The Employment Effects of Disability Benefits Without Work Restrictions

Bouke Klein Teeselink*[1] and Georgios Melios[2,3]

[1] *Department of Political Economy, King's College London*

[2] *Department of Behavioural Science, London School of Economics and Political Science*

[3] *Ethos Lab—Centre for Governance and Sustainability Research, Athens, Greece*

Abstract

Disability benefits typically reduce labor force participation through raising non-labor income (income effects) and phasing out benefits with earnings (substitution effects). Distinguishing these mechanisms matters for policy design, as income effects are non-distortionary while substitution effects generate deadweight loss. We isolate pure income effects for the UK's Personal Independence Payment (PIP) program, which supports disability-related costs without work restrictions or earnings tests. Following a 2013 reform, we use differential eligibility changes across health conditions and quasi-random regional assignment to different assessment providers in a difference-in-differences design with matched panel data. Despite substantial benefit receipt changes (1.8-5.9 percentage points across strategies), we only find positive employment effects in stricter assessment regions. For individuals who gain or lose benefits due to condition-specific eligibility changes, the effects are close to zero. These findings suggest that disability benefits designed as cost-of-living supplements can provide financial security without necessarily disincentivizing work.

Keywords: Disability benefits; Labor force participation; Income effects; United Kingdom;

JEL Codes: I38; H55; J22

*bouke.klein_teeselink@kcl.ac.uk. Melios: g.melios@lse.ac.uk. We thank Tom Waters and seminar participants from the Department for Work and Pensions for helpful comments. This research was supported by Ethos Lab—Centre for Governance and Sustainability Research, as part of the Horizon Europe project "BENEFITS", funded by the European Commission under Grant Agreement No. 101179032. The results presented are research outputs and do not necessarily reflect the views of the European Commission or any other organizations participating in the BENEFITS project.

1 Introduction

Disability benefits represent one of the largest social insurance expenditures in developed countries, with spending often exceeding two percent of GDP across OECD nations (OECD, 2010). While these programs provide important income security to individuals with health-related limitations, they also tend to reduce employment among recipients, with typical estimates ranging between 15 and 30 percentage points (Bound, 1989; Autor and Duggan, 2003, 2007; von Wachter et al., 2011; Maestas et al., 2013; French and Song, 2014; Kostol and Mogstad, 2014; Moore, 2015; Deshpande, 2016*a*; Autor et al., 2016; Garcia-Mandicó et al., 2020). This disemployment effect creates a tension between providing adequate support for vulnerable populations and maintaining employment incentives.

The decrease in employment resulting from disability benefit receipt can be decomposed into an income and a substitution effect, where the former refers to the tendency to reduce work hours because the additional non-labor income reduces the financial necessity of working, and the latter refers to disemployment resulting from the fact that most programs reduce benefit payments when recipients earn more labor income (Autor and Duggan, 2007; Gelber et al., 2017). Since most disability programs simultaneously provide income support and phase out with work activity, it is empirically challenging to separate these mechanisms (Autor and Duggan, 2007).[1] Nonetheless, the distinction is fundamental for policy design, as income effects represent non-distortionary transfers, while substitution effects generate deadweight loss by artificially constraining labor market choices (Deshpande, 2016*a*).

We exploit the unique design of the UK's Personal Independence Payment (PIP) to identify pure income effects of disability benefits. Introduced in 2013 to replace Disability Living Allowance, PIP provides non-means-tested support to individuals with long-term health conditions regardless of employment status. Unlike traditional disability programs, PIP eligibility does not require being out of work or demonstrating work incapacity—it simply aims to cover extra costs of living with a disability (Department for Work and Pensions, 2012). Recipients face no earnings restrictions or benefit reductions when working, which eliminates the substitution effects that characterize most disability insurance systems. This design creates an ideal setting to isolate income effects,

[1] See Deshpande (2016*b*); Gelber et al. (2017); Deuchert and Eugster (2019); Jones and Marinescu (2022); Becker et al. (2024) for evidence on income effects.

as any employment response operates solely through the non-labor income channel rather than program-imposed work disincentives.[2]

To identify the causal effect of PIP benefits on employment, we exploit two sources of quasi-experimental variation created by the DLA-to-PIP reform using data from the UK Household Longitudinal Study (UKHLS), a large-scale panel survey tracking a nationally representative sample of UK households from 2009-2019 (University of Essex, Institute for Social and Economic Research, 2024). Our primary identification strategy exploits the reform's differential impact across health conditions: PIP broadened eligibility for mental health conditions, increasing benefit receipt by 5.9 percentage points for individuals with pre-existing mental health problems, while tightening criteria for minor physical disabilities reduced benefit receipt by 5.6 percentage points for this population. To complement this approach and address potential concerns about unobserved shocks affecting specific disability groups, we also leverage the administrative assignment of UK regions to different assessment providers. This assignment creates geographic variation in the likelihood that otherwise identical people receive benefits, as regions assigned to the more stringent Atos assessor experience a 1.8 percentage point reduction in benefit receipt relative to those assigned to Capita. Using difference-in-differences designs, we compare employment trajectories before and after reform implementation across these treatment and control groups to estimate the causal effect of benefits on employment.

Our results show only modest employment responses across the different identification strategies, despite observing substantial variation in benefit receipt between treatment and control groups. Using our primary condition-based identification strategy, individuals with pre-existing mental health conditions who gained improved benefit access under PIP (5.9 percentage point increase in benefit receipt) show no evidence of reduced employment. If anything, the point estimates are slightly positive throughout the post-reform period, ranging from 0.3 to 1.6 percentage points, though not statistically significant. Similarly, those with minor physical disabilities who faced benefit restrictions (5.6 percentage point reduction in benefit receipt) show no employment increases despite losing benefits, with employment effects near zero (-0.3 percentage points). In contrast, our complementary assessor-based strategy yields different results: regions assigned to more stringent

[2]Substitution effects might still emerge if people strategically drop out of the labor force to signal worse health in order to obtain higher PIP benefits.

assessors experience a statistically significant 2.9 to 3.3 percentage point increase in employment probability, suggesting clear employment responses to differential benefit access in this context.

Additional analyses examining benefit substitution across seventeen categories of social transfers provide validation of our identification strategies and also eliminate alternative mechanisms underlying our employment results. Of 51 estimates testing effects on alternative benefit categories, only three show statistically significant responses. Importantly, we observe no effects on benefits unrelated to disability policy such as maternity benefits, trade union benefits, or national insurance credits, which provides strong evidence that our main results capture variation in PIP receipts rather than broader economic or policy confounds such as austerity. When examining aggregated income sources, we similarly find minimal offsetting changes in private benefits, investment income, or miscellaneous income, with clear effects concentrated in social benefit income.

To interpret these results, we note that disability benefit changes may generate offset effects that vary across populations. Standard income effects suggest that additional non-labor income reduces work necessity, potentially decreasing employment. In contrast, for people with severe health conditions, additional income can facilitate employment by enabling work accommodations or medical treatment that improves work capacity. In addition, the scope for strategic misrepresentation (people strategically exaggerating health problems to increase the chance of receiving PIP benefits), might differ across treatment groups. For assessor treatment, stricter assessments could disproportionately deny benefits to individuals who were exaggerating their limitations, leading these cases to return to work. In contrast, our condition-based identification strategies examine individuals whose benefit eligibility changed based on objectively measurable pre-existing health conditions, which largely eliminates the scope for strategic behavior. As such, the positive employment effects in regions with stringent assessors can reflect both standard income effects among a broad population with marginal work capacity and the filtering out of strategic claimants who were seeking benefits despite retaining substantial work capacity. The absence of employment effects among individuals with pre-existing mental health conditions suggests that when strategic misrepresentation is minimized, facilitating effects may dominate for this population.

Some of these population-specific mechanisms are partly supported by our heterogeneity analyses, which reveal that employment responses vary somewhat across demographic subgroups in ways consistent with different underlying constraints. In particular, for our assessor treatment, unmarried individuals show significantly larger employment responses to benefit loss (6.0 percentage points increase) compared to married individuals (1.5 percentage points increase), suggesting that spousal income buffers the employment effects of benefit loss (see also Autor et al., 2019). Yet, for most other analyses, we find relatively little heterogeneity.

Our findings contribute to a large literature documenting the employment effects of disability benefit programs. The majority of this research examines the U.S. Social Security Disability Insurance (SSDI) program, and find that disability benefits reduce employment by 15 to 30 percentage points (Bound, 1989; Autor and Duggan, 2003; von Wachter et al., 2011; Maestas et al., 2013; French and Song, 2014). Reforms in other countries likewise show that tightening benefit eligibility rules leads to an increase in labor force participation (see e.g. Kostol and Mogstad, 2014; Borghans et al., 2014; Autor et al., 2019). However, these studies typically examine programs that simultaneously increase non-labor income and impose work restrictions through earnings thresholds or incapacity requirements. Consequently, the large employment effects documented in this literature represent a combination of income and substitution effects, making it empirically difficult to isolate the pure income effects that are important for optimal transfer design (Low and Pistaferri, 2015).

A smaller but growing literature attempts to isolate the pure income effects of disability benefits, though findings remain mixed. Deshpande (2016*b*) exploits the removal of youth from SSI rolls and finds income effects of similar magnitude to substitution effects, while Gelber et al. (2017) use a regression kink design to estimate income effects of approximately 20 percentage points per dollar of SSDI benefits. However, Gelber et al. (2017) find that these large income effects are concentrated among higher earners, with no significant disemployment effects for individuals with low prior earnings. European evidence presents similarly mixed results: Deuchert and Eugster (2019) document substantial income effects of disability insurance in Switzerland, while Becker et al. (2024) show no significant income effects in Germany. Our findings contribute to this literature by providing the first evidence on pure income effects from a disability benefit explicitly designed to eliminate work disincentives. Unlike previous studies that rely on econometric decomposition or examine benefit removals that retain some substitution effects, PIP's unique design as a cost-of-living supplement allows us to isolate income effects directly.

These findings also inform the broader literature on cash transfers. Like PIP, cash transfer programs provide income support without explicit work restrictions, offering a natural laboratory for studying how non-labor income affects employment decisions. However, this literature exhibits considerable heterogeneity in estimated effects. Several studies document minimal employment responses: Jones and Marinescu (2022) find no significant effects from Alaska's Permanent Fund Dividend, Sauval et al. (2024) show limited impacts of Polish child benefits, and Cook and East (2023, 2024) find small effects of SNAP benefits. Conversely, other studies document more substantial employment reductions following transfer receipt (Hoynes and Schanzenbach, 2012; Verlaat et al., 2023; Gromadzki, 2024). Our results contribute to resolving this apparent inconsistency by highlighting how program design and target population characteristics jointly determine income effect magnitudes. The variation in findings across the transfer literature likely reflects differences in transfer size relative to household income, targeting mechanisms, and the severity of constraints faced by recipient populations, which highlights the importance of program-specific evaluation.

The rest of the paper is organized as follows. Section 2 provides background on the UK disability benefit system and the transition from Disability Living Allowance to Personal Independence Payment. Section 3 describes our data from the UK Household Longitudinal Study and presents summary statistics for our treatment and control groups. Section 4 outlines our identification strategy. Section 5 presents our main results. Section 6 concludes with a discussion of policy implications and directions for future research.

2 Policy Background

Public disability benefits represent a large and growing component of social expenditure in advanced economies. In the UK, the share of the working-age population claiming disability benefits has increased from less than 4% in 2000 to almost 8% in 2023 (Latimer et al., 2024). The increased costs of disability programs have led to a series of reforms aimed at fiscal sustainability and improved targeting (Banks et al., 2015).

The current UK disability benefit system consists primarily of two benefit streams: incapacity benefits, intended to compensate those unable to work due to a health condition, and disability benefits, designed to address additional living costs faced by disabled individuals regardless of employment status (Latimer et al., 2024). Disability Living Allowance (DLA), introduced in 1992, provided financial support irrespective of employment and was largely non-means-tested. By 2013, the UK government identified several inefficiencies within DLA, including insufficient targeting, rapidly increasing caseloads, and concerns about fiscal sustainability (Gray, 2016).

In response, the UK government initiated a reform that gradually replaced DLA with the Personal Independence Payment (PIP), beginning in April 2013. The explicit aims of PIP were to improve objectivity and fairness in assessments, better target resources toward those with the greatest medical needs, and improve fiscal sustainability by reducing forecasted expenditure on working-age disability benefits by approximately 20% (Cribb et al., 2022; Gray, 2016). Despite these intentions, spending on disability benefits has continued to grow, reaching £14.7 billion by 2021–22, up from £9.4 billion prior to the introduction of PIP (Cribb et al., 2022).

PIP differs from DLA primarily in its assessment methodology. While DLA assessments relied heavily on self-reporting and periodic reviews, PIP introduced a functional assessment conducted by health professionals. This assessment explicitly aims to evaluate claimants' capability to undertake daily living and mobility activities rather than their medical diagnoses alone. Claimants are assessed against specific activities (such as preparing food, personal care, communication, and mobility), each scored based on the level of difficulty or assistance required (Gray, 2016; Department for Work and Pensions, 2015).

Assessments involve a combination of face-to-face consultations, paper-based evaluations, and reviews of medical and supporting evidence. Decisions are subsequently made based on accumulated points, categorizing claimants into standard or enhanced award groups for either daily living, mobility, or both components (Gray, 2016). These assessments, however, have generated significant controversy due to concerns about transparency, fairness, and the consistency of outcomes, highlighted by the high rate of appeals and successful overturns at tribunal stages (Gray, 2016). Crucially for our identification strategy, the Department for Work and Pensions assigned different UK regions to one of two independent assessment providers: Independent Assessment Services (Atos) covered eight regions including England's major economic centers and Scotland, while Capita was responsible for four regions including Wales and Northern Ireland. This administrative allocation created quasi-experimental variation in assessment stringency that we exploit in our empirical analysis.

Despite substantial policy interest, rigorous evaluations of PIP's impacts on labor force participation remain sparse (see Cribb et al. (2022) for a notable exception). This study contributes by directly evaluating the causal impact of PIP on labor market outcomes.

3 Data

Our analysis uses data from the UK Household Longitudinal Study (UKHLS), a large-scale panel survey that has tracked a nationally representative sample of UK households annually since 2009 (University of Essex, Institute for Social and Economic Research, 2024). The survey collects a wide range of information on employment status, income sources, including disability benefit receipt, health conditions, and detailed demographic characteristics, making it particularly well-suited for analyzing the labor market effects of disability benefit reforms. The full data set contains 568,946 surveys across 98,106 individuals who are interviewed between 2009 and 2023.

We impose two sample restrictions to construct our final sample. First, we remove data from the first two waves, as the DLA/PIP data are either incomplete or missing altogether. Second, we focus on all individuals who are 59 years old or younger in 2013, as older individuals turn 65 in the post-treatment period, leading to retirement. After imposing these restrictions, our final sample consists of 259,087 observations for 33,051 individuals. For most analyses, we use only the period up to 2019 to avoid the COVID-19 period, which reduces the number of observations to 210,773. In the rest of this section, we report summary statistics for the full sample (including COVID-19), but the general patterns are similar if we restrict ourselves to the pre-COVID period.

