Essays on the Economics of Welfare Policy and Well-being

Charlotte Edney

Department of Economics

Lancaster University

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Economics

February 2022

Abstract

The objective of this thesis is the evaluation of government policy. The first two chapters are focused on policies that affect separated families; Child Maintenance, an important source of income for separated parents, and the Lone Parent Obligation reform.

Chapter 2 examines the effect of receiving child maintenance on youth behavioural and social outcomes, using new methods to assess the extent to which results are driven by selection on unobservable characteristics. The findings provide compelling estimates that, among boys, child maintenance receipt is associated with significantly fewer conduct problems and better pro-social skills. Chapter 3 evaluates the Lone Parent Obligation reform which imposed work search requirements for lone mothers with a youngest child aged 5 or 6. The findings suggest that the reform was successful in increasing maternal employment, an effect size of 8 percentage points. This was accompanied by an increase in the proportion of lone mothers who were searching for work by 9 percentage points. The results are similar across two datasets and are robust to several sensitivity checks, which add credibility to the estimates.

Finally, Chapter 4 focuses on the COVID-19 pandemic and mental health in the UK. It aims to isolate the effect of being in a government imposed lockdown from the threat of the virus. The results reveal that lockdown plays a large role in explaining mental health declines, with national case and death rates also found to be important. These effects are heterogeneous across the population, with younger adults, women, and lone mothers particularly vulnerable to the adverse effects of lockdown.

For my parents

Declaration

This thesis is submitted in partial fulfilment of the requirements for the degree Doctor of Philosophy in Economics at Lancaster University.

I hereby declare that the contents of this dissertation are original and the work has not been submitted, in whole or in part, for consideration for any other degree or professional qualification.

Charlotte Edney

September 2021

Acknowledgements

I am grateful to my supervisors, Professor Ian Walker, Dr Maria Navarro Paniagua and Dr Vincent O'Sullivan for their continuous support over the last four years. They gave an abundance of time, patience, and good humour. Their feedback and guidance have benefited me hugely in terms of developing my skills as a researcher in my academic studies and future career. I am grateful to my examiners Professor Yu Zhu and Dr Saurabh Singhal for their comments and to Dr Catherine Porter for her feedback at various milestones.

Lancaster University Economics department has proved a great environment to study and teach Economics. Thanks to Caren Wareing, Themis Pavlidis, Giuseppe Migali for their support and to Colin Green who persuaded me to do it.

I am grateful to the Economic and Social Research Council (ESRC) for providing me with the financial means to undertake this work, and to the Department for Work and Pensions for supporting me through a CASE studentship. The UK Research Institution (UKRI) provides opportunities for students to undertake various internships. I cannot emphasise enough the value of this scheme and I feel fortunate to have completed a three-month policy internship at the Department for Work and Pensions. Many thanks must go to my colleagues there who provided interesting and fruitful discussions around the topics covered in this thesis.

The PhD experience would have been far less enjoyable without the colleagues and friends made along the way who offered both academic help and friendship: Anwar Adem, Gerda Buchmueller, Alex Farnell, Alex Skouralis, Huan Yang, and Yingying Zhang. To my friends outside of academia, in particular Lizzie Lavan and Lucy Middleton, thank you for keeping me sane during the most intense periods.

And finally, the completion of this thesis would not have been possible without the following people. Thank you to my parents Graham and Jayne, and my sister Rosie, for their love, encouragement, and unwavering support. My grandparents Tom and Kathleen, and Colin and Joan, who is dearly missed. And to Andrew who kept me going throughout.

Table of Contents

List of Tables

Appendix Tables

List of Figures

Appendix Figures

1 Introduction

This thesis is concerned with evaluating government policy. The choices of policy and decision makers in Whitehall have the capacity to radically change the lives of people living in the UK. The first two chapters of this thesis evaluate welfare policies targeted at separated families. The first assesses the effect of lone mothers receiving transfers from her erstwhile partner on outcomes for the child experiencing separation, and the second examines a welfare policy change that sought to push single parents into work. The third chapter, investigates the effect of the COVID-19 pandemic on mental health, showcasing the potential for government policy to disrupt lives, albeit with the aim of protecting public health.

The circumstances in which a child grows up are important. There are around 3.5 million children in the UK who live in a separated family (Department for Work and Pensions, (2019)). For children experiencing parental separation, it can lead to emotional instability and upheaval, which is combined with a negative financial shock to the household. This leaves separated families at substantially higher risk of poverty and poorer living standards. The 2020 Joeseph Rowntree Foundation annual poverty report shows that lone parents have the highest poverty rate (households below average income measure) of all family types, which is almost double that of coupled families (Goulden, 2020). Though lone parent poverty rates have declined since the early 1990s, almost 50 percent of lone parents remain in poverty and little progress has been made since 2009 (ibid). Poverty may be one channel which causes the large disparities in educational attainment across family types, another important dimension for policymakers to consider.

Child maintenance is a policy that governments have used to both alleviate the financial burden on separated parents and to provide a saving to the public purse. Bradshaw (2006) shows that child poverty in separated families can be directly reduced by child maintenance. But the UK child maintenance system is in disarray, plagued with low compliance (in 2017/2018 only 48 percent of lone parents had any form of child maintenance arrangement (Department for Work and Pensions, (2019))), high running costs and arrears that look to be un-collectable. This does not appear to be a policy that is working to its full potential.

Despite evidence suggesting a beneficial effect of child maintenance on child poverty, little is known about the role of child maintenance on wider outcomes for children in the UK. Much of the evidence comes from the US and suggests that child maintenance may mitigate the adverse effects of separation, at least partially. Problematically, many of these studies suffer from bias associated with non-random selection into child maintenance.

Chapter two presents UK evidence on the effect of receiving child maintenance (referred to as child support in the US) on youth behavioural and social outcomes, measured using the Strengths and Difficulties Questionnaire administered to youths aged 10-15. A further problem that researchers must contend with in this area, is the small samples of separated families in survey data. For this chapter, I use longitudinal survey Understanding Society (UKHLS), which, crucially, has a large enough sample to facilitate the analysis, and is rich in information on child maintenance. In the absence of a credible instrument, I estimate pooled linear models and attempt to dispel concerns of selection bias by looking at coefficient movements with the addition of controls. I then employ the recently developed Oster (2019) test which gives an indication of the degree to which results are sensitive to selection on unobservables by inferring from selection on observables. Moreover, I utilise the panel dimension of the dataset with fixed effect modelling to account for unobserved heterogeneity, which provides an additional check on the susceptibility of results to time-invariant unobserved selection.

I find receiving child maintenance is associated with a reduction in youth conduct problems by 12 percent of a standard deviation and an improvement in pro-social skills by 16 percent of a standard deviation - effects that are driven entirely by boys. The findings are also robust to the Oster (2019) test, supporting the suggestion that the result is not, at least not entirely, driven by selection on unobservables. Receiving anything, as opposed to nothing, is found to be more important than the amount of child maintenance received.

Yet further issues stem from the very nature of separation – that we often do not observe the absent parent in survey data, they are rarely followed by survey teams when they leave the household. Being unable to control for non-resident parent characteristics means non-causal methods are even more likely to suffer from omitted variable bias. Understanding Society is a notable exception, in that the parent-with-care is asked questions about their ex-partner. This is not used in the main analysis, as the questions are only asked to a parent-with-care if they have contact with the non-resident parent. Nevertheless, using a sub-sample of respondents who do have contact, I test the importance of adding these controls, and find the estimates do not change much, in fact coefficient sizes get larger.

The policy implications are clear, increasing the number of children receiving child maintenance, could have significant benefits for the child in question. It also has long-term implications as fewer conduct problems in childhood, is associated with better outcomes in adulthood (see Feinstein (2000), Knapp et al. (2011), and Clark & Lepinteur (2019)). How to go about increasing child maintenance receipt is an important and demanding task for policymakers.

Lone parents face significant barriers to employment, which include the state of local labour market opportunities, the availability of affordable childcare, and the need for job flexibility in both hours and location. In 2012, 37 percent of lone parent households had no one in work compared with just 4.9 percent in coupled households (Office for National Statistics (2012)). This difference in employment rates between lone and coupled parents has narrowed over time but persists. The UK has one the widest lone-coupled parent employment gaps in the EU, in fact in Italy, Spain and France the employment rates of lone parents are above those of coupled parents.

There is a commonly held view among both the British public and their politicians that the solution to poverty is work. This was the main foundation on which a range of labour market activation policies were implemented in the 2000s. In Chapter three I examine one such reform; the 2012 final phase of the Lone Parent Obligation reform. In essence, the reform enforced work search requirements on welfare-claiming lone parents with a youngest child aged 5 or above. This chapter uses a quasi-experimental research design to provide causal evidence on the effect of this reform on lone mother’s employment outcomes, as well as welfare receipt. I use a difference-in-difference approach with married mothers of 5-6 year olds as the control group, and lone mothers of 5-6 year olds the treated group. For replicability purposes two datasets are used – the UK Labour Force Survey and Understanding Society. Numerous robustness tests are conducted, including the use of a group of single mothers with even younger children as an alternative control group.

The findings show the reform was effective in increasing maternal employment. The reform caused employment to rise by 8 percentage points. This was coupled with an increase in work search activity by 10 percentage points. Despite this, some lone mothers remained in receipt of benefit by shifting onto health-related benefits – an increase of around 6 percentage points (equivalent to 60-75 percent increase). However, this effect size is smaller than that found on employment. An event study analysis shows increases in employment start much later around five quarters post-reform and persist beyond that, whilst work search activity increases are concentrated in the quarters immediately after the reform. The chapter highlights the importance of the frequency of data collection – the quarterly Labour Force Survey gives a more granular view of changes in employment related outcomes over time. The results are similar across the two datasets and robust to several sensitivity checks.

The fourth chapter contributes to an emerging literature on the COVID-19 pandemic and mental health. Initially, the intention was to examine the impact of COVID on separated parents and their child maintenance payments, for which we submitted questions to be included in the Understanding Society COVID-19 survey. Although the questions were accepted, the response rate was not large enough to facilitate a rigorous economic analysis for inclusion in this thesis. Therefore, chapter three uses the full sample of the Understanding Society COVID-19 survey, of which I used 7 waves, which were collected in April, May, June, July, September, and November of 2020 and January 2021.

Existing evidence shows the pandemic led to a large decline in mental health, though many studies measure the overall effect of the pandemic by comparing mental health before the pandemic with a given time point during the pandemic. There is no UK research, that I know of, which examines the relationship between mental health, COVID case rates, death rates and being in a government-imposed lockdown. Using longitudinal data collected during the pandemic, linked with official data on COVID case and death rates, this chapter, by exploiting variation in both the severity of COVID and lockdowns, can separately unravel the importance of each on an individual's mental health. In addition, using geographic identifiers in the dataset, I can analyse the effect of both local authority and national COVID cases/deaths rates on mental health outcomes. The identification strategy utilises the Understanding Society panel dataset which follows individuals, first monthly and later bi-monthly, from April 2020 through to January 2021. This enables estimation of a fixed effects model, making the estimates robust to time-invariant heterogeneity, an improvement on pooled Ordinary Least Squares.

The analysis focuses on three main measures of mental health, and the findings suggest that lockdown is associated with significantly worse mental health, increasing mental distress by 0.7 points on the overall GHQ-12 scale, increases the count of mental health problems by half a problem, and is associated with a 7 percentage point increase in the incidence of severe mental health problems. Rising national COVID case and death rates are also predictive of worse mental health, where local measures are not.

Fitting with the theme of the previous two chapters, I look specifically at the effects on lone mother's mental health, though as mentioned sample sizes are small. I find lockdown places a significant burden on their mental health, the effect size is almost three times as large as for coupled mothers and twice as large as the full sample of respondents. Other heterogeneous effects exist; younger individuals and women are more adversely affected by lockdown than their counterparts.

During the pandemic, the government acknowledged that poor mental health was a concern, and introduced some measures, such as allowing exercise and social bubbles, to try to mitigate the fall in well-being. Whilst it is not possible to say what effect this had on mental health, the findings suggest that it was not enough to lessen the mental health burden, particularly for more vulnerable population groups such as lone mothers. This chapter contributes to a growing body of literature and sets the groundwork for important avenues of further research.

2 Chapter 2

Do financial contributions from non-resident parents improve outcomes for separated children?

Evidence from the UK child maintenance system.

I am grateful to my supervisors for their comments and feedback on this chapter. In particular, thanks to Ian Walker for his knowledge of the child maintenance system. In addition, thanks to participants at the North West Doctoral Training Programme 2020, Australian Gender Economics Workshop 2021, and the Royal Economic Society conference 2021. I am grateful to the Understanding Society team for providing support with data queries, to the UK Data Service for making the data available and to Alex Farnell, Andrew McKendrick, and Sarah Sinclair for helpful comments.

2.1 Introduction

Between 1996 and 2017 the number of lone parent families in the UK increased by 15 percent, with the biggest rise concentrated in the first decade. Over the last ten years the number has remained stable and, despite government attempts, levels of poverty among single parents relative to the population remains as large a problem as ever.

The UK government introduced a system of child maintenance[1] (CM) in 1993 whereby money is transferred from the non-resident parent (NRP) to the parent-with-care (PWC) according to, for the first time, a specific formula. An extract from the 1990 White Paper "Children Come First" demonstrates the government's desire to hold non-resident parents accountable for their children, whilst reducing the strain on the welfare state (and, by extension, taxpayers):

> "It is right that other taxpayers should help maintain children when the children's own parents, despite their own best efforts, do not have enough resources to do so themselves. That will continue to be the case. But it is not right that taxpayers, who include other families, should shoulder that responsibility instead of parents who are able to do it themselves."

Surprisingly, little quantitative evaluation of this policy exists, and none that relates to subsequent policy developments in the UK. This is in stark contrast to the US where a much larger literature exists around the effects of child support (see for example Roff (2010), Roff & Lugo-Gil (2012), Baughman (2017), Tannenbaum (2020), and Meyer et al. (2020)).

Parental separation is found to be negatively associated with a child's educational outcomes and well-being, and positively associated with the likelihood they engage in risky behaviours (surveys include Amato (2001), Haveman et al. (1995), Adamsons & Johnson (2013)). Some work has attempted to control for unobserved confounders using mainly fixed effects modelling which, on balance, suggests parental separation has negative consequences

[1] Child maintenance is commonly referred to as "child support" in the US and most other countries. This was also the case in the UK until it was replaced by child maintenance after several reforms in the 2000s.

for the child (see for example Amato (2014), Cherlin et al. (1998), and Aughinbaugh et al. (2005)).[2]

Separation has immediate effects on the household: typically, there would be a considerable loss in the child's household net income, as well as the absence of one parent, usually the father. Few papers in the UK, with the exception of Walker & Zhu (2011b), attempt to disentangle the two effects – something that is likely to be important for policy since the two might be systematically related. Upon separation the child will likely only have contact with one parent at a time so complementarities are forgone; at worst there could be no contact with the NRP at all and the child's sense of loss might be all the greater. Given the importance of family background for child outcomes (for example Brunello et al. (2017), Plug & Vijverberg (2003)), any negative effect from family dissolution might be large. In addition, there is the inevitable drop in living standards that the newly independent parent and their child(ren) face from both the loss in household income and the reduction in household economies of scale. Economic hardship is likely to be detrimental to children's development, in this case child maintenance should act as, at least, a partial buffer. It may also mitigate some emotional distress of the PWC, leading to higher quality parenting.

The contribution of this paper is threefold. First, I attempt to establish the correlations of child maintenance and a range of child behavioural and social outcomes in a rich and recent dataset. Second, I explore the extent to which the correlations might be causal by testing the sensitivity of estimates to potential selection on unobservables and use fixed effects estimation. And, finally, I conduct an analysis on a sub-sample of respondents to examine the role that NRP characteristics and new family dynamics play in determining child outcomes.

Measuring the direct effect that child maintenance has on child outcomes is a difficult task. Firstly, we must contend with the issue of selection into child maintenance payment. Any serious attempt to identify a causal effect would acknowledge that the characteristics of the NRP and the quality of the NRP's prior relationship is likely to be important. Furthermore, the amount of contact time, post separation, will matter. Even more complex is the relationship between the resident and non-resident parent. It may be that a bad relationship adversely affects the child(ren). While existing research addresses this by controlling for as many variables as

[2] Björklund & Sundström (2006) are an exception, by exploiting a sibling difference approach using a large Swedish administrative dataset they show that the strong correlation in the data completely disappears after controlling for family effects.

possible, the very nature of separation means this is not always extensive and the threat of omitted variable bias associated with selection on unobservables is non-negligible.

The sample comes from a large, detailed and long-running panel study and consists only of children aged 10-15 with separated parents. I rely on the panel nature of the data. Other strategies might include comparing children in separated families to intact families. This is not a strategy that is pursued because child maintenance income is likely to have different effects to other household income sources. Rather, using a rich dataset, I can control for the level of child-NRP contact, inter-parental relationship quality and, importantly other non-NRP characteristics. This last component is something that papers relating to CM, and the impacts of separation more generally, often do not have. The paper explores the sensitivity of these conventional non-causal estimates to selection on unobservables using the recent Oster (2019) test. In addition, I employ a fixed effects model to account for unobserved heterogeneity. This is also the first study to employ Oster's test in the context of CM and child outcomes.

The findings suggest that CM receipt is associated with a large reduction in youth conduct problems by 15 percent of a standard deviation, while the amount of CM received has a significant though much smaller effect. Breaking down this result by gender indicates the effect is driven by boys who see a statistically significant reduction in conduct problems by 21 percent of a standard deviation, while the effect size for girls is small and not statistically significant. For boys only, there is also an association between receiving CM and better pro-social skills corresponding to 16 percent of a standard deviation.

Including controls for the non-resident parent, using a sub-sample of the data, shows that the post-separation characteristics of the NRP are important and imply that the beneficial effects of receiving child maintenance may be offset by new family dynamics. I hesitate to claim that the estimated correlations reported here can be given a causal interpretation. However, being able to control for contact, relationship quality and non-resident parent characteristics across pooled models, random effects, and fixed effects, appears to offer robust and convincing evidence for the UK. The results from the Oster (2019) test indicate that selection is unlikely to be driving the result (at least, not fully).

The associations suggested above seem plausible and they are of interest in of themselves even in the absence of them being causal. The finding that the lower bound estimates, that allow for what is generally thought to be a reasonable degree of selection on unobservables, are significantly different from zero is consistent with the fixed effect findings from earlier work on the overall effects of separation – to the extent that the estimates of the CM effect counteracts the adverse effects of separation per se.

The paper is laid out as follows. Section 2.2 provides a background to the UK child maintenance system; Section 2.3 summarises the relevant literature; Section 2.4 presents the data and descriptive statistics; Section 2.5 outlines the methodology; and Section 2.6 the results. Section 2.7 concludes.

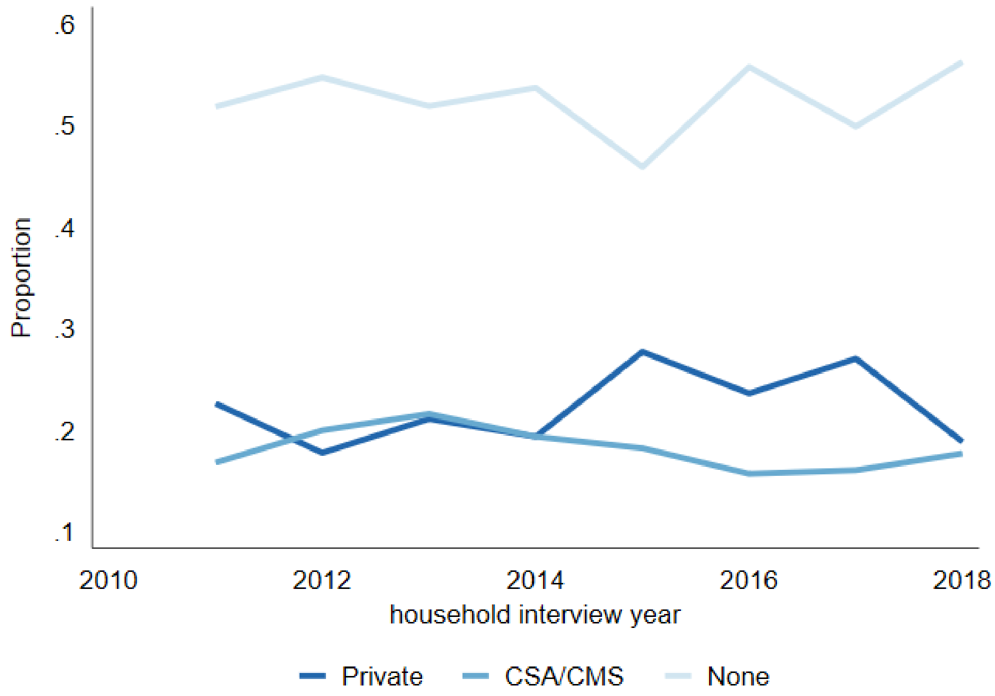
2.2 Background

The history of child maintenance in the UK has been long and complex. It began with the Child Support Act in 1991 and the introduction of a formula for calculating a so called "maintenance requirement". The formula was based on an income-shares method, where the CM obligation depended on the household income of both separated parents, and many deductions were possible for housing costs, new children, commuting cost and many more. The Child Support Handbook (2000) from the Child Poverty Action Group provides a useful guide to the many intricacies of the formula.

A new system emerged in 2003, which changed the way maintenance was calculated. The formula was modified to require only two key pieces of information, the NRPs income and their number of children. Skinner (2012) provides a summary of these two schemes. This scheme did not last 10 years before the Henshaw report publicised a multitude of system failings. This paved the way for a new CM scheme, the one that exists currently, to come into force. In contrast to the previous reform, the formula did not change radically. Instead, the main departure was the introduction of charging parents to use the statutory scheme, which was supposed to encourage private (also known as voluntary or informal) arrangements between parents. The first port of call for separated parents is the *CMS Options* service – a free and impartial support service to help parents choose their child maintenance arrangement. If parents cannot agree, refuse to engage with the other parent or mutually decide against an informal arrangement, CMS Options refers them to the CMS where they have two legally binding alternatives, which depend on the likelihood of the NRP paying. The PWC could opt for "Direct Pay" where the CMS calculates the maintenance and contacts the NRP to inform them of their liability. The NRP then pays child maintenance directly to the parent with care. The remaining option is "Collect & Pay", used when the non-resident parent is unlikely to pay child maintenance. The difference here is that the CMS takes on the enforcement and collection of the liability, along with the calculation. They can use deduction from earnings orders, lump sum deductions and liability orders – for which there is an additional “enforcement” charge. In serious cases of non-compliance the CMS recently gained the power to remove driving licences and passports. Using the CMS is costly, a £25 application fee is applied to both Direct Pay and Collect & Pay. If the parent-with-care opts for Collect & Pay s/he must sacrifice 4 percent of the liability to the CMS, while the NRP must pay an additional 20 percent of liability to the CMS. For example, if the calculated liability is £100, the non-resident parent pays £120 and the parent-with-care will only receive £96. For a detailed flow chart of how the new scheme works, see Figure A.2 in the appendix. It is worth noting that parents who make private arrangements can still use the formula as a guide for how much they should be paying/receiving.

Figure 2.1 shows the proportion of separated mothers on each type of arrangement between 2011 and 2018. Private and CSA/CMS arrangements each make up around 20 percent of all arrangements, but the largest group of separated mothers are those without any arrangement. This has remained the case over time. In 2014 there is a rise in private arrangements which coincides with the policy reform, though CSA/CMS arrangements are flat, despite the introduction of additional charges in order to use the service.

Figure 2.1 Use of different types of Child Maintenance arrangements over time



Child maintenance schemes can vary substantially from country to country. In some European countries, such as Denmark, Finland and Iceland, the state guarantees a minimum amount of maintenance with the hope of recovering it from the NRP. The US presents a more complicated case, with individual states controlling the level of benefit disregard and pass-through applied.[3] In essence, no disregard and pass-through have similar implications as a guaranteed maintenance scheme; the lone parent will receive only the social security benefit regardless of NRP compliance with his/her liability. The UK does not guarantee maintenance but it does allow a full benefit disregard and pass-through. Since 2010, welfare recipients receive their benefits in addition to any payment received from the NRP. Skinner et al. (2012) offers an interesting comparison of policy aims, schemes and support system of five countries.

2.3 Literature Review

Many studies observe that children from higher income families have better educational, income and employment outcomes in young adulthood. However, only a handful of studies address causality. Do the children of wealthier parents perform better because of what money can buy? Or do the unobserved abilities that made their parents higher earners manifest themselves in the way the child is raised? Or is it that ability is transmitted genetically or culturally, and those children become high earners like their parents, but not as a result of parents' higher incomes?

Researchers have tried to overcome this endogeneity problem and find mixed evidence of the effects of parental income on outcomes. Dahl & Lochner (2012) exploit variations in the Earned Income Tax Credit (EITC) in the US to find that a $1000 increase in income raises math and reading scores by six percent of a standard deviation. It is questionable whether the EITC, by encouraging low wage mothers to work longer hours, is affecting child outcomes purely through an income effect. On the other hand, Shea (2000) uses the union status of a father to instrument for parental income and finds no effect on investment in child human capital. The endogeneity of union status has been called into question in the context of union wage differentials. There, it has been suggested that union members have better unobservable labour market productivity than non-union workers. It is plausible to question whether

[3] Recently Colorado became the only state that allows a full disregard and pass-through. 20 states currently do not have a benefit disregard and pass-through. see: https://www.ncsl.org/research/human-services/state-policy-pass-through-disregard-child-support.aspx

unionised workers are also more productive in domestic production in a way that affects child outcomes.

Even if we could rely on the evidence above there are a number of reasons why a unit of child maintenance might not have the same effects as a unit of other forms of income to a PWC. Firstly, prefacing maintenance with "child" could encourage PWCs to spend the entirety of it on the child(ren) (as its intended recipients), therefore it may have a larger effect on well-being and child outcomes than other sources of income (Del Boca & Flinn (1994)). Evidence from child benefit provides an analogy as its ostensible purpose is also to support the child. Kooreman (2000), using Dutch data, finds that there is a larger marginal propensity to consume child clothing out of child benefit than from other income sources for single parent families. He interprets this as a "labelling effect".[4]

Secondly, the payment of child maintenance has a theoretically ambiguous impact on family relationships. These are potential mechanisms through which child outcomes are affected. Child maintenance payment could elicit further NRP involvement because they want to monitor how their money is spent (Weiss & Willis, (1985)). Some evidence suggests parents who pay child maintenance are more likely to have frequent contact with their children because they have a stronger commitment to the child ((Amato & Gilbreth (1999), Daniela Del Boca & Ribero (2001), Bradshaw et al. (1999), Wikeley et al, 2008). Whilst other evidence suggests that a binding child maintenance order may act as a substitute for contact (Ermisch (2008)), or even that maintenance might be the price of contact.

However, this evidence is not robust to selection on unobservables. Rossin-Slater & Wüst (2018) present causal evidence from Danish administrative data, which suggests an additional 1000DK increase in child maintenance obligation reduces father-child co-residence in at least one-year post-separation by 1.8 percentage points. Co-residence refers to the shared time of each child with the parents. To causally identify the effects of child maintenance income, they assign each non-resident father his child maintenance obligation in the separation year using only information on his income and number of children in the separation year. They use this as an instrument for his average child maintenance income over all years post-

[4] In contrast, Blow et al. (2012) find that, in the UK, child benefit is spent disproportionately on alcohol and "adult assignable goods" by middle and higher income lone parents. This suggests that child maintenance could be 'misspent' in the same way. However, both datasets pre-date child support. The results thus relate more to a mental accounting story.

separation. This should account for any behavioural responses to his child maintenance liability. Although, the methodology is convincing, co-residence is not the same as frequency of contact and the administrative dataset does not allow the authors to directly measure frequency of contact when the child and non-resident parent are living apart.

Not only is the relationship between NRP-child contact and child maintenance receipt ambiguous, it is often strongly related to the degree of inter-parental conflict. There may be opposing effects on child well-being depending on the level of contact and conflict between parents. If payment of child maintenance helps to maintain contact between the non-resident parent and child this could result in better relationships and reduced conflict between parents that would be beneficial for child outcomes. Alternatively, it might lead to more bitter interactions and conflict between parents, that would have negative implications for child outcomes.

Moreover, there has been some research that suggests the child might act as a signal to the non-resident parent. For example, a better behaving child might result in an increased likelihood that the NRP pays child maintenance or increases the amount paid, raising questions of reverse causality (Aughinbaugh (2001)).

Early work in the child maintenance and child well-being literature used multivariate analyses to obtain correlational evidence, with the majority of work being conducted in the US. Baydar & Brooks-Gunn (1994) examined a sample of children living with both biological parents in 1986 where the father has left by 1988. Adding controls for contact with the father and comparable test scores of children in 1986, they find children whose mothers do not receive child maintenance have significantly lower test scores than those who receive it. Knox & Bane (1994) use the Panel Study of Income Dynamics (PSID) and controlling for the absent father's education level, find a positive effect of child maintenance on grades completed and likelihood of high school graduation. More recently Nepomnyaschy et al. (2012) use the Fragile Families and Child Wellbeing study to examine the effects of both formal and informal child support on cognitive skills and behaviour of children aged 5. Highlighting the importance of selection into child maintenance, their empirical strategy rests on controlling for as many differences between those who receive and do not receive child support as possible. However, as with all these studies, there remains a concern of omitted variable bias. If there are unobserved characteristics, of either the mother or father, associated with the ability to claim/pay child maintenance which are correlated with child outcomes, estimates of the effect of CM will be biased. The negative estimated relationship found may be a reflection of some selection effect rather than some genuine causal effect.

In response some papers have tried to use instrumental variables to identify causal effects. Graham et al. (1994) use the US Current Population Survey (CPS). Controlling separately for whether the father lives in the same state, visitation rights of the father, and number of days of child-father contact they find significant and beneficial effects of child maintenance income on years of schooling, high school dropout, college entry and falling behind age cohort in school; an effect size at least five times as large as other family income. When an instrumental variable approach is taken, the effect size disappears completely. The authors predict child maintenance amount using a Tobit regression, using maternal socio-demographic characteristics and father contact and location as explanatory variables. They then use the predictions from this equation as an instrument for child maintenance. It suffers from validity concerns since the predicted amount is likely to be correlated with outcomes not only through the actual child maintenance received.

Knox (1996) uses the US National Longitudinal Study of Youth to examine the effect of child maintenance on children's achievement test scores and on the home environment. She finds evidence that an increase of $1000 per year in child maintenance payments improves the maths and reading test scores of elementary school-age children by around seven percent of a standard deviation (SD), a considerably greater effect than income from other sources. This finding remains stable for the instrumental variable specification. They instrument families' levels of total income and child maintenance income with state-level differences in employment rates, average child maintenance payments and the proportion of mothers receiving child maintenance payments in the state of the child's birth. The identification strategy acts through variations in the levels of child maintenance payments across states due to differences in the CM formula and enforcement. The instrument will not satisfy the exogeneity condition if economic conditions in the state influence child outcomes through any channel other than household income.

The NLSY oversamples young and more socially disadvantaged mothers, therefore Argys et al. (1998) update the analysis with a representative sample two years later and find the effects disappear using the same specification as Knox (1996). In addition, they estimate a new specification controlling for race and out-of-wedlock births and find the positive effect of child maintenance receipt on child cognitive outcomes is only present for blacks who have separated/divorced and whites who had a non-marital birth.

The only paper the author knows of in the UK which examines child maintenance and child outcomes is Walker & Zhu (2011). They use the British Household Panel Survey (BHPS) to obtain probit estimates of the effect of child maintenance on the likelihood of leaving school at 16 and of obtaining 5 or more good GCSE's – the criteria used in the UK by senior high schools for progression into an academic track. The findings indicate the amount of child maintenance has a strong positive effect - ten times as large as the effect from other household income sources. They then use the mothers' retrospective history as instruments for child maintenance income and household income (fertility, relationship and employment) and find the effect sizes are robust to the new specification. However, these instruments could be invalid if, for example, the mother's employment pre-childbirth is correlated with both educational outcomes and income.

While the literature in general finds a beneficial effect of child maintenance payments on child outcomes, researchers have not yet managed to persuasively ascertain that this relationship is not driven by selection. Although instrumental variable approaches have been used, it is difficult to find instruments that are credible. In light of this, I employ a fixed effects approach that removes the confounding associated with time-invariant unobservables. I also use an approach developed by Oster (2019) to test the robustness of coefficients to selection in least squares modelling.

2.4 Data and descriptive statistics

Understanding Society is a nationally representative sample which began in 2009 and is one of the largest panel surveys in the world, interviewing 40,000 households annually. The advantage of using Understanding Society over other data sources is not only the size of the sample but the rich information it contains about separation related topics. It collects information on child maintenance - amounts received, calculated entitlement, type of arrangement, timeliness, and reasons for no maintenance received. It also contains a non-resident parent module in which the respondent answers questions about their non-resident ex-partner, including their employment status, re-partnership, fertility and age, along with the level of contact and quality of relationship between the two parents. This information is only collected in the third, fifth, seventh and ninth waves which correspond to the years 2011-2019.

I use a sample of children, who live with their mother as a result of their mother being separated from the father and are between the ages of 10 and 15 and so provide responses to a youth self-completion module. Although it would be important to consider the outcomes of children who live with their father, these children account for less than 10 percent of the sample and separate analysis would be greatly underpowered, so these cases were, regrettably, excluded.

Ideally, it would be possible to control for in-wedlock and out-of-wedlock births as well as pre-separation characteristics of both the father and mother. Obtaining both mother and father pre-separation characteristics is possible but relies on separation happening within the survey time period. Complicating matters further, child maintenance variables are only observed in waves 3, 5, 7 and 9. There are only 102 youths who experience separation during the survey **and** whom are observed, in at least one of the four waves, in receipt of maintenance.

In the sample there are 2,744 individual children eligible for child maintenance between the ages of 10 and 15. The data could be treated as pooled cross-sections or the panel nature of the dataset can be exploited to estimate random and fixed effect models.

There are two measures used to calculate child maintenance: a binary indicator for child maintenance receipt and the actual amount received. Separately estimating the effect of receiving *any* child maintenance from variations in the amount received is important for policy makers and subsequent policy design. Child maintenance receipt is set equal to 0 if the respondent has no child maintenance arrangement or if there is an arrangement but the respondent reports no money being transferred.[5] It is equal to 1 if there is an arrangement that results in some money being paid, regardless of the type of arrangement and the level of compliance. The amount of child maintenance refers to weekly payments of child maintenance, excluding any alimony. The mother reports how much she receives in weekly child maintenance which is adjusted to 2018 prices and equivalised using the modified OECD equivalence scale. The equivalence scale assigns a weight of 1 to the first adult (person aged 14 or older) in the household, a weight of 0.5 to each additional adult, and a weight of 0.3 to each child (person aged 0-13). The data shows that 37 percent of separated mothers receive child maintenance. The equivalised average amount of maintenance received, conditional on

[5] Arrangement can refer to court order, CSA/CMS, or private/informal agreement

receiving anything, is £36 a week. Figure 2.2 shows that the distribution of child maintenance payments are heavily skewed towards zero.

I focus on five outcomes measuring social and behavioural problems of youths aged 10 - 15. They are based on an established strengths and difficulties questionnaire (SDQ) administered as part of the youth self-completion module. It screens for potential psychiatric disorders (Goodman (1997)). Youths are asked 25 questions on a three-point scale “not true”, “somewhat true” and “certainly true”. Each question is categorised into one of the five outcomes: conduct problems, emotional symptoms, hyperactivity/inattention, peer relationship problems, and pro-social behaviour. The responses to individual questions comprising each outcome are allocated points and the youth gets a total score per outcome, measured on a scale of 0 to 10 (see Table A.1 in the appendix for questions comprising each sub-scale). Figure 2.3 shows conduct, emotional and peer relationship problems are positively skewed, a higher number indicating more problems. For pro-social behaviour the scale is reversed, a higher number referring to better pro-social skills.

Figure 2.2 – Distribution of child maintenance receipt

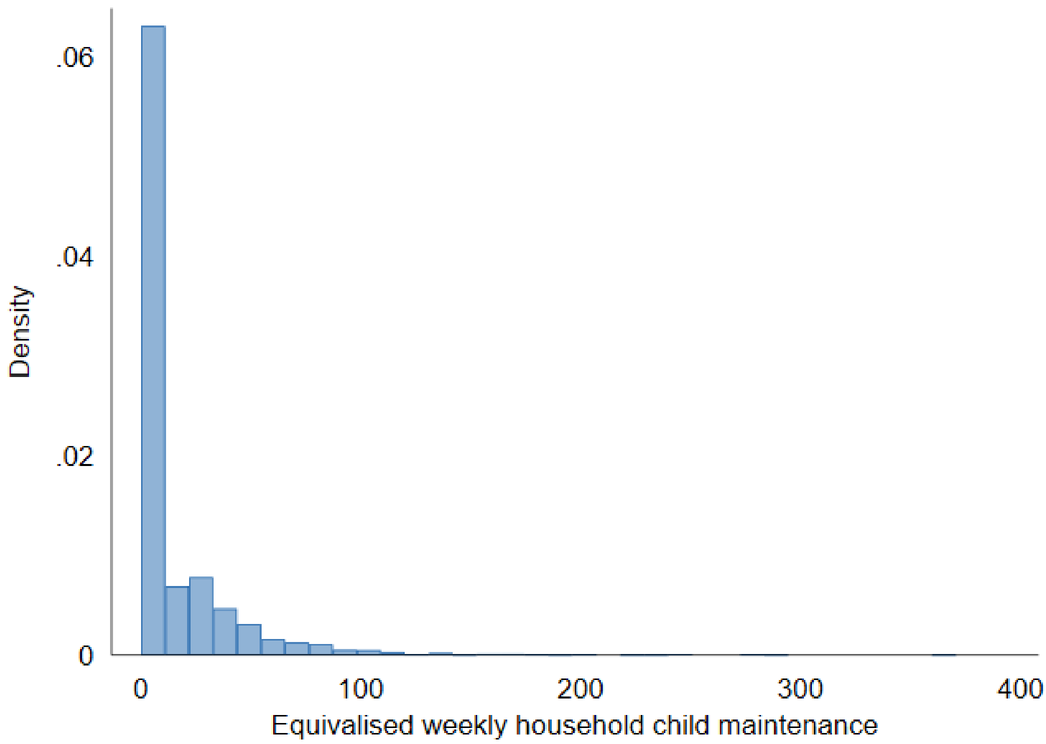


Figure 2.3 – Distribution of each of the five behavioural and social outcomes for youths aged 10-15

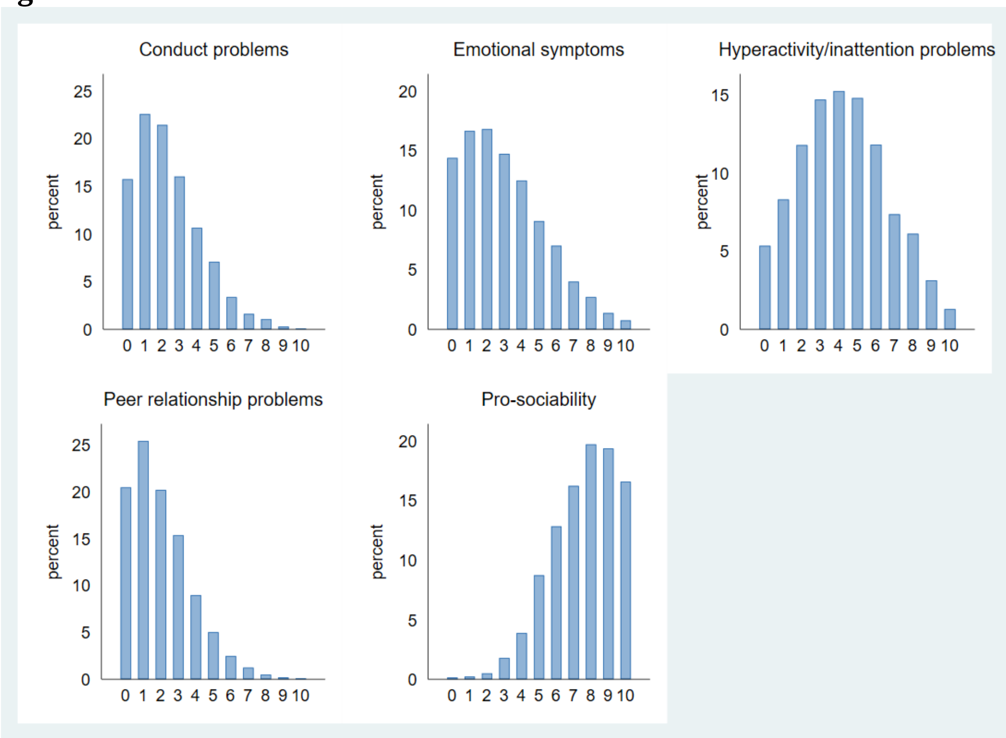
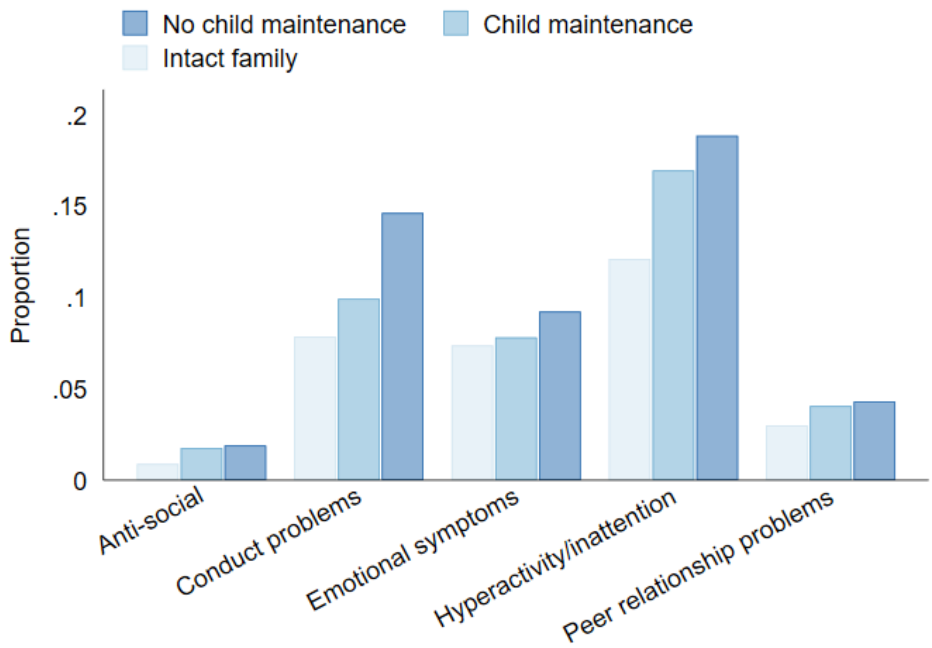


Figure 2.4 shows children who receive child maintenance have fewer conduct problems, fewer emotional symptoms and generally fewer negative outcomes than children who do not receive it. Consistent with the broader literature, children in intact families do better on all these outcomes.

The data contains a set of controls for the characteristics of the child, the mother and to a lesser extent the NRP. The child controls include age, gender, number of step-, half- and natural siblings, and a dummy variable for a step-parent being present.[6] Mother controls include age, non-white ethnicity, education level, monthly labour income, marital status and employment status. Having an informal child maintenance arrangement, such as buying clothes, toys, school trips, or paying for school fees etc., is also included as a control variable as it may be a substitute for child maintenance payments.

[6] The age of the child at the point of separation is likely to be important for outcomes but we cannot obtain this information for children whose parents separate outside the survey period. Unfortunately, this constitutes the majority of respondents.

Figure 2.4 – Social and behavioural outcomes by child maintenance receipt



There are also controls for child-NRP term-time contact and the quality of the relationship between the NRP and the PWC. Contact is measured as a continuous variable on a scale of 0 to 365, where 365 means they see each other every day and 0 means they never have any contact, reducing the variation from having lots of categories. Relationship quality is measured on a three-point scale where 0 refers to never sees them, 1 indicates an unfriendly relationship and 2 a friendly relationship.

Table 2.1 shows that there are many missing values for some of the variables used as controls, particularly NRP characteristics. Table 2.2 reports descriptive statistics for mothers by family type and whether they receive child maintenance. Of those who receive child maintenance only 10 percent are non-white, compared with 27 percent not receiving child maintenance. There is some evidence from the US that suggests this may be due to a higher incidence of out-of-wedlock births among some ethnic groups (black ethnicity) (J. W. Graham & Beller, 1996). And this corresponds with a lower likelihood of having a child maintenance award for previously unmarried women. Mothers who receive CM are significantly better educated, with higher income, are older, more likely to own their own home, be employed, and to have re-partnered. They are also less likely to receive Income Support. This is a strong indication of selection into treatment i.e. child maintenance receipt. There is a stronger selection effect when looking at contact and inter-parental conflict; 80 percent of mothers who receive child maintenance have contact with their ex-partner compared to 52 percent of mothers not receiving anything. This statistically significant difference is likely to capture some short-lived non-cohabiting relationships. Figure 2.5 also illustrates how receiving child maintenance is positively correlated with child-NRP contact; over 40 percent of NRPs who never have contact with their child also do not pay child maintenance.

