenue for future work.

Table 2.1: Summary statistics of key variables

	Mean	Standard deviation	Min	Max	Within-person standard deviation
GHQ12 caseness	1.459	2.686	0	12	1.853
SF12	49.882	8.778	0	75.5	6.583
Low overall autonomy	0.199	0.399	0	1	0.312
Low autonomy over job tasks	0.257	0.437	0	1	0.337
Low autonomy over work pace	0.244	0.430	0	1	0.321
Low autonomy over work manner	0.156	0.363	0	1	0.278
Low autonomy over task order	0.164	0.370	0	1	0.287
Low autonomy over work hours	0.527	0.499	0	1	0.407

Notes: The sample is constructed from observations in the even waves of the Understanding Society survey of people who report being in work in every wave in which they are observed. Low autonomy over job tasks, work pace, work manner, task order and work hours is scored 1 if a person reports having either a little autonomy or no autonomy on this dimension, and 0 otherwise. I define overall low autonomy as being equal to 1 if a person scores 1 (low) on 3 or more aspects of workplace autonomy. The GHQ12 caseness is the number of negative mental health symptoms a person has. The SF12 index scores the mental health of a person between 0 (worst) and 100 (best) based on their answers to a health survey. Within-person standard deviation is the standard deviation of the variable once a person-fixed effect is taken out.

Table 2.2: Autonomy scores across different occupations

Occupation	Low overall autonomy	Low autonomy over job tasks	Low autonomy over work pace	Low autonomy over work manner	Low autonomy over task order	Low autonomy over work hours
Other Specialist Managers	0.041	0.084	0.110	0.035	0.033	0.232
Production and Operation Managers	0.044	0.073	0.119	0.035	0.032	0.301
Architects, Engineers, and Related Professions	0.057	0.147	0.147	0.062	0.038	0.245
Finance and Sales Associate Professionals	0.066	0.149	0.165	0.061	0.052	0.259
Numerical Clerks	0.123	0.263	0.172	0.106	0.085	0.420
Other Clerks	0.128	0.217	0.164	0.113	0.085	0.353
Housekeeping and Restaurant Service Workers	0.186	0.310	0.229	0.177	0.141	0.391
Nursing and Midwifery Professionals	0.232	0.247	0.330	0.176	0.153	0.641
High-school Teachers	0.269	0.273	0.401	0.151	0.220	0.778
Personal Care and Related Workers	0.310	0.369	0.330	0.266	0.252	0.704
Salespeople and Demonstrators	0.313	0.326	0.339	0.225	0.272	0.788
Domestic Helpers, Cleaners and Launders	0.314	0.356	0.328	0.268	0.276	0.747
Porters and Doorkeepers	0.385	0.435	0.386	0.329	0.309	0.738
Client Information Clerks	0.397	0.545	0.344	0.294	0.389	0.650
Motor Vehicle Drivers	0.399	0.486	0.383	0.338	0.346	0.732

Notes: This table shows the average value of workplace autonomy variables by the 15 most common (3-digit ISCO) occupations. Low autonomy over job tasks, work pace, work manner, task order, and work hours score 1 if a person reports having either a little autonomy or no autonomy on this dimension, and 0 otherwise. I define overall low autonomy as being equal to 1 if a person scores 1 on 3 or more aspects of workplace autonomy.

Table 2.3: The effect of low autonomy on the mental health of men

	Panel A: effect on GHQ12 caseness					
Low autonomy:						
overall	0.444^{***}	0.370^{***}	0.409^{***}			
	(0.070)	(0.062)	(0.069)			
over job tasks				0.207^{***}	0.201^{***}	0.211^{***}
				(0.068)	(0.059)	(0.068)
over work pace				0.398^{***}	0.262^{***}	0.254^{***}
				(0.071)	(0.063)	(0.066)
over work manner				0.128	0.100	0.136
				(0.095)	(0.085)	(0.087)
over task order				0.070	0.097	0.151^{*}
				(0.089)	(0.079)	(0.083)
over work hours				-0.119^{**}	-0.057	0.049
				(0.049)	(0.046)	(0.059)
Observations	13,847	13,847	13,847	13,847	13,847	13,847
R^2	0.005	0.185	0.538	0.011	0.188	0.540
	Panel B: effect on SF12					
Low autonomy:						
overall	-1.794^{***}	-1.270^{***}	-1.245^{***}			
	(0.247)	(0.188)	(0.208)			
over job tasks				-0.355	-0.232	-0.386^{*}
				(0.235)	(0.182)	(0.205)
over work pace				-1.555^{***}	-0.990^{***}	-0.905^{***}
				(0.247)	(0.194)	(0.200)
over work manner				-0.203	-0.266	-0.164
				(0.327)	(0.263)	(0.263)
over task order				-0.751^{**}	-0.543^{**}	-0.703^{***}
				(0.301)	(0.236)	(0.250)
over work hours				-0.014	-0.082	-0.204
				(0.188)	(0.148)	(0.177)
Observations	13,892	13,892	13,892	13,892	13,892	13,892
R^2	0.012	0.338	0.649	0.018	0.340	0.651
Wave	Y	Y	Y	Y	Y	Y
Person FE	N	Y	Y	N	Y	Y
Other controls	N	N	Y	N	N	Y

Notes: The table presents results from regressions of measures of mental health on measures of work-related autonomy, with various controls (as in Equation 2.1), amongst men. Standard errors are clustered at the individual level. Other controls include education, marital status occupation, whether the respondent has biological children living with them, and lagged mental health, which is the value of the dependent variable for the same person in the previous wave. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2.4: The effect of low autonomy on the mental health of women

	Panel A: effect on GHQ12 caseness					
Low autonomy:						
overall	0.506***	0.417***	0.370***			
	(0.073)	(0.062)	(0.068)			
over job tasks				0.243***	0.175***	0.291***
				(0.073)	(0.064)	(0.069)
over work pace				0.222***	0.174***	0.067
				(0.072)	(0.062)	(0.070)
over work manner				0.365***	0.277***	0.180**
				(0.097)	(0.083)	(0.087)
over task order				−0.175**	−0.052	0.026
				(0.088)	(0.077)	(0.086)
over work hours				0.059	0.064	0.181***
				(0.057)	(0.052)	(0.064)
Observations	15,708	15,708	15,708	15,708	15,708	15,708
R^2	0.006	0.201	0.561	0.008	0.202	0.563
	Panel B: effect on SF12					
Low autonomy:						
overall	−1.295***	−0.867***	−0.878***			
	(0.228)	(0.178)	(0.195)			
over job tasks				−0.934***	−0.518***	−0.658***
				(0.227)	(0.185)	(0.197)
over work pace				−0.864***	−0.704***	−0.365*
				(0.226)	(0.178)	(0.200)
over work manner				−0.672**	−0.345	−0.380
				(0.300)	(0.248)	(0.249)
over task order				0.704***	0.388*	0.180
				(0.289)	(0.231)	(0.244)
over work hours				−0.046	−0.174	−0.561***
				(0.187)	(0.152)	(0.181)
Observations	15,692	15,692	15,692	15,692	15,692	15,692
R^2	0.008	0.307	0.635	0.011	0.309	0.636
Wave FE	Y	Y	Y	Y	Y	Y
Person FE	N	Y	Y	N	Y	Y
Other controls	N	N	Y	N	N	Y

Notes: The table presents results from regressions of measures of mental health on measures of work-related autonomy, with various controls (as in Equation 2.1), amongst women. See notes to Table 2.3 for a description of the controls used. *p<0.1; **p<0.05; ***p<0.01

Table 2.5: Causal effects of low autonomy on the GHQ12 caseness of men under different assumptions

δ	R-squared ratio	Low overall autonomy	Low autonomy over job tasks	Low autonomy over work pace	Low autonomy over work manner	Low autonomy over task order	Low autonomy over work hours
-1.3	1.1	0.418***	0.210***	0.295***	0.134**	0.126***	−0.003
		(0.022)	(0.050)	(0.068)	(0.064)	(0.048)	(0.019)
-1.3	1.3	0.430***	0.208***	0.345***	0.131***	0.097	−0.062
		(0.044)	(0.049)	(0.053)	(0.062)	(0.060)	(0.034)
-1	1.1	0.416***	0.210***	0.286***	0.134*	0.130***	0.007
		(0.020)	(0.053)	(0.073)	(0.069)	(0.048)	(0.017)
-1	1.3	0.426***	0.209***	0.331***	0.132**	0.105*	−0.046
		(0.037)	(0.046)	(0.054)	(0.058)	(0.054)	(0.030)
1	1.1	0.400***	0.212**	0.208	0.139	0.183**	0.112***
		(0.027)	(0.102)	(0.132)	(0.149)	(0.090)	(0.028)
1	1.3	0.369***	0.219	0.033	0.155	0.348	0.413***
		(0.091)	(0.281)	(0.278)	(0.466)	(0.300)	(0.125)
1.3	1.1	0.396***	0.213*	0.191	0.140	0.196	0.137***
		(0.034)	(0.116)	(0.146)	(0.173)	(0.105)	(0.036)
1.3	1.3	0.344**	0.228	−0.146	0.186	0.652	0.907
		(0.150)	(0.611)	(0.444)	(1.612)	(0.878)	(2.241)

Notes: The table shows the causal effect of work-related autonomy on men's GHQ12 caseness score under different assumptions about the importance of unobserved confounders relative to observed confounders. Calculations use Oster's (2019) method of calculating the causal effect under assumptions about 1) the total share of variance in the dependent variable explained by the independent variable and its confounders (max R-squared) and 2) the ratio of selection on unobservable confounders to observable confounders (δ). I assess the degree of selection on observables by comparing the R-squared and coefficients in the specification with the most controls to the specification with the least controls in Table 2.3. The GHQ12 caseness is derived from the the GHQ12 survey, and is the total number of negative mental health symptoms a person has. The SF12 index is derived from the SF12 survey, and is a weighting of scores on this index. The index varies between 0 and 100, with higher numbers indicating better mental health. Standard errors are calculated by block bootstrapping. *p<0.1; **p<0.05; ***p<0.01

Table 2.6: Causal effects of low autonomy on men's SF12 under different assumptions

δ	R-squared ratio	Low overall autonomy	Low autonomy over job tasks	Low autonomy over work pace	Low autonomy over work manner	Low autonomy over task order	Low autonomy over work hours
-1.3	1.1	-1.387^{***}	-0.377^{***}	-1.089^{***}	-0.177^{***}	-0.718^{***}	-0.146^{***}
		(0.068)	(0.050)	(0.068)	(0.064)	(0.048)	(0.019)
-1.3	1.3	-1.576^{***}	-0.366^{***}	-1.316^{***}	-0.190^{***}	-0.735^{***}	-0.079^{**}
		(0.157)	(0.049)	(0.053)	(0.062)	(0.060)	(0.034)
-1	1.1	-1.358^{***}	-0.379^{***}	-1.052^{***}	-0.174^{***}	-0.715^{***}	-0.157^{***}
		(0.054)	(0.053)	(0.073)	(0.069)	(0.048)	(0.017)
-1	1.3	-1.521^{***}	-0.369^{***}	-1.252^{***}	-0.187^{***}	-0.731^{***}	-0.097^{***}
		(0.131)	(0.046)	(0.054)	(0.058)	(0.054)	(0.030)
1	1.1	-1.097^{***}	-0.397^{***}	-0.701^{***}	-0.149	-0.685^{***}	-0.275^{***}
		(0.072)	(0.102)	(0.132)	(0.149)	(0.090)	(0.028)
1	1.3	-0.611^{**}	-0.449	0.079	-0.066	-0.589^{**}	-0.617^{***}
		(0.299)	(0.281)	(0.278)	(0.466)	(0.300)	(0.125)
1.3	1.1	-1.044^{***}	-0.402^{***}	-0.624^{***}	-0.142	-0.677^{***}	-0.303^{***}
		(0.097)	(0.116)	(0.146)	(0.173)	(0.105)	(0.036)
1.3	1.3	-0.222	-0.520	0.871^{**}	0.091	-0.419	-1.144
		(0.518)	(0.611)	(0.444)	(1.612)	(0.878)	(2.241)

Notes: The table presents the causal effects of work-related autonomy on the SF12 index under different assumptions amongst women. See notes to Table 2.5 for a full description of the method and variables.
*p<0.1; **p<0.05; ***p<0.01

Table 2.7: Causal effects of low autonomy on women's GHQ12 caseness under different assumptions

δ	R-squared ratio	Low overall autonomy	Low autonomy over job tasks	Low autonomy over work pace	Low autonomy over work manner	Low autonomy over task order	Low autonomy over work hours
-1.3	1.1	0.404^{***} (0.021)	0.277^{***} (0.052)	0.112^{***} (0.063)	0.239^{***} (0.061)	-0.036 (0.051)	0.145^{***} (0.021)
-1.3	1.3	0.450^{***} (0.044)	0.260^{***} (0.053)	0.166^{***} (0.053)	0.303^{***} (0.068)	-0.107^{*} (0.062)	0.102^{***} (0.039)
-1	1.1	0.397^{***} (0.018)	0.280^{***} (0.055)	0.103 (0.067)	0.227^{***} (0.064)	-0.024 (0.052)	0.152^{***} (0.018)
-1	1.3	0.437^{***} (0.037)	0.265^{***} (0.050)	0.151^{***} (0.053)	0.286^{***} (0.063)	-0.088 (0.056)	0.114^{***} (0.034)
1	1.1	0.335^{***} (0.025)	0.306^{***} (0.101)	0.014 (0.123)	0.107 (0.124)	0.102 (0.100)	0.224^{***} (0.027)
1	1.3	0.225^{***} (0.085)	0.370 (0.246)	-0.206 (0.296)	-0.279 (0.375)	0.481 (0.303)	0.411^{***} (0.124)
1.3	1.1	0.323^{***} (0.032)	0.312^{***} (0.114)	-0.006 (0.138)	0.077 (0.143)	0.133 (0.116)	0.240^{***} (0.035)
1.3	1.3	0.142 (0.137)	0.440 (0.431)	-0.469 (0.523)	-1.065 (3.500)	1.125 (2.406)	0.650^{**} (0.274)

Notes: The table presents the causal effects of work-related autonomy on the GHQ12 caseness score under different assumptions amongst women. See notes to Table 2.5 for a full description of the method and variables. *p<0.1; **p<0.05; ***p<0.01

Table 2.8: Causal effects of low autonomy on women's SF12 under different assumptions

δ	R-squared ratio	Low overall autonomy	Low autonomy over job tasks	Low autonomy over work pace	Low autonomy over work manner	Low autonomy over task order	Low autonomy over work hours
-1.3	1.1	-0.983^{***}	-0.738^{***}	-0.511^{***}	-0.472^{***}	0.343^{***}	-0.407^{***}
		(0.058)	(0.052)	(0.063)	(0.061)	(0.051)	(0.021)
-1.3	1.3	-1.124^{***}	-0.834^{***}	-0.685^{***}	-0.574^{***}	0.526^{***}	-0.228^{***}
		(0.134)	(0.053)	(0.053)	(0.068)	(0.062)	(0.039)
-1	1.1	-0.961^{***}	-0.722^{***}	-0.483^{***}	-0.455^{***}	0.311^{***}	-0.437^{***}
		(0.046)	(0.055)	(0.067)	(0.064)	(0.052)	(0.018)
-1	1.3	-1.082^{***}	-0.807^{***}	-0.637^{***}	-0.547^{***}	0.476^{***}	-0.277^{***}
		(0.112)	(0.050)	(0.053)	(0.063)	(0.056)	(0.034)
1	1.1	-0.773^{***}	-0.568^{***}	-0.197	-0.267^{**}	-0.016	-0.739^{***}
		(0.060)	(0.101)	(0.123)	(0.124)	(0.100)	(0.027)
1	1.3	-0.442^{*}	-0.206	0.510^{*}	0.332	-0.975^{***}	-1.514^{***}
		(0.245)	(0.246)	(0.296)	(0.375)	(0.303)	(0.124)
1.3	1.1	-0.736^{***}	-0.534^{***}	-0.132	-0.220	-0.095	-0.809^{***}
		(0.081)	(0.114)	(0.138)	(0.143)	(0.116)	(0.035)
1.3	1.3	-0.194	0.192	1.348^{**}	1.456	-2.519	-2.500^{***}
		(0.402)	(0.431)	(0.523)	(3.500)	(2.406)	(0.274)

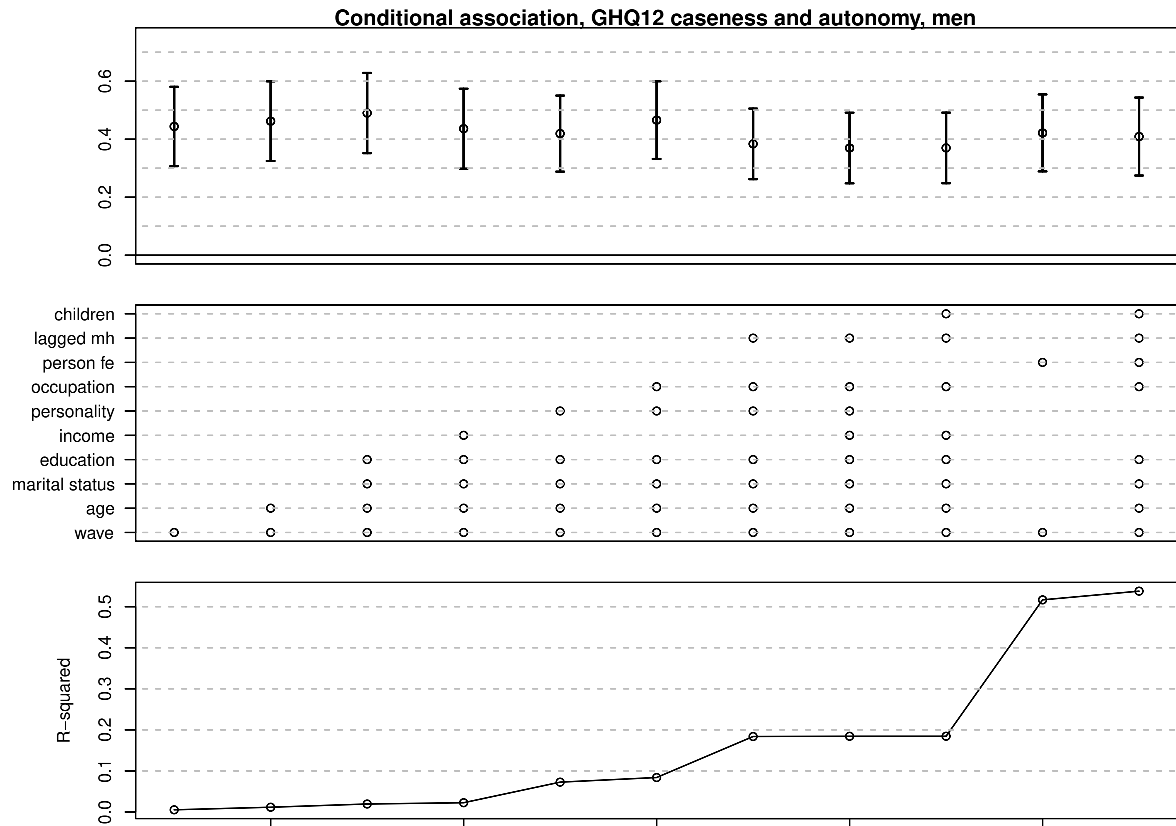
Notes: The table presents the causal effects of work-related autonomy on the SF12 index under different assumptions amongst women. See notes to Table 2.5 for a full description of the method and variables.
$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 2.9: The effect of low autonomy on the mental health of those who never change employers

	Men		Women	
Dependent variable:	GHQ12 caseness	SF12	GHQ12 caseness	SF12
	(1)	(2)	(3)	(4)
Low overall autonomy	0.404***	−1.114***	0.340***	−0.838***
	(0.075)	(0.231)	(0.075)	(0.213)
Observations	12,833	12,861	14,913	14,887
R^2	0.554	0.652	0.704	0.649
Wave	Y	Y	Y	Y
Person FE	Y	Y	Y	Y
Other controls	Y	Y	Y	Y

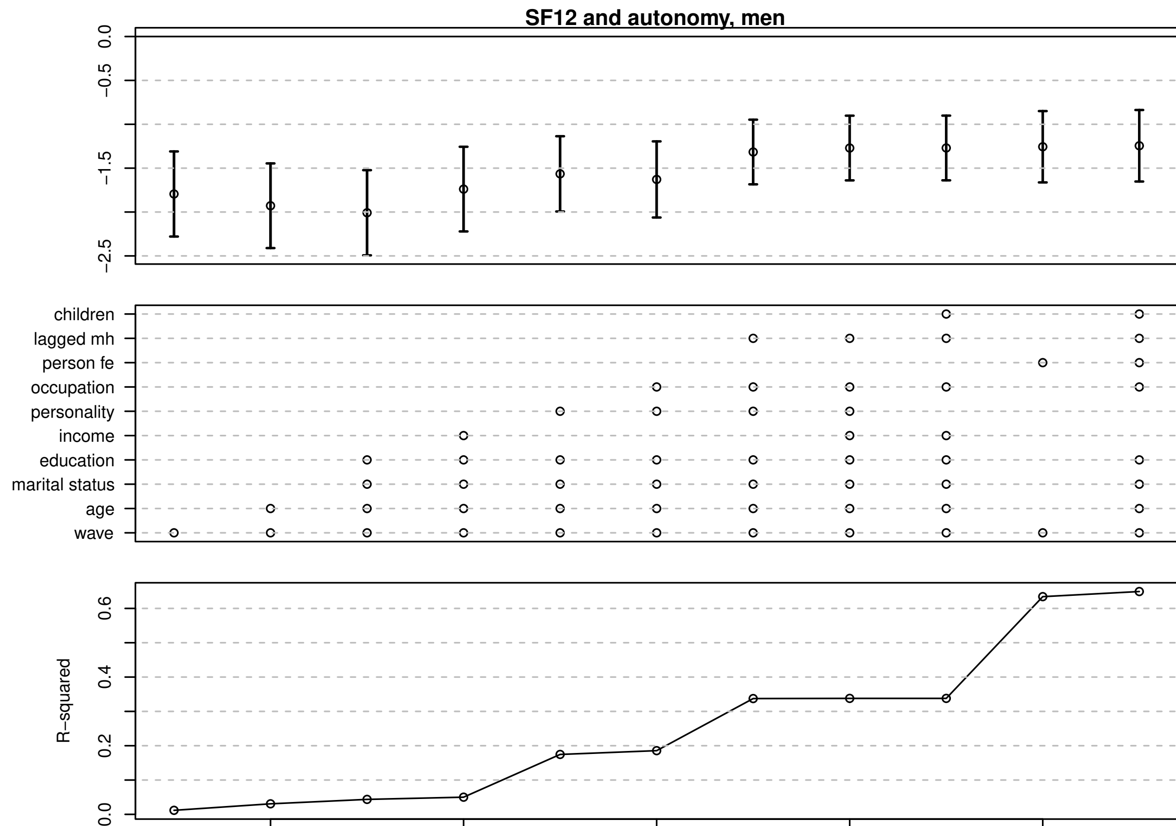
Notes: This table presents results of regressions of measures of mental health on work-related autonomy in a sample of those people who never change employer. As in equation 2.1, I control for person-fixed effects, wave fixed effects and controls (occupation, whether a person has biological children, education, marital status and lagged mental health). Standard errors are clustered at the individual level. *p<0.1; **p<0.05; ***p<0.01

Figure 2.1: Conditional associations between the GHQ12 caseness score and low work-related autonomy amongst men



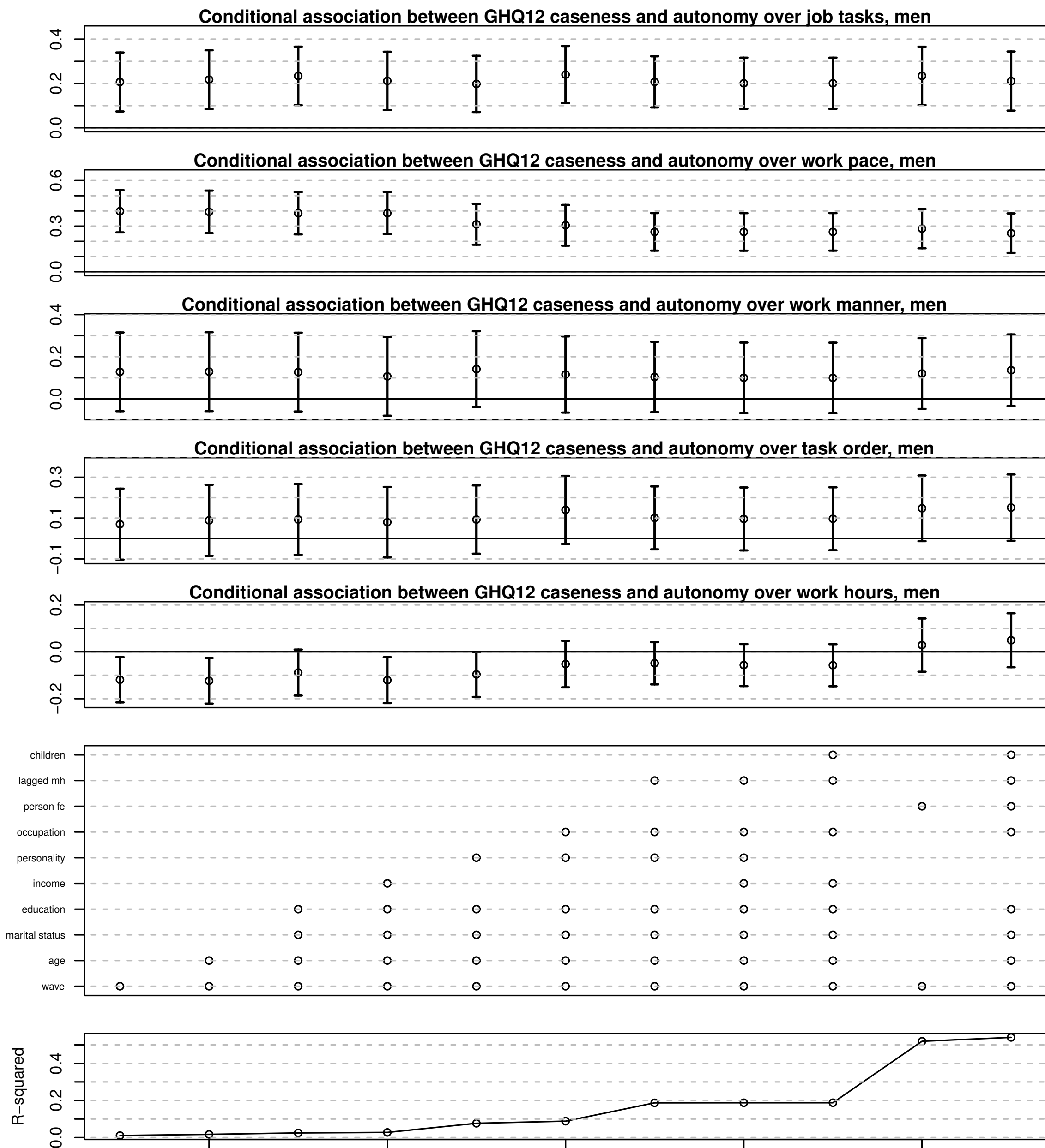
Notes: This figure shows how estimated associations between the GHQ12 caseness and having low work-related autonomy amongst men change as new controls are included. Conditional associations are estimated by regressing the GHQ12 caseness on the summary measure of autonomy, with various controls. The GHQ12 caseness is derived from the the GHQ12 survey, and is the total number of negative mental health symptoms a person has. The top panel reports coefficients on low autonomy, with 95% confidence intervals using standard errors clustered at the individual level. The middle panel shows which controls are included in each specification. The bottom panel shows the R-squared in each specification. Wave controls are a dummy for each wave of the survey. Age is age at last birthday and is controlled for with a dummy variable for each value it can take. Marital status and education are exhaustive, mutually exclusive values detailing the current marital status of a person and the highest qualification they achieve, respectively. I control for income using the log of real labor income. I control for personality using the "big five" personality traits, using a dummy for each possible combination they can take. Occupation is measured by 3-digit ISCO code. I include a dummy for each possible occupation. I control for person fixed effects with a full set of dummies. I control for children by including a dummy for whether a person has any biological children who live with them. Lagged mental health is the dependent variable the person reported in the previous wave of the survey.