We construct our main variables as follows. The main policy variable is a binary indicator that takes the value of 1 if someone receives either DLA or PIP benefits. We additionally consider the total amount of money received across all social benefits. The main outcome variable is a dummy variable for whether someone is in paid employment or not. Our treatments are based on (i) one's mental health pre-treatment, (ii) one's physical health pre-treatment, and (iii) which assessor region one is assigned to. Section 4 discusses our identification strategies in more detail. Our main mental health indicator takes the value of 1 if someone answered 'most of the time' or 'all of the time' to one of the following three questions: (i) "During the past 4 weeks, how much of the time have you accomplished less than you would like with your work or other regular daily activities as a result of any emotional problems?" (ii) "During the past 4 weeks, how much of the time did you do work or other activities less carefully than usual as a result of any emotional problems?" (iii) "Have you felt downhearted and depressed?"

For physical health, we focus on the question: During the past 4 weeks, how much of the time have you accomplished less than you would like with your work or other regular daily activities as a result of your physical health? We define as severe those who answer either 'all of the time' or 'most of the time', and as minor those who answer either 'some of the time' or 'a little of the time'. Those who answer 'none of the time' are classified as not having physical problems and are excluded from this analysis. To complement these condition-based treatments, we also exploit regional variation in assessor assignment. People living in North East, North West, Yorkshire and Humber, London, East of England, South East, South West, and Scotland were assigned to Independent Assessment Services (formerly Atos), while four regions (East Midlands, West Midlands, Wales, and Northern Ireland) were assigned to Capita.[3]

Additional variables we use, either as controls or as matching variables, are age, sex, number of children, total monthly personal income gross (log), and highest educational qualification. Total monthly personal income gross consists of income from savings and investments, total monthly labor income gross, and total income from benefits and other sources. We use the log-transformed version of this variable. Highest educational qualification is reported across 16 categories such as Higher degree, A-level, or GCSE/O level.

Table 1, Panel A presents summary statistics for our treatment and control groups, described in Section 4, in the period before PIP was implemented. We define these groups based on their health status/location in the last period before the implementation of PIP. As expected, our primary condition-based treatments show substantial baseline differences that reflect the targeting of disability benefits toward individuals with health limitations. Individuals with poor mental or physical health show markedly worse outcomes in all dimensions: benefit receipt rates are approximately 5 times higher, employment rates are 30 percentage points lower, life satisfaction scores are substantially reduced, and monthly incomes are £566 and £365 lower. These large pre-treatment differences underscore the importance of our matching procedure for ensuring valid comparisons. In contrast, our complementary assessor assignment groups appear well-balanced, with Capita-assigned regions showing only slightly higher pre-treatment benefit receipt (5.0

For the same individuals, Panel B displays the information from the period after the PIP was

[3]For brevity, we will refer to Independent Assessment Services (formerly Atos) as Atos in the rest of the paper.

Table 1: Summary Statistics by Treatment and Control Groups (Pre-Treatment Period)

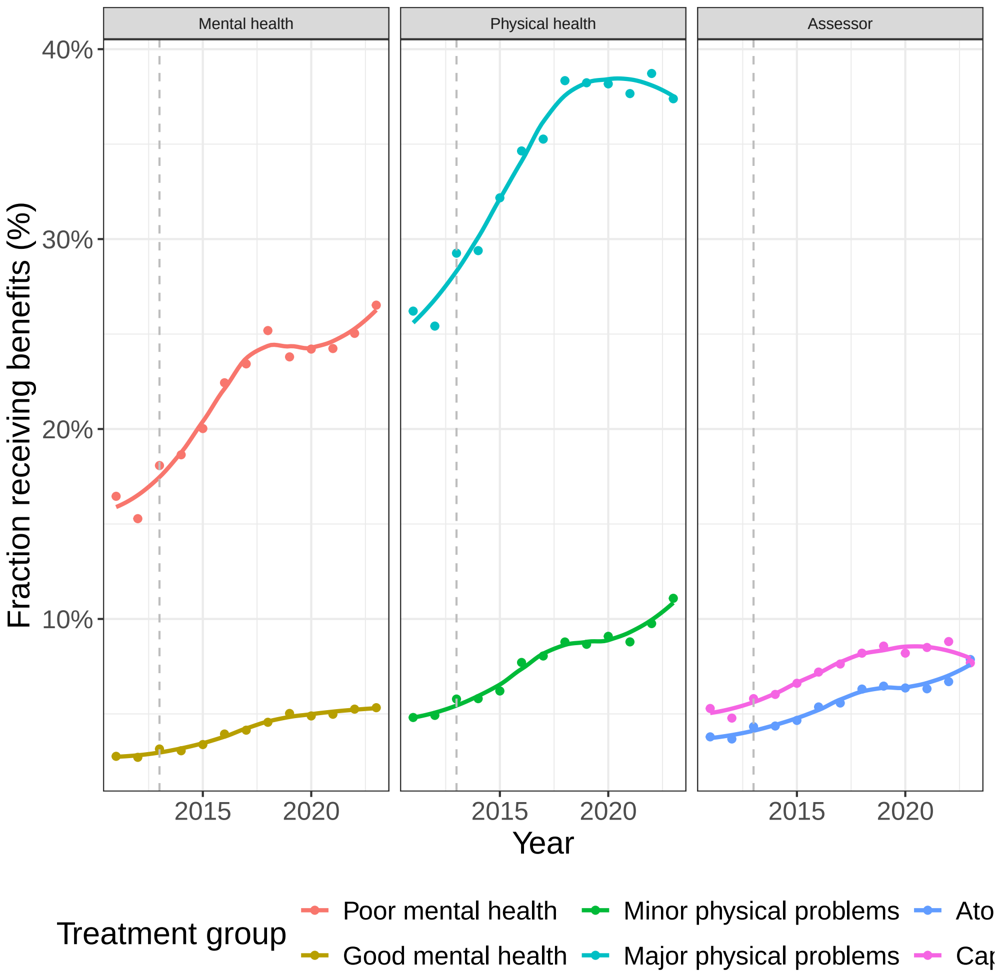
Panel A: Pre-treatment	**Mental Health**		**Physical Problems**		**Assessor**	
	Poor	Good	Major	Minor	Atos	Capita
PIP Benefits (%)	0.160	0.027	0.259	0.049	0.038	0.050
Employment (%)	0.461	0.758	0.394	0.695	0.730	0.701
Mental health problems (%)	0.767	0.028	0.467	0.144	0.106	0.117
Minor physical problems (%)	0.353	0.262	0.134	0.756	0.243	0.253
Major physical problems (%)	0.287	0.040	0.740	0.034	0.058	0.067
Female (%)	0.629	0.552	0.633	0.598	0.541	0.549
Age	38.877	38.481	42.892	39.276	38.589	38.090
Life satisfaction	3.680	5.173	3.847	4.781	5.008	5.021
Total income (monthly)	£1,246.52	£1,812.17	£1,242.76	£1,607.87	£1,847.32	£1,556.17
Observations	5,714	44,582	3,494	14,139	34,320	17,287
Panel B: Post-treatment	**Mental Health**		**Physical Problems**		**Assessor**	
	Poor	Good	Major	Minor	Atos	Capita
PIP Benefits (%)	0.148	0.028	0.227	0.051	0.037	0.051
Employment (%)	0.516	0.783	0.414	0.708	0.757	0.732
Mental health problems (%)	0.416	0.094	0.411	0.163	0.123	0.138
Minor physical problems (%)	0.363	0.299	0.335	0.454	0.291	0.312
Major physical problems (%)	0.281	0.069	0.432	0.114	0.086	0.098
Female (%)	0.648	0.563	0.648	0.609	0.558	0.564
Age	44.841	45.015	48.739	45.575	44.993	44.650
Life satisfaction	4.049	5.184	4.081	4.845	5.074	5.050
Total income (monthly)	£1,502.14	£2,191.68	£1,432.37	£1,901.67	£2,220.09	£1,913.76
Observations	19,753	164,602	12,530	52,136	127,840	58,746

Notes: The table presents summary statistics for treatment and control groups across three identification strategies using data from the UK Household Longitudinal Study (UKHLS) 2011-2023. Panel A shows pre-treatment characteristics (2011-2013) and Panel B shows post-treatment characteristics (2013-2023). Mental health groups compare individuals with poor mental health versus those with good mental health in the pre-treatment period. Physical health groups compare individuals with major physical problems versus minor physical problems in the pre-treatment period. Assessor groups are defined by regional assignment to Atos versus Capita assessment providers. PIP Benefits indicates receipt of either Disability Living Allowance (DLA) or Personal Independence Payment (PIP). The sample includes individuals aged 59 or younger in 2013, excluding the first two survey waves.

implemented. The large disparities between health-defined treatment and control groups persist, with individuals with poor mental health continuing to show benefit receipt rates over 5 times higher (14.8% versus 2.8%), employment rates 27 percentage points lower (51.6% versus 78.3%), and monthly incomes £690 lower than those with good mental health. Similarly, the gap between major and minor physical disability groups remains substantial, with benefit receipt differing by 18 percentage points (22.7% versus 5.1%) and employment rates by 29 percentage points (41.4% versus 70.8%). Notably, employment rates increase modestly and benefit receipt rates decline slightly across most groups, suggesting broader economic improvements during the study period.

Figure 1 shows the evolution of benefit receipt across all six treatment and control groups over the full 2011-2023 period. The graph shows that all groups exhibit upward trends in benefit receipt that reflect broader policy and economic changes. As expected, the health-based comparisons exhibit the substantial and persistent gaps anticipated from the summary statistics: individuals with poor mental health consistently show benefit receipt rates 4-5 times higher than those with

Figure 1: Evolution of benefits over time across groups

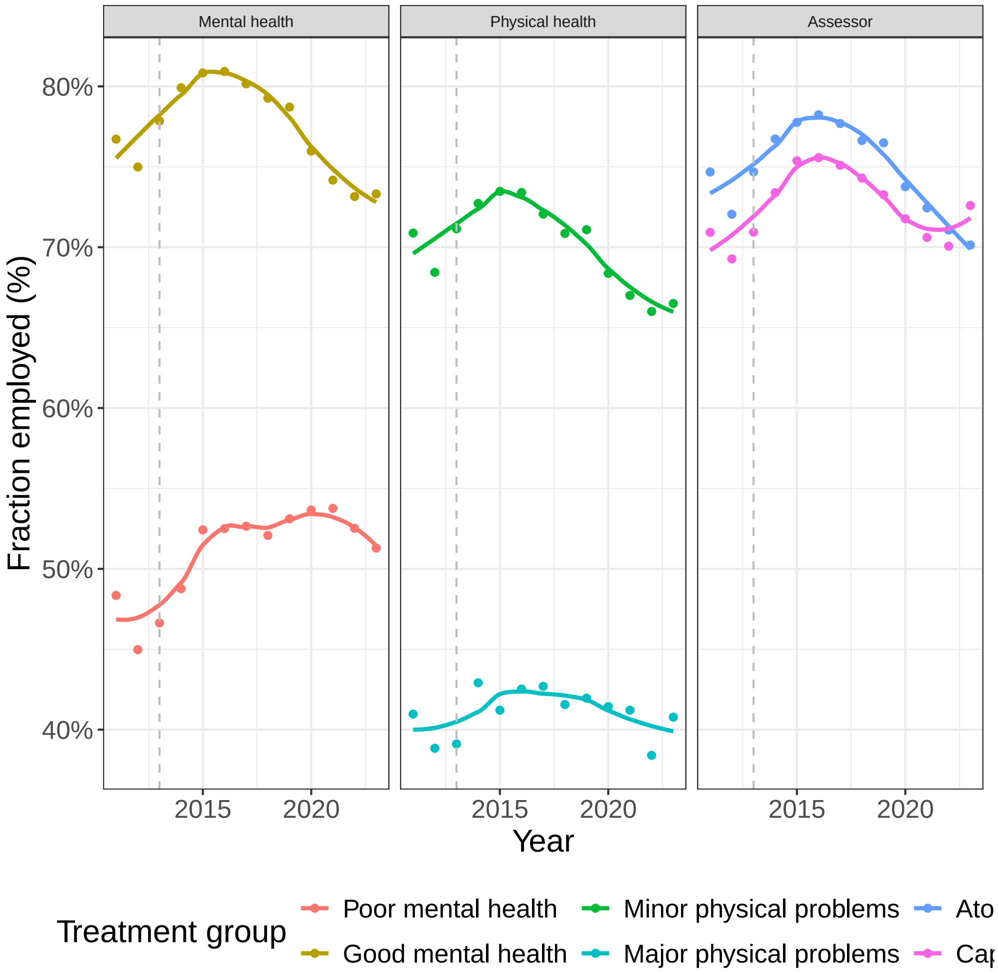


Notes: The figure shows the evolution of disability benefit receipt rates from 2011 to 2023 across three identification strategies. The left panel compares individuals with poor versus good pre-treatment mental health. The middle panel compares individuals with major versus minor pre-treatment physical health problems. The right panel compares regions assigned to Atos versus Capita assessment providers. Benefit receipt is measured as the percentage of individuals receiving either Disability Living Allowance (DLA) or Personal Independence Payment (PIP) in each year. All series are smoothed using locally weighted regressions (LOESS).

good mental health, while those with major physical problems maintain rates 5-6 times higher than those with minor problems. The assessor comparison continues to show the most balanced trajectories, with Capita-assigned regions maintaining a modest 1-2 percentage point higher benefit receipt rate than Atos regions throughout the period.

Figure 2 shows the corresponding employment trends. Employment rates generally exhibit an inverted-U pattern, rising modestly in the early-to-mid 2010s before declining after 2016, likely reflecting broader macroeconomic conditions including the post-financial crisis recovery and austerity

Figure 2: Evolution of employment over time across groups



Notes: The figure shows the evolution of employment rates from 2011 to 2023 across three identification strategies. The left panel compares individuals with poor versus good pre-treatment mental health. The middle panel compares individuals with major versus minor pre-treatment physical health problems. The right panel compares regions assigned to Atos versus Capita assessment providers. Employment is measured as the percentage of individuals in paid employment in each year. All series are smoothed using locally weighted regression (LOESS).

period. The health-based employment gaps are substantial, as individuals with good mental health maintain employment rates 25-30 percentage points higher than those with poor mental health, while those with minor physical problems consistently outperform those with major problems by similar margins. The assessor comparison again shows the most similar trajectories, with both Atos and Capita regions following nearly parallel paths, maintaining only small differences throughout the period.

4 Identification Strategy

4.1 Pre-Existing Conditions

Our identification strategy exploits differential policy impacts across disability categories created by the DLA-to-PIP transition. The reform simultaneously aimed to expand eligibility for individuals with mental health conditions while contracting eligibility for those with minor physical disabilities, creating quasi-experimental variation in benefit access.