Table 2.1: Summary statistics

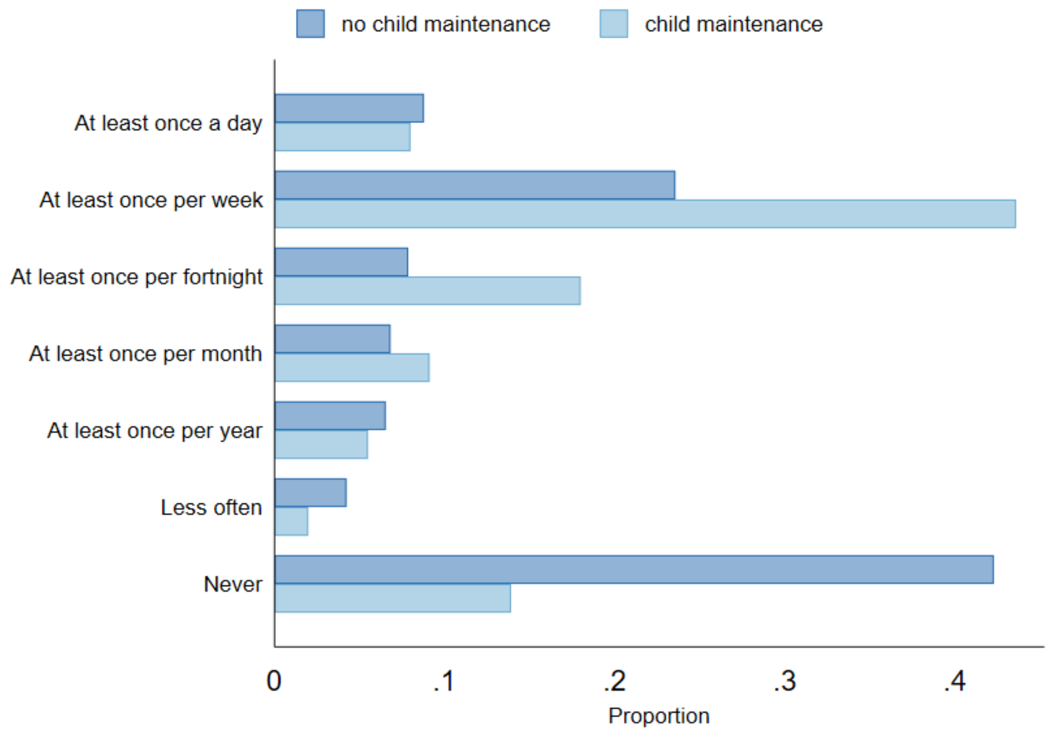
	N	Mean	St.Dev	min	max
Child age	4076	12.626	1.669	10	15
Female	4074	.503	.5	0	1
No. of natural siblings	4076	1.136	1.165	0	11
No. of half siblings	4076	.301	.674	0	5
No. of step siblings	4076	.053	.319	0	4
Has a step-parent	4019	.517	.5	0	1
Conduct problems	4019	2.352	1.866	0	10
Emotional symptoms	4018	3.015	2.33	0	10
Hyperactivity/inattention problems	4017	4.238	2.384	0	10
Peer relationship problems	4020	2.034	1.757	0	10
Pro-sociability	4022	7.543	1.887	0	10
Mother age	4075	40.393	6.32	25	62
Mother non-white	4068	.206	.405	0	1
Maternal labour income (Monthly 2018 prices)	3759	943	1183	0	10296
Informal help	3853	.291	.454	0	1
Receives child maintenance	3418	.378	.485	0	1
NRP age	3205	44.165	7.224	19	85
NRP has new partner	1237	.474	.5	0	1
NRP has new children	2095	.397	.489	0	1
NRP is employed	1242	.808	.394	0	1

Table 2.2: Descriptive Statistics

	(1) Intact family	(2) Receives CM	(3) Not receive CM	Difference (2) - (3)
Panel A – Child characteristics				
Female	0.50	0.49	0.51	-0.021
	(0.500)	(0.500)	(0.500)	
Age	12.46	12.66	12.63	0.034
	(1.695)	(1.648)	(1.671)	
No. of natural siblings	1.58	1.06	1.18	-0.124**
	(1.137)	(0.960)	(1.292)	
No. of half-siblings	0.04	0.31	0.28	0.037
	(0.232)	(0.651)	(0.652)	
No. of step-siblings	0.01	0.06	0.05	0.009
	(0.104)	(0.365)	(0.286)	
Has step-parent	0.16	0.55	0.49	0.061***
	(0.367)	(0.498)	(0.500)	
Panel B – Mother characteristics				
Age	42.84	41.32	40.02	1.302***
	(5.548)	(5.931)	(6.407)	
Non-white	0.25	0.10	0.27	-0.170***
	(0.431)	(0.302)	(0.445)	
Currently married	0.92	0.20	0.17	0.023
	(0.273)	(0.397)	(0.378)	
Employed	0.72	0.76	0.60	0.165***
	(0.451)	(0.425)	(0.490)	
Hours usually worked per week	26.46	27.72	26.61	1.107*
	(10.22)	(9.790)	(9.738)	
Monthly labour income (2018 prices)	1.24	1.22	0.84	372.999***
	(1.404)	(1.275)	(1.159)	
Uses childcare	0.24	0.33	0.22	0.109***
	(0.424)	(0.471)	(0.417)	
No formal qualifications	0.17	0.13	0.22	-0.093***
	(0.379)	(0.336)	(0.416)	
Has a degree	0.43	0.42	0.32	0.104***
	(0.495)	(0.494)	(0.466)	
Monthly equivalised hh income (2018 prices)	1741.73	1501.52	1256.63	244.883***
	(2572.0)	(867.2)	(712.9)	
On income support	0.02	0.09	0.15	-0.057***
	(0.148)	(0.288)	(0.355)	
Owns home	0.79	0.59	0.30	0.286***
	(0.406)	(0.492)	(0.459)	
Other parent helps in informal way	0.13	0.46	0.21	0.244***
	(0.337)	(0.498)	(0.411)	
Has some contact with ex-partner	0.80	0.75	0.46	0.292***
	(0.414)	(0.431)	(0.499)	
On friendly terms with ex-partner	1.23	1.30	1.18	0.220***
	(0.439)	(0.621)	(0.816)	
N	9185	1293	2125	3418

Note: mean coefficients; sd in parentheses

Figure 2.5 – Child maintenance receipt and the level of child contact with non-resident parent



Unusually, it is possible to control for NRP characteristics, such as age, new partnership and fertility, and employment status. But, unfortunately, these controls are only available for a sub-sample of mothers. The questionnaire routing means that the mother is only asked these questions about the NRP if she reports having some contact with him (see Figure A.1 in the Appendix for routing). Table 2.3 shows that there are important differences by contact. Mothers in contact with the NRP are more likely to have higher sociodemographic status, own their own home, are much more likely to receive child maintenance, and the average amount received is considerably higher (£15 a week compared with just £5).

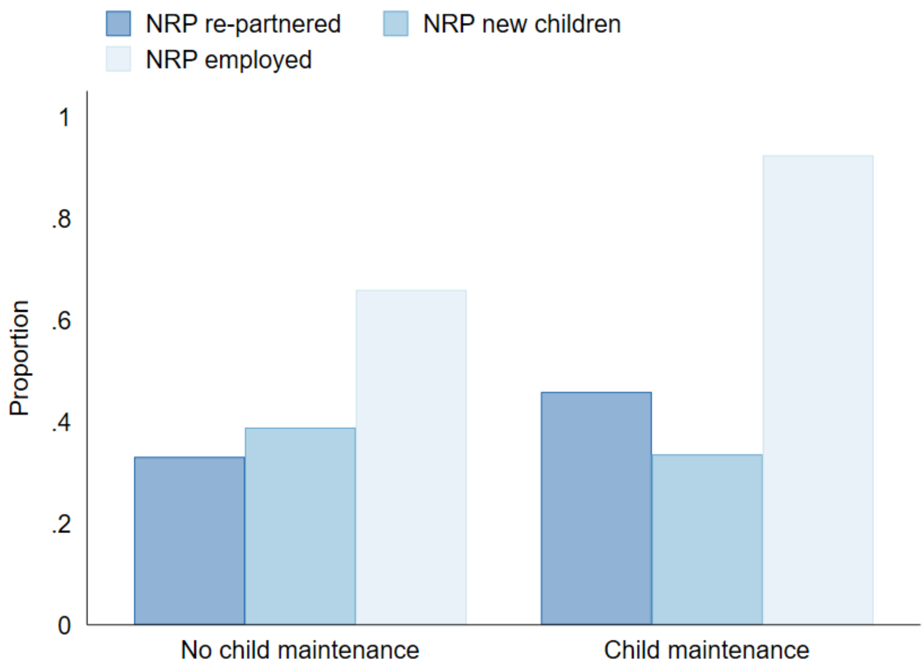
Table 2.3: Sub-sample summary statistics

	(1) Whole sample	(2) No contact	(3) Has contact (sub-sample)
Age	40.39	39.57	40.66
	(6.320)	(6.405)	(6.269)
Non-white	0.21	0.19	0.21
	(0.405)	(0.395)	(0.408)
Currently married	0.18	0.21	0.17
	(0.384)	(0.406)	(0.376)
Employed	0.65	0.61	0.66
	(0.477)	(0.488)	(0.473)
Hours usually worked per week	26.82	26.86	26.80
	(9.892)	(10.00)	(9.861)
Monthly labour income (2018 prices)	0.94	0.84	0.98
	(1.183)	(1.084)	(1.215)
Uses childcare	0.27	0.24	0.28
	(0.442)	(0.426)	(0.447)
No formal qualifications	0.19	0.20	0.19
	(0.393)	(0.398)	(0.392)
Has a degree	0.35	0.32	0.36
	(0.478)	(0.467)	(0.481)
Monthly equivalised hh income (2018 prices)	1332.70	1309.14	1341.25
	(759.3)	(704.2)	(778.3)
On income support	0.13	0.12	0.14
	(0.339)	(0.322)	(0.344)
Owns home	0.40	0.36	0.41
	(0.490)	(0.481)	(0.492)
Other parent helps in informal way	0.29	0.11	0.36
	(0.454)	(0.309)	(0.479)
On friendly terms with ex-partner	1.25	1.07	1.37
	(0.750)	(0.967)	(0.523)
Receives CM	0.38	0.24	0.42
	(0.485)	(0.427)	(0.494)
Weekly amount of CM received	12.68	5.86	14.99
	(28.36)	(15.83)	(31.16)
N	4076	1009	3067

Note: mean coefficients; sd in parentheses

Finally, Figure 2.6 shows that there are differences in child maintenance receipt associated with the NRP characteristics. Over 90 percent of non-resident fathers who pay child maintenance are employed, compared with a 65 percent employment rate for the non-payers.

Figure 2.6 - NRP characteristics by child maintenance receipt



2.5 Methodology

We are interested in the effect of mothers' child maintenance on outcomes for the child. Consider the following equation:

$$Y_{it} = \beta_0 + \beta_1 CM_{it} + \beta_2 Contact_{it} + \beta_3\, Relationship\ quality_{it} + \beta_4 X_{it} + Year_t + Region_i + \epsilon_{it} \qquad (1)$$

where Y_i is the child's behavioural/social outcome of interest measured on a continuous scale of 0 to 10, such as conduct problems. β_1 measures the effect of the parent receiving child maintenance, which is the coefficient of interest. β_2 and β_3 measure of the effect of term-time contact the child has with the NRP and quality of relationship between the separated parents, which is not censored by contact. β_4 contains a vector of control characteristics including mother's socio-demographic characteristics, child characteristics and NRP characteristics.[7] Interview year and region fixed effects are included.

The main concern with estimating equation (1) using ordinary least squares (OLS) is dealing with potential bias that confounds the estimates and prohibits researchers from arguing that the estimated effects are causal. Estimates, in particular β_1, will be biased if any right-

[7] For a sub-sample of respondents, censored by the mother being in contact with the NRP.

hand variable is correlated with the error term. If there are unobserved omitted factors that influence both child maintenance receipt and the outcome, then pooled OLS estimates will be biased and inconsistent. The direction of the bias is unclear because it depends on the nature of the distribution of unobservables that cannot be measured. Researchers have tried to address this issue by controlling for as many potential confounding variables that their data allows. However, of course, a limitation of this approach is the impossibility of knowing if all relevant control variables have been measured and included in the models. Indeed, there is no guarantee that adding more covariates will reduce such bias. Parenting ability might be such a case - if a 'better' mother is also more motivated to demand child maintenance, this is likely to be correlated with the outcome variables. For example, fewer conduct problems through better discipline, better at solving emotional crises, and better at encouraging pro-social behaviour in her children.

A further example would be the extent of NRP commitment to the child. Using child maintenance receipt as a binary indicator could proxy the level of commitment of the non-resident father. For example, a small amount of CM such as £1 a week, though possibly having a small effect on the child outcomes, might indicate more about a father's concern for the well-being of his child, which would otherwise go unobserved.

The novel aspect of this dataset is being able to control for child-NRP contact and parents' relationship quality, and by doing so, this may reduce omitted variable bias. However, they could be considered "bad controls" (Angrist & Pischke (2009)) if both the outcome and contact/relationship quality are simultaneously caused by child maintenance receipt. A plausible example is if the mothers' willingness for the non-resident father to have contact with the child or the mother's friendliness with him, is in part determined by him paying child maintenance.

A fixed effects model can exploit the panel nature of the data and should, arguably, be an improvement over OLS, accounting for some of the unobserved heterogeneity by differencing out time invariant unobservables. Due to the nature of the data, panel methods can only be used for the four waves that contain child maintenance questions. These questions are asked every two years and given that the sample of youths are aged between 10 and 15, a child can be observed a maximum of three times.

So long as the unobserved parenting ability or commitment does not change over time, a fixed effects model will reduce the bias from these omitted variables. If some mothers are more persistent in actively pursuing child maintenance awards, they could also be stricter in parenting (i.e. reducing conduct problems or increasing emotional symptoms). If this ambition/motivation does not change over time, then fixed effects will capture this heterogeneity. A fixed effects model should strip out unobserved heterogeneity among respondents so long as it remains constant over time. This is represented by α_i (see equation (2)) which captures the individual-specific intercept and is known as the entity fixed effect. The variation in the entity fixed effects comes from omitted variables that vary across entities but not over time.

$$Y_{it} = \beta_0 + \beta_1 CM_{it} + \beta_2 Cont_{it} + \beta_3\, Rel\ qual_{it} + \beta_4 X_{it} + Year_t + Region_i + \alpha_i + \epsilon_{it} \quad (2)$$

On the other hand, random effects should provide more precise estimates than fixed effect estimation. Thus there is a benefit from using a random effects model – provided there is some understanding of the process through which selection into child maintenance receipt occurs, and there is sufficient data. If there is reason to believe that differences across individuals have some influence on the outcome variable then a random effects specification should be used. The variance can be partitioned into two components: the between-individual variation and the within-child variation. The crucial assumption for the RE model, is that the unobserved time-invariant heterogeneity is uncorrelated with the regressors. This is necessary for the consistency of the random effects model but not for the consistency of fixed effects. If there is no correlation between the regressors and unobservables, then fixed and random effects are both consistent, but the fixed effects estimates are inefficient. It is possible to test this using a Hausman test, with the null hypothesis being that they are uncorrelated, and thus the random effects preferred.

However, measurement error is second potential source of bias. If there is error in mothers mis-reporting either CM receipt or amounts, in one or more of the time periods of treatment, then fixed effect estimates are likely to be more susceptible to attenuation bias associated with this measurement error than OLS estimates would be. This could be a problem either where the treatment is a binary indicator of CM receipt or if the treatment is the amount received.

There would also be an issue if there is not enough variation in the treatment over time - since this is the source of the identification. Although there is a maximum number of 3 observations per child, with an average of 1.3, child maintenance has some variability within individuals over time. Of the mothers not receiving child maintenance initially, almost 13 percent go on to receive in the future, whilst 28 percent of mothers who receive maintenance in one wave do not receive anything in the following wave they are observed.

There is still merit in using pooled OLS estimation. It is common to assess the extent to which selection on unobservables is driving OLS results by examining the stability of the treatment coefficient to the inclusion of additional control variables. However, Oster (2019), is critical of the intuition that coefficient stability upon adding controls can be informative about unobserved selection in the model. She shows that any change in coefficient size from adding an additional control variable should be scaled by the change in the R^2 to account for the amount of variation in the outcome that the control can explain. Hence, I make use of the test outlined in her paper which can be used in two separate ways.

First, the test can be used to infer the degree of unobserved selection that would need to exist to reduce the magnitude of the treatment effect to zero. This is the δ value. A value greater than 1, signifies that the unobservables would need to be more important than observables in order to explain away the result. An important parameter used in the Oster test is the maximum R^2 of the regression, which theoretically is 1. However, as Oster points out, this is unattainably high even in experimental studies, due to measurement error. Following an analysis of a range of top journal articles Oster (2019) concludes that the maximum R^2 should be at least a multiple of 1.3 times as high as the R^2 from the initial regression, while two times higher would be a more rigorous standard. In this context, considering the R^2 is very low, setting an R_{max} value of 1 appears much too high, although it is reported nevertheless. Even 1.3 is a small value when the dependent variable is constructed by aggregating ordinal responses to subjective scales. This is sure to impose a severe form of measurement error on the dependent variable. The threshold for robustness in this case is one – equal observed and unobserved selection.

The second capability of the Oster test is to bound estimates assuming a particular degree of unobserved selection. This is known as the β value. The aim is to bound the estimates based on the assumption that selection on unobservables is, at worst, equal to selection on observables ($\delta = 1$) . Formally, the bias-adjusted treatment effect of the coefficient of interest β* is:

$$\beta^* \approx \tilde{\beta} - \delta\left(\dot{\beta} - \tilde{\beta}\right)\frac{R_{max} - \tilde{R}}{\tilde{R} - \dot{R}}$$

The test does not enable causal inference, but it substantially augments the usefulness of OLS estimates.

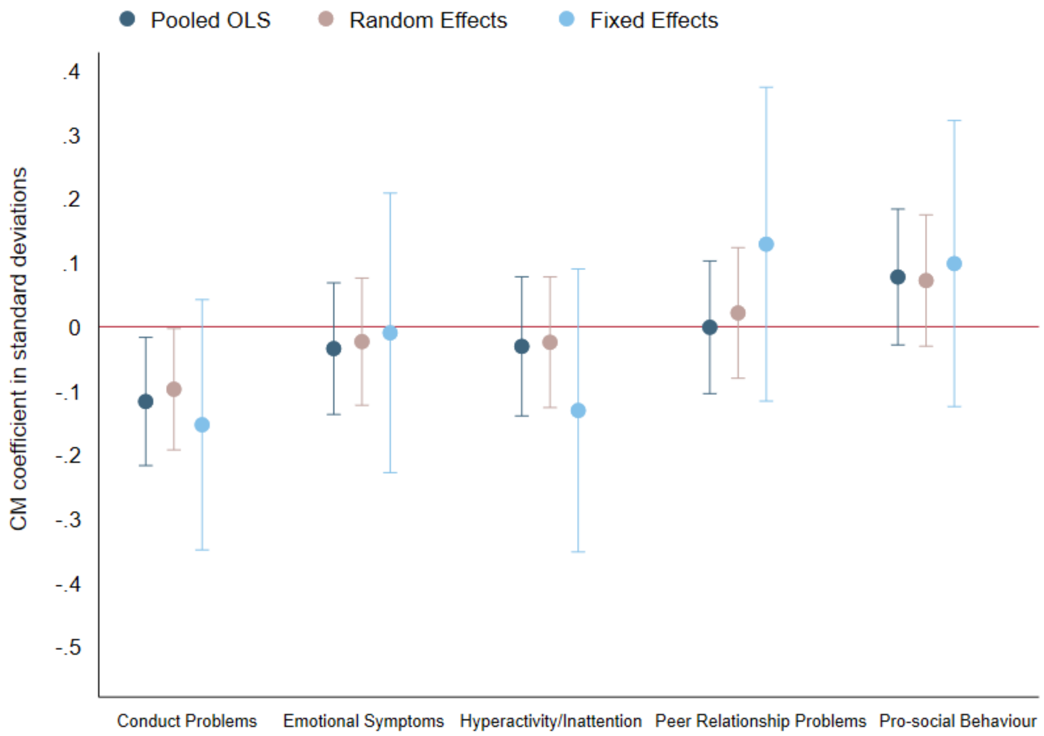
2.6 Results

In Section 2.6.1 the estimates for the binary indicator for CM receipt are reported, in Section 2.6.2 gender differences are examined, Section 2.6.3 presents estimates on a sub-sample of mothers using NRP characteristics as controls, and in Section 2.6.4 regressions of the amount of CM received are discussed.

2.6.1 Child maintenance receipt

Figure 2.7 summarises the pooled OLS, random and fixed effects estimate on CM receipt where the dependent variable has been standardised. While fixed effects estimates are in general larger, with bigger standard errors, the Hausman test rejects them in favour of random effects for all outcomes. Random effects and pooled OLS estimates are consistently similar in magnitude across outcomes, which is indicative of the models capturing any selection effects. Conduct is the only outcome that displays an effect size which is significantly different to zero and suggests that receiving child maintenance reduces conduct problems by around 12 percent of a standard deviation – this is a sufficiently large effect size to motivate consideration of policy interest in the finding. Conduct problems will be discussed in more detail below.

Figure 2.7 – Pooled OLS, random and fixed effects estimates of CM for each outcome



Note: The markers are point estimates taken from the preferred specification. The bars refer to 95% confidence intervals.

The pooled OLS (panel A), random effects (panel B) and fixed effects (panel C) regression coefficients on CM receipt are reported for the outcome conduct problems in Table 2.4. Results for the other four outcomes can be found in the Appendix (Tables A.2-A.5). Due to the skewed distribution of the outcome variable (see Figure 2.3), the results of a negative binomial regression are presented in the footnote of each table and indicate that non-linearity is not a cause for concern. Columns (1) to (7) incorporate controls for child characteristics, mother characteristics, child contact with father and quality of the mother's relationship with her ex-partner, as indicated at the bottom of each table. In column (4) a dummy variable is included for whether the mother receives any informal help from the non-resident parent, such as paying school fees or contributing to other costs such as clothing and school trips. In column (5), the NRP age is included and column (6) adds a control for child-NRP term-time contact.

Parental relationship quality is added as a control in column (7). There may be concerns about potential correlation between child-NRP contact and parents relationship quality. However, this is unlikely as contact is recorded at the child level (how often the child sees the NRP during term-time), whereas the relationship quality variable asks the mother how friendly she is with her ex-partner. The data shows over 50 percent of children whose mothers are very unfriendly or not very friendly with their ex-partner, still have term-time contact with the NRP.

The pooled OLS and random effects specifications both show that the inclusion of additional controls does not reduce the size of the CM coefficient. Indeed, almost all coefficients are bigger than the uncontrolled regression in column (1). The estimates are remarkably robust to the addition of socio-demographic controls for the mother and child along with the NRP age, and a dummy variable for informal receipt. An interesting finding is that the coefficient on child maintenance does not change when adding a control for whether the mother receives informal help, implying that informal help is not crowding out the effects of formal child maintenance. The fixed effects results in Panel C show the coefficient on CM receipt remains very similar across columns as more controls are added, suggesting that unobserved heterogeneity is being well captured.

I would expect that contact is important for child outcomes, particularly if it is proxying commitment of the NRP to the child. Including contact as a control has a theoretical basis, if contact is positively correlated with receiving CM and contact is associated with a reduction in conduct problems, then omitting it would bias the estimate of CM downwards. But contact can have a potentially ambiguous relationship with CM. The PWC can use CM as a bargaining tool in which she permits the NRP contact with the child only if payment is received. On the other hand, some NRPs may see paying CM as a substitute for having contact with the child. And for these reasons contact might be a “bad control” (Angrist & Pischke (2009)). Under the statutory formula an NRP could commit to a certain number of overnight stays in order to reduce their CM liability, though there is nothing to prevent the NRP from reneging on this later – particularly if the agreement is made privately as there is no state enforcement.

Weighing up these options it seems most likely that contact is an important omitted variable, though I cannot rule out the possibility of it being a bad control. However, should contact be a bad control, the effect of including it might reduce the size of the coefficient on CM. Ultimately, the addition of contact in column (6) of Table 2.4 does not change the size of the coefficient on CM in both pooled OLS estimates and random effect models, and it remains significant. This result is reasonable in convincing us that contact is not a “bad control”. Contact does not appear important in of itself in the regressions (see Table A.6 in appendix). This is in line with previous literature (Baydar & Brooks-Gunn (1994, Knox & Bane (1994) and Graham, Beller & Hernandez (1994)). The measure of contact is fairly precise so CM is unlikely to be picking up any effects that are not captured through measurement error in the contact variable. Receiving child maintenance is associated with a reduction in youth conduct problems of 0.215 (0.179 in random effects) on a ten-point scale, equivalent to 12 (10) percent of standard deviation.

Turning to relationship quality, the variable in the survey for is, at best, a proxy for the actual quality of relationship between parents. There is almost certainly a mismatch between the actual relationship quality of separated parents and what the child observes. They may act in a friendly fashion for the child's sake and indeed, "better" parents might be more likely to have a friendly relationship so as not to upset the child. Interpreted as measurement error, this will bias the coefficient towards zero, i.e it's effect will be attenuated. This bias may offset the omitted variable bias since the latter is likely to induce upward bias.

The causal pathway is not clear cut when it comes to parental relationship quality. There are two distinct types of relationship quality; pre-separation and post-separation relationship quality. Parents relationship quality pre-separation is likely to be correlated with the decision to pay/receive CM and thus may be an important omitted variable. Unfortunately, there are fewer than 250 observations where pre-separation relationship quality can be observed in the data. For these individual's relationship quality pre-separation is uncorrelated with relationship quality post-separation.

It is possible to observe the friendliness of parents relationship post-separation. In theory, this may be determined by CM – non-payment might lead to a more strained relationship. Indeed, the DWP actively state that child maintenance is intended to facilitate a friendlier relationship between parents, and this is expected to improve child outcomes. Adding post-separation relationship quality as a control in column (7) reduces the coefficient size on CM receipt by a third compared to the estimate in column (6) and it is no longer statistically significant (although they are not significantly different from each other). The result is unclear, including post-separation relationship quality in regression models could, if it is a potential mediator, give the impression that receiving CM has no effect, when in reality it is acting through relationship quality.

The coefficient on relationship quality is not significant (see Table A.6 in appendix) and it would appear relatively unimportant in explaining conduct problems. There is also the potential for reverse causality, where a child with bigger conduct problems might result in a strained relationship between parents, as they argue over their approach to tackle this behaviour. Therefore, it is difficult to theoretically predict the effect of relationship quality.

In light of problems raised by relationship quality and its likelihood of being a bad control, it seems reasonable to regard column 6 (where contact is included and relationship quality omitted) as the preferred specification. Although the Hausman test rejects in favour of random effects, the coefficient sizes are very similar across pooled OLS, random and fixed effects specifications. This is encouraging and suggests the results are capturing unobserved selection with the control variables, and the random effect specification estimates are not suffering from omitted variable bias. To further investigate the degree to which unobservables might be driving the result found in column (6), I perform the test first proposed by Altonji et al. (2005) and further developed by Oster (2019).

Table 2.4 - Regression of youth conduct problems on CM receipt

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Youth conduct problems			
Panel A - POLS							
Receives CM	-0.211***	-0.279***	-0.288***	-0.260***	-0.222**	-0.215**	-0.151
	(0.071)	(0.071)	(0.080)	(0.081)	(0.086)	(0.095)	(0.104)
R-squared	0.003	0.051	0.067	0.067	0.069	0.080	0.091
Panel B - Random effects							
Receives CM	-0.176***	-0.240***	-0.258***	-0.238***	-0.187**	-0.179**	-0.138
	(0.065)	(0.066)	(0.075)	(0.076)	(0.082)	(0.090)	(0.100)
Panel C - Fixed effects							
Receives CM	-0.166	-0.177	-0.260*	-0.241	-0.214	-0.285	-0.280
	(0.130)	(0.130)	(0.149)	(0.150)	(0.160)	(0.186)	(0.234)
R-squared	0.001	0.024	0.049	0.046	0.043	0.062	0.088
Observations	3,373	3,342	2,664	2,644	2,171	1,886	1,528
Hausman test	0.933	0.049	0.383	0.492	0.515	0.433	0.607
Panel D - Oster test δ's in POLS specification							
$R_{max} = 1$		-0.183	-0.766	0.702	0.170	0.173	0.232
$R_{max} = 2\tilde{R}^2$		-3.125	-9.941	9.265	2.193	1.954	2.299
$R_{max} = 1.3\tilde{R}^2$		-8.598	-28.519	27.398	6.669	5.998	7.379
Child characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes
Mother characteristics	No	No	Yes	Yes	Yes	Yes	Yes
Informal receipt	No	No	No	Yes	Yes	Yes	Yes
NRP age	No	No	No	No	Yes	Yes	Yes
Child-NRP contact	No	No	No	No	No	Yes	Yes
Mother-NRP friendliness	No	No	No	No	No	No	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Notes: The dependent variable is measured on a scale of 0-10 and is formed from 5 individual questions relating to youth conduct (see Table A.1 in appendix for individual components). Negative binomial regression coefficient is smaller but remains negative and statistically significant.

The Oster test deltas are reported in Panel D of Table 2.4 over three rows, with each row corresponding to different values of R_{max}. It suggests, in the preferred specification in column (6), that the share of variation explained by unobservable variables would need to be between 2 to 6 times as large as that of the share of variation explained by observable variables for the coefficient on child maintenance receipt to be driven to zero. This is an implausible degree of variation to attribute to the unobservables, and therefore I conclude that selection bias is not likely to be a sufficient explanation of the results.

It is also possible, using the Oster method, to generate a lower bound on the estimates of child maintenance. In Oster's paper she uses the example of maternal employment on wages, where there is relative consensus that ability is one of few omitted variables and selection on observables and unobservables operates in the same direction. However, in this analysis, it is ambiguous whether the selection runs in the same direction. Some of the important omitted variables discussed earlier include commitment to the child and ability/motivation to get child maintenance (from both PWC and NRP), in which selection would work in the same direction. But unobserved non-resident father characteristics, such as engaging in anti-social behaviour (smoking, drugs, gambling, alcohol etc.), might have the opposite effect and selection would move in the other direction. Therefore, bounds are reported for when $\delta = 1$ and $\delta = -1$. Since the direction of selection bias is unclear, I think of both the $\delta = 1$ and $\delta = -1$ as providing a lower bound on the CM estimate, but the two characterise the nature of omitted variable bias differently.

Table 2.5 presents the bounding estimates for the preferred specification (Column 6 of Table 2.4). The bounds for $\delta = 1$ and $\delta = -1$ are very similar and remarkably close in size to the baseline model (our preferred specification). The take away from the tightness of the bounds is that the model is not suffering from omitted variable bias.

The results indicate there is a plausibly negative association between child maintenance receipt and conduct problems. The Oster test is an indication of robustness of results, rather than a guarantee of causality, but taken in combination with the analysis above and the broad agreement between pooled OLS, random and fixed effects, I have some confidence in saying that selection is not driving the results. That is the delta estimates for multiples of R^2 of 1.3 or 2 are far above 1 – Oster's benchmark value, implying that selection on unobservables would need to be (six times in column 6) larger than the selection on observables in order to bias the coefficient on CM to 0.

Table 2.5 - Oster coefficient bounds on CM receipt coefficient for δ = 1 and δ = - 1

Dependent Variable	(1) Restricted model	(2) Baseline model	(3) Bound for β* for δ =1	(4) Bound for β* for δ = -1
Conduct problems	-0.211***	-0.215**	-0.202**	-0.232**
	(-0.071)	(0.095)	[-0.087]	[0.092]
R-squared	0.003	0.079		

Notes: Robust standard errors in parentheses, bootstrapped standard errors in square brackets; *** p<0.01, ** p<0.05, * p<0.1 The restricted model in column 1 refers to the case where conduct problems is regressed on CM receipt and an intercept. The baseline model in column 2 is the full specification where conduct problems is regressed on CM receipt and the full set of controls (from column 6 Table 2.4). In column 3 and 4 the bounds are reported for proportional unobserved selection that moves in the same and opposite direction to observable selection. Standard errors are bootstrapped with 1000 replications. All lower bounds are based on the R_{max} set to $1.3\tilde{R}$.

2.6.2 Gender differences

Table 2.6 shows that there are significant differences in the impact of CM receipt by gender of the child for conduct and pro-social skills. OLS and random effects estimates are found in columns (1) to (4). The fixed effects estimates have been included in column (5) and (6) for completeness. However, I do not attach much weight to these estimates given that the small sample size leads to negligible variation in receiving child maintenance between waves.[8]

For girls I find that CM does not have a beneficial effect for any of the tested outcomes. The previous finding that receiving CM reduces conduct problems is entirely driven by improvements in conduct for boys. This effect size is now -0.401 in the OLS regression (column 1), corresponding to 21 percent of a standard deviation. The coefficient estimate is almost double that of the coefficient in the regression with both genders included (see Table 2.4). The coefficient on pro-sociability for males is large and statistically significant, corresponding to 16 percent of a standard deviation, where for females it is much smaller and statistically insignificant. For all outcomes, the OLS and random effects estimates are consistently similar in size and not statistically different from each other, which is reassuring.

[8] The fixed effects estimates do contradict OLS and random effects estimates for the conduct problems outcome, suggesting that CM receipt reduces conduct problems only for females. However, I do not use these estimates for the reasons stated in the text. In addition, the coefficient sizes on the male and female sub-samples are not statistically different from each other and the female sub-sample is marginally statistically significant.

It is unclear what is driving the gender difference found in these results. Though there is some evidence in favour of a same-sex role model which improves outcomes ((Doherty & Needle (1991)), it is difficult to see how this would explain the gender difference found here unless child maintenance acts as a proxy for being a good role model. The only coefficient that shows a significant difference between genders is on child maintenance. There is some evidence, mainly correlational, which suggests parental separation affects boys and girls differently. Boys tend to exhibit more externalising behaviour such as conduct where girls experience more internalizing problems ((Bertrand & Pan, 2013), (Autor et al., 2019)). Receiving child maintenance may act as a buffer for conduct problems that arise from separation, which occur mainly in boys. Wasserman (2020) in the Future of Children: "Because of their higher risk of behavioural problems, boys may need additional inputs to produce the same outcome". Child maintenance may be one such input.

Table 2.6 - Gender differences using sub-samples of male and female youths for all outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS		RE		FE	
	Males	Females	Males	Females	Males	Females
Outcome of interest						
Conduct						
Receives CM	-0.401***	-0.104	-0.345**	-0.111	-0.323	-0.484*
	(0.143)	(0.129)	(0.137)	(0.122)	(0.254)	(0.256)
Emotions						
Receives CM	-0.035	-0.230	-0.041	-0.207	-0.108	-0.371
	(0.156)	(0.189)	(0.153)	(0.184)	(0.341)	(0.434)
Hyperactivity						
Receives CM	-0.128	-0.134	-0.123	-0.104	-0.170	-0.602*
	(0.191)	(0.193)	(0.182)	(0.179)	(0.380)	(0.346)
Peer relationships						
Receives CM	-0.149	0.111	-0.102	0.127	0.053	0.181
	(0.132)	(0.137)	(0.130)	(0.136)	(0.269)	(0.366)
Pro-social skills						
Receives CM	0.308**	0.084	0.283**	0.092	0.013	0.463
	(0.150)	(0.145)	(0.144)	(0.140)	(0.299)	(0.291)
Observations	941	947	941	947	941	947
R-squared	0.081	0.101			0.112	0.157

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

2.6.3 Sub-sample

So far, the entire sample has been used in the analysis. However, I have not been able to control for the set of NRP characteristics which include whether he has new children, a new partner, and his employment status. Existing data on NRPs is limited and surveys often do not follow a parent once they leave the household. This makes it challenging for researchers to control for NRP characteristics when looking at the effects of separation on child or mother outcomes. In our dataset it is only possible to do this for a sub-sample of the population - mothers who have contact with the NRP. This selective sample may differ in both observable and unobservable ways. In particular, the distribution of CM receipt is quite different - 57 percent of the sub-sample receive CM compared with just 44 percent of the whole sample. If a NRP changes his behaviour as a result of the CM obligation, ideally, pre-separation characteristics should be used as controls. Unfortunately, in this dataset it is not possible to do without reducing the sample size so much that it is underpowered and unable to conduct meaningful analysis.

Table 2.7 presents the coefficients on child maintenance and NRP current partnership status, employment status, and whether he has new children. Because this is a sub-sample, the estimates are not directly comparable to those of the whole sample (column 6 of Table 2.4). It is possible, however, to compare the sub-sample coefficient on CM with estimates from a regression only controlling for contact, as in the preferred specification in column 6 of Table 2.4. Each outcome variable is reported in the table and has two columns attached to it, the first is the sub-sample without controls for the NRP characteristics and the second is with controls so I can observe how stable the coefficient is to adding these controls. For reasons mentioned earlier, the relationship quality between the NRP and PWC is not included as a control. The results show that, as expected, for all significant outcomes the coefficient on child maintenance gets bigger with the inclusion of the characteristics of the non-resident parent, suggesting that previous estimates were a lower bound. The coefficient on CM is significant for youth conduct problems, emotional symptoms and pro-sociability. Interestingly, the CM coefficient on emotional symptoms and pro-sociability is not significant in the whole sample, but significant in the sub-sample. This could reflect effects being stronger for mothers who have contact with the NRP.

For conduct problems the coefficient on CM for pooled OLS and random effects are similar in size and statistically significant (see Panels A and B in column 2 of Table 2.7). The coefficients are also very similar in size to the fixed effects estimates in Panel C which are preferred by the Hausman test but lack precision. Receiving CM corresponds to a reduction in conduct problems by 15 percent of a standard deviation in pooled OLS. This effect size is highly robust to selection on observables. Although the fixed effects estimates are often favoured by the Hausman test, the estimates can still suffer from omitted variable bias if the unobserved heterogeneity is time varying.

The Oster δ's are reported in Table 2.8 for both the coefficient on child maintenance and the coefficients on the NRP characteristics. The Oster test on the outcome conduct problems in column 2 of Table 2.8, shows setting an R_{max} of $1.3\tilde{R}$, selection on unobservables would need to be 21 times as large as selection on observables in order to reduce the effect size to zero. In fact, even when assuming the maximum possible R_{max} of 1, the coefficient passes the robustness test.

Table 2.7 - Sub-sample regression of child outcomes on CM receipt

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A - Pooled OLS	Conduct	Conduct	Emotions	Emotions	Hyper	Hyper	Peers	Peers	Pro-social	Pro-social
Receives CM	-0.263**	-0.296**	-0.295*	-0.355**	-0.120	-0.118	-0.110	-0.083	0.280**	0.316**
	(0.124)	(0.129)	(0.163)	(0.177)	(0.183)	(0.194)	(0.131)	(0.134)	(0.135)	(0.141)
NRP partner		0.078		0.486***		0.359**		0.060		-0.063
		(0.132)		(0.175)		(0.177)		(0.137)		(0.136)
NRP children		0.063		-0.140		-0.025		-0.288**		-0.341**
		(0.128)		(0.163)		(0.183)		(0.136)		(0.134)
NRP employed		0.106		0.065		-0.148		-0.135		-0.120
		(0.163)		(0.212)		(0.245)		(0.177)		(0.182)
R-squared	0.142	0.143	0.135	0.145	0.062	0.067	0.067	0.073	0.148	0.156
Panel B - Random effects										
Receives CM	-0.248**	-0.288**	-0.278*	-0.347**	-0.133	-0.142	-0.089	-0.066	0.289**	0.325**
	(0.126)	(0.132)	(0.161)	(0.173)	(0.183)	(0.193)	(0.127)	(0.131)	(0.133)	(0.139)
NRP partner		0.090		0.485***		0.383**		0.052		-0.059
		(0.132)		(0.175)		(0.177)		(0.135)		(0.134)
NRP children		0.113		-0.086		0.007		-0.221*		-0.320**
		(0.126)		(0.163)		(0.185)		(0.134)		(0.131)
NRP employed		0.141		0.106		-0.119		-0.115		-0.126
		(0.167)		(0.212)		(0.240)		(0.173)		(0.179)
Panel C - Fixed effects										
Receives CM	-0.303	-0.329	0.189	0.070	-0.998**	-0.985**	0.359	0.356	0.825**	0.880**
	(0.336)	(0.335)	(0.421)	(0.404)	(0.461)	(0.485)	(0.345)	(0.319)	(0.344)	(0.356)
NRP partner		-0.262		-0.007		-0.075		-0.510		0.047
		(0.483)		(0.579)		(0.647)		(0.421)		(0.454)
NRP children		1.235***		-0.040		1.165		0.300		-1.226**
		(0.447)		(0.768)		(0.748)		(0.609)		(0.539)
NRP employed		0.004		1.567***		-0.821		0.842		0.043
		(0.629)		(0.534)		(0.555)		(0.516)		(0.730)
R-squared	0.317	0.348	0.360	0.387	0.250	0.265	0.285	0.317	0.247	0.267
N	843	843	843	843	842	842	840	840	843	843
Hausman test	0.019	0.021	0.019	0.020	0.461	0.555	0.074	0.032	0.982	0.981

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Note: The coefficient on contact is not reported however, is insignificant and close to zero for all specifications. In odd columns controls are included for child and mother characteristics, informal receipt, NRP age, and NRP-child contact. In even numbered columns controls for NRP's new partner, new children and employment status are included.

Emotional symptoms are significant and similar across the random effects and pooled OLS specifications. However, the Hausman test rejects the null, preferring the fixed effects model which are much closer to zero and not statistically significant. The Oster test on emotional symptoms shows that the delta is negative, meaning that if the observables are positively correlated with the treatment, the unobservables would need to be negatively correlated with the treatment. This is analogous to adding more controls strengthening the size of the CM coefficient on the outcomes, rendering it unlikely the result is driven by unobservables. The coefficient easily passes the test, with a δ of -42[9].

The coefficient on CM is not significant for youth hyperactivity/inattention and peer relationship problems, reported in columns (6) and (8) of Table 2.8. In the case of pro-social behaviour (column 10), on the other hand, the OLS and random effects coefficient sizes are large and statistically significant, implying that for separated families with closer ties, CM is associated with an improvement in children's pro-social behaviour. This could be due to parents being in contact having a positive effect through "leading by example".

The Oster bounds are reported in columns (3) and (4) of Table 2.9 and are presented for the coefficient on CM for the three significant outcomes: conduct problems, hyperactivity/inattention and pro-sociability. Since I cannot again be certain that δ should be positive I compute both the upper and lower bounds and, again, the bounds are close to the baseline estimate.

[9] A negative delta suggests that selection on unobservables would need to have the opposite sign as the selection on observables, such that failing to control for unobserved factors would lead to downward bias in the estimates.

Table 2.8 - Oster δ's of sub-sample regression of youth outcomes on CM receipt

	(1) Conduct	(2) Conduct	(3) Emotions	(4) Emotions	(5) Hyper	(6) Hyper	(7) Peers	(8) Peers	(9) Pro-social	(10) Pro-social
Dependent variable										
Receives CM	-0.263**	-0.296**	-0.295*	-0.355**	-0.120	-0.118	-0.110	-0.083	0.280**	0.316**
	(0.124)	(0.129)	(0.163)	(0.177)	(0.183)	(0.194)	(0.131)	(0.134)	(0.135)	(0.141)
$\boldsymbol{\delta}$										
$R_{max} = 1$	1.253	1.156	-1.178	-2.161	0.136	0.190	-0.659	-0.173	0.795	0.706
$R_{max} = 2\tilde{R}^2$	7.271	6.804	-6.750	-13.488	1.876	2.852	-8.345	-2.391	4.178	3.965
$R_{max} = 1.3\tilde{R}^2$	**22.269**	**21.021**	**-20.932**	**-42.329**	6.152	9.316	-27.428	-7.733	**12.847**	**12.257**
NRP re-partnered		0.078		0.486***		0.359**		0.060		-0.063
		(0.132)		(0.175)		(0.177)		(0.137)		(0.136)
$\boldsymbol{\delta}$										
$R_{max} = 1$		0.481		-2.003		1.328		-0.076		0.350
$R_{max} = 2\tilde{R}^2$		2.879		-11.179		17.296		-0.960		1.890
$R_{max} = 1.3\tilde{R}^2$		9.532		**-32.565**		**50.049**		-3.175		6.278
NRP new children		0.063		-0.140		-0.025		-0.288**		-0.341**
		(0.128)		(0.163)		(0.183)		(0.136)		(0.134)
$\boldsymbol{\delta}$										
$R_{max} = 1$		0.096		-0.230		-0.011		-0.232		0.611
$R_{max} = 2\tilde{R}^2$		0.577		-1.347		-0.154		-2.755		3.188
$R_{max} = 1.3\tilde{R}^2$		1.916		-4.434		-0.514		**-7.852**		**9.643**
N	843	843	843	843	842	842	840	840	843	843

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

Notes: The deltas are reported for all outcomes (columns) and dependent variables for completeness sake, the deltas of interest (where there is a significant coefficient) are in bold.

Table 2.9 - Sub-sample Oster bounds on CM coefficient for δ =1 and δ = -1

Outcome	(1) Restricted model	(2) Baseline model	(3) Bound for β* for δ =1	(4) Bound for β* for δ = -1
Conduct problems	-0.258	-0.296**	-0.313**	-0.282**
	(0.119)	(0.129)	(0.142)	(0.123)
R-squared	0.0055	0.143		
Emotional symptoms	-0.236	-0.355**	-0.406**	-0.312*
	(0.155)	(0.177)	(0.161)	(0.181)
R-squared	0.0028	0.145		
Pro-social behaviour	0.298	0.316**	0.324*	0.309*
	(0.125)	(0.141)	(0.147)	(0.125)
R-squared	0.0067	0.156		

Notes: The restricted model in column 1 refers to the case where the outcome is regressed on CM receipt and an intercept. The baseline model in column 2 is the full specification where the outcome is regressed on CM receipt and the full set of controls (from column 6 Table 2.4). In column 3 and 4 the bounds are reported for proportional unobserved selection that moves in the same and opposite direction to observable selection. Standard errors are bootstrapped with 1000 replications. Calculations are based on the R_{max} being set to $1.3\tilde{R}$.

The coefficient sizes on the NRP controls are also highly informative. It would appear the NRP having a new partner is significantly likely to increase youth emotional symptoms and hyperactivity/inattention. Indeed, for both outcomes the coefficient size on new partnership is larger and the opposite sign to that of CM receipt, suggesting that the NRP re-partnering has an adverse effect on youth emotional and hyperactivity/inattention problems, offsetting any beneficial effects of CM receipt. The effect size is large and statistically significant.

The NRP having new children also significantly reduces peer relationship problems, this could be because they have more interactions with younger step-siblings. This is at odds with the result found for pro-social behaviour which suggests new children reduce the child's pro-social behaviour, perhaps because they are fighting for the non-resident fathers' attention. The positive and significant effects of CM on pro-sociability are counteracted by the NRP having new children. These results highlight the importance of controlling for the NRP characteristics - in particular, for family dynamics post-separation. This study is the first, as far as I am aware, to control for these important omitted variables. Although these results may not be generalisable to the sample of mothers who have no contact with the NRP, they still represent an important and novel finding.

2.6.4 Child maintenance amounts

One might expect the amount of child maintenance to have a different effect on child outcomes to other sources of income, and this section attempts to test this hypothesis. In this dataset, it is possible to separate household income into three sources: weekly maternal labour income, weekly new partner income, and weekly household CM income. Partners' income is coded as 0 if there is no partner of the mother living in the household. There are 349 children to which this applies. New partner's income is included as a separate regressor as it is not necessarily the case that a parent in a new partnership would pool their income, particularly if one (or both) parent/s are receiving maintenance from outside the household.