Figure 2.2: Conditional associations between the SF12 index and low work-related autonomy amongst men



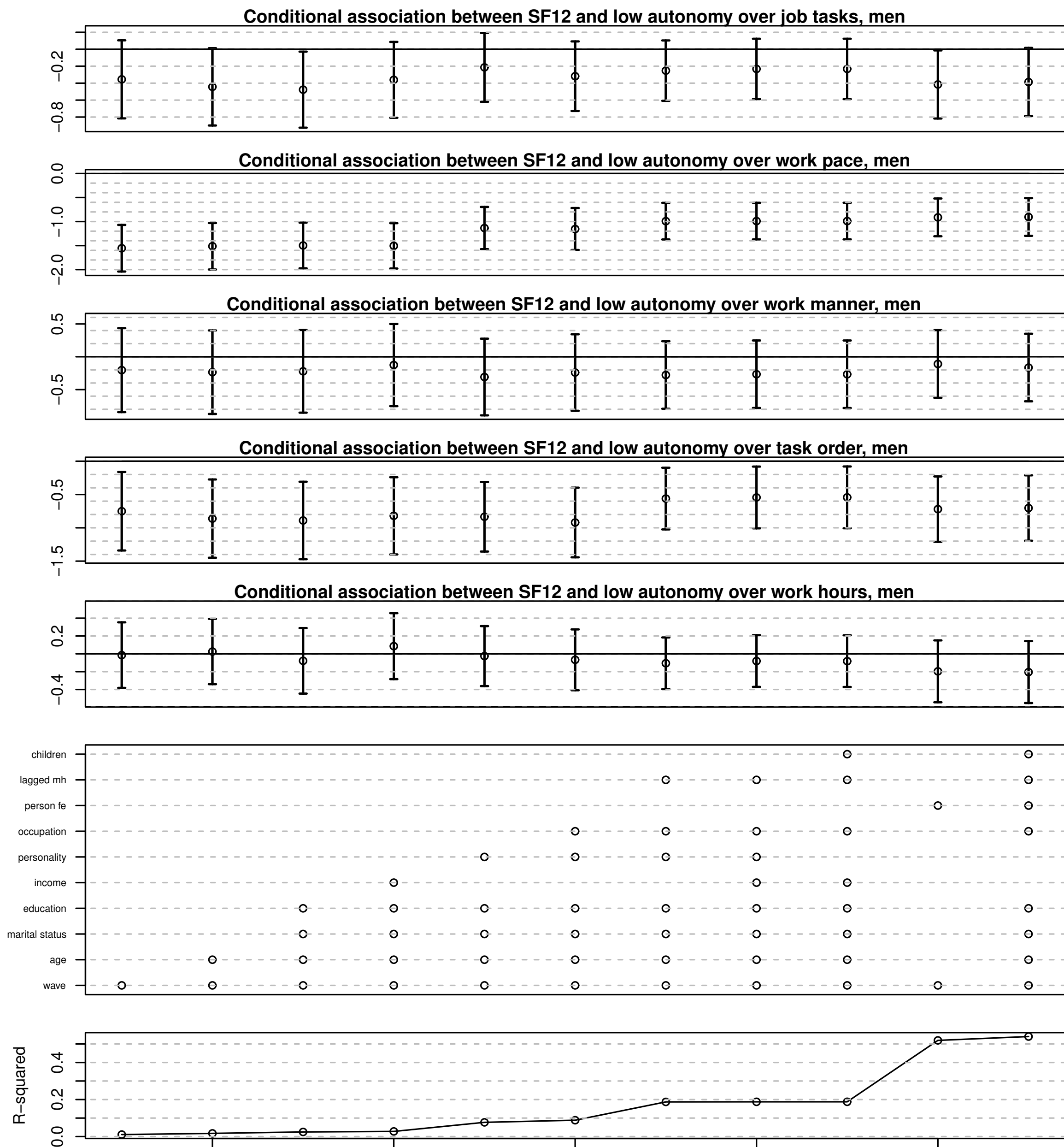
Notes: This figure shows how estimated associations between the SF12 index and levels of work-related autonomy amongst men change as new controls are included. Conditional associations are estimated by regressing the SF12 index on the summary measure of autonomy, with various controls. The SF12 index is derived from the SF12 survey, and is a weighting of scores on this index. The index varies between 0 and 100, with higher numbers indicating better mental health. The top panel reports coefficients on autonomy, with confidence intervals using standard errors clustered at the individual level. The middle panel shows which controls are included in each specification. The bottom panel shows the R-squared in each specification. Control variables are described in the note to Figure 2.1.

Figure 2.3: Conditional associations between the GHQ12 caseness score and low work-related autonomy across different dimensions amongst men



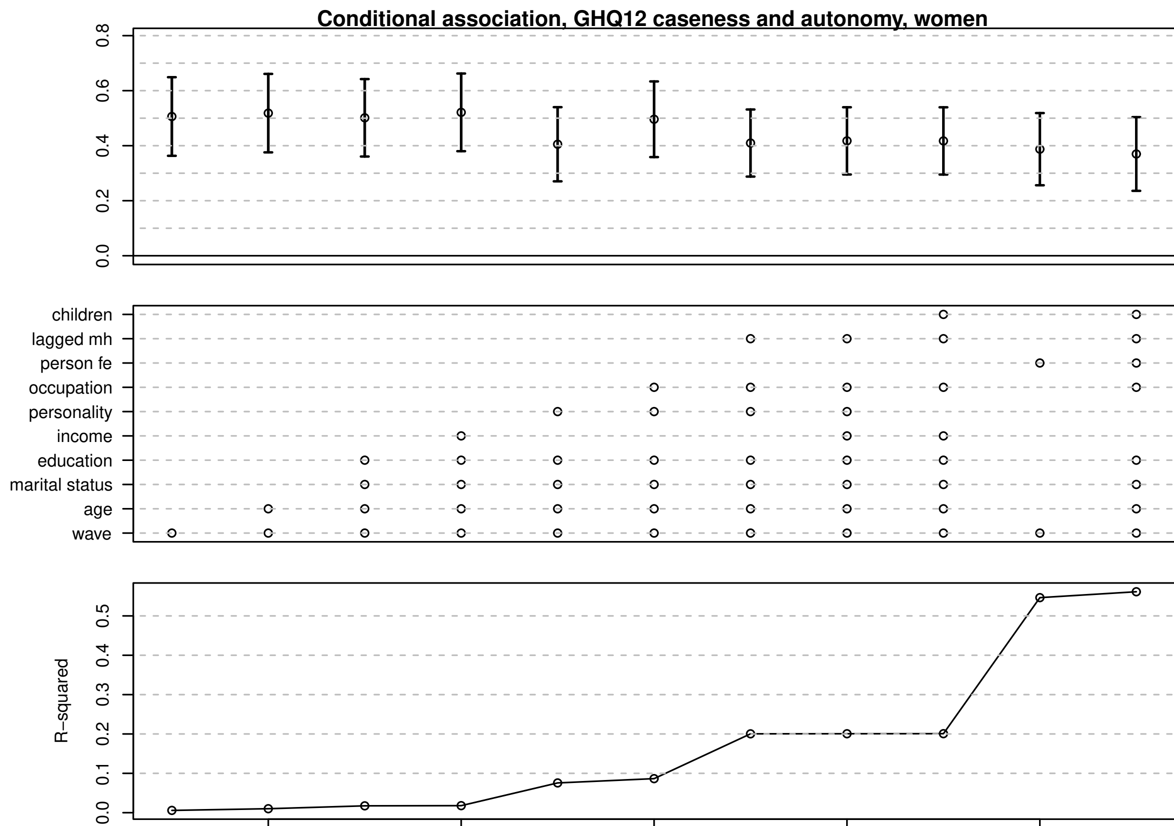
Notes: This figure shows how estimated associations between the GHQ12 caseness and levels of work-related autonomy, disaggregated across dimensions, amongst men change as new controls are included. Conditional associations are estimated by regressing the GHQ12 caseness on all measures of autonomy, with various controls. The GHQ12 caseness is described in the note to Figure 2.1. The top 5 panels report coefficients on autonomy, with confidence intervals using standard errors clustered at the individual level. The 6th panel shows which controls are included in each specification. The bottom panel shows the R-squared in each specification. Control variables are described in the note to Figure 2.1.

Figure 2.4: Conditional associations between the SF12 index and low work-related autonomy across different dimensions amongst men



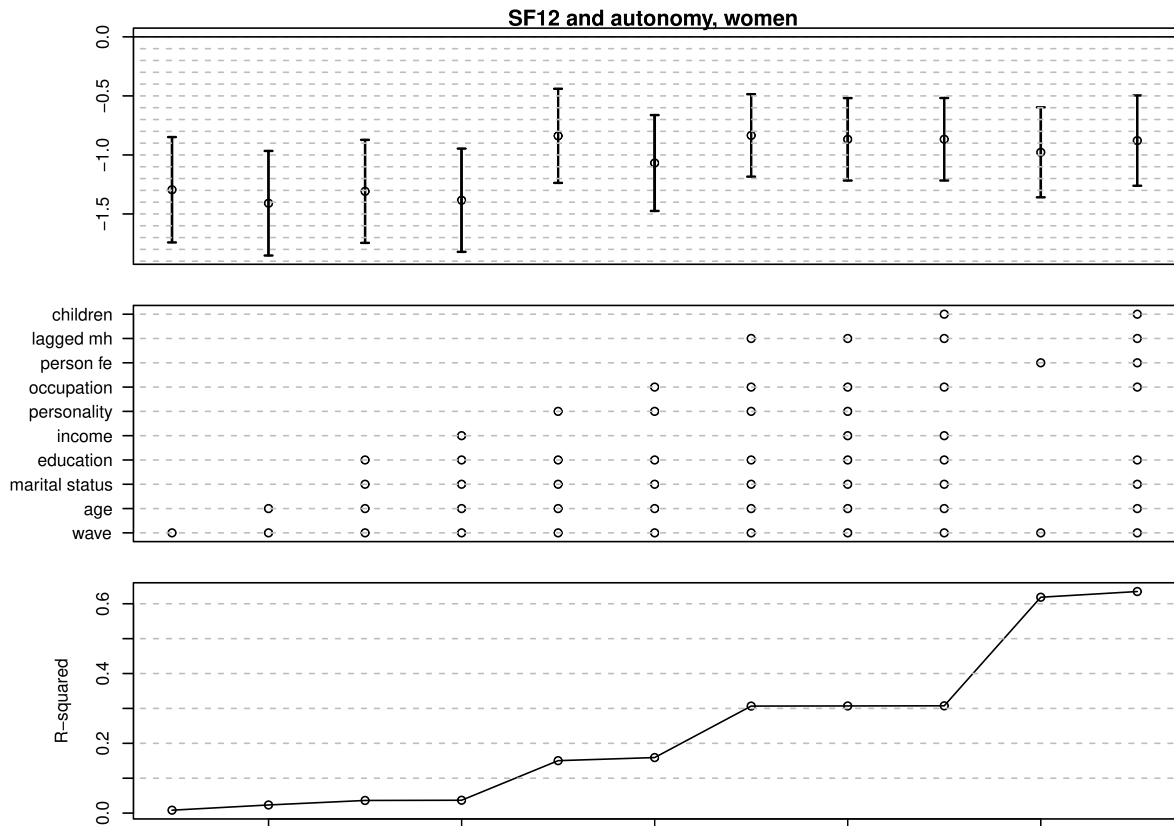
Notes: This figure shows how estimated associations between the SF12 index and levels of work-related autonomy, disaggregated across dimensions, amongst men change as new controls are included. Conditional associations are estimated by regressing the SF12 index on different measures of autonomy, with various controls. The SF12 index is described in the note to Figure 2.2. The top 5 panels report coefficients on autonomy, with confidence intervals using standard errors clustered at the individual level. The 6th panel shows which controls are included in each specification. The bottom panel shows the R-squared in each specification. Control variables are described in the note to Figure 2.1.

Figure 2.5: Conditional associations between the GHQ12 caseness score and low work-related autonomy amongst women



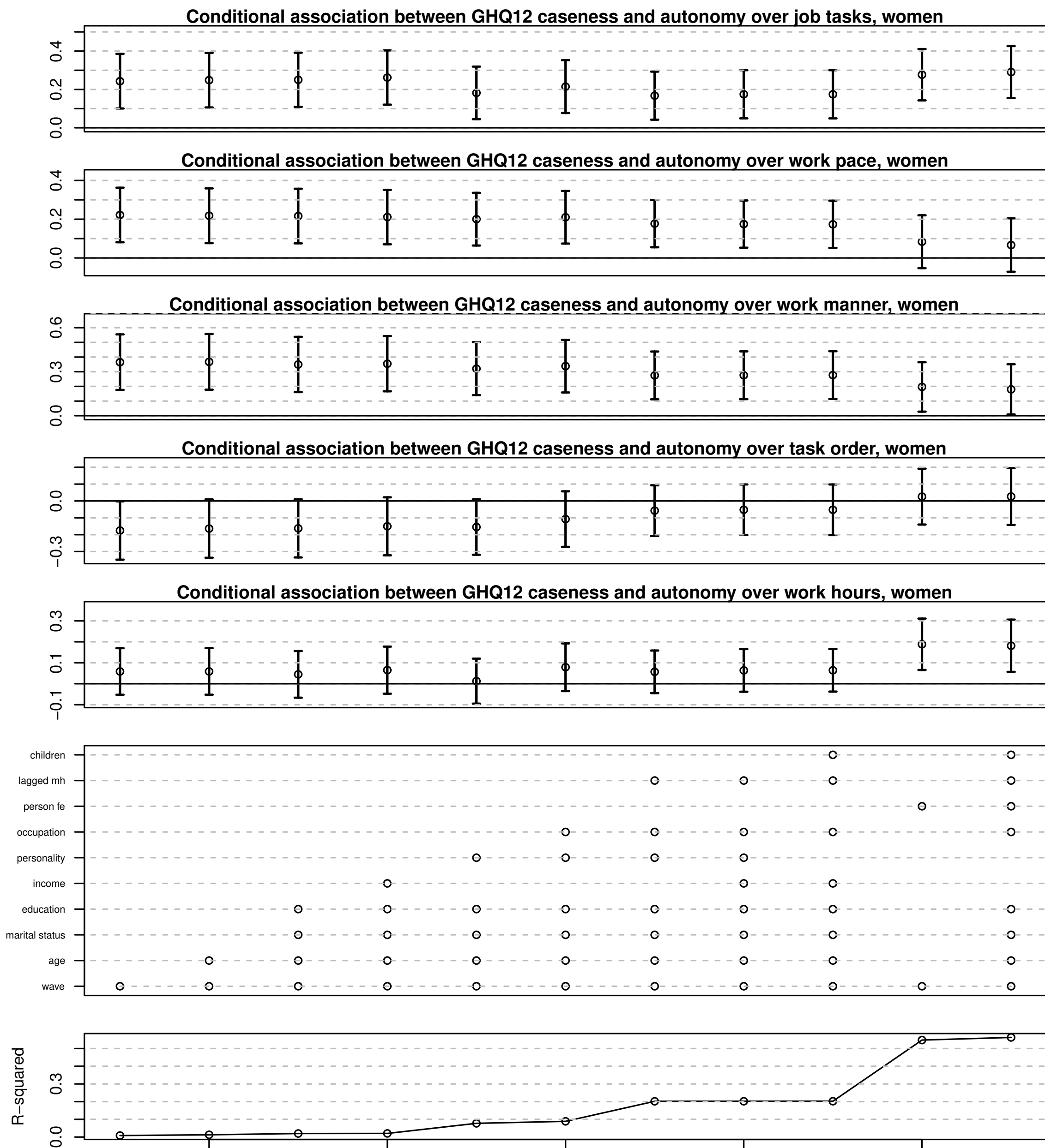
Notes: This figure shows how estimated associations between the GHQ12 caseness and having low work-related autonomy amongst women change as new controls are included. Conditional associations are estimated by regressing the GHQ12 caseness on the summary measure of autonomy, with various controls. The GHQ12 caseness is described in the note to Figure 2.1. The top panel reports coefficients on autonomy, with 95% confidence intervals using standard errors clustered at the individual level. The middle panel shows which controls are included in each specification. The bottom panel shows the R-squared in each specification. Control variables are described in the note to Figure 2.1.

Figure 2.6: Conditional associations between the SF12 index and low work-related autonomy amongst women



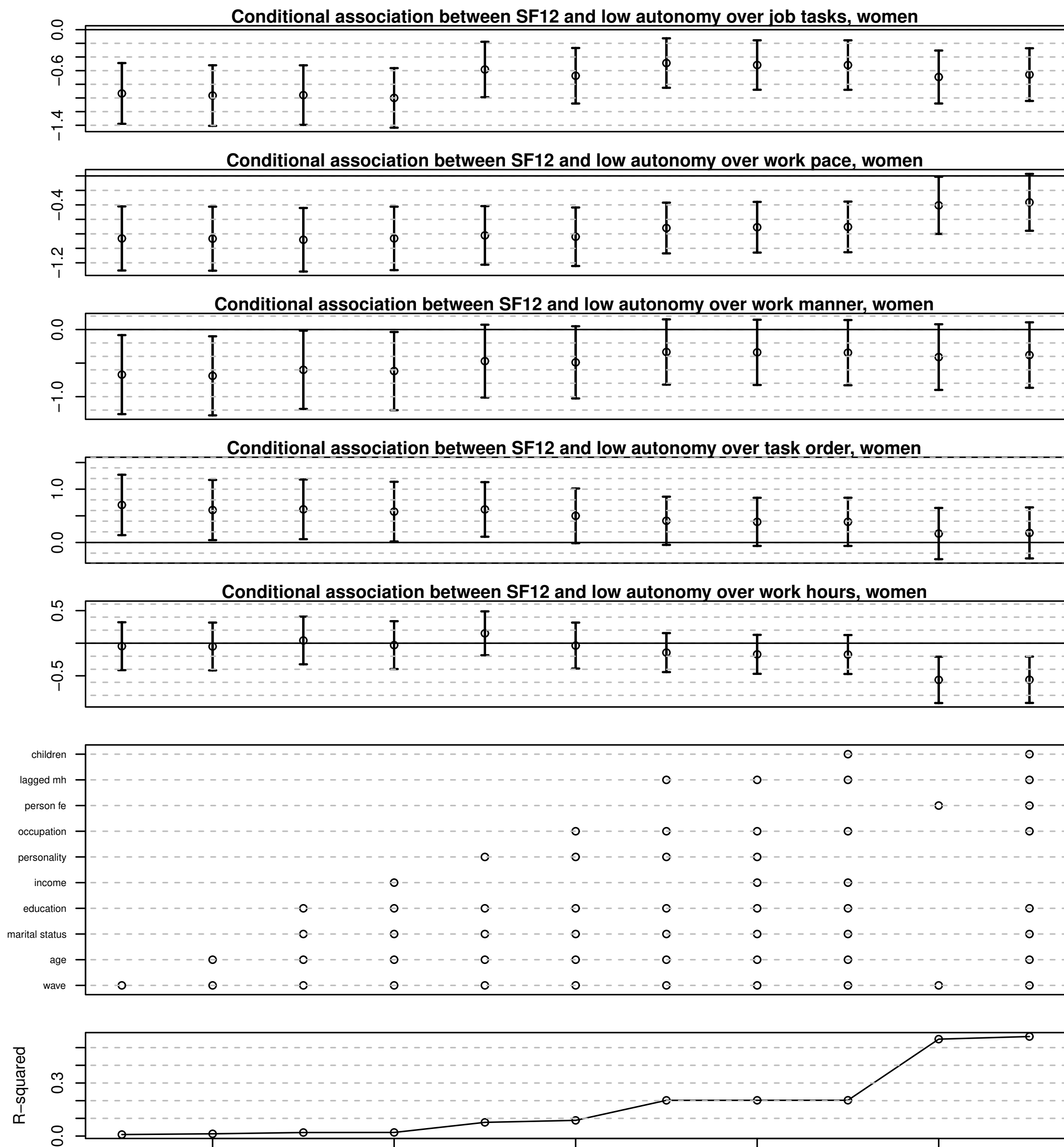
Notes: This figure shows how estimated associations between the SF12 index and having low work-related autonomy amongst women change as new controls are included. Conditional associations are estimated by regressing the SF12 index on the summary measure of autonomy, with various controls. The SF12 index is described in the note to Figure 2.2. The top panel reports coefficients on autonomy, with confidence intervals using standard errors clustered at the individual level. The middle panel shows which controls are included in each specification. The bottom panel shows the R-squared in each specification. Control variables are described in the note to Figure 2.1.

Figure 2.7: Conditional associations between the GHQ12 caseness score and low work-related autonomy across different dimensions amongst women



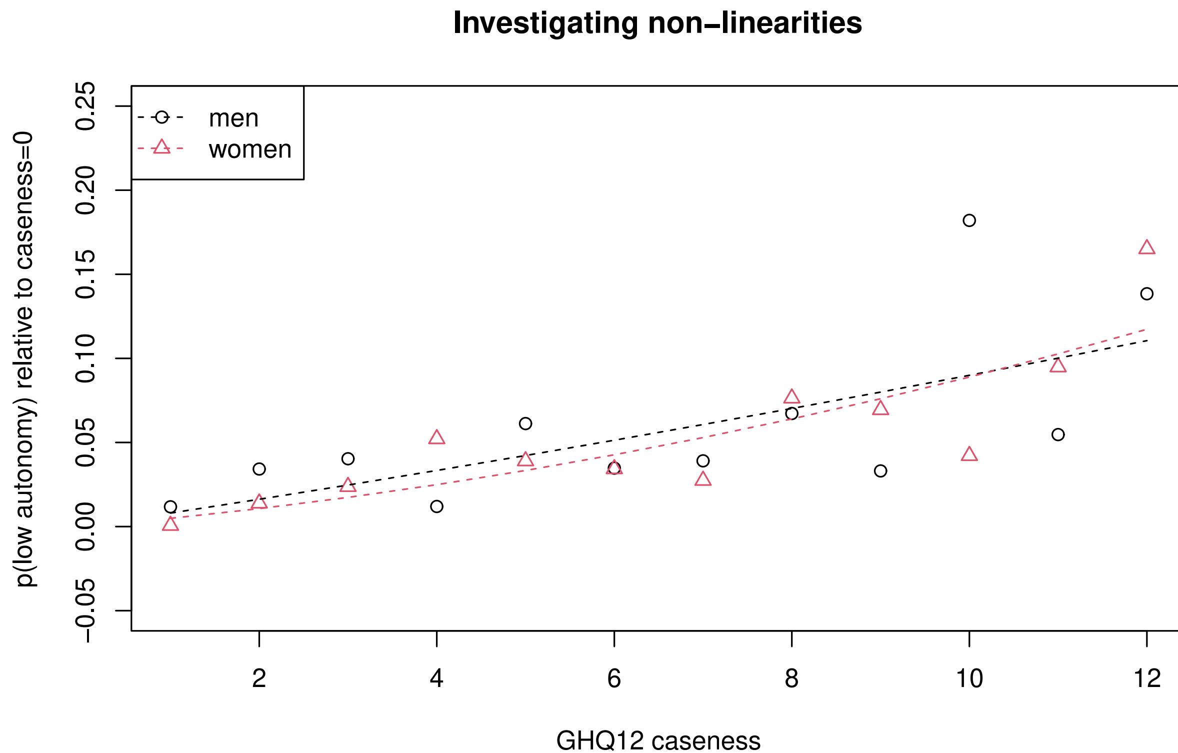
Notes: This figure shows how estimated associations between the GHQ12 caseness and having low work-related autonomy, disaggregated across multiple dimensions, amongst women change as new controls are included. Conditional associations are estimated by regressing the GHQ12 caseness on all measures of autonomy, with various controls. The GHQ12 caseness is described in the note to Figure 2.1. The top 5 panels report coefficients on autonomy, with confidence intervals using standard errors clustered at the individual level. The 6th panel shows which controls are included in each specification. The bottom panel shows the R-squared in each specification. Control variables are described in the note to Figure 2.1.

Figure 2.8: Conditional associations between the SF12 index and low work-related autonomy across different dimensions amongst women



Notes: This figure shows how estimated associations between the SF12 index and having low work-related autonomy, disaggregated across multiple dimensions, amongst women change as new controls are included. Conditional associations are estimated by regressing the SF12 index on different measures of autonomy, with various controls. The SF12 index is described in the note to Figure 2.2. The top 5 panels report coefficients on autonomy, with confidence intervals using standard errors clustered at the individual level. The 6th panel shows which controls are included in each specification. The bottom panel shows the R-squared in each specification. Control variables are described in the note to Figure 2.1.

Figure 2.9: Association between having low workplace autonomy and each value of the caseness score



Notes: This figure shows the estimated probability of having low autonomy at each level of the GHQ12 caseness relative to when the caseness score is 0. I estimate these probabilities from a linear probability model: I regress a dummy for having low autonomy on dummies for each value of the caseness score, controlling for wave fixed effects, person fixed effects, education, marital status, having biological children, and occupation dummies. The sample for this exercise is the sample of all those continuously in work. The lines are the estimated relationship from regressing a dummy for having low work-related autonomy on the caseness, the caseness squared, and the same controls.

Chapter 3

The effect of retirement eligibility on mental health in the United Kingdom: heterogeneous effects by occupation

3.1 Introduction

Due to the fiscal challenges of population ageing, many governments have increased the age of retirement eligibility. The distributional effect of these policy changes depends on how the health effects of retirement and retirement eligibility differ across occupations. I link the British Household Panel Survey/Understanding Society to O*NET, which contains detailed information about occupational characteristics. I use K-means clustering to define three occupation "clusters" and assess how the short-run effect of retirement eligibility on mental health differs across them using a Regression Discontinuity Design (RDD).

I find that there are two occupation clusters where workers' mental health improves more than the population average from retirement eligibility. The first is over-represented amongst skilled agricultural workers and craft and related trades, and includes mostly blue-collar occupations characterized by wearing Personal Protective Equipment (PPE), active body positioning, uncomfortable conditions, and job hazards. The second cluster with high benefits of retirement eligibility is over-represented amongst senior officials, professionals, associate professionals and clerks. This cluster is characterized by sedentary body positioning, high levels of communication, comfortable conditions and few job hazards, involves working with people, and is routine, competitive, and mostly white-collar. The mental health benefits of retirement eligibility are around half the population average for one cluster, which is over-represented in elementary occupations and service workers, and mixes blue- and white-collar jobs. These occupations

score the lowest for tight scheduling, highest for conflictual contact, and are the least competitive and routine.

My research relates to the literature estimating the effects of retirement eligibility and retirement on health outcomes (e.g., Neuman, 2008, Johnston and Lee 2009, Nishimura et al., 2018, Gorry et al., 2018, Müller and Shaikh, 2018, Rose, 2020, Geyer et al., 2020), including heterogeneous effects by occupation.[1] Most studies disaggregate the effect by one occupational characteristic, such as blue- and white-collar work (Kolodziej & García-Gómez, 2019), straining and non-straining jobs (Mazzonna and Peracchi, 2017), and physically demanding and less demanding jobs (Leckcivilize and McNamee, 2021). Research considering multiple occupational characteristics mostly considers one characteristic at a time (e.g., Eibich, 2015, Heller-Sahlgren, 2017). An exception is Carrino, Glaser, and Avendano (2020), who consider heterogeneity by levels of control and levels of physical/psychological demand simultaneously.

The novelty of this paper lies in employing K-means clustering to define occupational types, using rich data on occupational characteristics while maintaining precise RDD estimates. I contribute, firstly, by providing a richer characterization of the types of jobs where gaining eligibility to retire is more beneficial. Studies often examine effect heterogeneity across only one particular aspect of work (e.g., physical demands), but they cannot rule out that results are driven by other occupational characteristics that are correlated with the characteristic under consideration. My approach does not suffer from this concern. Secondly, by using a data-driven process to define occupational groupings, I do not impose *ad hoc* assumptions on the occupations that are restricted to have the same mental health effects. For example, previous literature has investigated whether being eligible to retire from white-collar and blue-collar jobs has differential effects (e.g., Kolodziej and García-Gómez, 2019). My approach separates white- and blue-collar jobs which have different characteristics. There are elevated benefits from becoming eligible to retire in some blue-collar jobs but not others; and some white-collar jobs but not others.