We define our treatment groups based on pre-reform health conditions reported in 2012, the last year before the DLA-to-PIP transition. Individuals with mental health conditions are identified using four criteria from the Understanding Society data: reporting impairment from mental health issues in (1) daily activities, (2) work capacity, or (3) social functioning, or (4) regularly experiencing depression, bad nerves, or anxiety.[4] Minor physical disabilities are defined as those who report accomplishing less than desired with work or daily activities due to physical health 'some of the time' or 'a little of the time'.[5]

To increase the comparability of our treatment and control groups, we run our analysis on a matched sample of individuals with similar pre-treatment characteristics. To do so, we first estimate propensity scores for the likelihood of receiving disability benefits in the last pre-treatment period based on an individual's health limitations affecting work capacity, age, mental health status, total monthly income (log-transformed), gender, and educational attainment. Then, using nearest neighbor matching without replacement, we identify the five non-recipients with the most similar propensity scores for each benefit recipient. These matched groups form our analytical sample.

Our empirical approach follows two steps. First, we estimate the following event-study specification to verify that individuals with relevant pre-existing conditions indeed experience changes in benefit receipt:

$$Benefits_{it} = \sum_{k=-2}^{K} \tau^k \times T_i \times D_t^k + \alpha_i + \delta_t + \varepsilon_{it} \tag{1}$$

Here, T_i indicates whether individual i reported a treatment condition (mental health problems or minor physical disabilities) in 2012. D_t^k indicates that year t is k years from the 2013 reform

[4] People who answer 'somewhat regularly' to any of these questions are classified as having mental health problems.

[5] We classify as major physical disabilities those answering 'all of the time' or 'most of the time', and exclude those answering 'none of the time' from this analysis.

implementation, with $k \in \{-2, -1, 0, 1, \ldots, 6\}$.[6] We include individual fixed effects α_i and year fixed effects δ_t, with standard errors clustered at the individual level. We consider additional specifications where we control for age, age squared, and the number of children in a household.

Identification relies on the parallel trends assumption that both types of areas would have evolved along similar lines had PIP not been introduced. To test this, we first use our event-study specification to examine whether both areas evolved along similar lines in the pre-treatment period. Second, we consider placebo treatments using earlier data from the BHPS. Nonetheless, we cannot fully dismiss the concern that treated individuals suffered different unobserved shocks that correlate with the policy implementation.[7] We therefore supplement the current analysis with an additional identification strategy based on regional assessor assignment in the next section.

Second, we use this eligibility-induced variation as an instrument for actual benefit receipt to estimate the effect of benefits on employment:

$$\text{Benefits}_{it} = \pi \times \text{Condition}_i \times \text{Post}_t + \tilde{\alpha}_i + \tilde{\delta}_t + \tilde{\varepsilon}_{it} \tag{2}$$

$$Y_{it} = \theta \times \widehat{\text{Benefits}}_{it} + \bar{\alpha}_i + \bar{\delta}_t + \bar{\varepsilon}_{it} \tag{3}$$

where Post_t indicates years after reform implementation, and Condition_i indicates either mental health problems or minor physical disability pre-treatment. The coefficient θ measures the local average treatment effect of benefit receipt on employment for individuals whose benefit status was affected by the reform. We complement the 2SLS analysis with a reduced form analysis that estimates the direct effect of our condition-based treatments on employment outcomes.

The validity of our instrumental variable approach rests on an exclusion restriction assumption in addition to the parallel trends requirement. The exclusion restriction requires that pre-existing health conditions affect changes in employment outcomes solely through their impact on benefit receipt, not through alternative channels. This assumption could be violated if the DLA-to-PIP reform altered employer attitudes toward hiring individuals with mental health conditions or minor physical disabilities, or if the policy change itself influenced job search behavior or work capacity independent of benefit access. We view this concern as limited, however, as the reform primarily

[6] We consider the sample period until the start of COVID-19.

[7] The robustness checks in Section 5.4 dismisses most concerns.

changed assessment procedures rather than public awareness of these conditions, and our robustness checks using placebo periods show no systematic employment effects in the absence of benefit changes.

Beyond our main employment outcomes, we also examine whether policy-induced changes in PIP receipt generate offsetting responses in other social benefits. This analysis serves two purposes: testing whether our identification strategies capture specific disability benefit effects rather than broader policy or economic confounds, and examining whether benefit substitution might explain the employment effects we observe. We apply our difference-in-differences design to seventeen additional benefit categories available in the UKHLS data.

4.2 Assessor Assignment

To complement our condition-based identification strategy and address potential concerns about unobserved shocks affecting specific disability groups, we exploit quasi-random variation in assessment stringency across UK regions. Following the 2013 DLA-to-PIP transition, the Department for Work and Pensions (DWP) allocated assessment contracts geographically: Independent Assessment Services (formerly Atos) was assigned eight regions (North East, North West, Yorkshire and Humber, London, East of England, South East, South West, and Scotland), while Capita was assigned four regions (East Midlands, West Midlands, Wales, and Northern Ireland).

This administrative allocation was determined by bureaucratic and logistical considerations rather than regional economic conditions or disability prevalence. As such, the assignment creates plausibly exogenous variation in assessment stringency if the two providers apply systematically different evaluation criteria, allowing us to identify treatment effects through differences in benefit access across otherwise similar individuals.

We estimate the following event-study specification to examine both differential benefit assessment stringency and employment effects:

$$Y_{it} = \sum_{k=-2}^{K} \tau^k \times Atos_i \times D_t^k + \alpha_i + \delta_t + \varepsilon_{it} \tag{4}$$

where Y_{it} is either benefit receipt or employment for individual i in year t. $Atos_i$ indicates residence in a region assigned to Independent Assessment Services. All other variables are the same as before.

We focus on reduced-form estimates for the employment effects instead of 2SLS estimates, because the treatment effect of assessor regions on benefits is relatively small, which would create weak instrument problems for a 2SLS analysis. We rely on the same matched sample as in the previous section.

5 Results

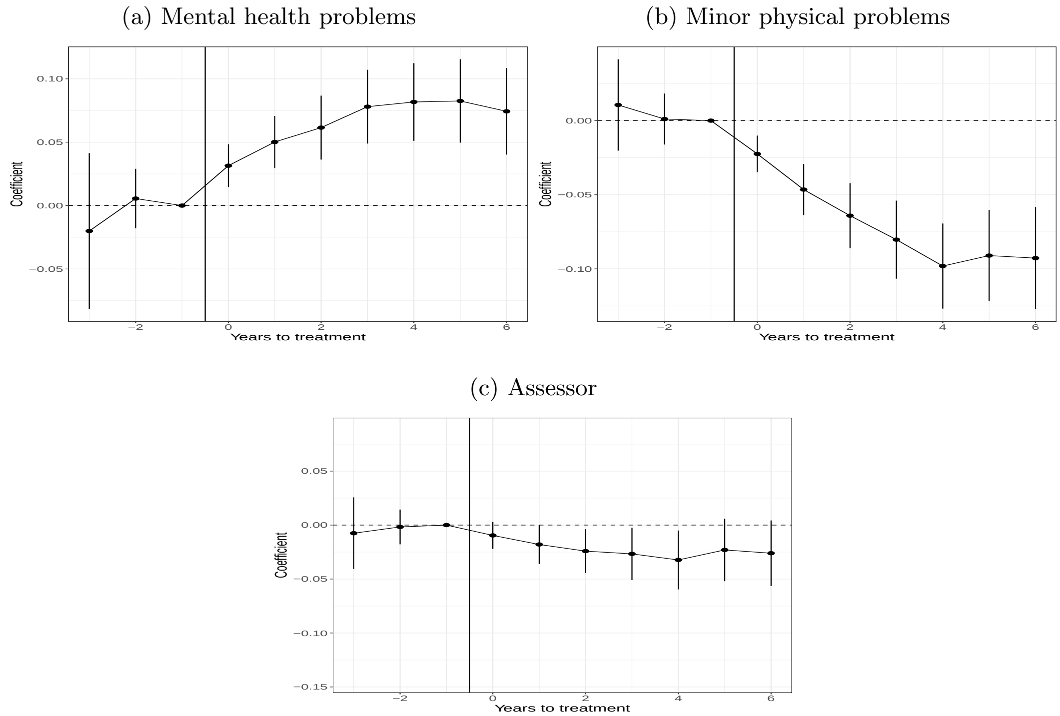
5.1 Pre-Existing Conditions

We begin by examining how the DLA-to-PIP reform affected benefit receipt and employment outcomes for individuals with pre-existing mental health conditions and minor physical disabilities. Panels A and B of Figure 3 present event study estimates showing how the reform affected benefit receipt for our treatment groups relative to matched control groups. The results confirm that the policy achieved its intended differential impacts across disability categories. For individuals with pre-existing mental health conditions, benefit receipt increases substantially following the reform implementation, whereas benefits gradually reduce for those with minor physical problems. The effects build gradually over time, consistent with the staggered rollout that began with new claimants in 2013 and incorporated existing DLA recipients through reassessment. All pre-reform coefficients are statistically indistinguishable from zero, supporting the parallel trends assumption.

Over the entire post-treatment period, individuals with mental health problems are on average 5.9 percentage points more likely to receive disability benefits compared to the control group ($p < 0.001$, see Table A17 in the Appendix). Conversely, individuals with minor physical disabilities experience a significant reduction in benefit access. The reform reduces their likelihood of receiving disability benefits by 5.6 percentage points ($p < 0.001$) relative to those with major physical disabilities.

Moving on to employment outcomes, Panels A and B of Figure 4 present the event study plots for the reduced-form employment effects for both treatment groups. The results provide little evidence that changes in benefit access (either increases or decreases) significantly affect employment outcomes. For individuals with mental health conditions who experienced expanded benefit access, we find no evidence of negative employment effects. If anything, the point estimates are consistently positive throughout the post-reform period. Panel A of Table 2 shows that the

Figure 3: Effect of treatments on likelihood of receiving PIP benefits



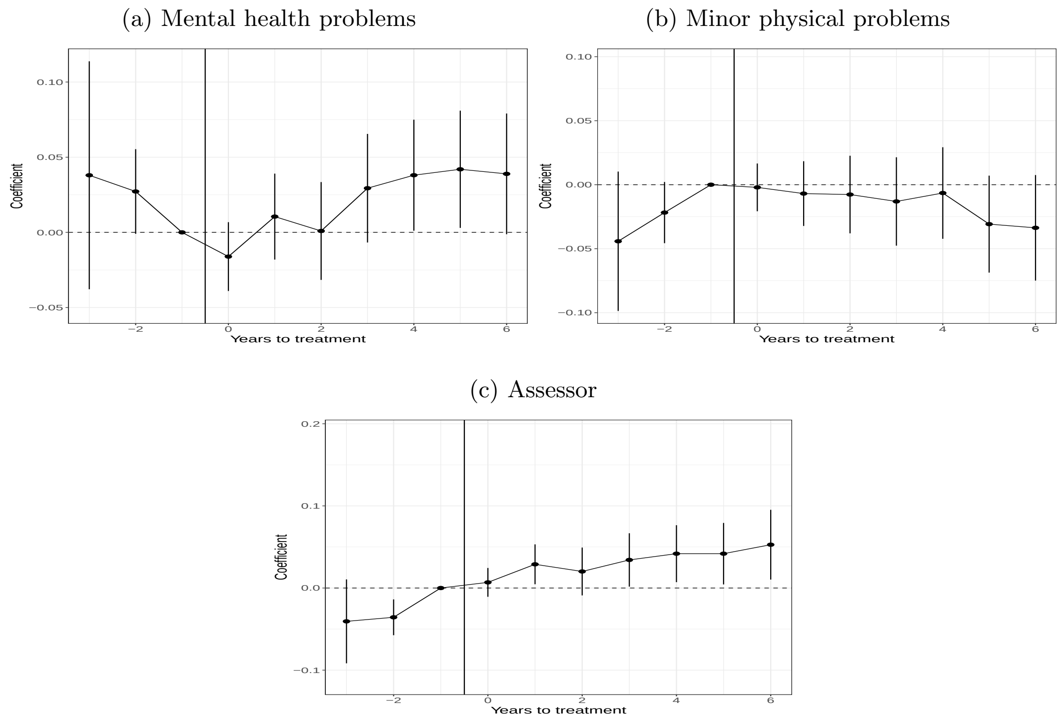
Notes: The figure shows event study estimates of the effect of three treatments on disability benefit receipt using matched difference-in-differences designs. Panel A compares individuals with pre-treatment mental health problems versus those with good mental health. Panel B compares individuals with pre-treatment minor physical problems versus those with major physical problems. Panel C compares regions assigned to Atos versus Capita assessment providers. The dependent variable is an indicator for receiving either Disability Living Allowance (DLA) or Personal Independence Payment (PIP). The horizontal axis shows years relative to the 2013 reform implementation, with negative values indicating pre-treatment periods. Point estimates and 95% confidence intervals are shown, with standard errors clustered at the individual level. All specifications include individual and year fixed effects and control for age, age-squared, and number of children.

average post-reform employment effect is statistically insignificant ($p > 0.218$).

Similarly, for individuals with minor physical disabilities who faced reduced benefit access, we find no significant employment response. Panel B of Figure 4 shows coefficients close to zero throughout the post-treatment period. The corresponding regression results in Panel B of Table 2 yield an insignificant treatment effect ($p > 0.807$).

To quantify the causal effect of benefit receipt on employment, we use the policy-induced variation in benefit eligibility as an instrument for actual benefit receipt. Panel A of Table 3 presents IV estimates for the mental health sample, yielding a coefficient of 0.276 ($p = 0.229$) in the baseline specification and 0.060 ($p = 0.781$) with additional controls. For the physical disability sample, Panel B shows IV coefficients of 0.050 ($p = 0.815$) and 0.051 ($p = 0.808$) across specifications. None

Figure 4: Treatment effects for employment



Notes: The figure shows event study estimates of the effect of three treatments on employment using matched difference-in-differences designs. The dependent variable is an indicator for being in paid employment. Treatment definitions, sample construction, and specification details are as in Figure 3. Point estimates and 95% confidence intervals are shown, with standard errors clustered at the individual level.

of these estimates are statistically significant, consistent with our reduced-form findings of minimal employment responses to benefit changes.

These minimal employment responses, despite substantial changes in benefit receipt, raise questions about potential offsetting mechanisms such as benefit substitution or broader policy confounds, which we examine systematically in Section 5.5.