The same concerns exist as in the previous section; the amount of CM a mother receives, depends not only on her own personal characteristics, but also the absent fathers'. Controlling for some of these socioeconomic characteristics will capture this directly, but some factors may not be observed. Are mothers who are able to extract **more** child maintenance from the NRP also better able to reduce conduct problems in children, address emotional problems, reduce hyperactivity and so on? And similarly, are fathers who pay more CM, also more interested in the child's well-being and development? Including the new partner's income could be problematic if more motivated mothers are also more motivated to find another partner (or even a higher earning one). It may also be a mediator through which CM payments work – the higher the amount, the more likely the mother is able to go out to find a new partner. The regressions are estimated including this potential bad control, and the results are found to be similar in size. There may also be more selection in CM amounts, as 'better' parents may be able to negotiate more maintenance.

The amount of CM could exhibit more measurement error than using a binary indicator for child maintenance receipt, since mothers may misreport the income they receive or, as is often the case, there may be unreliability in the amount received. The fixed effects model is particularly vulnerable to measurement error in the explanatory variable.

Modelling the effect of this income (along with maternal labour income and partner income) in a linear fashion might be overly restrictive and, therefore, the regressions include a squared term for each of the income variables. The theoretical basis for its inclusion is that as the amounts of child maintenance get larger, there is less benefit to the child from one additional £1 of child maintenance. Tables 2.10 and A.7-A.10 are laid out as described in the

CM receipt section - reporting the coefficients on household equivalised CM, maternal labour income and partner labour income.

Of the five variables of interest, the amount of CM only shows a significant and negative relationship in explaining differences in the case of conduct problems. For youth hyperactivity/inattention, peer relationship problems, pro-sociability and emotional symptoms, no effect is observed.

Panel A of Table 2.10 reports the pooled OLS effect of CM on conduct problems is small and indicate a negative effect. The pattern of coefficient stability across specifications is unsurprisingly very similar to the earlier CM receipt section for pooled OLS. The random effects estimates, however, lose significance much earlier when adding controls – adding the NRP age is sufficient. The coefficient drops slightly on the addition of informal receipt, but not by much, again suggesting that informal help is not a substitute for the amount of CM received.

Contact is added as a control in column (6), yet the coefficient remains stable. The significance of the quadratic term on CM, is indicative that, as theorised previously, the relationship is non-linear. Although this effect size is small (much smaller than the effect size for maintenance receipt), it is an order of magnitude larger than both maternal labour income and partner income, both of which have a small and not statistically significant effect on conduct. Due to lots of missing data in lone mothers reporting CM amounts, the sample sizes are smaller than in the earlier child maintenance receipt regressions and this is likely to be a problem for the fixed effects estimates in particular.

The Oster test results in Table 2.11 indicate that when choosing an R_{max}=1.3$\tilde{R}$, selection on unobservables would need to be twice as large as selection on observables to make the effect size disappear. This is not robust in the case of a more conservative assumption that R_{max}= 2$\tilde{R}$. The Oster lower bounds for this coefficient are found in Table 2.12 and are significantly different from zero and similar to the baseline estimate, at least for the case where δ = 1.

So far, the role of parents' relationship quality has not been discussed. As mentioned earlier, relationship quality was likely a bad control in the CM receipt context. For the amount of maintenance, however, this may not be the case. Because there is time variability in the amounts received, relationship quality may not be determined by, or at least be as strongly predicted by, the amount of CM paid at a point in time. On the other hand, the amounts also reflect the level of compliance. If the mother is not receiving as much as she should, or if the recovery of the child maintenance is slow and difficult and payments are irregular, this could lead to more conflict with her child's father. On balance, it is likely to be a bad control and will still suffer from measurement error. It is, nevertheless, added as a control in column 7, and decreases the size of the coefficient by 20 percent in the pooled OLS.

It is likely that a binary indicator of CM receipt is capturing some of the effects of commitment as well as being less vulnerable to measurement error. However, the comparison of CM amounts and receipt is informative for policy as it can shed light on the direction of policy: collecting more often or collecting more money. These results indicate the focus of policy should be on enforcement actions that encourage payment of CM, regardless of how small.

Table 2.10 - Regression of youth conduct problems on equivalised household CM amount

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A - POLS							
Eq. HH CM	-0.046***	-0.051***	-0.044**	-0.039**	-0.038**	-0.040**	-0.032
	(0.015)	(0.015)	(0.018)	(0.018)	(0.019)	(0.020)	(0.022)
Eq. HH CM sq.	0.001**	0.001**	0.001*	0.001*	0.001*	0.001*	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Maternal income			-0.002	-0.001	-0.000	-0.001	-0.003
			(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Partner income			-0.000	-0.001	0.000	-0.001	-0.000
			(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
R-squared	0.003	0.054	0.069	0.068	0.071	0.081	0.095
Panel B - Random effects							
Eq. HH CM	-0.035**	-0.041***	-0.036**	-0.033*	-0.026	-0.029	-0.024
	(0.014)	(0.015)	(0.017)	(0.017)	(0.018)	(0.019)	(0.021)
Eq. HH CM sq.	0.001*	0.001**	0.001*	0.001*	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Maternal income			-0.002	-0.002	-0.002	-0.001	-0.004
			(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Partner income			0.000	-0.000	0.001	-0.001	-0.001
			(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Panel C - Fixed effects							
Eq. HH CM	-0.007	-0.014	-0.012	-0.012	0.029	0.028	0.021
	(0.048)	(0.047)	(0.052)	(0.052)	(0.046)	(0.052)	(0.054)
Eq. HH CM sq.	0.000	0.000	0.000	-0.000	-0.001	-0.001	-0.002
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Maternal income			-0.002	-0.002	-0.004	0.000	-0.007
			(0.010)	(0.010)	(0.011)	(0.014)	(0.016)
Partner income			0.007	0.007	0.010	0.001	0.007
			(0.008)	(0.008)	(0.008)	(0.007)	(0.008)
R-squared	0.000	0.018	0.037	0.037	0.041	0.046	0.083
N	2,795	2,778	2,404	2,386	1,948	1,689	1,353
Hausman test	0.859	0.386	0.604	0.692	0.662	0.865	0.917
Child controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Mother controls	No	No	Yes	Yes	Yes	Yes	Yes
Informal receipt	No	No	No	Yes	Yes	Yes	Yes
NRP age	No	No	No	No	Yes	Yes	Yes
Child- NRP contact	No	No	No	No	No	Yes	Yes
Mother-NRP friendliness	No	No	No	No	No	No	Yes
NRP controls	No	No	No	No	No	No	No

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

Note: All income is indexed to 2018 prices and is reported in weekly terms and divided by 10, coefficients should be interpreted as an increase of £10 a week.

Table 2.11 - Oster δ's of sub-sample regression of youth conduct problems on CM amount

Conduct problems	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Eq. HH CM	-0.046***	-0.051***	-0.044**	-0.039**	-0.038**	-0.040**	-0.032
	-0.015	-0.015	-0.018	-0.018	-0.019	-0.02	-0.022
$R_{max} = 1$		0.222	0.203	0.114	0.066	0.074	0.076
$R_{max} = 2\tilde{R}^2$		3.695	2.665	1.523	0.848	0.827	0.715
$R_{max} = 1.3\tilde{R}^2$		10.869	8.261	4.799	2.680	2.614	2.308
Child controls	No	Yes	Yes	Yes	Yes	Yes	Yes
Mother controls	No	No	Yes	Yes	Yes	Yes	Yes
Informal receipt	No	No	No	Yes	Yes	Yes	Yes
NRP age	No	No	No	No	Yes	Yes	Yes
Child- NRP contact	No	No	No	No	No	Yes	Yes
Mother-NRP friendliness	No	No	No	No	No	No	Yes
NRP controls	No	No	No	No	No	No	No
N	2,795	2,778	2,404	2,386	1,948	1,689	1,353

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

Table 2.12 - Oster lower bounds on CM amount coefficient for δ =1 and δ = - 1

Conduct problems	(1) Restricted model	(2) Baseline model	(3) Bound for β* for δ =1	(4) Bound for β* for δ = -1
Eq. HH CM	-0.031	-0.040**	-0.058***	-0.035
	(0.0131)	(0.020)	[0.0159]	[0.0213]
R-squared	0.003	0.081		

Robust standard errors in parentheses, bootstrapped standard errors are in square brackets 1000 replications
*** p<0.01, ** p<0.05, * p<0.1
Notes: The restricted model in column 1 refers to the case where the outcome is regressed on CM receipt and an intercept. The baseline model in column 2 is the full specification where the outcome is regressed on CM receipt and the full set of controls (the preferred specification in column 6 of Table 2.10). In column 3 and 4 the bounds are reported for proportional unobserved selection that moves in the same and opposite direction to observable selection. Calculations are based on the R_{max} being set to $1.3\tilde{R}$.

The sub-sample analysis is also conducted on CM amounts (see Table A.11 in Appendix). In the sub-sample, equivalised household CM amount is not significant for any of the outcomes, including for conduct problems which was significant in the whole sample. The effects of the NRP having a new partner, new children and his employment status on child outcomes are in line with the results from the CM receipt sub-sample analysis.

2.7 Conclusion

This paper provides clear and compelling estimates of the impact of child maintenance on youth behavioural and social outcomes. The findings demonstrate the importance of analysing the various aspects of a child's well-being separately, because, interestingly, there appears to be no effect of child maintenance receipt on emotional symptoms, hyperactivity/inattention, peer relationships and pro-sociability. Receiving child maintenance is associated with a significant improvement in youth conduct. Investigating this further in a heterogeneity analysis using sub-samples of boys and girls shows that this result is driven by boys. The effect size is large at 25 percent of a standard deviation. Boys also see significant improvements in their pro-sociability while girls, on the other hand, do not appear to benefit from receiving child maintenance for any of our examined outcomes.

The methodology improves on past studies by employing fixed effects methods as well as including NRP characteristics for a sub-sample. The Oster test suggests selection on unobservables is not driving the result. The Oster bounds are not far from the OLS coefficient, and indeed this result holds when I explore the ambiguity of the direction of unobserved selection. The importance of the findings should not be understated. Behaviour in childhood and adolescence is an important predictor of later life outcomes. In the long-term, behavioural problems in childhood are associated with spending less time in employment in adulthood, an effect, similar to my results, only present for males ((Feinstein (2000) and Knapp et al. (2011)). Kokko and Pulkkinen (2000) show this relationship is mediated by shorter-term school outcomes (school success, interest in schoolwork and truancy). Child maintenance could have important long-term implications, reducing detrimental short- and long-term impacts for the child, and creating wider benefits to society.

The mechanism through which relationship quality affects outcomes is unclear but this could have important policy implications. If relationship quality is a mediator through which child maintenance has an effect on child conduct problems, then policy should not be aimed at expensive enforcement actions of child maintenance but rather facilitating relationship quality improvements among separated parents. A recent DWP project is already doing this which could create widespread benefits.[10]

The sub-sample analysis presents an interesting result that, in fact, the non-resident parents' new family is important in determining outcomes for the child "left behind", and this can undo the positive effects of receiving child maintenance.

There are many potential extensions to this work. Disentangling the effect of relationship quality, contact and child maintenance is an important area which needs more research. In addition, there are many outcomes which could be studied such as educational attainment and truancy. As more waves of data are released in Understanding Society it will become possible to follow the youths studied in this chapter into adulthood and examine the effect of child maintenance on longer term outcomes such as university attendance, employment, and earnings.

[10] See Reducing Parental Conflict programme: https://www.gov.uk/government/publications/reducing-parental-conflict-programme-information-for-stakeholders

Appendix A

Table A.1 - Calculation of each of the five outcome variables

Conduct problems scale	**Not true**	**Somewhat true**	**Certainly true**
I get very angry and often lose my temper	0	1	2
I usually do as I am told	2	1	0
I fight a lot. I can make other people do what I want	0	1	2
I am often accused of lying or cheating	0	1	2
I take things that are not mine from home, school or elsewhere	0	1	2
Emotional symptoms scale			
I get a lot of headaches, stomach-aches or sickness	0	1	2
I worry a lot	0	1	2
I am often unhappy, downhearted or tearful	0	1	2
I am nervous in new situations. I easily lose confidence	0	1	2
I have many fears, I am easily scared	0	1	2
Hyperactivity scale			
I am restless, I cannot stay still for long	0	1	2
I am constantly fidgeting or squirming	0	1	2
I am easily distracted, I find it difficult to concentrate	0	1	2
I think before I do things	2	1	0
I finish the work I'm doing	2	1	0
Peer problems scale			
I am usually on my own. I generally play alone or keep to myself	0	1	2
I have one good friend or more	2	1	0
Other people my age generally like me	2	1	0
Other children or young people pick on me or bully me	0	1	2
I get on better with adults than with people my own age	0	1	2
Prosocial scale			
I try to be nice to other people. I care about their feelings	0	1	2
I usually share with others (food, games, pens, etc.)	0	1	2
I am helpful if someone is hurt, upset or feeling ill	0	1	2
I am kind to young children	0	1	2
I often volunteer to help others (parents, teachers, children)	0	1	2

Figure A.1 - Question routing in the child maintenance questionnaire module

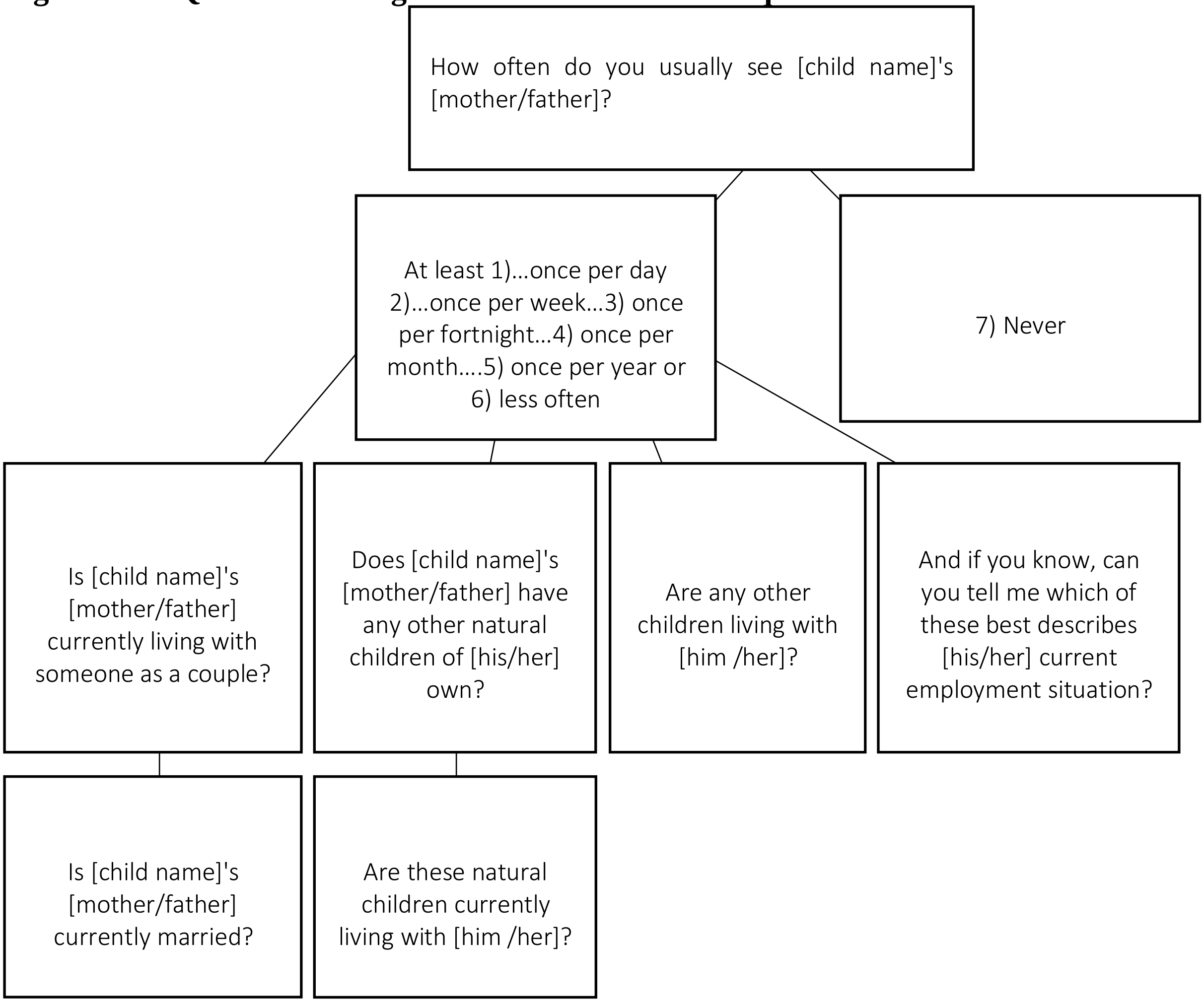


Figure A.2 - DWP flowchart of UK child maintenance scheme

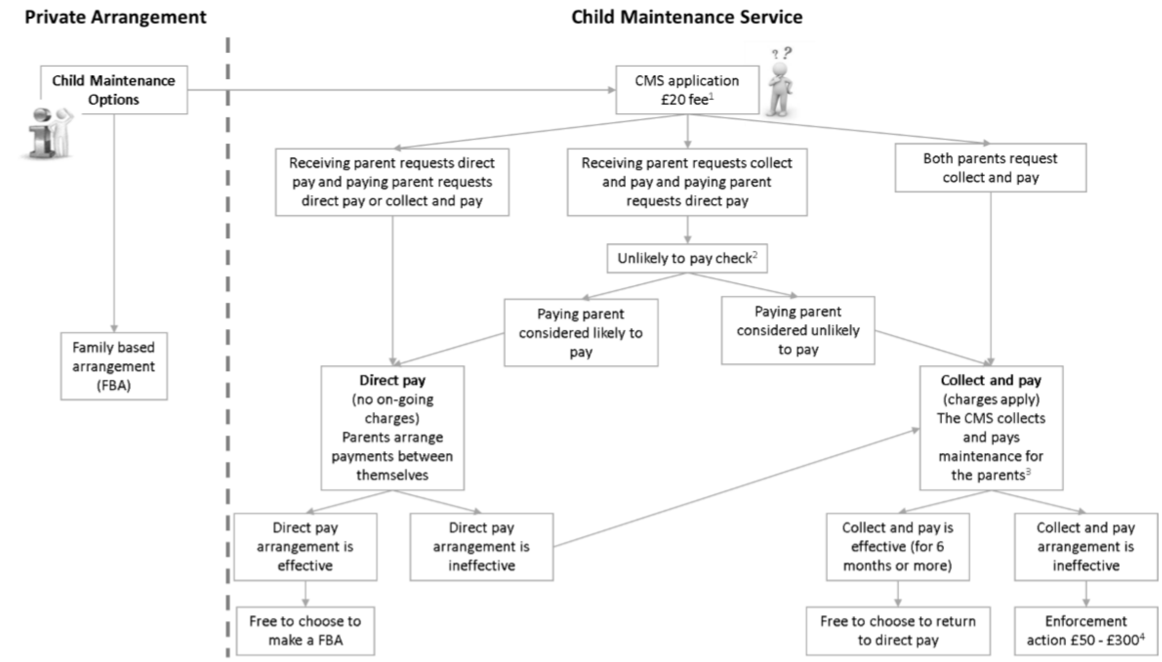


Figure 2 shows what happens when you make a child maintenance arrangement.

1 The application fee is waived for victims of domestic abuse and applicants under 19.
2 A paying parent's behaviour in the Child Maintenance Service is used to determine whether they are likely or unlikely to pay.
3 A 20% charge is added to the liability for the paying parent and 4% is deducted from the maintenance received by the receiving parent.
4 Enforcement charges of between £50 and £300 apply depending on the type of enforcement action taken.

Source: DWP report - Child Maintenance Reforms 30 Month Review of charging (2017)

Table A.2 - Regression of youth emotional symptoms on CM receipt

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Youth emotional symptoms						
Panel A - Pooled OLS							
Receives CM	-0.043	-0.109	-0.130	-0.112	-0.104	-0.080	-0.014
	(0.088)	(0.088)	(0.101)	(0.104)	(0.113)	(0.122)	(0.131)
R-squared	0.000	0.072	0.095	0.095	0.104	0.105	0.106
Panel B - Random effects							
Receives CM	-0.047	-0.092	-0.106	-0.090	-0.067	-0.054	0.001
	(0.084)	(0.084)	(0.098)	(0.100)	(0.108)	(0.118)	(0.127)
Panel C - Fixed effects							
Receives CM	-0.144	-0.102	0.024	0.038	0.079	-0.022	-0.021
	(0.177)	(0.173)	(0.197)	(0.198)	(0.211)	(0.259)	(0.318)
R-squared	0.001	0.037	0.078	0.078	0.084	0.084	0.155
N	3,372	3,341	2,663	2,643	2,175	1,887	1,529
Hausman test	0.567	0.691	0.375	0.419	0.482	0.490	0.508
Child characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes
Mother characteristics	No	No	Yes	Yes	Yes	Yes	Yes
Informal receipt	No	No	No	Yes	Yes	Yes	Yes
NRP age	No	No	No	No	Yes	Yes	Yes
Child contact with NRP	No	No	No	No	No	Yes	Yes
Mother-NRP friendliness	No	No	No	No	No	No	Yes
NRP characteristics	No	No	No	No	No	No	No

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Notes: As I am unable to use sampling weights in the random effects and fixed effects models, for comparability the unweighted coefficient is presented in pooled OLS. The weighted results however show a similar result. Negative binomial regression coefficient is similar in magnitude and remains statistically insignificant.

Table A.3 - Regression of youth hyperactivity/inattention on CM receipt

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Youth hyperactivity/inattention						
Panel A - Pooled OLS							
Receives CM	-0.054	-0.171*	-0.145	-0.131	-0.084	-0.073	-0.059
	(0.093)	(0.094)	(0.109)	(0.112)	(0.121)	(0.132)	(0.144)
R-squared	0.000	0.040	0.049	0.050	0.047	0.053	0.059
Panel B - Random effects							
Receives CM	-0.067	-0.165*	-0.136	-0.133	-0.103	-0.058	-0.067
	(0.086)	(0.087)	(0.102)	(0.104)	(0.113)	(0.124)	(0.137)
Panel C - Fixed effects							
Receives CM	-0.435**	-0.415**	-0.320	-0.309	-0.497**	-0.312	-0.538
	(0.171)	(0.173)	(0.205)	(0.204)	(0.218)	(0.268)	(0.333)
R-squared	0.006	0.024	0.044	0.045	0.059	0.071	0.108
N	3,372	3,341	2,661	2,641	2,172	1,884	1,527
Hausman test	0.022	0.486	0.759	0.714	0.347	0.455	0.864
Child characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes
Mother characteristics	No	No	Yes	Yes	Yes	Yes	Yes
Informal receipt	No	No	No	Yes	Yes	Yes	Yes
NRP age	No	No	No	No	Yes	Yes	Yes
Child-NRP contact	No	No	No	No	No	Yes	Yes
Mother-NRP friendliness	No	No	No	No	No	No	Yes
NRP characteristics	No	No	No	No	No	No	No

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Notes: As I am unable to use sampling weights in the random effects and fixed effects models, for comparability the unweighted coefficient is presented in pooled OLS. The weighted results however show a similar result. Negative binomial regression coefficient is similar in magnitude and remains statistically insignificant.

Table A.4 - Regression of youth peer relationship problems on CM receipt

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Youth peer relationship problems						
Panel A - Pooled OLS							
Receives CM	-0.067	-0.153**	-0.114	-0.079	-0.026	-0.002	0.032
	(0.067)	(0.068)	(0.077)	(0.080)	(0.087)	(0.093)	(0.101)
R-squared	0.000	0.031	0.041	0.042	0.049	0.050	0.054
Panel B - Random effects							
Receives CM	-0.051	-0.117*	-0.079	-0.053	0.007	0.038	0.021
	(0.064)	(0.066)	(0.075)	(0.077)	(0.086)	(0.091)	(0.099)
Panel C - Fixed effects							
Receives CM	0.069	0.106	0.050	0.045	0.157	0.226	-0.065
	(0.147)	(0.145)	(0.167)	(0.167)	(0.184)	(0.219)	(0.284)
R-squared	0.000	0.036	0.053	0.055	0.057	0.056	0.112
N	3,374	3,343	2,663	2,643	2,174	1,886	1,528
Hausman test	0.337	0.494	0.621	0.658	0.611	0.638	0.395
Child characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes
Mother characteristics	No	No	Yes	Yes	Yes	Yes	Yes
Informal receipt	No	No	No	Yes	Yes	Yes	Yes
NRP age	No	No	No	No	Yes	Yes	Yes
Child contact with NRP	No	No	No	No	No	Yes	Yes
Mother-NRP friendliness	No	No	No	No	No	No	Yes
NRP characteristics	No	No	No	No	No	No	No

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Notes: As I am unable to use sampling weights in the random effects and fixed effects models, for comparability the unweighted coefficient is presented for OLS. The weighted results however show a similar result. Negative binomial regression coefficient is similar in magnitude and remains statistically insignificant.

Table A.5 - Regression of youth pro-social behaviour on CM receipt

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Youth pro-social behaviour			
Panel A - Pooled OLS							
Receives CM	0.084	0.139*	0.128	0.120	0.089	0.146	0.163
	(0.072)	(0.072)	(0.081)	(0.085)	(0.093)	(0.102)	(0.114)
R-squared	0.000	0.085	0.103	0.102	0.107	0.105	0.109
Panel B - Random effects							
Receives CM	0.104	0.145**	0.127	0.122	0.090	0.136	0.151
	(0.069)	(0.069)	(0.079)	(0.082)	(0.090)	(0.099)	(0.110)
Panel C - Fixed effects							
Receives CM	0.368**	0.320**	0.285*	0.274	0.259	0.186	0.154
	(0.156)	(0.152)	(0.171)	(0.175)	(0.183)	(0.215)	(0.273)
R-squared	0.005	0.084	0.121	0.122	0.141	0.137	0.136
N	3,377	3,346	2,665	2,645	2,176	1,888	1,530
Hausman test	0.064	0.566	0.584	0.609	0.605	0.843	0.998
Child characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes
Mother characteristics	No	No	Yes	Yes	Yes	Yes	Yes
Informal receipt	No	No	No	Yes	Yes	Yes	Yes
NRP age	No	No	No	No	Yes	Yes	Yes
Child contact with NRP	No	No	No	No	No	Yes	Yes
Mother-NRP friendliness	No	No	No	No	No	No	Yes
NRP characteristics	No	No	No	No	No	No	No

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: As I am unable to use sampling weights in the random effects and fixed effects models, for comparability the unweighted coefficient is presented for OLS. The weighted results are significant and have a larger coefficient size.
Negative binomial regression coefficient is similar in magnitude and remains statistically insignificant.

Table A.6 - Contact and relationship quality coefficients from regressions of behavioural outcomes on CM receipt

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A - Pooled OLS	Conduct	Conduct	Emotions	Emotions	Hyper	Hyper	Peers	Peers	Pro-social	Pro-social
Receives CM	-0.215**	-0.151	-0.077	-0.010	-0.070	-0.053	-0.002	0.033	0.145	0.161
	(0.095)	(0.104)	(0.122)	(0.131)	(0.132)	(0.144)	(0.093)	(0.101)	(0.102)	(0.114)
Child-NRP contact	0.000	0.001	-0.000	-0.000	0.000	0.000	-0.000	-0.000	0.000	0.000
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Mother-NRP unfriendly		0.091		0.244		0.317		-0.051		-0.077
		(0.151)		(0.189)		(0.206)		(0.139)		(0.163)
Mother-NRP friendly		-0.138		0.263		0.069		-0.093		0.073
		(0.158)		(0.202)		(0.224)		(0.148)		(0.166)
R-squared	0.080	0.091	0.105	0.106	0.054	0.061	0.051	0.055	0.105	0.109
Panel B - Random effects										
Receives CM	-0.179**	-0.138	-0.051	0.005	-0.054	-0.060	0.038	0.021	0.135	0.148
	(0.090)	(0.100)	(0.118)	(0.127)	(0.124)	(0.137)	(0.092)	(0.100)	(0.099)	(0.110)
Child-NRP contact	0.000	0.000	-0.000	-0.000	-0.000	0.000	-0.000	-0.000	0.000	0.000
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Mother-NRP unfriendly		0.140		0.222		0.341*		-0.027		-0.086
		(0.147)		(0.186)		(0.198)		(0.139)		(0.160)
Mother-NRP friendly		-0.092		0.252		0.110		-0.022		0.049
		(0.154)		(0.199)		(0.210)		(0.149)		(0.163)
N	1,886	1,528	1,886	1,528	1,883	1,526	1,885	1,527	1,887	1,529
Child characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mother characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Informal receipt	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
NRP age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Child-NRP contact	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mother-NRP friendliness	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
NRP characteristics	No	No	No	No	No	No	No	No	No	No

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Note: The reference category for the friendliness variable is never sees the non-resident parent.

Table A.7 - Regression of youth emotional symptoms on equivalised household CM amounts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A - Pooled OLS							
Eq. HH CM	-0.008	-0.019	-0.013	-0.011	-0.018	-0.005	-0.006
	(0.021)	(0.020)	(0.023)	(0.023)	(0.026)	(0.027)	(0.027)
Eq. HH CM sq.	0.001	0.001	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Maternal income			-0.009**	-0.009**	-0.010**	-0.011**	-0.013**
			(0.004)	(0.005)	(0.005)	(0.005)	(0.006)
Partner income			0.006	0.006	0.005	0.005	0.006
			(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
R-squared	0.000	0.073	0.104	0.103	0.116	0.115	0.121
Panel B - Random effects							
Eq. HH CM	-0.007	-0.016	-0.012	-0.010	-0.016	-0.006	-0.007
	(0.020)	(0.020)	(0.022)	(0.023)	(0.025)	(0.027)	(0.028)
Eq. HH CM sq.	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Maternal income			-0.009*	-0.009*	-0.010**	-0.010**	-0.013**
			(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Partner income			0.005	0.005	0.004	0.004	0.005
			(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Panel C - Fixed effects							
Eq. HH CM	0.004	0.013	0.065	0.065	0.038	0.007	-0.065
	(0.091)	(0.093)	(0.093)	(0.093)	(0.103)	(0.119)	(0.128)
Eq. HH CM sq.	-0.003	-0.002	-0.004*	-0.004*	-0.004*	-0.004	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Maternal income			-0.007	-0.008	-0.008	-0.006	-0.036*
			(0.014)	(0.014)	(0.014)	(0.017)	(0.021)
Partner income			0.009	0.009	0.009	0.011	0.023
			(0.012)	(0.012)	(0.012)	(0.016)	(0.022)
R-squared	0.002	0.062	0.095	0.095	0.111	0.117	0.203
N	2,794	2,777	2,403	2,385	1,951	1,690	1,354
Hausman test	0.467	0.407	0.282	0.326	0.289	0.252	0.457
Child characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes
Mother characteristics	No	No	Yes	Yes	Yes	Yes	Yes
Informal receipt	No	No	No	Yes	Yes	Yes	Yes
NRP age	No	No	No	No	Yes	Yes	Yes
Child- NRP contact	No	No	No	No	No	Yes	Yes
Mother-NRP friendliness	No	No	No	No	No	No	Yes
NRP characteristics	No	No	No	No	No	No	No

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1
Notes: All income is indexed to 2018 prices and is reported in weekly terms and divided by 10, coefficients should be interpreted as an increase of £10 a week

Table A.8 - Regression of youth hyperactivity/inattention on equivalised household CM amounts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A - Pooled OLS							
Eq. HH CM	-0.009	-0.019	-0.011	-0.010	0.007	0.000	0.007
	(0.020)	(0.021)	(0.024)	(0.025)	(0.027)	(0.029)	(0.032)
Eq. HH CM sq.	0.001**	0.001**	0.001*	0.001*	0.001	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Maternal income			0.001	0.002	0.003	0.003	0.001
			(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
Partner income			0.002	0.002	0.004	0.002	-0.000
			(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
R-squared	0.001	0.041	0.055	0.055	0.051	0.059	0.068
Panel B - Random effects							
Eq. HH CM	0.000	-0.009	-0.008	-0.009	0.011	0.013	0.016
	(0.019)	(0.020)	(0.023)	(0.023)	(0.026)	(0.028)	(0.030)
Eq. HH CM sq.	0.001	0.001**	0.001*	0.001**	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Maternal income			0.001	0.002	0.002	0.002	0.000
			(0.005)	(0.005)	(0.005)	(0.006)	(0.006)
Partner income			-0.000	-0.000	0.001	-0.003	-0.003
			(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
Panel C - Fixed effects							
Eq. HH CM	-0.001	0.010	0.016	0.015	-0.020	0.022	-0.053
	(0.066)	(0.066)	(0.066)	(0.067)	(0.076)	(0.084)	(0.111)
Eq. HH CM sq.	0.001	0.001	0.002	0.001	0.003	0.002	0.004
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)
Maternal income			0.008	0.006	0.010	0.016	-0.001
			(0.013)	(0.013)	(0.013)	(0.014)	(0.022)
Partner income			-0.017	-0.017	-0.019	-0.037***	-0.038***
			(0.012)	(0.012)	(0.013)	(0.012)	(0.013)
R-squared	0.000	0.039	0.069	0.071	0.079	0.117	0.182
N	2,792	2,775	2,401	2,383	1,948	1,687	1,352
Hausman test	1.000	0.191	0.326	0.325	0.298	0.094	0.526
Child characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes
Mother characteristics	No	No	Yes	Yes	Yes	Yes	Yes
Informal receipt	No	No	No	Yes	Yes	Yes	Yes
NRP age	No	No	No	No	Yes	Yes	Yes
Child- NRP contact	No	No	No	No	No	Yes	Yes
Mother-NRP friendliness	No	No	No	No	No	No	Yes
NRP characteristics	No	No	No	No	No	No	No

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1
Notes: All income is indexed to 2018 prices and is reported in weekly terms and divided by 10, coefficients should be interpreted as an increase of £10 a week

Table A.9 - Regression of youth peer relationship problems on equivalised household CM amounts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A - Pooled OLS							
Eq. HH CM	-0.035**	-0.051***	-0.030*	-0.024	-0.022	-0.019	-0.016
	(0.016)	(0.016)	(0.018)	(0.019)	(0.021)	(0.022)	(0.023)
Eq. HH CM sq.	0.001***	0.001***	0.001***	0.001***	0.001***	0.001**	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Maternal income			-0.004	-0.004	-0.005	-0.005	-0.007
			(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Partner income			0.001	0.001	0.002	0.001	0.003
			(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
R-squared	0.002	0.035	0.050	0.049	0.058	0.062	0.067
Panel B - Random effects							
Eq. HH CM	-0.033**	-0.048***	-0.029	-0.025	-0.021	-0.020	-0.018
	(0.016)	(0.016)	(0.018)	(0.018)	(0.021)	(0.022)	(0.022)
Eq. HH CM sq.	0.001***	0.001***	0.001***	0.001***	0.001***	0.001**	0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Maternal income			-0.004	-0.003	-0.006	-0.005	-0.008*
			(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Partner income			-0.002	-0.002	-0.001	-0.000	0.001
			(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Panel C - Fixed effects							
Eq. HH CM	-0.093	-0.090	-0.051	-0.050	-0.040	-0.036	-0.013
	(0.064)	(0.065)	(0.068)	(0.068)	(0.075)	(0.083)	(0.075)
Eq. HH CM sq.	0.002	0.002	0.001	0.001	0.000	-0.001	-0.001
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
Maternal income			-0.011	-0.010	-0.018	-0.027**	-0.041***
			(0.011)	(0.011)	(0.012)	(0.011)	(0.015)
Partner income			-0.010	-0.010	-0.015*	-0.007	-0.001
			(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
R-squared	0.005	0.050	0.065	0.067	0.072	0.085	0.187
N	2,794	2,777	2,403	2,385	1,950	1,689	1,353
Hausman test	0.394	0.637	0.806	0.854	0.784	0.679	0.209
Child characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes
Mother characteristics	No	No	Yes	Yes	Yes	Yes	Yes
Informal receipt	No	No	No	Yes	Yes	Yes	Yes
NRP age	No	No	No	No	Yes	Yes	Yes
Child- NRP contact	No	No	No	No	No	Yes	Yes
Mother-NRP friendliness	No	No	No	No	No	No	Yes
NRP characteristics	No	No	No	No	No	No	No

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1
Notes: All income is indexed to 2018 prices and is reported in weekly terms and divided by 10, coefficients should be interpreted as an increase of £10 a week

Table A.10 - Regression of youth pro-social behaviour on equivalised household CM amounts

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A - Pooled OLS							
Eq. HH CM	0.022	0.020	0.019	0.020	0.008	0.011	0.007
	(0.016)	(0.016)	(0.018)	(0.019)	(0.022)	(0.023)	(0.026)
Eq. HH CM sq.	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Maternal income			0.000	0.000	-0.000	-0.000	0.002
			(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Partner income			0.002	0.002	-0.001	0.003	0.001
			(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
R-squared	0.001	0.092	0.105	0.104	0.108	0.106	0.106
Panel B - Random effects							
Eq. HH CM	0.024	0.021	0.020	0.021	0.007	0.009	0.006
	(0.016)	(0.016)	(0.018)	(0.019)	(0.022)	(0.023)	(0.026)
Eq. HH CM sq.	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Maternal income			0.001	0.001	0.001	0.001	0.003
			(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Partner income			0.003	0.003	0.000	0.004	0.002
			(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
Panel C - Fixed effects							
Eq. HH CM	0.179***	0.149**	0.142**	0.142**	0.153**	0.138*	0.094
	(0.064)	(0.066)	(0.066)	(0.066)	(0.071)	(0.078)	(0.108)
Eq. HH CM sq.	-0.005	-0.005	-0.004	-0.004	-0.004	-0.004	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Maternal income			0.023**	0.023**	0.025**	0.027**	0.057***
			(0.010)	(0.010)	(0.011)	(0.012)	(0.017)
Partner income			0.008	0.008	0.006	0.018	0.016
			(0.011)	(0.011)	(0.012)	(0.013)	(0.018)
R-squared	0.010	0.093	0.130	0.129	0.152	0.158	0.217
N	2,797	2,780	2,405	2,387	1,952	1,691	1,355
Hausman test	0.057	0.607	0.692	0.736	0.685	0.852	0.791
Child characteristics	No	Yes	Yes	Yes	Yes	Yes	Yes
Mother characteristics	No	No	Yes	Yes	Yes	Yes	Yes
Informal receipt	No	No	No	Yes	Yes	Yes	Yes
NRP age	No	No	No	No	Yes	Yes	Yes
Child- NRP contact	No	No	No	No	No	Yes	Yes
Mother-NRP friendliness	No	No	No	No	No	No	Yes
NRP characteristics	No	No	No	No	No	No	No

Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1
Notes: All income is indexed to 2018 prices and is reported in weekly terms and divided by 10, coefficients should be interpreted as an increase of £10 a week

Table A.11 - Sub-sample regression of CM on child outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A - Pooled OLS	Conduct	Conduct	Emotions	Emotions	Hyper	Hyper	Peers	Peers	Pro-social	Pro-social
Eq. hh CM amount	-0.025	-0.030	-0.028	-0.030	0.012	0.016	-0.033	-0.032	0.010	0.002
	(0.025)	(0.026)	(0.029)	(0.031)	(0.034)	(0.036)	(0.027)	(0.026)	(0.028)	(0.029)
NRP partner		0.087		0.505***		0.362*		0.036		-0.042
		(0.145)		(0.188)		(0.194)		(0.146)		(0.149)
NRP children		-0.010		-0.222		-0.141		-0.353**		-0.352**
		(0.138)		(0.172)		(0.203)		(0.145)		(0.147)
NRP employed		0.117		-0.022		-0.148		-0.204		0.047
		(0.174)		(0.220)		(0.264)		(0.192)		(0.193)
R-squared	0.151	0.152	0.156	0.168	0.073	0.078	0.085	0.095	0.151	0.160
Panel B - Random effects										
Receives CM	-0.019	-0.023	-0.034	-0.038	0.014	0.018	-0.039	-0.035	0.011	0.004
	(0.024)	(0.025)	(0.030)	(0.032)	(0.034)	(0.036)	(0.025)	(0.025)	(0.028)	(0.029)
NRP partner		0.134		0.518***		0.415**		0.056		-0.028
		(0.145)		(0.185)		(0.194)		(0.141)		(0.148)
NRP children		0.038		-0.185		-0.096		-0.274*		-0.325**
		(0.136)		(0.170)		(0.205)		(0.144)		(0.144)
NRP employed		0.137		0.026		-0.134		-0.197		0.041
		(0.178)		(0.219)		(0.259)		(0.185)		(0.187)
Panel C - Fixed effects										
Receives CM	0.080	0.067	-0.014	-0.074	-0.147	-0.186	-0.024	-0.026	0.339***	0.374***
	(0.085)	(0.087)	(0.189)	(0.184)	(0.184)	(0.179)	(0.088)	(0.089)	(0.105)	(0.125)
NRP partner		0.048		0.431		0.568		-0.209		-0.512
		(0.532)		(0.590)		(0.724)		(0.389)		(0.536)
NRP children		1.365*		-0.677		1.267		0.562		-1.765***
		(0.714)		(0.956)		(0.828)		(0.769)		(0.626)
NRP employed		-0.329		1.483**		-0.518		0.221		0.759
		(0.643)		(0.584)		(0.655)		(0.502)		(0.675)
R-squared	0.315	0.352	0.386	0.073	0.379	0.399	0.384	0.393	0.324	0.375
N	731	731	731	731	730	730	730	730	731	731

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1 Note: The coefficient on contact is not reported in the Table however, it is insignificant and close to zero for all specifications. The controls included in even columns are child characteristics, mother characteristics, informal receipt, NRP age, and NRP-child contact. In even numbered columns controls for NRP's new partner, new children and employment status are included and coefficients reported.

3 Chapter 3

Welfare conditionality and lone parents: quasi-experimental evidence from the UK Lone Parent Obligation reform

I am grateful for comments on this chapter from participants at the Lancaster University PhD Seminar Series, and the Work, Pensions and Labour Economics (WPEG) conference 2021: Alex Farnell, Stephen Jenkins, Andrew McKendrick, James Oswald, and Tom Stewart. Thanks go to the UK Data Service for making the data available.

3.1 Introduction

Lone mothers in the UK are a disadvantaged group with historically low employment rates and high poverty rates. During the last few decades, they have been the target of several labour market activation policies.[11] One such policy was the introduction of Lone Parent Obligations (LPO). The premise of this reform was to progressively reduce the age of the youngest child at which a lone parent could claim Income Support (IS) solely on the basis of being a lone parent. Removing eligibility for an unconditional benefit effectively imposed work search requirements on a lone parent as they would instead receive Job Seekers Allowance (JSA), a benefit which required claimants to look for and apply for jobs in order to remain eligible for payment.

There are several reasons why increasing lone mothers' employment is a desirable policy goal. Aside from the reduction in social security expenditure, maternal employment is generally seen as a way out of poverty for single-parent families, the vast majority of which are female-headed. It is thought to have beneficial effects on maternal mental well-being (Harkness 2016), and it might have a positive effect on child development through raising household income, enabling higher quality childcare, schooling or extra-curricular activities that contribute to child development.[12]

This paper contributes to the literature by providing causal evidence on the final phase of the LPO reform, introduced in 2010 and implemented fully in May 2012. It resulted in the age of the youngest child, at which a mother would no longer be eligible for income support, being reduced from 7 years old to 5 years old. Whilst some existing evidence suggests that the final phase of the reform has not increased the employment rates of lone parents, this chapter finds evidence to the contrary and examines other labour supply outcomes such as work search activity and entry and exit rates of employment. I also investigate whether there are heterogeneous responses in terms of presence of an older sibling, maternal education level and housing tenure.

[11] Some of which include compulsory work focused interviews, in-work credit, NDLP, an increase in the earnings disregard, WFTC replacing Family Credit and the National Childcare Strategy.
[12] A big literature has also shown it could have a negative effect of maternal work (see for example Mosca et al. (2021))

A difference-in-difference approach is taken and uses married or cohabiting mothers with a youngest child aged 5 or 6 as the comparison group. For comparability purposes, I use two datasets: the UK Labour Force Survey (LFS) cross section surveys and the Understanding Society panel. The estimates are found to be the same across datasets. The datasets each have their relative advantages which are discussed in Section 3.4.

The findings show that the reform increased lone mothers' employment by around 8 percentage points – an effect size remarkably close to that found in existing evidence on UK labour market reforms directed at lone mothers. In an event study framework, these effects on employment take time to come through after the reform, they start to appear after five quarters (in the LFS) and two-three years (in Understanding Society) after the reform. Work search activity increased in both datasets by 10 percentage points but this is concentrated in the periods immediately after the reform. The effects fade out after six quarters (in LFS) and one year (in Understanding Society). It is not possible to rule out changes in the composition of the sample being responsible for this. However, combining the two measures into one outcome, labour market participation, which is defined as searching for work or in work, shows a similar result, but one which is less likely to be contaminated by composition bias.

There were no effects on persistence of employment nor on exit rates from employment, although one caveat is the small sample sizes used to generate these estimates. There was however, a statistically significant increase in the proportion on health-related benefits by 6 percentage points in both datasets. In Section 3.7.3, I use pregnancy rates as an outcome and find no effect of the LPO, suggesting that the reform does not incentivise lone mothers to have more children in order to avoid losing eligibility to IS.

One concern with difference-in-difference models is that the treated and control groups may be different, and this difference might vary across time. There are several ways in which I demonstrate that this is not the case. First, the inclusion of a differential linear time trend in the preferred specification should go some way towards alleviating this concern. I also conduct several robustness checks. I use an alternative control group, which arguably is more similar to the treated group, and find the results do not change. I test whether trends in the labour market affect treated and untreated groups differently and find this is not the case. Thirdly, I implement a matched difference-in-difference, in which the results are unchanged from the main specifications.

Aside from fitting into the broader literature on the introduction of work search conditionalities for the most disadvantaged groups in society, this chapter is also important in informing future policy. In particular, the UK government has recently imposed job search requirements for lone mothers with a youngest child aged 3-4 for Universal Credit (UC) recipients, for which no evaluation currently exists but these results suggest that it may have the desired effect. Moreover, it supports the idea that the COVID hike in Universal Credit would have had a negative supply effect.