3.2 Data

I use a restricted version of the British Household Panel Survey/Understanding Society (BHPS/US),[2] which provides date of birth in months and occupation at the 4-digit ISCO-88 level, and O*NET.[3]

BHPS/US is a representative survey of the UK population with data on retirement status, household income (which I express in real equivalized log points) and the General Health Questionnaire (GHQ12). I use the GHQ12 to construct the GHQ12 caseness (McCabe et al., 1996), an index of mental health symptoms, and its sub-indices which measure the severity of symptoms of Anxiety and Depression, Loss

[1] There is also a literature which investigates heterogeneity across dimensions other than occupation, such as gender (Atalay & Barrett, 2014), marital status (Picchio and Ours, 2020, Della Giusta and Longhi, 2021), education (Della Giusta & Longhi, 2021), personality traits (Kesavayuth et al., 2016), quality of social life (Kettlewell & Lam, 2022), and the business cycle (Martinez Jimenez et al., 2021).

[2] University of Essex, Institute for Social and Economic Research. (2022). Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 15th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-16.

[3] O*NET 26.3 Database, by the US Department of Labor, Employment and Training Administration; https://creativecommons.org/licenses/by/4.0/.

of Confidence, and Social Dysfunction (Graetz, 1991). Higher values of mental health variables reflect worse mental health. I also consider a dummy variable for the caseness score being higher than 8, a key screening threshold (Anjara et al., 2020). Appendix C.1 details the construction of these indices. I use date of birth in months to construct expected age in months.[4]

The 57 O*NET "Work Context" variables report activities performed at work and the settings they are performed in. The O*NET survey interviews employees or experts in particular occupations. Respondents give a score on each variable for their occupation from an ordered list. The Department of Labor (DoL) reports a summary measure for each variable for each occupation, which I use as a measure of each occupation's score on each variable. Table 3.1 shows data on CEOs' levels of contact at work in the August 2020 wave of the O*NET data. Respondents choose from 5 ordered categories. The measure of CEOs' "contact at work" is calculated by assigning each response a number and calculating an "average" response:

$$(0 \times 1) + (0 \times 2) + (0.0268 \times 3) + (0.1483 \times 4) + (0.8249 \times 5) = 4.8 \tag{3.1}$$

For describing variation across occupational clusters, I aggregate variables using the DoL's "O*NET model",[5] which categorizes the 57 variables into 11 groups. I compute 11 indices as the average of the variables within them, choosing signs for each variable to give the index an intuitive economic meaning. I label the index to make this meaning explicit. For example, in the body positioning index, I assign a positive weight to sedentary body positioning (like sitting) and a negative weight to active body positioning (like standing) so the overall index is a measure of how sedentary body positioning is in that occupation. Table 3.2 presents a list of the indices, their constituent variables, and the sign of each variable. I use the economic interpretation of the indices in parentheses in the remainder of the paper. Each occupation's score on each O*NET variable is its average score over the period.[6]

While I study the effects of retirement eligibility on health outcomes in the United Kingdom, the O*NET data are collected in the United States. I therefore rely on similarity of the content of occupations between the two countries. Other researchers have applied the O*NET data to the European context (e.g., Goos et al., 2014, Hardy et al., 2018, Lewandowski, 2020), including the UK (Jolivet and Postel-Vinay, 2020). Applying the O*NET data to the UK is appropriate for several reasons: firstly, the UK and the US are at similar stages of economic development, with similar levels of education[7] and access to technology. These determinants of working conditions are therefore roughly constant. Secondly, there is significant overlap between firms' management in the US and the UK (Edwards et al., 2010) and we

[4]The survey does not report exact day of birth. I assume that day of birth is uniformly distributed. When a person is interviewed later in the month, their expected age is older than when they are interviewed earlier in the month. For respondents who are ever interviewed in their birth month, age at last birthday allows me to narrow down the day of the month that they were born on, so I tighten the bounds on their possible age accordingly.

[5]https://www.onetcenter.org/content.html?msclkid=203f5eb3bc3611ec8e6a560536451be2#cm4

[6]The rationale for averaging over the whole period is that results should not be affected by how characteristics of an occupation have changed since a person stopped working, i.e., a person's occupational cluster should not change because of changes to working conditions which have occurred since they retired. To determine the stability of occupational characteristics across the period I observe them, I estimate the correlation between an occupation's score on each O*NET work context variable in its first wave versus the most recent wave. For all variables the correlation is greater than 0.6, and for most it is above 0.7.

[7]Source: World Bank https://genderdata.worldbank.org/indicators/se-sch-life/?gender=total

might therefore expect management practices to be similar. Thirdly, when surveys similar to O*NET have been conducted in European countries, results have been closely correlated with those in the US (e.g., Handel, 2012, CEDEFOP, 2013). Hardy, Keister and Lewandowski (2018) provide further discussion of the applicability of O*NET to a European context.

I link the O*NET data to the BHPS/US data by current or last occupation at the 4-digit level using crosswalks provided by Hardy (2016).[8] My sample includes observations of people within 15 years of retirement eligibility who report a current or last occupation at least once within the survey, and which report values for all mental health variables and retirement status. I use this sample for descriptive analysis of the differences across people working in different types of occupations. Panel A of Table 3.3 presents summary statistics. There are 139,762 observations in the sample, although some observations have missing data on log household income. For each variable, I report the sample mean, standard deviation, and minimum and maximum values, as well as the mean in the full data set (i.e., including observations of respondents who are not within 15 years of retirement). In my main RDD estimates, I use a 5-year bandwidth. I present summary statistics for people within 5 years of retirement eligibility in Panel B of Table 3.3. Both samples are healthier than the data set as a whole on every measure of mental health. A greater share are retired, and male. They have roughly the same log household income.

3.3 Estimation

I define occupational groups using K-means clustering (Steinley, 2006, Bonhomme et al., 2022). I define a set of "centroids", each with a value for each O*NET variable, and assign each data point (work context scores for each person-wave observation) to a centroid, minimizing the total Euclidian distance between data points and the centroids they are assigned to. A "cluster" is a set of data points assigned to the same centroid.

For a given number of clusters N, I solve:

$$\min_{\{k_1,..k_I\}\in\{1,...,N\}} \Sigma_{i=1}^{I} ||h_i - \tilde{h}(k_i)|| \tag{3.2}$$

h_i is the vector of O*NET variables for observation i, k_i is the cluster assigned to observation i, and $\tilde{h}(k_i)$ is the mean vector of values for all points in cluster k.

I solve this problem using Lloyd's algorithm:

1. Allocate each data point to the cluster with the closest mean value.

[8]The crosswalk translates occupation codes across classification systems. It tells the researcher which code in, say, the ISCO-88 classification corresponds to each code in the SOC classification. Most papers (e.g., Carrino Glaser and Avendano 2020, Leckcivilize and McNamee 2021, Kolodziej and García-Gómez, 2019, Mazzonna and Peracchi 2017, and Bertoni Maggi and Weber 2018) which categorize people into different kinds of jobs in an RDD base the categorization on a person's current job or last job if not currently working. If a person has a history of working within the sample period, I classify their last occupation as the occupation they were last observed working in. Additionally, individuals who never work within the sample period report the last occupation they worked in if applicable.

2. Recalculate cluster means.

These steps are repeated until convergence. For my main results I use three clusters.[9] To verify my solution is the global minimum, I start the algorithm from 25 random centroids and select the smallest solution. Clustering differs from other dimension-reduction techniques such as factor and principal component analysis by producing discrete types, which is advantageous in this context because I flexibly interact cluster membership with the baseline RDD specification, avoiding unwarranted parametric assumptions, e.g., linearity of effect heterogeneity in factors.

I estimate the short-run effect of retirement eligibility using local linear regressions on the running variable, r_{it}, time to retirement eligibility in months, allowing for a discontinuity at r_{it} =0:

$$y_{it} = \alpha_0 + \delta_a r_{it} + \beta \mathbb{1}\{r_{it} < 0\} + \delta_b r_{it} \times \mathbb{1}\{r_{it} < 0\} + \tau_t + \epsilon_{it} \tag{3.3}$$

y_{it} is the outcome of interest for person i at time t. τ_t are wave fixed effects which can improve the precision of estimates (Lee and Lemieux, 2010). ϵ_{it} is the residual. α_0, δ_a, δ_b and β are regression coefficients, where β captures the discontinuity and the causal effect of interest. While the data are a panel, any characteristics which are fixed across individuals are not discontinuous in age, and therefore I do not include person fixed effects.[10]

The basic state pension pays a maximum of GBP 142 per week to people who have made sufficient tax contributions. The retirement eligibility age is the age at which a person can claim the basic state pension (the state pension age, SPA). In my sample, it is 65 for men, and between 60 and 65 for women.[11] The SPA is often also the age of retirement eligibility for occupational and private pensions (Hammond et al., 2016). Eligibility to claim pensions therefore changes discontinuously when a person reaches the SPA. For all men in my sample, this occurs on their 65th birthday, and therefore it is always possible to identify whether they are "treated", i.e., reached the SPA, from the "Age at last birthday" variable. Some women reach the SPA between birthdays. For some observations of those women it is not always possible to identify exactly which side of the cutoff their months to retirement eligibility falls. For those observations, I follow Dong's (2015) recommendation and drop them from the sample. Since healthcare is predominantly provided by the National Health Service for people of all ages, there is no confounding due to changes in health insurance eligibility. Within the age bandwidths studied, I am not aware of any other discontinuous changes which confound the relationship.

To estimate heterogeneous effects by occupation cluster, I allow the discontinuity and trend in the running variable to differ by occupation cluster membership, x_i:

[9] I choose three clusters in order to trade off the benefits of capturing a greater share of the variation in occupational characteristics against the costs of losing precision when the number of clusters increases. I discuss this tradeoff, and results using four clusters in Appendix C.6.

[10] Including person fixed effects could decrease precision significantly in this case because some individuals only appear in the bandwidth a small number of times (Lee and Lemieux 2010). For this reason, many researchers using a RDD do not use person fixed effects even when using panel data (e.g., Rose 2020, Watson 2020).

[11] Appendix C.2 provides details about the Pensions Acts which increased the age of retirement eligibility for women and how I incorporate the changing pension age into my analysis.

$$y_{it} = \alpha_0 + \delta_a r_{it} + \beta_1 \mathbb{1}\{r_{it} < 0\} + \delta_b r_{it} \times \mathbb{1}\{r_{it} < 0\} +$$
$$\Sigma_{j=2}^{J}(\delta_{j0}\mathbb{1}\{x_i = j\} + \delta_{aj} r_{it} \times \mathbb{1}\{x_i = j\} + \beta_j \mathbb{1}\{r_{it} < 0\} \times \mathbb{1}\{x_i = j\} + \delta_{cj} r_{it} \times \mathbb{1}\{r_{it} < 0\} \times \mathbb{1}\{x_i = j\}) +$$
$$\tau_t + \epsilon_{it} \tag{3.4}$$

Equation 3.4 is formed by interacting each term of Equation 3.3 with dummies for occupational cluster membership. The causal effect of retirement eligibility for those in cluster one is given by β_1 while the casual effect of retirement eligibility for those in cluster $j \neq 1$ is given by $\beta_1 + \beta_j$. I use a 5-year bandwidth[12] and a triangular kernel. Appendix C.3 and C.4 explore robustness of results to different kernels, bandwidths, and local polynomials. I cluster standard errors at the person level. Appendix C.5 investigates robustness to alternative standard error estimators.

Finally, while some researchers use retirement eligibility as an instrument for retirement (e.g., Heller-Sahlgren, 2017), I estimate the reduced-form effect of retirement eligibility on mental health, which has two key advantages. Firstly, it does not require the exclusion restriction assumption— that retirement eligibility only affects mental health via its effect on retirement behavior. Secondly, my approach directly assesses the effects of retirement eligibility, the key policy lever available to policymakers.

3.4 Results

3.4.1 Occupation groups

Table 3.4 presents the distribution of one-digit occupations across clusters. Occupations in cluster one are disproportionately in skilled agriculture, craft and related trades, and to a lesser extent plant and machinery operator workers. Cluster two jobs are disproportionately service jobs (e.g., travel attendants) plant and machinery operator workers, and elementary occupations (e.g., street vendors). Cluster three occupations are mostly white-collar, and largely comprise senior officials, professionals, associate professionals and clerks.

Figure 3.1 details the characteristics of the clusters resulting from K-means clustering. I plot the percentage of observations in each cluster who score "high" (in the top 25%) less the percentage of observations who score "low" (in the bottom 25%) on each O*NET index. Cluster one occupations are mostly outside, involve the least communication, working with people and conflictual contact, are the most likely to work in uncomfortable conditions, experience job hazards, wear PPE, and have tight scheduling. Cluster two jobs have the most conflictual contact, and score the lowest for tight scheduling. On the other hand, they do not score especially high or low on average for other job characteristics. Cluster three jobs score the lowest for wearing PPE, having uncomfortable conditions, and having high levels of job hazards. They score

[12] I present optimal bandwidths according to Calonico, Cattaneo and Titiunik (2014) in Table 3.6. The choice of bandwidth is based on trading off the decrease in variance from a larger bandwidth against the reduction in bias from reducing the bandwidth. Most optimal bandwidths are close to 5 years.

the highest for sedentary body positioning, communication, competitiveness and routineness, working indoors, and working with people.

Table 3.5 shows patterns of health, gender, and mental health across clusters. Employees in cluster one have lower household income, better mental health, and mostly blue-collar jobs and are disproportionately men. Cluster two mixes blue- and white-collar jobs and workers are disproportionately women. Cluster three employees have the highest household income and are 42% male.

3.4.2 Heterogeneous effects of retirement

Table 3.7 reports the effect of retirement eligibility overall and by cluster. Retirement eligibility has an overall beneficial effect on the overall GHQ12 caseness, the probability of the caseness being above clinically significant levels, symptoms of Social Dysfunction, symptoms of Loss of Confidence (but only significant at the 10% level), and symptoms of Anxiety and Depression. Effects are concentrated amongst those in cluster one and cluster three occupations. In cluster one, most of the benefit comes from symptoms of Anxiety and Depression; in cluster three, retirement eligibility affects symptoms of Anxiety and Depression and Social Dysfunction.

Men derive larger overall benefits from retiring. Disaggregating by cluster and gender, estimates are less precise and we should be more tentative about drawing conclusions. However, my results suggest occupational effect heterogeneity even within gender, especially amongst women. Therefore gender differences do not drive my main results. The overall muted effect of being eligible to retire from cluster two appears to be explained by a large beneficial effect amongst men and potentially an adverse effect amongst women (though the latter is not statistically significant). The beneficial effect of being eligible to retire from cluster one is present for men but not for women. The point estimates for cluster three are similar for both genders and are consistent with a beneficial effect which is large relative to the population average.

Differential patterns of retirement and changes in income are unlikely to drive my results: eligibility to retire from cluster two occupations has the smallest effect on mental health, but the largest effect on household income and second largest effect on the probability of retiring. My results are consistent with an adverse effect of working in cluster one and three occupations compared to retirement. The physical danger of cluster one occupations might increase stress about the risk of injury, and there are aspects of cluster three occupations associated with mental illness: they are sedentary, competitive and based indoors. Having the option of receiving the BSP may also improve mental health for those who work in these occupations, by improving the outside option for someone who is injured, fired, or otherwise loses their job. Alternatively, people who benefit most from retirement may disproportionately sort into cluster one and three occupations.

Appendix C.7 presents robustness tests. My results are qualitatively robust to dropping wave fixed effects and observations of people who retire before their retirement eligibility age.

3.5 Discussion and conclusion

This paper investigates how the effect of retirement eligibility on mental health outcomes varies across occupational characteristics. I use K-means clustering to estimate the heterogeneous effect of retirement eligibility across multiple dimensions simultaneously. The benefits of retirement are concentrated in identifiable occupational groups: one group of physical, dangerous, blue-collar jobs, and one group of professional, associate professional, or clerk jobs which are white collar and have high levels of contact with people.

My results provide novel evidence about the heterogeneous effects of retirement eligibility. Previous literature has estimated heterogeneous effects using a blue-collar and white-collar categorization (e.g., Kolodziej and García-Gómez, 2019). This approach does not allow for retirement eligibility to have heterogeneous effects within the blue-collar and white-collar categorizations, an important potential omission given that there is large variation in working conditions within blue-collar and white-collar occupations. I take a more detailed and data-driven approach to constructing occupational types. My approach suggests there are both blue- and white-collar occupations (clusters one and three) where the effect of retirement eligibility is elevated, and blue- and white-collar occupations (cluster two) where the effect is muted.

Secondly, previous literature has investigated mental health effects of psychologically straining jobs and physically straining jobs either separately (e.g., Eibich, 2015, Heller-Sahlgren, 2017, Mazzonna and Peracchi, 2017), or combining them into a single measure of job demand (Carrino et al., 2020). I show that there are two non-overlapping categories of occupations which are physically straining (cluster one) and psychologically straining (cluster three), and the causal effect of retirement eligibility varies between them. My results therefore indicate that the source of job strain is important and support differentiating between psychological and physical sources of strain.

Regarding policy, many developed economies are raising the age of retirement eligibility. The desirability of this policy may depend on who shoulders the burden. My results suggest that the largest mental health costs fall on those in the highest prestige, highest paying jobs (cluster three), followed by those in cluster one occupations, who have the lowest household income but better pre-retirement mental health.

My research has two limitations: firstly, I only identify the short-run effect of retirement eligibility;[13] secondly, I cannot distinguish between the hypothesis that the causal effect of being eligible to retire differs by occupation, and the hypothesis that people with different benefits of retirement sort into different occupations.

[13]An important rejoinder is that some studies have shown evidence of a "halo effect", a short-run beneficial effect which is not sustained (e.g., Heller-Sahlgren, 2017). To the extent that this is true, results RDD may not be a good indicator of the longer-run effect of retirement eligibility.

Table 3.1: O*NET measure of contact with people at work for CEOs, August 2020

Category	% of respondents
No contact with others	0
Occasional contact with others	0
Contact with others about half the time	2.68
Contact with others most of the time	14.83
Constant contact with others	82.49

Notes: Data are from O*NET. The O*NET study interviews either employees or experts in a given occupation and asks them to score the regularity of each variable within that occupation.

Table 3.2: O*NET indices and related variables

O*NET Index	Variables comprising the index
Work Setting (Working indoors)	Indoors, Environmentally Controlled (+) Indoors, Not Environmentally Controlled (+); Physical Proximity (-); Outdoors, Under Cover (-); In an Open Vehicle or Equipment (-); In an Enclosed Vehicle or Equipment (-); Outdoors, Exposed to Weather (-)
Communication (High communication)	Telephone (+); Electronic Mail (+); Letters and Memos (+); Face-to-Face Discussions (+); Contact with Others (+); Public Speaking (+)
Role Relationships (Work with people)	Coordinate or Lead Others (+); Work With Work Group or Team (+); Deal With External Customers (+)
Responsibility for Others (High responsibility)	Responsibility for Outcomes and Results (+); Responsible for Others' Health and Safety (+)
Conflictual Contact (High conflictual contact)	Deal With Unpleasant or Angry People (+); Deal with Physically Aggressive People (+); Frequency of Conflict Situations (+)
Environmental Conditions (Uncomfortable conditions)	Sounds/Noise Levels are Distracting or Uncomfortable (+); Very Hot or Cold Temperatures (+); Extremely Bright or Inadequate Lighting (+); Exposed to Contaminants (+); Cramped Work Space (+); Awkward Positions (+); Exposed to Whole Body Vibration (+)
Job Hazards (High job hazards)	Exposed to Radiation (+); Exposed to Disease or Infections (+); Exposed to High Places (+); Exposed to Hazardous Equipment (+); Exposed to Minor Burns, Cuts, Bites, or Stings (+); Exposed to Hazardous Conditions (+)

Notes: Indices are comprised of variables grouped by the O*NET model. The index is the average of the underlying measures. I sign each variable in the index to facilitate an easier interpretation of the overall index. The sign of each variable within the index is given by the (+) and (-) signs, and the economic interpretation of these indices is given in parentheses.

Table 3.2: O*NET indices and related variables

O*NET Index	Variables comprising the index
Body Positioning (Sedentary body positioning)	Spend Time Standing (-); Spend Time Sitting (+); Spend Time Climbing Ladders, Scaffolds, or Poles (-); Spend Time Walking and Running (-); Spend Time Kneeling, Crouching, Stooping, or Crawling (-); Spend Time Keeping or Regaining Balance (-); Spend Time Using Hands to Handle Control or Feel Objects Tools or Controls (-); Spend Time Bending or Twisting the Body (-); Spend Time Making Repetitive Motions (-)
Work Attire (Wear Personal Protective Equipment)	Wear Personal Protective Equipment (PPE) (+); Wear specialized PPE (+)
Cruciality of Position (Crucial position)	Freedom to Make Decisions (+); Consequence of Error (+); Impact of Decisions on Co-workers or Company Results (+); Frequency of Decision Making (+)
Routine vs. Challenging (Routine)	Degree of Automation (+); Importance of Being Exact or Accurate (+); Importance of Repeating Same Tasks (+); Structured versus Unstructured Work (+)
Level of Competition (Competitive)	Level of Competition (+)
Pace and Scheduling (Tight scheduling)	Time Pressure (+); Pace Determined by Speed of Equipment (+); Work Schedules (+); Duration of Typical Work Week (+)

Notes: Indices are comprised of variables grouped by the O*NET model. The index is the average of the underlying measures. I sign each variable in the index to facilitate an easier interpretation of the overall index. The sign of each variable within the index is given by the (+) and (-) signs, and the economic interpretation of these indices is given in parentheses.

Table 3.3: Sample summary statistics

Statistic	N	Mean	St. Dev.	Min	Max	Mean (full data set)
Panel A: 15 year bandwidth						
Age	139,762	62.8	7.9	45	80	59.9
GHQ12 caseness	139,762	1.49	2.8	0	12	1.74
1(Caseness>8)	139,762	0.06	0.25	0	1	0.08
Loss of Confidence	139,762	2.98	1.3	0	8	3.09
Social Dysfunction	139,762	12.5	2.1	5	24	12.6
Anxiety and Depression	139,762	7.15	2.5	0	16	7.42
Log household income	88,863	9.76	0.8	0	15	9.74
Retirement	139,762	0.49	0.50	0	1	0.38
Female	139,762	0.37	0.48	0	1	0.54
Panel B: 5 year bandwidth						
Age	47,235	62.9	3.4	52	70	
GHQ12 caseness	47,235	1.4	2.7	0	12	
1(Caseness>8)	47,235	0.06	0.24	0	1	
Loss of confidence	47,235	3.0	1.2	0	8	
Social dysfunction	47,235	12.5	2.0	6	24	
Anxiety and depression	47,235	7.1	2.5	1	16	
Log household income	28,723	9.8	0.79	0	14	
Retirement	47,235	0.50	0.50	0	1	
Female	47,235	0.36	0.48	0	1	

Notes: Data are from the Understanding Society/British Household Panel Survey. In Panel A, the sample includes all person-wave observations within 15 years of the retirement eligibility age, for individuals with data on their employment history. Sample statistics for each variable are reported after dropping missing values. Household income is adjusted for CPI and equivalence scales. The GHQ12 caseness score is the number of negative mental health symptoms a person currently experiences. Loss of Confidence, Social Dysfunction, and Anxiety and Depression measure the extent of negative mental health symptoms of these types. Retired is a binary variable equal to 1 if and only if a person's primary reported labor market status is retired. In Panel B I narrow the sample to observations of people within 5 years of the retirement eligibility age.

Table 3.4: Share of occupational employment by cluster

One-digit ISCO-88 occupation group	Cluster 1	Cluster 2	Cluster 3
Senior Officials	0.050	0.054	0.896
Professionals	0.010	0.184	0.807
Technicians and Associate Professionals	0.040	0.349	0.611
Clerks	0.014	0.061	0.925
Service Workers	0.058	0.874	0.069
Skilled Agriculture	0.925	0.014	0.061
Craft and Related Trades	0.798	0.121	0.081
Plant and Machinery Operator Workers	0.513	0.440	0.047
Elementary Occupations	0.413	0.529	0.058

Notes: The table shows the one-digit ISCO-88 occupation shares in each occupational cluster in the sample. I cluster person-wave observations by O*NET occupation data using K-means clustering.

Table 3.5: Patterns of income and health amongst those working in different occupational clusters

Cluster	Sample %	% Male	Average Monthly Household Income	Average Caseness Score	% White-collar
1	20.7	84.5	4,173	1.092	11.3
2	33.1	30.6	4,247	1.566	58.0
3	46.2	42.4	5,703	1.549	99.0

Notes: Cells are average values of each variable, conditional on working in a given occupation cluster. Variables are defined in the note to Table 3.3. Individuals are assigned to occupation clusters based on current occupation. While Understanding Society does not report whether a job is blue-collar or white-collar, the data do report the Goldthorpe-Hope scheme (see e.g., Erikson and Goldthorpe, 2002) an occupational categorization which can be collapsed to a blue-collar/white-collar categorization (Dhungel et al., 2021).

Table 3.6: Optimal bandwidths for each outcome

	MSE optimal bandwidth (years)
GHQ12 caseness	4.620
1(Caseness > 8)	4.891
Loss of Confidence	4.783
Social Dysfunction	4.363
Anxiety and Depression	4.741
Household Income	4.002
Retired	4.163

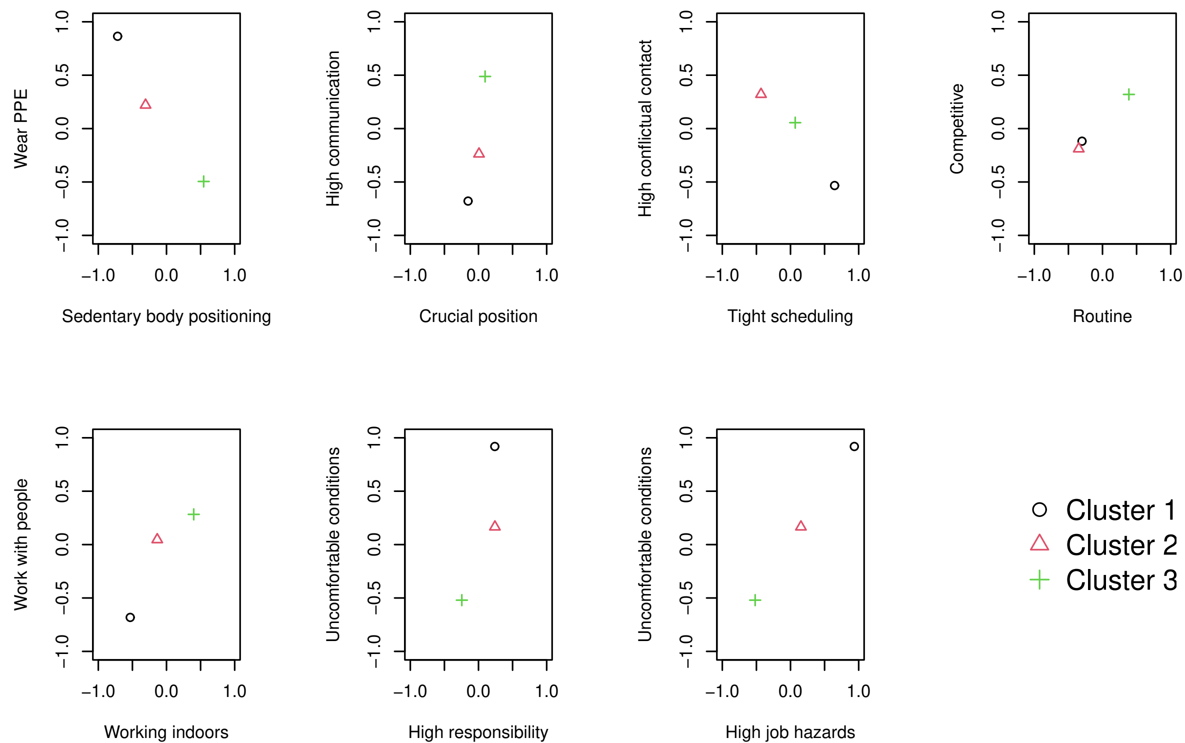
Notes: The sample is all person-wave observations within 15 years of the age of retirement eligibility that provide a value for each mental health variable and retirement. Optimal bandwidths are calculated using the procedures developed by Calonico, Cattaneo, Titiunik (2014), which minimize the mean squared error of the RDD estimator. Variables are defined in the note to Table 3.3.