5.2 Regional Assessor Assignment

Our second identification strategy exploits quasi-random variation in assessment stringency created by the administrative assignment of PIP evaluation contracts to different providers across UK regions. Panel C of Figure 3 shows that the two assessment providers apply systematically different evaluation standards. Individuals residing in regions assigned to Independent Assessment Services (Atos) face reduced benefit access compared to those in Capita-assigned regions. The effect becomes

Table 2: Treatment effects for employment

Panel A: Mental health problems (pre)		
Mental health problems (pre) x Post	0.016	0.003
	(0.013)	(0.013)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	23,074	23,074
Adjusted R-squared	0.689	0.697
Panel B: Minor physical health problems (pre)		
Minor physical disability (pre) x Post	−0.003	−0.003
	(0.012)	(0.011)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	18,764	18,764
Adjusted R-squared	0.697	0.704
Panel C: Assessor assignment		
Assessor = Atos x Post	0.029***	0.033***
	(0.011)	(0.011)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	21,125	21,125
Adjusted R-squared	0.688	0.696

Notes: The table shows difference-in-differences estimates of treatment effects on employment using matched samples. The dependent variable is an indicator for being in paid employment. Panel A compares individuals with pre-treatment mental health problems versus those with good mental health. Panel B compares individuals with pre-treatment minor physical problems versus those with major physical problems. Panel C compares regions assigned to Atos versus Capita assessment providers. All samples are constructed using nearest neighbor matching with five matches per treated unit, as described in Section 4. Column 1 includes no time-varying controls; Column 2 adds controls for age, age-squared, and number of children. All specifications include individual and year fixed effects. Standard errors are clustered at the individual level. *p<0.1; **p<0.05; ***p<0.01.

statistically significant two years after implementation and grows over time. Aggregated over the entire post-treatment period, assignment to Atos reduces the likelihood of receiving disability benefits by 1.8 percentage points relative to Capita assignment ($p = 0.028$, see Table A17 in the Appendix).

While this first-stage effect is statistically significant and economically meaningful, the resulting F-statistic falls below conventional thresholds for instrumental variable analyses, leading us to focus on reduced-form estimates that directly capture how otherwise similar individuals who live in different assessment regions see their labor market outcomes develop. Panel C of Figure 4 presents the reduced-form employment effects of regional assessor assignment. Unlike our condition-based results, we find clear evidence of employment responses to differential benefit access. Employment in Atos-assigned regions increases gradually relative to Capita-assigned regions, starting near zero in the immediate post-reform period and reaching approximately 5.2 percentage points by year 6

Table 3: The effect of benefits on employment, instrumented difference-in-differences

Panel A: Mental health problems (pre)		
Benefits (instrumented)	0.276	0.060
	(0.229)	(0.216)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	23,074	23,074
Adjusted R^2	0.671	0.695
Panel B: Minor physical health problems (pre)		
Benefits (instrumented)	0.050	0.051
	(0.212)	(0.210)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	18,764	18,764
Adjusted R^2	0.696	0.703

Notes: The table shows instrumental variable estimates of the causal effect of disability benefit receipt on employment for individuals with pre-existing conditions. The dependent variable is an indicator for being in paid employment. Benefit receipt (instrumented) is an indicator for receiving either Disability Living Allowance (DLA) or Personal Independence Payment (PIP), instrumented using the policy-induced variation in benefit eligibility for individuals with pre-existing mental health problems (Panel A) or pre-existing minor physical health problems (Panel B) following the DLA-to-PIP reform. Other definitions are as in Table A17.

after implementation. It should be noted that one pre-period estimate is statistically significant, suggesting some caution in the interpretation of these results.

Panel C of Table 2 presents the corresponding difference-in-differences estimates. Our baseline specification shows that individuals in Atos-assigned regions experience a 2.9 percentage point increase in employment probability relative to those in Capita-assigned regions ($p = 0.008$). Adding time-varying individual controls increases the coefficient to 3.3 percentage points while maintaining high statistical significance ($p = 0.003$). These results indicate economically meaningful employment responses to changes in benefit assessment stringency.

The contrast between our assessor-based and condition-based findings provides insight into the mechanisms underlying disability benefit employment effects. The positive employment response in regions with stricter assessment procedures may reflect both standard income effects among individuals with marginal work capacity and the filtering out of strategic claimants who were seeking benefits despite retaining substantial ability to work. Unlike our condition-based strategies that examine individuals whose eligibility changed based on objectively measurable pre-existing health conditions, the assessor variation affects benefit access through evaluation stringency, potentially selecting against applicants with weaker or more ambiguous claims.

This interpretation suggests that apparent work disincentives from disability benefits may partly reflect selection effects and strategic behavior rather than pure responses to income transfers. The absence of employment effects in our condition-based analysis, where scope for strategic misrepresentation is more limited, supports this view and indicates that when benefit changes are based on genuine health status rather than assessment procedures, employment responses may be minimal.

5.3 Heterogeneity

To better understand the mechanisms underlying our employment effects and assess the external validity of our findings, we examine heterogeneity across demographic characteristics that theory suggests may moderate responses to benefit changes. If differential assessment stringency operates primarily through income effects, we would expect larger responses among individuals with lower baseline incomes or weaker alternative support systems. Alternatively, if effects operate through changing perceptions of benefit accessibility, responses may vary systematically with demographic groups that differ in their interaction with social services. We examine heterogeneous treatment for both benefit receipt and employment outcomes.

Starting with the former, Figure A15 in the Appendix presents first-stage results by demographic subgroups for the mental health treatment. The increase in benefit receipt is consistently strong across all groups, with little variation in magnitudes across age, gender, and marital status. Similarly, the results for the minor physical problem treatment in Figure A17 show that the reduction in benefit receipt is most pronounced among older individuals (-7.4 p.p, $p < 0.002$) and least pronounced for young people (-3.9 p.p., $p < 0.001$), although the magnitudes are reasonably consistent across categories, and the effects are statistically significant for all. For the assessor region strategy, Figure A13 shows that the reduction in benefit receipt under Atos assessment is most pronounced among older individuals and women. This pattern suggests that more stringent assessment procedures may disproportionately affect groups that traditionally have higher baseline benefit receipt rates.

Moving on to the employment effects, Figures A16 and A18 shows no evidence of employment impacts across any demographic subgroup for the pre-existing condition treatments. While effects for the mental health group appear slightly larger for males than females, the difference is not statistically significant. The absence of heterogeneity in employment responses reinforces our conclusion that gaining or losing PIP-style benefits does not systematically discourage work.

For the stricter assessment treatment, Figure A14, we find little to no effect of the treatment on employment for most subgroups, with one clear exception: unmarried individuals. For this group, the employment effect is large (6.0 p.p., $p = 0.001$), which is four times as large as that of married individuals (1.5 p.p., $p = 0.239$). This result suggests that spousal income may buffer the employment effects of benefit changes, consistent with standard models of household labor supply (Autor et al., 2019).

5.4 Robustness Checks

We conduct several robustness checks to validate our identification strategies and assess potential threats to the interpretation of our results. These checks address three main concerns: whether our results reflect pre-existing differences rather than policy effects, the sensitivity of our findings to methodological choices, and the possibility of confounding factors that could bias our estimates. In addition to these robustness checks, Section 5.5 examines whether our treatments affect other social benefits unrelated to disability policy, which could suggest possible changes in underlying unobserved characteristics that correlate with our treatments.

First, we implement placebo tests using pre-reform BHPS data to verify parallel trends and rule out pre-existing regional differences (see Appendix A1.1 for methodological details). Figure A1 and Figure A2 show that 17 of 18 placebo coefficients across all identification strategies are statistically indistinguishable from zero and economically small. The single significant estimate likely reflects Type I error given the multiple testing context, and the overall pattern strongly supports the validity of our identification assumptions.

Second, we test whether our assessor results reflect broader regional economic shocks by re-estimating effects for healthy individuals only.[8] Tables A1 and A2 show no significant effects of assessor assignment on either benefits or employment for this population, providing evidence against confounding regional economic factors.

Third, we assess sensitivity to our matching procedure. Our main specification matches each treated unit to five control units based on pre-treatment characteristics. Tables A3 and A4 show that first-stage effects remain strong and significant when using either one or ten nearest neighbors,

[8] Healthy individuals are those who report their general health as either 'excellent' or 'very good'.

while employment effects maintain the same patterns as in our main results across specifications.

Fourth, we present results for our full unmatched sample. For the assessor identification strategy, Tables A7 and A8 show qualitatively similar results with parallel pre-trends. However, for the condition-based treatments, while first-stage effects remain strong, the event study plots in Figure A7 reveal pre-treatment trend divergence, which supports our decision to focus on the matched sample for these analyses.

Fifth, we examine sensitivity to alternative definitions of physical disability severity. We test moving individuals who report limitations "some of the time" from the minor to major physical health category. Appendix A1.6 shows highly similar results, indicating our findings are not sensitive to the precise threshold between minor and major health problems.

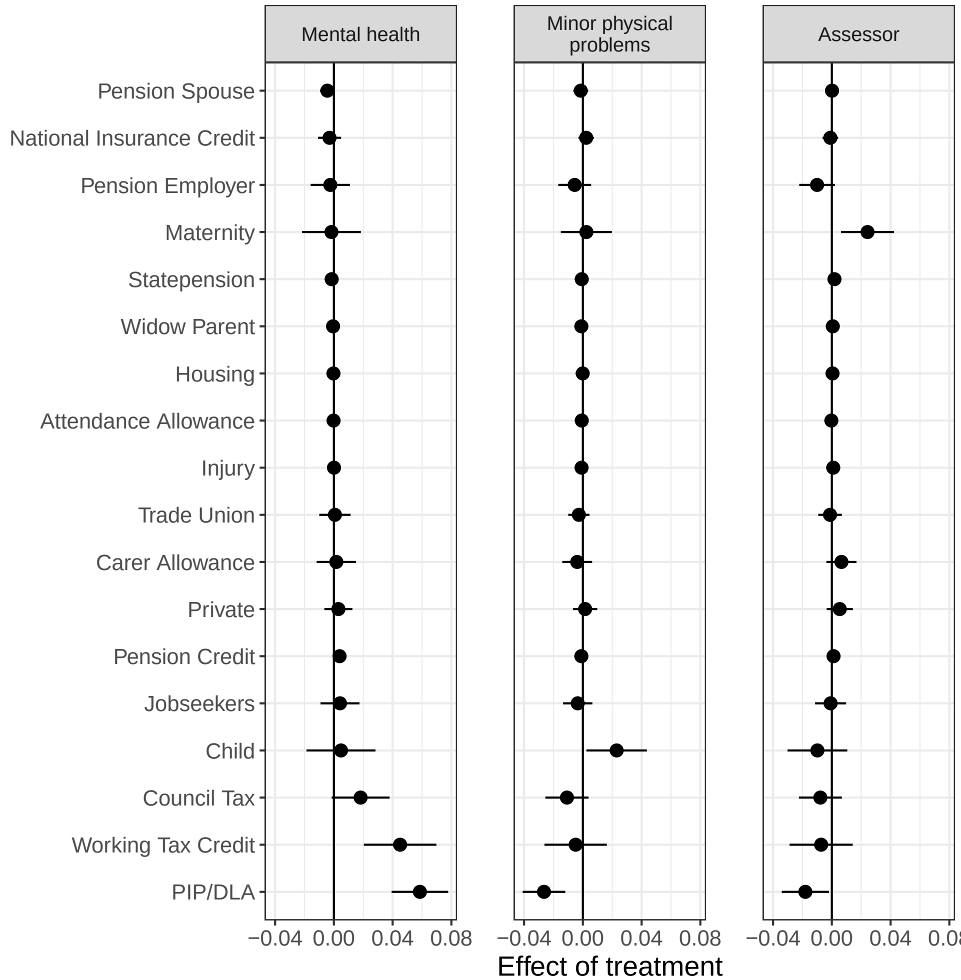
Finally, we test whether strategic relocation drives our assessor results by excluding the 3.2% of individuals who moved between Atos and Capita regions during our sample period. Tables A15 and A16 show that while first-stage results remain similar, the significant employment effect disappears. This finding suggests that strategic mobility may partially explain our assessor-based employment results, warranting caution in their interpretation and highlighting that even administratively determined assignments may not be entirely exogenous to individual behavior.

5.5 Do Disability Benefits Crowd Out Other Benefits?

Having established that our three identification strategies generate meaningful first-stage variation in disability benefit receipt, we examine whether these policy-induced changes affect receipt of other social benefits. This analysis serves two purposes: first, it tests whether disability benefits crowd out other forms of social support, which could partially explain the modest employment effects observed in previous sections (Borghans et al., 2014); second, it provides a validation check for our identification strategies by examining whether our treatments generate spurious effects on benefits unrelated to disability policy, which would suggest contamination from broader policy changes or economic shocks.

We examine effects on all benefit categories available in the UKHLS data except those pertaining to students, and benefits that were discontinued or underwent major reforms during our study period. Specifically, we exclude widow benefits, war disability benefits, income support, education grants, and incapacity benefits. This exclusion makes sure that our analysis focuses on benefits that

Figure 5: Effect of treatments on other benefits



Notes: The figure shows difference-in-differences estimates of the effects of our three treatments on receiving any type of benefit, using the matched sample described in Section 4. Each point estimate represents the coefficient from a separate regression with the respective benefit category as the dependent variable. Point estimates and 95% confidence intervals are shown, with standard errors clustered at the individual level. All specifications include individual and year fixed effects, and control for age, age-squared, and number of children.

remained stable in structure and eligibility throughout the 2011-2019. Our final sample consists of seventeen additional benefits.

Figure 5 presents the effects of our three identification strategies on seventeen categories of social benefits beyond PIP/DLA. We additionally report the PIP/DLA results for comparison. Across all approaches, the PIP/DLA estimate is either the most negative (assessor and physical problem treatments) or the most positive (mental health treatment). Of the remaining 51 estimates across alternative benefit categories, only three are statistically significant (maternity benefits for the assessor treatment, working tax credit for the mental health treatment, and child benefits for the physical disability treatment). However, these effects are smaller in magnitude than the corresponding changes in PIP/DLA receipt and do not follow systematic patterns suggesting widespread substitution across benefit categories. The absence of significant effects on other benefits—including council tax, housing, job seekers allowance, and national insurance credits—provides strong evidence that our identification strategies do not pick up other factors such as austerity, broader economic shocks, or policy confounds.

To examine potential substitution effects more extensively, we also analyze the monetary amounts across six aggregated income categories, namely social benefit income (our primary treatment), private benefit income, miscellaneous income, pension income, investment income, and net labor income. If disability benefits crowd out other social programs, we would expect to see offsetting reductions in alternative benefit categories when individuals gain or lose access to PIP.

Figure A19 in the Appendix present the effects of our three identification strategies on various income categories. Across all approaches, we observe clear first-stage effects on social benefit income, which validates our prior results: regions assigned to the more stringent Atos assessor show reduced social benefit income (approximately £29 annually), individuals with mental health conditions experience increased benefit receipt (approximately £43 annually), and those with minor physical disabilities face benefit reductions (approximately £41 annually). For all other income categories, we observe mostly null effects. The only exception is a positive and significant effect for labor income in the assessor treatment, which is consistent with the positive employment effects found in Section 5.2.

Taken together, the limited evidence of benefit substitution strengthens our identification strategy and suggests that people who lose/gain PIP do not substitute away from other types of benefits. These findings reinforce our interpretation that the modest employment effects documented in previous sections reflect responses to income transfers rather than artifacts of identification strategies or offsetting changes in alternative social support mechanisms.