The paper is structured as follows. Section 3.2 describes the LPO reforms and policy context, Section 3.3 reviews the literature, Section 3.4 illustrates the data used and Section 3.5 explains the methodology. Section 3.6 presents the empirical results, Section 3.7 tests the robustness of these results, and Section 3.8 concludes.

3.2 UK policy and reform

Income Support (IS), which has been recently (in 2013) incorporated into the new Universal Credit scheme, is a social security benefit to support individuals on low or no household income. It could be claimed by individuals aged 16 and over, unemployed or working fewer than 16 hours per week (and/or with a partner working fewer than 24 hours), and the individual must fall into a disadvantaged group, and lone parents were recognised as one such group.[13] While there was no requirement for the claimant to search for jobs, there were several other requirements to fulfil. Mandatory bi-annual Work-Focused-Interviews (WFIs) were an assessment of the claimant's skills and prospects for employment, to identify training or education needs to help further employment prospects.[14]

An individual who did not meet the criteria for IS could instead claim Job Seekers Allowance (JSA), which imposed stricter work search requirements. JSA has two separate components – contributory and means-tested. The contributory component (JSA(C)) is limited in duration, only being paid for six months. It also is based on the claimant's National Insurance (NI) contribution history (they must have paid enough class one NI contributions during the 2

[13] Other groups included: pregnant, lone foster parent, carer, on maternity, paternity or parental leave, in full-time education (not university) and aged between 16-20 and a parent, in full-time education (not university), aged between 16 and 20, and not living with a parent or someone acting as a parent, a refugee learning English, in custody or due to attend court or a tribunal.
[14] Failure to attend a WFI could result in a sizeable benefit sanction - 20 percent of the IS personal allowance. In the year 2008-2009 7.8 percent of all IS claims were sanctioned.

years preceding the claim). The means-tested element (JSA(IB)) can be claimed indefinitely and is calculated based on the claimant's savings and income.

Upon making an application for JSA, the claimant is invited for an interview at the Jobcentre and will sign what was known as a "Claimant Commitment". It sets out the type of work the claimant is looking for and the steps they will take to do this. This is used as a benchmark to check if the claimant is fulfilling their job seeking obligations, and the basis on which the Jobcentre can impose sanctions.

Though JSA was different to IS in job search requirements and threat of sanctions, both benefits were administered by the same agency and used the same means-testing calculation so, crucially, there was no difference in the size of payment. Both IS and JSA rates in 2012 for a lone parent were £71 per week. JSA sanctions were set at 20 percent of this personal allowance (£14.20) (Rutherford (2013). For both forms of benefit, the claimant faces 100% marginal deductions if they earn more than £20 per week and benefits were removed from the claimant if they started working 16 or more hours per week.

The first three phases of the Lone Parent Obligations (LPO) reform were set out in the 2007 report "In Work, Better Off: Next Steps to Full Employment" and gradually reduced the age of the youngest child at which a lone mother could receive IS. There were initially three phases. From November 2008, lone parents with a youngest child aged 12 or over would no longer be entitled to IS, falling to age 10 in October 2009, and age 7 in October 2010. The June 2010 Budget extended the reform to five-year olds from May 2012, which is the focus of this chapter.

A lone parent will either lose eligibility for IS on the birthday of the youngest child, or as a direct consequence of the reform coming into effect. This paper studies the latter. The process starts 12 months before the claimant is due to lose IS.[15] Quarterly WFIs were introduced in the year before the lone parent loses eligibility (Department for Work and Pensions, 2012). On the date that the lone parent loses IS, they are placed onto JSA. Lone mothers who fell into a disadvantaged category were exempt and remained on IS. Being in

[15] The lone parent is invited to an *Options and Choices Event* to give them more information along as having to attend quarterly WFIs. Eight weeks before their entitlement is due to end, the claimant receives a letter from the Job Centre which tells them when their last payment is due and invites them to an interview six weeks before their entitlement is due to end. The interview is designed to offer support and advice. Four weeks before IS is due to end, they receive another letter from the Benefit Delivery Centre. Five days before IS is due to end, the lone parent receives a final letter.

education and training was one such example. Alternatively, if the lone mother had poor health or a disability affecting their capability to work. In this instance, they were able to apply for health-related benefits, such as Employment Support Allowance (ESA), upon successfully meeting the criteria. It is worth noting, for an employed lone mother, not in receipt of IS, this resulted in a less generous safety net, incentivising her to remain in employment.

The reform did not affect any other benefits such as Child Tax Credit, Housing Benefit and Council Tax Benefit. Another important source of income for lone parents is child support, though since 2008 this has been fully disregarded from the benefit calculation.

3.3 Relevant literature

There has been a substantial amount of research, globally, on labour market activation policies. Because there is considerable variation across studies and countries, there is no real consensus on the effectiveness of active labour market policies. Heckman et al. (1999) and Martin & Grubb (2005) provide surveys. Most recently Card et al. (2010) conducted a meta-analysis of micro-econometric evaluations. Their findings suggest while subsidised public sector employment programmes tend not to be effective, imposing job search requirements may have some positive impacts, particularly in the short-term. The results also suggest that the measure of employment matters – some studies use duration of unemployment or time to exiting unemployment and tend to find larger positive effects than those using an indicator for employment.

For the UK, Dolton & O'Neill (1996) analysed the Restart programme.[16] The support and encouragement given to claimants is one aspect, but more importantly the claimant is faced with the threat of a sanction should they not attend the interview or fulfil job search requirements. The authors evaluated a controlled experiment in which the long-term unemployed were randomly assigned to treatment in the programme. Their evidence suggests the programme reduced unemployment duration, but further inspection indicated the Restart programme effect was not the same for different exit types. It increased the hazard rate to signing off from IS receipt, but it had a very small effect on training offers and receiving the

[16] The Restart programme was designed to reduce the time people spend unemployed and receiving unemployment benefits by offering six-monthly meetings between the individual and a counsellor.

Restart interview after 12 months of unemployment (control group), as opposed to after 6 months of unemployment (treatment group), may be thought to be detrimental to job prospects.

The LPO reform draws parallels with the introduction of Job Seekers Allowance in 1996. The UK government evaluation took the form of a 'before and after' approach using two cohorts of benefit claimants. Several reports (Trickey et al. (1998), McKay et al. (1999), Rayner et al. (2000), Smith et al. (2000)) found that movements off benefits increased as a result of "weeding out" low search activity claimants, fraudulent claims and increasing job search. Problematically, if JSA removed low search activity claimants from the claimant count, then average search intensity should increase post-JSA but purely due to changes in the composition of the sample. However, more robust empirical studies exist using quasi-experimental research designs. Petrongolo (2009) used an administrative dataset and found being on JSA had a significant and positive effect on the unemployment exit rate. The exit destinations paint a less rosy picture. Estimates of the exit rates to incapacity benefits were around 2-3 percent, and post-JSA earnings were considerably lower for the treated group. Due to data constraints, the author did not have the means to look at employment spells but had information on weeks worked and earnings for the years post-JSA. The results indicate that JSA had a negative effect on earnings, reflected in the negative effect on work and earnings conditional on employment. The finding that JSA reduced flows out of the claimant count is replicated in Manning (2009) who used the UK Labour Force Survey (LFS) to the same purpose. He found evidence corroborating the suggestion there was variation in the impact of JSA on claimant outflow depending on the level of initial work search activity and that JSA disproportionately removed individuals with low search activity from the claimant count.

The effect of imposing work search conditionalities on lone mothers is less clear. They face greater barriers to employment, balancing childcare with work, and may spend longer searching in order to find a job offering the necessary flexibility. Indeed, if the costs of childcare are greater than the wage received from employment then there is no incentive to work.

Initial evaluations of the LPO have been an important topic of discussion for the Department for Work and Pensions. An analysis of administrative data indicated that 83 percent of lone parents affected by phase 1 of the reform were no longer receiving IS. Of these, 16 percent moved into work of 16 hours or more, 56 percent moved onto JSA, 18 percent moved onto Employment and Support Allowance (ESA), 2 percent had re-partnered, 2 percent had moved onto another benefit and, for 6 percent the destination was unknown (Casebourne et al. (2010)).

Focusing on before the introduction of the reform, Soobedar (2009a) used a fuzzy regression discontinuity identification strategy in the Labour Force Survey (LFS) to look at employment outcomes when the lone mother lost eligibility upon her child turning 16. She found the existence of the age cut-off had significant labour supply disincentives for lone mothers with no qualifications, evidence in favour of the eventual implementation of the reforms. In a difference-in-difference approach Nielsen & Oakley (2011) found the first phase of the reform (where the cut-off age was reduced from 16 to 12) increased the probability of lone mothers being in work by 4.3 percentage points, falling to 2.7 percent when controlling for differential time trends, but still statistically significant.

Department for Work and Pensions research (Coleman & Riley, 2012) provides a detailed descriptive analysis of the destinations of lone mothers on IS. Using a quantitative survey analysing mothers in the third-rollout phase of the LPO, and a sample of around 900 lone mothers, the authors found over half of respondents move onto JSA and just over 20 percent get a job immediately after leaving income support. Longer-term outcomes indicate that a year later 33 percent of respondents had found a job and this work was sustained. Unfortunately, a move into work was not necessarily associated with lone mothers moving out of poverty, due to low-paid jobs and constraints on the number of hours they could work. Additional analysis in the report suggests longer spells on JSA were associated with lone mothers with lower qualifications and English as a second language. Childcare played a large part in lone mothers' decision to work and the demand for part-time or flexible working arrangements was high.

Two further recent papers both take a difference-in-difference approach, to examine the effects of various phases of the LPO. They took the approach of using lone mothers with younger children as the control group. Avram et al. (2018) used DWP administrative data and a difference-in-difference methodology to look at changes in lone mother employment and benefit receipt.[17] Although their findings suggest the reform encouraged lone mothers into employment (10pp), there was an even larger influx onto other benefits or non-claimant unemployment (18pp), a result which was magnified for lone mothers with low labour market

[17] Their data did not cover the most recent 2012 reform.

attachment (proxied for time spent on welfare benefits before the reform). Their analysis benefited from a large sample size and precise identification of when the individual was affected by the reform (by both timing of the reform and date of birth of the child). The data enabled them to follow recipients over time which enabled measurement of the persistence of the effect and the extent to which mothers anticipate the loss of income support. This is more important in the context where all lone mothers in the dataset are in receipt of IS.

Garaud (2014) used the Labour Force Survey (LFS) to the same purpose with a few distinctions. The dataset is, in some senses, more restrictive - the sample was lone mothers and their IS status was not exploited (making this an estimation of an intention-to-treat).[18] The dataset is cross-sectional and could not give any indication of the persistence of effects. Although he found a positive and statistically significant 3 percentage point increase in employment rates for the first rollout-phase, there is no effect for the second, third and final phase of the reform. The last is of interest for this chapter. The treatment group used was lone parents with a youngest child aged 3-4. This could be problematic given that a government policy extension in 2010 provided 15 hours of free childcare to mothers of children aged 3-4. Evidence suggests that childcare may be an important barrier to lone mother employment, and if the control group had been affected (whereas the treatment group have not), it could generate downward bias in the estimates.

Other UK policies targeted at improving maternal employment rates have had some success. Francesconi & Van der Klaauw (2007) found the Working Families Tax Credit (WFTC) increases lone mother labour participation by 5 percentage points on average.[19] The employment increase was associated with more usage of childcare (complementing the childcare element of the WFTC). This is consistent with other evidence on the WTFC (see for example Blundell et al. (2000), Brewer et al. (2005), and Gregg & Harkness (2003)). Interestingly, Francesconi & van der Klaauw (2007) found their results were more pronounced among single mothers with younger children, hypothesising this could be due to the childcare tax credit component of the scheme. Childcare appears to be a large barrier to entering employment for lone mothers. The authors also found considerable heterogeneity by ages of

[18] This is important - there is no mention of the fact for the first three phases the lone parent will receive treatment at different times depending on their income support status (those on income support will lose eligibility a year later than the reform date).
[19] Replacing Family Credit (FC) in 1999, it was substantially more generous in the amounts single mothers received.

the child where the effect sizes on employment for mothers with younger children were substantially larger than that of mothers with older children.

The different phases of the LPO may not have the same effect on employment outcomes, because the lone mother has a different set of incentives depending on the age of her youngest child. If her youngest child is older, she may have fewer issues with childcare, be less concerned about the impact of working on the child's development, or, on the other hand, she may have greater demand for the consumption (more expensive school trips, birthday gifts etc.) that employment could satisfy. Whilst there has been considerable evaluation of the first three phases of the LPO, evidence has not yet been convincingly extended to the final reform. There are two interesting dimensions to explore. Firstly, whether the policy has greater effectiveness for mothers with younger children and, secondly, whether there are heterogeneous effects by older sibling (proxying childcare), education level, and housing tenure status.

3.4 Data

Two datasets are used for the analysis to test for consistency in findings - Understanding Society and the Labour Force Survey (LFS). The two datasets have their relative advantages - the LFS is much larger and, due to the low incidence of lone parent families in household surveys, can offer larger sample sizes. Understanding Society, whilst somewhat smaller in size, has a much richer range of outcomes and it is possible to look at past employment status, duration of unemployment and employment spells. Also, when looking at heterogeneity, Understanding Society contains information on childcare and older siblings. Frequency of data collection is another distinction, the LFS collects responses quarterly, Understanding Society annually.

Understanding Society is a longitudinal dataset that began in 2009 and interviewed 40,000 households in the first wave. I use data from Waves 1 to 9 (2009 to 2019) and the sample consists of working age mothers with a youngest child aged 5 or 6. The sample includes 7699 mothers of 5-6 year olds, of which 1802 are lone mothers and 5897 are married or partnered. These are treated as individual cross-sections, although employment histories may be found in the data. There are several reasons for not exploiting the panel nature of the data. The main reason is that it would necessitate an entirely different identification strategy e.g. survival analysis/event history. Following mothers over time means that their child ages and would make it impossible to take a difference-in-difference approach. This is because you would be comparing mothers with younger children with mothers with older children in the pre- and post- period. And following them over time would mean that after May 2012 they are all treated by the reform. There may be interest in examining longer term outcomes in a different framework, but this is not within the score of this paper.

The Labour Force Survey (LFS) is a quarterly cross-sectional representative survey. It is a large dataset interviewing 40,000 households in each quarter. Using data from Q4 2008 to Q4 2015, the sample consists of working age lone mothers with a youngest child aged 5 or 6 in the household, identified by a variable capturing household composition. The total sample size is 42,609 mothers with a youngest child aged 5-6, of which 11,503 are lone mothers and 31,106 are married mothers.

The coding of key variables is the same for both datasets. A dummy variable indicating the reform takes the value 0 if the lone mother was interviewed prior to May 2012 and 1 if they were interviewed after. Outcome variables are whether the respondent receives IS or JSA, is employed (including self-employment), is searching for work, and receiving a health-related benefit. These variables are all binary coded. In Understanding Society it is possible to use respondents employment histories to examine effects on entry and exit rates into employment post-reform.

Table 3.1 presents summary statistics using Understanding Society and compares lone mothers (treated group) with married mothers (control group), both of whom have a youngest child aged 5 or 6. The former are more likely to be young and less educated. Lone mothers are also considerably less likely to own their own home, 45 percent rely on social housing, compared with just 15 percent of married mothers. These variables will be included as controls in the analysis. Childcare usage is similar between the two groups, and lone parents have a slight increase in childcare usage post-reform. Despite the importance of chilcare in predicting lone mothers employement, it is likely to be an endogeous variable, in that the reform may induce childcare usage to increase. For this reason childcare is not included as a contol in the regressions.

There might be a concern, using cross-sectional data that the composition of the sample changes after the policy – for example, lone parents might have new children in order to retain the same level of income in anticipation of having to search for jobs. In Table 3.2, the LFS presents a similar story. The biggest concern is that education levels between the two groups are quite different and the education level of the post-reform control group is quite a lot higher for married mothers. Due to this discrepancy in education level between the comparison and treated groups in both datasets, results are also presented using a sub-sample of married mothers with low education. Finding no differences between columns (1) and (3) and (2) and (4) indicate this is not the case.

Concentrating on differences between datasets, the LFS and Understanding Society are fairly similar in terms of sample composition (see Table 3.3). Lone mothers in the LFS are slightly younger and due to being coded slightly differently, the LFS does not allow us to see the proportion of respondents in social housing. Lone mothers are also slightly less educated in the LFS than in Understanding Society.

Table 3.1 - Understanding Society summary statistics

	Pre-reform		Post-reform	
	(1) Treatment - lone mother	(2) Control - married mother	(1) Treatment - lone mother	(2) Control - married mother
Age	34.53	37.56	35.41	38.05
	(6.978)	(5.452)	(6.771)	(5.481)
Non-white	0.30	0.22	0.28	0.26
	(0.458)	(0.415)	(0.448)	(0.440)
Education				
Degree	0.15	0.29	0.20	0.36
	(0.355)	(0.453)	(0.403)	(0.481)
Other higher	0.15	0.16	0.14	0.14
	(0.359)	(0.363)	(0.351)	(0.348)
A level	0.16	0.18	0.23	0.19
	(0.365)	(0.388)	(0.419)	(0.394)
GCSE	0.36	0.24	0.29	0.20
	(0.479)	(0.426)	(0.455)	(0.403)
Other qual	0.07	0.06	0.06	0.05
	(0.249)	(0.241)	(0.235)	(0.214)
No qual	0.12	0.07	0.08	0.05
	(0.326)	(0.258)	(0.264)	(0.222)
No. of kids <16 responsible for	1.78	2.06	1.82	2.08
	(0.855)	(0.878)	(0.845)	(0.863)
Long-standing illness/impairment	0.30	0.22	0.29	0.21
	(0.459)	(0.414)	(0.455)	(0.410)
Uses childcare	0.39	0.41	0.44	0.45
	(0.488)	(0.493)	(0.497)	(0.497)
Housing tenure				
Owned outright	0.05	0.08	0.05	0.08
	(0.213)	(0.266)	(0.208)	(0.270)
Owned with mortgage	0.23	0.65	0.21	0.63
	(0.421)	(0.476)	(0.408)	(0.482)
Social housing	0.45	0.15	0.45	0.15
	(0.497)	(0.358)	(0.498)	(0.362)
Renting	0.28	0.12	0.29	0.13
	(0.447)	(0.326)	(0.455)	(0.340)
Receives income support	0.44	0.04	0.15	0.01
	(0.497)	(0.190)	(0.356)	(0.115)
Receives job seekers allowance	0.01	0.01	0.08	0.01
	(0.0888)	(0.0924)	(0.275)	(0.0886)
Employed	0.47	0.68	0.62	0.70
	(0.499)	(0.468)	(0.486)	(0.456)
No. of hours normally worked per week	25.57	25.02	25.29	26.04
	(10.23)	(10.40)	(9.703)	(10.53)
Looked for work in last 4 weeks	0.25	0.13	0.29	0.11
	(0.434)	(0.331)	(0.456)	(0.309)
On health-related benefit	0.10	0.07	0.17	0.06
	(0.295)	(0.253)	(0.373)	(0.242)
N	757	2137	1045	3760

mean coefficients; std dev in parentheses

Table 3.2 - Labour Force Survey summary statistics

	Pre reform		Post reform	
	(1) Treatment - lone mother	(2) Control - married mother	(1) Treatment - lone mother	(2) Control - married mother
Age	33.73	37.10	33.90	37.53
	(7.693)	(6.574)	(7.704)	(6.726)
Non-white	0.17	0.14	0.16	0.17
	(0.373)	(0.351)	(0.366)	(0.373)
Education				
Degree or equivalent	0.11	0.26	0.15	0.37
	(0.308)	(0.439)	(0.353)	(0.483)
Higher educ	0.08	0.11	0.09	0.10
	(0.277)	(0.317)	(0.280)	(0.298)
GCE A Level or equiv.	0.20	0.19	0.22	0.18
	(0.401)	(0.391)	(0.415)	(0.383)
GCSE grades A-C or equiv.	0.34	0.27	0.32	0.21
	(0.474)	(0.441)	(0.467)	(0.409)
Other qualifications	0.12	0.10	0.12	0.08
	(0.330)	(0.304)	(0.326)	(0.278)
No qualification	0.14	0.07	0.11	0.05
	(0.349)	(0.252)	(0.307)	(0.226)
No. of kids <16 responsible for	1.89	2.08	1.87	2.06
	(0.983)	(0.888)	(0.972)	(0.878)
Health problem	0.30	0.22	0.32	0.22
	(0.458)	(0.416)	(0.468)	(0.414)
Housing tenure				
Owned outright	0.04	0.08	0.03	0.07
	(0.191)	(0.269)	(0.160)	(0.262)
Owned with mortgage	0.22	0.67	0.19	0.61
	(0.413)	(0.472)	(0.392)	(0.487)
Part rent/rent free	0.01	0.01	0.01	0.01
	(0.114)	(0.0992)	(0.0941)	(0.0986)
Renting	0.73	0.24	0.78	0.30
	(0.444)	(0.430)	(0.417)	(0.459)
Receives income support	0.41	0.01	0.12	0.01
	(0.491)	(0.107)	(0.329)	(0.0817)
Receives JSA	0.01	0.01	0.11	0.01
	(0.102)	(0.0890)	(0.317)	(0.0759)
Employed	0.51	0.71	0.61	0.72
	(0.500)	(0.454)	(0.487)	(0.450)
Total usual hours in main job	26.36	27.08	25.49	28.22
	(11.77)	(13.01)	(11.55)	(12.79)
Looked for work in last 4 weeks	0.30	0.19	0.40	0.19
	(0.458)	(0.394)	(0.489)	(0.393)
On health-related benefit	0.08	0.05	0.11	0.04
	(0.269)	(0.219)	(0.319)	(0.188)
N	6127	15922	5376	15184

mean coefficients; std dev in parentheses

Table 3.3 - Sample differences between Understanding Society and LFS

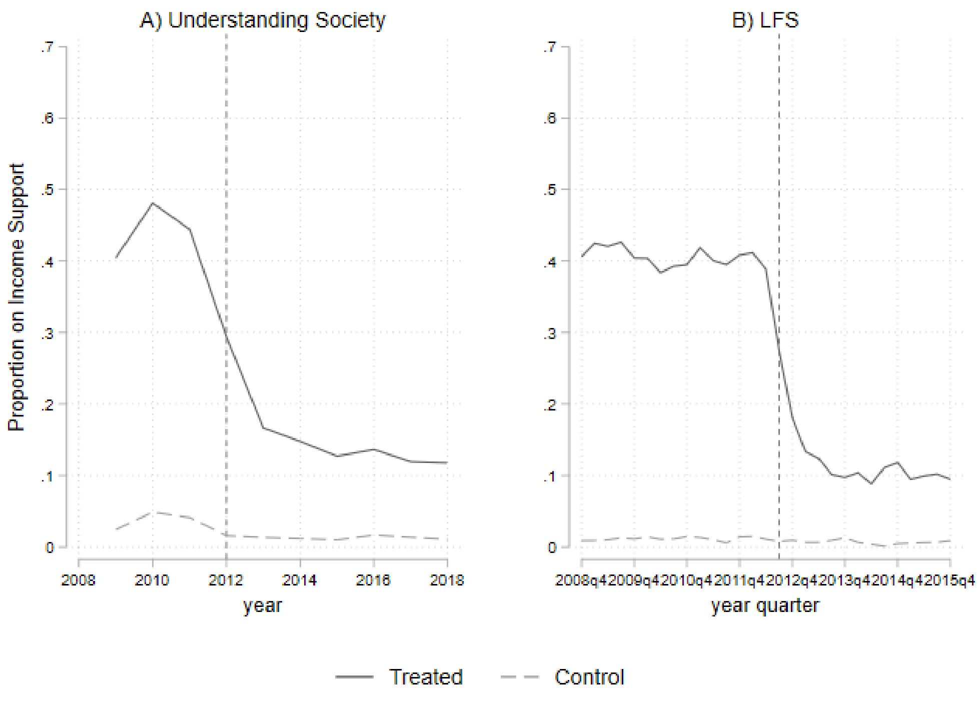
	USOC		LFS	
	(1) Lone mother	(2) Married mother	(3) Lone mother	(4) Married mother
Age	35.04	37.87	33.81	37.31
	(6.870)	(5.475)	(7.698)	(6.652)
Non-white	0.29	0.25	0.16	0.16
	(0.452)	(0.432)	(0.370)	(0.362)
Education				
Degree	0.18	0.34	0.12	0.31
	(0.384)	(0.472)	(0.331)	(0.464)
Other higher	0.15	0.15	0.08	0.11
	(0.354)	(0.354)	(0.278)	(0.308)
A level	0.20	0.19	0.21	0.18
	(0.398)	(0.392)	(0.408)	(0.388)
GCSE	0.32	0.22	0.33	0.24
	(0.466)	(0.412)	(0.471)	(0.427)
Other qual	0.06	0.05	0.12	0.09
	(0.241)	(0.224)	(0.328)	(0.292)
No qual	0.09	0.06	0.12	0.06
	(0.293)	(0.236)	(0.347)	(0.252)
No. of kids <16 resp. for	1.79	2.04	1.88	2.07
	(0.842)	(0.850)	(0.978)	(0.883)
Health problem	0.30	0.22	0.31	0.22
	(0.457)	(0.411)	(0.463)	(0.415)
Housing tenure				
Owned outright	0.05	0.08	0.03	0.08
	(0.210)	(0.269)	(0.177)	(0.266)
Owned with mortgage	0.22	0.64	0.21	0.64
	(0.413)	(0.480)	(0.404)	(0.480)
Social housing	0.45	0.15	-	-
	(0.498)	(0.360)		
Renting	0.28	0.13	0.75	0.27
	(0.452)	(0.335)	(0.432)	(0.445)
Receives income support	0.27	0.02	0.27	0.01
	(0.445)	(0.147)	(0.446)	(0.0953)
Receives job seekers allowance	0.06	0.02	0.06	0.01
	(0.313)	(0.139)	(0.235)	(0.0829)
Employed	0.56	0.69	0.56	0.71
	(0.497)	(0.461)	(0.497)	(0.452)
No. of hours worked per week	25.40	25.68	25.91	27.64
	(9.895)	(10.49)	(11.67)	(12.92)
Looked for work in last 4 weeks	0.27	0.11	0.34	0.19
	(0.445)	(0.318)	(0.474)	(0.393)
Uses childcare	0.42	0.43	-	-
	(0.494)	(0.496)		
N	1802	5897	11503	31106

mean coefficients; std dev in parentheses

Figure 3.1 plots IS receipt between 2009 and 2018 for both datasets; the red dashed line indicating the year of policy implementation. The proportion of lone mothers receiving IS falls from around 40 percent pre-reform to 15 percent in period following the reform, a finding substantiated by both datasets.[20] As expected, it was accompanied by an increase in the proportion of lone mothers receiving JSA post-reform (see Figure 3.2). Whilst both datasets show disparities in IS receipt between the treatment and control group, crucially the pre-reform trend looks similar, and visually the LFS and Understanding Society are very much in line with each other. More generally, there has been a declining trend in JSA receipt (see Appendix B Figure A.1). This could be explained in part by the introduction of Universal Credit from 2013 – though official statistics suggest that new benefit claimants did not start to receive it until early 2015. There was also a job boom which saw unemployment fall from 2.7 million in 2011 to under 1.7 million in November 2015.

Figure 3.3 presents the trends for another outcome variable, the receipt of health-related benefits. The treatment and control groups show fairly similar trends pre-reform and there seems to be a slightly higher incidence of health-related benefits in Understanding Society.

Figure 3.1 – Trends in Income Support receipt



[20] The 15 percent remaining on IS are likely to be receiving income support on the grounds of being in another disadvantaged group.

Figure 3.2 – Trends in Job Seekers Allowance receipt

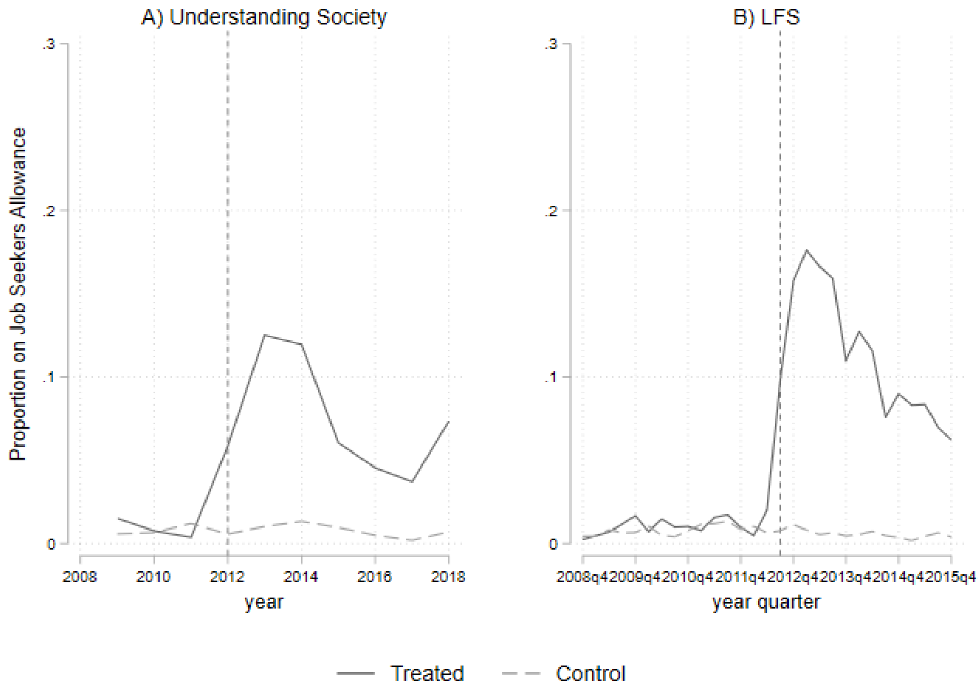
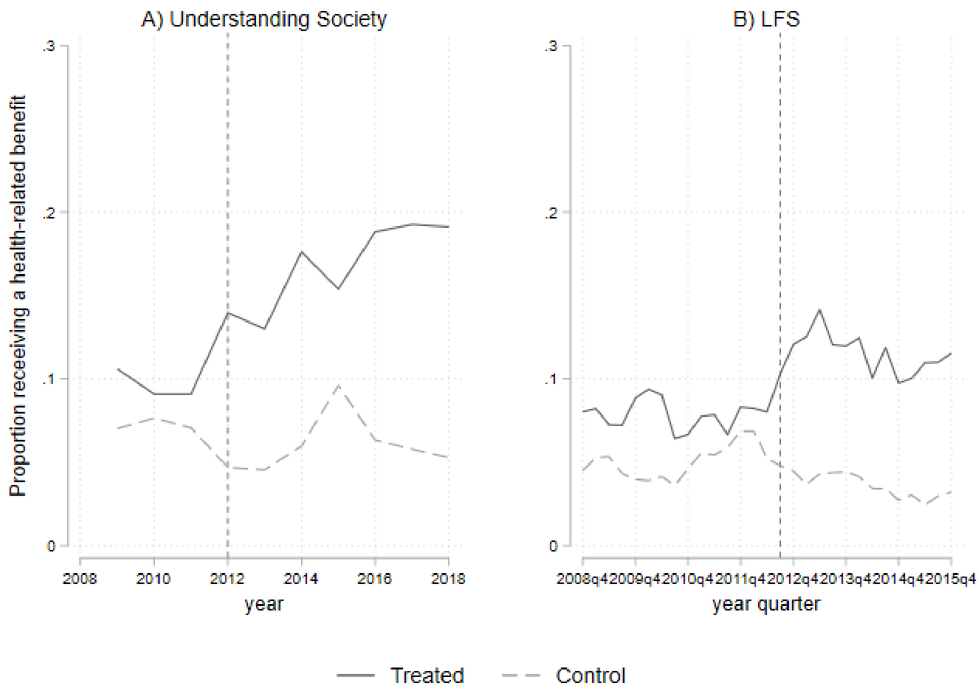
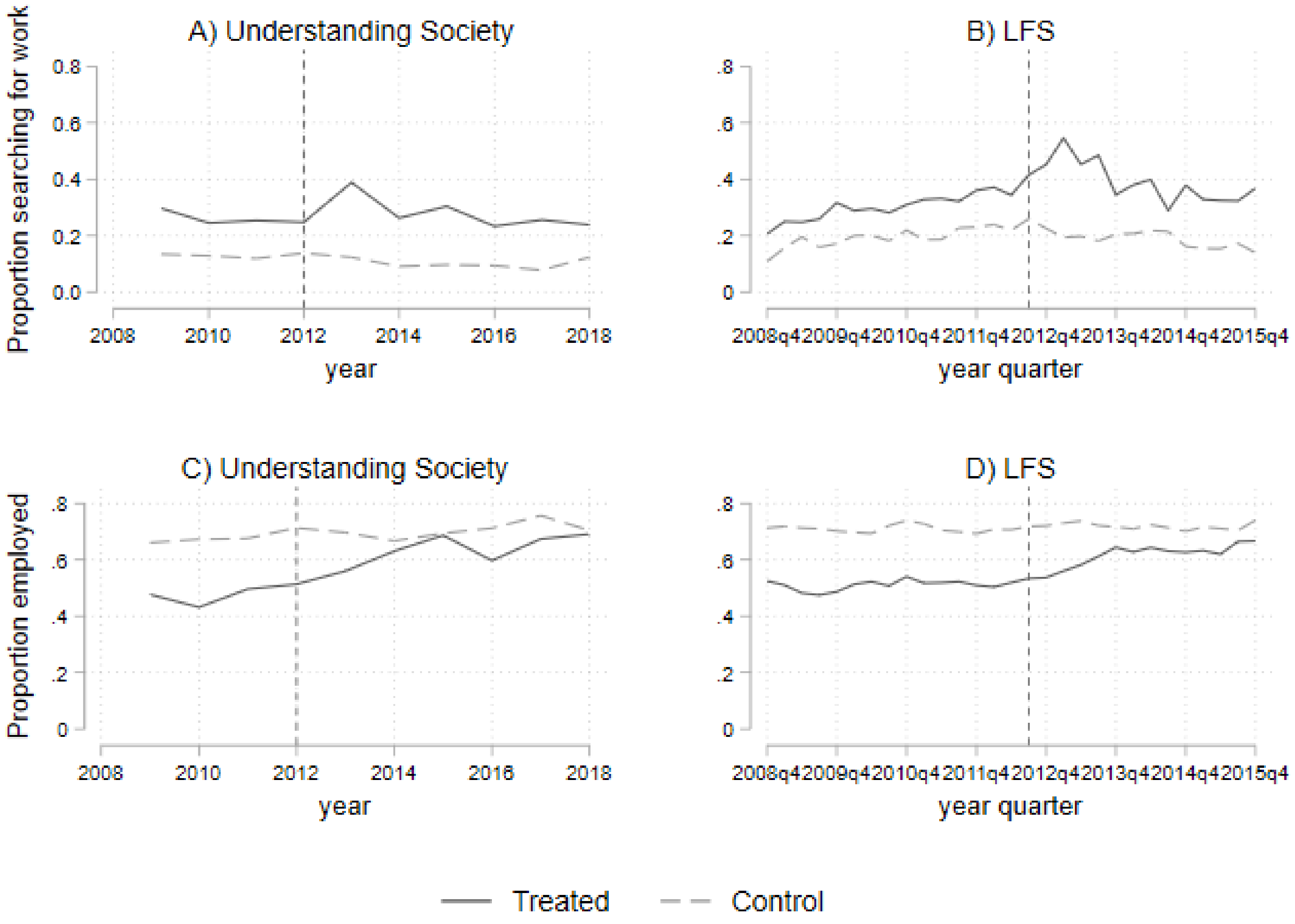


Figure 3.3 – Trends in health-related benefit receipt



The top row of Figure 3.4 presents the trends in work search activity across the two datasets. Work search activity is coded as 0 if the respondent has not been looking for work in

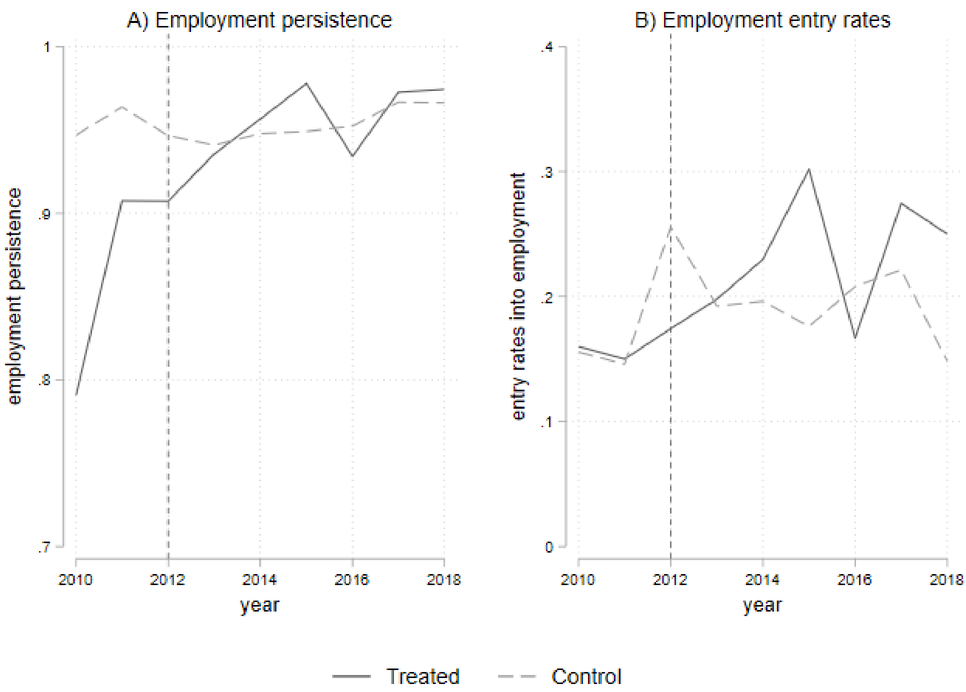
Figure 3.4 – Trends in work search activity and employment rates



the last 4 weeks and 1 if they have. On the bottom, panel C and D correspond to employment trends, as measured by a binary indicator for whether the respondent is employed or not (regardless of the number of hours worked). Both datasets show similar trends, with small jumps after the implementaton of the reform. The LFS is noisier due to the quarterly nature of the data collection but, graphically, they look remarkably consistent.

Figure 3.5 corresponds to employment persistence and entry into employment. Here the sample sizes are further reduced as the sample is split into being employed/unemployed in the previous wave – outcomes which are only available in Understanding Society. The smaller sample is likely responsible for the noisiness of the data points.

Figure 3.5 – Employment exit and entry rates (Understanding Society only)



3.5 Empirical strategy

I follow a difference-in-difference approach using married mothers with a youngest child aged 5 or 6 as the control group. Using a sample of mothers both in receipt and not in receipt of IS, the estimation is an intention-to-treat. In constructing the treated and control groups, lone mothers with a youngest child aged 5 and 6 year olds become ineligible for IS after the reform, whereas married mothers of the same age children are unaffected. This is at odds with the control group used in Garaud (2014) and the concept in Avram et al. (2018) who use lone parents with younger children as the control group (which would be 3-4 year olds in this instance). The reasoning behind selecting a different control group stems from the 2010 introduction of 15 hours of free childcare for 3–4-year-olds. If this policy encourages lone mothers into employment, and these effects are long-lasting or take time to trickle through, then this might mask the true effect of the policy. Conversely, I do not use lone mothers with a youngest child over the age of 7 as they were affected by a previous phase of the LPO reform in the pre-treatment period (2010).

The main estimating equation is a flexible difference-in-difference model which can account for differential time trends between the treatment and control group. The following equation is estimated using Ordinary Least Squares (OLS):

$$\begin{aligned} \mathrm{y_{it}} = \alpha + \gamma\,\mathrm{Treated_i} + \delta\mathrm{Reform_t} + \beta\mathrm{Treated}_i \text{ x } \mathrm{Reform_t} + \\ \varphi\mathrm{Treated}_i \text{ x } \mathrm{Time\ trend}_t + \tau\mathrm{Time\ trend}_t + \theta\mathrm{X_{it}} + Year_t + Region_i + \varepsilon_{\mathrm{it}} \quad (1) \end{aligned}$$

where y_{it} is the outcome variable of interest, Reform is a binary indicator of whether the respondent was interviewed pre- or post-reform, and β is the coefficient of interest on the interaction of the two. To account for differences in group-specific compositional changes over time, individual characteristics are included: age, age squared, ethnicity, education level, number of children and an indicator for health problems, as well as geographic variables including region of residence, regional employment, and regional economic inactivity rates to capture changes in the local labour market. Year and region fixed effects are also included. A restriction that the LPO reform has the same effect in each period post-reform is imposed - although this assumption is later relaxed.

The main identifying assumption is that trends in the outcome would be the same in both treatment and control groups in the absence of treatment. Besley and Burgess (2004) show that large and statistically significant treatment effects may be sensitive to differential time trends in a difference-in-difference setting. Thus, an interaction of a linear time trend and treatment dummy is included to test for differential trends between treatment and control groups. Results tables are presented both with and without it. This relates to another important assumption that during the period there were no other policies implemented or shocks occurring that affected the control and treatment group differentially.

Another source of bias could arise as a result of lone mothers' behavioural responses to announcement of the policy reform - 2 years prior to its actual implementation. If the announcement caused lone mothers to start looking for employment then the estimates would not identify the policy effect. This is tested and discussed in Section 3.6.2.

3.6 Results

3.6.1 Employment and benefit outcomes

Table 3.4 reports the estimated coefficient on the difference-in-difference estimator for receipt of IS. Estimates from both Understanding Society (Panel A) and the LFS (Panel B) are presented alongside each other. Column (1) presents the basic specification with only year and region fixed effects included, column (2) includes all controls described in the methodology section, and column (3), the preferred specification, adds a differential time trend. Reassuringly, moving from column (1) to (2) shows negligible changes in the coefficient size for both outcomes in both datasets. Looking at the results from Understanding Society, Panel A suggests the reform results in a 20 percentage point reduction in IS receipt. A result corroborated by the LFS in Panel B of 24 percentage points.

Table 3.5 reports the estimates for the outcome JSA receipt again presenting both datasets as before. Panel A indicates lone mothers are 13 percentage points more likely to move onto JSA, a coefficient size double that of column (1). An important point to note, is that estimates are remarkably similar in the two datasets. The coefficient on JSA in column (3) is exactly the same and statistically significant in both datasets, though in the LFS it is measured with greater precision.

The estimates in column (4) and (5) serve as a sensitivity check, obtained from a sub-sample of married mothers with education below A level ("low education sample"). This generates small differences in coefficient sizes, suggesting that the reform reduces IS receipt in this group by around 18 percentage points, compared with 20 percentage points in column (3), the baseline specification, and JSA receipt by 11.7 compared with 12.9.

Table 3.4 - Difference-in-difference estimates (OLS) on the likelihood of respondent being in receipt of Income Support

	(1)	(2)	(3)	(4) Low education sample (2)	(5) Low education sample (3)
Dependent variable: Income Support receipt					
A) Understanding Society					
Reform x Treated	-0.270***	-0.263***	-0.202***	-0.228***	-0.183***
	(0.022)	(0.021)	(0.034)	(0.024)	(0.037)
Treated	0.405***	0.377***	0.412***	0.455***	0.481***
	(0.019)	(0.018)	(0.023)	(0.024)	(0.029)
N	7,528	7,379	7,379	3,304	3,304
R-squared	0.225	0.269	0.270	0.259	0.259
B) LFS					
Reform x Treated	-0.278***	-0.290***	-0.245***	-0.344***	-0.228***
	(0.008)	(0.008)	(0.017)	(0.013)	(0.029)
Treated	0.393***	0.377***	0.401***	0.505***	0.496***
	(0.006)	(0.007)	(0.011)	(0.013)	(0.019)
N	42,609	36,181	36,181	12,157	12,157
R-squared	0.262	0.297	0.297	0.230	0.231
Year & region fixed effects	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	Yes	Yes	Yes
Employment & inactivity rates	No	Yes	Yes	Yes	Yes
Differential time trend	No	No	Yes	No	Yes

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

Notes: In the low education sample, the control group is restricted to married mothers of 5-6 year olds with educational qualifications at GCSE level or below.

Table 3.5 - Difference-in-difference estimates (OLS) on the likelihood of respondent being in receipt of Job Seekers Allowance

	(1)	(2)	(3)	(4) Low education sample (2)	(5) Low education sample (3)
Dependent variable: Job Seekers Allowance receipt					
A) Understanding Society					
Reform x Treated	0.074***	0.075***	0.129***	0.078***	0.117***
	(0.009)	(0.009)	(0.021)	(0.011)	(0.024)
Treated	-0.000	-0.013***	0.018*	0.002	0.024*
	(0.004)	(0.004)	(0.010)	(0.008)	(0.013)
N	7,685	7,528	7,528	3,354	3,354
R-squared	0.042	0.061	0.065	0.068	0.070
B) LFS					
Reform x Treated	0.105***	0.097***	0.129***	0.113***	0.170***
	(0.0046)	(0.0049)	(0.0098)	(0.0085)	(0.019)
Treated	0.002	-0.003*	0.014***	0.004	0.032***
	(0.001)	(0.002)	(0.004)	(0.007)	(0.008)
N	42,609	36,172	36,172	11,387	11,387
R-squared	0.065	0.066	0.067	0.067	0.068
Year & region fixed effects	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	Yes	Yes	Yes
Employment & inactivity rates	No	Yes	Yes	Yes	Yes
Differential time trend	No	No	Yes	No	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Notes: In the low education sample, the control group is restricted to married mothers of 5-6 year olds with educational qualifications at GCSE level or below.

Table 3.6 reports estimates for work search activity again presenting both datasets next to each other.[21] Here there is greater discrepancy between datasets. Starting with job search activity, the estimates increase in size upon including a differential time trend, Understanding Society more so than the LFS. However, the estimates in the preferred specification in column (3) are remarkably comparable, 10.5 percentage points vs 9.7 percentage points and the larger sample size in the LFS helps with precision. The low education sample is considerably larger

[21] I try to look at job search intensity but the sample size becomes underpowered, particularly in Understanding Society.

in the LFS, indeed more than double the size of the initial coefficient indicating 9.7 percentage points could already be a lower bound.