Table 3.7: The effect of retirement eligibility by occupational cluster (using RDD)

	GHQ12 caseness	1(caseness > 8)	Loss of Confidence	Social Dysfunction	Anxiety and Depression	Log Household Income	Retired
				Panel A: Men and Women			
All clusters	-0.115^{***}	-0.008^{**}	-0.039^{*}	-0.122^{***}	-0.116^{***}	0.134^{***}	0.226^{***}
	(0.044)	(0.004)	(0.020)	(0.033)	(0.040)	(0.017)	(0.008)
Cluster 1	−0.126	−0.006	−0.049	−0.095	-0.200^{**}	0.139^{***}	0.299^{***}
	(0.089)	(0.008)	(0.041)	(0.066)	(0.081)	(0.032)	(0.015)
Cluster 2	−0.048	−0.004	−0.032	−0.084	−0.029	0.190^{***}	0.239^{***}
	(0.078)	(0.007)	(0.036)	(0.058)	(0.071)	(0.030)	(0.013)
Cluster 3	-0.158^{***}	-0.012^{**}	−0.040	-0.166^{***}	-0.136^{***}	0.094^{***}	0.176^{***}
	(0.066)	(0.006)	(0.030)	(0.049)	(0.060)	(0.025)	(0.011)
Obs	47,235	47,235	47,235	47,235	47,235	32,641	47,235
				Panel B: Men			
All clusters	-0.159^{***}	-0.011^{***}	-0.049^{**}	-0.150^{***}	-0.109^{**}	0.138^{***}	0.233^{***}
	(0.050)	(0.004)	(0.024)	(0.037)	(0.047)	(0.019)	(0.009)
Cluster 1	-0.173^{*}	−0.011	−0.061	-0.126^{*}	-0.231^{***}	0.139^{***}	0.308^{***}
	(0.092)	(0.008)	(0.044)	(0.068)	(0.087)	(0.035)	(0.017)
Cluster 2	-0.162^{*}	−0.011	−0.058	-0.132^{*}	−0.019	0.196^{***}	0.243^{***}
	(0.098)	(0.008)	(0.046)	(0.072)	(0.092)	(0.035)	(0.018)
Cluster 3	-0.150^{**}	−0.012	−0.037	-0.177^{***}	−0.082	0.095^{***}	0.174^{***}
	(0.076)	(0.007)	(0.036)	(0.056)	(0.072)	(0.028)	(0.014)
Obs	30,486	30,486	30,486	30,486	30,486	25,576	30,486
				Panel C: Women			
All clusters	−0.013	−0.002	−0.014	−0.052	-0.135^{*}	0.090	0.223^{***}
	(0.086)	(0.008)	(0.037)	(0.065)	(0.073)	(0.043)	(0.013)
Cluster 1	0.020	0.008	−0.012	0.030	−0.164	0.194	0.277^{***}
	(0.230)	(0.021)	(0.100)	(0.175)	(0.196)	(0.124)	(0.034)
Cluster 2	0.115	0.005	−0.0002	−0.013	−0.041	0.166^{**}	0.244^{***}
	(0.131)	(0.012)	(0.057)	(0.100)	(0.112)	(0.067)	(0.019)
Cluster 3	−0.150	−0.011	−0.032	−0.120	-0.224^{**}	0.044	0.188^{***}
	(0.128)	(0.012)	(0.056)	(0.097)	(0.109)	(0.059)	(0.019)
Obs	16,749	16,749	16,749	16,749	16,749	7,065	16,749

Notes: Results are the short-run causal effect of being eligible to retire for those in each occupational cluster. I use a local linear regression around the cutoff of zero months to retirement eligibility. The running variable is expected months to retirement eligibility. A bandwidth of 5 years is used, and observations are weighted using a triangular kernel. For heterogeneous effects by cluster, I also interact with occupational cluster membership as in Equation 3.4, and use the sum of coefficients to infer the causal effects presented in the Table, i.e., the causal effect of retirement eligibility for cluster one is β_1 in Equation 3.4, and $\beta_1 + \beta_j$ for cluster j. I control for wave fixed effects. Standard errors are clustered at the individual level. I calculate the standard error of the causal effect by using the heteroskedasticity-robust variance-covariance matrix clustered at the individual level. Where I allow the effect to differ by gender, I estimate the effect separately in sample of men and women. Variables are defined in the note to Table 3.3. *p<0.1; **p<0.05; ***p<0.01.

Figure 3.1: Percent of people in each cluster scoring high minus percent scoring low



Notes: I cluster person-wave observations by O*NET occupation data using K-means clustering. The measures of cluster characteristics are calculated by subtracting the percentage of people who score "low" on an O*NET variable within a cluster from the percentage of people scoring "high", where scoring "high" or "low" is determined by being in the top 25% or bottom 25% respectively of the distribution of scores. Individuals are assigned to occupation clusters based on current or last occupation. O*NET indices are calculated based on the O*NET model, which groups variables into categories. I calculate the indices as the simple average over the variables in the index, which are signed to facilitate an intuitive interpretation (see Table 3.2).

Appendix A

The welfare effects of work restrictions for disability benefit claimants: evidence from the UK

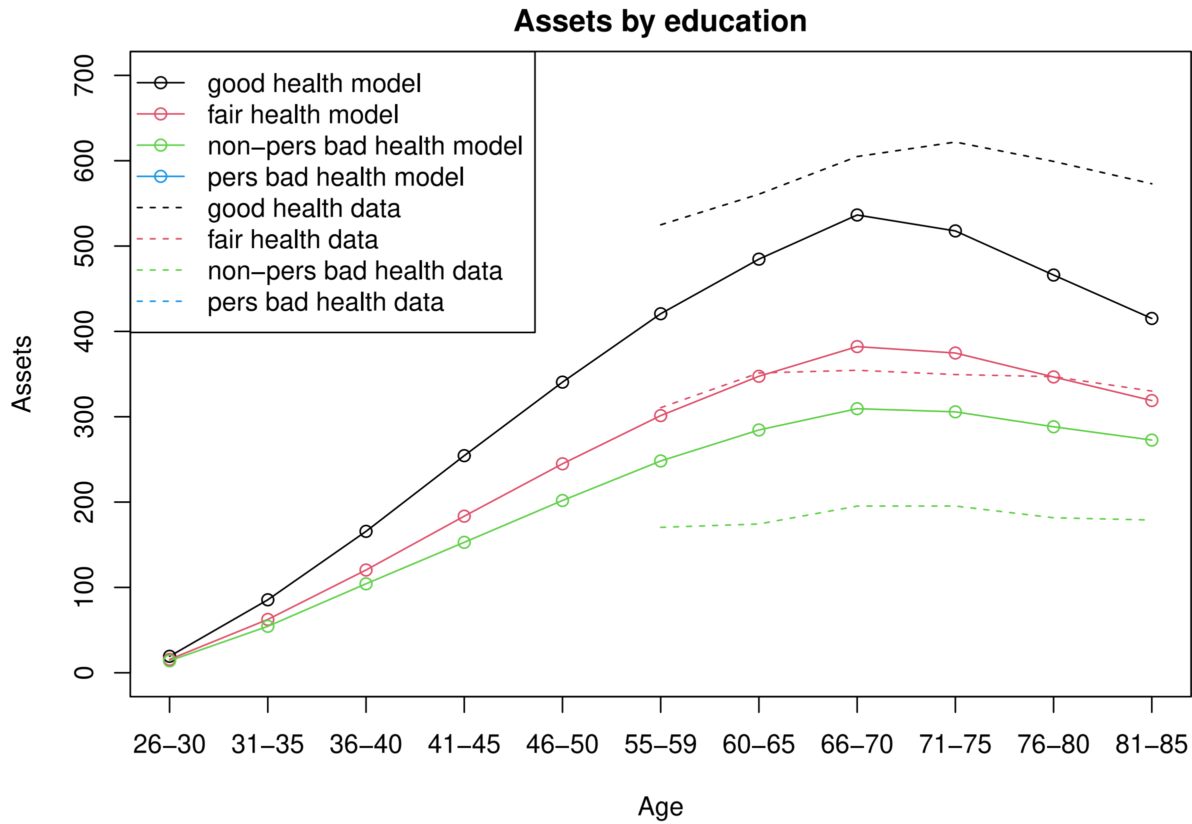
A.1 Results with discount rate fixed across education types

In the model, I estimate the discount rates separately for each education category in order to match asset accumulation by education. This might raise some concerns about welfare calculations. In particular, allowing the patience parameter to differ by education means that agents with different amounts of education will evaluate policy changes which shift resources across the life cycle differently. This effect may obscure important information about the distributional effects of reform.

In order to investigate how important different patience parameters are in driving results, I re-estimate my model while fixing the patience parameter at 0.98 for all levels of education. 0.98 is a commonly used value in the literature for the UK (see e.g., Salvati, 2021). The resulting model is less effective in capturing the degree of asset inequality (see Figure A.1). I overestimate asset accumulation for those with no qualifications or other qualification, and underestimate asset accumulation for those with a degree or equivalent.

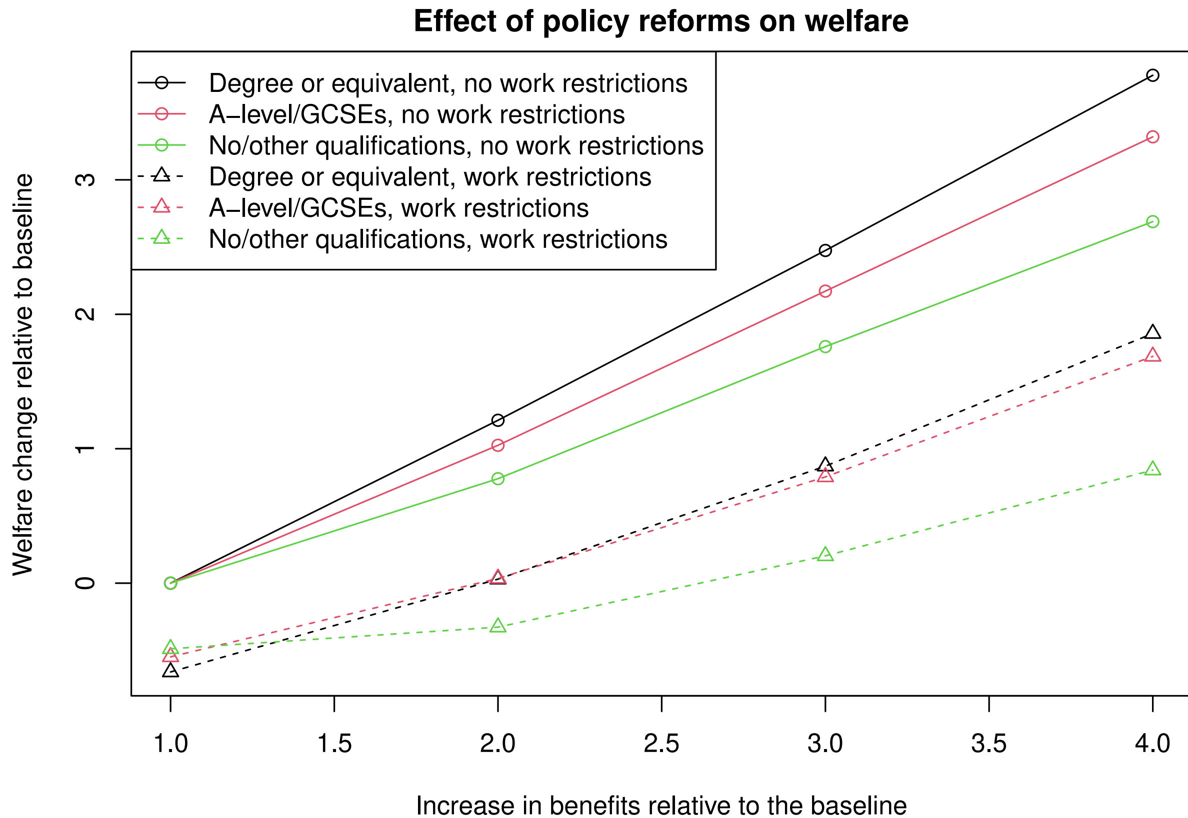
Given the newly estimated model, I re-run my policy experiments as described in Section 1.6. Table A.1 and Figure **??** report the results. Overall, they are consistent with the results reported in the text. Increases in benefits are always welfare-improving, and introducing work restrictions always decreases welfare, holding the level of benefit generosity fixed. As with my main model, this result is consistent even for large increases in benefits. Interestingly, even when the discount rate is fixed across education types, it is consistently those agents with more education who value more generous benefits. However, those with the least education see the largest welfare loss from the introduction of work restrictions when benefits are increased above the baseline level.

Figure A.1: Asset accumulation by education in the model where the discount rate is fixed at 0.98 for all education types.



Notes: I re-estimate my model while fixing the discount rate at 0.98 for all education levels. See notes to Table 1.5 for a description of education types and the construction of moments.

Figure A.2: Welfare effects of reforming disability benefits in the model where the discount rate is fixed at 0.98 for all education types.



Notes: I repeat the policy experiments as described in section 1.6, with my model solved holding discount rates fixed. See notes to Figure 1.11 for a description of key variables.

Table A.1: Effects of policy reform where the discount rate is fixed across education

	Baseline	Doubling benefits	Work restrictions
DLA/PIP payment per person (GBP 1000s)	0.236	0.576	0.145
Average labor income (GBP 1000s)	20.354	20.350	20.310
Average assets aged 65, (GBP 1000s)			
good health	627.388	617.226	626.417
fair health	633.773	625.437	632.063
bad health	614.177	614.021	610.188
Welfare equivalent % change in (lifetime) consumption			
Degree or equivalent		1.212	−0.660
A-level/GCSE		1.025	−0.548
No/other qualification		0.777	−0.488

Notes: I repeat the policy experiments as described in Section 1.6, with my model solved holding patience parameters fixed. See notes to Table 1.11 for a description of key variables.

A.2 Model solution

Conditional on surviving to the age of t =85, an individual's optimal decision is to allocate resources between consumption and bequests:

$$V(t=85, k_{i85}, h_{i85}, p_{i85}, b_{i85}, e_i) = \max_{c_{i85}} u(c_{i85}, 0, 0, h_{i85}, p_{i85}) + \beta_i u_B(k_{i86}) \tag{A.1}$$

subject to the budget constraint:

$$k_{i86} = k_{i85}(1+r) + \mathbb{1}\{b_{i85}=1\}y^{\text{PIP}} + \text{BEN}(85, h_{i85}, 0) - c_{i85}$$

Retired agents younger than 85 have the value function:

$$V(\Omega_{it}^R) = \max_{c_{it}} u(c_{it}, 0, 0, h_{it}, p_{it}) + \beta_i(\text{p_surv}(t, h_{it})\mathbb{E}[V(\Omega_{it+1}^R)] + (1-\text{p_surv}(t, h_{it}))u_B(k_{it+1})) \tag{A.2}$$

subject to the budget constraint:

$$k_{it+1} = k_{it}(1+r) + y^{\text{PIP}}\mathbb{1}(b_{it}=1) + \text{BEN}(t, h_{it}, \ell_{it}w_{it}) - c_{it}$$

The expectation function integrates over the conditional distribution of state variables in the next period. Age evolves deterministically, assets are deterministic conditional on the agent's choice, h and p evolve stochastically according to the exogenous health process and b_{it+1} is a stochastic function of health given that an agent currently claims (see equation 1.13). The choice of consumption which maximizes lifetime utility given the state space implies a policy function: $c(\Omega_{it}^R)$.

For working-age agents, the value function is:

$$V(\Omega_{it}^W) = \max_{d_{it}} \{U(\Omega_{it}^W, d_{it}) + \nu_{itd}\} \tag{A.3}$$

The agent selects the best of all possible discrete choices. $U(\Omega_{it}^W, d_{it})$ is a function which maps the state space onto the expected lifetime utility of each discrete choice when the value of the preference shock is zero.

$$U(\Omega_{it}^W, d_{it}) = \max_{c_{it}} u(c_{it}, \ell(d_{it}), \text{app}(d_{it}), h_{it}, p_{it}) + \beta_i \text{EMax}[V(\Omega_{it+1}^W)]$$

subject to:

$k_{it+1} = (1+r)k_{it} + \ell(d_{it})w(t, h_{it}, \epsilon_{it}, e_i) + y^{\text{PIP}}\mathbb{1}(b_{it} = 1) + \text{BEN}(t, h_{it}, \ell_{it}w_{it}) + \text{SPOU}_{it} - T(\ell(d_{it})w_{it}) - c_{it}$

The EMax function is the expectation of the value function next period, integrating over state variables and preference shocks. By the law of iterated expectations and type 1 extreme value preference shocks:

$$\text{EMax}[V(\Omega_{it+1}^W)] = \mathbb{E}_{\Omega'}[\log(\Sigma_{d'} \exp[U(\Omega_{it+1}^W = \Omega', d_{it+1} = d')])] \tag{A.4}$$

Conditional on an agent's discrete choice, the policy function is given by $c(\Omega_{it}^W, d_{it})$. Given the state space and the type 1 extreme value distribution of preference shocks, the probability that an agent chooses discrete choice d' is given by:

$$\Pr(d_{it} = d' | \Omega_{it}^W) = \frac{\exp(U(\Omega_{it}^W, d'))}{\Sigma_d \exp(U(\Omega_{it}^W, d))} \tag{A.5}$$

I solve by backward induction, i.e., sequentially solving for the agent's optimal behavior for all points in the state space starting with the oldest age and working backwards. I discretize assets into 50 points, and wage shocks into 5 points. Policy functions are solved for at each discrete point in the state space, and in simulations, I use linear interpolation between points on the capital grid to simulate consumption decisions.

A.3 Health process parameters

Table A.2 shows the estimated parameters for the health process.

Table A.2: Health process parameter estimates

	Parameter	Estimate	Standard error
$\gamma^2_{0,e=1,h_{it}=1}$	Intercept, education=1, $h_{it} = 1$	3.86***	(0.088)
$\gamma^2_{1,e=1,h_{it}=1}$	Age coefficient , education=1, $h_{it} = 1$	0.02***	(0.002)
$\gamma^2_{2,e=1,h_{it}=1}$	Persistence coefficient, education=1, $h_{it} = 1$	−0.31***	(0.048)
$\gamma^2_{0,e=2,h_{it}=1}$	Intercept, education=2, $h_{it} = 1$	−3.42***	(0.162)
$\gamma^2_{1,e=2,h_{it}=1}$	Age coefficient , education=2, $h_{it} = 1$	0.02***	(0.003)
$\gamma^2_{2,e=2,h_{it}=1}$	Persistence coefficient, education=2, $h_{it} = 1$	−0.35***	(0.089)
$\gamma^2_{0,e=3,h_{it}=1}$	Intercept, education=3, $h_{it} = 1$	−3.12***	(0.489)
$\gamma^2_{1,e=3,h_{it}=1}$	Age coefficient , education=3, $h_{it} = 1$	0.02***	(0.009)
$\gamma^2_{2,e=3,h_{it}=1}$	Persistence coefficient, education=3, $h_{it} = 1$	−0.34	(0.292)
$\gamma^2_{0,e=1,h_{it}=2}$	Intercept, education=1, $h_{it} = 2$	−1.35***	(0.27)
$\gamma^2_{1,e=1,h_{it}=2}$	Age coefficient , education=1, $h_{it} = 2$	0.02***	(0.005)
$\gamma^2_{2,e=1,h_{it}=2}$	Persistence coefficient, education=1, $h_{it} = 2$	1.10***	(0.153)
$\gamma^2_{0,e=2,h_{it}=2}$	Intercept, education=2, $h_{it} = 2$	−1.01***	(0.377)
$\gamma^2_{1,e=2,h_{it}=2}$	Age coefficient , education=2, $h_{it} = 2$	0.02***	(0.007)
$\gamma^2_{2,e=2,h_{it}=2}$	Persistence coefficient, education=2, $h_{it} = 2$	0.95***	(0.202)
$\gamma^2_{0,e=3,h_{it}=2}$	Intercept, education=3, $h_{it} = 2$	−0.75	(0.538)
$\gamma^2_{1,e=3,h_{it}=2}$	Age coefficient , education=3, $h_{it} = 2$	0.01	(0.01)
$\gamma^2_{2,e=3,h_{it}=2}$	Persistence coefficient, education=3, $h_{it} = 2$	0.89***	(0.314)
$\gamma^2_{0,e=1,h_{it}=3}$	Intercept, education=1, $h_{it} = 3$	−0.84***	(0.06)
$\gamma^2_{1,e=1,h_{it}=3}$	Age coefficient , education=1, $h_{it} = 3$	0.03***	(0.001)
$\gamma^2_{2,e=1,h_{it}=3}$	Persistence coefficient, education=1, $h_{it} = 3$	0.92***	(0.039)
$\gamma^2_{0,e=2,h_{it}=3}$	Intercept, education=2, $h_{it} = 3$	−0.60***	(0.11)
$\gamma^2_{1,e=2,h_{it}=3}$	Age coefficient , education=2, $h_{it} = 3$	0.03***	(0.002)
$\gamma^2_{2,e=2,h_{it}=3}$	Persistence coefficient, education=2, $h_{it} = 3$	0.67***	(0.069)
$\gamma^2_{0,e=3,h_{it}=3}$	Intercept, education=3, $h_{it} = 3$	−0.72**	(0.285)
$\gamma^2_{1,e=3,h_{it}=3}$	Age coefficient , education=3, $h_{it} = 3$	0.03***	(0.005)
$\gamma^2_{2,e=3,h_{it}=3}$	Persistence coefficient, education=3, $h_{it} = 3$	0.81***	(0.204)
$\gamma^3_{0,e=1,h_{it}=1}$	Intercept, education=1, $h_{it} = 1$	−5.86***	(0.19)
$\gamma^3_{1,e=1,h_{it}=1}$	Age coefficient , education=1, $h_{it} = 1$	0.02***	(0.003)
$\gamma^3_{2,e=1,h_{it}=1}$	Persistence coefficient, education=1, $h_{it} = 1$	−0.77***	(0.123)

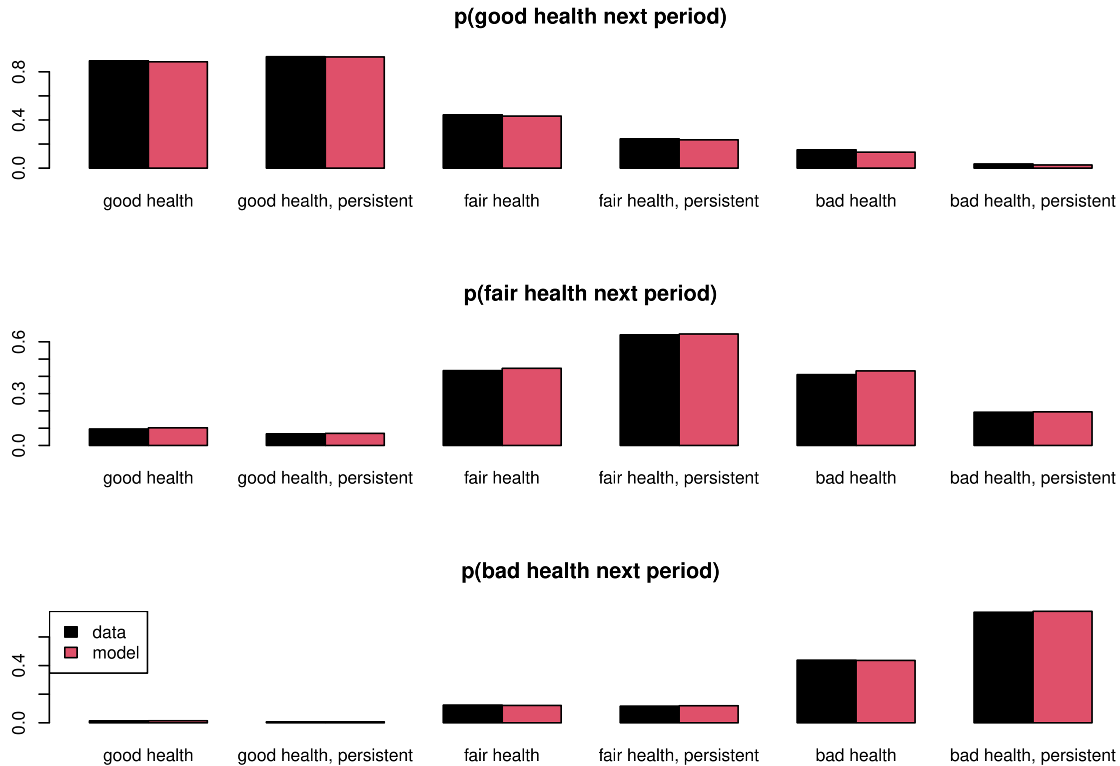
Notes: The health process parameters are estimated by maximizing the likelihood of observed conditional health status transitions. *p<0.1; **p<0.05; ***p<0.01

Table A.2: Health process parameter estimates

Parameter		Estimate	Standard error
$\gamma^3_{0,e=2,h_{it}=1}$	Intercept, education=2, $h_{it} = 1$	-5.84^{***}	(0.254)
$\gamma^3_{1,e=2,h_{it}=1}$	Age coefficient , education=2, $h_{it} = 1$	0.02^{***}	(0.004)
$\gamma^3_{2,e=2,h_{it}=1}$	Persistence coefficient, education=2, $h_{it} = 1$	-0.76^{***}	(0.154)
$\gamma^3_{0,e=3,h_{it}=1}$	Intercept, education=3, $h_{it} = 1$	-5.10^{***}	(0.332)
$\gamma^3_{0,e=3,h_{it}=1}$	Intercept, education=3, $h_{it} = 1$	-5.10^{***}	(0.332)
$\gamma^3_{1,e=3,h_{it}=1}$	Age coefficient , education=3, $h_{it} = 1$	0.02^{***}	(0.006)
$\gamma^3_{2,e=3,h_{it}=1}$	Persistence coefficient, education=3, $h_{it} = 1$	-0.56^{**}	(0.227)
$\gamma^3_{0,e=1,h_{it}=2}$	Intercept, education=1, $h_{it} = 2$	-3.61^{***}	(0.101)
$\gamma^3_{1,e=1,h_{it}=2}$	Age coefficient , education=1, $h_{it} = 2$	0.03^{***}	(0.002)
$\gamma^3_{2,e=1,h_{it}=2}$	Persistence coefficient, education=1, $h_{it} = 2$	0.66^{***}	(0.05)
$\gamma^3_{0,e=2,h_{it}=2}$	Intercept, education=2, $h_{it} = 2$	-3.08^{***}	(0.181)
$\gamma^3_{1,e=2,h_{it}=2}$	Age coefficient, education=2, $h_{it} = 2$	0.02^{***}	(0.003)
$\gamma^3_{2,e=2,h_{it}=2}$	Persistence coefficient, education=2, $h_{it} = 2$	0.64^{***}	(0.083)
$\gamma^3_{0,e=3,h_{it}=2}$	Intercept, education=3, $h_{it} = 2$	-1.80^{***}	(0.411)
$\gamma^3_{1,e=3,h_{it}=2}$	Age coefficient , education=3, $h_{it} = 2$	0.01^{***}	(0.007)
$\gamma^3_{2,e=3,h_{it}=2}$	Persistence coefficient, education=3, $h_{it} = 2$	0.37^{*}	(0.211)
$\gamma^3_{0,e=1,h_{it}=3}$	Intercept, education=1, $h_{it} = 3$	-0.57^{**}	(0.27)
$\gamma^3_{1,e=1,h_{it}=3}$	Age coefficient , education=1, $h_{it} = 3$	0.03^{***}	(0.004)
$\gamma^3_{2,e=1,h_{it}=3}$	Persistence coefficient, education=1, $h_{it} = 3$	2.16^{***}	(0.133)
$\gamma^3_{0,e=2,h_{it}=3}$	Intercept, education=2, $h_{it} = 3$	-0.65^{*}	(0.348)
$\gamma^3_{1,e=2,h_{it}=3}$	Age coefficient , education=2, $h_{it} = 3$	0.03^{***}	(0.005)
$\gamma^3_{2,e=2,h_{it}=3}$	Persistence coefficient, education=2, $h_{it} = 3$	2.07^{***}	(0.164)
$\gamma^3_{0,e=3,h_{it}=3}$	Intercept, education=3, $h_{it} = 3$	-0.57	(0.472)
$\gamma^3_{1,e=3,h_{it}=3}$	Age coefficient , education=3, $h_{it} = 3$	0.02^{***}	(0.008)
$\gamma^3_{2,e=3,h_{it}=3}$	Persistence coefficient, education=3, $h_{it} = 3$	2.33^{***}	(0.237)

Notes: The health process parameters are estimated by maximizing the likelihood of observed conditional health status transitions. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Figure A.3: Estimated versus simulated health transitions, over 50s

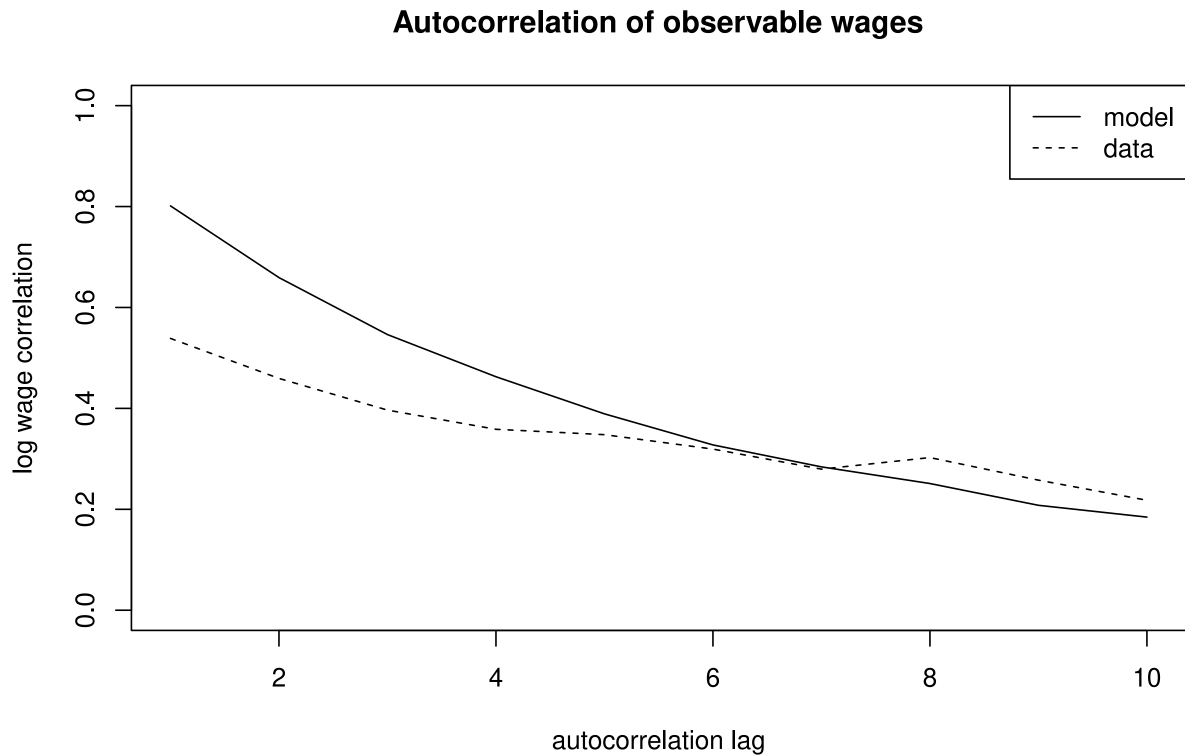


Notes: Transitions between health states for over 50s, data and simulations. Persistent health states are health states an agent has been in for more than one year. The health process is estimated by maximizing log likelihood (see equation 1.20). Simulated moments are weighted by education and age to be comparable with survey data.