6 Conclusion

This paper examines whether disability benefits necessarily reduce employment by exploiting the unique design of the UK's Personal Independence Payment program. Unlike traditional disability insurance that conditions benefits on non-employment or low earnings, PIP provides income support without work restrictions, allowing recipients to work at any earnings level without benefit reduction. This design enables us to isolate pure income effects from substitution effects. Using quasi-experimental variation from differential eligibility changes across health conditions and regional assessor assignment, we provide some of the first credible evidence on employment effects of disability benefits when work disincentives are removed by design.

We find mixed employment responses despite substantial changes in benefit receipt. Using our primary condition-based identification strategy, individuals with pre-existing mental health conditions who experienced a 5.9 percentage point increase in benefit receipt show no evidence of reduced employment, with point estimates that are, if anything, slightly positive. Similarly, those with minor physical disabilities who faced a 5.6 percentage point reduction in benefits show no employment gains. These findings provide evidence that improved access to disability benefits need not discourage labor market participation. However, our complementary assessor assignment strategy shows a contrasting pattern: regions assigned to more stringent assessment procedures experience a significant 2.9 to 3.3 percentage point increase in employment probability when benefits decrease by 1.8 percentage points, suggesting that employment responses may depend on the mechanism through which benefit access changes.

One potential interpretation of the heterogeneous effects across identification strategies is that stricter assessors reduce the scope for strategic misrepresentation, whereas the other treatments, which rely on pre-existing conditions, do not. The strategic misrepresentation concern is that people seek out PIP benefits by exaggerating their health problems or functional limitations in order to receive benefits when they are capable of working. This has been a major concern in recent policy debates about PIP (Department for Work and Pensions, 2025). Stricter assessment procedures may deter such behavior or filter out individuals with weaker claims, leading to the observed employment increases in Atos regions as those with marginal work capacity are denied benefits and return to employment. In contrast, our condition-based identification strategies examine individuals whose eligibility changes were based on objectively measurable pre-existing health conditions rather than assessment stringency, reducing the scope for strategic manipulation of benefit receipt. These findings indicate that apparent work disincentives from disability benefits may partly reflect selection effects and strategic behavior rather than pure responses to income transfers. Formally investigating strategic behavior in this context would be an interesting direction for future research.

Our findings have important implications for optimal disability insurance design. Low and Pistaferri (2020) argue that US disability benefit reforms should be focused on reducing false rejections and supporting labor market attachment. Since our results suggest that PIP benefits do not have significant labor market effects for people with pre-existing mental health problems or minor physical disabilities, PIP appears to be quite an efficient welfare program. However, our assessor results also suggest that stricter assessment procedures may be a reasonable policy option to reduce strategic misrepresentation and improve targeting, as tighter eligibility screening appears to successfully return individuals with marginal work capacity to employment. The absence of negative employment effects among mental health beneficiaries or people with pre-existing physical disabilities point toward a system where work capacity and benefit receipt can be optimally separated. For UK disability insurance design, this suggests maintaining PIP's structure of work-neutral benefits while balancing two policy objectives: potentially expanding coverage for individuals with objectively verifiable conditions where strategic misrepresentation concerns are minimal, while simultaneously implementing more rigorous assessment procedures to better screen out strategic claimants. The efficiency costs of providing support appear minimal when beneficiaries retain full flexibility to work according to their capabilities and circumstances, but only when the system successfully distinguishes between genuine need and strategic behavior.

Several caveats are worth noting. First, while PIP eliminates formal work restrictions, implementation gaps may create implicit disincentives if assessors view employment as evidence of reduced need. The high appeal rates and assessment inconsistencies documented in program evaluations suggest such gaps may partially explain why our estimated effects, while small, are not zero. Second, our analysis captures short- to medium-term responses; longer-term effects may differ as individuals and employers fully adjust. Finally, our design cannot fully address whether individuals strategically adjust their reported conditions in response to eligibility changes, a concern explicitly raised by policymakers that merits future research.

Despite these limitations, our study provides compelling evidence that disability benefits need not create substantial work disincentives when properly designed. By removing the artificial link between benefit receipt and work incapacity, programs like PIP can provide income security while preserving employment incentives and individual agency. As aging populations and rising disability caseloads strain social insurance systems worldwide, our findings suggest that reimagining disability support as a supplement for extra costs rather than a replacement for earnings offers a promising path forward.

References

Autor, D., Duggan, M., Greenberg, K. and Lyle, D. S. (2016), 'The impact of disability benefits on labor supply: Evidence from the VA's disability compensation program', *American Economic Journal: Applied Economics* **8**(3), 31–68.

Autor, D. H. and Duggan, M. G. (2003), 'The rise in the disability rolls and the decline in unemployment', *Quarterly Journal of Economics* **118**(1), 157–205.

Autor, D. H. and Duggan, M. G. (2007), 'Distinguishing income from substitution effects in disability insurance', *American Economic Review* **97**(2), 119–124.

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

Banks, J., Blundell, R. and Emmerson, C. (2015), 'Disability benefit receipt and reform: Reconciling trends in the united kingdom', *Journal of Economic Perspectives* **29**(2), 173–190.

Becker, S., Gehlen, A., Geyer, J. and Haan, P. (2024), Income effects of disability benefits, Technical report, IZA Discussion Papers.

Borghans, L., Gielen, A. C. and Luttmer, E. F. (2014), 'Social support substitution and the earnings rebound: Evidence from a regression discontinuity in disability insurance reform', *American economic Journal: economic policy* **6**(4), 34–70.

Bound, J. (1989), 'The health and earnings of rejected disability insurance applicants', *American Economic Review* **79**(3), 482–503.

Cook, J. B. and East, C. N. (2023), The effect of means-tested transfers on work: evidence from quasi-randomly assigned snap caseworkers, Technical report, National Bureau of Economic Research.

Cook, J. B. and East, C. N. (2024), Work requirements with no teeth still bite: Disenrollment and labor supply effects of snap general work requirements, Technical report, National Bureau of Economic Research.

Cribb, J., Karjalainen, H. and Waters, T. (2022), Living standards of working-age disability benefits recipients in the uk, Technical report, IFS Working Paper.

Department for Work and Pensions (2012), Personal independence payment: assessment thresholds and consultation, Impact assessment, Department for Work and Pensions, London.

Department for Work and Pensions (2015), 'Personal independence payment: Official statistics'. Available at `https://www.gov.uk/government/statistics/personal-independence-payment-official-statistics`.

Department for Work and Pensions (2025), 'Pathways to work: Reforming benefits and support to get britain working
green paper'. Presented to Parliament by the Secretary of State for Work and Pensions.
URL: *https://www.gov.uk/government/consultations/pathways-to-work-reforming-benefits-and-support-to-get-britain-working-green-paper/pathways-to-work-reforming-benefits-and-support-to-get-britain-working-green-paper*

Deshpande, M. (2016*a*), 'Does welfare inhibit success? the long-term effects of removing low-income youth from the disability rolls', *American Economic Review* **106**(11), 3300–3330.

Deshpande, M. (2016*b*), 'The effect of disability payments on household earnings and income: Evidence from the ssi children's program', *Review of Economics and Statistics* **98**(4), 638–654.

Deuchert, E. and Eugster, B. (2019), 'Income and substitution effects of a disability insurance reform', *Journal of Public Economics* **170**, 1–14.

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

Garcia-Mandicó, S., García-Gómez, P., Gielen, A. C. and O'Donnell, O. (2020), 'Earnings responses to disability insurance stringency', *Labour Economics* **66**, 101880.

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

Gray, P. (2016), *The Second Independent Review of Personal Independence Payment Assessment*, Dandy Booksellers.

Gromadzki, J. (2024), 'Labor supply effects of a universal cash transfer', *Journal of Public Economics* **239**, 105248.

Hoynes, H. W. and Schanzenbach, D. W. (2012), 'Work incentives and the food stamp program', *Journal of Public Economics* **96**(1-2), 151–162.

Jones, D. and Marinescu, I. (2022), 'The labor market impacts of universal and permanent cash transfers: Evidence from the alaska permanent fund', *American Economic Journal: Economic Policy* **14**(2), 315–340.

Kostol, A. R. and Mogstad, M. (2014), 'How financial incentives induce disability insurance recipients to return to work', *American Economic Review* **104**(2), 624–655.

Latimer, E., Pflanz, F. and Waters, T. (2024), *Health-related benefit claims post-pandemic: UK trends and global context*, number R333, IFS Report.

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

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

Maestas, N., Mullen, K. J. and 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.

Moore, T. J. (2015), 'Disability insurance and the dynamics of the incentive-insurance tradeoff', *American Economic Review* **105**(10), 2986–3029.

OECD (2010), Sickness, disability and work: Breaking the barriers. a synthesis of findings across oecd countries, Technical report, Organisation for Economic Co-operation and Development.

Sauval, M., Duncan, G. J., Gennetian, L. A., Magnuson, K. A., Fox, N. A., Noble, K. G. and Yoshikawa, H. (2024), 'Unconditional cash transfers and maternal employment: Evidence from the baby's first years study', *Journal of Public Economics* **236**, 105159.

University of Essex, Institute for Social and Economic Research (2024), 'Understanding society: Waves 1-14, 2009-2023 and harmonised bhps: Waves 1-18, 1991-2009'. SN: 6614.
URL: *http://doi.org/10.5255/UKDA-SN-6614-20*

Verlaat, T., Todeschini, F. and Ramos, X. (2023), 'The employment effects of generous and unconditional cash support'.

von Wachter, T., Song, J. and Manchester, J. (2011), 'Trends in employment and earnings of allowed and rejected applicants to the social security disability insurance program', *American Economic Review* **101**(7), 3308–3329.

A1 Robustness Checks

A1.1 Placebo tests BHPS data

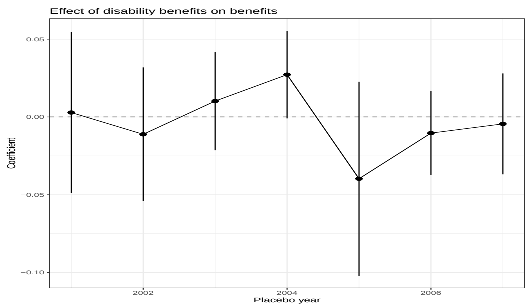
We conduct a set of placebo tests using data from the British Household Panel Survey (BHPS), the precursor survey to our main data source (UKHLS), from 1998-2008. This period falls well before the DLA-to-PIP reform was implemented, which allows us to conduct placebo tests for our main treatments in the absence of any actual policy change. Since the regional assessor assignments were not operative during this earlier period, we should observe no systematic differences in Disability Living Allowance receipt between treatment and control groups if our identification strategy is valid.

We estimate placebo treatment effects for hypothetical reform years implemented in any year between 2001 to 2007 using BHPS data. For these reforms, we use the same regional treatments as in our main analysis. Because the survey questions are not identical, we use slightly different variables than in our main specification. To measure benefits, we now exclusively use a dummy variable that takes the value of 1 if someone receives disability living allowance payments. We measure mental health problems by a binary question asking whether someone experiences health problems related to anxiety or depression. We measure minor physical health problems by people who report to experience 'fair' or 'good' health over last 12 months, with major problems referring to those who report 'poor' or 'very poor' health. All other variables remain the same as in our main analysis. We again estimate a difference-in-differences model for a matched sample that is matched on whether health limits their ability to work, age, mental health problems, total income, gender, and educational qualification.

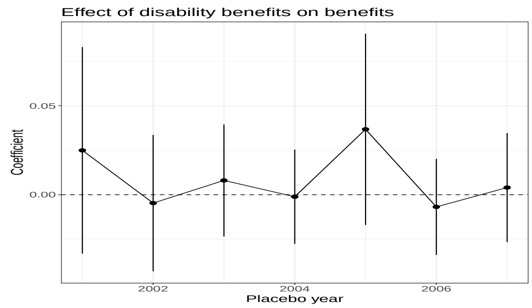
For each placebo check, we measure treatment conditions in the last year before the placebo treatment. For example, when we run the placebo treatment for 2005, we define pre-existing region, mental health, and physical health based on 2004 values.

Figure A1: Placebo tests benefits

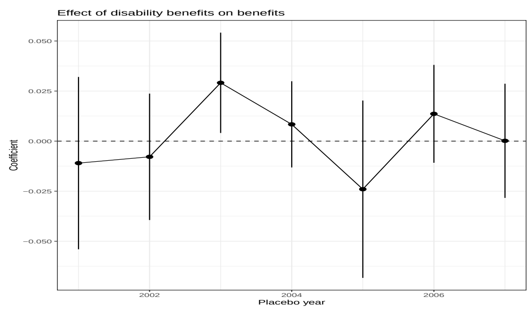
(a) Mental health problems



(b) Minor physical problems



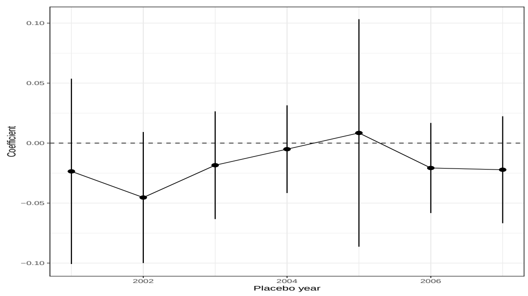
(c) Assessor



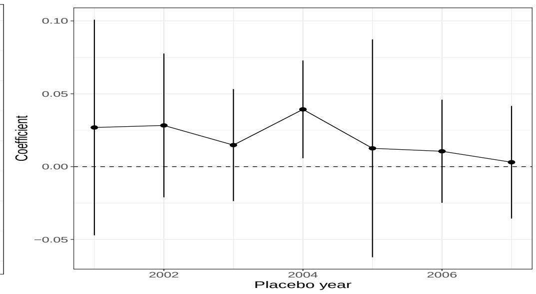
Notes: The figure shows placebo test estimates for hypothetical reform years between 2002 and 2007 using data from the British Household Panel Survey (BHPS), 1998-2008. Panel A compares individuals with mental health problems (anxiety or depression) in the last period before the placebo treatment versus those without. Panel B compares individuals with minor physical health problems pre treatment (reporting 'fair' or 'good' health) versus those with major problems (reporting 'poor' or 'very poor' health). Panel C tests the effect of regional assessor assignment (Atos versus Capita regions) on disability benefit receipt in periods before the actual DLA-to-PIP reform was implemented. The dependent variable is an indicator for receiving Disability Living Allowance (DLA). All estimates use matched difference-in-differences designs with samples matched on health limitations affecting work capacity, age, mental health status, total income, gender, and educational qualification. Point estimates and 95% confidence intervals are shown, with standard errors clustered at the individual level.