In Table 3.7, employment as the outcome shows the greatest differences between the datasets which arises when controlling for the differential time trend (column 3). This time it does not appear to be a precision issue – the coefficient on employment in Understanding Society is 1.7 percentage points and not statistically significant, considerably smaller than the statistically significant 7.9 percentage points in the LFS. This could be related to the limited number of periods available pre-reform in Understanding Society. There is also a concern that the low education sample is very different in the LFS, almost double the coefficient on the main sample. Considering the larger sample size and number of pre-reform periods available, the LFS gives the most compelling estimate, though these effect sizes are examined in Section 3.6.2 in a time-event context which shows that the results do not contradict each other.

Another important destination of lone parents is health-related benefits. Table 3.8 shows the likelihood of being on health-related benefits is significantly higher (6pp) for the treatment group post-reform, suggesting that lone mothers move onto another type of benefit, an effect size slightly smaller than that of the effect size for employment (in the Labour Force Survey).

However, the results above may be obscuring the effect of lone mothers remaining in employment (particularly non-IS claimant lone mothers). In Table 3.9, I test this by conditioning on the previous wave employment status, although this is only possible in longitudinal Understanding Society. The results in Panel A and B show that lone mothers neither increase their entry into employment nor are less likely to exit employment. This may be unsurprising given the small sample sizes and insignificance of the coefficient when using employment as an outcome.

Table 3.6 - Difference-in-difference estimates (OLS) on the likelihood of respondent searching for work

	(1)	(2)	(3)	(4) Low education sample (2)	(5) Low education sample (3)
Dependent variable: Work Search Activity					
A) Understanding Society					
Reform x Treated	0.060*	0.053	0.105*	0.060	0.102
	(0.036)	(0.036)	(0.063)	(0.040)	(0.071)
Treated	0.129***	0.096***	0.127***	0.089***	0.114**
	(0.026)	(0.026)	(0.040)	(0.032)	(0.045)
N	2,465	2,379	2,379	1,471	1,471
R-squared	0.051	0.083	0.083	0.109	0.110
B) LFS					
Reform x Treated	0.0909***	0.0932***	0.0973***	0.135***	0.180***
	(0.0158)	(0.0165)	(0.0329)	(0.0238)	(0.0494)
Treated	0.111***	0.102***	0.104***	0.117***	0.139***
	(0.0101)	(0.0108)	(0.0178)	(0.0173)	(0.0253)
N	14,000	12,060	12,060	5,479	5,479
R-squared	0.048	0.087	0.087	0.139	0.139
Year & region fixed effects	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	Yes	Yes	Yes
Employment & inactivity rates	No	Yes	Yes	Yes	Yes
Differential time trend	No	No	Yes	No	Yes

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

Notes: In the low education sample, the control group is restricted to married mothers of 5-6 year olds with educational qualifications at GCSE level or below.

Table 3.7 - Difference-in-difference estimates (OLS) on the likelihood of respondent being in employment

	(1)	(2)	(3)	(4) Low education sample (2)	(5) Low education sample (3)
Dependent variable: Employed					
A) Understanding Society					
Reform x Treated	0.116***	0.083***	0.017	0.116***	-0.001
	(0.027)	(0.025)	(0.042)	(0.031)	(0.053)
Treated	-0.195***	-0.112***	-0.150***	0.116***	-0.001
	(0.021)	(0.020)	(0.028)	(0.031)	(0.053)
N	7,684	7,527	7,527	3,354	3,354
R-squared	0.035	0.203	0.204	0.213	0.215
B) LFS					
Reform x Treated	0.0910***	0.106***	0.0795***	0.107***	0.130***
	(0.0105)	(0.0104)	(0.0210)	(0.0235)	(0.0475)
Treated	-0.193***	-0.116***	-0.130***	-0.052***	-0.041
	(0.007)	(0.008)	(0.012)	(0.017)	(0.026)
N	42,401	36,011	36,011	11,288	11,288
R-squared	0.036	0.190	0.190	0.226	0.226
Year & region fixed effects	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	Yes	Yes	Yes
Employment & inactivity rates	No	Yes	Yes	Yes	Yes
Differential time trend	No	No	Yes	No	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Notes: In the low education sample, the control group is restricted to married mothers of 5-6 year olds with educational qualifications at GCSE level or below.

Table 3.8 - Difference-in-difference estimates (OLS) on the likelihood of respondent being on health-related benefits

	(1)	(2)	(3)	(4) Low education sample (2)	(5) Low education sample (3)
Dependent variable: Health related benefits					
A) Understanding Society					
Reform x Treated	0.076***	0.082***	0.065**	0.086***	0.095***
	(0.017)	(0.016)	(0.029)	(0.022)	(0.036)
Treated	0.028**	0.002	-0.008	0.008	0.014
	(0.012)	(0.012)	(0.019)	(0.019)	(0.025)
N	7,685	7,528	7,528	3,354	3,354
R-squared	0.021	0.124	0.125	0.145	0.145
B) LFS					
Reform x Treated	0.050***	0.043***	0.061***	0.037**	0.057*
	(0.0060)	(0.0060)	(0.012)	(0.016)	(0.033)
Treated	0.028***	0.0097**	0.019***	0.019	0.029
	(0.0039)	(0.0041)	(0.0068)	(0.014)	(0.018)
N	42,609	36,172	36,172	11,387	11,387
R-squared	0.016	0.176	0.176	0.212	0.212
Year & region fixed effects	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	Yes	Yes	Yes
Employment & inactivity rates	No	Yes	Yes	Yes	Yes
Differential time trend	No	No	Yes	No	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Notes: In the low education sample, the control group is restricted to married mothers of 5-6 year olds with educational qualifications at GCSE level or below.

Table 3.9 - Difference-in-difference estimates (OLS) on the entry and exit rates into and out of employment (Understanding Society only)

	(1)	(2)	(3)	(4) Low education sample (2)	(5) Low education sample (3)
A) Entry into employment					
Reform x Treated	0.033	0.027	-0.022	0.113**	0.068
	(0.039)	(0.039)	(0.059)	(0.044)	(0.065)
Treated	-0.007	0.016	-0.025	-0.012	-0.048
	(0.030)	(0.031)	(0.049)	(0.037)	(0.053)
N	2,154	2,091	2,091	1,230	1,230
R-squared	0.016	0.076	0.077	0.073	0.074
B) Persistence in employment					
Reform x Treated	0.068***	0.062**	0.030	0.064**	0.028
	(0.026)	(0.026)	(0.035)	(0.032)	(0.044)
Treated	-0.071***	-0.052**	-0.077**	-0.027	-0.057
	(0.024)	(0.024)	(0.031)	(0.032)	(0.042)
N	3,912	3,870	3,870	1,337	1,337
R-squared	0.011	0.027	0.027	0.062	0.063
Year & region fixed effects	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	Yes	Yes	Yes
Employment & inactivity rates	No	Yes	Yes	Yes	Yes
Differential time trend	No	No	Yes	No	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: Panel A shows the entry rates into employment i.e. conditional on lone mother being unemployed in the previous wave. Panel B looks at the probability she remains in employment given that she was employed in the previous wave. In the low education sample, the control group is restricted to married mothers of 5-6 year olds with educational qualifications at GCSE level or below.

Behavioural responses at the intensive margin are likely to be important but difficult to disentangle. Under JSA rules, usually jobseekers must be available to work for 40 hours a week. Single parents can limit the amount of hours they work to take account of caring responsibilities. Despite this, they must be able to work for a minimum of 16 hours per week – effectively full-time work. Table B.1 in the Appendix shows some discrepencies between the two datasets. The smaller Understanding Society suggests a positive and significant effect of increasing full-time work (relative to part-time) work by 10 percentage points. The LFS shows no effect with the coefficient estimate being very small, though positive and statistically signifiance in the low education sample. It is not possible to conclusively determine the effect of the reform on the intensive margin of work.

Finally, to allow for labour market outcomes to be correlated I re-run the above analysis clustering standard errors at the government office region. This is the lowest level of geography available in the end user licence dataset. This does not change the significance of any of the results presented above.

3.6.2 Anticipation effects

If lone mothers respond to the policy announcement in 2010, before the policy is implemented in 2012, then our estimates could be masking the true effect of the policy reform, and subject to downward bias. This is plausible given that claimant lone mothers are reminded they will lose IS in quarterly WFIs in the 12 months leading up to losing IS eligibility. Relaxing the assumption that the policy has the same effect in each quarter, if there are anticipation effects I would expect the coefficient on interacted period and reform in 2011 and 2010 to be positive and statistically significant. Table 3.10 shows this is not the case - lone mothers are not responding to the loss of IS eligibility prior to May 2012. LFS results are presented in Table 3.11 and are consistent with this interpretation. Interestingly, because the LFS data is collected quarterly, it can shine a light on the outcome in each quarter pre-reform, capturing any short-term anticipation effects. This result is depicted in Figure 3.6 (LFS) and Figure 3.7 (Understanding Society) for the two main outcomes of interest: work search activity and employment. The effect on employment becomes statistically significant around 5 quarters after the reform and peaks 2 years post-reform, while the increase in job search activity (conditional on being unemployed) was almost instantaneous and lasts for around a year. Having quarterly data (Figure 3.6) in addition to annual data captures both job search activity and employment better, as work search activity increases in the four quarters of the reform, and is accompanied by increases in employment around 5 quarters after the reform. This lends credibility to the idea that some lone mothers respond to the reform by first searching for a job and going on to find a job in the next period. Understanding Society data is not as granular but still shows that work search activity increases by the largest amount in the year following the reform and two and three years following the reform show the largest increases in employment. The coefficient sizes on work search activity four quarters after the reform (LFS), compared with one year after (Understanding Society) are similar in size, 18 versus 12 percentage points.

Though the initial coefficient estimate on employment (Panel A of Table 3.7) was not statistically significant and was small in magnitude, the evidence in Figures 3.6 and 3.7 suggest that the frequency of data collection may be an important factor in this finding. The coefficient sizes on employment 5 quarters after the reform (LFS), compared with 1 year after (Understanding Society) are 10 versus 14 percentage points. It is likely that this event-study finding is the reason Garaud (2014) finds no effect on employment. His sample is comprised of the available data which was quarters up to and including Q1 of 2013, meaning he only analyses 3 quarters following the policy implementation. Turning to my results, Table 3.11 (and Figure 3.6) shows at Q1 2013 there is no effect of the LPO on employment, but increases in employment take time to emerge in later periods (five quarters post-reform) and persist for several periods after. Garaud (2014) does not examine work search activity.

Table 3.10 - Difference-in-difference estimates (OLS) with dynamic treatment effects for all outcomes in Understanding Society

		(1) IS	(2) JSA	(3) Job search	(4) Job	(5) Health rel. benefit	(5) Entry	(6) Persistence
Pre-reform period	2010 x Treated	0.061	-0.003	-0.024	-0.070	-0.026	-	
		(0.052)	(0.013)	(0.077)	(0.056)	(0.034)		
	2011 x Treated	0.035	-0.008	-0.007	0.001	-0.024	0.017	0.097
		(0.052)	(0.013)	(0.078)	(0.056)	(0.035)	(0.076)	(0.069)
Post-reform period	2012 x Treated	-0.092*	0.053***	-0.021	-0.023	0.043	-0.064	0.123*
		(0.052)	(0.020)	(0.080)	(0.058)	(0.037)	(0.079)	(0.069)
	2013 x Treated	-0.211***	0.113***	0.123	0.012	0.045	0.010	0.143**
		(0.049)	(0.026)	(0.086)	(0.059)	(0.037)	(0.082)	(0.068)
	2014 x Treated	-0.234***	0.105***	0.028	0.093	0.089**	0.032	0.159**
		(0.049)	(0.027)	(0.087)	(0.060)	(0.040)	(0.086)	(0.067)
	2015 x Treated	-0.241***	0.045**	0.036	0.135**	0.033	0.130	0.175***
		(0.049)	(0.020)	(0.093)	(0.059)	(0.040)	(0.095)	(0.066)
	2016 x Treated	-0.241***	0.035*	-0.004	0.017	0.089**	-0.041	0.123*
		(0.051)	(0.020)	(0.089)	(0.061)	(0.042)	(0.087)	(0.070)
	2017 onwards	-0.268***	0.040**	-0.005	0.086	0.078*	0.069	0.160**
		(0.048)	(0.018)	(0.087)	(0.058)	(0.040)	(0.089)	(0.065)
	Treated	0.344***	-0.009	0.113*	-0.086*	-0.004	0.021	-0.139**
		(0.042)	(0.012)	(0.065)	(0.046)	(0.725)	(0.065)	(0.063)
	N	7,379	7,535	2,379	7,534	7,535	2,091	3,870
	R Squared	0.269	0.065	0.084	0.205	0.125	0.076	0.029

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: All models include year and region fixed effects, individual-level controls for age, age squared, ethnicity, education level, number of children, region of residence, and an indicator for health problems, employment and inactivity rates, and an interaction term between the treatment group and time trend.

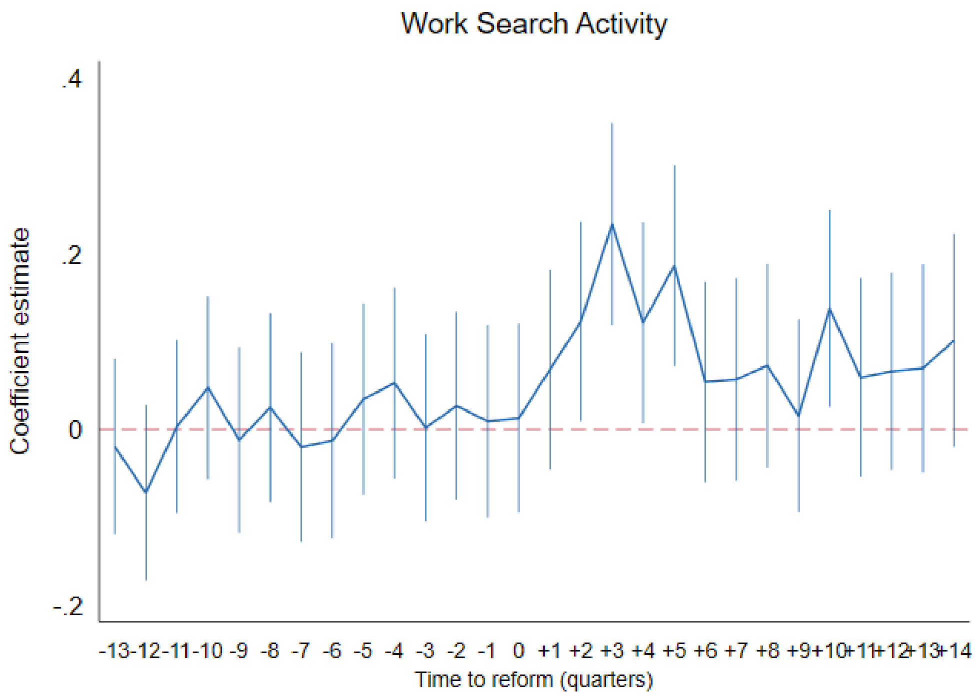
Table 3.11 - Difference-in-difference estimates (OLS) with dynamic treatment effects for all outcomes in the Labour Force Survey

		(1) IS	(2) JSA	(3) Job search	(4) Job	(5) Health-related benefit
Pre-reform period	2011 Q4 x Treated	0.0148	-0.0103	0.0267	-0.00943	-0.0193
		(0.0363)	(0.00802)	(0.0545)	(0.0395)	(0.0209)
	2012 Q1 x Treated	0.0197	-0.0195***	0.00922	-0.0293	-0.0135
		(0.0362)	(0.00712)	(0.0559)	(0.0391)	(0.0211)
	2012 Q2 x Treated	-0.0106	0.00178	0.0126	-0.00929	-0.00507
		(0.0358)	(0.00982)	(0.0547)	(0.0388)	(0.0210)
Post-reform period	2012 Q3 x Treated	-0.136***	0.0774***	0.0678	0.0240	0.0211
		(0.0343)	(0.0173)	(0.0579)	(0.0391)	(0.0213)
	2012 Q4 x Treated	-0.236***	0.127***	0.123**	0.0290	0.0375*
		(0.0322)	(0.0207)	(0.0577)	(0.0395)	(0.0217)
	2013 Q1 x Treated	-0.272***	0.146***	0.233***	0.0414	0.0610***
		(0.0312)	(0.0219)	(0.0585)	(0.0399)	(0.0225)
	2013 Q2 x Treated	-0.282***	0.133***	0.121**	0.0503	0.0700***
		(0.0304)	(0.0203)	(0.0583)	(0.0393)	(0.0224)
	2013 Q3 x Treated	-0.298***	0.130***	0.186***	0.101***	0.0297
		(0.0298)	(0.0189)	(0.0582)	(0.0384)	(0.0214)
	2013 Q4 x Treated	-0.298***	0.100***	0.0538	0.122***	0.0428*
		(0.0298)	(0.0169)	(0.0581)	(0.0379)	(0.0219)
	Treated	0.365***	-0.00890**	0.0975***	-0.102***	0.0170
		(0.0251)	(0.00365)	(0.0346)	(0.0271)	(0.0139)
	N	36,172	36,172	12,060	36,011	36,172
	R Squared	0.300	0.070	0.089	0.191	0.177

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

Note: Only a subset of time period reform interactions are presented here. Figures 3.6 and 3.7 display the full set. All models include year and region fixed effects, individual-level controls for age, age squared, ethnicity, education level, number of children, region of residence, and an indicator for health problems, employment and inactivity rates, and an interaction term between the treatment group and time trend.

Figure 3.6 – Estimated impact of LPO reform for period before, during and after the reform (LFS)



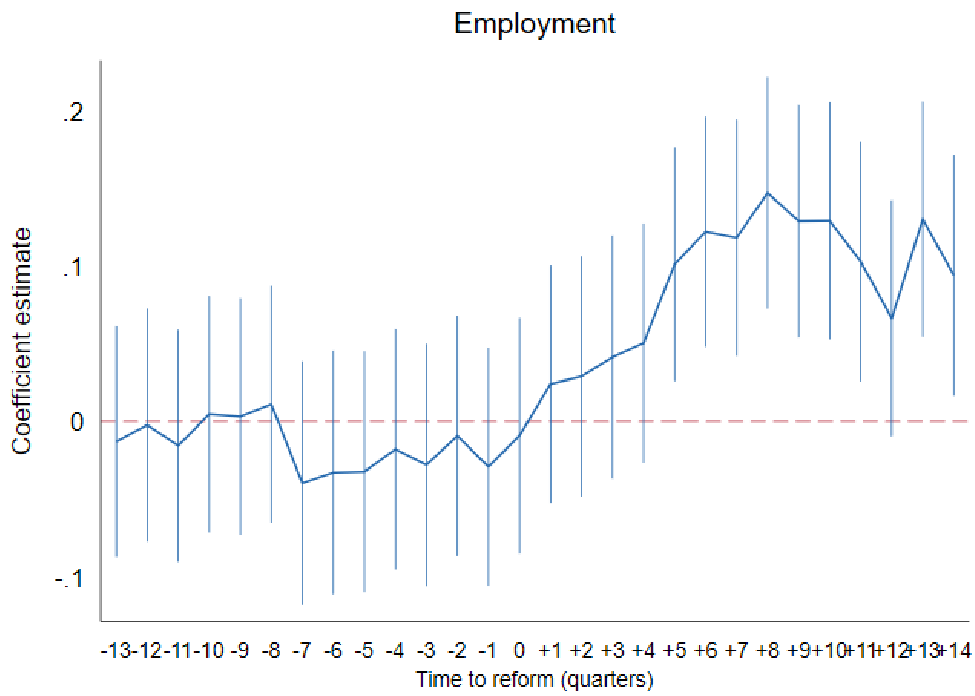
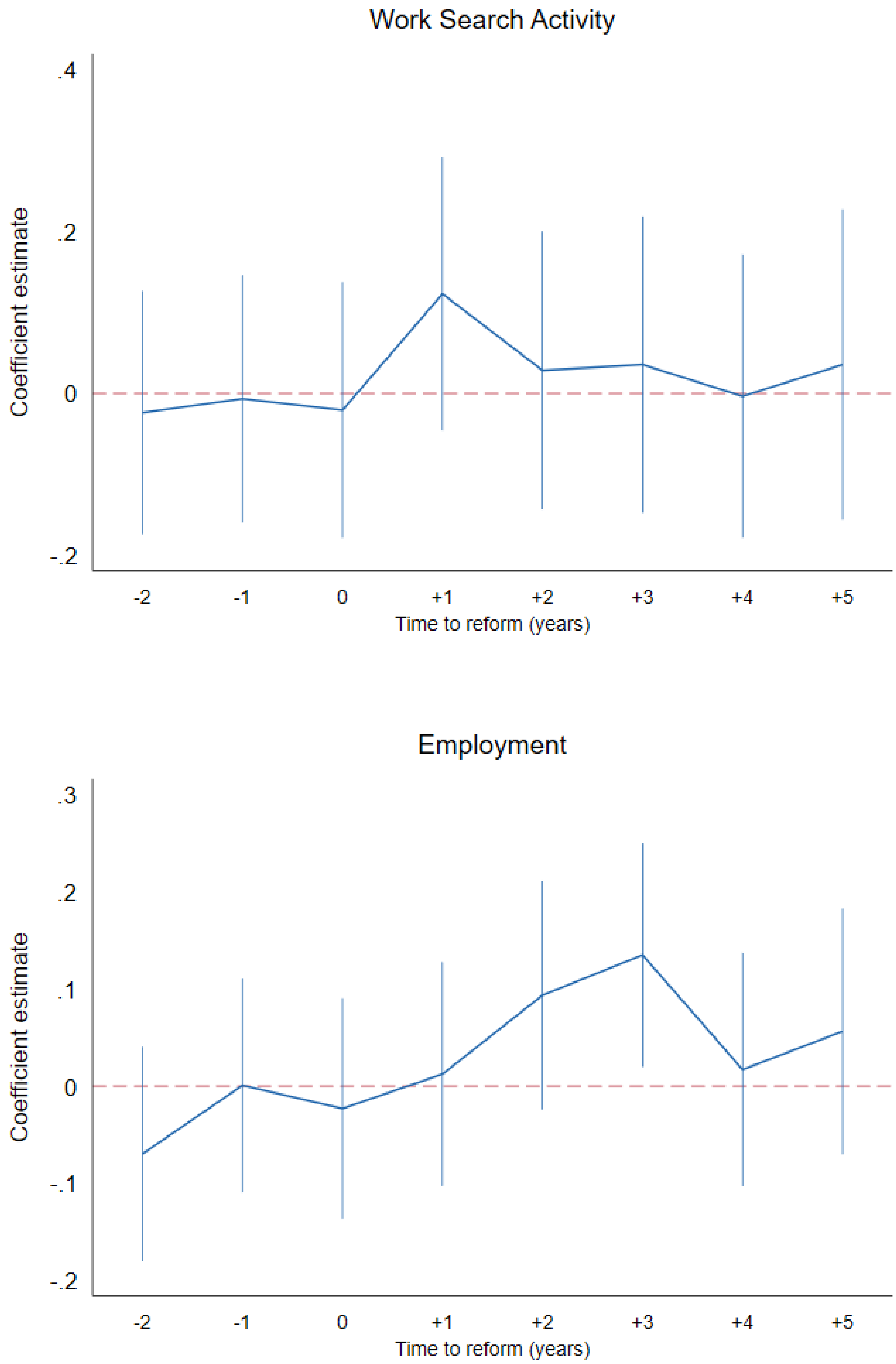


Figure 3.7 – Estimated impact of LPO reform for period before, during and after the reform (Understanding Society)



3.6.3 Heterogeneity

The LPO reform could have created varied labour market responses depending on the characteristics of the lone mother. There is some evidence to suggest that childcare is the greatest barrier to work for lone mothers and lower educated mothers may find that the return to being employed is not large enough to justify it. In response to existing evidence, I explore these dimensions of heterogeneity using a sub-sample of respondents in the LFS only, as cell sizes in Understanding Society become very small. The benefit of this approach is that it allows all coefficients in the model to change and I am able to compare the coefficient size for each of the different sub-samples.

The results are presented in Table 3.12, columns (1) and (2) examine heterogeneity in terms of having older children, (3) and (4) in terms of housing tenure, (5) and (6) education level. A differential time trend is included in all columns. The presence of an older sibling acts as a proxy for childcare, assuming an older sibling makes it more likely they have babysitting duties. However, I find no evidence of heterogeneity by presence of an older child for both work search activity and employment. Differences between low and high educated individuals may be important for policymakers in designing active labour market programmes. Here I find no differences in work search activity nor employment responses between low and high educated lone mothers.

However, the effect on employment is considerably smaller and statistically different for owner/occupiers versus non-owner/occupiers, for owner/occupiers there is no effect on employment where non-owners/occupiers are driving the large effect on employment. One channel through which this could be working is that these lone mothers may have kept the marital home after the relationship ended, are more affluent and do not respond to the removal or removal of a safety net, of the unconditional benefit IS.

The results are also presented using a triple interaction of the treatment group, reform and dimension of heterogeneity. The two methods do not show any significant differences (Table B.2 in Appendix).

Table 3.12 - Estimates of difference-in-difference coefficient on the impact of LPO reform on work search activity and employment by presence of older children, housing tenure and education level – LFS using sub-samples

	(1) Older children	(2) No older children	(3) Difference	(4) Owner occupier	(5) Not owner-occupier	(6) Difference	(7) Low educated	(8) High educated	(9) Difference
Work search activity									
Reform * Treated	0.0681*	0.107*	0.039	0.0933	0.0706*	-0.022	0.135*	0.0942**	-0.041
	(0.0404)	(0.0591)		(0.0926)	(0.0393)		(0.0701)	(0.0371)	
Treated	0.0784***	0.132***		0.0918*	0.0906***		0.0587*	0.122***	
	(0.0206)	(0.0360)		(0.0479)	(0.0224)		(0.0304)	(0.0211)	
N	8,814	3,326		4,544	7,596		2,065	9,995	
Employment									
Reform * Treated	0.0850***	0.0705**	-0.015	-0.00557	0.0929***	0.098**	0.0610	0.0800***	0.019
	(0.0285)	(0.0341)		(0.0338)	(0.0307)		(0.0659)	(0.0229)	
Treated	-0.179***	-0.128***		0.00479	-0.119***		-0.104***	-0.145***	
	(0.0160)	(0.0206)		(0.0186)	(0.0187)		(0.0336)	(0.0135)	
N	25,146	11,130		21,305	14,971		2,934	33,077	

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: Columns (3) reports the raw difference between the previous two columns (2) – (1) with asterisks denoting the level of statistical significance from a Wald test of the null hypothesis that there is no difference between the coefficients. The same logic applies to columns (6) and (9). All models include year and region fixed effects, individual-level controls for age, age squared, ethnicity, education level, number of children, region of residence, and an indicator for health problems, employment and inactivity rates, and an interaction term between the treatment group and time trend. Low education refers to educational qualifications at GCSE level or below.

3.7 Robustness checks

There are several threats to causal interpretation of the results. Here I investigate whether the parallel trends assumption holds and the sensitivity of the results to the control group used.

3.7.1 Parallel trends

The main identifying assumption of the difference-in-difference strategy is that the treatment and control group would follow the same trend in the absence of treatment. This is not directly testable, but my analysis gives a good indication the assumption holds. Firstly, Figures 3.1-3.4 show that graphically the trends appear similar in both datasets. Secondly, relaxing the assumption of linearity in the post-reform periods, shows that the pre-treatment time periods interacted with the treatment are not statistically significant for any of the outcomes. Also controlling for differential trends in the preferred specification should mean that any differential trend is captured in the results. However, there are several other tests which can be performed to further justify why the parallel trends assumption is not violated.

First of all, I estimate a placebo model which simulates that the reform happened in one of the pre-treatment periods, and estimate the regression using only pre-reform data. There is only data available from three periods (corresponding to three years) before the reform in Understanding Society compared with thirteen periods (corresponding to three years) in the LFS. I thus use the LFS for this analysis. The results in Table B.3 show that the placebo regression year had no significant effect on any of the outcomes, reinforcing the validity of the parallel trends assumption.

In addition, perhaps there is a concern around the timing of the introduction of the LPO in 2012, towards the end of a global recession. Any increases in employment could be due to the recovery. Between 2013 and 2018 there was a job boom and this recovery of the labour market could lead to bias in the estimates if it affected treatment and control groups differently. Another placebo test can be conducted using lone mothers (treatment group) and married mothers (control group) who are both unaffected by the reform. Because of the three initial phases of the reform, every single age group has been affected at some point during the time period the data covers. Thus, in order to keep the multiple pre- and post- reform periods, the best option is using younger children as the placebo control group. Mothers with 2-3 year olds are selected as the control group, this is because 4 year olds might start to be affected by the quarterly WFIs lone parents attend the year before losing eligibility. The results are found in the appendix in Table B.4-B.6 and find that, when controlling for differential trends, there are no differences between treatment and control for all outcomes.[22]

Taken together, these results give a good indication that the parallel trends assumption holds. However, the treatment and control groups are quite different characteristically. Lone mothers have fewer educational qualifications, are less likely to own their home and generally less affluent (see Table 3.1). In fact, when using a low education sample of married mothers as the control group, the estimates are much larger for work search activity and employment than when using the full sample.

Kahn-Lane and Lang (2019) point out that difference-in-difference estimates will be more plausible if treatment and control groups are similar in levels and not just trends. I thus combine propensity score matching with difference-in-difference methodology, with the rationale of finding a group of married mothers who are similar to the treated group of lone mothers. Doing this with cross-sectional data is prohibitive, indeed it may introduce "post-treatment bias" (Rosenbaum, 1984) if covariates are included which are affected by the programme. Despite the shortcomings associated with using two cross-sectional datasets in this chapter, I estimate a matched difference-in-difference under the assumption that covariates are exogeneous to the treatment.[23] Any large deviations from the main results would be indicative of potential selection bias. Table B.7 shows that the results from the matched sample do not differ from the main estimates.

3.7.2 Sensitivity to control group

Another potential threat to identification stems from the selection of the control group, whether married mothers with a youngest child aged 5-6 are a good counterfactual. Garaud (2014) uses a control group consisting of lone mothers with younger children (3-4 years old). In 2010 the government introduced 15 hours of free childcare to mothers of 3-4 year olds, which would have affected the employment rates of the control group relative to treatment group (who were unaffected as their youngest child was too old). This would bias estimates downwards which

[22] The specifications without the differential time trend are significant but the effect sizes are small and close to zero. Combined with the insignificance of the preferred specification that includes a differential time trend this is evidence in favour of the economic situation not driving the main results.

[23] The covariates are individual-level characteristics which include age, age squared, education level, number of children, ethnicity, and an indicator for having health problems.

could suggest that the null result found by Garaud (2014) is, at least in part, due to the chosen control group. I test the sensitivity of results to the control group by using lone mothers of 3-4 year olds and find that despite concerns about the control group used, there are no major differences in the estimates when using lone mothers of 3-4 year olds (see results in Table B.8). The estimate for employment is slightly lower, 7 percentage points compared with 9 percentage points in the LFS, but not statistically different from each other. The health-related benefits also become statistically insignificant in Understanding Society, though the effect size is similar and measured with less precision in a smaller sample.

It is worthy of note that concerns about the similarity in levels of control and treatment group, do not hold up when running this sensitivity check. Lone parents with a youngest child aged 3-4 are characteristically similar to lone parents with a child aged 5-6, making them a better "match" to the treatment group.

3.7.3 Other checks

One source of bias in using cross-sectional data can be caused by time-variant unobservables, which effectively introduce omitted variable bias. The treatment should be exogenous to any changes in the composition of the sample for the results to be interpreted causally. In this context, there would be bias introduced if the demographics of the lone parent and married mother groups changed over time and this was associated with a greater likelihood of being in employment. One way of checking this is to compare the pre- and post-reform periods for the treatment and control group. Descriptively, Table 3.1 and 3.2 (for both datasets) show that the characteristics of lone mothers in the pre-reform and post-reform periods are similar, this holds for the control group of married mothers.

However, the outcome variable, work search activity, is conditional on the respondent being employed. Since employment is a transitory state and may change directly because of the reform, the above-described composition bias will be a problem. The result is still of interest, particularly in the event-study framework, but I also test using an outcome of labour force participation where it is equal to 1 if the respondent is either in work or searching for work and 0 if not. This effectively combines the two measures. The results in Panel B of Table B.9 (see Appendix) show similar effect sizes to those on employment of around 9 percentage points. The event study results are presented in Figure B.1 (see Appendix). The LFS result is comparable to employment outcome event-study graphs and increases in labour market participation are significant for each period in the three years (13 quarters) post-reform. However, in Understanding Society, the event study results are no longer statistically significant for any period post-reform than when only using employment as the outcome variable – this may suggest that work search activity may not be fading out due to compositional biases but instead drives the insignificant effect found here.

There is anecdotal evidence from the UK which suggests welfare recipients may change their behaviour to avoid losing welfare. One plausible scenario in the LPO context is that lone mothers may have new children, which would allow them to retain their eligibility for income support. This could introduce composition bias, but, more problematically for policymakers, may be an unintended consequence of the reform. To investigate this possibility, I estimate the same OLS difference-in-difference using a new outcome, namely whether the respondent has had a pregnancy since the last wave of Understanding Society.[24] Past pregnancy is chosen for two reasons. Most importantly, it means that the individual remains in the defined treatment group of lone mothers with a youngest child aged 3-4. Secondly, it captures mothers-to-be, whereas using birth rates would not identify this group. Table B.10 (see Appendix) shows that the difference-in-difference estimator is close to zero and not statistically significant. Lone mothers do not appear to have new children in order to remain on income support.[25]

An alternate way of retaining IS was to move into education or training however, both Understanding Society and the LFS results suggest this does not appear to be the case. Results can be made available on request.

3.8 Conclusion

The objective of the LPO was to increase employment among lone mothers with young children. This paper provides causal evidence of the effects of the roll-out of the 2012 final phase of the reform, affecting lone mothers with children aged 5 or 6. Reducing the age of the youngest child at which the mother loses eligibility to 5, imposed job search requirements onto lone mothers, where before there were none.

[24] The LFS does not contain any variables around birth or pregnancy rates, and so is not used in this analysis.
[25] I also test this in an event study context and rule out any changes in fertility in anticipation of the policy.

I use a difference-in-difference approach to examine the effect of the reform on work search activity, employment, receipt of other benefits, persistence of employment and employment exit rates. The findings suggest lone mothers have responded with an 8 percentage point increase in employment. Work search activity has increased by 9 percentage points and although a significant proportion of lone mothers move onto a health-related benefit (6 percentage points), this is only slightly lower than that of the effect on employment.

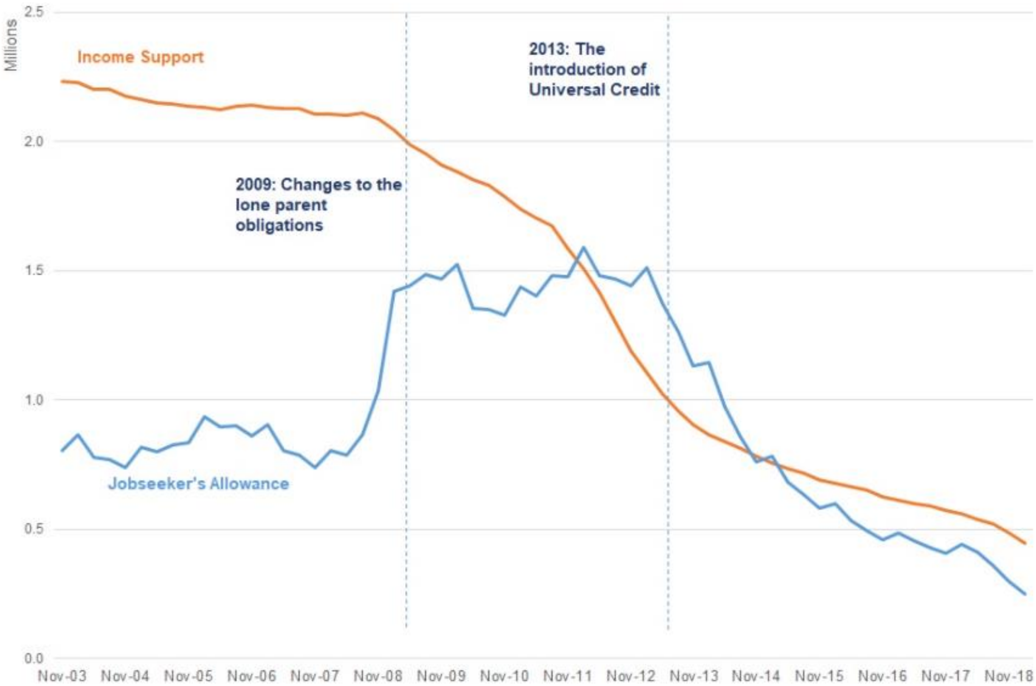
Using two datasets, I demonstrate that the results are remarkably similar. It also highlights the importance of the frequency of data collection in estimating the policy impact. This is demonstrated in Figure 3.6 and 3.7 which show the largest increases in work search activity are concentrated within the first year of the reform, while increases in employment take longer to appear, with the magnitude of the effect peaking between two-three years after the reform. Sensitivity tests suggest that the effects found are not due to time-varying shocks affecting the treatment and control groups differentially, the results are robust to the chosen comparison group, and lone mothers are not adjusting their behaviour in anticipation of the loss of IS. Heterogeneity analysis is conducted, and I find the effects on employment are driven by mothers who do not own their own home. This could be indicative that more affluent lone mothers have fewer incentives to search for and gain employment, as they potentially rely on alimony/child maintenance as a source of income.

These findings mirror those of Avram et al. (2018) who find that the loss of entitlement increases the probability of a lone parent being in work by around 10 percentage points 9 months later. Though this estimate is for the previous three LPO reforms (in 2009, 2010 and 2011) and lone mothers with a younger (unaffected) youngest child are used as the control group, the estimate is still of interest for comparison purposes, and I find a similar effect size present after 12 months in the LFS. In contrast, they find that the effect on movement into health benefits is greater than that of movements into employment, where I find larger effects on employment. This could be due to the relatively older ages of lone mothers in their sample (as they are looking at mothers with a youngest child aged 7 years old or older). The result from Garaud (2014) could not be replicated due to the limited number of quarters that he includes post-reform, at which the effects on employment are yet to appear.

The main takeaway for policymakers is that the reform has been effective in increasing employment among lone mothers. However, the measure of success should not depend solely on increases in work search activity and employment, but also any welfare consequences of the reform. However, no evidence on previous phases of the LPO has considered this. Further research is vital to examine how the reform has affected poverty and standards of living. Another important dimension, which could be the subject of future work, is how the reform affects the outcomes of the children of lone mothers, contributing to existing literature on maternal employment and child outcomes.

Appendix B

Figure B.1 Trends in JSA and Income Support over time with introduction of LPO and Universal Credit
Saurabh



Source: DWP Benefit Statistics

Table B.1 Difference-in-difference estimates (OLS) on the likelihood of respondent being in full-time work

	(1)	(2)	(3)	(4) Low education sample (2)	(5) Low education sample (3)
Dependent variable: Working full-time (ref. category part-time work)					
A) Understanding Society					
Reform x Treated	0.016	0.022	0.101***	-0.032	0.035
	(0.022)	(0.022)	(0.034)	(0.034)	(0.053)
Treated	0.056***	0.058***	0.103***	0.106***	0.146***
	(0.018)	(0.018)	(0.023)	(0.031)	(0.039)
N	5,090	5,023	5,023	1,827	1,827
R-squared	0.021	0.045	0.046	0.066	0.068
B) LFS					
Reform x Treated	-0.037***	-0.0095	0.0069	0.0590	0.154*
	(0.013)	(0.013)	(0.027)	(0.0474)	(0.0912)
Treated	0.0748***	0.0733***	0.0821***	0.0720*	0.119**
	(0.00888)	(0.00971)	(0.0157)	(0.0423)	(0.0589)
N	22,764	19,210	19,210	4,296	4,296
R-squared	0.015	0.053	0.053	0.075	0.075
Year & region fixed effects	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	Yes	Yes	Yes
Employment & inactivity rates	No	Yes	Yes	Yes	Yes
Differential time trend	No	No	Yes	No	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Notes: In the low education sample, the control group is restricted to married mothers of 5-6 year olds with educational qualifications at GCSE level or below.

Table B.2 - Estimates of difference-in-difference coefficient on the impact of LPO reform on work search activity and employment by presence of older sibling, education level and housing tenure

	(1) Job search activity	(2) Job search activity	(3) Employed	(4) Employed
Labour Force Survey				
a) Older Sibling				
Treated*Reform*Older Sibling	-0.053	-0.054	0.051**	0.050**
	(0.035)	(0.035)	(0.020)	(0.020)
Reform*Older Sibling	0.002	0.002	0.025**	0.025**
	(0.021)	(0.021)	(0.011)	(0.011)
Treated*Older Sibling	-0.036	-0.035	-0.017	-0.016
	(0.023)	(0.023)	(0.014)	(0.014)
Reform*Treated	0.128***	0.137***	0.076***	0.036
	(0.029)	(0.039)	(0.016)	(0.023)
b) Owner/occupier				
Treated*Reform*Owner/occupier	0.059	0.060	-0.103***	-0.106***
	(0.049)	(0.049)	(0.021)	(0.021)
Reform*Owner/occupier	-0.038**	-0.040**	0.020*	0.019
	(0.017)	(0.017)	(0.012)	(0.012)
Treated*Owner/occupier	-0.000	-0.001	0.124***	0.126***
	(0.028)	(0.028)	(0.015)	(0.015)
Reform*Treated	0.080***	0.088***	0.124***	0.068***
	(0.019)	(0.033)	(0.014)	(0.022)
c) Low education				
Treated*Reform*Low education	0.043	0.045	0.022	0.022
	(0.037)	(0.037)	(0.033)	(0.033)
Reform*Low education	-0.026	-0.029	-0.022	-0.022
	(0.020)	(0.020)	(0.021)	(0.021)
Treated*Low education	-0.074***	-0.075***	0.054***	0.056***
	(0.022)	(0.022)	(0.021)	(0.021)
Treated*Reform	0.086***	0.104***	0.086***	0.045**
	(0.018)	(0.032)	(0.011)	(0.020)
Year & region fixed effects	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes
Employment & inactivity rates	Yes	Yes	Yes	Yes
Differential time trend	No	Yes	No	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Notes: Low education refers to educational qualifications at GCSE level or below.

Table B.3 - Estimates on the impact of the LPO reform on all outcomes using a placebo reform year of 2010

	(1) IS	(2) JSA	(3) Job search	(4) Job	(5) Health related benefit
B) LFS					
Placebo Reform*Treated	0.0107	-0.00216	-0.0243	-0.0412	-0.0244
	(0.0279)	(0.00691)	(0.0453)	(0.0307)	(0.0168)
Treated	0.372***	-0.00311	0.0694***	-0.115***	0.0151
	(0.0164)	(0.00310)	(0.0255)	(0.0182)	(0.00963)
N	17,492	17,492	6,165	17,426	17,492
R Squared	0.346	0.015	0.068	0.191	0.157
Year & region fixed effects	Yes	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes	Yes
Employment & inactivity	Yes	Yes	Yes	Yes	Yes
Differential time trend	Yes	Yes	Yes	Yes	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: Results come from the Labour Force Survey as it has many pre-reform periods to facilitate the analysis.

Table B.4 - Estimates from a placebo test of the impact of the LPO reform on Income Support and Job Seekers Allowance receipt using an alternative control and treatment group both unaffected by the reform

	(1)	(2)	(3)
Dependent variable: Income Support receipt			
A) Understanding Society			
Reform x Treated	-0.044**	-0.038*	-0.026
	(0.022)	(0.022)	(0.038)
N	10,430	10,198	10,197
R-squared	0.345	0.386	0.386
B) LFS			
Reform x Treated	-0.043***	-0.039***	0.0078
	(0.0085)	(0.0090)	(0.018)
N	60,461	52,344	52,344
R-squared	0.380	0.399	0.399
Dependent variable: Job Seekers Allowance receipt			
C) Understanding Society			
Reform x Treated	0.012***	0.012***	-0.003
	(0.005)	(0.005)	(0.008)
N	10,690	10,438	10,429
R-squared	0.005	0.015	0.015
D) LFS			
Reform x Treated	0.002	0.002	-0.005
	(0.002)	(0.002)	(0.004)
N	60,461	52,344	52,344
R-squared	0.002	0.012	0.012
Year & region fixed effects	Yes	Yes	Yes
Personal characteristics	No	Yes	Yes
Employment & inactivity rates	No	Yes	Yes
Differential time trend	No	No	Yes

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

Note: The control group are married mothers with a youngest child aged 2 or 3 years olds. The treatment group are lone mothers with a youngest child aged 2 or 3 years old.

Table B.5 - Estimates from a placebo test of the impact of the LPO reform on work search activity and employment using an alternative control and treatment group both unaffected by the reform

	(1)	(2)	(3)
Dependent variable: Work Search Activity			
A) Understanding Society			
Reform x Treated	0.021	0.022	-0.013
	(0.025)	(0.025)	(0.043)
N	4,387	4,232	4,232
R-squared	0.029	0.051	0.052
B) LFS			
Reform x Treated	0.044***	0.036***	0.018
	(0.010)	(0.011)	(0.022)
N	24,910	21,550	21,550
R-squared	0.020	0.038	0.038
Dependent variable: Employed			
C) Understanding Society			
Reform x Treated	-0.004	-0.017	0.025
	(0.024)	(0.022)	(0.039)
N	10,686	10,435	10,426
R-squared	0.053	0.230	0.230
D) LFS			
Reform x Treated	0.028***	0.015	-0.005
	(0.009)	(0.009)	(0.018)
N	60,262	52,185	52,185
R-squared	0.047	0.208	0.208
Year & region fixed effects	Yes	Yes	Yes
Personal characteristics	No	Yes	Yes
Employment & inactivity rates	No	Yes	Yes
Differential time trend	No	No	Yes

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

Note: The control group are married mothers with a youngest child aged 2 or 3 years olds. The treatment group are lone mothers with a youngest child aged 2 or 3 years old.