A.4 Simulated versus data moments

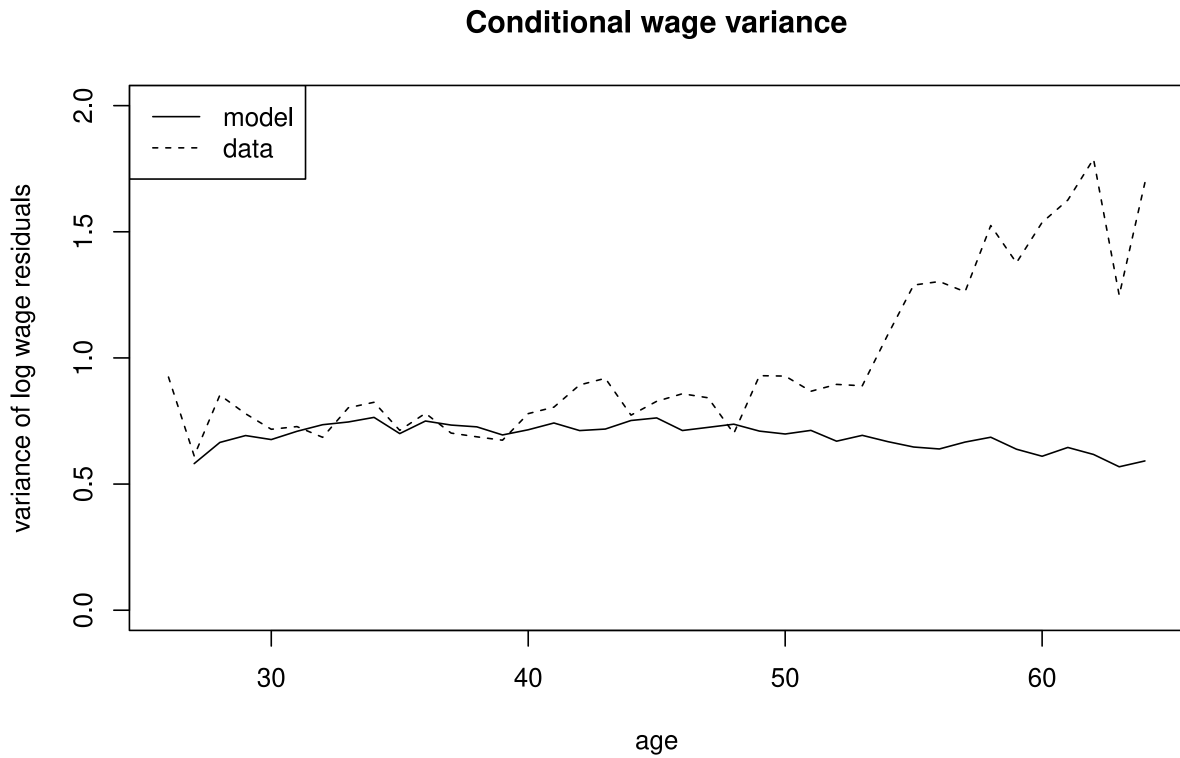
Figure A.3 shows the estimated and simulated health transitions for over 50s. As with under 51s, I am broadly successful in capturing patterns of transitions. Persistence of health state plays a key role in determining the probability of recovery. Figures A.4 and A.5 show the behavior of log wage residuals. I capture the broad magnitude of log wage persistence out to 10 periods, although imposing an AR(1) process means that I do not fully capture how the decay in the correlation changes across the curve. I am broadly successful at capturing the variance in log wages in the early part of the life cycle, but I cannot match the increase in the variance of wages. This result is likely because my model imposes a constant variance on log wages across the life cycle.

Figure A.4: Estimated versus simulated log wages persistence



Notes: estimated log wage residual persistence conditional on working. Residuals are estimated by regressing log wages on age, age squared and health status. I do this in both the data and simulations. Simulated moments are weighted by education and age to be comparable with survey data.

Figure A.5: Estimated versus simulated variance in log wages



Notes: estimated variance of the log wage residual persistence conditional on working. Simulated moments are weighted by education and age to be comparable with survey data. The note to Figure A.4 explained how the residuals are derived.

Appendix B

The causal effects of workplace autonomy on mental health

B.1 Mental health variable questions and scoring

The SF12 questions are as follows:

1. In general, would you say your health is: 1) excellent 2) very good 3) good 4) fair 5) or poor?
2. Does your health now limit you in performing moderate activities? If so, how much? 1) Yes, limited a lot 2) Yes, limited a little 3) No, not limited at all
3. Does your health limit climbing several flights of stairs? 1) Yes, limited a lot 2) Yes, limited a little 3) No, not limited at all
4. During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health? Accomplished less than you would like: 1) All of the time 2) Most of the time 3) Some of the time 4) A little of the time 5) None of the time
5. During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of your physical health? Were limited in the kind of work or other activities: 1) All of the time 2) Most of the time 3) Some of the time 4) A little of the time 5) None of the time
6. During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)? Accomplished less than you would like: 1) All of the time 2) Most of the time 3) Some of the time 4) A little of the time 5) None of the time

7. During the past 4 weeks, how much of the time have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)? Did work or other activities less: 1) All of the time 2) Most of the time 3) Some of the time 4) A little of the time 5) None of the time carefully than usual
8. During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)? 1) Not at all 2) A little bit 3) Moderately 4) Quite a bit 5) Extremely
9. How much of the time during the past 4 weeks have you felt calm and peaceful? 1) All of the time 2) Most of the time 3) Some of the time 4) A little of the time 5) None of the time
10. How much of the time during the past 4 weeks did you have a lot of energy? 1) All of the time 2) Most of the time 3) Some of the time 4) A little of the time 5) None of the time
11. How much of the time during the past 4 weeks have you felt downhearted and depressed? 1) All of the time 2) Most of the time 3) Some of the time 4) A little of the time 5) None of the time
12. During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives)? 1) All of the time 2) Most of the time 3) Some of the time 4) A little of the time 5) None of the time

Table B.1 shows how the answers to these questions are scored to calculate the SF12 indices for physical health and mental health. Each answer is associated with a different score. Note that there is an omitted category associated with each answer which scores zero, and an intercept for each index. Initial construction of the scoring system is from a regression of a person's score on the longer 36-item index on a full set of dummies for responses to the 12-item questionnaire (Ware Jr et al., 1996). The coefficients from Ware Jr, Kosinski, and Keller's (1996) regression are reproduced in Table B.11. However, the Understanding Society data set produces the final index as a derived variable, which is the variable I use.

The questions used to construct the GHQ12 variables are listed below. Each question has one of four answers of varying severity. If the question is posed so that affirmative answers indicate good mental health, the answers are one of the following: "Better than usual", "Same as usual", "Less than usual", "Much less than usual". If the question is posed so that affirmative answers indicate poor mental health, the answers are one of the following: "Not at all", "No more than usual", "Rather more than usual", "Much more than usual". For each question, a respondent is considered to have a symptom of mental illness if they respond with the least healthy or second least healthy answer (for example, "Less than usual" or "Much less than usual" if the question is phrased so that affirmative answers indicate good health and "Rather more than usual" or "Much more than usual" if the question is phrased so that affirmative answers indicate bad health). The GHQ12 caseness score uses all 12 questions, and is the sum of all symptoms for which a person reports having.

1. Have you recently been able to concentrate on whatever you're doing?

Table B.1: Scoring system for SF12 indices

Question	Mental health score	Physical health score
Intercept	60.8	4.347
Moderate activities limited a lot	3.93	-7.23
Moderate activities limited a little	1.87	-3.46
Climbing several flights of stairs limited a lot	2.68	-6.24
Climbing several flights of stairs limited a little	1.43	-2.74
Accomplish less than you would like	1.44	-4.62
Limited in the kind of activities	1.67	-5.52
Pain interferes with normal work extremely	1.47	-11.26
Pain interferes with normal work quite a bit	1.77	-8.38
Pain interferes with normal work moderately	1.49	-6.51
Pain interferes with normal work a little bit	0.90	-3.8
Health in general is poor	-1.71	-10.76
Health in general is fair	-0.17	-8.070
Health in general is good	0.03	-4.304
Health in general is very good	-0.6	-1.614
Have a lot of energy none of the time	-6.02	-2.45
Have a lot of energy a little of the time	-4.89	-2.02
Have a lot of energy some of the time	-3.30	-1.162
Have a lot of energy a good bit of the time	-1.65	-1.14
Have a lot of energy most of the time	-0.92	-0.423
Health interferes with social activities all of the time	-6.30	-0.34
Health interferes with social activities most of the time	-8.26	-0.94
Health interferes with social activities some of the time	-5.63	-0.18
Health interferes with social activities a little of the time	-3.14	0.11
Accomplish less than you would like	-6.83	3.04
Didn't do activities as carefully as usual	-5.70	2.32
Felt calm and peaceful none of the time	-10.19	3.47
Felt calm and peaceful a little of the time	-7.93	2.90
Felt calm and peaceful some of the time	-6.31	2.37
Felt calm and peaceful a good bit of the time	-4.10	1.37
Felt calm and peaceful most of the time	-1.95	0.67
Felt downhearted and blue all of the time	-16.15	3.61
Felt downhearted and blue most of the time	-10.78	3.42
Felt downhearted and blue a good bit of the time	-8.10	2.34
Felt downhearted and blue some of the time	-4.59	1.28
Felt downhearted and blue a little of the time	-1.96	0.41

Notes: These are the scores given to each possible answer to the short-form survey used in calculating the SF12 health measures. The SF12 indices are calculated by summing all the numbers corresponding to a person's answers (and the intercept). The index can range between 0 and 100.

2. Have you recently lost much sleep over worry?
3. Have you recently felt that you were 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 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?

B.2 Robustness to different summary measures of work-related autonomy

Tables B.2 and B.3 report regression results for different summary measures of overall low autonomy.

Table B.2: Robustness of the Causal Effects of Low Overall Autonomy on the Mental Health of Men to Different Cutoffs

	Panel A: effect on GHQ12 caseness							
Dimensions of low autonomy (≥ 1)	0.237***	0.266***						
	(0.047)	(0.053)						
Dimensions of low autonomy ($\geq$2)			0.448***	0.452***				
			(0.057)	(0.056)				
Dimensions of low autonomy (≥ 4)					0.539***	0.504***		
					(0.087)	(0.076)		
Dimensions of low autonomy ($\geq$5)							0.497***	0.405***
							(0.112)	(0.094)
Observations	13,847	13,847	13,847	13,847	13,847	13,847	13,847	13,847
	Panel B: effect on SF12							
Dimensions of low autonomy ($\geq$1)	−1.227***	−0.767***						
	(0.177)	(0.163)						
Dimensions of low autonomy ($\geq$2)			−1.769***	−1.184***				
			(0.207)	(0.172)				
Dimensions of low autonomy ($\geq$4)					−1.942***	−1.251***		
					(0.300)	(0.236)		
Dimensions of low autonomy ($\geq$5)							−1.836***	−0.950***
							(0.392)	(0.294)
Observations	13,892	13,892	13,892	13,892	13,892	13,892	13,892	13,892
Wave	Y	Y	Y	Y	Y	Y	Y	Y
Person fixed effects	N	Y	N	Y	N	Y	N	Y
Other controls	N	Y	N	Y	N	Y	N	Y

Notes: This table shows the results of regressing measures of mental health on measures of low autonomy as measured by different cutoff values for my summary variable amongst men. For each regression, the explanatory variable is a dummy variable indicating whether a person has low autonomy across a greater number of dimensions than a given number. The GHQ12 caseness and SF12 are defined in the note to Table 3.3. Standard errors are clustered at the individual level. Other controls are defined in the note to Table 2.3. *p<0.1; **p<0.05; ***p<0.01

Table B.3: Robustness of the causal effects of low overall autonomy on the Mental Health of Women to Different Cutoffs

	Panel A: effect on GHQ12 caseness							
Dimensions of low autonomy (≥1)	0.289***	0.335***						
	(0.055)	(0.059)						
Dimensions of low autonomy (≥2)			0.460***	0.388***				
			(0.059)	(0.057)				
Dimensions of low autonomy (≥4)					0.537***	0.515***		
					(0.088)	(0.075)		
Dimensions of low autonomy (≥5)							0.494***	0.507***
							(0.096)	(0.092)
Observations	15,708	15,708	15,708	15,708	15,708	15,708	15,708	15,708
	Panel B: effect on SF12							
Dimensions of low autonomy (≥1)	−0.863***	−0.802***						
	(0.185)	(0.172)						
Dimensions of low autonomy (≥2)			−1.294***	−1.074***				
			(0.189)	(0.165)				
Dimensions of low autonomy (≥4)					−1.210***	−1.151***		
					(0.272)	(0.219)		
Dimensions of low autonomy (≥5)							−1.081***	−1.076***
							(0.342)	(0.269)
Observations	15,692	15,692	15,692	15,692	15,692	15,692	15,692	15,692
Wave	Y	Y	Y	Y	Y	Y	Y	Y
Person fixed effects	N	Y	N	Y	N	Y	N	Y
Other controls	N	Y	N	Y	N	Y	N	Y

Notes: This table shows the results of regressing measures of mental health on measures of low autonomy as measured by different cutoff values for my summary variable amongst women. For each regression, the explanatory variable is a dummy variable indicating whether a person has low autonomy across a greater number of dimensions than a given number. The GHQ12 caseness and SF12 are defined in the note to Table 3.3. Standard errors are clustered at the individual level. Other controls are defined in the note to Table 2.3. *p<0.1; **p<0.05; ***p<0.01

B.3 Robustness to inclusion of respondents who are not consistently in work

As a robustness test, I re-run my analysis including observations who are not in work in every wave they are observed. The rationale for excluding these workers from my main sample is that including them introduces a selection problem. For these workers, their work-related autonomy is only observed in some periods, and employment and mental health may be jointly determined.

Including these respondents increases the total sample size to 69,824 person-wave observations of 22,945 unique individuals. I estimate the same set of regressions as for the main sample. Results are presented in Table B.4. Point estimates are generally similar to those calculated from the main sample, but controls have a larger effect on the estimate. Results are broadly robust, in that low overall autonomy is always associated with worse mental health; low autonomy over job tasks causes worse mental health in nearly all specifications; and low autonomy over work hours causes worse mental health for women in my preferred specifications.

There are two likely reasons for the greater effect of controls in this sample. Firstly, workers who experience spells of non-work are observably and unobservably different in their mental health, the types of jobs they do, and variables such as education. This variation generates additional associations between mental health and work-related autonomy which can be explained with controls. One implication is that my results in the main text which refer specifically to workers who are consistently in work likely do not generalize to all workers. Secondly, the average number of observations per individual in this sample is 2.11, versus 3.6 in the main sample. As a result, controls such as person fixed effects mechanically absorb larger portions of the variation in the dependent variable.

B.4 Trajectory of mental health about changes in work-related autonomy

In this section, I explore how mental health evolves around changes in work-related autonomy. To do so, I limit my sample to just those individuals who experience at exactly one transition in work-related autonomy, either from high autonomy to low autonomy or low autonomy to high autonomy, and estimate the conditional change in mental health symptoms around that change. Specifically, I estimate the following regressions:

$$\text{mental_health}_{it} = \Sigma_{t \neq -2} \beta_\tau \mathbb{1}(t - T_i = \tau) + X'_{it}\gamma + \eta_t + \mu_i + \epsilon_{it} \tag{B.1}$$

Here X_{it} include all of the controls used in the most saturated model except lagged mental health (occupation, education, marital status, having children and age). T_i is the wave of the survey in which person i is first "treated" (i.e., either working in a low autonomy job if this is the treatment, or working in a high autonomy job is this is the treatment). η_t and μ_i are wave and person fixed effects respectively. The β_τ coefficients report how the worker's mental health changes relative to the period when they first

Table B.4: Robustness of the causal effect of low autonomy on the mental health of men using the full sample of workers

	Panel A: effect on GHQ12 caseness					
Low autonomy:						
overall	0.457***	0.326***	0.324***			
	(0.047)	(0.051)	(0.051)			
over job tasks				0.155***	0.131**	0.119**
				(0.051)	(0.051)	(0.051)
over work pace				0.409***	0.240***	0.236***
				(0.054)	(0.051)	(0.051)
over work manner				0.023	0.098	0.099
				(0.069)	(0.065)	(0.065)
over task order				0.186***	0.141**	0.153**
				(0.064)	(0.061)	(0.061)
over work hours				−0.063*	−0.016	0.001
				(0.036)	(0.044)	(0.044)
Observations	27,564	27,564	27,564	27,564	27,564	27,564
R^2	0.005	0.605	0.645	0.008	0.606	0.646
	Panel B: effect on SF12					
Low autonomy:						
overall	−1.975***	−1.085***	−1.156***			
	(0.167)	(0.155)	(0.155)			
over job tasks				−0.375**	−0.200	−0.229
				(0.171)	(0.155)	(0.154)
over work pace				−1.684***	−0.905***	−0.897***
				(0.181)	(0.154)	(0.153)
over work manner				0.032	−0.136	−0.167
				(0.232)	(0.197)	(0.196)
over task order				−0.892***	−0.675***	−0.713***
				(0.214)	(0.184)	(0.184)
over work hours				−0.151	0.104	0.044
				(0.136)	(0.133)	(0.133)
Observations	27,636	27,636	27,636	27,636	27,636	27,636
R^2	0.014	0.664	0.675	0.020	0.665	0.676
Wave	Y	Y	Y	Y	Y	Y
Person FE	N	Y	Y	N	Y	Y
Other controls	N	N	Y	N	N	Y

Notes: I re-estimate my main results with a larger sample, including all observations of workers, instead of only those who are continuously in work. See notes to Table 2.3 for descriptions of the regression and key variables. *p<0.1; **p<0.05; ***p<0.01

Table B.5: Robustness of the causal effect of low autonomy on the mental health of women using the full sample of workers

	Panel A: effect on GHQ12 caseness					
Low autonomy:						
overall	0.525^{***}	0.322^{***}	0.333^{***}			
	(0.047)	(0.048)	(0.048)			
over job tasks				0.266^{***}	0.271^{***}	0.274^{***}
				(0.050)	(0.050)	(0.049)
over work pace				0.265^{***}	0.154^{***}	0.138^{***}
				(0.051)	(0.051)	(0.051)
over work manner				0.187^{***}	0.023	0.021
				(0.065)	(0.062)	(0.062)
over task order				-0.036	0.048	0.078
				(0.061)	(0.060)	(0.060)
over work hours				0.055	0.093^{**}	0.106^{**}
				(0.039)	(0.045)	(0.045)
Observations	33,734	33,734	33,734	33,734	33,734	33,734
R^2	0.006	0.576	0.586	0.008	0.577	0.587
	Panel B: effect on SF12					
Low autonomy:						
overall	-1.620^{***}	-0.844^{***}	-0.887^{***}			
	(0.149)	(0.138)	(0.138)			
over job tasks				-0.898^{***}	-0.659^{***}	-0.673^{***}
				(0.158)	(0.143)	(0.142)
over work pace				-0.914^{***}	-0.538^{***}	-0.503^{***}
				(0.160)	(0.147)	(0.146)
over work manner				-0.324	0.042	0.033
				(0.204)	(0.179)	(0.178)
over task order				0.202	0.133	0.075
				(0.198)	(0.174)	(0.173)
over work hours				-0.359^{***}	-0.446^{***}	-0.512^{***}
				(0.130)	(0.130)	(0.130)
Observations	33,723	33,723	33,723	33,723	33,723	33,723
R^2	0.013	0.689	0.715	0.015	0.690	0.716
Wave	Y	Y	Y	Y	Y	Y
Person FE	N	Y	Y	N	Y	Y
Other controls	N	N	Y	N	N	Y

Notes: I re-estimate my main results with a larger sample, including all observations of workers, instead of only those who are continuously in work. See notes to Table 2.3 for descriptions of the regression and key variables. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

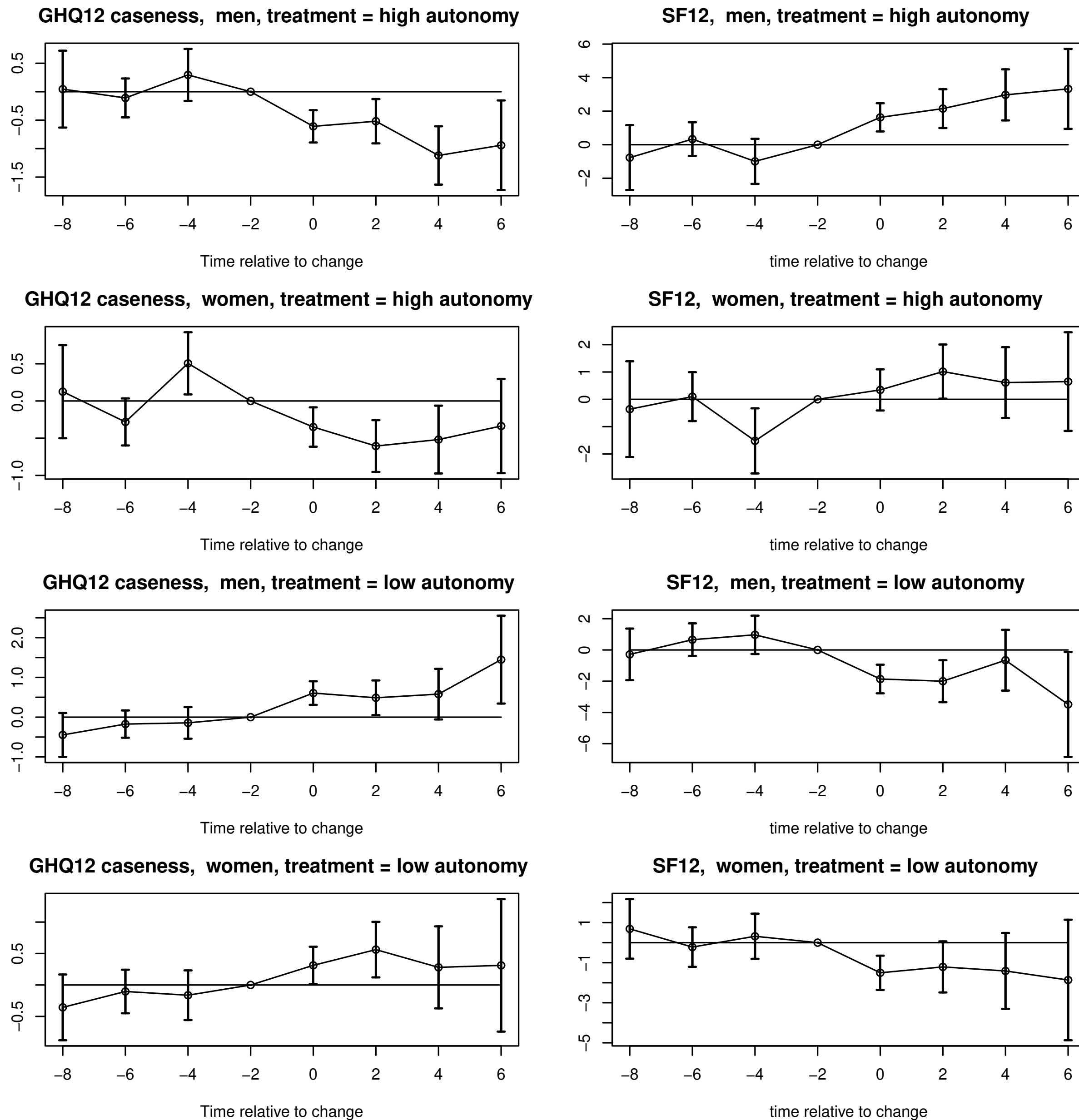
report being treated, holding constant my control variables. I plot β_τ, and their 95% confidence intervals, in Figure B.1.

In principle, the timing of the transition from high autonomy to low autonomy work and vice versa and the change in mental health may be informative about the causal connection between the two. If the majority of the change in mental health happens after the change in working conditions, this is consistent with a causal effect of working conditions on mental health; if the majority of the change in mental health happens before the change in working conditions, it seems more likely that the association reflects sorting based on mental health (Boyce & Oswald, 2012).

In general, the event studies suggest that the majority of the change in mental health occurs in the same two year time period as the change in work-related autonomy. Men who transition from low autonomy to high autonomy experience an improvement in their mental health in the first period in which they are observed working in a high autonomy job, and men who transition from high autonomy jobs to low autonomy jobs experience worse mental health in the first period in which they are observed working in a low autonomy job. There is little evidence of pre-trends for any of the event studies of men. For women, while their mental health improves at the time they transition from low autonomy to high autonomy work, there is also evidence of an improvement in mental health in the four years leading to this transition. However, for women transitioning to low autonomy work the pre-trends are small an insignificant. For the most part these results are consistent with work-related autonomy having an effect on mental health rather than the other way around.

Some caution is required in drawing firm conclusions from this exercise. My event studies cannot speak to the ordering of the changes in work-related autonomy and mental health within the two-year window between observations. They are, however, mostly consistent with the causal effect I argue for in the main text.