Figure A2: Placebo tests employment

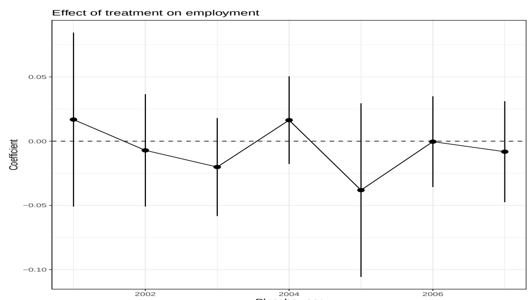
(a) Mental health problems



(b) Minor physical problems



(c) Assessor



Notes: The figure shows placebo test estimates for hypothetical reform years between 2002 and 2007 using employment as the dependent variable. Panel A tests the effect of regional assessor assignment on employment rates, Panel B compares employment outcomes between individuals with and without pre-treatment mental health problems, and Panel C compares employment outcomes between individuals with minor versus major physical health problems. The dependent variable is an indicator for being in paid employment. All other definitions, sample construction, and specification details are as in Figure A1.

A1.2 Assessor effects for healthy people

Table A1: The effect of assessor on benefits, healthy individuals

Assessor = Atos x Post	−0.0003	0.0001
	(0.014)	(0.014)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	14,498	14,498
Adjusted R-squared	0.729	0.729

Notes: The table shows difference-in-differences estimates of the assessor treatment on disability benefits, using only healthy individuals. All other definitions are as in Table A3.

Table A2: The effect of benefits on employment, healthy individuals

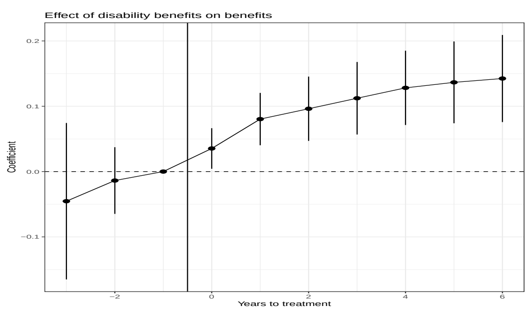
Benefits (instrumented)	27.765	14.627
	(700.001)	(208.643)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	14,280	14,280
Adjusted R^2	−191.469	−52.735

Notes: The table shows difference-in-differences estimates of the assessor treatment on employment, using only healthy individuals. All other definitions are as in Table A3.

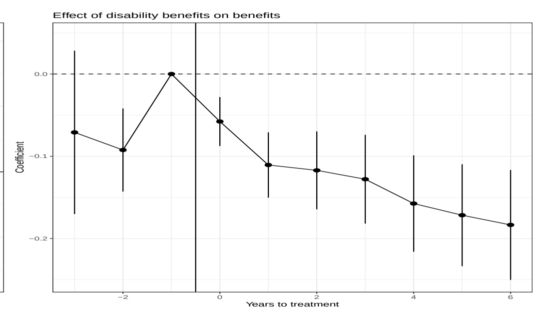
A1.3 Matching variations

Figure A3: Effect of treatments on likelihood of receiving PIP benefits, 1 match

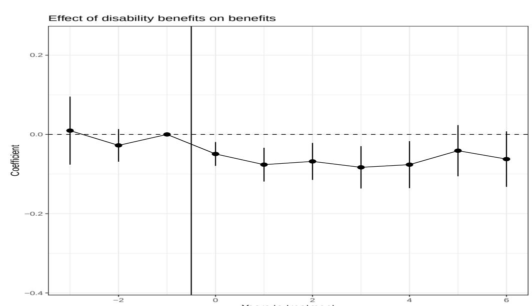
(a) Mental health problems



(b) Minor physical problems



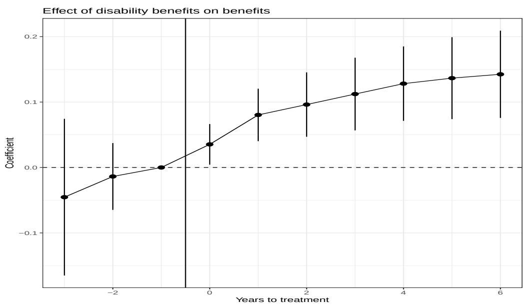
(c) Assessor



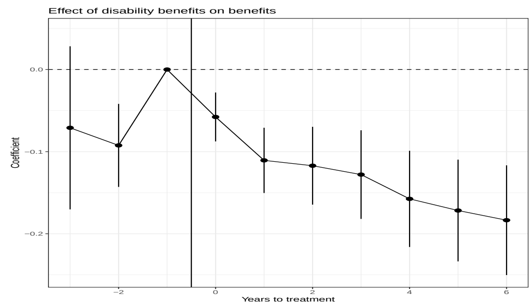
Notes: The figure shows event study estimates of the effect of three treatments on disability benefit receipt using matched samples with one nearest neighbor match per treated unit. Panel A compares individuals with pre-treatment mental health problems versus those with good mental health. Panel B compares individuals with pre-treatment minor physical problems versus those with major physical problems. Panel C compares regions assigned to Atos versus Capita assessment providers. The dependent variable is an indicator for receiving either Disability Living Allowance (DLA) or Personal Independence Payment (PIP). The horizontal axis shows years relative to the 2013 reform implementation, with negative values indicating pre-treatment periods. Point estimates and 95% confidence intervals are shown, with standard errors clustered at the individual level. All specifications include individual and year fixed effects and control for age, age-squared, and number of children.

Figure A4: Effect of treatments on employment, 1 match

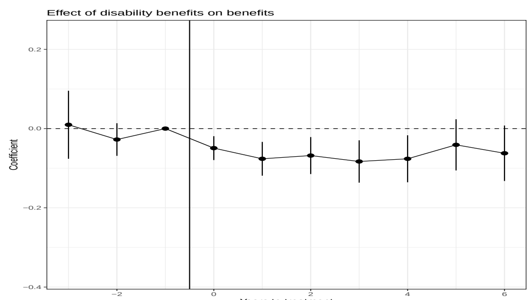
(a) Mental health problems



(b) Minor physical problems

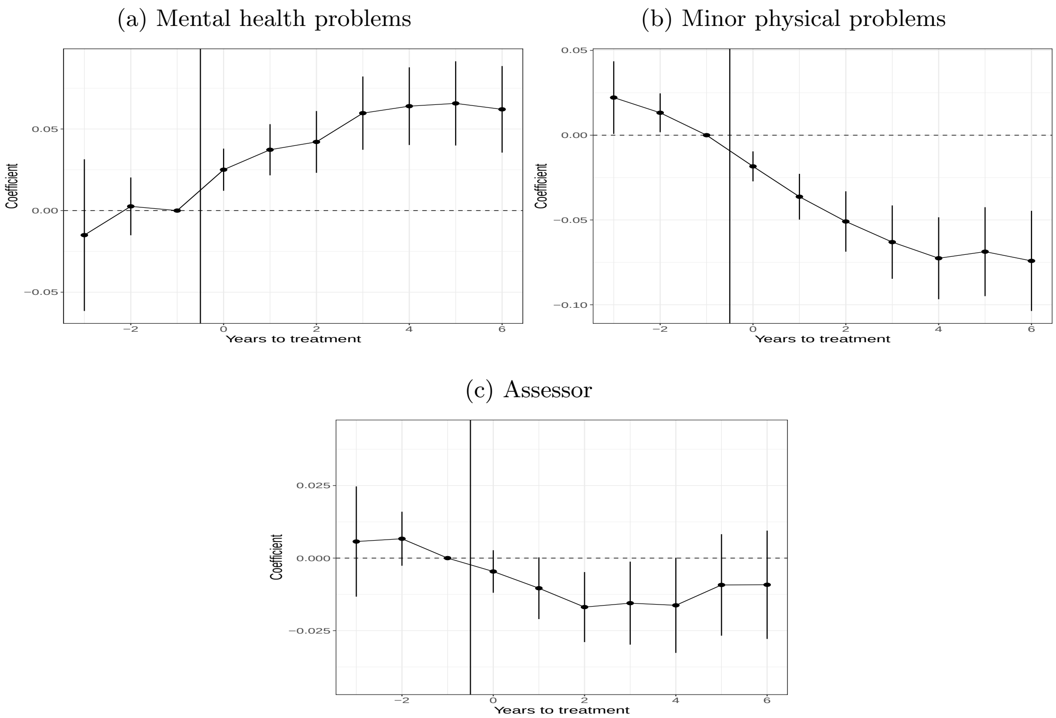


(c) Assessor



Notes: The figure shows event study estimates of the effect of three treatments on employment using matched samples with one nearest neighbor match per treated unit. All other definitions are as in Figure A3.

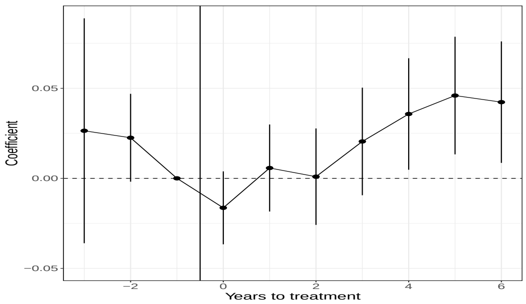
Figure A5: Effect of treatments on likelihood of receiving PIP benefits, 10 matches



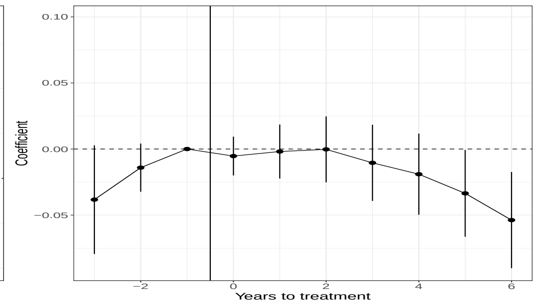
Notes: The figure shows event study estimates of the effect of three treatments on disability benefit receipt using matched samples with 10 nearest neighbor matches per treated unit. All other definitions are as in Figure A3.

Figure A6: Effect of treatments on employment, 10 matches

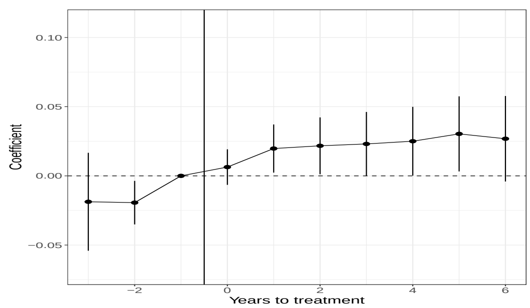
(a) Mental health problems



(b) Minor physical problems



(c) Assessor



Notes: The figure shows event study estimates of the effect of three treatments on employment using matched samples with 10 nearest neighbor matches per treated unit. All other definitions are as in Figure A3.

Table A3: Treatment effects for benefits, 1 match

Panel A: Mental health problems (pre)		
Mental disability (pre) x Post	0.096***	0.096***
	(0.019)	(0.019)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	7,704	7,704
Adjusted R-squared	0.74	0.741
Panel B: Minor physical health problems (pre)		
Minor physical disability (pre) x Post	−0.090***	−0.084***
	(0.019)	(0.019)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	6,224	6,224
Adjusted R-squared	0.751	0.752
Panel C: Assessor assignment		
Assessor = Atos x Post	−0.057***	−0.055***
	(0.019)	(0.019)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	6,965	6,965
Adjusted R-squared	0.741	0.742

Notes: The table shows difference-in-differences estimates of treatment effects on disability benefit receipt using matched samples with one nearest neighbor match per treated unit. The dependent variable is an indicator for receiving either Disability Living Allowance (DLA) or Personal Independence Payment (PIP). Panel A compares individuals with pre-treatment mental health problems versus those with good mental health. Panel B compares individuals with pre-treatment minor physical problems versus those with major physical problems. Panel C compares regions assigned to Atos versus Capita assessment providers. Matching is based on propensity scores estimated from health limitations affecting work capacity, age, mental health status, total monthly income (log-transformed), gender, and educational attainment in the pre-treatment period. Column 1 includes no time-varying controls; Column 2 adds controls for age, age-squared, and number of children. All specifications include individual and year fixed effects. Standard errors are clustered at the individual level. *p<0.1; **p<0.05; ***p<0.01.

Table A4: Treatment effects on benefits, 10 matches

Panel A: Mental health problems (pre)		
Mental disability (pre) x Post	0.047***	0.046***
	(0.007)	(0.008)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	43,124	43,124
Adjusted R-squared	0.737	0.737
Panel B: Minor physical health problems (pre)		
Minor physical disability (pre) x Post	−0.042***	−0.042***
	(0.007)	(0.007)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	31,881	31,881
Adjusted R-squared	0.736	0.736
Panel C: Assessor assignment		
Assessor = Atos x Post	−0.011**	−0.011**
	(0.005)	(0.005)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	39,484	39,484
Adjusted R-squared	0.732	0.732

Notes: The table shows difference-in-differences estimates of treatment effects on disability benefit receipt using matched samples with 10 nearest neighbor match per treated unit. All other definitions are as in Table A3.

Table A5: Treatment effects on employment, 1 match

Panel A: Mental health problems (pre)		
Mental disability (pre) x Post	−0.027	−0.024
	(0.019)	(0.019)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	7,582	7,582
Adjusted R-squared	0.742	0.746
Panel B: Minor physical health problems (pre)		
Minor physical disability (pre) x Post	0.001	−0.013
	(0.019)	(0.019)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	6,111	6,111
Adjusted R-squared	0.744	0.748
Panel C: Assessor assignment		
Assessor = Atos x Post	0.050***	0.051***
	(0.019)	(0.019)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	6,864	6,864
Adjusted R-squared	0.743	0.748

Notes: The table shows difference-in-differences estimates of treatment effects on employment using matched samples with one nearest neighbor match per treated unit. All other definitions are as in Table A3.

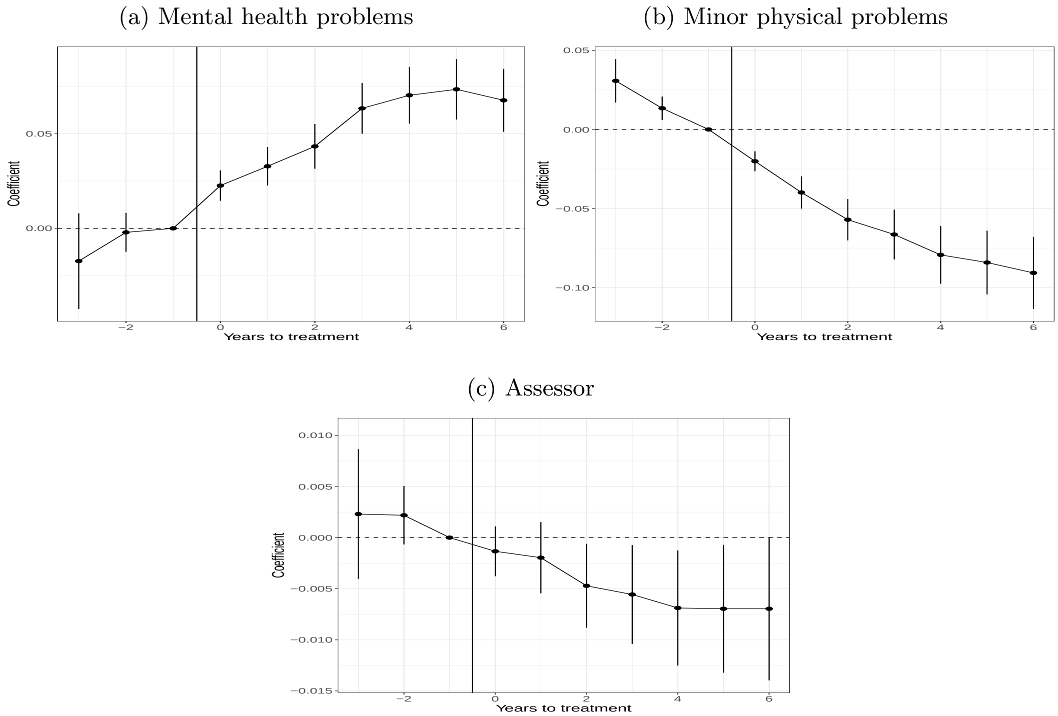
Table A6: Treatment effects on employment, 10 matches

Panel A: Mental health problems (pre)		
Mental disability (pre) x Post	0.024**	0.004
	(0.011)	(0.010)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	42,529	42,529
Adjusted R-squared	0.65	0.662
Panel B: Minor physical health problems (pre)		
Minor physical disability (pre) x Post	−0.005	−0.007
	(0.009)	(0.009)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	31,418	31,418
Adjusted R-squared	0.665	0.676
Panel C: Assessor assignment		
Assessor = Atos x Post	0.019**	0.022***
	(0.008)	(0.008)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	38,958	38,958
Adjusted R-squared	0.647	0.659

Notes: The table shows difference-in-differences estimates of treatment effects on employment using matched samples with 10 nearest neighbor match per treated unit. All other definitions are as in Table A3.