Table B.6 - Estimates from a placebo test of the impact of the LPO reform on health-related benefit receipt using an alternative control and treatment group both unaffected by the reform

	(1)	(2)	(3)
Dependent variable: Health related benefits			
A) Understanding Society			
Reform x Treated	0.014	0.011	-0.012
	(0.012)	(0.012)	(0.021)
N	10,690	10,438	10,429
R-squared	0.010	0.089	0.090
B) LFS			
Reform x Treated	-0.001	-0.002	0.012
	(0.004)	(0.004)	(0.009)
N	60,461	52,344	52,344
R-squared	0.004	0.118	0.118
Year & region fixed effects	Yes	Yes	Yes
Personal characteristics	No	Yes	Yes
Employment & inactivity rates	No	Yes	Yes
Differential time trend	No	No	Yes

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

Note: The control group is defined as married mothers with a youngest child aged 2 or 3 years old. The treatment group are lone mothers with a youngest child aged 2 or 3 years old.

Table B.7 - Comparison of the difference-in-difference estimate when using the matched sample versus main sample

Outcome	**IS**		**JSA**		**Work search activity**		**Job**		**Health rel. benefit**	
	Main sample (1)	Matched sample (2)	Main sample (3)	Matched sample (4)	Main sample (5)	Matched sample (6)	Main sample (7)	Matched sample (8)	Main sample (9)	Matched sample (10)
A) Understanding Society										
Reform x Treated	-0.263***	-0.252***	0.075***	0.0770***	0.053	0.0186	0.083***	0.0592***	0.082***	0.0891***
	(0.021)	(0.0154)	(0.009)	(0.00811)	(0.036)	(0.0328)	(0.025)	(0.0227)	(0.016)	(0.0147)
N	7,379	7,367	7,528	7,523	2,379	2,363	7,527	7,522	7,528	7,523
R-squared	0.269	0.186	0.061	0.031	0.083	0.028	0.203	0.020	0.124	0.012
B) Labour Force Survey										
Reform x Treated	-0.290***	-0.281***	0.097***	0.0987***	0.0932***	0.0858***	0.106***	0.104***	0.043***	0.0479***
	(0.008)	(0.00661)	(0.0049)	(0.00370)	(0.0165)	(0.0159)	(0.0104)	(0.0103)	(0.0060)	(0.00590)
N	36,181	36,133	36,172	36,133	12,060	12,004	36,011	35,977	36,172	36,133
R-squared	0.297	0.213	0.066	0.050	0.087	0.031	0.190	0.011	0.176	0.005

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

Note: Odd numbered columns refer to the main specification without matching. Even number columns are using a matched sample which is estimated using the *diff* command in Stata. All models include year and region fixed effects, personal characteristics, and employment and inactivity rates.

Table B.8 - Estimates on the impact of the LPO reform on all outcomes using an alternative control group of lone mothers with a youngest child aged 3-4

	(1) IS	(2) JSA	(3) Job search	(4) Job	(5) Health related benefit	(6) Entry	(7) Persistence
A) Understanding Society							
Reform*Treated	-0.219***	0.134***	0.128*	0.025	0.054	0.015	0.022
	(0.050)	(0.022)	(0.070)	(0.053)	(0.034)	(0.059)	(0.051)
N	3,704	3,711	1,851	3,722	3,723	1,457	1,296
R Squared	0.205	0.061	0.062	0.188	0.144	0.068	0.046
B) LFS							
Reform*Treated	-0.233***	0.116***	0.0947***	0.0739***	0.0547***	-	-
	(0.024)	(0.010)	(0.034)	(0.025)	(0.014)	-	-
N	21,164	21,164	10,766	20,991	21,164	-	-
R Squared	0.170	0.060	0.081	0.213	0.171	-	-
Year & region fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Personal characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Employment & inactivity	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Differential time trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes

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

Table B.9 - Difference-in-difference estimates (OLS) on labour market participation rates

	(1)	(2)	(3)	(4) Low education sample (2)	(5) Low education sample (3)
Dependent variable: Labour market participation					
A) Understanding Society					
Reform x Treated	0.099***	0.073***	0.059	0.112***	0.043
	(0.026)	(0.024)	(0.041)	(0.031)	(0.052)
Treated	-0.109***	-0.051***	-0.058**	-0.056**	-0.097***
	(0.020)	(0.019)	(0.027)	(0.028)	(0.037)
N	7,498	7,348	7,348	3,284	3,284
R-squared	0.021	0.182	0.182	0.198	0.198
B) LFS					
Reform x Treated	0.0983***	0.103***	0.0908***	0.132***	0.184***
	(0.010)	(0.010)	(0.019)	(0.024)	(0.049)
Treated	-0.104***	-0.046***	-0.053***	-0.008	0.017
	(0.007)	(0.007)	(0.011)	(0.020)	(0.028)
N	42,456	36,059	36,059	11,331	11,331
R-squared	0.017	0.172	0.172	0.248	0.248
Year & region fixed effects	Yes	Yes	Yes	Yes	Yes
Personal characteristics	No	Yes	Yes	Yes	Yes
Employment & inactivity rates	No	Yes	Yes	Yes	Yes
Differential time trend	No	No	Yes	No	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Notes: In the low education sample, the control group is restricted to married mothers of 5-6 year olds with educational qualifications at GCSE level or below.

Figure B.2 - Estimated impact of LPO reform on labour market participation rates for period before, during, and after the reform

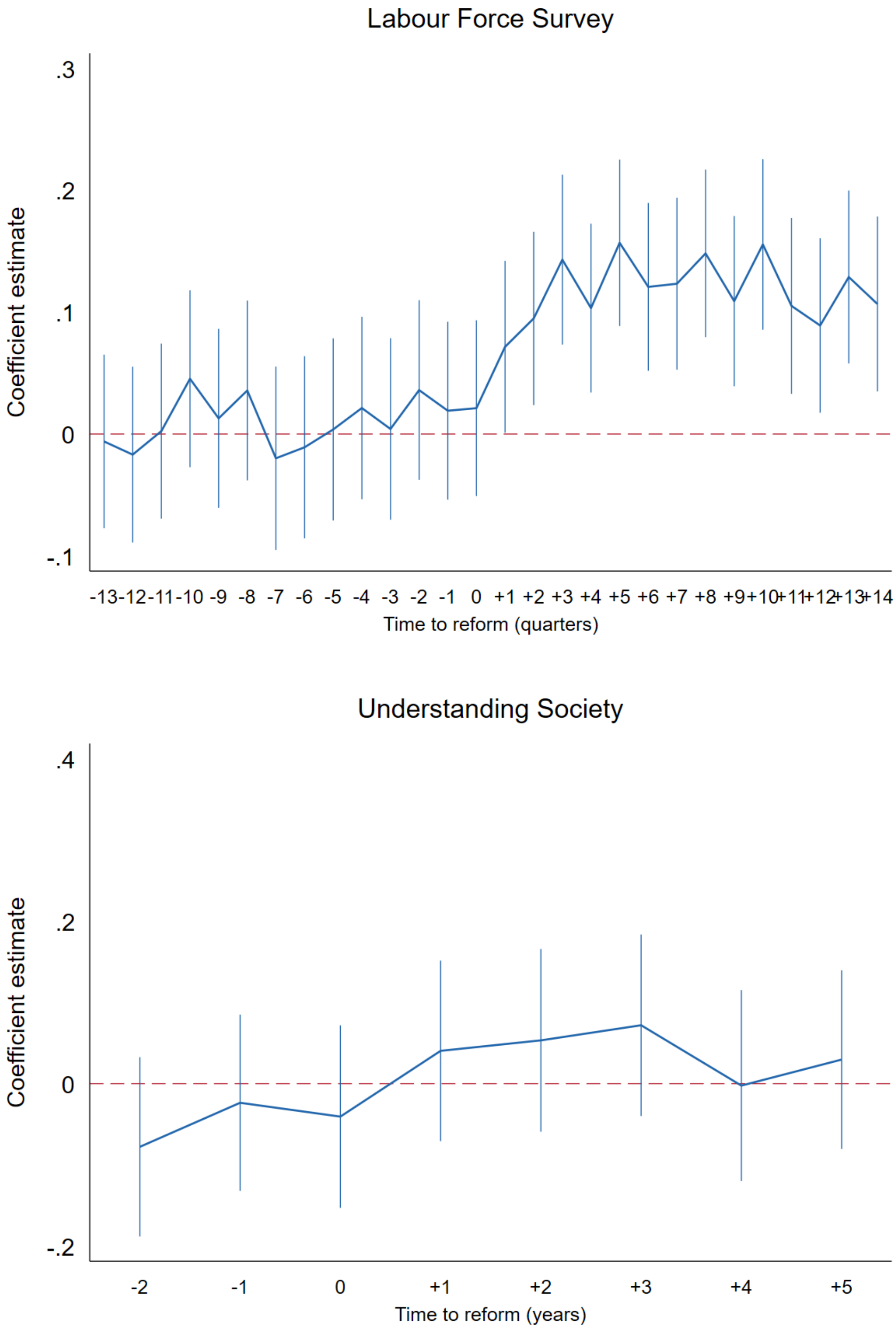


Table B.10 - Difference-in-difference estimates (OLS) on pregnancy rates

	(1)	(2)	(3)
Dependent variable: Pregnancy rates			
A) Understanding Society			
Reform x Treated	0.000	-0.001	-0.004
	(0.015)	(0.015)	(0.022)
Treated	-0.002	-0.031**	-0.033*
	(0.013)	(0.013)	(0.018)
N	6,366	6,250	6,250
R-squared	0.005	0.056	0.056
Year & region fixed effects	Yes	Yes	Yes
Personal characteristics	No	Yes	Yes
Employment & inactivity rates	No	Yes	Yes
Differential time trend	No	No	Yes

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Notes: In the low education sample, the control group is restricted to married mothers of 5-6 year olds with educational qualifications at GCSE level or below.

4 Chapter 4

Lockdown, COVID Severity and Mental Health

I am grateful to the Understanding Society team for making available the Special Licence dataset which includes the Local Authority District geographical identifiers. I also thank the UK Data Service for making the data available.

4.1 Introduction

Previous literature has documented the significant and negative impact of the COVID-19 pandemic on mental health – effects that are highly heterogenous across the population. In the UK, mental health decreases have been coupled with having one of the highest COVID death rates in Europe. At its peak, in January 2021, there were 1,325 daily COVID deaths. To control the virus the government imposed two national lockdowns, less than a year apart, and implemented various public health measures. Whilst the effectiveness of these responses can be evaluated through falling death rates, case rates, and hospital admissions, mental health outcomes remain, to some extent, invisible.

The contribution of this paper is to discern how mental health evolves over the course of the pandemic in the UK (between April 2020 and January 2021) and, more specifically, the relationship between mental health, the threat of the virus and the impact of policy responses to it. It is a novel addition to existing literature, which so far only addresses the overall pandemic effect on mental health (see Banks & Xu (2021) for a summary). For instance, in these studies, there is no attempt to control for the severity of COVID, nor the imposition of lockdown, and mental health trajectories during the pandemic have not been studied in detail. Moreover, many of these studies compare a point in time before COVID, with a point in time during COVID, usually when we were in lockdown and had high COVID case and death rates. In contrast, this chapter, by exploiting variation in both the severity of COVID and lockdowns, can unpick the importance of each on an individual's mental health. Additionally, the data allows us to use geographic variation in the severity of COVID to examine whether COVID severity in an individual's local area is better predictive of mental health than national severity.

For this, the Special Licence version of Understanding Society COVID-19 survey is used, which is a large longitudinal dataset, collected from the beginning of the pandemic in April 2020, up to January 2021. It contains the Local Authority District (LA) each respondent resides in which I link to official government data on COVID severity. Severity is measured in two separate ways – weekly case rates and weekly death rates per 100,000 people. For each measure, there are two parameters, one which captures the local case/death rates and another for the national case/death rates.

Following the approach in Banks and Xu (2020) I examine three separate mental health outcomes, all measured through responses to the General Health Questionnaire (GHQ-12): the overall GHQ-12 well-being score, number of mental health problems, and an indicator for having any severe mental health problems. I exploit the panel nature of the data using fixed effects methods which, by accounting for unobserved heterogeneity, improve upon random effects estimation and a pooled Ordinary Least Squares regression.

The findings suggest that being in lockdown has a large and significant effect on mental health, worsening mental well-being on the overall GHQ-12 score of between 0.7-0.85 points, which is equivalent to a 5.6-6.7 percent increase from the mean. This result holds for the measures of more serious mental health problems. It is associated with an additional half a problem when counting an individuals number of reported mental health problems (GHQ-12 components), equivalent to a 20 percent increase from the mean. Lockdown also increases the probability of having a severe mental health problem by 7 percentage points, which is a particularly large effect, corresponding to a 40 percent increase in severe mental health problems.

The effect of COVID severity is more complex. The estimates imply that local case rates and local death rates do not predict any of the three mental health outcomes. However, national COVID severity measures are found to have an effect, which may be due to the greater visibility of national case rates/death rates and their correlation with the introduction of national policy responses e.g. travel restrictions and public health measures. The effect of national rates is only present for the two outcomes which capture more serious mental health problems.

The findings are important, not only for evaluating policy decisions, but also for allowing policymakers to predict the potential ramifications of their decisions by better understanding how much mental health would decline given a lockdown or a particular number of cases or deaths. For example, the results imply an increase in national deaths by 20 per 100,000 (the peak number of UK deaths) leads to, on average, a third of an additional mental health problem and an increase in the incidence of severe mental health problems by 10 percentage points at the individual level. This compares to the respective lockdown effects noted above of half a problem and 7 percentage point increase in the incidence of severe mental health problems.

I draw noteworthy findings from a heterogeneity analysis, which generally support existing evidence on COVID and widening mental health inequalities. When examining the effect of lockdown on various sub-groups, the results indicate women, lone mothers and younger age groups are the groups who suffer the most mentally. In terms of the effect of COVID severity on mental health among these sub-groups, the results are more mixed.

Poorer mental health may have significant consequences, not only for the individual, but also for the economy through, for example, productivity losses and increasing demand for services that treat symptoms. The individual consequences of poor mental health may have a long-term impact over the course of an individual's life. Moreover, the effects of poor mental health may create spillovers into wider society; some evidence has even related these spillovers to the COVID-19 pandemic. For example, Krekel et al. (2020) find that past and present happiness is associated with an increase the overall compliance with public health measures.

This chapter is structured as follows: Section 4.2 reviews the recent literature, Section 4.3 describes the data used, Section 4.4 describes the methodology, Section 4.5 contains the regression results, Section 4.6 contains the results of a robustness check and Section 4.7 concludes.

4.2 Existing evidence

Some researchers were quick to generate survey questionnaires in response to COVID-19, but none provided such real-time information as the Google Trends data. Analysing changes in Google searches over time, Brodeur et al. (2021) use difference-in-difference methodology to show that in the first month of lockdown there were increases in search terms for boredom, sadness and loneliness, compared to the same day in the previous year. Other research has also used Google data to this effect: Knipe et al. (2020) find falls in searches for suicide, though Stefan Foa et al. (2020) and Jacobson et al. (2020) find most of the rise in negative search terms took place before the start of the lockdown and in fact fell during the lockdown.

In a similar vein, Klotzbücher & Armbruster (2020) use high frequency data of the number of calls to a counselling service in Germany. They find in the first week of lockdown calls increased by 20 percent. This was driven by increases in physical and mental health concerns, more specifically loneliness, anxiety, and suicide ideation. In contrast, there was no significant increase in calls relating to economic issues.

Whilst giving valuable and speedy insights such high frequency data is limited by self-selection whereby certain types of people are more likely to call helplines or search on Google.

Additionally, the search terms which are used as outcome variables are chosen for convenience and are not necessarily related to the commonly used measures of mental health.

Current evidence on the effects of the pandemic generally rely on one of two types of survey data - those where data is collected both before and during the pandemic, and those in which the sample is first interviewed once the pandemic began. The former allows a comparison of before and after and thus the overall impact of the pandemic itself, whilst the latter can only say something about mental health trajectories over the course of the pandemic.

4.2.1 Comparing mental health before and during COVID

The surveys that collect mental health before and during the pandemic are useful to facilitate a before-after comparison of mental health outcomes. Both cross-sectional and longitudinal surveys have been used to this effect. The ONS Opinions and Lifestyle is a cross-sectional survey which collected data prior to the pandemic but adapted to weekly sampling period from March 2020, to understand more about life during the pandemic. It has a sample of around 4000-4500 adult responses each week, and the ONS produce a weekly bulletin. Their descriptive analysis (Office for National Statistics (2021)) shows that during the first couple of weeks after the lockdown happiness fell sharply, and anxiety rose. However, over the long-term, all four measures of mental well-being - life satisfaction, feeling worthwhile, happiness, and anxiety improved. Despite this improvement, they have not yet returned to pre-pandemic averages. What is particularly interesting is that their graphical analyses show an association between lockdown and worse self-reported mental health, though the causal direction is unknown.

Whilst this cross-sectional data is useful, longitudinal data sources have the advantage of being able to analyse within-person changes in mental health. Existing participants in the Understanding Society study were asked to complete a succession of short surveys. The content of the survey encompasses the changing impact of COVID on participants. It began in April, one month after the first lockdown was introduced on March 23rd, and has continued, first monthly (Apr-Jul), and later bi-monthly (Sep-Jan). A benefit to researchers is that it is possible to obtain individual and household characteristics measured pre-pandemic since responses to these surveys could be linked to the main waves of Understanding Society. Using these pre-pandemic waves, Banks & Xu (2020) estimate an individual's "counterfactual" mental health for April 2020 had the pandemic not occurred and calculate the difference between the estimated and actual reported mental health. They find mental health worsened by 8.1 percent over the first two months of lockdown. The effect was heterogeneous across the population. Ethnic minorities, 16-24 year olds, females and those with pre-existing mental health issues were some of the groups more negatively affected. Banks et al. (2021) update their analysis using the September wave of Understanding Society COVID survey. They find that although mental health improved, relative to April 2020, it was still 0.3 points on the GHQ-12 scale, equivalent to 2.2 percent below the estimated 'not in pandemic' counterfactual. The authors conclude: "there is still much uncertainty surrounding the pandemic's second and third waves and how the associated lockdowns of economics and social activities will affect mental health…".

Using the same dataset, and a similar methodology, Pierce et al. (2020) find a significant worsening of mental health in April by 0.5 points on the overall 36-point GHQ-12 scale, which was more acute in younger adults and women. By comparison the effect size found by Banks and Xu (2020) was 1 point.[26] Etheridge & Spantig (2020) document the gender differences in mental health and find that social factors are an important mechanism. Examining a longer time horizon (until June 2020), Daly et al. (2020) find the negative relationship between the pandemic and mental health persists.

Anaya et al. (2021) attempt to obtain causal estimates of the impact of lockdown and find lockdown results in a 0.76 point decline on the GHQ-12 scale. Their difference-in-difference strategy relies on comparing mental health outcomes pre and post March 2020 with a comparison group of those interviewed pre and post the same date in 2019. This controls for timing, but for nothing else that might confound their estimates. In particular, like Banks & Xu (2020), they cannot unravel lockdown from cases.

4.2.2 Analysing mental health changes during COVID

The UCL Social Survey is a panel dataset which commenced on the 21st March as the pandemic took hold. The data is mainly used in a descriptive analysis, whose results are updated in regular reports. In the 35th report Fancourt et al. (2021) examine the first 64 weeks since the lockdown was announced on the 23rd March 2020. They document a steady improvement in mental health since the beginning of lockdown, which continues to fall during summer when the lockdown

[26] This larger estimate stems from differences in the methodologies of the two papers. While Pierce et al. (2020) is similar in that they estimate counterfactual mental health using fixed effects methods, they do not control for seasonal trends and their modelling approach is different.

is eased. New restrictions in September led to a slight worsening of mental health. The authors report no increase in suicide ideation or self-harm. They also show trajectories of the main factors causing respondents stress, life satisfaction, abuse, loneliness, happiness, and other worries over the course of the pandemic. In addition, they provide a detailed breakdown by nation, age, gender, ethnicity, whether have children, partnership status, key worker status, household income and physical health diagnosis. Looking at the happiness variable, women display lower happiness than men throughout the pandemic, as do ethnic minority groups, young adults, people on lower household income and those with a physical health diagnosis. Individuals who have children are less happy than people without children, though this difference dissipated over time, and coincided with schools re-opening. O'Connor et al. (2021) collect their own survey during the first six weeks of lockdown and find that suicide ideation increased, though symptoms of anxiety, entrapment and levels of defeat decreased.

Existing research discussed has focused on the overall effect of the pandemic on worsening mental health but cannot estimate the extent to which the worsening of mental health emanates from being in lockdown or being in a pandemic. The association between lockdown and mental health is ambiguous. Lockdowns improve mental health as forward-looking people respond positively to government action to reduce the threat of transmission. On the other hand, the constraints on activities during a lockdown may have detrimental mental health consequences.

Some studies, which specifically examine the COVID-19 pandemic, exploit variation in the timing of lockdowns across US states/regions to identify the impact of lockdown on mental health. Adams-Prassl et al. (2020) use two survey waves collected in March and April 2020 and find the mental health scores (using the WHO-5 measure[27]) for individuals in US states in lockdown were 0.085 standard deviations lower than individuals in states that had not imposed a lockdown. Interestingly, the effect is entirely driven by women. They can unravel the health threat from the lockdown effect by controlling for county-level COVID cases and deaths. The addition of these controls does not change the coefficient on lockdown, which they take to mean that the effect of lockdown cannot be explained through higher COVID case and

[27] The WHO-5 is a self-reported measure of mental well-being consisting of five statements "I have felt cheerful and in good spirits", "I have felt calm and relaxed", "I have felt active and vigorous", "I woke up feeling fresh and rested" and "My daily life has been filled with things that interest me". The respondent can answer on a 6-point scale (all of the time, most of the time, more than half of the time, less than half of the time, some of the time, at no time).

death rates as well through financial worries and childcare responsibilities. This methodology would be difficult to implement in the UK context because there was little variation in lockdown timing and geography, and all nations of the UK almost always announced similar changes to restrictions at similar times.[28]

The literature is not constrained to the UK, researchers around the world have documented the relationship between the global pandemic and mental health. Across Europe, in France and Italy there was an increase in anxiety symptoms ((Ramiz et al. (2021), (Castellini et al., 2021)). Using longitudinal data researchers in Spain found depression increased during the first lockdown, with no change in anxiety (González-Sanguino et al. (2021)). In Germany, using high frequency panel data Schmidtke et al. (2021) find that both first and second waves of the pandemic significantly reduce workers mental health. Further afield, research from Australia, Singapore, and China has shown an increase in psychological distress associated with the pandemic ((Newby et al., 2020), (Cheng et al., 2021) and (Ahmed et al., 2020)). However not all studies agree. Researchers in the Netherlands find a slight decrease in anxiety and depression during the first lockdown (van der Velden et al. (2021)). Wang et al. (2020) also found significant reductions in mental distress during the first 4 weeks of the pandemic in China. Studies outside the UK are important but may differ in terms of the policy and cultural context and the intensity of exposure to the pandemic.

The existing evidence is clear: the pandemic has had a significant negative impact on mental health. But so far, none of the UK research papers have decomposed the incidence of mental health problems by COVID severity and lockdown, or controlled for weather variation. This chapter aims to fill this gap in the literature. I use a special licence version of the Understanding Society COVID survey which contains detailed geographical identifiers. The contribution is twofold. Firstly, I investigate the impact of COVID case rates and death rates on mental health, a facet that has been so far neglected in the literature. Secondly, I examine whether COVID severity (measured by cases and deaths) in an individual's local area (local authority district) matters more for mental health than national COVID severity.

I analyse the importance of lockdown, local COVID severity, and national COVID severity on an individual's mental health, all of which are measured during the pandemic –

[28] A tiered system was implemented in late October before the UK moved into a second national lockdown, but the duration was only around 4 weeks. There is little data within that time frame which is both geographically granular enough and has measures of mental health to facilitate the analysis mentioned above. Secure access Annual Population Survey (APS) is an exception, and current data is available up to the end of 2020.

between April 2020 and January 2021. I estimate individual fixed effects models which account for any unobserved heterogeneity between individuals and, in turn, may help to eliminate differences across local authorities. In addition, I include controls for the weather which should circumvent any concerns of seasonality of mental health. The scope of the paper is not the impact of COVID itself on mental health (i.e. a before-after comparison), but focuses instead on within-pandemic changes – due to the rising case and death rates as well as national lockdowns.

Understanding how the case rate, death rate and lockdown, each affect mental health, should be important to policy makers. Investigating to what extent lockdowns are damaging to mental health may be useful in the future, particularly when mental health has been highlighted as a reason to limit the duration of lockdowns by anti-lockdown campaigners.

4.3 Data

I draw on data from several different sources. First, survey data is used to obtain information relating to individual mental health. Second, I utilise publicly available official government data on LA COVID cases, LA COVID deaths and UK-wide testing capacity. Finally, I link the LA-level COVID data to the survey data based on each individual's geographical identifier and date of their interview.

4.3.1 Understanding Society

Whilst cross-sectional data can be informative, panel data is necessary to study within-person changes in mental health. This facilitates a fixed effect model as the basis of the identification strategy. Therefore, the individual level data used here comes from the longitudinal COVID-19 Understanding Society (UKHLS) survey. The survey began in April 2020 with around 15,000 individual responses. I use 7 waves of data, which were collected throughout 2020 in April, May, June, July, September, November, and in January 2021. Understanding Society is a household survey and all household members over the age of 16 were invited to take part. Respondents were sent an online survey around the last week of the month. This means that there is little variation in timing of responses within waves, so this limits the scope of examining variation within a survey wave.

From the sample of respondents in Wave 9 of the main Understanding Society survey, the response rate to wave 1 of the COVID-19 survey was 48.6 percent.[29] In raw numbers it corresponds to 17,007 respondents, falling to 15,360 after dropping individuals with missing information. The pooled sample is 89,260 across all seven waves of data, after keeping only those individuals with non-missing GHQ scores. There are 8,585 individuals who are observed in every wave, while 2,366 are present only in the first wave of the survey.

As well as collecting demographic information on age, gender and ethnicity, the COVID-19 survey asks respondents about their economic circumstances. Some of these variables are pandemic specific, such as being put on furlough and keyworker status, others are questions that are also asked in the pre-COVID main surveys e.g. employment status and partnership status.[30]

There are also questions relating to the health impact of the virus, such as whether the respondent has had COVID symptoms, has been tested for COVID, and a measure of clinical vulnerability to the virus which is calculated by the survey team. The COVID survey has been linked to the most recent pre-COVID wave of Understanding Society. The most recent wave depended on the stage of fieldwork and could be either refer to wave 10 (2018-2020) or 11 (2019-2021). I use this linkage to retrieve information on individuals most recent pre-pandemic self-reported mental health and highest educational attainment, which are used as controls in the pooled OLS analysis.[31]

The data contains the Local Authority District (LA) that an individual resides in. There are 366 LAs represented in the data, and the per wave average number of observations for each is 55. The average population size is 177,298, the largest being Birmingham with a population of 1,141,816, while the Orkney Islands is the smallest with just 22,270 people.

I use three separate measures of mental health, all of which are taken from the well-known 12 item self-reported General Health Questionnaire (GHQ-12), asked in every wave of the COVID survey. The GHQ was developed as a screening tool of psychological symptoms in a non-clinical environment (Goldberg & Williams (1988)). There are variations of the scale, but the shortened 12-item version includes six positive and six negative questions. The GHQ-

[29] This is a comparable response rate to other large voluntary government surveys in the UK.
[30] The definition of furlough changed throughout the pandemic and the survey does not facilitate being able to look at changes in furlough due to the question routing. Thus, the variable I construct essentially measures whether the individual was ever furloughed in the survey time period.
[31] These controls drop out of the fixed effects analysis.

12 has been well validated and used in a wide range of contexts (D. P. Goldberg et al. (1997), (Graetz, 1991)) and consequently is a commonly used measure of well-being among economists.[32]

The first measure I define as the overall GHQ-12 score. This assigns a value to the response to each of the 12 components from 0 to 3 (see Appendix C for list of questions and response items). Thus, it ranges from 0 to 36, and can be thought as an indicator of overall well-being. A higher number on the scale indicates greater mental distress. It is also beneficial to capture the incidence of more severe mental health issues. First, I use a measure known as "number of problems" which assesses each component and counts the number of symptoms where the response is "worse than usual" or "much worse than usual". This is reported on a scale of 0 to 12 where a higher number implies more problems. I also construct a binary indicator for having even more severe problems, following the approach in Banks and Xu (2020). The variable equals one if the individual responds "much worse than usual" in any of the 12 components that make up the GHQ-12, and otherwise equals zero.

Figure 4.1 shows how average mental health has changed during the pandemic for all three outcomes. On average, females display poorer mental health than males, in particular young women. For all ages, average mental health improves in July, this corresponds with the lifting of lockdown and the announcement of the "Eat Out To Help Out" scheme that encouraged spending in restaurants, pubs, bars, and cafes[33]. Banks and Xu (2020) show that across all nine main waves of Understanding Society there is an improvement in average mental health over the spring and summer months. This could have implications for the analysis, as the lifting of lockdown is co-linear with summer months (July, Sept), when mental health typically improves. Is this mental health improvement attributable to the lockdown, or could it be due to the seasonality of mental health? One way to address this is to control for the weather. I therefore merge in daily rainfall and temperature data from the Met Office (2021). This weather data is measured and collected by weather stations located all over the UK. To approximate the weather in the LA of interest, the distance between the weather station and the centroid of the LA is calculated and weather measurements from the closest weather station

[32] Researchers have used the GHQ-12 to study, for example, the relationship between psychological distress and income inequality (Wildman, 2003), debt (Brown et al., 2005), medium-sized lottery wins (Gardner & Oswald, 2007), employment (García-Gómez et al., 2010) (Lagomarsino & Spiganti, 2020), education (Cornaglia et al., 2015), and crime (Dustmann & Fasani, 2016) .

[33] (Fetzer, 2020) finds the scheme significantly accelerates the spread of COVID-19.

are used. The rainfall and temperature measurements correspond to the day of the interview. Figure 4.1 also shows that by the end of September mental health began to worsen again, more so for young adults. This was at a time of rising cases and deaths. And by the end of November, when the new lockdown was announced, mental health had returned to the levels reported at the start of the pandemic in April.

Summary statistics of the key variables are presented in Table 4.1. It is clear that they pose a concern for the representativeness of the sample. In the unweighted sample, only 41 percent of respondents are male, and the education level is higher than average population levels (48 percent have a degree or equivalent). In light of this, the COVID-19 survey longitudinal weights are used to account for both unequal selection and non-response. Examining the weighted summary statistics they look, reassuringly, much closer to the population. They are also more in line with summary statistics from the most recent Wave 9 of the main Understanding Society. However, if there is selection into the COVID-19 survey which depends on previous self-reported mental health this will still be problematic for the estimates. Reassuringly, Daly et al. (2020) find that this is not the case for the 2017-2019 waves. It also worthy of note that, because some people are assigned a zero weight, there is a considerable drop in the sample size when using the weighted data, although the sample remains sufficiently large for our purposes. [34]

[34] The survey data team assign a zero longitudinal weight if the respondent leaves the survey and returns in a later wave. A detailed explanation of how the weights are constructed can be found in the Understanding Society User Guide.

Figure 4.1 - Mental health (Overall GHQ-12 score, number of problems and having a severe problem) across all waves of Understanding Society COVID survey

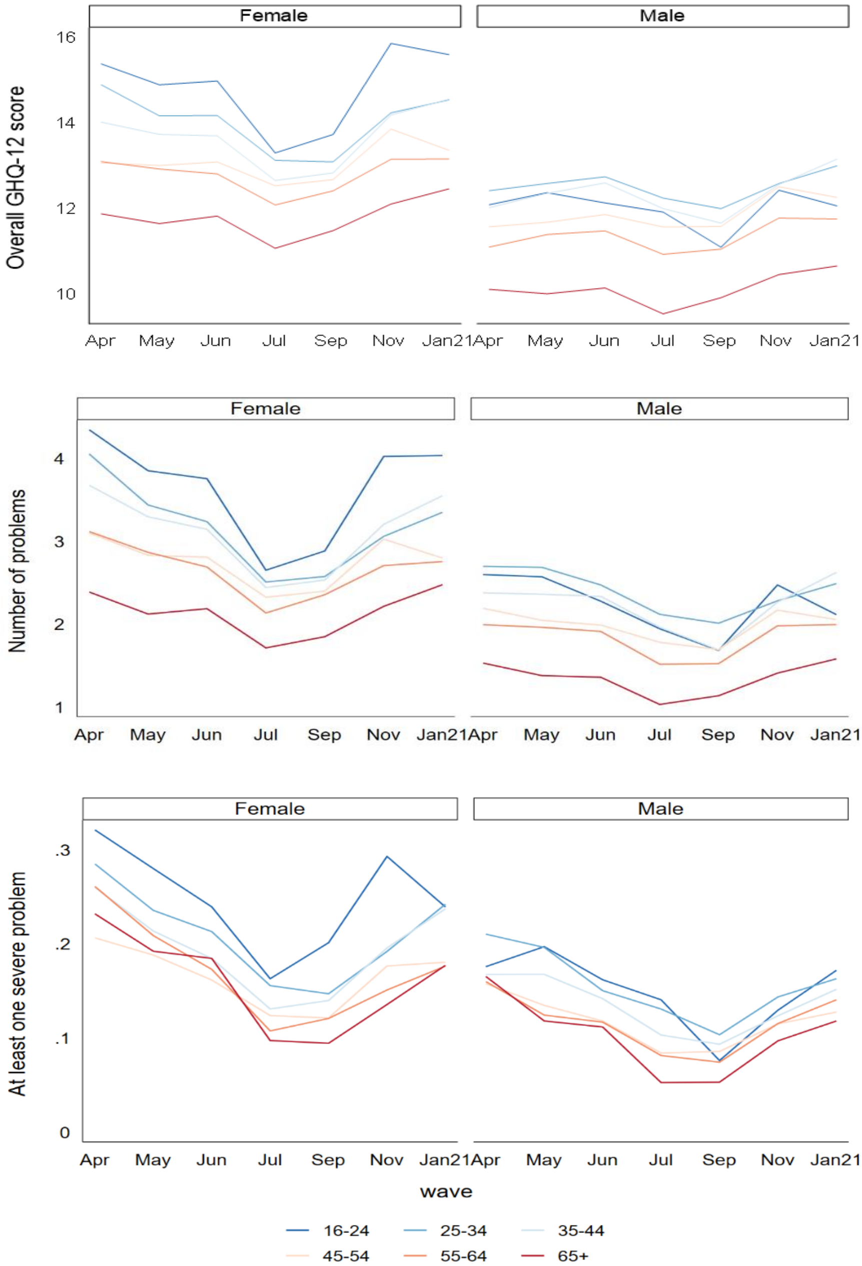


Table 4.1 – Summary Statistics

	Unweighted			Weighted		
	Mean	Std. Dev	N	Mean	Std. Dev	N
Age	53.11	(16.58)	93161	48.80	(17.98)	64068
Male	0.41	(0.49)	93152	0.48	(0.50)	64068
Non-white	0.11	(0.31)	90814	0.09	(0.28)	63855
In couple	0.70	(0.46)	93160	0.61	(0.49)	64067
Number of kids	0.38	(0.79)	88277	0.40	(0.85)	64028
Not clinically vulnerable	0.59	(0.49)	92889	0.61	(0.49)	63930
Clinically vulnerable	0.36	(0.48)	92889	0.34	(0.47)	63930
Clinically extremely vulnerable	0.05	(0.22)	92889	0.05	(0.21)	63930
No qualifications	0.16	(0.36)	82497	0.17	(0.38)	55571
GCSE or equivalent	0.25	(0.43)	82497	0.30	(0.46)	55571
A-Level or equivalent	0.12	(0.32)	82497	0.13	(0.34)	55571
Degree or equivalent	0.48	(0.50)	82497	0.40	(0.49)	55571
Has job	0.59	(0.49)	91805	0.61	(0.49)	63814
Key worker	0.33	(0.47)	91841	0.34	(0.47)	63852
Furlough	0.10	(0.30)	93161	0.14	(0.35)	64068
Had COVID test	0.09	(0.29)	93122	0.082	(0.27)	64044
Had COVID symptoms	0.06	(0.23)	93115	0.05	(0.22)	64038
Average overall GHQ-12 score	12.28	(5.94)	89260	12.74	(6.24)	63055
Average number of problems	2.37	(3.34)	89260	2.56	(3.52)	63055
One severe problem	0.16	(0.36)	93161	0.17	(0.38)	64068

4.3.2 Official COVID data

To examine the impact of COVID on mental health, it is necessary to link the official COVID data to the survey data. There are several different measures of COVID severity that have been collected, each with advantages and disadvantages, and this section discusses them.

A logical starting point is to look at the most widely reported measures, case rates and death rates, which people could easily follow in the media. Case rates were timely; as soon as cases were confirmed they were published shortly afterwards. In the official government statistics, a case was measured as a positive COVID-19 virus test (either lab reported or rapid lateral flow test) and is allocated to the person's area of residence.[35]

However, testing capacity at the beginning of the pandemic was limited. It will be problematic for our estimates if certain LAs were more likely to have tests than other LAs, which may inflate the case rate in those LAs. It may also be an issue when estimating a fixed effects regression that looks at the within-changes to an individual's mental health throughout

[35] If a person has more than one positive test, they are only counted as one case.

the pandemic. Unfortunately, while the government releases data on daily testing capacity, it is only at the UK level. Still useful, this is included as a control variable in the main analysis and later used in a robustness check.

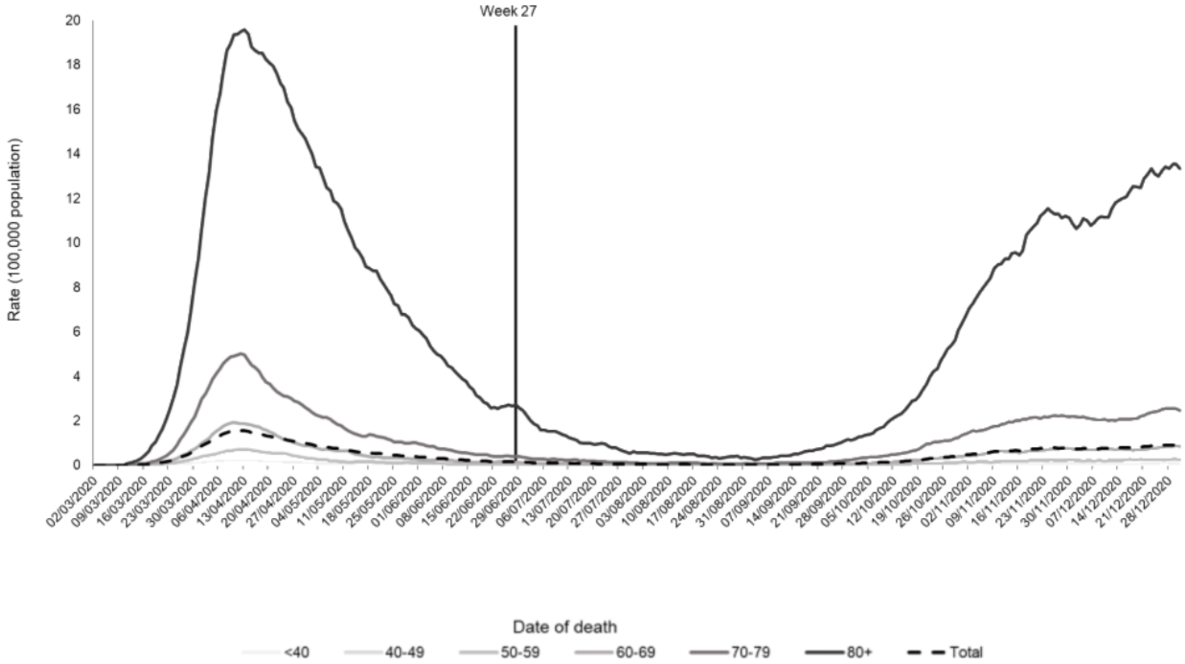
Whilst case rates are dependent on testing capacity, the measurement of deaths has remained the same throughout – making it a more consistent measure. There are two measures of deaths in the UK. The first is collected by the Department of Health and Social Care, which measures the number of deaths in the UK that occur within 28 days of a positive test. These gov.uk figures are timely but do not include deaths of people not tested or where the test was negative, making it a more simplistic measure of deaths. On the other hand, this was the information people had to hand and, although not an accurate measure of deaths, it was the most widely reported and, arguably, more salient to mental health.

The second measure of death rates comes from ONS data and is based on death certificates. Every death has an ‘underlying’ cause recorded and any ‘contributory’ causes – together these are called ‘mentions’. For some health conditions the ONS uses the ‘underlying’ cause to record deaths. COVID-19, like other coronavirus strains such as influenza, can directly or indirectly cause deaths and, therefore, the ONS uses ‘mentions’, whether underlying or contributing, to define COVID-19 deaths. Whilst a more accurate measure, it was not the one reported on the government dashboard or in government briefings. There was also a considerable time delay in these cases being reported and, for these reasons, the former measure is preferred.

A limitation of using the death rates may arise from the correlation between LA characteristics and the death rate. An area with a much larger population of elderly people, high levels of deprivation, worse health infrastructure or lower health expenditures, may show a higher death rate whilst having a similar case rate to a more affluent LA. In this chapter, I assume that the COVID severity right-hand side variable is exogenous. Using the death rate would call this into question as more deprived areas may be correlated with worse mental health. Using a fixed effects model eliminates this concern as long as LA characteristics remain constant of time – which would be reasonable to assume over such a short period. It remains an important question as to how LA characteristics affect the COVID case and death rate, but not one which is examined within the scope of this paper as LA identifiers are dropped from the individual fixed effects regression.

All age groups are at risk of catching COVID, however, there may be a rationale for using age-specific COVID mortality rates as older individuals are considerably more at risk of serious health complications and death. Figure 4.2 shows the strong age gradient in COVID-19 mortality and individuals may be more concerned with age-specific death rates than overall case rates or death rates. This data is not used in this chapter for two reasons. Firstly, it is not available at the local level, and secondly any differential age effects should be accounted for by controlling for age and can be further unpicked in the heterogeneity analysis.

Figure 4.2 – Seven-day rolling average mortality rates in laboratory-confirmed cases of COVID-19, by age group. Source: COVID-19 confirmed deaths in England (to 31 December 2020): report (ONS)



I downloaded the raw numbers of COVID cases and deaths using the gov.uk dashboard.[36] Cases and deaths are measured in two ways, both available from the dashboard. The number of cases can either be measured by the date of the test specimen or the date published. Deaths are either measured by the date on the death certificate or the date published. I use the specimen date and date of death as they are not as heavily impacted by reporting errors and corrections - which the gov.uk data description is transparent about. The data are updated daily and are reported for each LA in the respective nations and for the UK as a whole. Daily cases are small and give rise to a lot of variation (e.g. a weekend effect) and potential reporting errors. Therefore, I sum the cases and deaths by week of the year, to get weekly case and deaths.

[36] The dashboard can be found at: https://coronavirus.data.gov.uk/

These raw numbers are then converted into weekly rates per 100,000 people by using the ONS population estimates of each LA for 2019.

The average number of weekly national cases across the survey period was 75 cases per 100,000 people. During lockdown, the average case rate was 85 per 100,000 and falls to 50 cases when not in lockdown. National death rates were much lower, an overall weekly average of 11 deaths per 100,000, 15 per 100,000 during periods of lockdown, and less than one death per 100,000 when not in lockdown. However, reporting the national average conceals the considerable geographic variation in COVID severity. Figures 4.3 and 4.4 visualise the variations in COVID severity over the course of the pandemic. Though I have COVID data for each local authority, the figures show the aggregated **regional** case and death rates so as not to overcrowd the graph.[37] Figure 4.3 shows in some regions case rates were considerably lower than in other more badly affected regions. For example, just as the new lockdown was announced in January 2021, cases peaked at almost 900 per 100,000 people in London, in sharp contrast to a case rate of only 220 per 100,000 in Scotland. A similar comparison can be made with death rates. There were two peaks in the death rate, in April and November, which can be seen in Figure 4.4. In early 2020 the South-East and East of England saw the highest death rates in November (around 20 per 100,000). As cases and deaths fell towards zero, the lockdown restrictions eased, and there was then little regional variation in case and death rates.

A comparison of Figures 4.3 and 4.4 suggests that testing capacity was influential in the low case rates observed in early 2020. Death rates peak at similar levels in April and November while there was a spike in case rates in the later stages of the pandemic (when testing capacity had been significantly increased). Whilst correlated, a high case rate does not necessarily imply a high death rate, because death rates can depend on other factors such as local demographics or hospital capacity. As an example, London had the highest case rate in the first week of January but by no means the largest death rate - this is attributed to the West Midlands and Wales. In summary, I use both case rate and death rate measures, keeping in mind the drawbacks of each measure. In turn, any differences could reflect that each measure may be telling us different things. People may pay more attention to, or are more concerned

[37] For any readers unfamiliar with UK geographies - regions are a larger aggregation of which there are 13 and each of the 366 local authorities map just to one of the regions.

with, death rates than with case rates or vice versa, though there will be some degree of correlation between measured cases and later deaths.

Figures 4.5 and 4.6 illustrate weekly LA cases and deaths per 100,000 people. The numbers refer to weekly cases and deaths in the final week of the month, which corresponds to the week in which respondents were interviewed in the COVID survey. Areas with higher case rates are in regions with higher case rates in general. In addition, Scotland appears to be relatively less badly affected across all time periods. The figures also document how cases and deaths are at odds in the early stages of lockdown due to the testing capacity and the improvements across the majority of LAs in the summer months, though there is some variation in COVID cases and deaths such as higher case rates in the North East in September.