Figure B.1: The evolution of mental health around a transition between high and low autonomy work



Notes: I estimate the path of mental health about a transition from high autonomy to low autonomy work or low autonomy work to high autonomy work as in equation B.1. I plot the coefficients β_τ which describe average mental health compared to the first period in which a person is treated, conditional on covariates. The sample is all people who ever experience the relevant transition (either from low autonomy work to high autonomy work, or from high autonomy work to low autonomy work), and transition at most once. I define overall low autonomy as being equal to 1 if a person scores 1 (low) on 3 or more aspects of workplace autonomy. The GHQ12 caseness is the number of negative mental health symptoms a person has. The SF12 index scores the mental health of a person between 0 (worst) and 100 (best) based on their answers to a health survey.

Appendix C

The effect of retirement eligibility on mental health in the United Kingdom: heterogeneous effects by occupation

C.1 GHQ12 questions and sub-indices

The GHQ12 is a set of 12 questions about mental health symptoms, each of which has four possible answers reflecting different levels of severity. If the question is posed so that affirmative answers indicate good mental health, the answers are one of the following: "Better than usual", "Same as usual", "Less than usual", "Much less than usual". If the question is posed so that affirmative answers indicate poor mental health, the answers are one of the following: "Not at all", "No more than usual", "Rather more than usual", "Much more than usual". To construct the caseness, I first convert each question into a binary variable: a person scores 0 if their symptoms are either the best or second best possible, and 1 if their symptoms are either the worst or second worst possible. I aggregate these variables into the GHQ12 caseness score. The sub-indices are Loss of Confidence, Anxiety and Depression and Social Dysfunction. In calculating the sub-indices, one treats the individual variables as cardinal measures of mental health symptoms, scored from 1 to 4, and sum over all the symptoms of a particular type

The questions used to construct the GHQ12 variables are listed below. The sub-index a question belongs to is in parentheses after each question.

1. Have you recently been able to concentrate on whatever you're doing? (Social Dysfunction)
2. Have you recently lost much sleep over worry? (Anxiety and Depression)

3. Have you recently felt that you were playing a useful part in things? (Social Dysfunction)
4. Have you recently felt capable of making decisions about things? (Social Dysfunction)
5. Have you recently felt constantly under strain? (Anxiety and Depression)
6. Have you recently felt you couldn't overcome your difficulties? (Anxiety and Depression)
7. Have you recently been able to enjoy your normal day-to-day activities? (Social Dysfunction)
8. Have you recently been able to face up to problems? (Social Dysfunction)
9. Have you recently been feeling unhappy or depressed? (Anxiety and Depression)
10. Have you recently been losing confidence in yourself? (Loss of Confidence)
11. Have you recently been thinking of yourself as a worthless person? (Loss of Confidence)
12. Have you recently been feeling reasonably happy, all things considered? (Social Dysfunction)

The GHQ12 has been used to screen for psychological illness in a variety of contexts (e.g., Gureje and Obikoya, 1990, Anjara et al., 2020). There is also evidence that screening using sub-indices can help identify the risk of different psychiatric illnesses (Gelaye et al., 2015). Both the caseness score and sub-indices have been used as outcome measures in various economics studies (e.g., Gathergood, 2013, Belloni et al., 2022).

C.2 Institutional details on women's retirement ages

For women born before April 1950, the SPA is 60. The Pension Acts of 1995 and 2011 increased the retirement eligibility age for women born after this date. The increase was phased in over time so that the later a woman is born, the older she is on the date she reaches SPA eligibility, until the retirement age is equalized between men and women. For women, there is therefore a mapping between date of birth and date of reaching SPA. The retirement dates for UK women by date of birth are given in Table C.1.[1]

For each woman in my sample, I construct her expected months to retirement by taking a weighted average of retirement dates she could face and subtracting the current date. Dong (2015) shows that measurement error in the running variable does not lead to bias in RDD results if a local linear regression is used and the measurement error is mean zero. Additionally, following Dong's recommendation, I drop observations where it cannot be determined which side of the cutoff they fall. Note that all such observations are of women, because all of the men in my sample reach retirement age at the age of 65, and the survey reports age at last birthday.

[1]https://www.gov.uk/government/publications/state-pension-age-timetable/state-pension-age-timetable

Table C.1: Women's retirement date by date of birth.

Date of birth	Date State Pension Age reached
6 April 1950 – 5 May 1950	6 May 2010
6 May 1950 – 5 June 1950	6 July 2010
6 June 1950 – 5 July 1950	6 September 2010
6 July 1950 – 5 August 1950	6 November 2010
6 August 1950 – 5 September 1950	6 January 2011
6 September 1950 – 5 October 1950	6 March 2011
6 October 1950 – 5 November 1950	6 May 2011
6 November 1950 – 5 December 1950	6 July 2011
6 December 1950 – 5 January 1951	6 September 2011
6 January 1951 – 5 February 1951	6 November 2011
6 February 1951 – 5 March 1951	6 January 2012
6 March 1951 – 5 April 1951	6 March 2012
6 April 1951 – 5 May 1951	6 May 2012
6 May 1951 – 5 June 1951	6 July 2012
6 June 1951 – 5 July 1951	6 September 2012
6 July 1951 – 5 August 1951	6 November 2012
6 August 1951 – 5 September 1951	6 January 2013
6 September 1951 – 5 October 1951	6 March 2013
6 October 1951 – 5 November 1951	6 May 2013
6 November 1951 – 5 December 1951	6 July 2013
6 December 1951 – 5 January 1952	6 September 2013
6 January 1952 – 5 February 1952	6 November 2013
6 February 1952 – 5 March 1952	6 January 2014
6 March 1952 – 5 April 1952	6 March 2014
6 April 1952 – 5 May 1952	6 May 2014
6 May 1952 – 5 June 1952	6 July 2014
6 June 1952 – 5 July 1952	6 September 2014
6 July 1952 – 5 August 1952	6 November 2014
6 August 1952 – 5 September 1952	6 January 2015
6 September 1952 – 5 October 1952	6 March 2015
6 October 1952 – 5 November 1952	6 May 2015
6 November 1952 – 5 December 1952	6 July 2015
6 December 1952 – 5 January 1953	6 September 2015

Notes: Data show the date of state pension eligibility by date of birth for women in the UK. Source: Department for Work and Pensions. Women born before 6 April 1950 reach retirement eligibility aged 60. Women born after 5 December 1953 reach retirement eligibility aged 65.

Table C.1: Women's retirement date by date of birth.

Date of birth	Date State Pension age reached
6 January 1953 – 5 February 1953	6 November 2015
6 February 1953 – 5 March 1953	6 January 2016
6 March 1953 – 5 April 1953	6 March 2016
6 April 1953 – 5 May 1953	6 July 2016
6 May 1953 – 5 June 1953	6 November 2016
6 June 1953 – 5 July 1953	6 March 2017
6 July 1953 – 5 August 1953	6 July 2017
6 August 1953 – 5 September 1953	6 November 2017
6 September 1953 – 5 October 1953	6 March 2018
6 October 1953 – 5 November 1953	6 July 2018
6 November 1953 – 5 December 1953	6 November 2018

Notes: Data show the date of state pension eligibility by date of birth for women in the UK. Source: Department for Work and Pensions. Women born before 6 April 1950 reach retirement eligibility aged 60. Women born after 5 December 1953 reach retirement eligibility aged 65.

C.3 Sensitivity of all-sample estimates to changes in specification

Figures C.1 to C.4 display how the causal estimates of the effect of being eligible to retire on mental health outcomes differ across specifications. For example, in Figure C.1, the top panel shows the estimated effect of gaining eligibility to retire on the GHQ12 caseness. Each point is an estimated causal effect of retirement eligibility on the GHQ12 caseness (and confidence interval) for a given set of estimation choices. The three panels below show the bandwidth used in estimation, whether a quadratic term in the running variable is included, and which type of kernel used.

Across almost all specifications, being eligible to retire has a beneficial effect on mental health as measured by GHQ12 caseness, symptoms of Anxiety and Depression, and Social Dysfunction. The estimates are remarkably stable, and the only specifications in which they are not statistically significant are when a small bandwidth and quadratic specification are used.

Using a quadratic specification with measurement error in the running variable biases the estimated effect.[2] However, in most of my estimates, the size of the bias is small. It depends on the variance of the

[2]Dong (2015) shows that with a second-order polynomial estimation of RDD and measurement error in the running variable, the true causal effect is given by:

$$\tau = \beta_1 - \mu_1 \delta_b + (2\mu_1^2 - \mu_2)\delta_d \qquad \text{(C.1)}$$

where β_1, and δ_b have the same interpretation as in Equation 3.3, δ_d is the coefficient on the squared term of the polynomial, and μ_i is the ith moment of the measurement error in the running variable. This equation implies that if the running variable is observed with mean-zero error and a squared term is not included, measurement error in the running variable does not bias the RDD estimate.

measurement error, and the magnitude of the estimated curvature of the trend in potential outcomes for those who are treated. Given my assumptions about the error in months to retirement being uniformly distributed, this variance is always less than $\frac{1}{12}$. The estimated curvature tends to be small. For example, when a 5-year bandwidth and triangular kernal are used to estimate the effect on the overall caseness, the coefficient on the squared term of those above the cutoff is 0.000004631, giving a total bias smaller than $0.000004631 \times \frac{1}{12} = 0.00000386$.

C.4 Sensitivity of estimates by occupation type to changes in specification

This section discusses and Figures C.5 to C.8 show how the effect of retirement eligibility on mental health by occupational cluster depends on estimation choices. For example, in Figure C.5, the top three panels show the estimated effect of gaining eligibility to retire from each of the three occupational clusters on the GHQ12 caseness. Each point is therefore an estimated causal effect of retirement eligibility on the GHQ12 caseness (and confidence interval) for a given set of estimation choices. The three panels below show the bandwidth used in estimation, whether a quadratic term in the running variable is included, and which type of kernel used.

The effect of retirement eligibility on the GHQ12 caseness index is greatest in cluster three. Figure C.5 shows how this effect varies by estimation choices. Being eligible to retire has a beneficial effect on the GHQ12 caseness measure amongst those in cluster three in every specification, and a beneficial effect amongst those in cluster one in most specifications (the exceptions are at small bandwidths). However, some estimates do not attain statistical significance. Including a quadratic term in the specification pushes estimates towards zero.

Figure C.6 shows the range of estimates by specification for the causal effect on symptoms of Anxiety and Depression. Again, the beneficial effect of retirement eligibility is evident in all specifications for those in cluster three, and most specifications for cluster one. Again, the inclusion of a quadratic term is most important in changing the results.

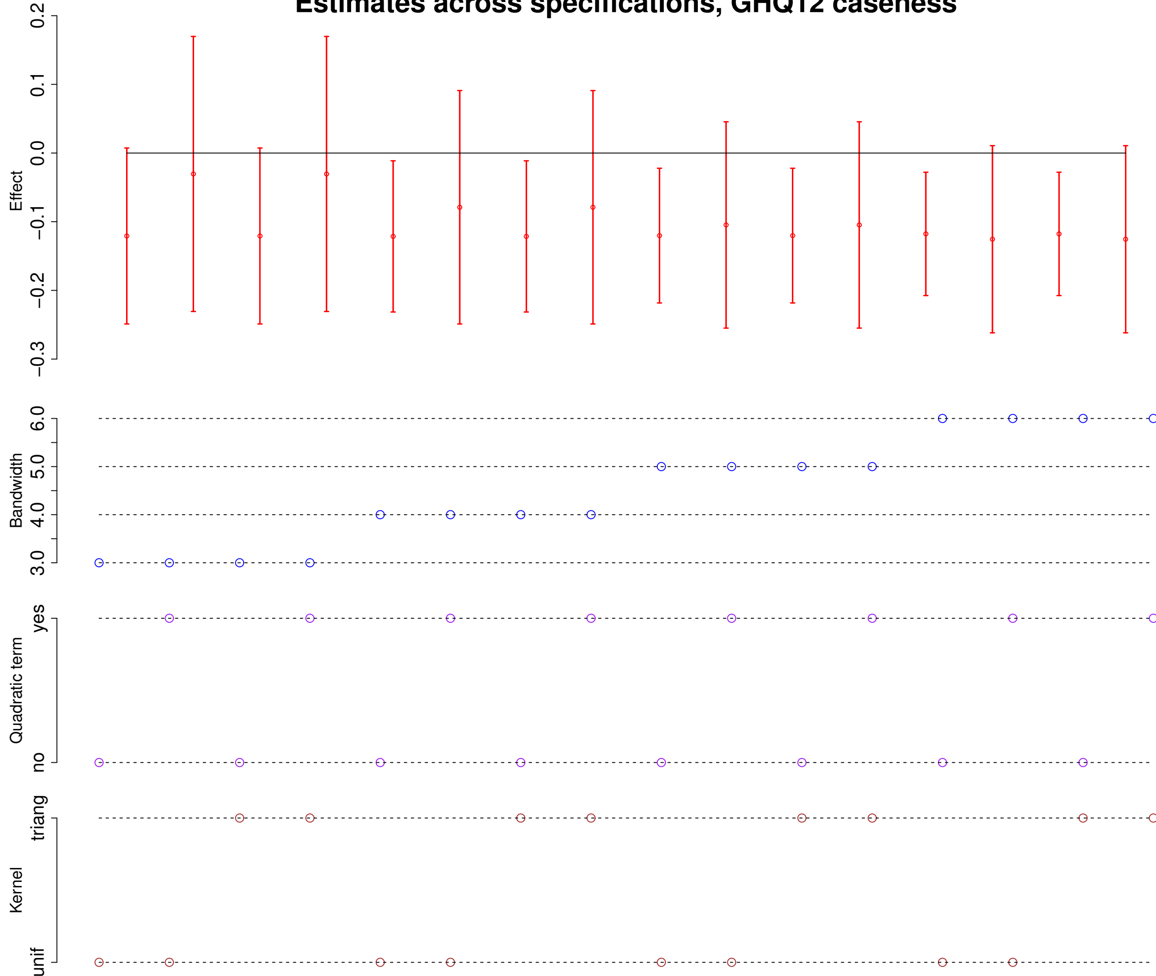
In Figure C.7, I show the pattern of estimates for symptoms of Loss of Confidence. There are no results which are statistically significant. Results are also close to zero in most specifications.

In my preferred specification (Equations 3.3 and 3.4 with a local linear regression, 5-year bandwidth, and triangular kernel), there is a beneficial effect of being eligible to retire on symptoms of Social Dysfunction for those in cluster three. This result is relatively consistent across specifications (see Figure C.8). In most specifications, the estimated effect is statistically significant.

C.5 Standard error estimates

Table C.2 shows how the standard error changes with different estimators for the specification which investigates the average effect of retirement eligibility. For the most part, standard errors are similar when

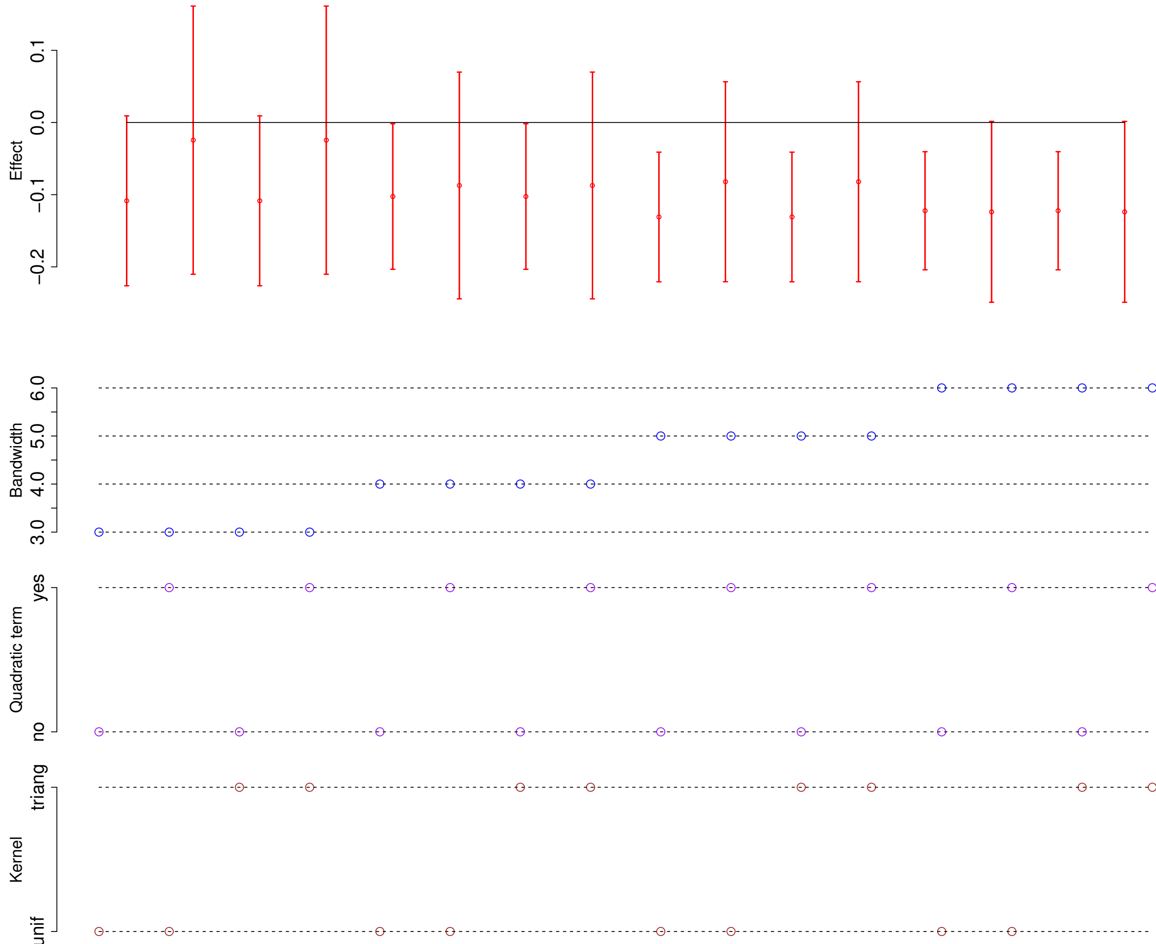
Figure C.1: Effect of retirement eligibility on GHQ12 caseness across specifications



Notes: Results are the average short-run causal effect of being eligible to retire on the GHQ12 caseness. I estimate the causal effect using the discontinuity in the GHQ12 caseness as a person become eligible for retirement. I use a bandwidth of between 3 and 6 years. I use a triangular or uniform kernel. I use a local linear or local quadratic estimator. I calculate the standard error of the causal effect by using the heteroskedasticity-robust variance-covariance matrix clustered by individual. Variables are defined in the note to Table 3.3.

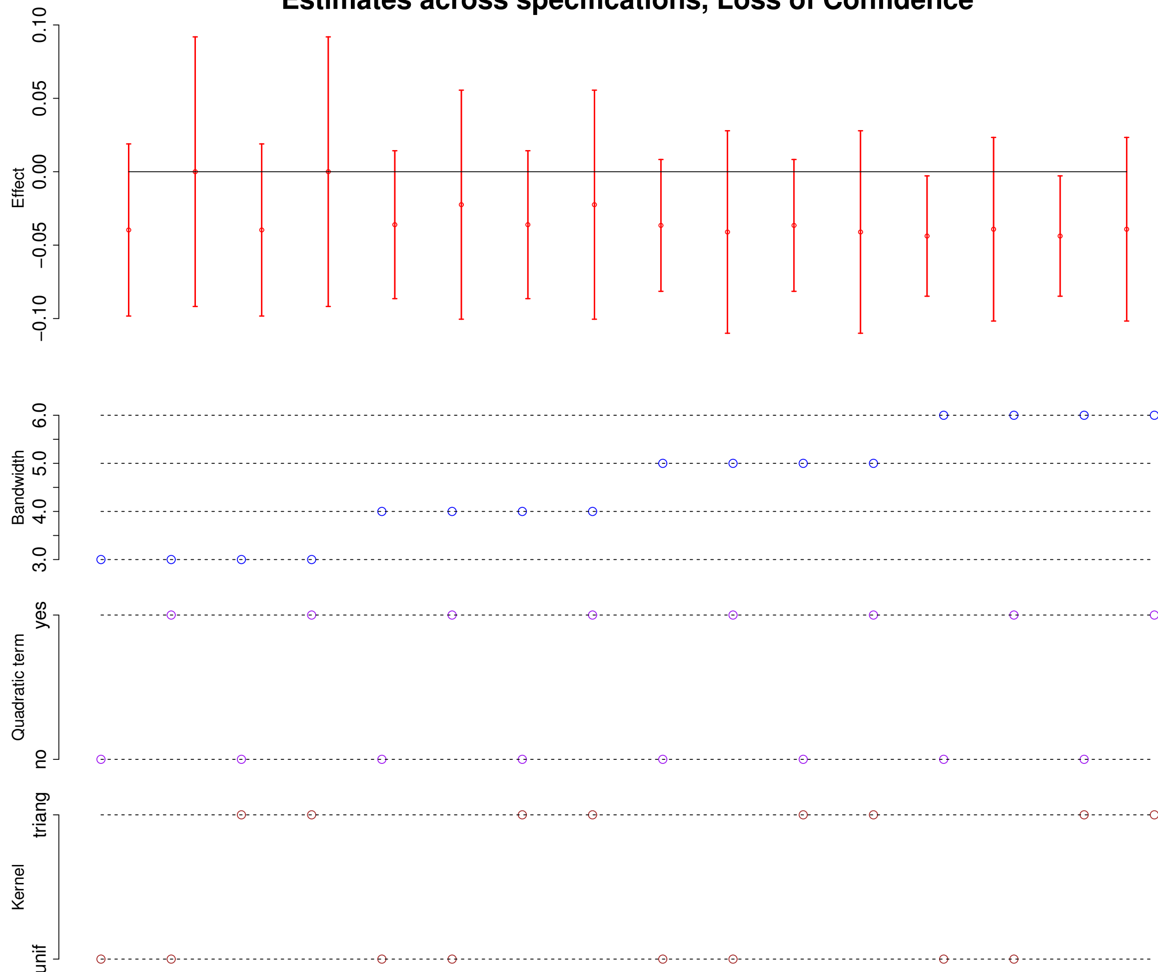
Figure C.2: Effect of retirement eligibility on Anxiety and Depression across specifications

Estimates across specifications, Anxiety and Depression



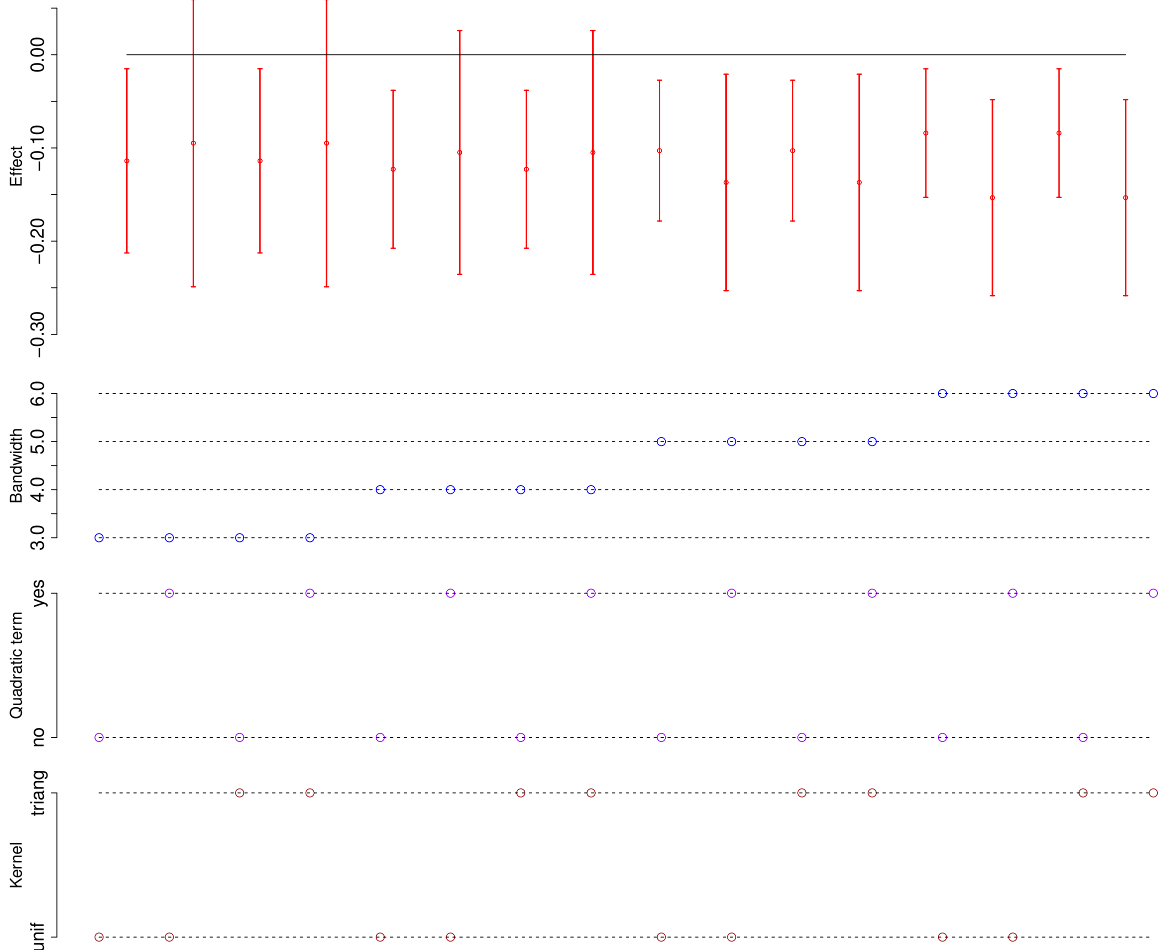
Notes: Results are the average short-run causal effect of being eligible on symptoms of Anxiety and Depression to retire in the population. The range of estimation specifications is described in the note to Figure C.3. Variables are defined in the note to Table 3.3.

Figure C.3: Effect of retirement eligibility on Loss of Confidence across specifications



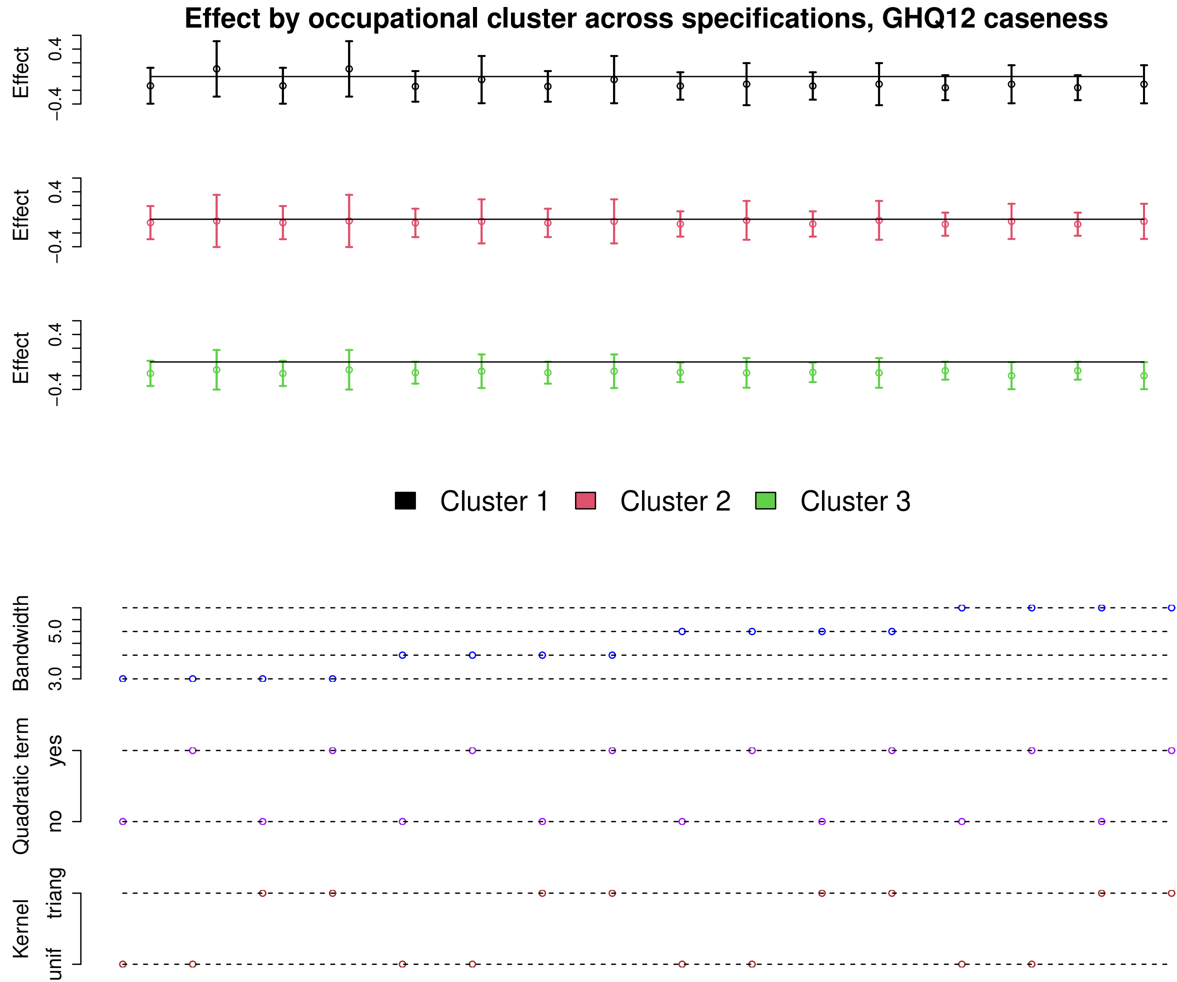
Notes: Results are the average short-run causal effect of being eligible to retire on symptoms of Loss of Confidence. The range of estimation specifications is described in the note to Figure C.3. Variables are defined in the note to Table 3.3.