A1.4 No matching

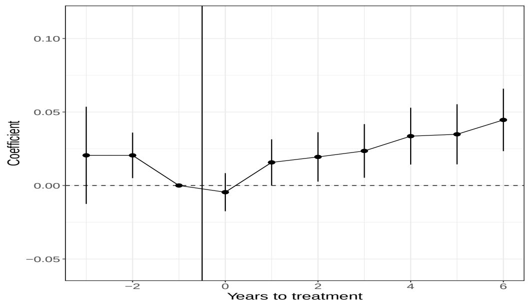
Figure A7: Treatment effects for benefits, no matching



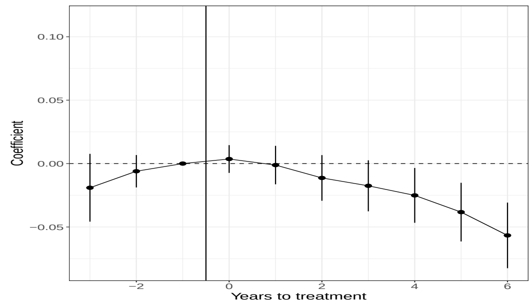
Notes: The figure shows event study estimates of the effect of three treatments on disability benefit receipt using the full unmatched sample. All other definitions are as in Figure A3.

Figure A8: Treatment effects for employment, no matching

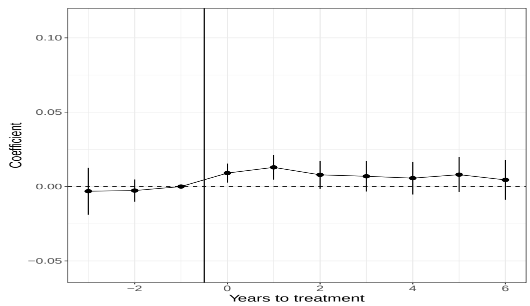
(a) Mental health problems



(b) Minor physical problems



(c) Assessor



Notes: The figure shows event study estimates of the effect of three treatments on employment using the full unmatched sample. All other definitions are the same as in Figure A3.

Table A7: The effect of assessor on benefits, non-matched DID

Panel A: Mental health problems (pre)		
Mental disability (pre) x Post	0.050^{***}	0.050^{***}
	(0.005)	(0.005)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	190,174	190,174
Adjusted R-squared	0.712	0.712
Panel B: Minor physical health problems (pre)		
Minor physical disability (pre) x Post	-0.045^{***}	-0.045^{***}
	(0.005)	(0.005)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	66,840	66,840
Adjusted R-squared	0.73	0.73
Panel C: Assessor assignment		
Assessor = Atos x Post	-0.004^{**}	-0.004^{***}
	(0.002)	(0.002)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	198,232	198,232
Adjusted R-squared	0.708	0.708

Notes: The table shows difference-in-differences estimates of treatment effects on disability benefit receipt using the full unmatched sample. All other definitions are as in Table A3.

Table A8: The effect of assessor on employment, non-matched DID

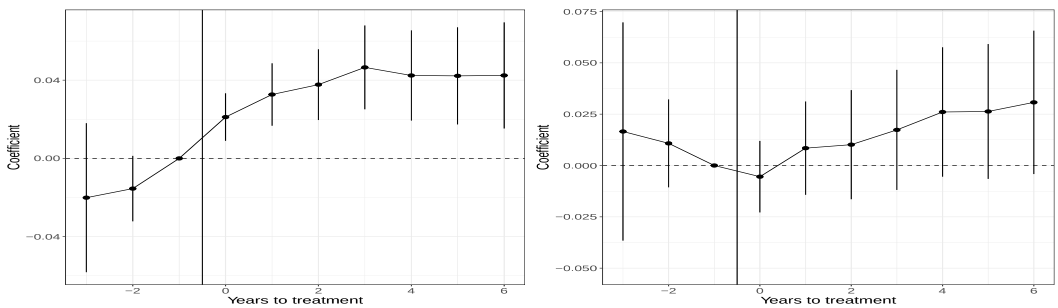
Panel A: Mental health problems (pre)		
Mental disability (pre) x Post	0.012^{*}	0.011^{*}
	(0.007)	(0.006)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	185,902	185,902
Adjusted R-squared	0.609	0.636
Panel B: Minor physical health problems (pre)		
Minor physical disability (pre) x Post	0.013^{**}	-0.006
	(0.007)	(0.006)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	65,693	65,693
Adjusted R-squared	0.662	0.682
Panel C: Assessor assignment		
Assessor = Atos x Post	0.010^{***}	0.013^{***}
	(0.004)	(0.004)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	186,753	186,753
Adjusted R-squared	0.616	0.642

Notes: The table shows difference-in-differences estimates of treatment effects on employment using the full unmatched sample. All other definitions are as in Table A3.

A1.5 Looser/stricter definitions mental health

Figure A9: Treatment effects for benefits, looser definition mental health

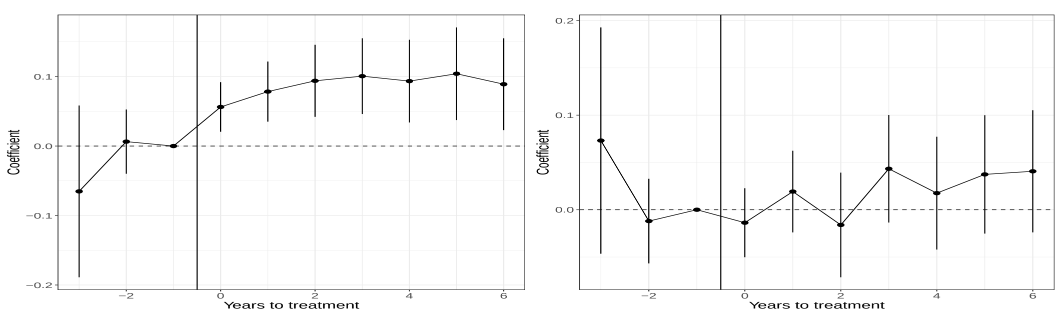
(a) Benefits (b) Employment



Notes: The figure shows event study estimates of the effect of pre-existing mental health on disability benefit receipt and employment, using a looser definition of mental health conditions. Mental health problems are defined more inclusively than in the main specification, including individuals who report being limited by mental health issues "some of the time" in addition to "most of the time" or "all of the time" in daily activities, work capacity, or social functioning. All other definitions are as in Figure A3.

Figure A10: Treatment effects for benefits, stricter definition mental health

(a) Benefits (b) Employment



Notes: The figure shows event study estimates of the effect of pre-existing mental health on disability benefit receipt and employment, using a stricter definition of mental health conditions. Mental health problems are include only individuals who report being limited by mental health issues "all of the time" in daily activities, work capacity, or social functioning, excluding those who report limitations "most of the time". All other definitions are as in Figure A3.

Table A9: The effect of pre-existing mental health problems on benefits, looser definition mental health

Mental health problems (pre) x Post	0.038***	0.038***
	(0.007)	(0.007)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	23,406	23,406
Adjusted R-squared	0.751	0.751

Notes: The table shows difference-in-differences estimates of the effect of pre-existing mental health problems on disability benefit receipt using a looser definition of mental health conditions. Mental health problems are defined more inclusively than in the main specification, including individuals who report being limited by mental health issues "some of the time" in addition to "most of the time" or "all of the time" in daily activities, work capacity, or social functioning. All other definitions are as in Table A3.

Table A10: The effect of pre-existing mental health problems on employment, looser definition mental health

Mental health problems (pre) x Post	0.012	0.006
	(0.011)	(0.010)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	23,074	23,074
Adjusted R-squared	0.689	0.697

Notes: The table shows difference-in-differences estimates of the effect of pre-existing mental health problems on employment using a looser definition of mental health conditions. Mental health problems are defined more inclusively than in the main specification, including individuals who report being limited by mental health issues "some of the time" in addition to "most of the time" or "all of the time" in daily activities, work capacity, or social functioning. All other definitions are as in Table A3.

Table A11: The effect of pre-existing mental health problems on benefits, stricter definition mental health

Mental health problems (pre) x Post	0.085***	0.085***
	(0.020)	(0.020)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	23,406	23,406
Adjusted R-squared	0.751	0.751

Notes: The table shows difference-in-differences estimates of the effect of pre-existing mental health problems on disability benefit receipt using a stricter definition of mental health conditions. Mental health problems are include only individuals who report being limited by mental health issues "all of the time" in daily activities, work capacity, or social functioning, excluding those who report limitations "most of the time". All other definitions are as in Table A3.

Table A12: The effect of pre-existing mental health problems on employment, stricter definition mental health

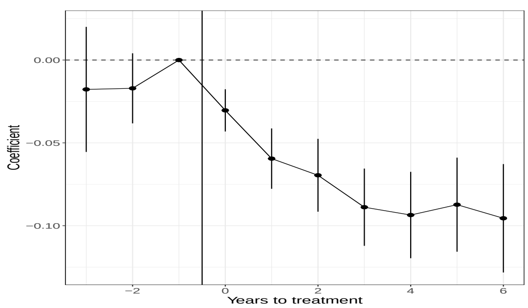
Mental health problems (pre) x Post	0.022	0.013
	(0.021)	(0.021)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	23,074	23,074
Adjusted R-squared	0.689	0.697

Notes: The table shows difference-in-differences estimates of the effect of pre-existing mental health problems on employment using a stricter definition of mental health conditions. Mental health problems are include only individuals who report being limited by mental health issues "all of the time" in daily activities, work capacity, or social functioning, excluding those who report limitations "most of the time". All other definitions are as in Table A3.

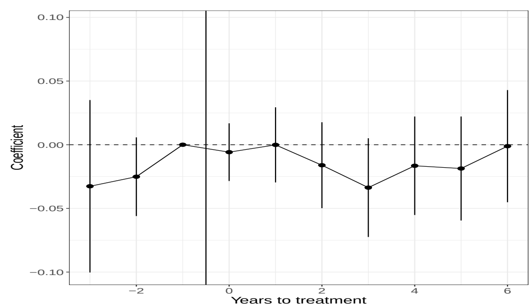
A1.6 Different definition minor physical problems

Figure A11: Treatment effects for benefits, alternative definition minor physical problems

(a) Benefits



(b) Employment



Notes: The figure shows event study estimates of the effect of pre-existing minor physical health problems on disability benefit receipt and employment, using an alternative definition of minor physical health conditions. Minor conditions are now defined as those who report limitations "a little of the time" in daily activities due to physical health, and those who report "some of the time", "most of the time", or "all of the time" are categorized as major physical health problems.. All other definitions are as in Figure A3.

Table A13: The effect of pre-existing minor physical health problems on benefits, alternative definition

Minor physical disability (pre) x Post	-0.062^{***}	-0.061^{***}
	(0.008)	(0.008)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	19,059	19,059
Adjusted R-squared	0.753	0.753

Notes: The table shows difference-in-differences estimates of the effect of pre-existing minor physical health problems on disability benefit receipt using an alternative definition of minor physical health conditions. Minor conditions are now defined as those who report limitations “a little of the time” in daily activities due to physical health, and those who report “some of the time”, “most of the time”, or “all of the time” are categorized as major physical health problems. All other definitions are as in Table A3.

Table A14: The effect of pre-existing minor physical health problems on employment, alternative definition

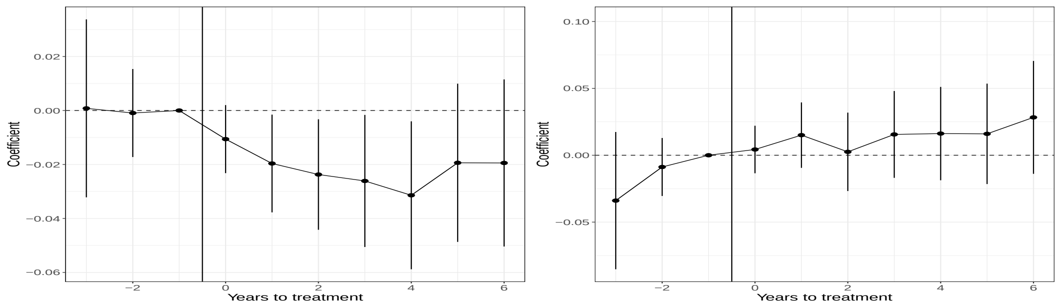
Minor physical disability (pre) x Post	-0.002	-0.002
	(0.013)	(0.013)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	18,764	18,764
Adjusted R-squared	0.697	0.704

Notes: The table shows difference-in-differences estimates of the effect of pre-existing minor physical health problems on employment using an alternative definition of minor physical health conditions. Minor conditions are now defined as those who report limitations “a little of the time” in daily activities due to physical health, and those who report “some of the time”, “most of the time”, or “all of the time” are categorized as major physical health problems. All other definitions are as in Table A3.

A1.7 No movers

Figure A12: Effect of assessor on benefits and employment, no movers

(a) Benefits (b) Employment



Notes: The figure shows event study estimates of the effect of the assessor treatment on benefits and employment using matched samples, excluding individuals who moved between Atos and Capita regions during the sample period. All other definitions are as in Figure A3.

Table A15: The effect of assessor on benefits, no movers

Assessor = Atos x Post	-0.018^{**}	-0.017^{**}
	(0.008)	(0.008)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	20,853	20,853
Adjusted R-squared	0.745	0.746

Notes: The table shows difference-in-differences estimates of the effect of regional assessor assignment on disability benefit receipt, excluding individuals who moved between Atos and Capita regions during the sample period. All other definitions are as in Table A3.