Figure 4.3 – Regional weekly cases (cases per 100,000 people)

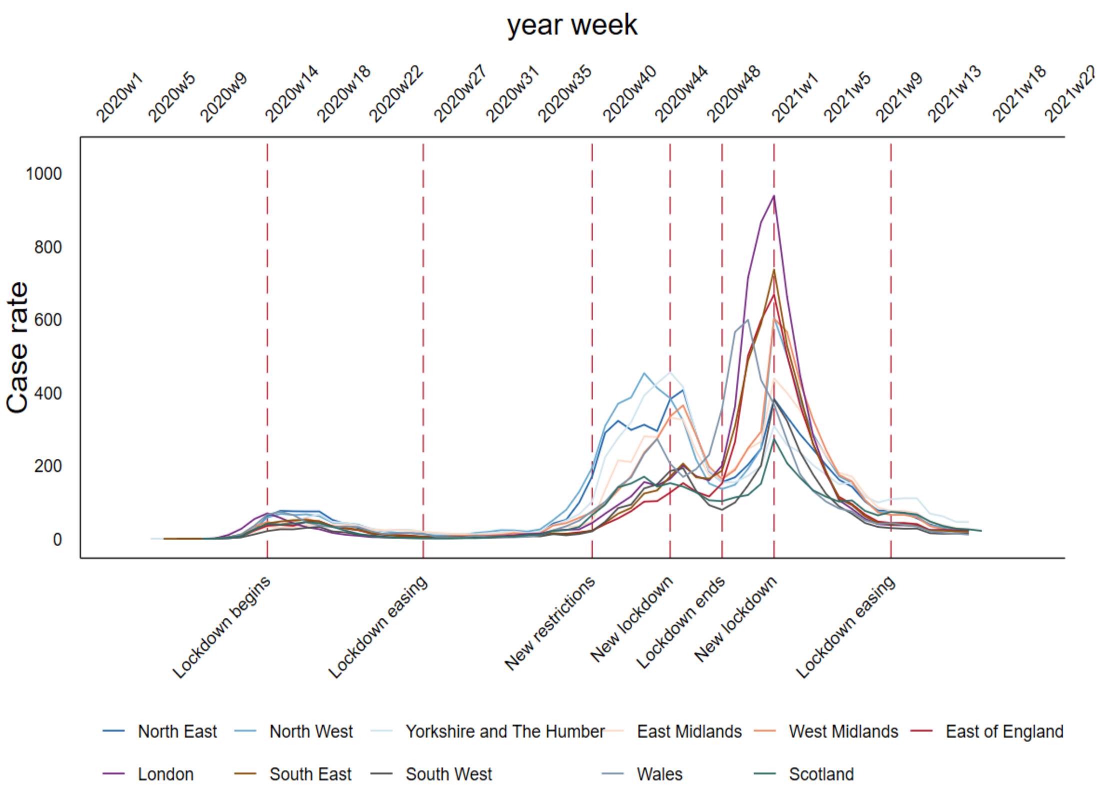


Figure 4.4 – Regional weekly deaths (deaths per 100,000 people)

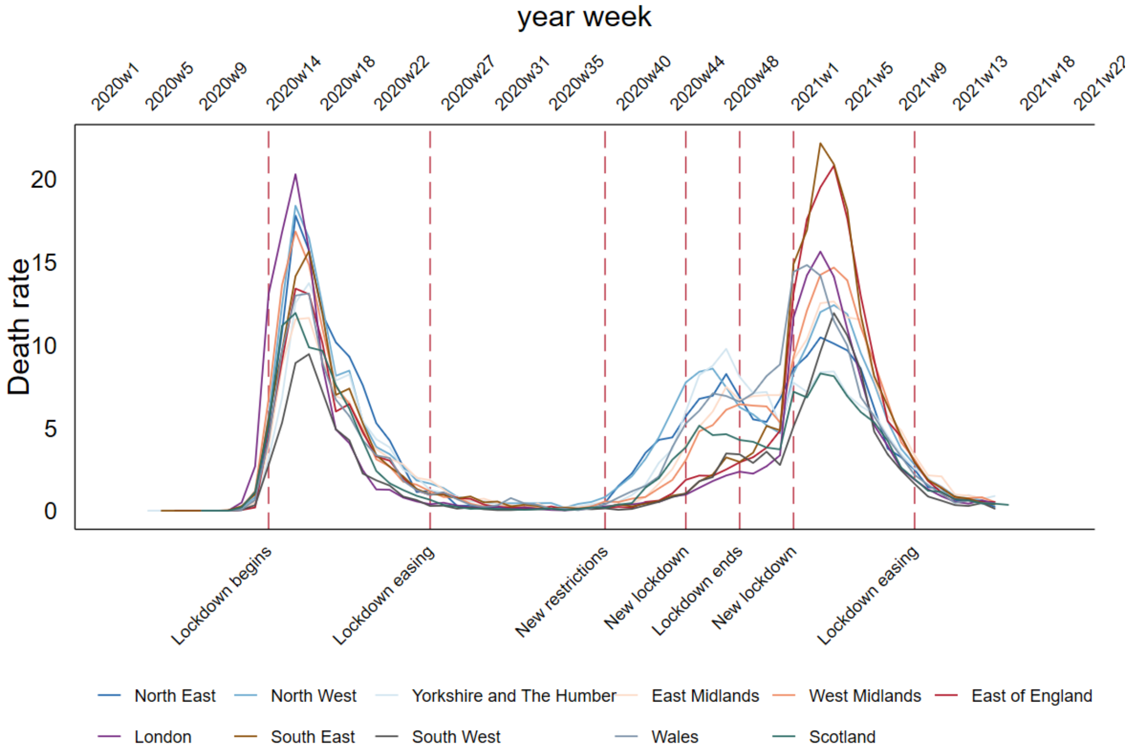


Figure 4.5 – Geographical representation of weekly COVID case rates

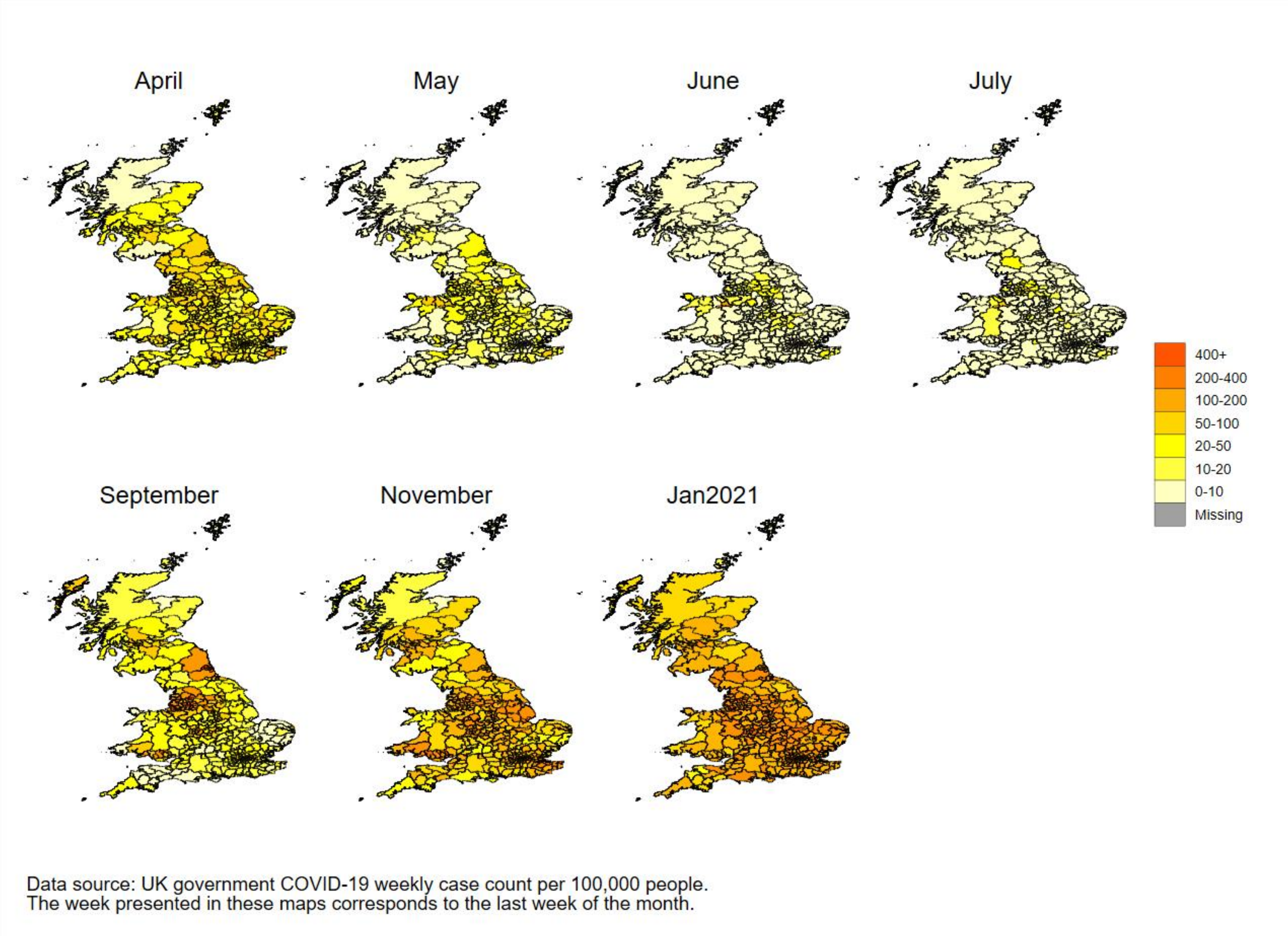
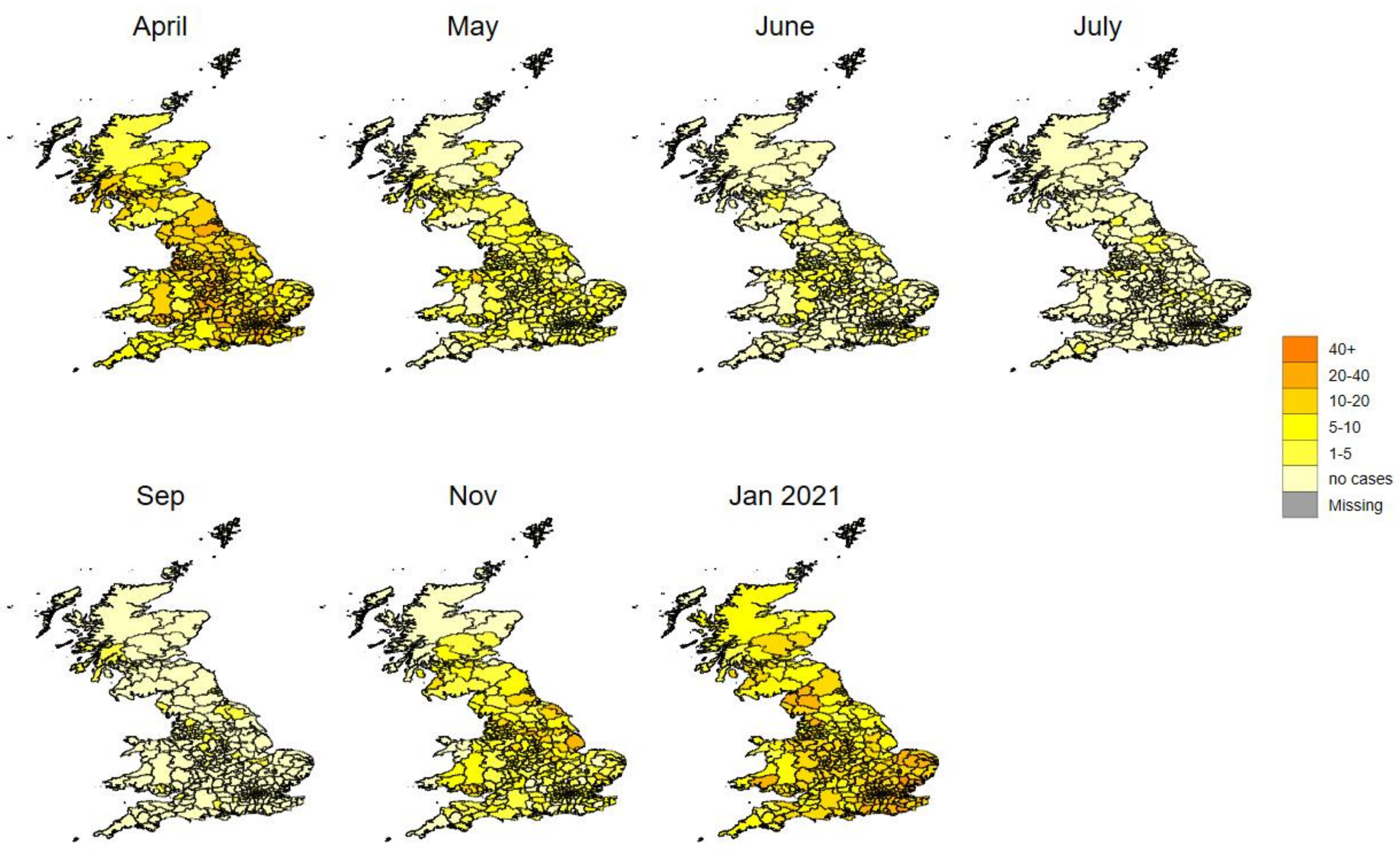


Figure 4.6 - Geographical representation of weekly COVID death rates



Data source: UK government COVID-19 weekly death count per 100,000 people.
The week presented in these maps corresponds to the last week of the month.

4.4 Methodology

Consider the following equation:

$$\mathrm{y_{it}} = \alpha_i + \gamma\text{Local COVID severity}_{\mathrm{it}} + \theta\text{National COVID severity}_{\mathrm{it}} + \partial\text{National Lockdown}_{\mathrm{t}} + \beta\mathrm{X_{it}} + \rho_t + \varepsilon_{\mathrm{it}} \qquad (1)$$

Estimating equation (1) as a pooled linear model will identify differences across individuals over time under the assumption of strict exogeneity. However, it is likely that there are individual specific factors influencing mental health which are not observed in survey data, which could range from personality traits (such as resilience) to family circumstances (for example having an at-risk family member or friend). These unobserved factors could be a source of omitted variable bias. A fixed effects model is an improvement on pooled OLS as it only requires that the time varying covariates must not correlate with the time varying error term. For example, if mental health is correlated with resilience, fixed effect estimates will not be biased so long as resilience is time-invariant. I estimate both models but, for the reasons listed above, the fixed effects model is preferred and the pooled OLS results are relegated to Appendix C.

In equation (1), y_{it} refers to the outcome of interest - one of the three measures of mental health: overall GHQ-12 score, number of mental health problems, and having at least one severe mental health problem. In our fixed effects estimation, α_i captures the individual fixed effect. This is likely to subsume the local authority fixed effect, assuming that individuals do not move to a different local authority during the sample period.[38]

Each of the three mental health outcomes are regressed on the measure of local COVID severity, national COVID severity, an indicator for being in lockdown. As mentioned previously, I use two distinct measures of COVID severity – weekly case rates and weekly death rates. For each of these measures there are two variables of importance. The first is the local rate, the effect of which is captured by parameter γ, and identification comes from changes in LA COVID severity over time. The second is the national rate, the effect of which is captured by parameter θ. This varies over time and across the nation in which the respondent

38 Unfortunately, it is not possible to identify movers in the Understanding Society COVID survey. However, between wave 8 and 9 of the main survey 4 percent of respondents moved. Of these, 65 percent moved within 10km of their previous address.

resides - England, Scotland, or Wales.[39] Both are measured in the same way and reported as on the government dashboard – in cases/deaths per 100,000 people to account for differences in population size across local authorities and nations.

The dummy variable for lockdown is constructed from either being in a full national lockdown or not in lockdown, only varying across time, not across nations or LAs. The lockdown variable does not capture the gradual easing of certain restrictions, for example, non-essential shops re-opening. The survey months when the UK was locked down were April, May, June, November of 2020 and January 2021.[40] There was also a tiered system which was implemented at the end of October 2020. The system did not remain in place for long, in fact, a national lockdown was re-imposed four weeks later. Whilst this generated some geographic variation in the stringency of lockdown rules; unfortunately, the survey data does not cover this period, so it is not possible to exploit the tiered system as part of the identification strategy.

X_{it} contains a vector of time-varying personal characteristics which include age, age squared, number of children, partnership status, education, employment, key worker, even been on furlough, had COVID, had symptoms, clinical vulnerability, and previous self-reported mental health.[41] I also include weather controls for rainfall and temperature which vary by LA and interview day. Survey wave fixed effects are captured by equation term ρ_t in order to control for common factors at the time of the survey. These time fixed effects control for the announcement of the discovery of and progress on testing of the COVID-19 vaccine and the subsequent roll-out of the vaccine.

A concern is that the testing capacity at the beginning of lockdown was low and is the reason for the arbitrarily low reported case rates. Whilst the case rates are what people actually observe (thus there is an argument against it being interpreted as measurement error), there was arguably a lot of media attention around testing capacity limitations. To account for this, I control for UK-wide testing capacity, which, in the available data, varies over time but not across LAs. Testing capacity could also be an important explanatory variable because an individual's mental health may depend on their ability to get tested.

[39] Northern Ireland is excluded from the analysis as it does not share a land border with the UK.
[40] There were small differences in lockdown rules across the three nations, but, as a general rule, lockdowns eased at the beginning of July and restrictions were re-introduced in November. If the data had been collected across the month, rather than in the final week of the month, we would be able to exploit within-wave variation in lockdown. Sadly, this is not the case.
[41] Gender and ethnicity were not included as controls due to be time-invariant and thus dropped from the fixed effects estimation. In the pooled OLS results found in Appendix C they are included as controls.

The model does not include any dynamic effects. Whilst this has been considered, it is not included, in part, due to the design of Understanding Society. The data collection period changes from monthly to bi-monthly from July meaning it is not readily amenable to time series analysis as dynamics become difficult to incorporate. The advantage of our model is its simplicity and ease of interpretation.

Finally, it is evident from the current literature that the pandemic has had differential effects across various population sub-groups. Thus, I examine heterogeneity by age, gender, ethnicity and being a lone mother. A feature of these population groups is that they do not vary over time and will be dropped in the fixed effects estimation which renders interaction effects of no use. For this reason, I use a sub sample analysis to analyse the differential effects of lockdown, case rates, and death rates on these sub-groups.

4.5 Results

I report the fixed effects estimates for all three outcomes using COVID case rates in Table 4.2 and COVID death rates in Table 4.3. The results of interacting lockdown with case rates are presented in Table 4.4 and the interaction term of lockdown and death rates in Table 4.5. The pooled OLS results can be found in Appendix C.[42]

4.5.1 Case rates

Table 4.2 presents the estimates for COVID case rates. In columns (1) and (2) the estimates are shown when only including the local case rates, columns (3) and (4) present the national cases, and columns (5) and (6) include both local and national case rates. The dummy indicator for being in lockdown is included in all specifications. Column 6 is the preferred specification because it includes the full set of controls and both local and national case rates.[43] The case rate has been scaled to improve readability and coefficients on case rates should be interpreted as an increase of 100 cases per 100,000 people. Indeed, an increase of 100 cases per 100,000 people is within the realms of possibility given the average case rate in the sample period was 75 cases per 100,000.

[42] Random effects models are also estimated and are found to be similar across all of the main specifications in Table 4.2 and 4.3. The Hausman test cannot be implemented on weighted regressions. Thus, I estimate random and fixed effect models in the unweighted sample and find the Hausman test always rejects in favour of the fixed effects estimates.

Both the national and local case rates do not appear to be important for overall well-being, nor the number of mental health problems, the estimates are not statistically significant (see columns 6 Panel A and B). However, an increase in the national case rate by 100 cases leads to an increase in the likelihood of having any severe mental health problem by 3.5 percentage points. This corresponds to an 18 percent increase in the incidence of severe mental health problems.

Understanding the size of the effect depends on whether an additional 100 cases per 100,000 is a reasonable increase in the case rate. In fact, the case rate was over 800 at peak times during the pandemic. An increase of 100 cases per 100,000 is a reasonable increase to consider since, at its peak, case rates were 800 per 100,000. Such an increase would lead to a 28 percent increase in the chance of having one severe mental health problem – a large effect.

Across all specifications, being in lockdown is found to increase mental health problems, the coefficient is large, statistically significant, and very stable across all regressions. A national lockdown is associated with increasing the overall GHQ-12 score (increasing mental distress) by 0.71 points, all else equal. This is equivalent to a 5.8 percent change from the mean (12.74). Turning to the number of mental health problems, being in lockdown is associated with an increase of half a mental health problem. Lockdown also has a large effect on severe mental health problems – a 6 percentage point increase in the likelihood of having a severe problem which is double the effect of an increase in the case rate of 100 in 100,000 people.

Table 4.2 - Fixed effects estimates of mental health outcomes regressed on lockdown, local and national COVID case rates

	(1) Basic	(2) Full	(3) Basic	(4) Full	(5) Basic	(6) Full
Panel A – Overall well-being GHQ-12						
Local case rate	0.155***	0.0632			0.0255	0.0558
	(0.0414)	(0.0632)			(0.0735)	(0.0749)
National case rate			0.199***	0.0807	0.175**	0.0285
			(0.0426)	(0.106)	(0.0747)	(0.126)
Lockdown	0.704***	0.712***	0.690***	0.708***	0.689***	0.706***
	(0.0644)	(0.0660)	(0.0650)	(0.0697)	(0.0648)	(0.0692)
N	62,951	53,077	62,951	53,077	62,951	53,077
Panel B – Number of problems						
Local case rate	0.0100	0.0755*			0.0328	0.0535
	(0.0273)	(0.0402)			(0.0476)	(0.0484)
National case rate			0.00111	0.134**	-0.0307	0.0842
			(0.0276)	(0.0661)	(0.0469)	(0.0807)
Lockdown	0.578***	0.571***	0.581***	0.555***	0.580***	0.553***
	(0.0392)	(0.0425)	(0.0395)	(0.0427)	(0.0396)	(0.0427)
N	62,951	53,077	62,951	53,077	62,951	53,077
Panel C – Any severe problem						
Local case rate	-0.00701**	0.0148***			0.00286	0.00557
	(0.00309)	(0.00488)			(0.00578)	(0.00587)
National case rate			-0.0106***	0.0404***	-0.0133**	0.0352***
			(0.00314)	(0.00812)	(0.00586)	(0.00986)
Lockdown	0.0719***	0.0713***	0.0731***	0.0643***	0.0730***	0.0640***
	(0.00484)	(0.00546)	(0.00495)	(0.00582)	(0.00493)	(0.00577)
N	63,960	53,679	63,960	53,679	63,960	53,679
Full set controls	No	Yes	No	Yes	No	Yes

Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: Columns (1), (3) and (5) present the basic specification which only includes the three variables reported in the Table. Columns (2), (4) and (6) present the results using the full set of control variables which include: age, age squared, education level, gender, an indicator for non-white, in couple, number of kids, employed, key worker, on

furlough, had COVID symptoms, tested for COVID, previous wave self-reported mental health, clinically vulnerable, rainfall, temperature and UK-wide testing capacity. Both the local and national case rates are measured on a scale of 1 in 1000 cases or can be interpreted as an increase of 100 in 100,000 cases.

4.5.2 Death rates

Table 4.3 follows the same structure as Table 4.2 but presents the estimates for local and national COVID death rates. Again, neither local nor national deaths predict overall mental well-being. However, national death rates are associated with an increase in more serious mental health problems – both in terms of the number of problems and the incidence of having a severe mental health problem. An increase in deaths by 100 per 100,000 leads to almost two additional mental health problems, but this is not a realistic increase in the death rate. Thus, an increase in the death rate of 10 per 100,000 leads to an increase in the number of mental health problems by just under one fifth of a problem. It also corresponds to a (statistically significant) 4.9 percentage point increase in the likelihood of having a severe mental health problem.

Regardless of the specification, the coefficient estimate on lockdown is statistically significant for all three outcomes. Imposing a lockdown is associated with an 0.73 point increase in mental distress on the overall GHQ-12 scale, which corresponds to a 6.7 percent change from the mean. It is associated with an additional half a mental health problem, which is large, and a 3.5 percentage point increase in the likelihood of having a severe mental health problem.

Table 4.3 - Fixed effects estimates of mental health outcomes regressed on lockdown, local and national COVID death rates

	(1) Basic	(2) Full	(3) Basic	(4) Full	(5) Basic	(6) Full
Panel A – Overall well-being GHQ-12						
Local death rate	0.549	-0.502			-0.802	-0.977
	(0.454)	(0.563)			(0.794)	(0.823)
National death rate			1.326**	0.00229	2.111**	0.974
			(0.589)	(0.803)	(1.029)	(1.175)
Lockdown	0.730***	0.765***	0.681***	0.731***	0.682***	0.732***
	(0.0678)	(0.0725)	(0.0702)	(0.0804)	(0.0701)	(0.0804)
N	62,900	53,042	62,951	53,077	62,900	53,042
Panel B - Number of problems						
Local death rate	0.909***	0.257			-0.578	-0.591
	(0.271)	(0.329)			(0.441)	(0.439)
National death rate			1.768***	1.155**	2.325***	1.738***
			(0.360)	(0.495)	(0.591)	(0.662)
Lockdown	0.521***	0.576***	0.468***	0.517***	0.468***	0.516***
	(0.0395)	(0.0436)	(0.0407)	(0.0472)	(0.0407)	(0.0472)
N	62,900	53,042	62,951	53,077	62,900	53,042
Panel C – Any severe problem						
Local death rate	0.209***	0.218***			0.0160	0.0437
	(0.0381)	(0.0464)			(0.0673)	(0.0678)
National death rate			0.318***	0.400***	0.301***	0.356***
			(0.0509)	(0.0686)	(0.0895)	(0.100)
Lockdown	0.0553***	0.0613***	0.0485***	0.0491***	0.0485***	0.0491***
	(0.00523)	(0.00603)	(0.00587)	(0.00713)	(0.00585)	(0.00712)
N	63,909	53,644	63,960	53,679	63,909	53,644
Full set controls	No	Yes	No	Yes	No	Yes

Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: Columns (1), (3) and (5) present the basic specification which only includes the three variables reported in the Table. Columns (2), (4) and (6) present the results using the full set of control variables which include: age, age squared, education level, gender, an indicator for non-white, in couple, number of kids, employed, key worker, on furlough, had COVID symptoms, tested for COVID, previous wave self-reported mental health, clinically vulnerable, rainfall, temperature and UK-wide testing capacity. Both the local and national death rates are measured on a scale of 1 in 1000 deaths or can be interpreted as an increase of 100 in 100,000 deaths.

4.5.3 Summary of headline estimates

To summarise, whether using cases or deaths, the local measure of COVID severity has no effect on any of the three mental health outcomes. One explanation could be national rates are more visible in the media. Moreover, they may act as a signal for future lockdowns and more stringent public health measures. The three outcomes show different results, overall mental well-being does not suffer when national cases/deaths are rising, whereas the incidence of more serious mental health problems are strongly related to national case and death rates. Across all specifications, it is clear that lockdown has a large negative and statistically significant impact on mental health. Also, it is encouraging that the coefficient estimate on lockdown is similar regardless of whether COVID case rates or COVID death rates are used.

To put these results into context, Banks & Xu (2020), as of April 2020, find the overall effect of the pandemic was a 0.9 point increase in the overall GHQ-12 score, one additional mental health problem, and the incidence of reporting any severe mental health problem doubled. In my results the coefficient on lockdown is around 0.7 – only slightly lower than the estimate they find. Whilst it is not possible to make a direct comparison between the estimates, not least because the Banks finding is specific to counterfactual mental health in April 2020, it could indicate that the overall effect of the pandemic was only slightly above that of the effect I find for being in lockdown. Implying that a substantial part of the mental health burden arises from the lockdown effect.

Finally, the pooled OLS results, found in Appendix C (Tables C.1 and C.2), are broadly similar to the fixed effects results. Although fixed effects modelling is preferable, there may be a concern that, without sufficient variation in both outcome and explanatory variables, the estimates lack precision. This is not the case here as the fixed effects are reported with at least as much precision as the OLS results.

4.5.4 Lockdown effects by COVID severity

In this sub-section, I examine mental health regressed on an interaction between the indicator for being in lockdown and the COVID severity measure. Figure 4.3 shows when the country is not in lockdown there is little variation in the number of cases (as they are all very close to zero). However, being in lockdown generates much greater geographical variation in case and death rates. Cases and deaths may have different impacts on mental health whether in lockdown or not. Lockdown may reassure people that government action will reduce future cases and death. On the other hand, higher cases/deaths in lockdown may cause even greater worry or the numerous press briefings and media reporting may draw more attention to the worrying COVID situation. I thus interact COVID case and death rates with being in lockdown.

Firstly, consider the interaction between (local and national) COVID case rates and lockdown, for which the estimates can be found in Table 4.4. The interaction term is not statistically significant for two outcomes, overall well-being GHQ-12 and number of problems – indicating that there is no effect of being in lockdown and having a higher national/local case rate. The interaction term is positive and statistically significant for having a severe mental health problem, which when in lockdown and an additional 100 cases per 100,000 people implies a 3.5 percentage point increase in the likelihood of having a severe problem.

Table 4.4 - Interaction between lockdown and case rates (fixed effects estimates)

	Overall GHQ-12		No. of problems		Any severe problem	
	(1)	(2)	(3)	(4)	(5)	(6)
Lockdown * Local case rate		0.176		0.105		-0.000240
		(0.132)		(0.0814)		(0.0105)
Lockdown * National case rate		-0.229		0.0548		0.0356**
		(0.202)		(0.125)		(0.0152)
Lockdown	0.706***	0.734***	0.553***	0.484***	0.0640***	0.0482***
	(0.0692)	(0.0940)	(0.0427)	(0.0562)	(0.00577)	(0.00824)
Local case rate	0.0558	-0.0565	0.0535	-0.0161	0.00557	0.00529
	(0.0749)	(0.0892)	(0.0484)	(0.0574)	(0.00587)	(0.00794)
National case rate	0.0285	0.177	0.0842	0.0311	0.0352***	0.00887
	(0.126)	(0.186)	(0.0807)	(0.121)	(0.00986)	(0.0136)
N	53,077	53,077	53,077	53,077	53,679	53,679

Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: Column (1), (3) and (5) report the results from Table 4.2 without the inclusion of the interaction terms. The full set of controls are included in all cases: age, age squared, education level, gender, an indicator for non-white, in couple, number of kids, employed, key worker, on furlough, had COVID symptoms, tested for COVID, previous wave self-reported mental health, clinically vulnerable, rainfall, temperature and UK-wide testing capacity.
Both the local and national case rates are measured on a scale of 1 in 1000 cases or can be interpreted as an increase of 100 in 100,000 cases.

Using death rates as the explanatory variable in Table 4.5 shows that in general being in lockdown and having a higher death rates (local or national) also does not correspond to poorer mental health outcomes. Though the estimates are large, they become statistically insignificant when interpreting both the main and interaction effects together.[44]

Table 4.5 - Interaction between lockdown and death rates (fixed effects estimates)

	Overall GHQ-12		No. of problems		Any severe problem	
	(1)	(2)	(3)	(4)	(5)	(6)
Lockdown * Local area death rate		8.584		-0.323		0.675
		(7.958)		(4.327)		(0.616)
Lockdown * National death rate		-202.6**		-85.10		-5.540
		(92.90)		(56.41)		(6.356)
Lockdown	0.732***	1.417***	0.516***	0.818***	0.0491***	0.0663***
	(0.0804)	(0.346)	(0.0472)	(0.211)	(0.00712)	(0.0248)
Local death rate	-0.977	-9.533	-0.591	-0.272	-0.628	-0.600
	(0.823)	(8.030)	(0.439)	(4.309)	(0.622)	(0.623)
National death rate	0.974	203.7**		86.91		5.897
	(1.175)	(92.95)		(56.39)		(6.356)
N	53,042	53,042	53,042	53,042	53,644	53,644

Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: Column (1), (3) and (5) report the un-interacted results from Table 4.2.The full set of controls are included in all cases: age, age squared, education level, gender, an indicator for non-white, in couple, number of kids, employed, key worker, on furlough, had COVID symptoms, tested for COVID, previous wave self-reported mental health, clinically vulnerable, rainfall, temperature and UK-wide testing capacity. The local and national death rates are measured on a scale of 1 in 1000 deaths or can be interpreted as an increase of 100 in 100,000 deaths.

[44] I visualise these interacted effects in Figures C.1 and C.2 in Appendix C for the case rate and death rate interaction. National case rates are found on the top row, local case rates on the bottom row, while the un-interacted estimate from Tables 4.2 and 4.3 is used for a comparison and represented by the blue line. Figure C.1 shows across all three outcomes, and for both national and local measures, there appears to be no effect on mental health of not being in lockdown and higher case rates - the slope of the line is relatively flat. There are no statistically significant differences between being in lockdown and not being in lockdown for any of the outcomes with the exception of having any severe problems. Figure C.2 shows no statistically significant differences between being in lockdown with a higher death rate and not being in lockdown with higher death rates for any of the three outcomes. However, the graph does show that for national deaths there is a slight difference in the slope of the lines, which may give an indication that higher deaths outside of lockdown are worse for mental health than those within lockdown. It is not possible to say however, that this is statistically significant difference between slopes.

4.5.5 Heterogeneity

This section presents the results of sub-sample analyses by gender, non-white ethnicity, being a lone mother, and age category. In each of the following Tables Panel A regressions use the COVID case rate while Panel B presents results of the COVID death rate measure.

4.5.5.1 Gender

It has been argued in the media that the mental health of women has been more badly affected by COVID than that of men. There are stark differences in the lockdown effect between men and women in the estimates in Table 4.6. Looking at the case rates regression in Panel A, for women, the coefficient on lockdown is larger, and statistically different, than the coefficient from the male sample for all three outcomes. In fact, for both overall well-being and the incidence of having a severe problem, the female sample estimate is more than double that of male sample. This may reflect the differential effect of lockdown on the two genders as evidence has shown women are likely to take on additional childcare, home-schooling, and household chores. It may also be the case that women feel the consequences of social isolation more than men. These findings are consistent with the coefficients on lockdown in Panel B (using deaths as the measure of COVID severity) - they are similar in size, and the effect of lockdown is always larger in the female sample compared with the male sample.

Women also suffer a much greater mental health penalty from a higher national death rate in terms of the number of mental health problems. In the female sample, an increase in the national death rate by 10 deaths per 100,000, is associated with almost an additional three tenths of a mental health problem. In the male sample the coefficient is much smaller and statistically insignificant. But when looking at the most serious measure of mental health, there are no statistically significant differences between genders for national deaths. This could indicate that women may suffer mentally over more of the individual components of the GHQ-12 (e.g. sleep, concentration etc.) when national deaths are higher.

In terms of COVID case rates there are no statistically significant differences between genders for overall well-being and the number of mental health problems. But slight differences on having a severe mental health problem are present in the results. Higher national cases are associated with a greater likelihood of having a severe mental health problem and, although the effect is significant for both genders, in this case the estimate is larger for men, 4.4 percent compared with 2.6 percent for women. But this is only marginally statistically different.

Table 4.6 - Heterogeneity by gender (fixed effects estimates)

	Overall GHQ-12		No. of problems		Any severe problem	
	(1)	(2)	(3)	(4)	(5)	(6)
	Female	Male	Female	Male	Female	Male
Panel A – Case rates						
Local case rate	0.0783	0.0353	0.0544	0.0529	0.0139*	-0.00383
	(0.0985)	(0.112)	(0.0667)	(0.0682)	(0.00790)	(0.00852)
National case rate	0.101	-0.0744	0.177	-0.0276	0.0260*	0.0444***
	(0.180)	(0.175)	(0.115)	(0.112)	(0.0135)	(0.0144)
Lockdown	0.958***	0.444***	0.689***	0.412***	0.0811***	0.0462***
	(0.0922)	(0.104)	(0.0627)	(0.0581)	(0.00824)	(0.00806)
N	31,035	22,042	31,035	22,042	31,433	22,246
Panel B – Death rates						
Local death rate	-0.553	-1.405	-0.442	-0.791	0.139	-0.0564
	(1.021)	(1.298)	(0.636)	(0.594)	(0.0883)	(0.100)
National death rate	2.048	-0.425	2.927***	0.354	0.326**	0.373***
	(1.534)	(1.805)	(0.895)	(0.979)	(0.140)	(0.141)
Lockdown	0.910***	0.556***	0.589***	0.449***	0.0616***	0.0365***
	(0.104)	(0.124)	(0.0680)	(0.0659)	(0.00983)	(0.0105)
N	31,011	22,031	31,011	22,031	31,409	22,235

Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: The full set of controls are included in all cases: age, age squared, education level, gender, an indicator for non-white, in couple, number of kids, employed, key worker, on furlough, had COVID symptoms, tested for COVID, previous wave self-reported mental health, clinically vulnerable, rainfall, temperature and UK-wide testing capacity.
Local and national case and death rates are measured on a scale of 1 in 1000 cases/deaths or can be interpreted as an increase of 100 in 100,000 cases/deaths.

4.5.5.2 Ethnicity

Table 4.7 shows that being in a lockdown does not appear to affect white and non-white ethnicities differentially in terms of their overall well-being and the number of mental health problems. The coefficient on lockdown is slightly higher for non-white ethnicities but is not measured with enough precision to say they are statistically different from each other. Nor are they quantitively different. However, examining the outcome of having a severe mental health problem, the coefficient on lockdown for the non-white ethnicity sample is smaller than, but not statistically different, from, the large and statistically significant coefficient in the white ethnicity sample (in both Panel A and B).

Though the national death rate is insignificant and large for non-whites for the number of mental health problems and overall well-being, the estimate is measured with considerably more error, due to smaller sample sizes. However, national deaths are three times as damaging to severe mental health problems for non-whites – a differential effect that is statistically significant.

Table 4.7 - Heterogeneity by non-white ethnicity (fixed effects estimates)

	Overall GHQ-12		No. of problems		Any severe problem	
	(1)	(2)	(3)	(4)	(5)	(6)
	White	Non-white	White	Non-white	White	Non-white
Panel A – Case rates						
Local case rate	0.0909	-0.308	0.0703	-0.123	0.00901	-0.0332*
	(0.0699)	(0.370)	(0.0473)	(0.220)	(0.00596)	(0.0195)
National case rate	0.0968	-0.736	0.118	-0.270	0.0347***	0.0432
	(0.127)	(0.546)	(0.0833)	(0.309)	(0.01000)	(0.0410)
Lockdown	0.698***	0.865***	0.546***	0.661***	0.0677***	0.0284
	(0.0698)	(0.332)	(0.0445)	(0.160)	(0.00563)	(0.0309)
N	48,451	4,626	48,451	4,626	48,959	4,720
Panel B – Death rates						
Local death rate	-0.319	-10.37*	-0.403	-3.310	0.0752	-0.575
	(0.788)	(5.771)	(0.430)	(2.852)	(0.0678)	(0.405)
National death rate	0.578	9.043	1.657**	3.905	0.299***	1.291**
	(1.149)	(7.412)	(0.676)	(3.183)	(0.0963)	(0.645)
Lockdown	0.738***	0.699*	0.517***	0.528***	0.0554***	-0.0126
	(0.0790)	(0.407)	(0.0489)	(0.184)	(0.00667)	(0.0402)
N	48,416	4,626	48,416	4,626	48,924	4,720

Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: The full set of controls are included in all cases: age, age squared, education level, gender, an indicator for non-white, in couple, number of kids, employed, key worker, on furlough, had COVID symptoms, tested for COVID, previous wave self-reported mental health, clinically vulnerable, rainfall, temperature and UK-wide testing capacity.
Local and national case and death rates are measured on a scale of 1 in 1000 cases/deaths or can be interpreted as an increase of 100 in 100,000 cases/deaths.

4.5.5.3 Single vs coupled mothers

It seems likely that lone mothers find the experience of lockdown more isolating. Moreover, they are not able to share the burden of caring for their child with a second parent. One caveat is that the small sample of lone mothers may cause a loss of precision in the fixed effects estimates. Despite this, the results in Table 4.8 show that lone mothers are significantly more adversely affected by lockdown than coupled mothers. The magnitude of the effect of lockdown is almost always around 3 times the estimate in the lone mother sample – a large and statistically significant difference. It implies that lockdown is associated with an increase in mental distress on the overall well-being scale of 2.5 points, almost one and a half additional mental health problem, and a 15 percentage point increase in the likelihood of a severe problem.

Table 4.8 - Heterogeneity by household type (fixed effects estimates)

	Overall GHQ-12		No. of problems		Any severe problem	
	(1) Lone mother	(2) Coupled mother	(3) Lone mother	(4) Coupled mother	(5) Lone mother	(6) Coupled mother
Panel A – Case rates						
Local case rate	-0.252	0.168	-0.0619	0.177*	0.0538	0.00418
	(0.217)	(0.111)	(0.233)	(0.101)	(0.0360)	(0.0117)
National case rate	0.375	0.0608	0.123	0.118	-0.0143	0.0521***
	(0.592)	(0.272)	(0.420)	(0.211)	(0.0589)	(0.0172)
Lockdown	2.474***	0.682***	1.473***	0.565***	0.150***	0.0510***
	(0.493)	(0.144)	(0.307)	(0.102)	(0.0420)	(0.0113)
N	1,215	10,131	1,215	10,131	1,234	10,312
Panel B – Death rates						
Local death rate	-1.550	-2.414	-0.110	-1.137	0.375	0.0163
	(4.511)	(1.596)	(3.182)	(0.976)	(0.407)	(0.127)
National death rate	1.417	4.017	2.738	3.407**	-0.212	0.525***
	(7.466)	(2.590)	(4.366)	(1.690)	(0.623)	(0.201)
Lockdown	2.534***	0.644***	1.307***	0.500***	0.147**	0.0311**
	(0.684)	(0.181)	(0.373)	(0.105)	(0.0580)	(0.0146)
N	1,214	10,126	1,214	10,126	1,233	10,307

Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: The full set of controls are included in all cases: age, age squared, education level, gender, an indicator for non-white, in couple, number of kids, employed, key worker, on furlough, had COVID symptoms, tested for COVID, previous wave self-reported mental health, clinically vulnerable, rainfall, temperature and UK-wide testing capacity.
Local and national case and death rates are measured on a scale of 1 in 1000 cases/deaths or can be interpreted as an increase of 100 in 100,000 cases/deaths.

In contrast, the results also suggest that coupled mothers are more adversely affected by increases in the death rate, for the mental health outcomes that capture more serious mental health problems.

4.5.5.4 Age categories

The sample is split into just four age categories in order not to heavily restrict sample sizes. I expect that lockdown will have a lower impact on the elderly, conditional on the same family composition. Yet Table 4.9 reveals there is no clear pattern in the relationship between age and being in lockdown.

The differences for overall-wellbeing are also not entirely clear cut, as the estimates on lockdown differ slightly between Panel A and B. In both panels the age category most affected by lockdown is 31–45-year-olds, an effect size corresponding to a 1 to1.1 point increase on the overall well-being GHQ-12 measure of mental distress (column 2). In Panel A, 16-30 year olds are also negatively affected by lockdown by almost twice as much as the 60+ age category. However, a caveat is that this does not hold in Panel B, where there are no significant differences between 16-30 year olds and 60+ category. With regard to the number of mental health problems, 16–30-year-olds and 31-45 year olds both see an increase in the number of mental health problems (columns 5 and 6) by around two thirds of a problem – compared with just over one third for both 46-60 year olds and ages 60+ (columns 7 and 8). But, looking at the third outcome, the incidence of severe mental health problems, the effect is much more similar across age categories and there is little heterogeneity.

As reported in the main specifications, overall well-being is not impacted by COVID case and death rates for any of the age categories (columns 1-4). However, there are some differences across ages with regard to the number of mental health problems. Among the two youngest age categories, an additional 10 deaths per 100,000 leads to around an additional two fifths of a mental health problem, where there is no statistically significant effect for the two oldest age groups. This is an interesting finding, but it does not hold for having a severe mental health problem. For this outcome, the age group 31-45 are most affected by both national case rates and death rates, followed by the 60+ age group. As might be predicted, neither case rates nor death rates have an impact on having a severe mental health problem for those in the youngest age category - 16-30 year olds.

Table 4.9 - Heterogeneity by age category (fixed effects estimates)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Age category	16-30	31-45	46-60	60+	16-30	31-45	46-60	60+	16-30	31-45	46-60	60+
	Overall GHQ-12				Number of problems				Any severe problem			
Panel A – Case rates												
Local case rate	-0.0624	0.0759	0.132	0.0440	-0.0699	0.0429	0.182**	0.0316	0.00855	0.00163	0.00949	0.00485
	(0.301)	(0.132)	(0.0813)	(0.0695)	(0.180)	(0.0803)	(0.0753)	(0.0455)	(0.0208)	(0.00768)	(0.0109)	(0.00616)
National case rate	0.126	0.196	-0.184	0.105	0.158	0.320**	-0.204	0.132	0.0207	0.0557***	0.0233	0.0446***
	(0.420)	(0.262)	(0.180)	(0.150)	(0.263)	(0.159)	(0.140)	(0.0867)	(0.0309)	(0.0186)	(0.0146)	(0.0140)
Lockdown	0.904***	1.086***	0.444***	0.508***	0.768***	0.777***	0.377***	0.394***	0.0624***	0.0762***	0.0556***	0.0610***
	(0.220)	(0.163)	(0.0931)	(0.0805)	(0.130)	(0.0932)	(0.0658)	(0.0528)	(0.0200)	(0.0104)	(0.00804)	(0.00796)
N	4,307	9,940	17,043	21,787	4,307	9,940	17,043	21,787	4,388	10,113	17,226	21,952
Panel B – Death rates												
Local death rate	-3.856	-1.541	-0.0384	0.448	-1.947	-1.275	-0.0306	0.220	0.168	-0.122	0.0149	0.116
	(3.328)	(1.621)	(1.077)	(0.943)	(1.554)	(1.009)	(0.630)	(0.525)	(0.222)	(0.146)	(0.103)	(0.135)
National death rate	7.271*	2.455	-1.730	-1.404	4.109*	3.546**	0.0821	0.580	0.398	0.539***	0.231	0.329***
	(4.211)	(2.537)	(1.618)	(1.341)	(2.121)	(1.585)	(1.016)	(0.749)	(0.330)	(0.201)	(0.145)	(0.018)
Lockdown	0.691***	1.108***	0.546***	0.622***	0.643***	0.732***	0.367***	0.390***	0.0343	0.0651***	0.0487***	0.0454***
	(0.264)	(0.182)	(0.105)	(0.0905)	(0.148)	(0.111)	(0.0614)	(0.0574)	(0.0247)	(0.0129)	(0.0100)	(0.000)
N	4,302	9,931	17,033	21,776	4,302	9,931	17,033	21,776	4,383	10,104	17,216	21,941

Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: The full set of controls are included in all cases: age, age squared, education level, gender, an indicator for non-white, in couple, number of kids, employed, key worker, on furlough, had COVID symptoms, tested for COVID, previous wave self-reported mental health, clinically vulnerable, rainfall, temperature and UK-wide testing capacity. Local and national case and death rates are measured on a scale of 1 in 1000 cases/deaths or can be interpreted as an increase of 100 in 100,000 cases/deaths.

4.6 Robustness

Though the reported case rates are what people see in the press, the figures in the early stages of the pandemic are not comparable with later figures – because of the limited availability of testing capacity in the UK. Including testing capacity as a control variable is one way to account for this and indeed its inclusion does change the results materially.[45] Ideally, local testing capacity data would be available, since there is likely to be discrepancies between local authorities in their testing capacity, which may depend on their location, health funding and infrastructure (number of labs). Unfortunately, the data, accessible via the government dashboard, are available only at the UK level.