Figure C.4: Effect of retirement eligibility on Social Dysfunction across specifications



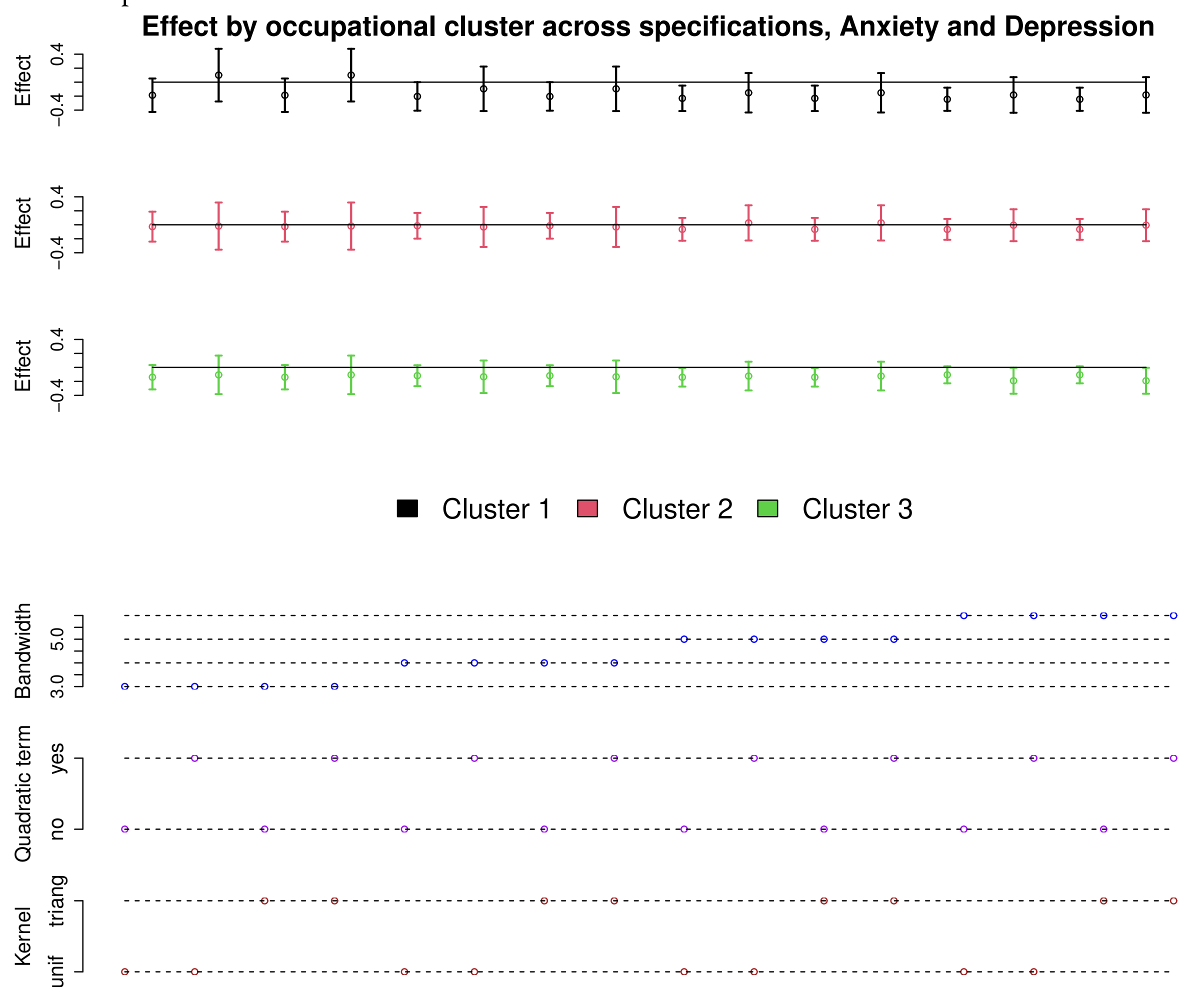
Notes: Results are the average short-run causal effect of being eligible to retire on symptoms of Social Dysfunction. The range of estimation specifications is described in the note to Figure C.3. Variables are defined in the note to Table 3.3.

Figure C.5: Effect of retirement eligibility on GHQ12 caseness across occupational clusters across different specifications



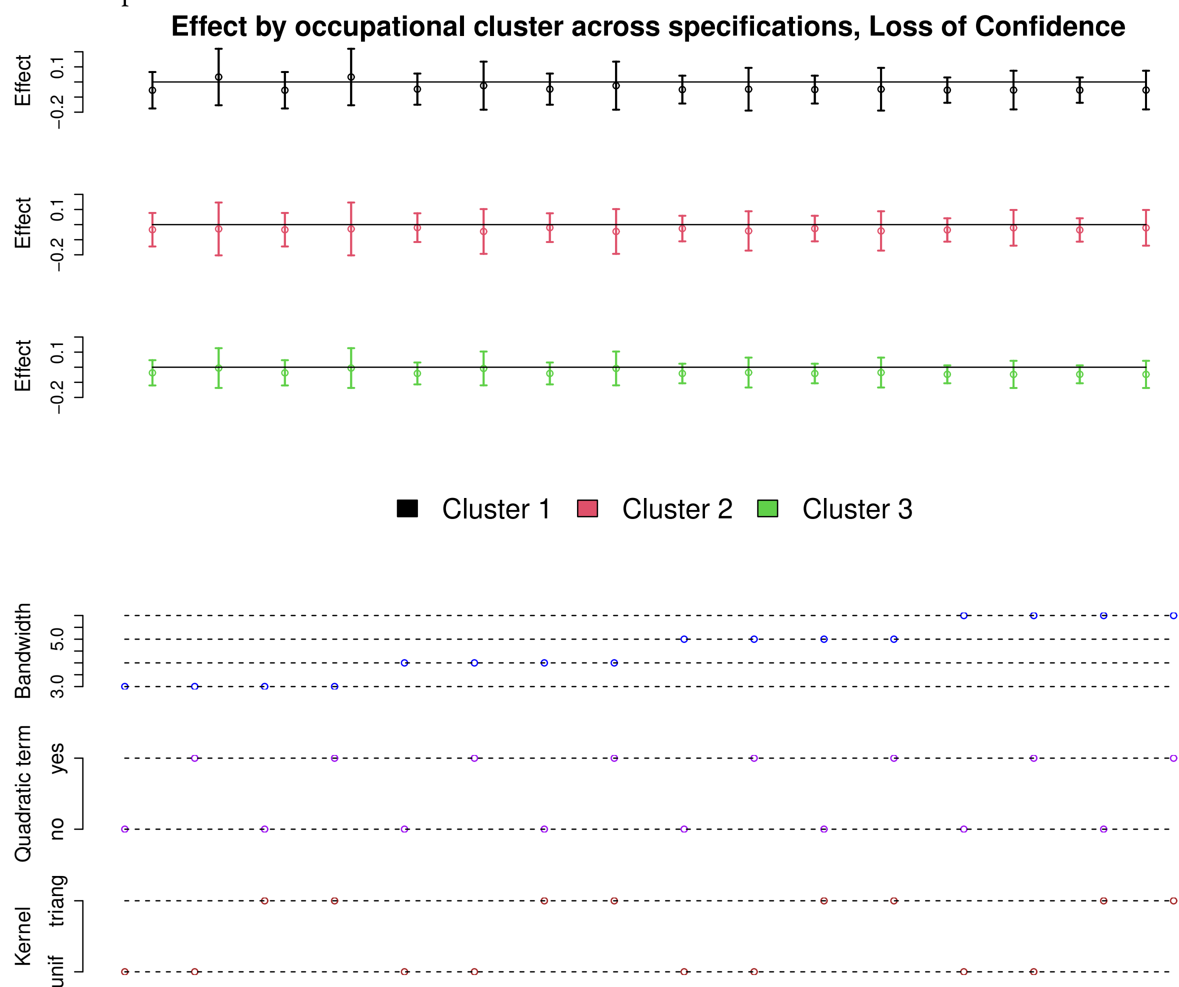
Notes: Results are obtained using a local linear or quadratic regression around the cutoff of zero months to retirement, interacting each term in the regression with occupational cluster membership. The causal effects are the discontinuity interacted with the discontinutiy interacted with cluster membership (except for cluster one, which is the omitted value). The running variable is expected months to retirement. A bandwidth of 3 to 6 years is used; observations are weighted using a triangular or uniform kernel. Standard errors are heteroskedasticity robust and clustered at the individual level. Variables are defined in the note to Table 3.3.

Figure C.6: Effect of retirement eligibility on Anxiety and Depression across occupational clusters across different specifications



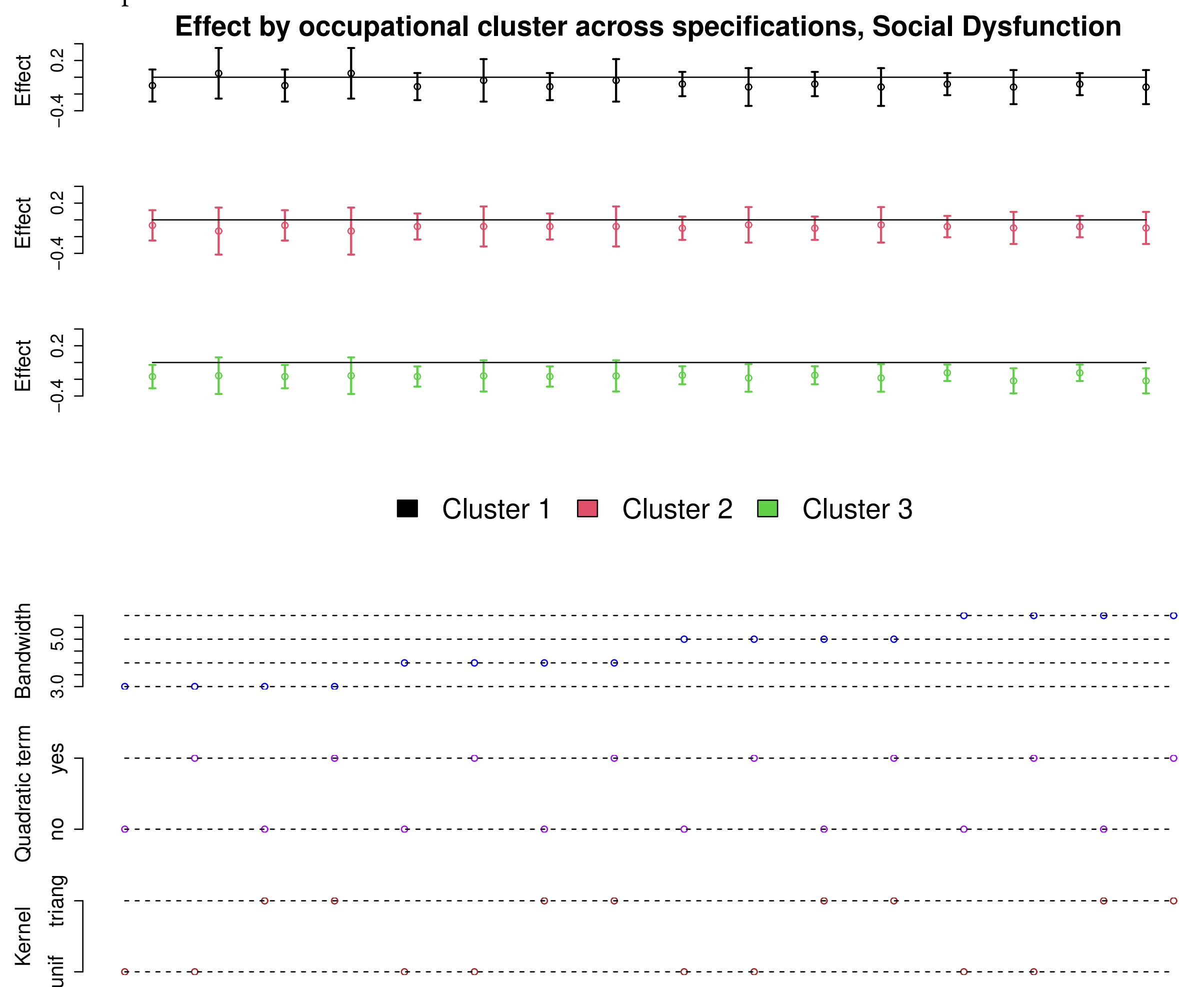
Notes: The range of specifications for estimation is described in the note to Figure C.4. Standard errors are heteroskedasticity robust and clustered at the individual level. Variables are defined in the note to Table 3.3.

Figure C.7: Effect of retirement eligibility on Loss of Confidence across occupational clusters across different specifications



Notes: The range of specifications for estimation is described in the note to Figure C.4. Standard errors are heteroskedasticity robust and clustered at the individual level. Variables are defined in the note to Table 3.3.

Figure C.8: Effect of retirement eligibility on Social Dysfunction across occupational clusters across different specifications



Notes: The range of specifications for estimation is described in the note to Figure C.4. Standard errors are heteroskedasticity robust and clustered at the individual level. Variables are defined in the note to Table 3.3.

calculated in different ways. The exception is the "robust" standard error (Calonico et al., 2014). This method adjusts for the bias associated with larger bandwidths. I cannot calculate CCT standard errors for local linear estimators which include interaction terms. I therefore use OLS standard errors clustered at the individual level, and examine how results change as the bandwidth decreases (see Appendix C.3 and C.4).

Note that I do not cluster standard errors by the running variable, age in months. The reason is that standard errors clustered by the running variable have poor coverage properties, i.e., if the regression is correctly specified, the standard errors will tend to be too small. (Kolesár & Rothe, 2018).

Table C.2: Standard errors using different calculations, for each dependent variable

	GHQ12 caseness	1(caseness $>$ 8)	Loss of Confidence	Social Dysfunction	Anxiety and Depression	Log Household Income	Retired
OLS	0.044	0.004	0.020	0.033	0.040	0.017	0.008
OLS clustered	0.044	0.004	0.020	0.033	0.040	0.017	0.008
cct conventional	0.050	0.004	0.023	0.038	0.046	0.020	0.009
cct bias-correct	0.050	0.004	0.023	0.038	0.046	0.020	0.009
cct robust	0.077	0.007	0.035	0.058	0.072	0.031	0.014
Honest KR	0.050	0.004	0.023	0.038	0.046	0.020	0.009

Notes: Standard errors are calculated for an RDD estimator where the running variable is expected months to retirement eligibility. CCT standard errors, CCT bias-correct standard errors and CCT robust standard errors are as detailed by Calonico, Cattaneo, and Titiunik (2014). Honest standard errors are those described by Kolesár and Rothe (2018). The estimation is described in the notes to Table 3.7. Variables are defined in the notes to Table 3.3.

C.6 Analysis using four occupational clusters

In this section, I discuss the tradeoffs involved in choosing the number of clusters, and show that the results when the number of clusters increases to four are consistent with my main results.

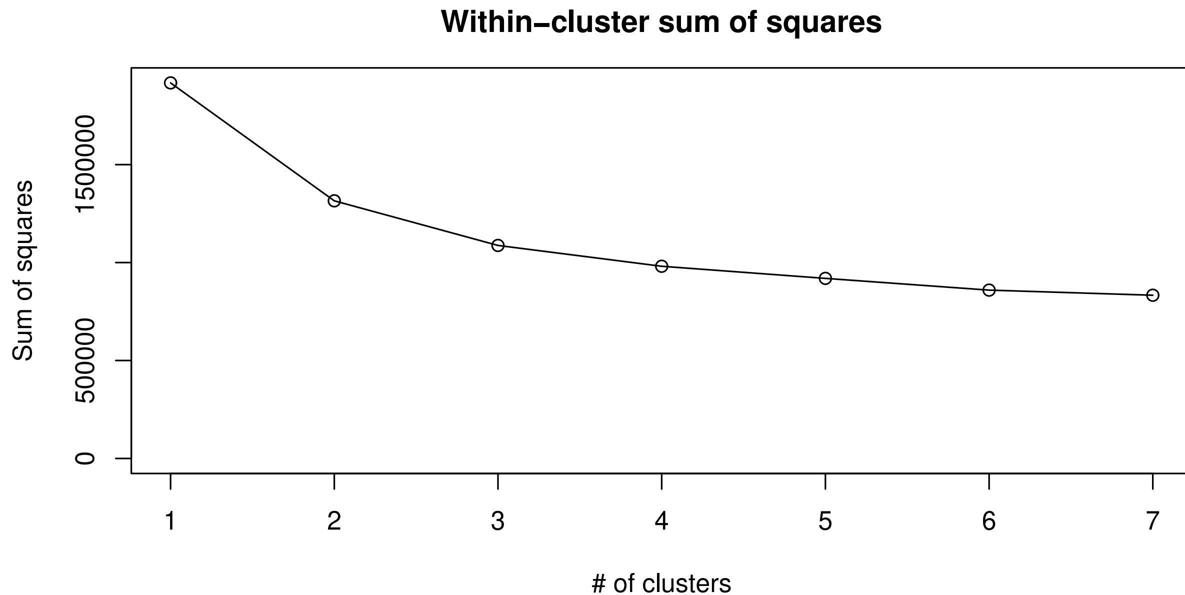
Using a larger number of clusters yields a finer, more precise definition of an occupational type, but also reduces the statistical power and ability to detect heterogeneous effects. One way of quantifying the effect of increasing clusters is to observe how the share of variation explained by the cluster structure changes as the number of clusters increases. Figure C.9 shows how the within-cluster sum of squares changes as the number of clusters increases. A smaller within-cluster sum of squares indicates that the cluster structure explains a larger share of the variation between occupations. We can see that the returns to increasing the number of clusters decreases significantly between three and four clusters, i.e., the reduction in within-cluster sum of squares when we move from three to four clusters is small relative to the reduction in within-cluster sum of squares when we move from two to three clusters. This indicates that using three or four clusters is likely to deliver most of the benefits of clustering in explaining the structure of occupational characteristics while preserving precise estimates.

Because the within-cluster sum of squares is similar for three and four clusters, I use three clusters within the main text to attain the most precise estimates. However, I also reproduce my main results using four clusters. The occupational mix of clusters produced, and the O*NET characteristics, are shown in Table C.3 and Figure C.10 respectively. Cluster one is over-represented amongst service workers, plant and machinery operator workers, and elementary occupations. These jobs score the lowest for cruciality of positioning and tight scheduling, are the least routine, and the least competitive. Cluster two occupations are over-represented amongst technicians and associate professionals, and to a lesser extent amongst senior officials and professionals. These jobs score highly for cruciality of position, conflictual contact, competitiveness, and working with people. Cluster three is over-represented amongst senior officials, professionals, and clerks. These jobs have the most sedentary body positioning, score highly for communication, are the most routine, and are the most indoor, low-hazard, comfortable jobs. Cluster four is most over-represented amongst skilled agricultural jobs, and in craft and related trades. These jobs score highest for wearing PPE, lowest for communication, highest for tight scheduling, lowest for working with people, lowest for working indoors, and highest for uncomfortable conditions and high job hazards.

There is not a direct mapping from the three cluster summary to the four cluster summary. Loosely, in terms of occupational employment and O*NET characteristics, cluster one (with three clusters) is similar in characteristics to cluster four (with four clusters); cluster three (three clusters) is similar to clusters two and three (with four clusters); and cluster two (with three clusters) is most similar to cluster one (four clusters).

Table C.4 shows my main results where four clusters are used. With regard to the overall GHQ12 caseness, there is a higher benefit of retirement amongst people who work in cluster four occupations, and a higher (though statistically insignificant) benefit of retirement for people in clusters two and three. The benefit of retirement eligibility for those in cluster four comes predominantly from symptoms of Anxiety and Depression, while the benefit in cluster two comes from symptoms of Loss of Confidence, Social

Figure C.9: Within-cluster sum of squares at different number of clusters

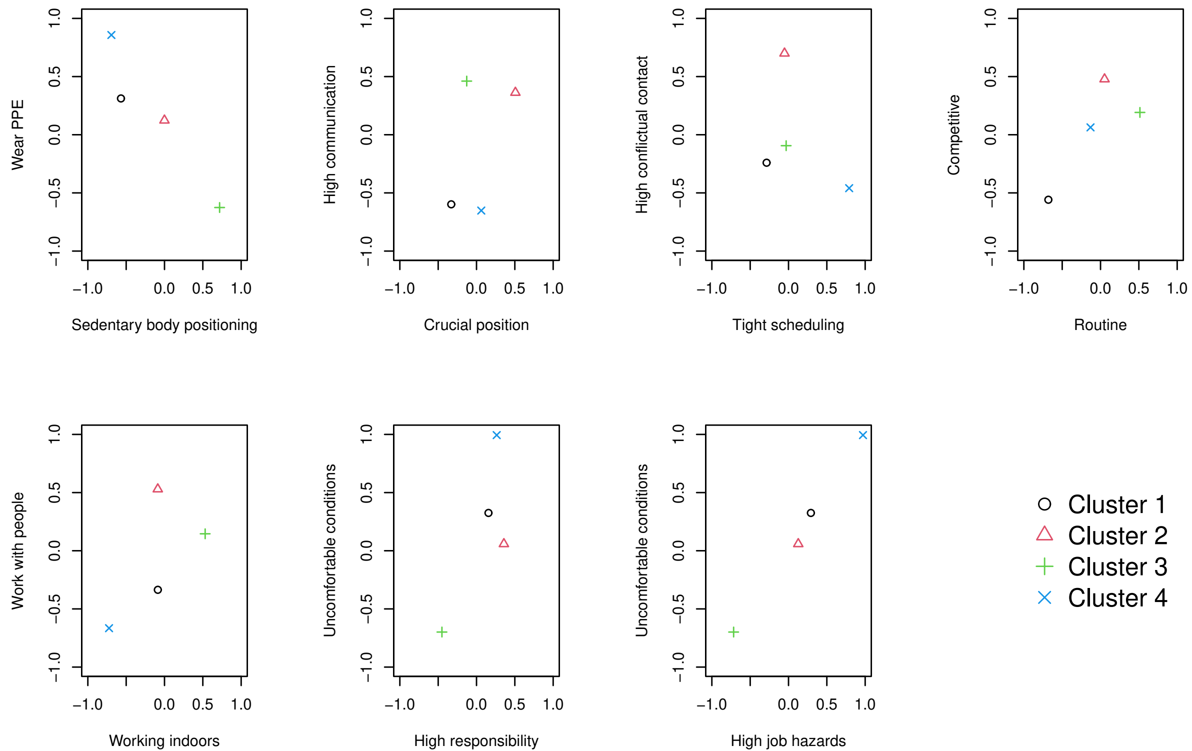


Notes: I cluster person-wave observations by O*NET occupation data using K-means clustering. Within-cluster sum of squares is the sum of the Euclidian distance between data points and their centroid. I normalize by dividing by the total Euclidian distance between all data points.

Dysfunction, and Anxiety and Depression. The benefit of retirement eligibility for those in cluster two is concentrated amongst men, who gain significantly more. For women in cluster four, their benefit of retirement eligibility is much larger than that for men, although for the GHQ12 caseness, this effect does not attain statistical significance.

Although we should not expect a direct mapping from cluster three results to cluster four results, the broad pattern is consistent. Cluster four (four clusters) workers have a higher than average benefit of retirement eligibility, concentrated in symptoms of Anxiety and Depression. This result is consistent with those in cluster one (three clusters) occupations having a higher than average benefit of retirement eligibility. Similarly, cluster two (four clusters) workers have somewhat elevated benefits from retirement eligibility, spread more evenly across symptoms of Anxiety and Depression, Loss of Confidence and Social Dysfunction. This result is consistent with those in cluster three (three clusters) having higher benefits from retirement eligibility spread across symptoms of Social Dysfunction and Anxiety and Depression. The inconsistencies are that, based on the three cluster results, we would think that either cluster two or three would have the highest overall benefit from retirement eligibility in the four-cluster analysis. Actually cluster four does. We would also not expect cluster two employees to gain in terms of symptoms of Loss of Confidence.

Figure C.10: Percent of people in each cluster scoring high minus percent scoring low, four clusters



Notes: I cluster person-wave observations by O*NET occupation data using K-means clustering. The measures of cluster characteristics are calculated by subtracting the percentage of people who score "low" on an O*NET variable within a cluster from the percentage of people scoring "high", where scoring "high" or "low" is determined by being in the top 25% or bottom 25% respectively of the distribution of scores. Individuals are assigned to occupation clusters based on current or last occupation. O*NET indices are calculated based on the O*NET model, which groups variables into categories. I calculate the indices as the simple average over the variables in the index, which are signed to facilitate an intuitive interpretation (see Table 3.2).

Table C.3: Share of occupational employment by cluster

	cluster 1	cluster 2	cluster 3	cluster 4
Senior Officials	0.022	0.339	0.612	0.027
Professionals	0.014	0.380	0.598	0.008
Technicians and Associate Professionals	0.091	0.513	0.366	0.030
Clerks	0.046	0.039	0.906	0.009
Service Workers	0.639	0.302	0.042	0.017
Skilled Agriculture	0.027	0.032	0.029	0.912
Craft and Related Trades	0.072	0.083	0.059	0.786
Plant and Machinery Operator Workers	0.530	0.050	0.029	0.391
Elementary Occupations	0.534	0.137	0.045	0.284

Notes: The table shows the one-digit ISCO-88 occupation shares in each occupational cluster in the sample. I cluster person-wave observations by O*NET occupation data using K-means clustering.