Table A16: The effect of assessor on employment, no movers

Assessor = Atos x Post	0.011	0.013
	(0.011)	(0.011)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	20,579	20,579
Adjusted R-squared	0.689	0.698

Notes: The table shows difference-in-differences estimates of the effect of regional assessor assignment on employment, excluding individuals who moved between Atos and Capita regions during the sample period. All other definitions are as in Table A3.

A2 Additional Tables

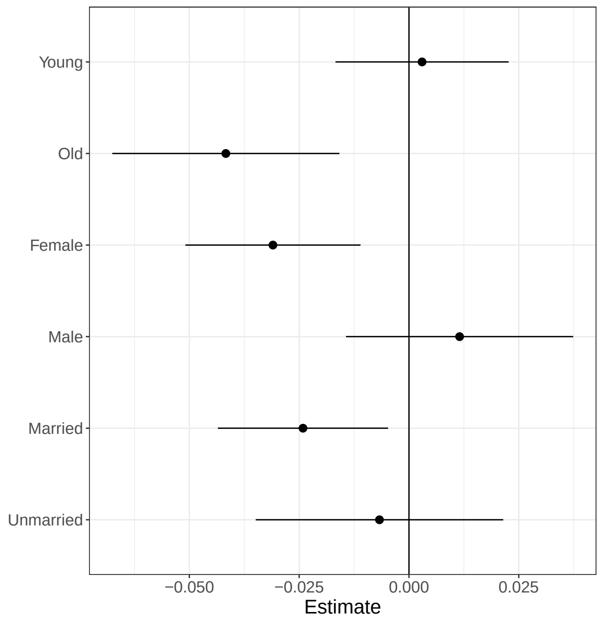
Table A17: Treatment effects for benefits

Panel A: Mental health problems (pre)		
Mental health problems (pre) x Post	0.058^{***}	0.059^{***}
	(0.010)	(0.010)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	23,406	23,406
Adjusted R-squared	0.751	0.751
Panel B: Minor physical health problems (pre)		
Minor physical disability (pre) x Post	-0.056^{***}	-0.056^{***}
	(0.009)	(0.009)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	19,059	19,059
Adjusted R-squared	0.753	0.753
Panel C: Assessor assignment		
Assessor = Atos x Post	-0.018^{**}	-0.018^{**}
	(0.008)	(0.008)
Controls	No	Yes
Person fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Observations	21,408	21,408
Adjusted R-squared	0.748	0.748

Notes: The table shows difference-in-differences estimates of treatment effects on disability benefit receipt using matched samples. The dependent variable is an indicator for receiving either Disability Living Allowance (DLA) or Personal Independence Payment (PIP). Panel A compares individuals with pre-treatment mental health problems versus those with good mental health. Panel B compares individuals with pre-treatment minor physical problems versus those with major physical problems. Panel C compares regions assigned to Atos versus Capita assessment providers. All samples are constructed using nearest neighbor matching with five matches per treated unit based on pre-treatment characteristics including health limitations affecting work capacity, age, mental health status, total monthly income (log-transformed), gender, and educational attainment. Column 1 includes no time-varying controls; Column 2 adds controls for age, age-squared, and number of children. All specifications include individual and year fixed effects. Standard errors are clustered at the individual level. *p<0.1; **p<0.05; ***p<0.01.

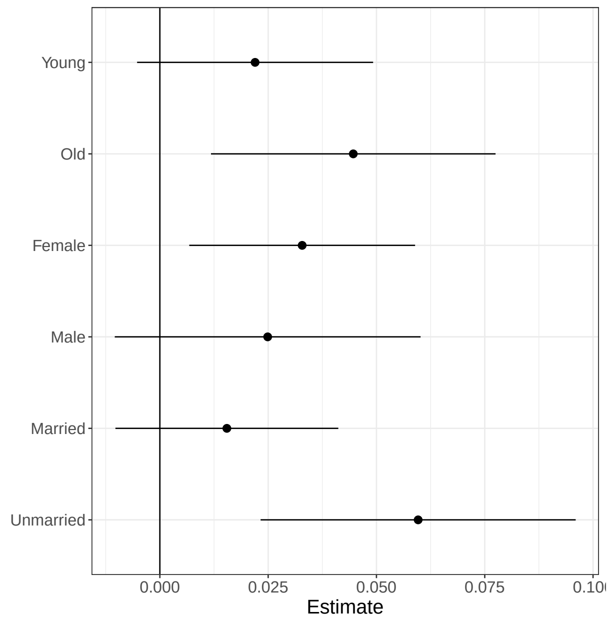
A3 Additional Figures

Figure A13: Heterogeneous effects of assessor on benefits



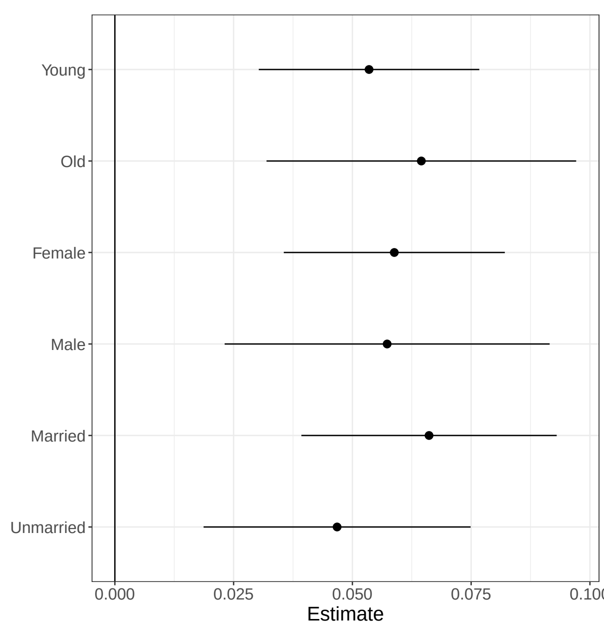
Notes: The figure shows heterogeneous treatment effects of regional assessor assignment on disability benefit receipt across demographic subgroups. Each point estimate represents the coefficient from a separate difference-in-differences regression comparing regions assigned to Atos versus Capita assessment providers for the specified subgroup. The dependent variable is an indicator for receiving either Disability Living Allowance (DLA) or Personal Independence Payment (PIP). Subgroups are defined as follows: Young/Old split at median age in the pre-treatment period; Male/Female based on gender; Married/Unmarried based on marital status in the pre-treatment period. All estimates use the matched sample described in Section 4 with five matches per treated unit. Point estimates and 95% confidence intervals are shown, with standard errors clustered at the individual level. All specifications include individual and year fixed effects and control for age, age-squared, and number of children.

Figure A14: Heterogeneous effects of assessor on employment



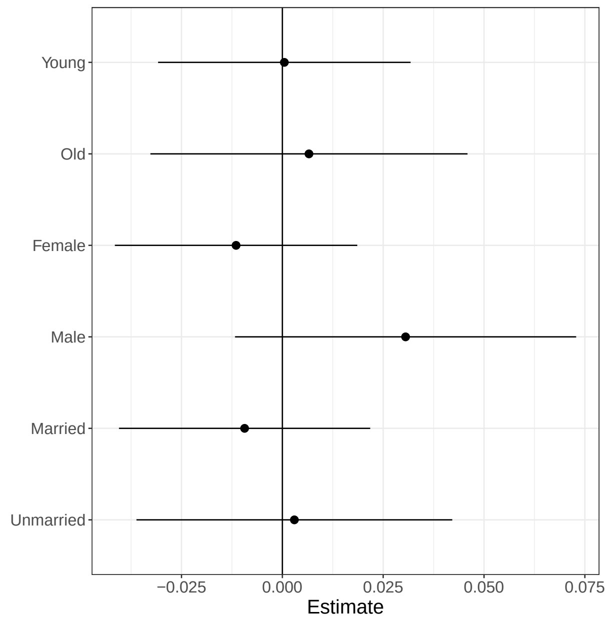
Notes: The figure shows heterogeneous treatment effects of regional assessor assignment on employment across demographic subgroups. Other definitions are as in Figure A13.

Figure A15: Heterogeneous effects of pre-treatment mental disability on benefits



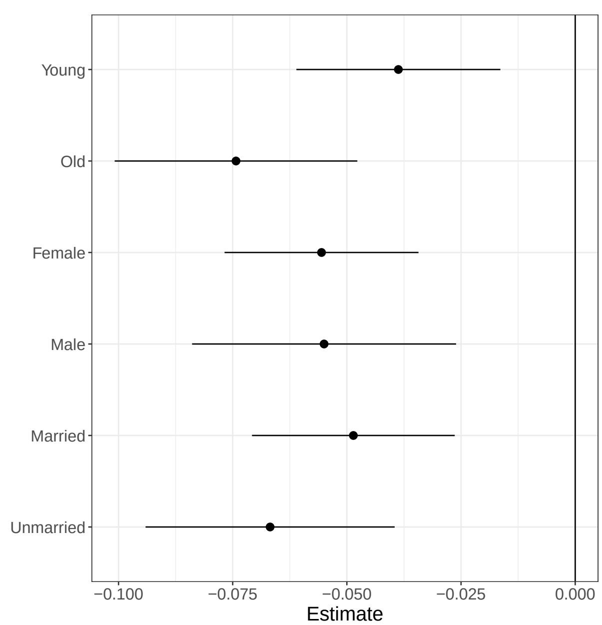
Notes: The figure shows heterogeneous treatment effects of pre-existing mental health problems on disability benefit receipt across demographic subgroups. Other definitions are as in Figure A13.

Figure A16: Heterogeneous effects of pre-treatment mental disability on employment



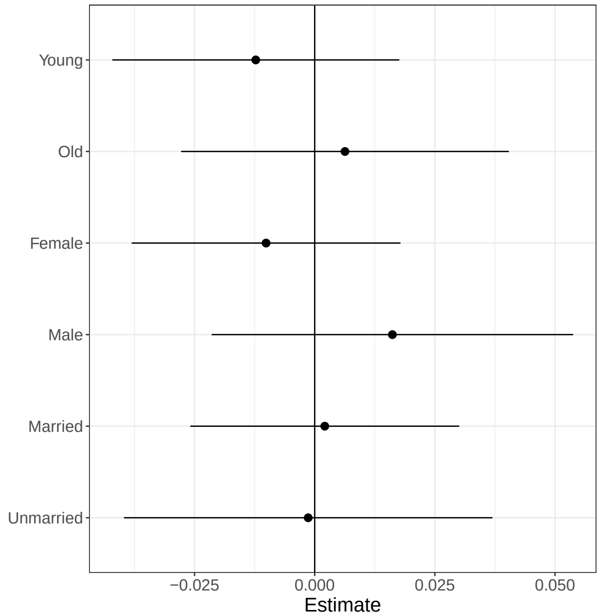
Notes: The figure shows heterogeneous treatment effects of pre-existing mental health problems on employment across demographic subgroups. Other definitions are as in Figure A13.

Figure A17: Heterogeneous effects of pre-treatment minor physical disability on benefits



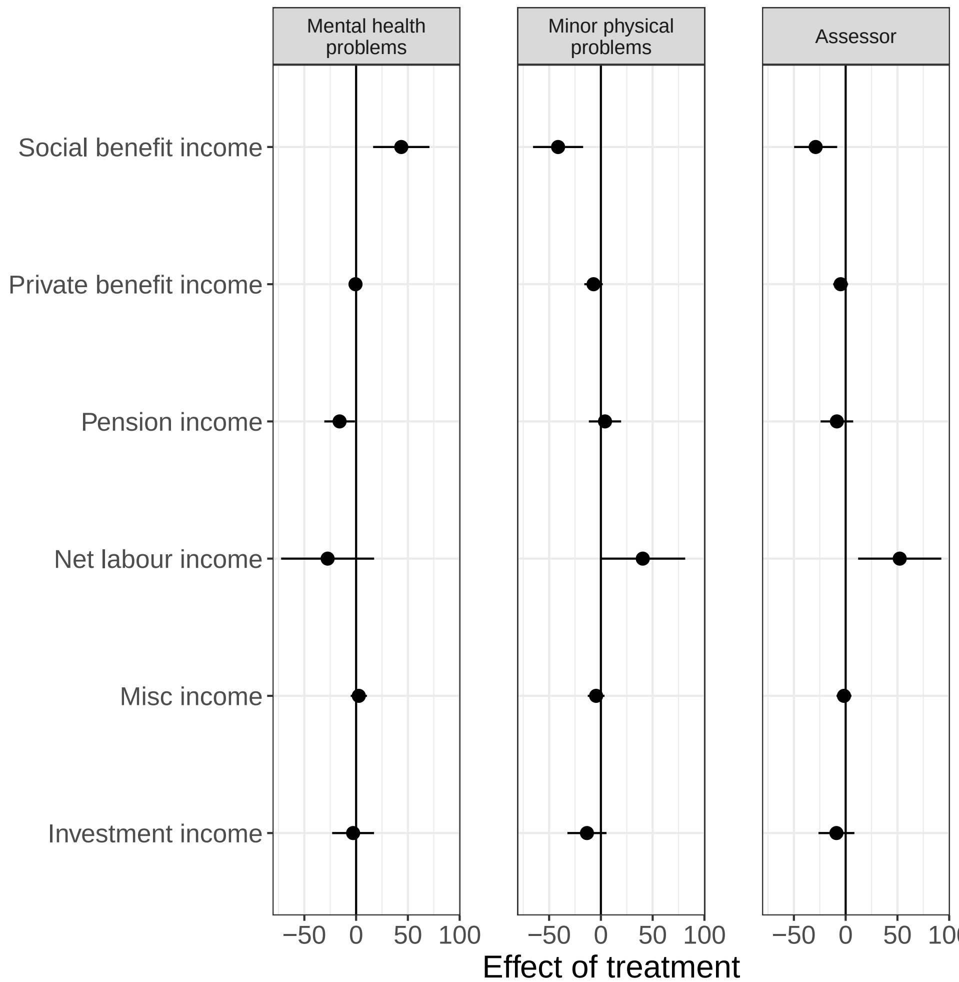
Notes: The figure shows heterogeneous treatment effects of pre-existing minor physical disabilities on disability benefit receipt across demographic subgroups. Other definitions are as in Figure A13.

Figure A18: Heterogeneous effects of pre-treatment minor physical disability on employment



Notes: The figure shows heterogeneous treatment effects of pre-existing minor physical disabilities on employment across demographic subgroups. Other definitions are as in Figure A13.

Figure A19: Effect of treatments on other benefits (amounts)



Notes: The figure shows difference-in-differences estimates of the effects of our three treatments on various income sources, using the matched sample described in Section 4. Each point estimate represents the coefficient from a separate regression with the respective income category as the dependent variable. Social benefit income includes disability benefits (DLA/PIP) and other social transfers. Private benefit income includes employer-provided benefits and insurance payments. Net labor income represents earnings from employment. Pension income, investment income, and miscellaneous income capture additional income streams. All other definitions are as in Figure 5.