As a rough robustness check, I scale the case rates by a testing factor, which roughly predicts case rates if tests had been available. I do this by calculating a testing factor, by dividing the current testing capacity by the "full testing capacity" where "full testing capacity" is defined as the testing capacity as of January of 2021. The cases are then estimated by dividing the case rate by the testing factor,

$$Predicted\ cases = \frac{Case\ rate}{Testing\ Factor}$$

where,

$$Testing\ Factor = \frac{Testing\ capacity}{Jan\ 2021\ Testing\ capacity}$$

The results in Table 4.10 compare the main results from Table 4.2 to the results when scaling cases by testing capacity. Broadly the results are in line but there are a few minor differences. The effect of lockdown on overall well-being is slightly larger when scaling by the testing capacity. The coefficient on the national case rate becomes statistically significant for the number of problems where before it was not. These discrepancies could be an indication of the sensitivity of results to testing capacity, although, scaling the cases assumes the case rate is linearly related to testing capacity, which may be an overly simplistic assumption.

[45] Results without controlling for testing capacity have not been reported. But often the sign is the reverse of what is expected and not consistent with the result on death rates.

Table 4.10 - Scaling by a testing factor (fixed effects estimates)

	GHQ-12 Overall score		Number of problems		Any severe problem	
	Main sample	New	Main sample	New	Main sample	New
Local case rate	0.0558	-0.0138	0.0535	0.00568	0.00557	0.00347
	(0.0749)	(0.0313)	(0.0484)	(0.0200)	(0.00587)	(0.00359)
National case rate	0.0285	-0.0327	0.0842	0.0551*	0.0352***	0.0212***
	(0.126)	(0.0479)	(0.0807)	(0.0305)	(0.00986)	(0.00438)
Lockdown	0.706***	0.816***	0.553***	0.539***	0.0640***	0.0528***
	(0.0692)	(0.0645)	(0.0427)	(0.0386)	(0.00577)	(0.00576)
N	53,077	53,077	53,077	53,077	53,679	53,679

Clustered standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1
Note: The full set of controls are included in all cases: age, age squared, education level, gender, an indicator for non-white, in couple, number of kids, employed, key worker, on furlough, had COVID symptoms, tested for COVID, previous wave self-reported mental health, clinically vulnerable, rainfall, temperature and UK-wide testing capacity.
Both the local and national case rates are measured on a scale of 1 in 1000 cases or can be interpreted as an increase of 100 in 100,000 cases.

4.7 Conclusion

This chapter connects both the severity of COVID and being in lockdown, with mental health in the UK. I attempt to separate out the overall effects of the pandemic on mental health into the incarceration effect and the threat of the virus, and can quantify the collateral damage of lockdown, case, and death rates. I do this by exploiting geographic variation in local and national COVID severity, and temporal variation in lockdown, estimating fixed effects models in a credible identification strategy which should eliminate some of the biases associated with unobserved heterogeneity.

The results show that lockdown has a clear significant and negative impact on mental health, which is statistically significant and similar in magnitude across all specifications. Lockdown is associated with between a 5-6 percentage point increase in the incidence of severe mental health problems, and half an additional mental health problem. Lockdown also increases mental distress by 0.7 points on the overall GHQ-12 measure of well-being, an effect size comparable the 0.76 point increase found in Anaya et al. (2021), despite misgivings with their approach.

Local COVID cases and deaths are relatively less important in predicting mental health than national rates. However, rising national cases and deaths are associated with declines in more serious mental health outcomes but not for overall well-being. The results can give an estimate of the potential mental health consequences should cases/deaths increase by some given amount. This is an important policy implication not only for (potential) future pandemics but also in evaluating government decision making during the pandemic.

The effect of cases, deaths and lockdown differ across sub-groups of the populations. Younger individuals, women, lone mothers, and non-white ethnicities are much more adversely affected by being in lockdown. This is an unsurprising result, and though it confirms findings from previous work of the impact of the pandemic on mental health, it serves to isolate the lockdown effect alone from the total 'pandemic' effect. And I find that those "at risk", such as lone mothers, are significantly more adversely affected by being in lockdown.

Recently, the Lancet assembled its own "COVID-19 Commission Mental Health Task Force" to summarise and analyse key findings from the literature and offer recommendations on the basis of it. The authors of the report ((Aknin et al. (2021)), which includes much of the literature discussed in this chapter, conclude their analysis by suggesting that mental health returned to previous levels by mid-2020. The findings from my chapter, would suggest that this improvement in mental health was a consequence of the lifting of lockdown in mid-2020. There have been many calls for more action on mental health, and whilst an eye must be kept on potential long-term effects of mental health, the evidence here suggests the time for action was as lockdowns occurred.

The main contribution of this work is being able to quantify how the severity of COVID affects mental health, but there are many potential extensions. There are additional ways of measuring COVID severity aside from case and death rates - hospital admissions and the R number are two interesting dimensions. Given the many adverts encouraging us to "Stay home to save lives", with the aim of reducing the strain on the NHS, one might think that hospital capacity could predict mental health. Whilst this may be an additional measure, considerations would need to be made in that people often travel (or are taken) to hospitals outside their local authority or even region, depending on capacity and how poorly they are. The R number is another distinct measure as it is likely to better capture the threat of infection both now and in the future – cases/deaths may be low but the risk of infection high. However, the R number data is not available at the local authority level, it is only available for each of the nations. It is unclear whether this will be rectified in future data releases.

I use a crude measure of lockdown restrictions. In reality, things were not nearly as black and white as our lockdown indicator would suggest. And often nations differed in the intricacies of their rules, such as the number of people you were allowed to meet outside/inside. The Oxford COVID-19 Government Response Tracker (OxCGRT) has collected publicly available information to create indices which measure a governments response to the pandemic over a range of indicators, for example containment and health measures, economic support, and lockdowns. The main advantage is to facilitate cross-country comparisons. Though at first these indicators were only available for the UK, work is currently underway to calculate indices separately for each of the four nations. Once this becomes available it could serve as an additional specification in the analysis which can capture the gradual easing of lockdown, varying across nation and over time.

This chapter uses simple measures of mental health, the outcomes examined do not measure clinical disorders such as anxiety or depression, which may be preferred, though the pandemic is likely to have affected diagnoses as access to mental health services was restricted. Moreover, there are other dimensions of mental health, such as alcohol consumption, anxiety, and sleep quality, which have not been explored in this chapter.

Another worthy project would be to delve deeper into the welfare consequences of high case rates, death rates, and lockdowns – disaggregating into health, economic and other factors. Recently, the vaccine rollout, despite many people receiving their vaccine whilst in lockdown, could have had ambiguous effects on mental health, and is something that could be analysed as more data becomes available. To conclude, there is an overwhelming case for future research which could form a large portfolio of work.

Appendix C

GHQ-12 Questions

The next questions are about how you have been feeling recently…

Have you recently been able to concentrate on whatever you're doing?
{Better than usual, Same as usual, Less than usual, Much less than usual}

Have you recently lost much sleep over worry?
{Not at all, No more than usual, Rather more than usual, Much more than usual}

Have you recently felt that you were playing a useful part in things?
{More so than usual, Same as usual, Less so than usual, Much less than usual}

Have you recently felt capable of making decisions?
{More so than usual, Same as usual, Less so than usual, Much less capable}

Have you recently felt constantly under strain?
{Not at all, No more than usual, Rather more than usual, Much more than usual}

Have you felt you couldn't overcome your difficulties?
{Not at all, No more than usual, Rather more than usual, Much more than usual}

Have you recently been able to enjoy your normal day-to-day activities?
{More so than usual, Same as usual, Less so than usual, Much less than usual}

Have you recently been able to face up to problems?
{More so than usual, Same as usual, Less so than usual, Much less able}

Have you recently been feeling unhappy or depressed?
{Not at all, No more than usual, Rather more than usual, Much more than usual}

Have you recently been losing confidence in yourself?
{Not at all, No more than usual, Rather more than usual, Much more than usual}

Have you recently been thinking of yourself as a worthless person?
{Not at all, No more than usual, Rather more than usual, Much more than usual}

Have you recently been feeling happy, all things considered?
{More so than usual, About the same as usual, Less so than usual, Much less than usual}

Table C.1 - Pooled OLS estimates of mental health outcomes regressed on local & national COVID case rates

	(1) Basic	(2) Full	(3) Basic	(4) Full	(5) Basic	(6) Full
Panel A – Overall well-being GHQ-12						
Local case rate	0.117**	0.0497			0.217**	0.0873
	(0.0542)	(0.0708)			(0.106)	(0.0841)
National case rate			0.0669	-0.0800	-0.144	-0.156
			(0.0549)	(0.120)	(0.111)	(0.144)
Lockdown	0.763***	0.752***	0.783***	0.789***	0.776***	0.784***
	(0.0777)	(0.0760)	(0.0770)	(0.0809)	(0.0774)	(0.0808)
N	62,951	53,077	62,951	53,077	62,951	53,077
Panel B – Number of problems						
Local case rate	-0.0161	0.0589			0.111*	0.0633
	(0.0345)	(0.0445)			(0.0635)	(0.0546)
National case rate			-0.0745**	0.0371	-0.182***	-0.0183
			(0.0334)	(0.0737)	(0.0628)	(0.0927)
Lockdown	0.618***	0.612***	0.638***	0.620***	0.634***	0.616***
	(0.0463)	(0.0476)	(0.0457)	(0.0490)	(0.0460)	(0.0491)
N	62,951	53,077	62,951	53,077	62,951	53,077
Panel C – Any severe problem						
Local case rate	-0.00733**	0.0132**			0.0106*	0.00773
	(0.00314)	(0.00513)			(0.00559)	(0.00598)
National case rate			-0.0155***	0.0295***	-0.0258***	0.0227**
			(0.00341)	(0.00848)	(0.00609)	(0.0100)
Lockdown	0.0751***	0.0749***	0.0778***	0.0709***	0.0774***	0.0704***
	(0.00506)	(0.00567)	(0.00511)	(0.00595)	(0.00511)	(0.00592)
N	63,960	53,679	63,960	53,679	63,960	53,679
Full set controls	No	Yes	No	Yes	No	Yes

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

Note: Columns (1), (3) and (5) present the basic specification which only includes the three variables reported in the Table. Columns (2), (4) and (6) present the results using the full set of control variables which include: age, age squared, education level, gender, an indicator for non-white, in couple, number of kids, employed, key worker, on furlough, had COVID symptoms, tested for COVID, previous wave self-reported mental health, clinically vulnerable, rainfall, temperature and UK-wide testing capacity.

Both the local and national case rates are measured on a scale of 1 in 1000 cases or can be interpreted as an increase of 100 in 100,000 cases.

Table C.2 - Pooled OLS estimates of mental health outcomes regressed on local and national COVID death rates

	(1) Basic	(2) Full	(3) Basic	(4) Full	(5) Basic	(6) Full
Panel A – Overall well-being GHQ-12						
Local death rate	0.341	-0.0107			0.162	0.729
	(0.652)	(0.733)			(1.351)	(1.214)
National death rate			0.490	-0.819	0.299	-1.537
			(0.654)	(0.926)	(1.532)	(1.601)
Lockdown	0.781***	0.767***	0.775***	0.822***	0.774***	0.820***
	(0.0844)	(0.0903)	(0.0809)	(0.0980)	(0.0809)	(0.0979)
N	62,900	53,042	62,951	53,077	62,900	53,042
Panel B - Number of problems						
Local death rate	0.746**	0.539			0.0649	0.449
	(0.357)	(0.407)			(0.706)	(0.655)
National death rate			1.227***	0.623	1.136	0.186
			(0.395)	(0.545)	(0.832)	(0.892)
Lockdown	0.561***	0.592***	0.533***	0.588***	0.532***	0.585***
	(0.0480)	(0.0535)	(0.0463)	(0.0577)	(0.0464)	(0.0577)
N	62,900	53,042	62,951	53,077	62,900	53,042
Panel C – Any severe problem						
Local death rate	0.190***	0.220***			0.0540	0.118
	(0.0405)	(0.0495)			(0.0725)	(0.0726)
National death rate			0.281***	0.328***	0.226**	0.212**
			(0.0536)	(0.0724)	(0.0954)	(0.106)
Lockdown	0.0599***	0.0641***	0.0542***	0.0569***	0.0542***	0.0568***
	(0.00546)	(0.00617)	(0.00603)	(0.00736)	(0.00602)	(0.00733)
N	63,909	53,644	63,960	53,679	63,909	53,644
Full set of controls	No	Yes	No	Yes	No	Yes

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

Note: Columns (1), (3) and (5) present the basic specification which only includes the three variables reported in the Table. Columns (2), (4) and (6) present the results using the full set of control variables which include: age, age squared, education level, gender, an indicator for non-white, in couple, number of kids, employed, key worker, on furlough, had COVID symptoms, tested for COVID, previous wave self-reported mental health, clinically vulnerable, rainfall, temperature and UK-wide testing capacity.

Figure C.1 - Margins plot of the lockdown and case rate interaction terms

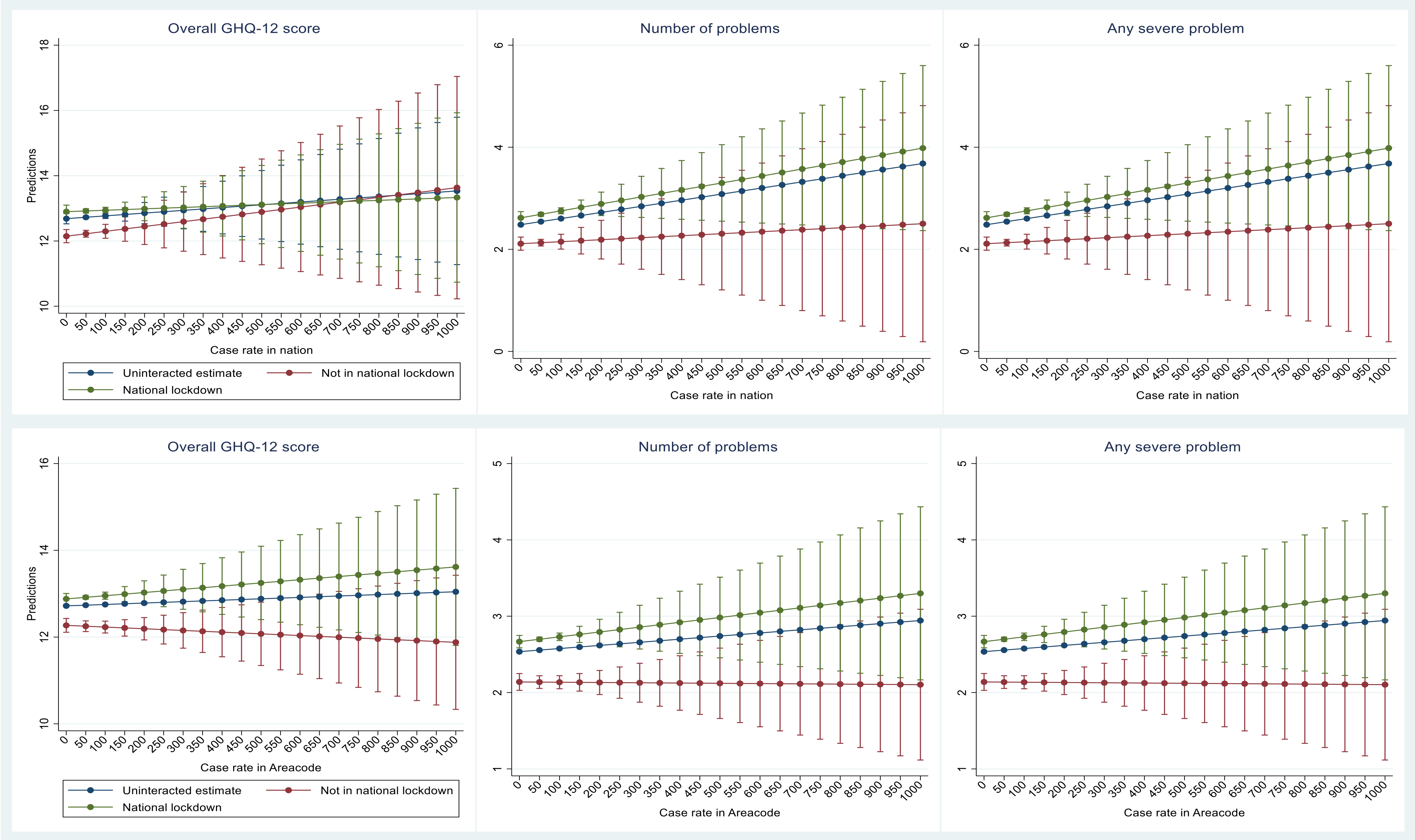
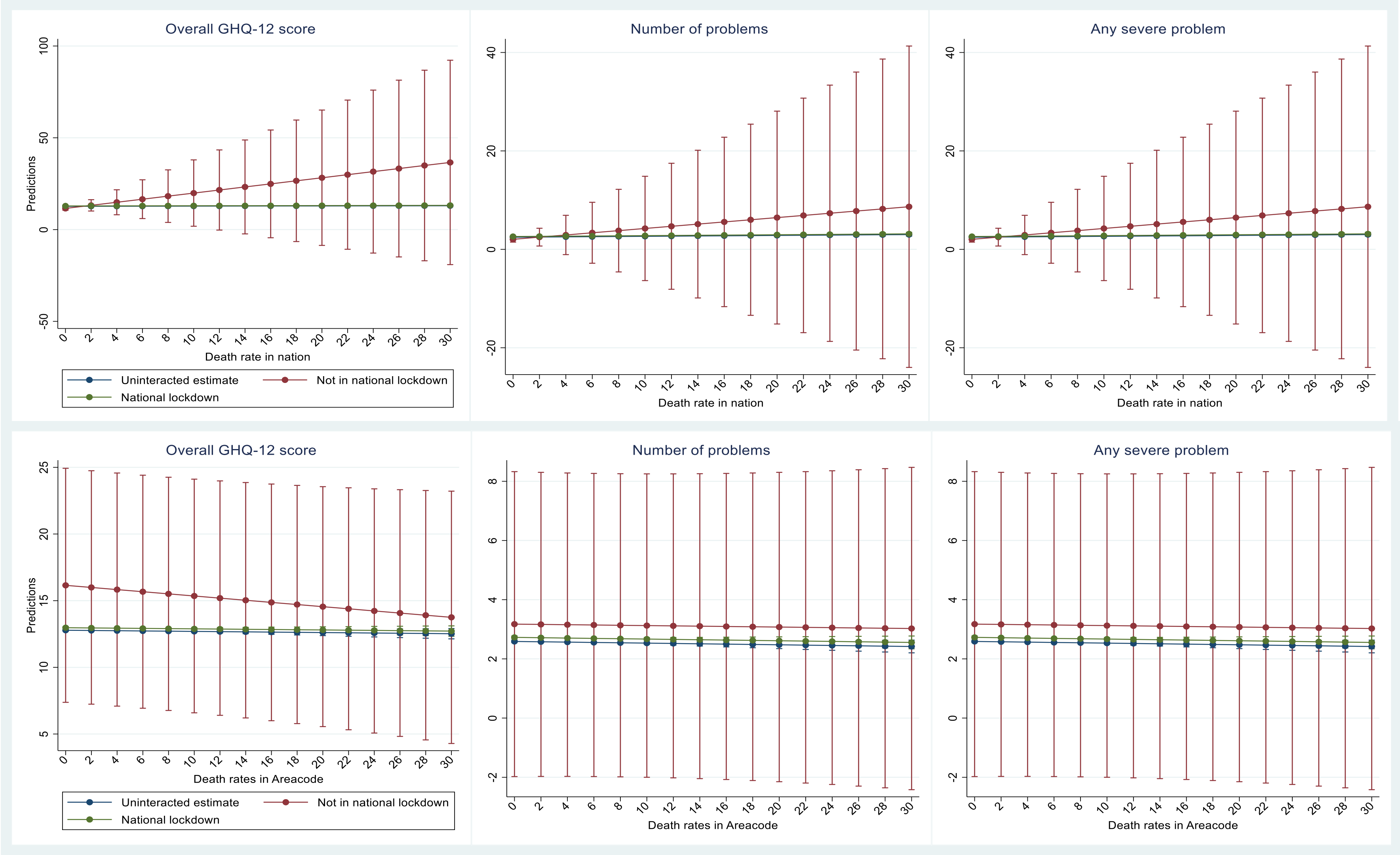


Figure C.2 – Margins plot of the lockdown and death rate interaction terms



5 Conclusion

This dissertation contributes to a wide range of literature: child maintenance, the effectiveness of labour market activation policies and the recently emerging literature on pandemics and mental health.

In Chapter 2, I examined the role of child maintenance in improving outcomes of young people aged 10-15. For boys, receiving child maintenance reduces youth conduct problems and improves pro-social skills. The importance of this finding should not be understated. Receiving child maintenance has implications both for the individual over the course of their life and wider benefits to society. Reducing conduct problems at a young age is found to reduce the likelihood of future unemployment and becoming involved in crime. Peer effects could also be an important mechanism – fewer disruptive pupils in school can improve outcomes for everyone, effectively a kind of positive externality that could form the basis of future analysis. There is further scope to extend this work to other dimensions of child outcomes such as engaging in risky behaviours and educational attainment. I have applied for access to linked dataset Understanding Society and the National Pupil Database, which can be used in future work to analyse the effects of child maintenance on educational outcomes and truancy. As an additional point – the sheer number of mothers who do not receive the maintenance they should suggests policy should be designed to improve receipt of child maintenance rather than focusing on the amounts of maintenance.

Chapter 3 considers the effect of a labour market activation policy, the LPO, which imposed, conditional on the age of the youngest child, job search requirements for welfare claimant lone mothers. Given the initially low employment rates, the reform is found to be successful at incentivising lone mothers to work. For policymakers, the reform, in increasing maternal employment, could be considered a success. Indeed, tackling worklessness has many gains for the government, reducing dependence on the welfare state and increasing revenue through greater tax collection. Whether the reform contributes to reducing lone mother poverty is another important question warranting further research.

In Chapter 4, I focus on the COVID-19 pandemic in the UK and its relationship with mental health. The results provide evidence that being in lockdown had a detrimental effect on mental health. This effect was most pronounced among women, younger individuals, and lone mothers. The level of COVID severity (in terms of the number of cases and deaths) was important, but only when measured nationally. The findings suggest that as lockdown was lifted mental health improved, the implication being that key time for action in tacking mental health was during lockdown. However, the results cannot say anything about the permanence of these adverse mental health effects. Thus, I still recommend investment in more mental health services, improving access to those who need it, and further research on the long-term mental health effects of the pandemic.

These findings should present significant value to policymakers, especially as, at the time of writing, the civil service has never been more focused on the use of evidence-based policymaking. Despite this shift towards having an evidence base on which policy is developed, the lack of data available for studying separated parents is astonishing. Administrative data on child maintenance is used by the Department for Work and Pensions, but access for external researchers is limited – with a suggestion from civil servants that the data is “too messy” as three different CM schemes operate concurrently. Though the most recent CM policy change incentivises parents to make private arrangements, a major disadvantage is that these types of arrangements are not recorded in the admin data and thus the government knows very little about separated families on these arrangements. To their credit, the DWP had questions included in the longitudinal Understanding Society survey in an attempt to correct for the absence of data on private arrangements, but so far I have not seen the data exploited in a meaningful way. Other data sources are limited - the Family and Children Study (FACS) ended in 2008, the Family Resources Survey is small in sample size and is cross-sectional, the Labour Force Survey does not ask about child maintenance, nor are the cohort studies large enough in size to facilitate robust empirical identification strategies. Even in Understanding Society the drawback of covering so many topics is that there leaves little room for deeper questions asked to the separated family population. Given these considerable data constraints and empirical challenges that researchers face when attempting to answer research questions, it is not an attractive area of research and one which is often overlooked.

As separation and “non-traditional” blended families are becoming increasingly common, we need to do a better job of understanding and capturing these varying family types in conventional research and making data accessible to researchers is more important than ever.

Bibliography

Adams-Prassl, A., Boneva, T., Golin, M., & Rauh, C. (2020). The impact of the coronavirus lockdown on mental health: evidence from the US. *HCEO Working Paper Series*, *030*, 1–20.

Adamsons, K., & Johnson, S. K. (2013). An updated and expanded meta-analysis of nonresident fathering and child well-being. *Journal of Family Psychology*. https://doi.org/10.1037/a0033786

Ahmed, M. Z., Ahmed, O., Aibao, Z., Hanbin, S., Siyu, L., & Ahmad, A. (2020). Epidemic of COVID-19 in China and associated Psychological Problems. *Asian Journal of Psychiatry*. https://doi.org/10.1016/j.ajp.2020.102092

Aknin, L., Neve, J. De, Dunn, E. W., Fancourt, D., Goldberg, E., Helliwell, J., Jones, S., Karam, E., Layard, R., Lyubomirsky, S., Rzepa, A., Saxena, S., Thornton, E., VanderWeele, T., Whillans, A., Zaki, J., Karadag Caman, O., & Ben Amor, Y. (2021). *Mental Health During the First Year of the COVID-19 Pandemic: A Review and Recommendations for Moving Forward*. 11–28. https://doi.org/https://doi.org/10.31234/osf.io/zw93g

Altonji, J. G., Elder, T. E., & Taber, C. R. (2005). An evaluation of instrumental variable strategies for estimating the effects of catholic schooling. *Journal of Human Resources*. https://doi.org/10.3368/jhr.xl.4.791

Amato, P. R. (2001). Children of divorce in the 1990s: An update of the Amato and Keith (1991) meta-analysis. In *Journal of Family Psychology*. https://doi.org/10.1037/0893-3200.15.3.355

Amato, P. R. (2014). The Consequences of Divorce for Adults and Children: An Update. *Drustvena Istrazivanja*. https://doi.org/10.5559/di.23.1.01

Amato, P. R., & Gilbreth, J. G. (1999). Nonresident Fathers and Children ' s Well-Being : A Meta-Analysis Published by : National Council on Family Relations Nonresident Fathers and Children ' s Well-Being : A Meta-Analysis. *Journal of Marriage and Family*, *61*(3), 557–573.

Anaya, L., Howley, P., Waqas, M., Yalonetzky, G., Anaya, L., Howley, P., & Waqas, M. (2021). *Locked down in distress : a causal estimation of the mental-health fallout from the COVID-19 pandemic in the UK* (21/10; Issue June). http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Angrist, J., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Argys, L. M., Peters, H. E., Brooks-Gunn, J., & Smith, J. R. (1998). The impact of child support on cognitive outcomes of young children. *Demography*. https://doi.org/10.2307/3004049

Aughinbaugh, A. (2001). Signals of child achievement as determinants of child support.

American Economic Review. https://doi.org/10.1257/aer.91.2.140

Aughinbaugh, A., Pierret, C. R., & Rothstein, D. S. (2005). The impact of family structure transitions on youth achievement: Evidence from the children of the NLSY79. *Demography*. https://doi.org/10.1353/dem.2005.0023

Autor, D., Figlio, D., Karbownik, K., Roth, J., & Wasserman, M. (2019). Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economic Journal: Applied Economics*. https://doi.org/10.1257/app.20170571

Avram, S., Brewer, M., & Salvatori, A. (2018). Can't work or won't work: Quasi-experimental evidence on work search requirements for single parents. *Labour Economics*. https://doi.org/10.1016/j.labeco.2017.10.002

Banks, J., Fancourt, D., & Xu, H. (2021). Mental Health and the COVID-19 Pandemic. In *World Happiness Report 2021*. https://worldhappiness.report/ed/2021/mental-health-and-the-covid-19-pandemic/

Banks, J., & Xu, X. (2020). The Mental Health Effects of the First Two Months of Lockdown during the COVID-19 Pandemic in the UK*. *Fiscal Studies, 41*(3), 685–708. https://doi.org/10.1111/1475-5890.12239

Baughman, R. A. (2017). The impact of child support on child health. *Review of Economics of the Household*. https://doi.org/10.1007/s11150-014-9268-3

Baydar, N., & Brooks-Gunn, J. (1994). The Dynamics of Child Support and Its Consequences for Children. In I. Garfinkel, S. McLanahan, & P. Robins (Eds.), *Child Support and Child Well-being* (pp. 257–279). Urban Institute Press.

Bertrand, M., & Pan, J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics*. https://doi.org/10.1257/app.5.1.32

Björklund, A., & Sundström, M. (2006). Parental separation and children's educational attainment: A siblings analysis on Swedish register data. *Economica*. https://doi.org/10.1111/j.1468-0335.2006.00529.x

Blow, L., Walker, I., & Zhu, Y. (2012). Who benefits from child benefit? *Economic Inquiry*. https://doi.org/10.1111/j.1465-7295.2010.00348.x

Blundell, R., Duncan, A., McCRAE, J., & Meghir, C. (2000). The Labour Market Impact of the Working Families' Tax Credit. *Fiscal Studies*. https://doi.org/10.1111/j.1475-5890.2000.tb00581.x

Bradshaw, J., Skinner, C., Stimson, C., & Williams, J. (1999). *Absent Fathers?* Psychology Press.

Brewer, M., Duncan, A., Shephard, A., & Suarez, M. J. (2005). Did working families' tax credit work? The final evaluation of the impact of in-work support on parents' labour supply and take-up behaviour in the UK. In *October*.

Brodeur, A., Clark, A. E., Fleche, S., & Powdthavee, N. (2021). COVID-19, lockdowns and well-being: Evidence from Google Trends. *Journal of Public Economics*, *193*, 104346. https://doi.org/10.1016/j.jpubeco.2020.104346

Brown, S., Taylor, K., & Wheatley Price, S. (2005). Debt and distress: Evaluating the psychological cost of credit. *Journal of Economic Psychology*. https://doi.org/10.1016/j.joep.2005.01.002

Brunello, G., Weber, G., & Weiss, C. T. (2017). Books are Forever: Early Life Conditions, Education and Lifetime Earnings in Europe. *Economic Journal*. https://doi.org/10.1111/ecoj.12307

Card, D., Kluve, J., & Weber, A. (2010). Active labour market policy evaluations: A meta-analysis. *Economic Journal*. https://doi.org/10.1111/j.1468-0297.2010.02387.x

Casebourne, J., Davies, M., Foster, S., Lane, P., Purvis, A., & Whitehurst, D. (2010). *Lone Parent Obligations: destinations of lone parents after Income Support eligibility ends*.

Castellini, G., Rossi, E., Cassioli, E., Sanfilippo, G., Innocenti, M., Gironi, V., Silvestri, C., Voller, F., & Ricca, V. (2021). A longitudinal observation of general psychopathology before the COVID-19 outbreak and during lockdown in Italy. *Journal of Psychosomatic Research*. https://doi.org/10.1016/j.jpsychores.2020.110328

Cheng, T. C., Kim, S., & Koh, K. (2021). The Impact of COVID-19 on Subjective Well-Being: Evidence from Singapore. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3695403

Cherlin, A. J., Chase-Lansdale, P. L., & McRae, C. (1998). Effects of parental divorce on mental health throughout the life course. *American Sociological Review*. https://doi.org/10.2307/2657325

Clark, A. E., & Lepinteur, A. (2019). The causes and consequences of early-adult unemployment: Evidence from cohort data. *Journal of Economic Behavior and Organization*. https://doi.org/10.1016/j.jebo.2019.08.020

Coleman, N., & Riley, T. (2012). *Lone Parent Obligations: following lone parents' journeys from benefits to work*. DWP Research Report 818. http://research.dwp.gov.uk/asd/asd5/rrs-index.asp

Cornaglia, F., Crivellaro, E., & McNally, S. (2015). Mental health and education decisions. *Labour Economics*. https://doi.org/10.1016/j.labeco.2015.01.005

Dahl, G. B., & Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *American Economic Review*. https://doi.org/10.1257/aer.102.5.1927

Daly, M., Sutin, A., & Robinson, E. (2020). Longitudinal changes in mental health and the COVID-19 pandemic: Evidence from the UK Household Longitudinal Study. *Psychological Medicine*. https://doi.org/10.2331/suisan.35.791

Del Boca, D., & Flinn, C. J. (1994). Expenditure decisions of divorced mothers and income composition. *Journal of Human Resources*, *29*(3), 742–761. https://doi.org/10.2307/146251

Del Boca, Daniela, & Ribero, R. (2001). The effect of child-support policies on visitations and transfers. *American Economic Review*, *91*(2), 130–134. https://doi.org/10.1257/aer.91.2.130

Department for Work and Pensions. (2012). *Income Support Lone Parent Regime: Official Statistics*.

Department for Work and Pensions. (2019). *Estimates of the separated family population statistics*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/796120/separated-families-population-statistics-2014-to-2015-2015-to-2016-2016-to-2017.pdf

Dolton, P., & O'Neill, D. (1996). Unemployment duration and the Restart effect: Some experimental evidence. *Economic Journal*. https://doi.org/10.2307/2235254

Dustmann, C., & Fasani, F. (2016). The Effect of Local Area Crime on Mental Health. *Economic Journal*. https://doi.org/10.1111/ecoj.12205

DWP. (2017). *Child Maintenance Reforms - 30 Month Review of chargins*.

Ermisch, J. (2008). Child support and non-resident fathers' contact with their children. *Journal of Population Economics*. https://doi.org/10.1007/s00148-006-0125-4

Etheridge, B., & Spantig, L. (2020). *The gender gap in mental well-being during the Covid-19 outbreak: Evidence from the UK*.

Fancourt, D., Bu, F., Mak, H. W., Paul, E., & Steptoe, A. (2021). COVID-19 social study. In *Nuffield Foundation* (Issue June). https://www.nuffieldfoundation.org/project/covid-19-social-study

Feinstein, L. (2000). The Relative Economic Importance of Academic Psychological and Behavioural Attributes Developed in Childhood. *Centre for Economic Performance*.

Fetzer, T. (2020). Subsidizing the Spread of COVID19: Evidence from the Uk's Eat-Out-To-Help-Out Scheme. In *CAGE Working Paper* (Issue 517).

Francesconi, M., & Van der Klaauw, W. (2007). The socioeconomic consequences of "in-work" benefit reform for British lone mothers. *Journal of Human Resources*, *42*(1), 1–31. https://doi.org/10.3368/jhr.xlii.1.1

Garaud, P. (2014). *Lone Parent Obligations: an impact assessment*.

García-Gómez, P., Jones, A. M., & Rice, N. (2010). Health effects on labour market exits and entries. *Labour Economics*. https://doi.org/10.1016/j.labeco.2009.04.004

Gardner, J., & Oswald, A. J. (2007). Money and mental wellbeing: A longitudinal study of medium-sized lottery wins. *Journal of Health Economics*.

https://doi.org/10.1016/j.jhealeco.2006.08.004

Goldberg, D. P., Gater, R., Sartorius, N., Ustun, T. B., Piccinelli, M., Gureje, O., & Rutter, C. (1997). The validity of two versions of the GHQ in the WHO study of mental illness in general health care. *Psychological Medicine*. https://doi.org/10.1017/S0033291796004242

Goldberg, David P, & Williams, P. (1988). *A user's guide to the General Health Questionnaire*. NFER-Nelson Publishing Company Limited.

González-Sanguino, C., Ausín, B., Castellanos, M. A., Saiz, J., & Muñoz, M. (2021). Mental health consequences of the Covid-19 outbreak in Spain. A longitudinal study of the alarm situation and return to the new normality. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*. https://doi.org/10.1016/j.pnpbp.2020.110219

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *38*(5), 581–586. https://doi.org/10.1111/j.1469-7610.1997.tb01545.x

Goulden, C. (2020). *UK Poverty 2019/2020*.

Graetz, B. (1991). Multidimensional properties of the General Health Questionnaire. *Social Psychiatry and Psychiatric Epidemiology*. https://doi.org/10.1007/BF00782952

Graham, J., Beller, A., & Hernandez, P. M. (1994). The Effects of Child Support on Educational Attainment. In I. Garfinkel, S. McLanahan, & P. Robins (Eds.), *Child Support and Child Well-being* (pp. 317–354). Urban Institute Press.

Graham, J. W., & Beller, A. H. (1996). Child Support in Black and White: Racial Differentials in the Award and Receipt of Child Support during the 1980s. *Social Science Quarterly*.

Gregg, P., & Harkness, S. (2003). Welfare Reform and Lone Parents Employment in the UK. *Department of Economics, University of Bristol, UK, Leverhulme Centre for Market and Public Organisation, 2003, 25 Pp.*

Haveman, R., Wolfe, B., Haveman, R., & Wolfe, B. (1995). The Determinants of Children's Attainments: A Review of Methods and Findings. *Journal of Economic Literature*.

Heckman, J. J., Lalonde, R. J., & Smith, J. A. (1999). Chapter 31 The economics and econometrics of active labor market programs. In *Handbook of Labor Economics*. https://doi.org/10.1016/S1573-4463(99)03012-6

Jacobson, N. C., Lekkas, D., Price, G., Heinz, M. V., Song, M., James O'Malley, A., & Barr, P. J. (2020). Flattening the mental health curve: COVID-19 stay-at-home orders are associated with alterations in mental health search behavior in the United States. *JMIR Mental Health*, *7*(6). https://doi.org/10.2196/19347

Klotzbücher, V., & Armbruster, S. (2020). Lost in lockdown? COVID-19, social distancing, and mental health in Germany. In *Discussion Paper Series* (No. 2020–04). https://doi.org/10.4324/9780367854973-4

Knapp, M., King, D., Healey, A., & Thomas, C. (2011). Economic outcomes in adulthood and their associations with antisocial conduct, attention deficit and anxiety problems in childhood. *Journal of Mental Health Policy and Economics*.

Knipe, D., Evans, H., Marchant, A., Gunnell, D., & John, A. (2020). Mapping population mental health concerns related to COVID-19 and the consequences of physical distancing: A Google trends analysis. *Wellcome Open Research*, *5*, 1–17. https://doi.org/10.12688/wellcomeopenres.15870.2

Knox, V. W. (1996). The effects of child support payments on developmental outcomes for elementary school-age children. *Journal of Human Resources*. https://doi.org/10.2307/146148

Knox, V. W., & Bane, M. J. (1994). Child Support and Schooling. In I. Garfinkel, S. McLanahan, & P. Robins (Eds.), *Child Support and Child Well-being* (pp. 285–316). Urban Institute Press.

Kooreman, P. (2000). The labeling effect of a child benefit system. *American Economic Review*. https://doi.org/10.1257/aer.90.3.571

Krekel, C., Swanke, S., Neve, J. De, Fancourt, D., Powdthavee, N., Prati, A., Wratil, C., & Pinchuk, K. (2020). *Are Happier People More Compliant ? Global Evidence From Three Large-Scale Surveys During Covid-19 Lockdowns*.

Lagomarsino, E., & Spiganti, A. (2020). No gain in pain: psychological well-being, participation, and wages in the BHPS. *European Journal of Health Economics*. https://doi.org/10.1007/s10198-020-01234-4

Manning, A. (2009). You can’t always get what you want: The impact of the UK Jobseeker’s Allowance. *Labour Economics*. https://doi.org/10.1016/j.labeco.2008.09.005

Martin, J. P., & Grubb, D. (2005). What Works and for Whom: A Review of OECD Countries’ Experiences with Active Labour Market Policies. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.348621

McKay, S., Smith, A., Youngs, R., & Walker, R. (1999). Unemployment and Jobseeking after the introduction of Jobseeker’s Allowance. *Department of Social Security, Research Report No. 99*.

Meyer, D. R., Cancian, M., & Waring, M. K. (2020). Use of child support enforcement actions and their relationship to payments. *Children and Youth Services Review*. https://doi.org/10.1016/j.childyouth.2019.104672

Mosca, I., O’Sullivan, V., & Wright, R. E. (2021). The educational attainment of the children of stay-at-home mothers: evidence from the Irish Marriage Bar. *Oxford Economic Papers*, *73*(2), 534–560. https://doi.org/10.1093/oep/gpaa031

Nepomnyaschy, L., Magnuson, K. A., & Berger, L. M. (2012). Child support and young Children’s Development. *Social Service Review*. https://doi.org/10.1086/665668

Newby, J. M., O’Moore, K., Tang, S., Christensen, H., & Faasse, K. (2020). Acute mental health responses during the COVID-19 pandemic in Australia. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0236562

Nielsen, R., & Oakley, M. (2011). *The impact of increased conditionality for out-of-work lone parents: Evidence from the UK Labour Force Survey*.

O'Connor, R. C., Wetherall, K., Cleare, S., McClelland, H., Melson, A. J., Niedzwiedz, C. L., O'Carroll, R. E., O'Connor, D. B., Platt, S., Scowcroft, E., Watson, B., Zortea, T., Ferguson, E., & Robb, K. A. (2021). Mental health and well-being during the COVID-19 pandemic: Longitudinal analyses of adults in the UK COVID-19 Mental Health & Wellbeing study. *British Journal of Psychiatry*, *218*(6), 326–333. https://doi.org/10.1192/bjp.2020.212

Office for National Statistics. (2012). *Working and workless households in the UK: 2012*. 29 August 2012. https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/bulletins/workingandworklesshouseholds/2012-08-29

Office for National Statistics. (2021). *Coronavirus and the social impacts on Great Britain: 27 August 2021*. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/bulletins/coronavirusandthesocialimpactsongreatbritain/29january2021

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business and Economic Statistics*, *37*(2), 187–204. https://doi.org/10.1080/07350015.2016.1227711

Petrongolo, B. (2009). The long-term effects of job search requirements: Evidence from the UK JSA reform. *Journal of Public Economics*. https://doi.org/10.1016/j.jpubeco.2009.09.001

Pierce, M., Hope, H., Ford, T., Hatch, S., Hotopf, M., John, A., Kontopantelis, E., Webb, R., Wessely, S., McManus, S., & Abel, K. M. (2020). Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. *The Lancet Psychiatry*, *7*(10), 883–892. https://doi.org/10.1016/S2215-0366(20)30308-4

Plug, E., & Vijverberg, W. (2003). Schooling, Family background, and adoption: Is it nature or is it nurture? *Journal of Political Economy*. https://doi.org/10.1086/374185

Ramiz, L., Contrand, B., Rojas Castro, M. Y., Dupuy, M., Lu, L., Sztal-Kutas, C., & Lagarde, E. (2021). A longitudinal study of mental health before and during COVID-19 lockdown in the French population. *Globalization and Health*. https://doi.org/10.1186/s12992-021-00682-8

Rayner, E., Shah, S., White, R., Dawes, L., & Tinsley, K. (2000). Evaluating Jobseeker's Allowance: A Summary of the Research Findings. *Department of Social Security, Research Report No. 116*.

Roff, J. (2010). Welfare, child support, and strategic behavior. Do high orders and low disregards discourage child support awards. *Journal of Human Resources*. https://doi.org/10.3368/jhr.45.1.59

Roff, J., & Lugo-Gil, J. (2012). A model of child support and the underground economy. *Labour Economics*. https://doi.org/10.1016/j.labeco.2012.03.006

Rossin-Slater, M., & Wüst, M. (2018). Parental responses to child support obligations: Evidence from administrative data. *Journal of Public Economics*. https://doi.org/10.1016/j.jpubeco.2018.06.003

Rutherford, T. (2013). *Historical Rates of Social Security Benefits*.

Schmidtke, J., Hetschko, C., Schöb, R., Stephan, G., Eid, M., & Lawes, M. (2021). The Effects of the COVID-19 Pandemic on the Mental Health and Subjective Well-Being of Workers: An Event Study Based on High-Frequency Panel Data. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3905073

Shea, J. (2000). Does parents' money matter? *Journal of Public Economics, 77*(2), 155–184. https://doi.org/10.1016/S0047-2727(99)00087-0

Skinner, C. (2012). Child Maintenance in the United Kingdom. *European Journal of Social Security*, *14*(4).

Skinner, C., Hakovirta, M., & Davidson, J. (2012). A Comparative Analysis of Child Maintenance Schemes in Five Countries. *European Journal of Social Security*. https://doi.org/10.1177/138826271201400407

Smith, A., Youngs, R., Ashworth, K., McKay, S., Walker, R., Elias, P., & McKnight, A. (2000). Understanding the Impact of Jobseeker's Allowance. *Department of Social Security, Research Report No. 111*.

Soobedar, Z. (2009). *Labour Supply Disincentives of Income Support: An Analysis of Single Mothers with No Qualifications in the UK* (Working Paper No. 656).

Stefan Foa, R., Gilbert, S., & Otto Fabian, M. (2020). COVID-19 and Subjective Well-Being: Separating the Effects of Lockdowns from the Pandemic. *The Lancet Psychiatry*. https://ssrn.com/abstract=3674080

Tannenbaum, D. I. (2020). The effect of child support on selection into marriage and fertility. *Journal of Labor Economics*. https://doi.org/10.1086/705928

Trickey, H., Kellard, K., Walker, R., Ashworth, K., & Smith, A. (1998). Unemployment and Jobseeking: Two Years On. In *Department of Social Security, Research Report No. 87*.

van der Velden, P. G., Hyland, P., Contino, C., von Gaudecker, H. M., Muffels, R., & Das, M. (2021). Anxiety and depression symptoms, the recovery from symptoms, and loneliness before and after the COVID-19 outbreak among the general population: Findings from a Dutch population-based longitudinal study. *PloS One*. https://doi.org/10.1371/journal.pone.0245057

Walker, I., & Zhu, Y. (2011a). Child Support and Educational Outcomes: Evidence from the British Household Panel Survey. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1414930

Walker, I., & Zhu, Y. (2011b). Do Dads Matter? Or Is It Just Their Money that Matters? Unpicking the Effects of Separation on Educational Outcomes. In *Household Economic Behaviors*. https://doi.org/10.1007/978-1-4419-9431-8_7

Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., McIntyre, R. S., Choo, F. N., Tran, B., Ho, R., Sharma, V. K., & Ho, C. (2020). A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. *Brain, Behavior, and Immunity*. https://doi.org/10.1016/j.bbi.2020.04.028

Wasserman, M. (2020). The disparate effects of family structure. *Future of Children*. https://doi.org/10.1353/foc.2020.0008

Weiss, Y., & Willis, R. J. (1985). Children as Collective Goods and Divorce Settlements. *Journal of Labor Economics*, *3*(3), 268–292. https://doi.org/10.1086/298056

Wildman, J. (2003). Income related inequalities in mental health in Great Britain: Analysing the causes of health inequality over time. *Journal of Health Economics*. https://doi.org/10.1016/S0167-6296(02)00101-7