Table C.4: The effect of retirement by occupational cluster (four clusters)

	GHQ12 caseness	1(caseness > 8)	Loss of Confidence	Social Dysfunction	Anxiety and Depression	Log Household Income	Retired
Panel A: All							
Cluster 1 RD	−0.066	−0.003	−0.004	−0.032	−0.045	0.163***	0.268***
	(0.099)	(0.008)	(0.045)	(0.075)	(0.086)	(0.036)	(0.015)
Cluster 2 RD	−0.094	−0.003	−0.082**	−0.121*	−0.142*	0.137***	0.205***
	(0.091)	(0.007)	(0.041)	(0.071)	(0.083)	(0.037)	(0.015)
Cluster 3 RD	−0.108	−0.009	−0.034	−0.091	−0.067	0.105***	0.159***
	(0.078)	(0.006)	(0.035)	(0.060)	(0.072)	(0.030)	(0.013)
Cluster 4 RD	−0.238**	−0.008	−0.067	−0.099	−0.311***	0.142***	0.323***
	(0.107)	(0.008)	(0.049)	(0.079)	(0.098)	(0.036)	(0.017)
Obs	47,326	47,326	47,551	47,547	47,514	32,641	52,547
Panel B: Men							
Cluster 1 RD	−0.101	−0.001	0.015	−0.002	−0.066	0.132***	0.303***
	(0.124)	(0.010)	(0.058)	(0.093)	(0.111)	(0.037)	(0.019)
Cluster 2 RD	−0.143	−0.007	−0.075	−0.164**	−0.095	0.165***	0.212***
	(0.108)	(0.008)	(0.051)	(0.082)	(0.102)	(0.040)	(0.019)
Cluster 3 RD	−0.087	−0.007	−0.052	−0.050	−0.069	0.087***	0.185***
	(0.088)	(0.007)	(0.042)	(0.069)	(0.085)	(0.032)	(0.016)
Cluster 4 RD	−0.223**	−0.010	−0.062	−0.133*	−0.310***	0.137***	0.319***
	(0.106)	(0.008)	(0.050)	(0.078)	(0.101)	(0.037)	(0.018)
Obs	30,571	30,571	30,687	30,678	30,672	25,576	34,323
Panel C: Women							
Cluster 1 RD	−0.005	−0.005	−0.028	−0.055	−0.014	0.219*	0.246***
	(0.160)	(0.014)	(0.071)	(0.124)	(0.135)	(0.130)	(0.021)
Cluster 2 RD	−0.002	0.003	−0.107	−0.018	−0.264*	−0.112	0.234***
	(0.161)	(0.014)	(0.070)	(0.131)	(0.139)	(0.091)	(0.023)
Cluster 3 RD	−0.170	−0.018	−0.004	−0.183	−0.099	0.155	0.146***
	(0.153)	(0.013)	(0.065)	(0.119)	(0.132)	(0.105)	(0.022)
Cluster 4 RD	−0.655	−0.031	−0.181	−0.028	−0.653*	0.281**	0.428***
	(0.479)	(0.043)	(0.189)	(0.358)	(0.362)	(0.142)	(0.052)
Obs	16,755	16,755	16,864	16,869	16,842	7,065	18,014

Notes: results are the short-run causal effect of being eligible to retire for those in each occupational cluster. I use a local linear regression about the cutoff of zero months to retirement, interacted with occupational cluster membership as in Equation 3.4. I control for wave fixed effects. Cluster one is the omitted cluster. The causal effect for cluster one individuals is the discontinuity at the age of retirement eligibility. The causal effects for clusters two to four are the causal effect plus the coefficients on the terms which interact cluster membership and the discontinuity. I calculate the standard error of the causal effect by using the heteroskedasticity-robust variance-covariance matrix clustered at the individual level. I use a bandwidth of 5 years. I use a triangular kernel. Variables are defined in the note to Table 3.3.

C.7 Robustness tests

I conduct a number of robustness tests. Firstly, I drop wave fixed effects. Results are shown in Table C.5. The patterns are very similar to those of the baseline results. In the full sample, there is a beneficial effect of being eligible to retire on all measures of mental health and this benefit is higher for people in occupations in clusters one and three. In cluster one, the benefit is especially pronounced for symptoms of Anxiety and Depression. In cluster three, benefits are more evenly drawn from symptoms of Anxiety and Depression and Social Dysfunction.

Secondly, I investigate how my results change when I estimate the causal effect amongst only those who work up to retirement. I drop all individuals where 1) they are observed not working at any point within the bandwidth before the age of retirement eligibility or 2) it cannot be determined whether they were working at all points within the bandwidth before their age of retirement eligibility. This exercise involves dropping all individuals who never reach the age of retirement eligibility, and all individuals who enter the sample with fewer than 5 years until retirement eligibility. Correspondingly, standard errors are larger.

Results are shown in Table C.6. Across all measures of mental health, there is a beneficial effect of retirement eligibility in the full sample. The effect is significant for overall caseness, symptoms of Social Dysfunction, and Anxiety and Depression. On the other hand, there is not as pronounced a benefit of retirement for those in cluster one occupations in this specification, though they still experience a reduction in symptoms of Anxiety and Depression, which is significant at the 10% level. There is a larger benefit of retirement on symptoms of mental illness for those in cluster three, with the benefit coming from symptoms of Social Dysfunction and Anxiety and Depression. These results are consistent with the benefits of retirement eligibility accruing mostly to those who are still working at the point they become eligible to retire in cluster three, but partially to those who are not working at the age of retirement eligibility in cluster one. If those who benefit from being eligible to retire are working, this suggests the benefit runs through the decision to retire, i.e., the benefit of retiring from cluster three jobs is high. On the other hand, it is possible that many people in cluster one who benefit from being eligible to retire have stopped working by the time they reach retirement age, and benefit predominantly from the added economic security that BSP benefits give them.

Table C.5: The effect of retirement eligibility by occupational cluster, excluding wave fixed effects

	GHQ12 caseness	1(caseness > 8)	Loss of Confidence	Social Dysfunction	Anxiety and Depression	Log Household Income	Retired
Panel A: Average effect of retirement eligibility							
All clusters	−0.119***	−0.008**	−0.038*	−0.126***	−0.116***	0.127***	0.224***
	(0.044)	(0.004)	(0.020)	(0.033)	(0.041)	(0.017)	(0.008)
Panel B: Effect of retirement eligibility across occupational clusters							
Cluster 1	−0.135	−0.007	−0.052	−0.103	−0.205**	0.134***	0.298***
	(0.089)	(0.008)	(0.041)	(0.066)	(0.082)	(0.032)	(0.015)
Cluster 2	−0.051	−0.005	−0.030	−0.089	−0.023	0.178***	0.234***
	(0.079)	(0.007)	(0.036)	(0.058)	(0.072)	(0.030)	(0.013)
Cluster 3	−0.161**	−0.012**	−0.039	−0.167***	−0.140**	0.089***	0.175***
	(0.067)	(0.006)	(0.030)	(0.049)	(0.061)	(0.025)	(0.011)
Obs	47,235	47,235	47,235	47,235	47,235	32,641	47,235

Notes: Results are the short-run causal effect of being eligible to retire for those in each occupational cluster. I use a local linear estimator as described in the note of Table 3.7, although in this specification, I omit wave fixed effects. Variables are defined in the note to Table 3.3. *p<0.1; **p<0.05; ***p<0.01

Table C.6: The effect of retirement eligibility by occupational cluster amongst people who work until the SPA

	GHQ12 caseness	1(caseness > 8)	Loss of Confidence	Social Dysfunction	Anxiety and Depression	Log Household Income	Retired
			Panel A: Average effect of retirement eligibility				
All clusters	-0.120^{**}	-0.007	-0.043	-0.135^{***}	-0.183^{***}	-0.025	0.430^{***}
	(0.059)	(0.005)	(0.027)	(0.044)	(0.054)	(0.020)	(0.009)
			Panel B: Effect of retirement eligibility across occupational clusters				
Cluster 1	-0.080	0.001	-0.036	-0.056	-0.180^{*}	0.027	0.503^{***}
	(0.116)	(0.010)	(0.053)	(0.087)	(0.105)	(0.035)	(0.017)
Cluster 2	-0.039	-0.003	-0.070	-0.058	-0.134	0.036	0.434^{***}
	(0.104)	(0.009)	(0.047)	(0.078)	(0.095)	(0.037)	(0.015)
Cluster 3	-0.212^{***}	-0.014	-0.030	-0.241^{***}	-0.235^{***}	-0.110^{***}	0.383^{***}
	(0.092)	(0.008)	(0.042)	(0.069)	(0.083)	(0.030)	(0.013)
Obs	27,568	27,568	27,568	27,568	27,568	18,372	27,568

Notes: results are the short-run causal effect of being eligible to retire for those in each occupational cluster. The sample is all those people who have a full work history within the bandwidth up to the age of retirement eligibility, and are always observed to be in work. I use a local linear estimator as described in the note of Table 3.7. Variables are defined in the note to Table 3.3. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Bibliography

Adam, S., Browne, J., Phillips, D., & Roantree, B. (2021). Frictions and taxpayer responses: Evidence from bunching at personal tax thresholds. *International Tax and Public Finance*, *28*(3), 612–653.

Ahammer, A., & Packham, A. (2022). Disability insurance screening and worker outcomes.

Aizer, A., Devereux, P., & Salvanes, K. (2022). Grandparents, moms, or dads? why children of teen mothers do worse in life. *Journal of Human Resources*, *57*(6), 2012–2047.

Aloisi, A., & De Stefano, V. (2022). Essential jobs, remote work and digital surveillance: Addressing the covid-19 pandemic panopticon. *International Labour Review*, *161*(2), 289–314.

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.

Anjara, S., Bonetto, C, Van Bortel, T., & Brayne, C. (2020). Using the ghq-12 to screen for mental health problems among primary care patients: Psychometrics and practical considerations. *International journal of mental health systems*, *14*(1), 1–13.

Atalay, K., & Barrett, G. F. (2014). The causal effect of retirement on health: New evidence from australian pension reform. *Economics Letters*, *125*(3), 392–395.

Autor, D., Kostøl, A., Mogstad, M., & Setzler, B. (2019). Disability benefits, consumption insurance, and household labor supply. *American Economic Review*, *109*(7), 2613–2654.

Bain, P., & Taylor, P. (2000). Entrapped by the 'electronic panopticon'? worker resistance in the call centre. *New technology, work and employment*, *15*(1), 2–18.

Ball, S., & Low, H. (2014). Do self-insurance and disability insurance prevent consumption loss on disability? *Economica*, *81*(323), 468–490.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.

Bell, M. (2015). Mental health at work and the duty to make reasonable adjustments. *Industrial Law Journal*, *44*(2), 194–221.

Belloni, M., Carrino, L., & Meschi, E. (2022). The impact of working conditions on mental health: Novel evidence from the uk. *Labour Economics*, 102176.

Benítez-Silva, H., Buchinsky, M., Man Chan, H., Cheidvasser, S., & Rust, J. (2004). How large is the bias in self-reported disability? *Journal of Applied Econometrics*, *19*(6), 649–670.

Benıtez-Silva, H., Buchinsky, M., & Rust, J. (2006). Using a life-cycle model to predict induced entry effects of a 1 for 2 benefit offset in the ssdi program.

Bentley, R. J., Kavanagh, A., Krnjacki, L., & LaMontagne, A. D. (2015). A longitudinal analysis of changes in job control and mental health. *American Journal of Epidemiology*, *182*(4), 328–334.

Bertoni, M., Maggi, S., & Weber, G. (2018). Work, retirement, and muscle strength loss in old age. *Health economics*, *27*(1), 115–128.

Blundell, R., Borella, M., Commault, J., & De Nardi, M. (2023). *Old age risks, consumption, and insurance* (tech. rep.). Institute for Fiscal Studies.

Blundell, R., Britton, J., Costa Dias, M., & French, E. (2017). The impact of health on labor supply near retirement. *Available at SSRN 3103887*.

Bonhomme, S., Lamadon, T., & Manresa, E. (2022). Discretizing unobserved heterogeneity. *Econometrica*, *90*(2), 625–643.

Borella, M., De Nardi, M., & Yang, F. (2019). *The lost ones: The opportunities and outcomes of non-college educated americans born in the 1960s* (tech. rep.). National Bureau of Economic Research.

Boyce, C. J., & Oswald, A. J. (2012). Do people become healthier after being promoted? *Health economics*, *21*(5), 580–596.

Brewer, M., Saez, E., & Shephard, A. (2010). Means-testing and tax rates on earnings. *Dimensions of Tax Design: the mirrlees review*, *1*, 90–201.

Britton, J., & Gruber, J. (2020). Do income contingent student loans reduce labor supply? *Economics of Education Review*, *79*, 102061.

Browne, J., & Phillips, D. (2017). *Estimating the size and nature of responses to changes in income tax rates on top incomes in the uk: A panel analysis* (tech. rep.). IFS Working Papers.

Burgess, J., & Connell, J. (2004). Emerging developments in call centre research. *Labour & Industry: a journal of the social and economic relations of work*, *14*(3), 1–13.

Bursztyn, L., Rao, A., Roth, C. P., & Yanagizawa-Drott, D. H. (2020). *Misinformation during a pandemic* (tech. rep.). National Bureau of Economic Research.

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

Campolieti, M., & Riddell, C. (2012). Disability policy and the labor market: Evidence from a natural experiment in canada, 1998–2006. *Journal of Public Economics*, *96*(3), 306–316. https://doi.org/https://doi.org/10.1016/j.jpubeco.2011.09.001

Campos-Mercade, P., Meier, A. N., Schneider, F. H., & Wengström, E. (2021). Prosociality predicts health behaviors during the covid-19 pandemic. *Journal of public economics*, *195*, 104367.

Carr, A. (2013). *Positive psychology: The science of happiness and human strengths*. Routledge.

Carrino, L., Glaser, K., & Avendano, M. (2020). Later retirement, job strain, and health: Evidence from the new state pension age in the united kingdom. *Health economics*, *29*(8), 891–912.

Castillo, M., Jordan, J. L., & Petrie, R. (2019). Discount rates of children and high school graduation. *The Economic Journal*, *129*(619), 1153–1181.

CEDEFOP. (2013). Quantifying skill needs in europe. occupational skills profiles: Methodology and application.

Chen, M. K., Rossi, P. E., Chevalier, J. A., & Oehlsen, E. (2019). The value of flexible work: Evidence from uber drivers. *Journal of political economy*, *127*(6), 2735–2794.

Churchill, S. A., Munyanyi, M. E., Prakash, K., & Smyth, R. (2020). Locus of control and the gender gap in mental health. *Journal of Economic Behavior & Organization*, *178*, 740–758.

Coad, A., & Binder, M. (2014). Causal linkages between work and life satisfaction and their determinants in a structural var approach. *Economics Letters*, *124*(2), 263–268.

Cole, H. L., Kim, S., & Krueger, D. (2019). Analysing the effects of insuring health risks: On the trade-off between short-run insurance benefits versus long-run incentive costs. *The Review of Economic Studies*, *86*(3), 1123–1169.

Dal Bianco, C. (2022). Disability insurance and the effects of return-to-work policies. *Review of Economic Dynamics*.

Davalos, M. E., & French, M. T. (2011). This recession is wearing me out! health-related quality of life and economic downturns. *Journal of Mental Health Policy and Economics*.

De Nardi, M. (2004). Wealth inequality and intergenerational links. *The Review of Economic Studies*, *71*(3), 743–768.

De Nardi, M., Fella, G., & Paz-Pardo, G. (2021). *Wage risk and government and spousal insurance* (tech. rep.). National Bureau of Economic Research.

De Nardi, M., French, E., & Jones, J. B. (2016). Medicaid insurance in old age. *American Economic Review*, *106*(11), 3480–3520.

De Nardi, M., Pashchenko, S., & Porapakkarm, P. (2022). *The lifetime costs of bad health* (tech. rep.). National Bureau of Economic Research.

Della Giusta, M., & Longhi, S. (2021). Stung by pension reforms: The unequal impact of changes in state pension age on uk women and their partners. *Labour Economics*, *72*, 102049.

Deshpande, M., & Li, Y. (2019). Who is screened out? application costs and the targeting of disability programs. *American Economic Journal: Economic Policy*, *11*(4), 213–48.

Deshpande, M., & Lockwood, L. M. (2022). Beyond health: Nonhealth risk and the value of disability insurance. *Econometrica*, *90*(4), 1781–1810.

Dhungel, B., Murakami, T., Wada, K., & Gilmour, S. (2021). Mortality risks among blue-and white-collar workers: A time series study among japanese men aged 25-64 years from 1980 to 2015. *Journal of occupational health*, *63*(1), e12215.

Dong, Y. (2015). Regression discontinuity applications with rounding errors in the running variable. *Journal of Applied Econometrics*, *30*(3), 422–446.

Dube, A., Naidu, S., & Reich, A. D. (2022). *Power and dignity in the low-wage labor market: Theory and evidence from wal-mart workers* (tech. rep.). National Bureau of Economic Research.

Edwards, T., Edwards, P., Ferner, A., Marginson, P., & Tregaskis, O. (2010). Multinational companies and the diffusion of employment practices from outside the country of origin. *Management International Review*, *50*(5), 613–634.

Egan, M., Matvos, G., & Seru, A. (2022). When harry fired sally: The double standard in punishing misconduct. *Journal of Political Economy*, *130*(5), 1184–1248.

Eibich, P. (2015). Understanding the effect of retirement on health: Mechanisms and heterogeneity. *Journal of health economics*, *43*, 1–12.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual review of sociology*, *40*, 31–53.

Erikson, R., & Goldthorpe, J. H. (2002). Intergenerational inequality: A sociological perspective. *Journal of Economic Perspectives*, *16*(3), 31–44.

Fadlon, I., & Nielsen, T. H. (2021). Family labor supply responses to severe health shocks: Evidence from danish administrative records. *American Economic Journal: Applied Economics*, *13*(3), 1–30.

French, E., & Song, J. (2014). The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy*, *6*(2), 291–337.

Frenkel, E., Kugelmass, S., Nathan, M., & Ingraham, L. J. (1995). Locus of control and mental health in adolescence and adulthood. *Schizophrenia Bulletin*, *21*(2), 219–226.

Galama, T. J., Lleras-Muney, A., & Van Kippersluis, H. (2018). The effect of education on health and mortality: A review of experimental and quasi-experimental evidence.

Gathergood, J. (2013). An instrumental variable approach to unemployment, psychological health and social norm effects. *Health economics*, *22*(6), 643–654.

Gelaye, B., Tadesse, M. G., Lohsoonthorn, V., Lertmeharit, S., Pensuksan, W. C., Sanchez, S. E., Lemma, S., Berhane, Y., Vélez, J. C., Barbosa, C., et al. (2015). Psychometric properties and factor structure of the general health questionnaire as a screening tool for anxiety and depressive symptoms in a multi-national study of young adults. *Journal of affective disorders*, *187*, 197–202.

Gelber, A., Moore, T., Pei, Z., & Strand, A. (2022). Disability insurance income saves lives.

Gelber, A., Moore, T. J., & Strand, A. (2017). The effect of disability insurance payments on beneficiaries' earnings. *American Economic Journal: Economic Policy*, *9*(3), 229–261.

Geyer, J., Haan, P., Hammerschmid, A., & Peters, M. (2020). Labor market and distributional effects of an increase in the retirement age. *Labour Economics*, *65*, 101817.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, *104*(4), 1091–1119.

Goldsmith-Pinkham, P., Hull, P., & Kolesár, M. (2021). On estimating multiple treatment effects with regression. *arXiv preprint arXiv:2106.05024*.

Goos, M., Manning, A., & Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American economic review*, *104*(8), 2509–26.

Gorry, A., Gorry, D., & Slavov, S. N. (2018). Does retirement improve health and life satisfaction? *Health economics*, *27*(12), 2067–2086.

Gourinchas, P.-O., & Parker, J. A. (2002). Consumption over the life cycle. *Econometrica*, *70*(1), 47–89.

Graetz, B. (1991). Multidimensional properties of the general health questionnaire. *Social psychiatry and psychiatric epidemiology*, *26*(3), 132–138.

Gureje, O., & Obikoya, B. (1990). The ghq-12 as a screening tool in a primary care setting. *Social psychiatry and psychiatric epidemiology*, *25*(5), 276–280.

Haller, A., Staubli, S., & Zweimüller, J. (2020). *Designing disability insurance reforms: Tightening eligibility rules or reducing benefits* (tech. rep.). National Bureau of Economic Research.

Hammond, R, Baxter, S, Bramley, R, Kakkad, A, Mehta, S, & Sadler, M. (2016). Considerations on state pension age in the united kingdom. *British Actuarial Journal*, *21*(1), 165–203.

Handel, M. J. (2012). Trends in job skill demands in oecd countries.

Hansen, S., Lambert, P. J., Bloom, N., Davis, S. J., Sadun, R., & Taska, B. (2023). *Remote work across jobs, companies, and space* (tech. rep.). National Bureau of Economic Research.

Hardy, W. (2016). Hardy, 2006 [Accessed: 2021-08-05].

Hardy, W., Keister, R., & Lewandowski, P. (2018). Educational upgrading, structural change and the task composition of jobs in europe. *Economics of Transition*, *26*(2), 201–231.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of political economy*, *63*(4), 309–321.

He, H., Neumark, D., & Weng, Q. (2021). Do workers value flexible jobs? a field experiment. *Journal of Labor Economics*, *39*(3), 709–738.

Heller-Sahlgren, G. (2017). Retirement blues. *Journal of health economics*, *54*, 66–78.

Holston, K., Laubach, T., & Williams, J. C. (2017). Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics*, *108*, S59–S75.

Hosseini, R., Kopecky, K. A., & Zhao, K. (2021). How important is health inequality for lifetime earnings inequality?

Hoynes, H. W., & Moffitt, R. (1999). Tax rates and work incentives in the social security disability insurance program: Current law and alternative reforms. *National Tax Journal*, *52*(4), 623–654.

Jenkinson, C., & Layte, R. (1997). Development and testing of the uk sf-12. *Journal of health services research & policy*, *2*(1), 14–18.

Johannsen, R., & Zak, P. J. (2020). Autonomy raises productivity: An experiment measuring neurophysiology. *Frontiers in Psychology*, *11*, 963.

Johnston, D. W., & Lee, W.-S. (2009). Retiring to the good life? the short-term effects of retirement on health. *Economics Letters*, *103*(1), 8–11. https://doi.org/https://doi.org/10.1016/j.econlet.2009.01.015

Jolivet, G., & Postel-Vinay, F. (2020). A structural analysis of mental health and labor market trajectories.

Keane, M., Capatina, E., & Maruyama, S. (2020). *Health shocks and the evolution of earnings over the life-cycle* (tech. rep.).

Kellogg, M. (2022). Household self-insurance and the value of disability insurance in the united states. *Available at SSRN 4187881*.

Kesavayuth, D., Rosenman, R. E., & Zikos, V. (2016). Retirement, Personality, And Well-Being. *Economic Inquiry*, *54*(2), 733–750. https://ideas.repec.org/a/bla/ecinqu/v54y2016i2p733-750.html

Kettlewell, N., & Lam, J. (2022). Retirement, social support and mental well-being: A couple-level analysis. *The European Journal of Health Economics*, *23*(3), 511–535.

Knapp, M., McDaid, D., & Parsonage, M. (2011). Mental health promotion and mental illness prevention: The economic case.

Kolesár, M., & Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, *108*(8), 2277–2304.

Kolodziej, I. W., & García-Gómez, P. (2019). Saved by retirement: Beyond the mean effect on mental health. *Social Science & Medicine*, *225*, 85–97.

Kontodimopoulos, N., Pappa, E., Niakas, D., & Tountas, Y. (2007). Validity of sf-12 summary scores in a greek general population. *Health and quality of life outcomes*, *5*(1), 1–9.

Kostol, A. R., & Mogstad, M. (2014). How financial incentives induce disability insurance recipients to return to work. *American Economic Review*, *104*(2), 624–55.

Krekó, J., Prinz, D., & Weber, A. (2022). Take-up and labor supply responses to disability insurance earnings limits.

Le Grand, J. (2003). *Motivation, agency, and public policy: Of knights and knaves, pawns and queens*. OUP Oxford.

Leckcivilize, A., & McNamee, P. (2021). The effects of retirement on health and health behaviour among retirees and their partners: Evidence from the english longitudinal study of ageing. *Journal of Population Ageing*, 1–32.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, *48*(2), 281–355.

Lewandowski, P. (2020). Occupational exposure to contagion and the spread of covid-19 in europe.

Lockwood, G., Henderson, C., & Thornicroft, G. (2012). The equality act 2010 and mental health. *The British Journal of Psychiatry*, *200*(3), 182–183.

Low, H., & Pistaferri, L. (2015). Disability insurance and the dynamics of the incentive insurance trade-off. *American Economic Review*, *105*(10), 2986–3029.

Low, H., & Pistaferri, L. (2020). Disability insurance: Theoretical trade-offs and empirical evidence. *Fiscal Studies*, *41*(1), 129–164.

Lyness, K. S., Gornick, J. C., Stone, P., & Grotto, A. R. (2012). It's all about control: Worker control over schedule and hours in cross-national context. *American Sociological Review*, *77*(6), 1023–1049.

Maestas, N., Mullen, K. J., Powell, D., Von Wachter, T., & Wenger, J. B. (2023). The value of working conditions in the united states and implications for the structure of wages. *American Economic Review*, *113*(7), 2007–2047.

Maestas, N., Mullen, K. J., & Strand, A. (2013). Does disability insurance receipt discourage work? using examiner assignment to estimate causal effects of ssdi receipt. *American Economic Review*, *103*(5), 1797–1829.

Markussen, S., Røed, K., & Schreiner, R. C. (2018). Can compulsory dialogues nudge sick-listed workers back to work? *The Economic Journal*, *128*(610), 1276–1303.

Martinez Jimenez, M., Hollingsworth, B., & Zucchelli, E. (2021). Health, retirement and economic shocks.

Mas, A., & Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, *107*(12), 3722–3759.

Mazzonna, F., & Peracchi, F. (2017). Unhealthy retirement? *Journal of Human Resources*, *52*(1), 128–151.

McCabe, C. J., Thomas, K. J., Brazier, J. E., & Coleman, P. (1996). Measuring the mental health status of a population: A comparison of the ghq–12 and the sf–36 (mhi–5). *The British Journal of Psychiatry*, *169*(4), 517–521.

McCrone, P., Dhanasiri, S., Patel, A., Knapp, M., & Lawton-Smith, S. (2008). Paying the price. *The cost of mental health care in England to*, *2026*, 1–165.

Meghir, C., & Phillips, D. (2008). *Labour supply and taxation* (tech. rep.). IFS Working Papers.

Michaud, A., & Wiczer, D. (2018). Occupational hazards and social disability insurance. *Journal of Monetary Economics*, *96*, 77–92.

Moore, T. J. (2015). The employment effects of terminating disability benefits. *Journal of Public Economics*, *124*, 30–43.

Müller, T., & Shaikh, M. (2018). Your retirement and my health behavior: Evidence on retirement externalities from a fuzzy regression discontinuity design. *Journal of Health Economics*, *57*, 45–59. https://doi.org/https://doi.org/10.1016/j.jhealeco.2017.10.005

Neuman, K. (2008). Quit your job and get healthier? the effect of retirement on health. *Journal of Labor Research*, *29*, 177–201.

Nishimura, Y., Oikawa, M., & Motegi, H. (2018). What explains the difference in the effect of retirement on health? evidence from global aging data. *Journal of Economic Surveys*, *32*(3), 792–847.

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

Picchio, M., & Ours, J. C. v. (2020). Mental health effects of retirement. *De Economist*, *168*(3), 419–452.

Ravesteijn, B., Kippersluis, H. v., & Doorslaer, E. v. (2018). The wear and tear on health: What is the role of occupation? *Health economics*, *27*(2), e69–e86.

Rawls, J. (2020). *A theory of justice: Revised edition*. Harvard university press.

Rose, L. (2020). Retirement and health: Evidence from england. *Journal of Health Economics*, *73*, 102352.

Rotter, J. B., & Mulry, R. C. (1965). Internal versus external control of reinforcement and decision time. *Journal of personality and social psychology*, *2*(4), 598.

Royston, S. (2017). *Broken benefits: What's gone wrong with welfare reform*. Policy Press.

Salvati, F. (2021). *Essays on human capital formation over the life cycle* [Doctoral dissertation, UCL (University College London)].

Silver, D., & Zhang, J. (2022). *Impacts of basic income on health and economic well-being: Evidence from the va's disability compensation program* (tech. rep.). National Bureau of Economic Research.

Stansfeld, S. A., Fuhrer, R., Head, J., Ferrie, J., & Shipley, M. (1997). Work and psychiatric disorder in the whitehall ii study. *Journal of psychosomatic research*, *43*(1), 73–81.

Stansfeld, S. A., Fuhrer, R., Shipley, M. J., & Marmot, M. G. (1999). Work characteristics predict psychiatric disorder: Prospective results from the whitehall ii study. *Occupational and environmental medicine*, *56*(5), 302–307.

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*(1), 1–34.

Tibubos, A. N., & Kröger, H. (2020). A cross-cultural comparison of the ultrabrief mental health screeners phq-4 and sf-12 in germany. *Psychological Assessment*, *32*(7), 690.

Wallace, S., Nazroo, J., & Bécares, L. (2016). Cumulative effect of racial discrimination on the mental health of ethnic minorities in the united kingdom. *American journal of public health*, *106*(7), 1294–1300.

Ware Jr, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical care*, 220–233.

Watson, T. (2020). Retirement blues and boons: Health and spillover effects of retirement.

Woodcock, J., & Graham, M. (2019). The gig economy. *A critical introduction. Cambridge: Polity*.